

**Title: Low alpha-defensin gene copy number increases the risk for IgA nephropathy development and poor renal survival**

**Authors:** Zhen Ai<sup>1,2,†</sup>, Ming Li<sup>1,2,†</sup>, Wenting Liu<sup>1,2†</sup>, Jia-Nee Foo<sup>3</sup>, Omniah Mansouri<sup>4</sup>, Peiran Yin<sup>1,2</sup>, Qian Zhou<sup>1,2</sup>, Xueqing Tang<sup>1,2</sup>, Xiuqing Dong<sup>1,2</sup>, Shaozhen Feng<sup>1,2</sup>, Ricong Xu<sup>1,2</sup>, ZhongZhong<sup>1,2</sup>, Jian Chen<sup>5</sup>, Jianxin Wan<sup>6</sup>, Tanqi Lou<sup>7</sup>, Jianwen Yu<sup>1,2</sup>, Qin Zhou<sup>1,2</sup>, Jinjin Fan<sup>1,2</sup>, Haiping Mao<sup>1,2</sup>, Daniel Gale<sup>8</sup>, Jonathan Barratt<sup>9</sup>, John AL Armour<sup>4</sup>, Jianjun Liu<sup>3,10,11‡</sup> and Xueqing Yu<sup>1,2,\*</sup>

**Affiliations:**

<sup>1</sup>Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510080, China.

<sup>2</sup>Key Laboratory of Nephrology, Ministry of Health and Guangdong Province, Guangzhou, Guangdong 510080, China.

<sup>3</sup>Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore.

<sup>4</sup>School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK.

<sup>5</sup>Department of Nephrology, Fuzhou General Hospital of Nanjing Military Command, Fuzhou, Fujian 350025, China.

<sup>6</sup>Department of Nephrology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian 350005, China.

<sup>7</sup>Department of Nephrology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510630, China.

<sup>8</sup> University College London Centre for Nephrology, Royal Free Hospital, London NW3 2PF, UK.

<sup>9</sup> Department of Infection, Immunity and Inflammation, University of Leicester, Leicester LE1 9HN, U.K.

<sup>10</sup> School of Biological Sciences, Anhui Medical University, Hefei, Anhui 230032, China.

<sup>11</sup> Saw Swee Hock School of Public Health, National University of Singapore, Singapore 119077, Singapore.

\*Corresponding author: yuxq@mail.sysu.edu.cn.

†These authors contributed equally.

‡Co-corresponding author: liuj3@gis.a-star.edu.sg.

**One Sentence Summary: Low copy numbers of *DEFA1A3* CNVs increases the risk for IgA nephropathy development and poor renal survival.**

**Abstract:** IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide whose etiology is poorly understood. Although being a major source of genetic variation in the human genome, copy number variations (CNVs), particularly the complex multi-allelic ones, are poorly studied and understood in terms of their involvement in disease development. Here we performed a comprehensive association analysis of *DEFA1A3* CNV locus on 8p23.1 in two independent IgA nephropathy (IgAN) cohorts of Southern Chinese Han, consisting of a total of 1189 cases and 1187 healthy controls. We discovered three independent associations within the locus: the total copy number of *DEFA1A3* ( $P=3.99\times 10^{-9}$ , OR=0.88 per copy) as well as the copy

numbers of *DEFA3* ( $P=6.55\times 10^{-5}$ , OR=0.82) and a noncoding deletion variant (*211bp*) ( $P=3.50\times 10^{-16}$ , OR=0.75) (fixed-effects meta-analysis). While showing very strong association with increased risk for IgAN ( $P=9.56\times 10^{-20}$ ), the low total copy numbers of the three variants (Genetic Score) also showed significant association with poor long-term renal survival of IgAN ( $P=0.03$ , HR=3.69, after controlling the effects of known prognostic factors) and high serum level of galactose-deficient IgA1 ( $P=0.03$ ). As a replication, we confirmed the associations of *DEFA1A3* ( $P=4.42\times 10^{-4}$ , OR=0.82) and *DEFA3* copy numbers ( $P=4.30\times 10^{-3}$ , OR=0.74) with IgAN in a Caucasian cohort (531 cases and 198 controls). Interestingly, we also found the association of *211bp* copy number with membranous nephropathy ( $P=1.11\times 10^{-7}$ , OR=0.74 in 493 Chinese cases and 500 matched controls), but no associations were found with diabetic kidney disease (in 806 Chinese cases and 786 matched controls). By explaining 4.96% of risk variance, the *DEFA1A3* CNV locus is the strongest genetic susceptibility locus for IgAN identified so far and also influences the renal survival of IgAN, supporting the possibility of *DEFA1A3* being a novel therapeutic target and prognostic biomarker.

**[Main Text:]**

## **Introduction**

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide (1-3). The clinicopathologic patterns and the prognosis of IgAN show extremely wide variability, ranging from asymptomatic to rapidly progressive forms (4). The pathogenesis of IgAN is not well understood, but it's generally considered to be a complex and multifactorial disorder with both genetic and environmental factors contributing to its development (5). Thus far, there have been four genome-wide association studies (GWAS) for IgAN, and multiple susceptibility loci have

been identified (6-9). However, these GWAS loci can only explain 5% to 7.6% of total disease risk (9), and none of them has effect on disease prognosis.

Copy number variations (CNVs) have been shown to be a major source of genetic variation in human genome (10, 11) and have significant phenotypic impact by altering gene dosage, disrupting coding sequences or perturbing long-range gene regulation (12). Although CNVs have been suggested to play important roles in disease development (11, 13, 14), the associations of CNVs, particularly the role complex multi-allelic CNVs in complex diseases are poorly studied and understood, mainly due to experimental difficulties in measuring the copy number of such variants accurately (15). The alpha-defensin locus (*DEFA1A3*) is one of the complex multi-allelic CNVs, presented as tandem repeats of a 19kb sequence unit, each of which contains one copy of *DEFA1* or *DEFA3* (alpha-defensin 1 or 3 that differ by a single non-synonymous coding variant within the 3<sup>rd</sup> exon) as well as several internal bi-allelic polymorphisms (16-18). The protein products of *DEFA1A3*, HNP1-3, are microbicidal peptides that are important components of innate immunity (20, 22-24). HNP1-3 can also act as immune modulators that chemoattract native T cells, immature dendritic cells and monocytes and induce the release of chemokines and cytokines (25). HNP1-3 is the most abundant of neutrophil granule proteins, and there is growing evidence to suggest that neutrophils may play important role in mediating glomerular injury in IgAN (26, 27). Alpha-defensins have also been suggested to play a role in the regulation of the complement system and pro-inflammatory cytokine production both of which have been shown to play important pathogenetic role in IgAN (19, 28, 29). Interestingly, the *DEFA1A3* CNV locus is located within the region of the new IgAN susceptibility locus on 8p23.1 that was first discovered in the Chinese population and later validated by independent studies (5, 6, 8), which,

together with its biological functions, suggests that the *DEFA1A3* CNV locus may be involved in IgAN development.

To evaluate the role of *DEFA1A3* CNV in IgAN, we performed a comprehensive association analysis of the *DEFA1A3* CNV locus in two independent IgAN cohorts of Southern Chinese Han, consisting of a total of 1189 cases and 1187 healthy controls. We discovered three strong independent associations within the *DEFA1A3* CNV locus and further demonstrated that the *DEFA1A3* CNVs could also influence the prognosis of IgAN. As a replication, we further confirmed the associations in Caucasian IgAN samples. Interestingly, we found the association with membranous nephropathy (MN), but not diabetic kidney disease (DKD) in Chinese population.

## **Results**

### **Distributions of *DEFA1A3* CNVs in Chinese population**

Two independent cohorts of IgAN samples were analyzed. The first cohort consisted of 197 cases and 199 matched healthy controls in terms of age, gender and geographic origin (Southern Chinese Han), and the second cohort included 992 cases and 988 healthy controls that were only matched for geographic origin (Southern Chinese Han) (Table S1). We measured the copy numbers (CNs) of *DEFA1*, *DEFA3* and *DEFA1A3* (*DEFA1+DEFA3*) as well as the CNs of three bi-allelic polymorphisms, 5bp indel (129bp vs. 124bp), 7bp duplication (282bp vs. 275bp) and the 4bp deletion (211bp vs. 215bp) using well-established paralogue ratio test (PRT)-based methods (17, 30) (Fig.S1). The PRT-based methods have been shown to have much greater accuracy for measuring the CNs of multi-allelic CNVs, such as beta-defensins, than real-time PCR

analysis(31, 32) and have been successfully applied for the studies of chemokines (*CCL3L1/CCL4L1*) (33), immunoglobulin-receptor (*FCGR3*)(34), and beta-defensins (*DEFB*)(31, 35).

The total CN of *DEFA1A3* ranges from 2 to 16 copies with a median CN of 6 per genome in our Chinese samples (Fig. S2). The range and median CN are 0-15 and 5 copies for *DEFA1*, 0-5 and 1 copy for *DEFA3*, 0-6 and 1 copy for *129bp*, 0-14 and 5 copies for *124bp*, 0-7 and 2 copies for *282bp*, 1-14 and 5 copies for *275bp*, 0-7 and 1 copy for *211bp*, and 0-15 and 5 copies for *215bp*. Our results are consistent with the reported median CNs of 6 in the HapMap CHB samples and 7 in CEU [US] and ECACC [UK] samples(16, 17), but lower than the reported median CN of 9 in the Japanese population (the JPT samples of the HapMap project)(19). In addition, the total *DEFA1A3*CNs of the 24HapMap CHB samples measured in this study showed a near-perfect correlation with the sequencing read depth of these samples from the 1000 Genome Project ( $R^2=0.98$ ,  $P<0.0001$ )(Fig.S3), confirming the high accuracy of our CN measurements.

