The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/34556/1/Author%20accepted%20manuscript-Zhen%20Ai%20et%20al%202016.pdf

Low α-defensin gene copy number increases the risk for IgA nephropathy development and renal dysfunction

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

Affiliations:

[1]Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510080, China.

[2]Key Laboratory of Nephrology, Ministry of Health and Guangdong Province, Guangzhou, Guangdong 510080, China.

[3]Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore.

[4]School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK.

[5]Department of Nephrology, Fuzhou General Hospital of Nanjing Military Command, Fuzhou, Fujian 350025, China.

[6]Department of Nephrology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian 350005, China.

[7]Department of Nephrology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510630, China.

[8]UCL Centre for Nephrology, Royal Free Hospital,London,NW3 2PF,UK

[9]Department of Infection, Immunity and Inflammation, University of Leicester, Leicester LE1 9HN, UK.

[10]School of Biological Sciences, Anhui Medical University, Hefei, Anhui 230032, China.

[11]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 119077, Singapore.

[†]These authors contributed equally.

[*]Corresponding author: yuxq@mail.sysu.edu.cn.

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

Abstract

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide. Although a major source of genetic variation, copy number variations (CNVs) and their involvement in disease development have not been well studied. Here, we performed association analysis of the *DEFA1A3* CNV locus in two independent IgAN cohorts of Southern Chinese Han (total1189 cases and 1187 controls). We discovered three independent copy number associations within the locus: *DEFA1A3* (P=3.99×10^{-9}, OR=0.88), *DEFA3* (P=6.55×10^{-5}, OR=0.82) and a noncoding deletion variant (*211bp*) (P=3.50×10^{-16}, OR=0.75) (OR per copy, fixed-effects meta-analysis). While showing strong association with increased risk for IgAN (P=9.56×10^{-20}), low total copy numbers of the three variants also showed significant association with renal dysfunction in

patients with IgAN (P=0.03, HR=3.69, after controlling for the effects of known prognostic factors) as well as high serum IgA1 (P=0.02) and a high proportion of galactose-deficient IgA1 (P=0.03). For replication, we confirmed the associations of *DEFA1A3* (P=4.42×10^{-4}, OR=0.82) and *DEFA3* copy numbers (P=4.30×10^{-3}, OR=0.74) with IgAN in a Caucasian cohort (531 cases and 198controls) and found the *211*bp variant to be much rarer in Caucasians. Interestingly, we also observed an association of the *211*bp copy number with membranous nephropathy (P=1.11×10^{-7}, OR=0.74 in 493 Chinese cases and 500 matched controls), but not with diabetic kidney disease (in 806 Chinese cases and 786matched controls). By explaining 4.96% of disease risk and influencing the renal dysfunction in IgAN, the *DEFA1A3* CNV locus is a potential candidate for therapeutic target and prognostic marker development.

Introduction

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide(*1*). The clinicopathologic patterns and the prognosis of IgAN show wide variability, ranging from asymptomatic to rapidly progressive forms(*2*). The pathogenesis of IgAN is not well understood, but it is generally considered to be a complex and multifactorial disorder with both genetic and environmental factors contributing to its development(*3*). Thus far, there have been five genome-wide association studies (GWAS) for IgAN, and multiple susceptibility loci have been identified (*4-8*). However, these GWAS loci can only explain 5% to 7.6% of total disease risk (*7*), and none of them has an effect on disease prognosis.

Copy number variations (CNVs) have been shown to be a major source of genetic variation in the human genome(*9, 10*) and have significant phenotypic impact by altering gene dosage,

disrupting coding sequences or perturbing long-range gene regulation(*11*). Although CNVs have been suggested to play important roles in disease development(*10, 12, 13*), the associations of CNVs, particularly the role complex multi-allelic CNVs may play in complex diseases, are poorly studied and understood, mainly due to experimental difficulties in measuring the copy number of such variants accurately(*14*).The α-defensin locus (*DEFA1A3*) is one of the complex multi-allelic CNVs, presents as tandem repeats of a 19kb sequence unit, each of which contains one copy of *DEFA1* or *DEFA3*(differing by a single non-synonymous coding variant within the third exon) as well as several internal bi-allelic polymorphisms(*15-17*). The protein products of *DEFA1A3*, HNP1-3 (human neutrophil peptide 1to 3), are the most abundant of neutrophil granule proteins, and there is evidence to suggest that neutrophils may play an important role in mediating glomerular injury in IgAN(*18-20*). HNP1-3 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 pathogenic roles in IgAN(*21-23*). Interestingly, the *DEFA1A3* CNV is located within the new IgAN susceptibility locus on 8p23.1 that was first discovered in the Chinese population and later validated by independent studies (*4, 7, 8*), which, together with its biological functions, suggests the hypothesis 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 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. For replication, we further confirmed the

associations in Caucasian IgAN samples. Interestingly, we found an association with membranous nephropathy (MN), but not diabetic kidney disease (DKD) in the Chinese population.