Correlation analysis revealed that the CNs of *DEFA1*, *124bp*, *275bp* and *215bp* are correlated with each other ( $P<0.0001$ ,  $r>0.65$ ) (TableS2).

### **Association analysis of the *DEFA1A3* CNVs in Chinese IgAN samples**

We tested the associations of the *DEFA1A3*CNVs with IgAN susceptibility using logistic regression analysis and by treating the two cohorts (see above) as independent samples through a fixed-effects meta-analysis. As demonstrated in our previous studies (36), although sub-population structure exists between Northern and Southern Chinese Han populations, the samples from Southern China are genetically homogenous. No control for population stratification was included in our association analysis. We found highly significant associations

for the CNs of *DEFA1A3* ( $P=3.99\times 10^{-9}$ , OR=0.88), *DEFA1* ( $P=6.71\times 10^{-5}$ , OR=0.92), *DEFA3* ( $P=6.55\times 10^{-8}$ , OR=0.82), *124bp* ( $P=2.89\times 10^{-8}$ , OR=0.87), *275bp* ( $P=2.51\times 10^{-7}$ , OR=0.89) and *211bp* ( $P=3.50\times 10^{-16}$ , OR=0.75) (OR per copy). Suggestive association was also observed for the CNs of *129bp* ( $P=1.37\times 10^{-2}$ , OR=0.89). These variants showed consistent association between the two independent cohorts without any evidence of genetic heterogeneity, and all results have been adjusted for age and gender (Table S3). We performed step-wise conditional analysis (Table 1) and found that the associations of the *211bp*, *DEFA1A3* and *DEFA3* CNs are independent. A multivariate analysis has also confirmed the independence of these three associations, *211bp* ( $P=1.71\times 10^{-9}$ , OR=0.78), *DEFA1A3* ( $P=2.20\times 10^{-3}$ , OR=0.93) and *DEFA3* ( $P=3.33\times 10^{-3}$ , OR=0.85). All the three associations remain statistically significant after Bonferroni correction ( $0.05/9=5.56\times 10^{-3}$ ). No evidence was found for interaction among the three independent CNVs (data not shown).

The patients carried significantly lower CNs of all the three CNVs than the controls, indicating that the low copy numbers of these variants have risk effect on IgAN. On average, a patient carries 1 copy of *211bp*, 6 copies of *DEFA1A3* and 1 copy of *DEFA3*, which are all significantly lower than the median CNs in the healthy controls, 2 copies of *211bp* ( $P=3.36\times 10^{-16}$ ), 7 copies of *DEFA1A3* ( $P=4.74\times 10^{-8}$ ) and 1 copy of *DEFA3* ( $P=2.56\times 10^{-4}$ ) (Table S4; Fig.1). We found that for each of the three CNVs, the risk is reduced progressively with the increase of CN (Fig.2). For *211bp*, individuals carrying zero copies (about 26.8% of the population) have a 5.9-fold higher risk than those carrying five or more copies (about 1.9% of the population) ( $P=2.17\times 10^{-6}$ ). For *DEFA1A3*, individuals carrying four or fewer copies (about 7.6% of the population) have a 3.8-fold higher risk than those carrying thirteen or more copies (about 2.0% of the

population)( $P=2.13\times 10^{-4}$ ). For *DEFA3*, individuals carrying zero copies (about 12.5% of the population) have a 3.2-fold higher risk than those carrying four or more copies (about 1.5% of the population) ( $P=3.57\times 10^{-3}$ ). The multi-variant genetic score (GS) (the standardized weighted sum of the CNs) of the three variants (ranging from 1.35 to 19.67) (Fig.S4) also showed very strong association with IgAN susceptibility ( $P=9.56\times 10^{-20}$ ). These three variants of the *DEFAIA3* CNV locus can explain up to 4.96% of phenotypic variance cumulatively.

### **Analysis of clinical phenotypes and renal survival in Chinese IgAN patients**

All the 1189 IgAN patients analyzed in this study have information for clinical and pathological features that are summarized in Table S5. In addition, of the 1189 IgAN patients, 382 patients had follow-up information with a median follow-up time of 4.5 years (interquartile range, 3.8-5.4 years). Of the 382 patients, 27 patients progressed to end stage renal disease (ESRD) (eGFR<15ml/min/1.73m<sup>2</sup>, or dialysis, or renal transplantation), and 9 patients were found to have doubling of serum creatinine after diagnosis.

Interestingly, the survival analysis of renal outcomes (ESRD or doubling of serum creatinine after diagnosis) using Cox regression model in the 382 IgAN patients with follow-up data revealed a significant association of the multi-variant GS (ranging from 1.80 to 17.14) of the *DEFAIA3* CNVs with the renal survival of IgAN ( $P=0.01$ , HR=0.79) after adjustment for known prognostic factors, including age, gender, proteinuria, hypertension and serum creatinine at biopsy(3, 37) (Table 2). We further divided all the 382 patients into five groups based on the quintile of their GS values and found that the patients in the 1<sup>st</sup> quintile of GS (average GS=3.06) showed significant worse renal survival than the patients of the 5<sup>th</sup> quintile (average



GS=9.57) ( $P=0.03$ , HR=3.69, after adjustment for known prognostic factors) (Fig. 3; Table 2). Given that the GS is intrinsically correlated with the copy numbers of *DEFA1A3* CNVs, our results have demonstrated that patients carrying low copies of *DEFA1A3* CNVs have increased risk for poor renal outcomes.

The analysis of clinical phenotypes using the multi-variant GS revealed the association of the *DEFA1A3* CNV locus with hyperuricemia ( $P=0.01$ ) and thickening of arterial wall ( $P=0.008$ ). However, the evidence could not survive the conservative Bonferroni correction for the multiple testing of 15 clinical phenotypes (although they are not totally independent) (Table S6), and the finding needs to be confirmed by further study.

### ***DEFA1A3* CNVs as the driver of the GWAS locus on 8p23.1**

We genotyped *rs2738048*, the previously reported susceptible SNP on 8p23.1 by GWAS (6) in the same 1189 cases and 1187 controls and performed association analysis using the same method above. We confirmed the association of *rs2738048* ( $P=9.58 \times 10^{-4}$ , OR=0.80) in our samples (Table 1), but the association was weaker than those of *211bp*, *DEFA1A3* and *DEFA3*. Furthermore, conditioning on the association effect of either *211bp* or *DEFA1A3* can fully abolish the association at *rs2738048*; whereas all the associations of the *DEFA1A3* CNVs remained highly significant after conditioning the association effect at *rs2738048* (Table S7). Consistently, the genotypes of *rs2738048* showed strong correlation with the CNs of *211bp* and *DEFA1A3* (Table S2; Fig.S5). These results clearly demonstrate that the previously reported association at *rs2738048* is due to the association effect of the *DEFA1A3* CNV locus.

### **Replication study in Caucasian IgAN and Chinese non-IgAN renal disease samples**

As a replication, we investigated the associations of *211bp*, *DEFA1A3* and *DEFA3* copy numbers in a Caucasian IgAN cohort of 531 cases and 198 controls. Both *DEFA1A3* ( $P = 4.42 \times 10^{-4}$ ; OR, 0.82) and *DEFA3* ( $P = 4.30 \times 10^{-3}$ ; OR, 0.74) showed significant associations in the Caucasian cohort with consistent genetic effects as in the Chinese cohort (Table 3). However, *211bp* was much rarer in the Caucasian population, and consequently, the association of *211bp* did not reach statistical significance ( $P = 7.56 \times 10^{-2}$ ), although its genetic effect was consistent between Caucasian (OR, 0.79) and Chinese samples (OR, 0.75) (Table 3 and fig. S6).

To evaluate whether the association effects of *DEFA1A3* CNVs (*211bp*, *DEFA1A3* and *DEFA3*) are unique to IgAN, we further investigated the associations in two Chinese cohorts of non-IgAN renal disease, diabetic kidney disease (DKD) (806 cases and 786 age and gender-matched controls) and membranous nephropathy (MN) (493 cases and 500 age and gender-matched controls). Interestingly, we detected the significant association of the copy number of *211bp* variant with MN ( $P=1.11 \times 10^{-7}$ , OR=0.74 per copy), but did not observe any association of *DEFA1A3* CNVs with DKD (Table 3).

### **Association analysis of *DEFA1A3* CNVs with HNP1-3 and IgA1 protein expression levels**

To investigate the effect of *DEFA1A3* CNVs on HNP1-3 expression, we first measured the copy numbers of the *DEFA1A3* CNVs (*211bp*, *DEFA1A3* and *DEFA3*) as well as the serum and urine protein (HNP1-3) levels of *DEFA1A3* in 96 IgAN patients and 62 healthy controls. There were no differences in the copy number distributions of *DEFA1A3* CNVs between the e patients and controls (Wilcoxon test  $P>0.05$ ). In both serum and urine, the protein levels of HNP1-3 were

significantly higher in the patients than the healthy controls ( $P < 0.001$ ) (Fig.4), which is consistent with the fact that mature HNP1-3 are largely produced by activated neutrophils as inflammation response (20). However, we could not find any association between *DEFA1A3* CNVs and protein levels of either serum HNP1-3 or urine HNP1-3 (Table S8). In addition, we also performed an in vitro protein analysis of HNP1-3 in neutrophils isolated from 83 IgAN patients and 79 health controls after stimulating neutrophils by lipopolysaccharide (LPS). While there are no differences in *DEFA1A3* copy number distributions between these patients and controls (Wilcoxon test  $P > 0.05$ ), the extracellular HNP1-3 levels were significantly lower in neutrophils isolated from the patients than the ones from controls ( $P = 0.006$ ), after a stimulation by LPS (100ng/ml for 6 hours) (Fig.4). However, we did not detect the association between *DEFA1A3* CNVs and extracellular HNP1-3 level released by neutrophils after stimulation of LPS.