Results

Distribution of *DEFA1A3* CNVs in the Chinese population

Two independent cohorts of IgAN samples of southern Chinese Han were analyzed. The first cohort had197 cases and 199 matched healthy controls in terms of age, gender and geographic origin, and the second cohort included 992 cases and 988 healthy controls that were matched for geographic origin(Table S1). We measured the copy numbers of *DEFA1, DEFA3* and *DEFA1A3* (*DEFA1+DEFA3*) as well as the copy numbers 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(*16, 24*) (Fig.S1). The PRT-based methods have been shown to have much greater accuracy for measuring the copy numbers of multi-allelic CNVs than real-time PCR analysis(*25, 26*) and have been successfully applied to studies of genes encoding chemokines (*CCL3L1/CCL4L1*)(*27*), immunoglobulin-receptor (*FCGR3*)(*28*), and beta-defensins (*DEFB*)(*25, 29*).

The copy number of *DEFA1A3* ranged from 2 to 16 copies with a median copy number of 6 per genome in our Chinese samples (Fig.S2, Table S2), which was consistent with the reported median copy numbers of 6 in the HapMap CHB (Chinese Han from Beijing) samples and 7 in CEU (CEPH Utah; United States) and ECACC (European Collection of Cell Culture; UK) samples(*15, 16*), but lower than the reported median copy number of 9 in the Japanese

population (the JPT samples of the HapMap project)(*22*). In addition, the *DEFA1A3* copy numbers of 24 HapMap 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 accuracy of our copy number measurements. Correlation analysis revealed that the copy numbers of *DEFA1*, *124bp*, *275bp* and *215bp* correlated with each other (*P*<0.0001)(TableS3).

Association analysis of the *DEFA1A3* CNVs in Chinese cohort of patients with IgAN

We tested the associations of the *DEFA1A3* CNVs with susceptibility to IgAN using logistic regression analysis and by treating the two cohorts (see above) as independent samples through a fixed-effects meta-analysis. Although sub-population structure existed between northern and southern Chinese Han populations, the samples from southern China were genetically homogeneous (*30*), and no control for population stratification was therefore included in our association analysis. We found highly significant associations for the copy numbers of *DEFA1A3* (*P*=3.99×10⁻⁹, OR=0.88), *DEFA1*(*P*=6.71×10⁻⁵, OR=0.92), *DEFA3* (*P*=6.55×10⁻⁸, OR=0.82), *124bp* (*P*=2.89×10⁻⁸, OR=0.87), *275bp* (*P*=2.51×10⁻⁷, OR=0.89) and *211bp* (*P*=3.50×10⁻¹⁶, OR=0.75) (OR per copy, adjusted for age and gender) without any evidence of genetic heterogeneity between the two independent cohorts (Table S4). Step-wise conditional analysis demonstrated the independent associations of the *211bp*, *DEFA1A3* and *DEFA3* and further multivariate analysis also confirmed the independence of these three associations, *211bp* (*P*=1.71×10⁻⁹, OR=0.78), *DEFA1A3* (*P*=2.20×10⁻³, OR=0.93) and *DEFA3* (*P*=3.33×10⁻³, OR=0.85) (Table 1). All of the three associations remained statistically significant after

Bonferonni correction ($0.05/9=5.56 \times 10^{-3}$). No evidence was found for interaction among the three independent CNVs (Table S5).

The patients carried significantly lower copy numbers of three CNVs than did the controls, indicating that low copy numbers of these variants had a risk effect for IgAN. On average, patients carried 1 copy of *211bp*, 6 copies of *DEFA1A3* and 1 copy of *DEFA3*, which were all significantly lower than the median copy numbers in the healthy controls, which were 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 S6; Fig.1).We found that for all three CNVs, the risk was reduced progressively with the increase in copy number (Fig.2). For example, individuals carrying zero copies of *211bp* (about 26.8% of the population) had a 5.9-fold higher risk than those carrying five or more copies (about 1.9% of the population) ($P=2.17 \times 10^{-6}$). The multi-variant genetic score (GS) (the standardized weighted sum of the copy numbers) of the three variants (ranging from 1.35 to 19.67) (Fig.S4) also showed strong association with IgAN susceptibility ($P=9.56 \times 10^{-20}$). These three variants of the *DEFA1A3* CNV locus could explain up to 4.96% of phenotypic variance cumulatively.

Analysis of clinical phenotypes in a Chinese cohort of patients with IgAN

All of the 1189 IgAN patients analyzed in this study had clinical and pathological information available (summarized in Table S7). 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^2, or dialysis, or renal transplantation), and 9 patients were found to have a doubling of serum creatinine after diagnosis.