Since the formation of the galactose-deficient IgA1-containing immune complexes is the key pathogenic factor contributing to the development of IgAN(38), we also investigated the effects of *DEFA1A3* CNVs on the expression levels of serum IgA1 and galactose-deficient IgA1 in 96 IgAN patients and 62 controls. As expected, serum levels of IgA1 and galactose-deficient IgA1 were both significantly higher in the IgAN patients than the health controls ( $P < 0.01$ ) (Fig.S7). Although there were no significant evidences for individual *DEFA1A3* CNVs, the multi-variant GS of these variants did show significant negative association with the serum levels of both IgA1 ( $P = 0.02$ ) and galactose-deficient IgA1 ( $P = 0.03$ ) (Table 4).

## **Discussion**

Our current study has discovered three independent risk factors for IgAN within the *DEFA1A3* CNV locus, the CNs of *DEFA1A3*, *DEFA3* and *211bp* by analyzing two independent Chinese IgAN samples and further confirming the associations in a Caucasian IgAN cohort. Although each extra copy of these CNVs is only associated with moderate reduction of IgAN risk (by 12 to 25%), the overall impact of the whole *DEFA1A3* CNV locus is very strong, because each person can carry multiple copies of these variants. By explaining 4.96 % of risk variance (much higher than any previously reported SNPs), the *DEFA1A3* CNV locus is the strongest genetic susceptibility factor that has been discovered for IgAN so far. Interestingly, our study has also revealed that the low copy number of *DEFA1A3* CNV is associated with poor renal survival of IgAN as well as higher expression of galactose-deficient IgA1, a known pathogenic risk factor for IgAN. Consistently, our analysis of HNP1-3 expressions has also demonstrated that HNP1-3 expression will be elevated as inflammation response as seen in the serum and urine samples of patients, and study has demonstrated that the low copy number of the *DEFA1A3* CNV is associated with the neutrophils from health individuals produce more HNP1-3 than IgAN patients when they are stimulated by LPS. Taken together, our genetic and expression analysis results have suggested that HNP1-3 provide protective effect against IgAN development and the progression of renal dysfunction.

Our study has also revealed a strong association of the *DEFA1A3* CNV (211bp) with membranous nephropathy, but not diabetic kidney disease, although our diabetic kidney disease cohort had a bigger sample size and thus more power than did the membranous nephropathy cohort to detect the association. The genetic effect (OR) of 211bp was consistent between membranous nephropathy (OR, 0.74) and IgAN (OR, 0.75) without evidence of heterogeneity ( $I^2$

= 0; Q test = 0.84), suggesting the shared protective effect of 211bp variant against both membranous nephropathy and IgAN. Both IgAN and membranous nephropathy are common causes of primary glomerulonephritis, and both are mediated by immune complexes. In contrast, diabetic kidney disease is a chronic microvascular complication of diabetes due to metabolic dysregulation (35). Although the precise immunopathogenesis of IgAN and membranous nephropathy remains unclear, HLA-DQA1 has been shown to be associated with both IgAN and membranous nephropathy (6, 36). Together, these results suggest that there are shared pathogenic pathways between IgAN and membranous nephropathy and that the *DEFA1A3* CNV locus may have a broad impact on the development of primary glomerulonephritis.

It is interesting to see the independent association effect of *DEFA3* CN beyond the total *DEFA1A3* CN. While *DEFA1* has been in a multi-copy array for at least 25 million years, *DEFA3* is a more recent human-specific variant arose from *DEFA1* through a non-synonymous coding mutation in the 3<sup>rd</sup>exon(16). Previous studies have suggested that HNP3 is generally less active than HNP1, although HNP3 is expressed at about twice the level of HNP1 for each of copy of the gene (20, 45, 46). It is also intriguing that the 211bp CN shows an independent association effect beyond that of the *DEFA1A3* CN. The 211bp is a 4bp deletion variant (TATC) within the 2<sup>nd</sup>intron of *DEFA1A3* first investigated by us. Its molecular function is unclear, but analysis by Human Splicing Finder (47) suggests that this 4bp deletion may create a new splice acceptor site (CAG), raising the possibility that this 4bp deletion may create another novel splice isoform different from the mRNAs of *DEFA1* and *DEFA3*, which may in turn influence the activities of HNP1-3 protein.

Our large-scale protein analysis, however, failed to reveal a direct correlation between the DEFA1A3 copy numbers and HNP1–3 levels in serum and urine or the production of HNP1–3 by neutrophils upon LPS stimulation. Previous studies of DEFA1A3 expression and its correlation with its copy number were limited. The mRNA analysis in leukocytes and lymphoblastoid cell lines (16, 19) as well as plasma protein analysis in patients with severe sepsis did not show correlation (21), but a small study (eight subjects) showed the correlation of HNP1–3 amounts inside neutrophils with the copy numbers of DEFA1A3 (18). The lack of correlation between the *DEFA1A3* copy numbers and protein level in serum, urine and neutrophils is probably due to the fact that HNP1-3 are mainly produced by bone marrow precursors of neutrophils in promyelocytes and early myelocytes (20). In addition, as what have been demonstrated in other CNV loci (49, 50), a full understanding of the haplotype structure (allelic structure) of the whole *DEFA1A3* CNV locus may be necessary for demonstrating the correlation.

*DEFA1A3* CNVs could potentially influence IgAN development and renal progression in several ways. First, as *DEFA1A3* is primarily expressed in promyelocytes within bone marrow, it may have physiological effects on bone marrow microenvironment (39, 40). Second, recent studies of IgAN pathogenesis highlighted the important role of impaired immune regulation along the mucosa-bone marrow axis and suggested that galactose-deficient IgA1 seems to be overproduced in bone marrow(41-43). *DEFA1A3* CNVs may influence the impaired mucosa-bone marrow axis through its impact on the overproduction of galactose-deficient IgA1. Third, HNP1-3 plays a very important role in innate immunity, and neutrophils are the first defenders for infection.

*DEFA1A3* CNVs may influence the dysregulated mucosal immune response in IgAN by enhancing inflammatory signals(44).

Our study has also demonstrated that the *DEFA1A3* CNVs are the primary driver of the previously reported association within the GWAS locus on 8p23.1(6). The *DEFA1A3* CNVs showed stronger associations than the previously reported SNP *rs2738048*, and conditioning on the effect of the *DEFA1A3* CNVs can fully abolish the association at *rs2738048*, but not vice versa. So, the association at *rs2738048* is secondary to the association effect of the *DEFA1A3* CNVs. While this manuscript was under preparation, a new IgAN GWAS was published online, revealing another independent SNP *rs10086568* within the GWAS locus on 8p23.1 (9). It would be interesting to investigate whether the association of this SNP is also due to the effect of *DEFA1A3*CNVs in a future study.

Two major limitations of this study should be noted. First, our current findings are largely of genetic associations. The potential mechanisms underlying the involvement of the *DEFA1A3* in the development of IgAN are interesting and biologically plausible but do need to be elucidated by future functional investigations. Second, the current study did not investigate the haplotype structure of *DEFA1A3* CNVs because current genotyping and sequencing technologies could not determine the haplotype structure of multiallelic CNVs with long repeats in unrelated samples. Rapid development of long-read next-generation sequencing analysis may offer opportunities to characterize the haplotype structure of *DEFA1A3* CNVs and understand in detail how different *DEFA1A3* CNVs work together to influence the level and functionality of HNP proteins and susceptibility to developing IgAN.

In summary, our study is the first to show the association of the *DEFA1A3* CNV locus with IgAN risk and renal progression and one of the few cases where the association of a complex multi-allelic CNV has been demonstrated with robust evidence. Our study has clearly demonstrated that this multi-allelic CNV is the strongest genetic susceptibility locus for IgAN identified so far and may have a broad impact on the development of other inflammatory renal diseases. By demonstrating its protective effect against IgAN development and renal progression, our study has also suggested the possibility of HNP1-3 being a novel therapeutic target and prognostic biomarker.

## **Materials and Methods**

### **Study Design**

The objective of this study was to study the effect of the *DEFA1A3* CNV locus on IgAN development by investigating the association between the copy number of the *DEFA1A3* CNV locus with IgAN susceptibility and renal outcomes. First, two independent cohorts from Southern Chinese Han consisting of a total of 1189 IgAN cases and 1187 controls were employed for the association analysis. All IgAN cases were biopsy-proven. The copy numbers of *DEFA1A3* CNVs were measured using well-established paralogue ratio test (PRT)-based method. Second, as a replication, the three risk variants of *DEFA1A3* CNV locus discovered in Chinese IgAN samples were further analyzed in additional Caucasian IgAN (531 cases and 198 age-, gender-matched controls), Chinese DKD (806 cases and 786 age-, gender-matched controls) and Chinese MN



(493 cases and 500 age-, gender-matched controls) cohorts using the PRT-based method.

Clinicopathologic information and follow-up data were collected from Chinese IgAN patients to further assess the clinical impact of *DEFAIA3* CNVs on renal progression. We also investigated the relationship between the *DEFAIA3* CNV locus and the previously published GWAS locus (*rs2738048*) on 8q23.1 by genotyping the SNP using TaqMan SNP genotyping assay in the same two independent Chinese cohorts. Last, we investigated the effects of *DEFAIA3* CNVs on the expression of HNP1-3 and galactose-deficient IgA1. There was no randomization or blinding.

### **Study Subjects**

All IgAN cases were histopathologically diagnosed as primary IgAN by renal biopsy according to the following criteria: (i) immunofluorescence showing at least 2+ (scale 0 to 3+) mesangial deposition of IgA, with IgA comprising the dominant immunoglobulin deposited in the glomeruli, and (ii) excluding individuals with cirrhosis, Henoch-Schönleinpurpura nephritis, hepatitis B-associated glomerulonephritis, HIV infection and systemic lupus erythematosus.

Clinical and pathologic information were collected at diagnosis, including blood pressure, serum creatinine, albumin, cholesterol, triglyceride, uric acid, serum IgA level and proteinuria.

Pathological changes recorded included crescent, thickening of arterial wall and interstitial inflammation. In accordance to the Oxford Classification of IgAN, samples were graded by the four pathological features, mesangial hypercellularity (M), endocapillary hypercellularity (E), segmental glomerulosclerosis (S), and tubular atrophy/interstitial fibrosis (T), resulting in a MEST score. All DKD cases were diagnosed as persistent albuminuria (>30mg/g) and/or low estimated glomerular filtration rate (eGFR<60ml/min/1.73m<sup>2</sup>) in patients with diabetes (51). Diabetes was defined as fasting plasma glucose ≥7.0 mmol/L and/or 2-h postprandial plasma

glucose  $\geq 11.1$ mmol/L, by the use of hypoglycemic agents despite fasting plasma glucose, or any self-reported history of diabetes. All MN cases were adult biopsy-proven idiopathic MN. All the healthy controls were clinically verified with normal urinalysis (without red blood cells and protein in urine) and normal serum creatinine levels. Gender, age, geographical origin and ethnicity information were collected from both cases and controls through questionnaires. All subjects completed a written informed consent form, and the study protocol was approved by the Institutional Review Board at The First Affiliated Hospital of Sun Yat-sen University.

### **Copy Number Measurements**

The copy numbers of *DEFA1A3* CNVs were measured using published methods(17, 48) which use paralogue ratio tests(30) to determine both the total numbers of *DEFA1A3* genes, and the copy numbers of each of four gene variants; we have already validated the accuracy of these methods, including calibration against 1000 Genomes Project read-depth data(17, 48). We measured the total copy numbers of *DEFA1A3* (*DEFA1* + *DEFA3*) (by *MLT1A0* and *DEFA4PRT*) and the individual copy numbers of *DEFA1* and of *DEFA3* (by *DefHae3* ratio), as well as the allelic copy numbers of three internal bi-allelic polymorphisms; 5bp indel (129bp vs. 124bp, measured by *Indel5* ratio), 7bp duplication (282bp vs. 275bp, measured by *7bpdup* ratio) and 4bp deletion (211bp vs. 215bp, measured by *Del4* ratio) (Fig. S1). These internal variants were chosen for three reasons. First, as a part of PRT methodology, the measurements of allelic variant ratio of internal variants can help to determine the copy number more accurately(17). Second, previous studies have suggested that some internal variants, such as the coding variant *DEFA1* and *DEFA3*, may influence the expression and functionality of HNP proteins(16). Third, as reported, some variants within the *DEFA1A3* locus, particularly 5bp indel, show good

correlation with major haplotype classes of *DEFA1A3* locus at least in European population (48). In brief, we combined the information from multiplex PRTs (*MLT1A0* and *DEFA4* PRTs) and four allelic ratio measurements (*DefHae3/Indel5/7bpdup/Del4*) into a maximum-likelihood framework to assign each sample to an integer copy number for all the *DEFA1A3* CNVs with high confidence, as described(17).

These measurements were all well-established PRT-based methods as previously described(17)except for the *Del4* assay, which we developed as part of this work. The 4bp deletion (TATC) is a newly-characterized variant within the second intron of *DEFA1A3*. The PCR for the *Del4* assay was performed by using 1uM of the primers TGCTCTCATTTTTGCATTCC and NED-TTTCTCCAAAGACTTGATTCCAA and 27 cycles of 95°C for 30 seconds, 57°C for 30 seconds, and 68°C for 60 seconds, followed by a 70°C hold for 40 minutes to generate amplicon products of 211bp/215bp from 10ng genomic DNA input. The products were mixed with 10ul HiDi formamide with ROX-500 marker (Applied Biosystems) for fragment analysis using an ABI 3730xl DNA Analyzer. GeneMapper 4.1 (Applied Biosystems) was used to extract the peak areas of the separated PCR products. The ratio between the 211bp and 215bp products (*Del4* ratio) was estimated, and the copy numbers of the 211bp and 215bp alleles were calculated by using this *Del4* ratio and the total copy number of *DEFA1A3*.

For the PRT analysis of the total number of *DEFA1A3* CNV, seven CEPH/UTAH samples were included as reference samples of known copy number: *NA11931* (7 copies), *NA07347* (8 copies), *NA11930* (5 copies), *NA11993* (7 copies), *NA06993* (8 copies), *NA12248* (6 copies) and

*NA12249* (8 copies). The copy numbers for these reference samples had been determined unequivocally in previous work, including analysis of segregation and read depth (17, 48). The total *DEFA1A3* copy numbers among the cases and controls were determined by calibration of *MLT1A0/DEFA4* PRTs against these reference samples, and the copy numbers of the other variants (*DEFA1* vs. *DEFA3*, 129bp vs. 124bp, 282bp vs. 275bp, and 211bp vs. 215bp) were determined by using the total *DEFA1A3* copy numbers and the ratio between the CNs of two variants.

### **Genotyping Analysis of *rs2738048***

Genotyping analysis of *rs2738048* was performed by using the TaqMan SNP genotyping assay from ABI. TaqMan reactions were carried out in 5ul volumes containing 10-20ng DNA according to the manufacturer's protocols. Fluorescence data were obtained in the ABI PRISM7900HT, and genotypes were called using SDS2.4 software (Applied Biosystems). We examined the clustering patterns of genotypes and confirmed them to be of good quality.

### **Association Analysis of *DEFA1A3* CNVs and *rs2738048***

Frequency tables were employed to describe the distributions of the *DEFA1A3* CNVs, and then Wilcoxon rank sum tests and Chi-square tests were used to determine the significance of the distribution differences in *DEFA1A3* CNVs and *rs2738048* between the IgAN cases and controls. Spearman correlation analyses were used to test the correlations between the *DEFA1A3* CNVs and *rs2738048*.

Two independent cohorts (Table S1) were analyzed to evaluate the associations of *DEFA1A3* CNVs and SNP *rs2738048* with IgAN susceptibility. First, the association analysis was performed using PROC LOGISTIC in SAS 9.2 software (SAS Institute), where each of the *DEFA1A3* CNVs and *rs2738048* were included as univariate with adjustment by age and gender. Second, the results from the two cohorts were combined using fixed-effects meta-analysis in PLINK(52). The degree of heterogeneity was estimated using the heterogeneity index ( $I^2$ ) and Cochran's Q test in the combined analysis. Third, conditional analyses were performed using a stepwise logistic regression model to find independent associations. A multivariate analysis was then conducted to confirm the independent associations. Interaction analysis was performed by adding the product term of two CNVs into the logistic regression analysis. OR values were measured as OR per copy of *DEFA1A3* CNV.  $P < 0.05$  after Bonferroni correction for multiple testing was considered as statistically significant.

To assess the cumulative effects of *DEFA1A3* CNVs, we built a composite genetic score (GS) of the three risk-associated variants (*DEFA1A3*, *DEFA3* and *211bp*). To account for the different effect sizes of the three variants, we calculated the standardized weighted sum of their copy numbers where the copy number of each *DEFA1A3* CNV was weighted by its predicted effect size (natural log of OR from multi-variant logistic regression analysis), and the sum of the weighted copy numbers of the three variants was then divided by the average effect size to rescale the score (standardization). Based on the distribution of copy numbers in the patients with follow-up data, we have divided the patients into five groups (quintiles). The percentage of the total variance explained was estimated by calculating Nagelkerke's pseudo  $R^2$  using the fmsb

package, from the result of entering diploid copies of independent *DEFAIA3* CNVs and affection status into the glm function in R (v 2.15.1).

### **Analysis of Clinical Phenotypes**

*DEFAIA3* CNVs were further studied for association with clinicopathologic features. Multiple linear regression models were applied for the correlation analysis of quantitative phenotypes including natural log-transformed proteinuria and serum IgA. A binary logistic regression model was applied for the correlation analysis of categorical variables, such as hyperuricemia and thickening of arterial wall. An ordinal logistic regression model was used for analysis of degrees of interstitial inflammation, tubular atrophy/interstitial fibrosis. The renal survival analysis of *DEFAIA3* CNVs was performed using univariate (for the GS only) and multivariate (GS and known prognostic factors) Cox regression analysis. The renal survival analysis was also performed using Kaplan-Meier analysis, where the patients were divided into five groups (quintiles) according to their GS values, and log rank test was employed to evaluate the differences of renal survival rates between groups. All the statistical analyses were performed using SPSS 16.0 software (SPSS, Chicago, IL) and SAS 9.2 software (SAS Institute).

### **Analysis of HNP1-3 serum and urine level**

The serum and urine level of HNP1–3 were measured with the HNP1–3 ELISA Test Kit (Hycult biotechnology, Uden, The Netherlands) according to the manufacturer’s instructions. Serum was diluted 1:50 in appropriate buffer, while urine was analyzed without dilution. Both serum and urine samples were transferred in duplicate samples to microtitre plates for incubation procedures. Absorbance was measured at 450nm using the SpectraMaxPlus384 Microplatereader

(Molecular devices, USA). The concentration of HNP1–3 was calculated according to the standard curve.

### **Isolation, Culture and Stimulation of Human Neutrophils**

Whole blood (5 ml) was collected using EDTA-coated Vacutainer and processed immediately for neutrophil purification. Neutrophils were isolated by single-step centrifugation of whole blood onto Polymorphprep (Axis-Shield) according to the manufacturer's recommendation. Neutrophils were resuspended in RPMI 1640 medium supplemented with 10% fetal bovine serum at a final concentration of  $1 \times 10^6$ /ml. The purity of the neutrophils was on average >95%. Trypan blue exclusion showed greater than 96% viability. Neutrophils were maintained at 37°C in the presence of 5% CO<sub>2</sub> and then stimulated with LPS from *Escherichia coli* (100 ng/ml) (Sigma, catalog no. L4391) for 6 hours. Neutrophils from similar cultures not exposed to stimulants were used as negative controls. After incubation, cells were centrifuged, and the culture supernatants were stored at –80°C until use for the determination of HNP1–3. The concentration of HNP1–3 in culture supernatants was measured with the HNP1–3 ELISA Kit (Hycult Biotech) as previously described. The ratio for HNP1–3 level in stimulated supernatants normalized with that of negative controls was used in cell culture experiments.

### **Analysis of serum IgA1 and galactose-deficient IgA1 levels**

Serum IgA1 and galactose-deficient IgA1 (using *Helix pomatia* agglutinin, HPA, a lectin that binds terminal GalNAc in galactose-deficient IgA1) were measured by ELISA. The serum samples were diluted into 1:40000 for serum IgA1 and 1: 500 for galactose-deficient IgA1. For

the measurement of serum IgA1, a standard consisting of native IgA1 (Calbiochem) was used at concentrations from 600 ng/ml to 4.68 ng/ml. Rabbit anti-human IgA F(ab')<sub>2</sub> and HRP-labeled F(ab')<sub>2</sub> fragment of goat IgG anti-human IgA1 were purchased from Dako and Southern Biotech, respectively. Neuraminidase and biotin-labeled HPA were purchased from Roche and Sigma, respectively. The absorbance was measured at 490nm with a SpectraMaxPlus384 Microplatereader (Molecular devices, USA). The serum concentration of IgA1 was calculated according to the standard curve. The relative galactose-deficient IgA1 level was calculated as the A<sub>490</sub> value of HPA over the A<sub>490</sub> level of IgA1.

### **Supplementary Materials**

Fig. S1. Schematic map of the *DEFA1A3* CNV locus based on the Sanger sequencing of about 3.4kb *DEFA1A3* region in CHB samples.

Fig. S2. Distribution of *DEFA1A3* CNVs in the combined Chinese IgAN cohort (1189 cases/ 1187 controls)

Fig. S3. Correlation of *DEFA1A3* copy numbers measured by two methods

Fig. S4. Distribution of genetic scores in the combined Chinese IgAN cohort (1189 cases/ 1187 controls)

Fig. S5. Distribution of average copy numbers of *DEFA1A3* CNVs according to genotypes of *rs2738048* in the combined Chinese IgAN cohort

Fig.S6. Distribution of three risk *DEFA1A3* CNVs in the Caucasian and Chinese IgAN cohorts

Fig.S7. Serum level of IgA1 and galactose-deficient IgA1: (A) IgA1; (B) galactose-deficient IgA1



Table S1. Summary of the Southern Chinese samples in the *DEFAIA3* CNVs association analysis

Table S2. Spearman correlation analysis between *DEFAIA3* CNVs and *rs2738048* in the combined Chinese IgAN cohorts

Table S3. Logistic regression analysis of *DEFAIA3* CNVs and *rs2738048* in the Chinese IgAN cohorts

Table S4. Comparison of genomic copy numbers of the *DEFAIA3* CNVs between the IgAN cases and healthy controls in the combined Chinese IgAN cohort

Table S5. The clinical and pathological features at the time of diagnosis for 1189 Chinese IgAN patients

Table S6. Association analysis between the genetic scores and clinicopathologic features in 1189 Chinese IgAN patients

Table S7. Logistic regression analysis of *rs2738048* and *DEFAIA3* CNVs in the combined Chinese IgAN cohort

Table S8. Multivariate linear regression analysis between the expression level of serum HNP1-3, urine HNP1-3, neutrophil HNP1-3 (LPS stimulated) and *DEFAIA3* CNVs

## References

1. G. D'Amico, The commonest glomerulonephritis in the world: IgA nephropathy. *The Quarterly journal of medicine* **64**, 709-727 (1987).
2. J. Barratt, J. Feehally, IgA nephropathy. *Journal of the American Society of Nephrology : JASN* **16**, 2088-2097 (2005).

3. R. J. Wyatt, B. A. Julian, IgA nephropathy. *The New England journal of medicine* **368**, 2402-2414 (2013).
4. G. D'Amico, Natural history of idiopathic IgA nephropathy: role of clinical and histological prognostic factors. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **36**, 227-237 (2000).
5. S. I. Hsu, S. B. Ramirez, M. P. Winn, J. V. Bonventre, W. F. Owen, Evidence for genetic factors in the development and progression of IgA nephropathy. *Kidney international* **57**, 1818-1835 (2000).
6. X. Q. Yu, M. Li, H. Zhang, H. Q. Low, X. Wei, J. Q. Wang, L. D. Sun, K. S. Sim, Y. Li, J. N. Foo, W. Wang, Z. J. Li, X. Y. Yin, X. Q. Tang, L. Fan, J. Chen, R. S. Li, J. X. Wan, Z. S. Liu, T. Q. Lou, L. Zhu, X. J. Huang, X. J. Zhang, Z. H. Liu, J. J. Liu, A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet* **44**, 178-182 (2012).
7. J. Feehally, M. Farrall, A. Boland, D. P. Gale, I. Gut, S. Heath, A. Kumar, J. F. Peden, P. H. Maxwell, D. L. Morris, S. Padmanabhan, T. J. Vyse, A. Zawadzka, A. J. Rees, M. Lathrop, P. J. Ratcliffe, HLA has strongest association with IgA nephropathy in genome-wide analysis. *Journal of the American Society of Nephrology : JASN* **21**, 1791-1797 (2010).
8. A. G. Gharavi, K. Kiryluk, M. Choi, Y. Li, P. Hou, J. Xie, S. Sanna-Cherchi, C. J. Men, B. A. Julian, R. J. Wyatt, J. Novak, J. C. He, H. Wang, J. Lv, L. Zhu, W. Wang, Z. Wang, K. Yasuno, M. Gunel, S. Mane, S. Umlauf, I. Tikhonova, I. Beerman, S. Savoldi, R. Magistroni, G. M. Ghiggeri, M. Bodria, F. Lugani, P. Ravani, C. Ponticelli, L. Allegri, G. Boscutti, G. Frasca, A. Amore, L. Peruzzi, R. Coppo, C. Izzi, B. F. Viola, E. Prati, M.

- Salvadori, R. Mignani, L. Gesualdo, F. Bertinetto, P. Mesiano, A. Amoroso, F. Scolari, N. Chen, H. Zhang, R. P. Lifton, Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* **43**, 321-327 (2011).
9. K. Kiryluk, Y. Li, F. Scolari, S. Sanna-Cherchi, M. Choi, M. Verbitsky, D. Fasel, S. Lata, S. Prakash, S. Shapiro, C. Fischman, H. J. Snyder, G. Appel, C. Izzi, B. F. Viola, N. Dallera, L. Del Vecchio, C. Barlassina, E. Salvi, F. E. Bertinetto, A. Amoroso, S. Savoldi, M. Rocchietti, A. Amore, L. Peruzzi, R. Coppo, M. Salvadori, P. Ravani, R. Magistrone, G. M. Ghiggeri, G. Caridi, M. Bodria, F. Lugani, L. Allegri, M. Delsante, M. Maiorana, A. Magnano, G. Frasca, E. Boer, G. Boscutti, C. Ponticelli, R. Mignani, C. Marcantoni, D. Di Landro, D. Santoro, A. Pani, R. Polci, S. Feriozzi, S. Chicca, M. Galliani, M. Gigante, L. Gesualdo, P. Zamboli, G. G. Battaglia, M. Garozzo, D. Maixnerova, V. Tesar, F. Eitner, T. Rauen, J. Floege, T. Kovacs, J. Nagy, K. Mucha, L. Paczek, M. Zaniew, M. Mizerska-Wasiak, M. Roszkowska-Blaim, K. Pawlaczyk, D. Gale, J. Barratt, L. Thibaudin, F. Berthoux, G. Canaud, A. Boland, M. Metzger, U. Panzer, H. Suzuki, S. Goto, I. Narita, Y. Caliskan, J. Xie, P. Hou, N. Chen, H. Zhang, R. J. Wyatt, J. Novak, B. A. Julian, J. Feehally, B. Stengel, D. Cusi, R. P. Lifton, A. G. Gharavi, Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* **46**, 1187-1196 (2014).
10. C. Wellcome Trust Case Control, N. Craddock, M. E. Hurles, N. Cardin, R. D. Pearson, V. Plagnol, S. Robson, D. Vukcevic, C. Barnes, D. F. Conrad, E. Giannoulatou, C. Holmes, J. L. Marchini, K. Stirrups, M. D. Tobin, L. V. Wain, C. Yau, J. Aerts, T. Ahmad, T. D. Andrews, H. Arbury, A. Attwood, A. Auton, S. G. Ball, A. J. Balmforth, J. C. Barrett, I. Barroso, A. Barton, A. J. Bennett, S. Bhaskar, K. Blaszczyk, J. Bowes, O. J.

Brand, P. S. Braund, F. Bredin, G. Breen, M. J. Brown, I. N. Bruce, J. Bull, O. S. Burren,  
J. Burton, J. Byrnes, S. Caesar, C. M. Clee, A. J. Coffey, J. M. Connell, J. D. Cooper, A.  
F. Dominiczak, K. Downes, H. E. Drummond, D. Dudakia, A. Dunham, B. Ebbs, D.  
Eccles, S. Edkins, C. Edwards, A. Elliot, P. Emery, D. M. Evans, G. Evans, S. Eyre, A.  
Farmer, I. N. Ferrier, L. Feuk, T. Fitzgerald, E. Flynn, A. Forbes, L. Forty, J. A.  
Franklyn, R. M. Freathy, P. Gibbs, P. Gilbert, O. Gokumen, K. Gordon-Smith, E. Gray,  
E. Green, C. J. Groves, D. Grozeva, R. Gwilliam, A. Hall, N. Hammond, M. Hardy, P.  
Harrison, N. Hassanali, H. Hebaishi, S. Hines, A. Hinks, G. A. Hitman, L. Hocking, E.  
Howard, P. Howard, J. M. Howson, D. Hughes, S. Hunt, J. D. Isaacs, M. Jain, D. P.  
Jewell, T. Johnson, J. D. Jolley, I. R. Jones, L. A. Jones, G. Kirov, C. F. Langford, H.  
Lango-Allen, G. M. Lathrop, J. Lee, K. L. Lee, C. Lees, K. Lewis, C. M. Lindgren, M.  
Maisuria-Armer, J. Maller, J. Mansfield, P. Martin, D. C. Massey, W. L. McArdle, P.  
McGuffin, K. E. McLay, A. Mentzer, M. L. Mimmack, A. E. Morgan, A. P. Morris, C.  
Mowat, S. Myers, W. Newman, E. R. Nimmo, M. C. O'Donovan, A. Onipinla, I. Onyiah,  
N. R. Ovington, M. J. Owen, K. Palin, K. Parnell, D. Pernet, J. R. Perry, A. Phillips, D.  
Pinto, N. J. Prescott, I. Prokopenko, M. A. Quail, S. Rafelt, N. W. Rayner, R. Redon, D.  
M. Reid, Renwick, S. M. Ring, N. Robertson, E. Russell, D. St Clair, J. G. Sambrook, J.  
D. Sanderson, H. Schuilenburg, C. E. Scott, R. Scott, S. Seal, S. Shaw-Hawkins, B. M.  
Shields, M. J. Simmonds, D. J. Smyth, E. Somaskantharajah, K. Spanova, S. Steer, J.  
Stephens, H. E. Stevens, M. A. Stone, Z. Su, D. P. Symmons, J. R. Thompson, W.  
Thomson, M. E. Travers, C. Turnbull, A. Valsesia, M. Walker, N. M. Walker, C.  
Wallace, M. Warren-Perry, N. A. Watkins, J. Webster, M. N. Weedon, A. G. Wilson, M.  
Woodburn, B. P. Wordsworth, A. H. Young, E. Zeggini, N. P. Carter, T. M. Frayling, C.

- Lee, G. McVean, P. B. Munroe, A. Palotie, S. J. Sawcer, S. W. Scherer, D. P. Strachan, C. Tyler-Smith, M. A. Brown, P. R. Burton, M. J. Caulfield, A. Compston, M. Farrall, S. C. Gough, A. S. Hall, A. T. Hattersley, A. V. Hill, C. G. Mathew, M. Pembrey, J. Satsangi, M. R. Stratton, J. Worthington, P. Deloukas, A. Duncanson, D. P. Kwiatkowski, M. I. McCarthy, W. Ouwehand, M. Parkes, N. Rahman, J. A. Todd, N. J. Samani, P. Donnelly, Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713-720 (2010).
11. Y. C. Tang, A. Amon, Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**, 394-405 (2013).
  12. B. E. Stranger, M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S. W. Scherer, S. Tavaré, P. Deloukas, M. E. Hurles, E. T. Dermitzakis, Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853 (2007).
  13. M. J. Mason, C. Speake, V. H. Gersuk, Q. A. Nguyen, K. K. O'Brien, J. M. Odegard, J. H. Buckner, C. J. Greenbaum, D. Chaussabel, G. T. Nepom, Low HERV-K(C4) copy number is associated with type 1 diabetes. *Diabetes* **63**, 1789-1795 (2014).
  14. T. J. Aitman, R. Dong, T. J. Vyse, P. J. Norsworthy, M. D. Johnson, J. Smith, J. Mangion, C. Robertson-Lowe, A. J. Marshall, E. Petretto, M. D. Hodges, G. Bhangal, S. G. Patel, K. Sheehan-Rooney, M. Duda, P. R. Cook, D. J. Evans, J. Domin, J. Flint, J. J. Boyle, C. D. Pusey, H. T. Cook, Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851-855 (2006).
  15. S. Cantsilieris, S. J. White, Correlating multiallelic copy number polymorphisms with disease susceptibility. *Human mutation* **34**, 1-13 (2013).

16. P. M. Aldred, E. J. Hollox, J. A. Armour, Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Human molecular genetics* **14**, 2045-2052 (2005).
17. F. F. Khan, D. Carpenter, L. Mitchell, O. Mansouri, H. A. Black, J. Tyson, J. A. Armour, Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. *BMC Genomics* **14**, 719 (2013).
18. R. M. Linzmeier, T. Ganz, Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* **86**, 423-430 (2005).
19. F. J. Cheng, X. J. Zhou, Y. F. Zhao, M. H. Zhao, H. Zhang, Alpha-defensin DEFA1A3 gene copy number variation in Asians and its genetic association study in Chinese systemic lupus erythematosus patients. *Gene* **517**, 158-163 (2013).
20. T. Ganz, Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* **3**, 710-720 (2003).
21. Q. Chen, M. Hakimi, S. Wu, Y. Jin, B. Cheng, H. Wang, G. Xie, T. Ganz, R. M. Linzmeier, X. Fang, Increased genomic copy number of DEFA1/DEFA3 is associated with susceptibility to severe sepsis in Chinese Han population. *Anesthesiology* **112**, 1428-1434 (2010).
22. L. Hazlett, M. Wu, Defensins in innate immunity. *Cell and tissue research* **343**, 175-188 (2011).
23. J. Jarczak, E. M. Kosciuczuk, P. Lisowski, N. Strzalkowska, A. Jozwik, J. Horbanczuk, J. Krzyzewski, L. Zwierzchowski, E. Bagnicka, Defensins: natural component of human innate immunity. *Human immunology* **74**, 1069-1079 (2013).

24. R. I. Lehrer, W. Lu, alpha-Defensins in human innate immunity. *Immunol Rev* **245**, 84-112 (2012).
25. L. M. Rehaume, R. E. Hancock, Neutrophil-derived defensins as modulators of innate immune function. *Critical reviews in immunology* **28**, 185-200 (2008).
26. K. N. Lai, Future directions in the treatment of IgA nephropathy. *Nephron* **92**, 263-270 (2002).
27. K. N. Lai, J. C. Leung, Heat-aggregated IgA prepared from patients with IgA nephropathy increases calcium mobilization and superoxide production of human neutrophils in vitro. *Nephron* **64**, 129-135 (1993).
28. K. Kiryluk, J. Novak, The genetics and immunobiology of IgA nephropathy. *The Journal of clinical investigation* **124**, 2325-2332 (2014).
29. T. W. Groeneveld, T. H. Ramwadhoebe, L. A. Trouw, D. L. van den Ham, V. van der Borden, J. W. Drijfhout, P. S. Hiemstra, M. R. Daha, A. Roos, Human neutrophil peptide-1 inhibits both the classical and the lectin pathway of complement activation. *Molecular immunology* **44**, 3608-3614 (2007).
30. J. A. Armour, R. Palla, P. L. Zeeuwen, M. den Heijer, J. Schalkwijk, E. J. Hollox, Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic acids research* **35**, e19 (2007).
31. M. C. Aldhous, S. Abu Bakar, N. J. Prescott, R. Palla, K. Soo, J. C. Mansfield, C. G. Mathew, J. Satsangi, J. A. Armour, Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Human molecular genetics* **19**, 4930-4938 (2010).

32. P. Fode, C. Jespersgaard, R. J. Hardwick, H. Bogle, M. Theisen, D. Dodoo, M. Lenicek, L. Vitek, A. Vieira, J. Freitas, P. S. Andersen, E. J. Hollox, Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. *PloS one* **6**, e16768 (2011).
33. S. Walker, S. Janyakhantikul, J. A. Armour, Multiplex Parologue Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics* **93**, 98-103 (2009).
34. E. J. Hollox, J. C. Detering, T. Dehnugara, An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. *Human mutation* **30**, 477-484 (2009).
35. S. Abu Bakar, E. J. Hollox, J. A. Armour, Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 853-858 (2009).
36. J. Chen, H. Zheng, J. X. Bei, L. Sun, W. H. Jia, T. Li, F. Zhang, M. Seielstad, Y. X. Zeng, X. Zhang, J. Liu, Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *American journal of human genetics* **85**, 775-785 (2009).
37. E. Alamartine, J. C. Sabatier, C. Guerin, J. M. Berliet, F. Berthoux, Prognostic factors in mesangial IgA glomerulonephritis: an extensive study with univariate and multivariate analyses. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **18**, 12-19 (1991).
38. J. Novak, B. A. Julian, J. Mestecky, M. B. Renfrow, Glycosylation of IgA1 and pathogenesis of IgA nephropathy. *Seminars in immunopathology* **34**, 365-382 (2012).
39. A. Glenthoj, A. J. Glenthoj, N. Borregaard, ProHNPs are the principal alpha-defensins of human plasma. *European journal of clinical investigation* **43**, 836-843 (2013).



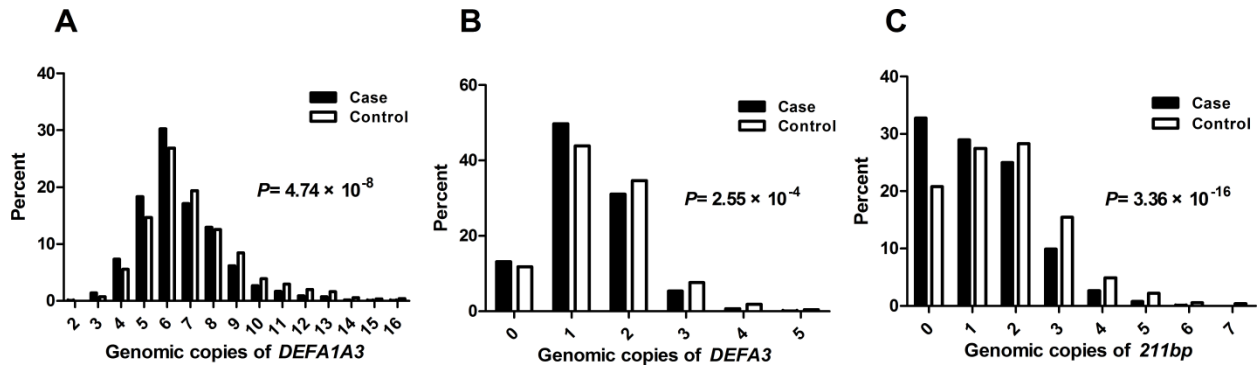
40. F. Emmertsen, A. Glenthoj, J. Sonderskov, P. Kampmann, H. Sengelov, N. Borregaard, ProHNPs are specific markers of normal myelopoiesis. *Blood cancer journal* **4**, e193 (2014).
41. K. N. Lai, Pathogenesis of IgA nephropathy. *Nature reviews. Nephrology* **8**, 275-283 (2012).
42. Y. Suzuki, Y. Tomino, Potential immunopathogenic role of the mucosa-bone marrow axis in IgA nephropathy: insights from animal models. *Seminars in nephrology* **28**, 66-77 (2008).
43. Y. Suzuki, H. Suzuki, D. Sato, T. Kajiyama, K. Okazaki, A. Hashimoto, M. Kihara, K. Yamaji, K. Satake, J. Nakata, M. Aizawa, J. Novak, Y. Tomino, Reevaluation of the mucosa-bone marrow axis in IgA nephropathy with animal models. *Advances in oto-rhino-laryngology* **72**, 64-67 (2011).
44. K. Kiryluk, J. Novak, A. G. Gharavi, Pathogenesis of immunoglobulin A nephropathy: recent insight from genetic studies. *Annual review of medicine* **64**, 339-356 (2013).
45. M. E. Klotman, T. L. Chang, Defensins in innate antiviral immunity. *Nat Rev Immunol* **6**, 447-456 (2006).
46. B. Ericksen, Z. Wu, W. Lu, R. I. Lehrer, Antibacterial activity and specificity of the six human {alpha}-defensins. *Antimicrobial agents and chemotherapy* **49**, 269-275 (2005).
47. F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres, C. Beroud, Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* **37**, e67 (2009).

48. H. A. Black, F. F. Khan, J. Tyson, J. Al Armour, Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. *BMC genomics* **15**, 614 (2014).
49. M. Mueller, P. Barros, A. S. Witherden, A. L. Roberts, Z. Zhang, H. Schaschl, C. Y. Yu, M. E. Hurles, C. Schaffner, R. A. Floto, L. Game, K. M. Steinberg, R. K. Wilson, T. A. Graves, E. E. Eichler, H. T. Cook, T. J. Vyse, T. J. Aitman, Genomic pathology of SLE-associated copy-number variation at the FCGR2C/FCGR3B/FCGR2B locus. *American journal of human genetics* **92**, 28-40 (2013).
50. J. van der Heijden, W. B. Breunis, J. Geissler, M. de Boer, T. K. van den Berg, T. W. Kuijpers, Phenotypic variation in IgG receptors by nonclassical FCGR2C alleles. *Journal of immunology* **188**, 1318-1324 (2012).
51. Kdoqi, KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Diabetes and Chronic Kidney Disease. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **49**, S12-154 (2007).
52. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575 (2007).

**Acknowledgments:** We thank the staff in The First Affiliated Hospital of Sun Yat-sen University for help with sample collection, DNA extraction and sample storage. **Funding:** This work was funded by Guangdong Department of Science & Technology Translational Medicine Center grant (2011A080300002), the Specialized Research Fund for the Doctoral Program of

Higher Education of China (20130171130008), the Science and Technology Planning Project of Guangdong Province, China (2013B051000019), Guangzhou Committee of Science and Technology, China(2012J5100031), Young Scientists Fund of National Natural Science Foundation of China (81200489), Yong Scholars Fund for the Doctoral Program of Higher Education of China (20120171120087), grant from the BBSRC (BB/1006370/1) to JALA, and the Agency for Science & Technology and Research (A\*STAR) of Singapore. OM is supported by the Division of Higher Education, Kingdom of Saudi Arabia (S4674). **Author contributions:** X. T., X.Q., J.C., J.W. and T.L. performed clinical characterization and recruitment of subjects and contributed samples; P.Y., W.L. and Z.Z. performed DNA extraction; Z.A. ,O.M., W. L and S. F. performed experiments; Z.A., J.F. and Q.Z. analyzed and interpreted data; M.L., S.F. and R.X. contributed to analytical support and discussion; X.Y., J.L. and J.A.A. provided the platform, carried out the study protocol and organized the study; Z.A., M.L., J.L. and X.Y prepared the manuscript; J.L. and X.Y. conceived and supervised the project. **Competing interests:** All authors declare no competing financial interests.

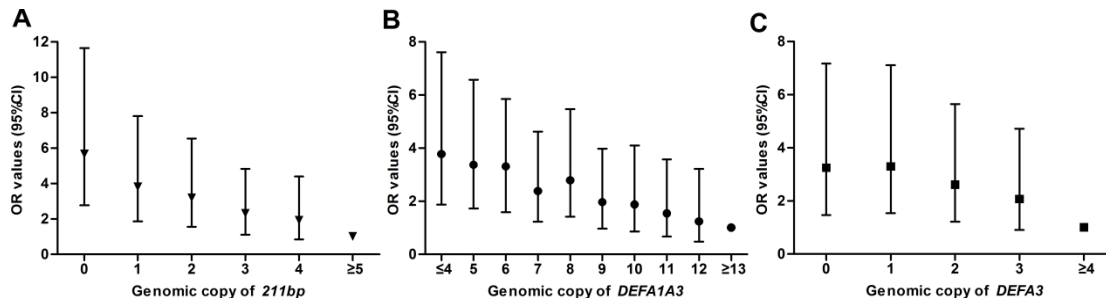
**Figures:**



**Fig.1. Distributions of the copy numbers of the three variants of the *DEFA1A3* CNV locus in 1189 Chinese IgAN cases and 1187 controls**

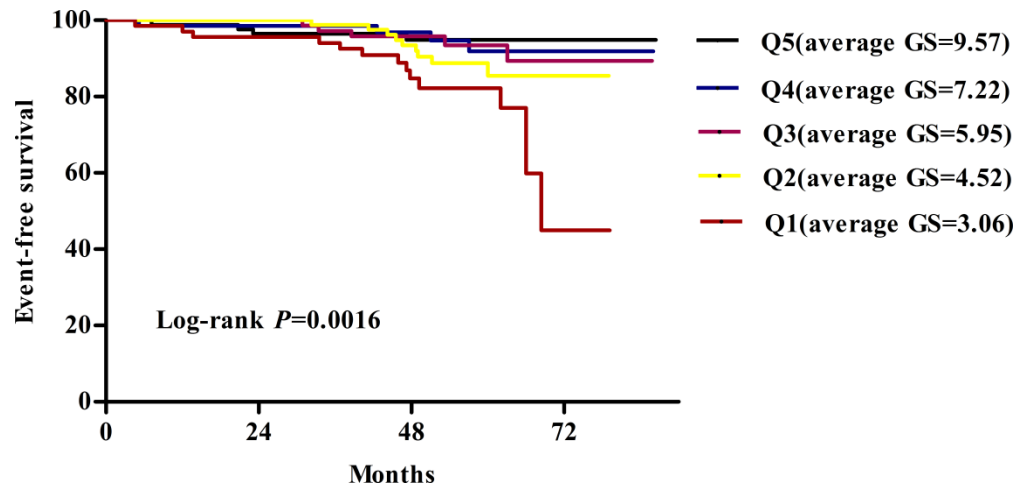
Wilcoxon rank sum tests were used to determine the significance of the distribution difference.

Lower copies of *DEFA1A3* CNVs were more frequent in IgAN patients than controls: (A) total *DEFA1A3*; (B) *DEFA3*; (C) *211bp*.



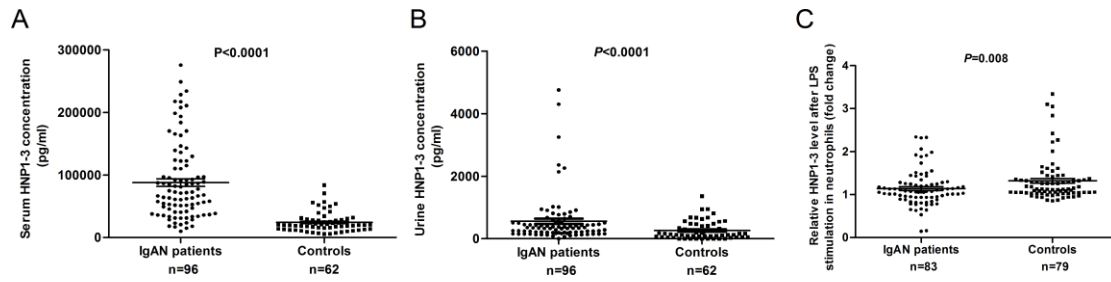
**Fig.2. Distribution of OR values according to different copy numbers of the *DEFA1A3* CNVs**

OR values were calculated from logistic regression analysis in the combined Chinese IgAN cohorts (1189 cases/1187 controls). The overall risk was reduced steadily with the increase of copy number for each variant: **(A)** 211bp; **(B)** *DEFA1A3*; **(C)** *DEFA3*.



**Fig. 3. Renal survival analyses by using the Kaplan-Meier method**

IgAN patients with follow-up data were divided into five groups based on the quintiles of genetic score (GS): Q1 (1<sup>st</sup> quintile), Q2 (2<sup>nd</sup> quintile), Q3 (3<sup>rd</sup> quintile), Q4 (4<sup>th</sup> quintile) and Q5 (5<sup>th</sup> quintile). The percentage of event-free survival (Y axis) was plotted against the follow-up time (X axis) for each of the five groups. Significant difference of renal survival was observed among the five groups of patients with different GS (Log Rank  $P=0.0008$ ).



**Fig.4. Serum, urine and neutrophil level of HNP1-3: (A) Serum HNP1-3; (B) Urine HNP1-3; (C) Extracellular HNP1-3 in isolated neutrophils after stimulation of LPS**

Wilcoxon rank sum tests were used to determine the significance of the expression differences. Serum and urine level of HNP1-3 were both significantly higher in IgAN patients ( $P < 0.001$ ), while extracellular level of HNP1-3 in isolated neutrophils after stimulation of LPS was significantly lower in IgAN patients ( $P < 0.01$ ).

**Table 1. Logistic regression analysis of *DEFA1A3* CNVs and *rs2738048* in combined Chinese IgAN cohort**

Variable	Unconditional <sup>b</sup>		Condition on <i>211bp</i>		Condition on <i>DEFA1A3 &amp; 211bp</i>		Multivariate analysis	
	<i>P</i>	OR (95%CI) <sup>a</sup>	<i>P</i>	OR (95%CI) <sup>a</sup>	<i>P</i>	OR (95%CI) <sup>a</sup>	<i>P</i>	OR (95%CI) <sup>a</sup>
<i>DEFA1A3</i>	3.99×10 <sup>-9</sup>	0.88(0.84,0.92)	3.07×10 <sup>-3</sup>	0.93(0.89,0.98)			2.20×10 <sup>-3</sup> ,	0.93(0.88, 0.97)
<i>DEFA1</i>	6.71×10 <sup>-5</sup>	0.92(0.88,0.96)	NS		NS			
<i>DEFA3</i>	6.55×10 <sup>-5</sup>	0.82(0.76,0.88)	1.35×10 <sup>-3</sup>	0.85(0.77,0.94)	3.14×10 <sup>-3</sup>	0.86(0.78,0.95)	3.33×10 <sup>-3</sup>	0.86(0.77,0.95)
<i>129bp</i>	1.37×10 <sup>-2</sup>	0.88(0.80,0.99)	NS		NS			
<i>124bp</i>	2.89×10 <sup>-8</sup>	0.87(0.82,0.91)	3.07×10 <sup>-2</sup>	0.94(0.89,0.99)	NS			
<i>275bp</i>	2.51×10 <sup>-7</sup>	0.89(0.83,0.91)	NS		NS			
<i>211bp</i>	3.50×10 <sup>-16</sup>	0.75(0.70,0.80)					1.71×10 <sup>-9</sup>	0.78(0.72,0.84)
<i>rs2738048</i>	9.58×10 <sup>-4</sup>	0.80(0.70,0.92)	NS		NS			

<sup>a</sup>: OR per copy of CNV

<sup>b</sup>: from the meta-analysis of the two independent cohorts

NS: not significant,  $P \geq 0.05$ .



**Table2. Survival analysis of the Genetic Score of the *DEFA1A3* CNVs in Chinese IgAN**

	Univariate		Multivariate <sup>a</sup>	
	HR (95%CI)	P-value	HR(95%CI)	P-value
Genetic Score	0.78 (0.66,0.92)	0.004	0.79 (0.66,0.95)	0.012
Quintile of Genetic Score		0.003		0.005
Q5 (n=84, average GS=9.57)	1 (Ref)		1 (Ref)	
Q4 (n=71, average GS=7.22)	1.21 (0.30,4.83)	0.790	0.51 (0.11,2.33)	0.388
Q3 (n=76, average GS=5.95)	1.41 (0.38,5.27)	0.606	1.08 (0.29,4.06)	0.912
Q2 (n=82, average GS=4.52)	2.08 (0.63,6.91)	0.232	1.25 (0.37,4.25)	0.722
Q1 (n=69, average GS=3.06)	5.60 (1.84,17.05)	0.002	3.69 (1.16,11.82)	0.028

<sup>a</sup>: multivariate analysis adjusted for age, gender, proteinuria, hypertension and serum creatinine at biopsy

**Table3. Logistic regression analysis of *DEFA1A3* CNVs in Caucasian IgAN, Chinese DKD and MN cohorts**

Variant	Caucasian IgAN cohort (531 cases/198 controls)		Chinese DKD cohort (806 cases/786 controls)		Chinese MN cohort (493 cases/500 controls)	
	<i>P</i>	OR (95%CI) <sup>a</sup>	<i>P</i>	OR (95%CI) <sup>a</sup>	<i>P</i>	OR (95%CI) <sup>a</sup>
<i>DEFA1A3</i>	4.42×10 <sup>-4</sup>	0.82(0.74,0.92)	4.48×10 <sup>-1</sup>	0.98(0.94,1.03)	8.76×10 <sup>-1</sup>	1.00(0.94,1.07)
<i>DEFA3</i>	4.30×10 <sup>-3</sup>	0.74(0.60,0.91)	7.78×10 <sup>-1</sup>	0.99(0.89,1.09)	6.50×10 <sup>-2</sup>	1.16(0.99,1.35)
<i>211bp</i>	7.56×10 <sup>-2</sup>	0.79(0.61,1.02)	1.58×10 <sup>-1</sup>	0.94(0.87,1.02)	1.11×10 <sup>-7</sup>	0.74(0.67,0.83)

<sup>a</sup>: OR per copy of CNV .

**Table4. Multivariate linear regression analysis between serum IgA1, galactose-deficient IgA1 and *DEFA1A3* CNVs in 158 subjects**

Variables	Galactose-deficient IgA1 <sup>a</sup>				Serum IgA1 <sup>b</sup>			
	Model 1		Model 2		Model 3		Model 4	
	$\beta^c$	<i>P</i>	$\beta^c$	<i>P</i>	$\beta^c$	<i>P</i>	$\beta^c$	<i>P</i>
<i>DEFA1A3</i> copies	-0.12	0.20			-0.13	0.16		
<i>DEFA3</i> copies	-0.09	0.26			-0.13	0.14		
<i>211bp</i> copies	-0.08	0.35			-0.08	0.39		
GS			-0.17	0.03	0.26	0.002	-0.19	0.02
Case (vs. control)	0.21	0.008	0.22	0.006			0.27	0.001

<sup>a</sup> The dependent variable in the analysis was the natural logarithmic transformed value of galactose-deficient IgA1

<sup>b</sup> The dependent variable in the analysis was the square root transformed value of serum IgA1

$\beta^c$  Standardized  $\beta$