The survival analysis of renal outcomes (ESRD or doubling of serum creatinine after diagnosis) using the Cox regression model in the 382 IgAN patients revealed a significant association of the multi-variant GS (ranging from 1.80 to 17.14) with renal degeneration in IgAN (P=0.01, HR=0.79) after adjustment for known prognostic factors, including age, gender, proteinuria, hypertension and serum creatinine at biopsy[31, 32] (Table 2).We further divided the 382 patients into five groups based on the quintile of their GS values and found that the patients in the 1st quintile of GS (average GS=3.06) showed significantly worse renal function than the patients of the 5th quintile (average GS=9.57) (P=0.03, HR=3.69, after adjustment for known prognostic factors) (Fig. 3; Table 2). While confirmation is needed from further studies, our results suggested that patients carrying low copies of *DEFA1A3* CNVs had an increased risk for poor renal outcomes.

The analysis of clinical phenotypes revealed the associations of the multi-variant GS with hyperuricemia (P=0.01) and thickening of the arterial wall (P=0.008). However, the evidence could not survive the conservative Bonferroni correction for testing 15 clinical phenotypes (Table S8), and the findings need to be confirmed by further studies.

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

We genotyped *rs2738048* and *rs12716641* (the previously reported susceptible single-nucleotide polymorphisms (SNPs) within *DEFA1A3* CNV region by GWAS of the Chinese population)[4, 8] in the same 1189 cases and 1187 controls and performed association analysis using the same method described above. We confirmed the associations of *rs2738048* (P=9.58×10^{-4}, OR=0.80) and *rs12716641* (P=6.99×10^{-6}, OR=0.72) in our samples (Table 1), but the associations were weaker than those of *211bp*, *DEFA1A3* and *DEFA3*. Furthermore, conditioning on the

association effect of either *211bp* or *DEFA1A3* fully abolished the association at *rs2738048,* whereas all the associations of the *DEFA1A3* CNVs remained highly significant after conditioning the association effect at *rs2738048* (Table S9). Consistently, the genotypes of *rs2738048* showed strong correlation with the copy numbers of *211bp* and *DEFA1A3* (Table S3; Fig.S5). However, the association effect of *rs12716641* remained significant even after conditioning on all three risk variants of *DEFA1A3* ($P=2.54\times10^{-4}$, OR=0.76) (Table S9). These results clearly demonstrated that the previously reported association at *rs2738048* was due to the association effect of the *DEFA1A3* CNV locus, while the association effect of *rs12716641* was independent of the *DEFA1A3* CNV.

Replication study in Caucasian patients with IgAN and Chinese non-IgAN patients with kidney disease

For replication, we investigated the associations of *211bp*, *DEFA1A3* and *DEFA3* copy numbers in a Caucasian IgAN cohort of 531 cases and 198 controls using the same genotyping method. Both *DEFA1A3* ($P=4.42\times10^{-4}$, OR=0.82) and *DEFA3* ($P=4.30\times10^{-3}$, OR=0.74) showed significant associations in the Caucasian cohort with consistent genetic effects as in the Chinese cohort (Table 3). *211bp* was, however, much rarer in the Caucasian population, and consequently the association of *211bp* did not reach statistical significance ($P=7.56\times10^{-2}$), although its genetic effect was consistent between Caucasian (OR=0.79) and Chinese samples (OR=0.75) (Table 3, Fig.S6).

We further investigated the associations of *211bp*, *DEFA1A3* and *DEFA3* in two Chinese cohorts with non-IgAN renal disease, that was, diabetic kidney disease(806 cases and 786 age and gender-matched controls) and membranous nephropathy (493 cases and 500 age and gender-

matched controls) using the same method. Interestingly, we detected significant association of the *211bp* variant with membranous nephropathy ($P=1.11\times10^{-7}$, OR=0.74 per copy), but no association with diabetic kidney disease (Table4).

Association analysis of *DEFA1A3* CNVs with HNP1-3 and IgA1

To investigate the effect of *DEFA1A3* CNVs on HNP1-3 expression, we first measured the copy numbers of *DEFA1A3* CNVs (*211bp*, *DEFA1A3* and *DEFA3*) as well as the serum and urine amounts of HNP1-3 in 96 IgAN patients and 62 healthy controls. There were no significant differences in copy number distributions of *DEFA1A3* CNVs between patients and controls (Wilcoxon test $P>0.05$). In both serum and urine, HNP1-3 was significantly higher in patients than healthy controls ($P<0.001$) (Fig.4), which was consistent with the fact that HNP1-3 were largely produced by activated neutrophils during an inflammatory response (*33*). However, we could not find any correlation between *DEFA1A3* CNVs and the amounts of either serum or urine HNP1-3 (TableS10). In addition, we also analyzed HNP1-3 in the neutrophils isolated from 83 IgAN patients and 79 healthy controls after stimulating neutrophils with lipopolysaccharide (LPS). While there were no differences in *DEFA1A3* copy number distributions between the patients and controls (Wilcoxon test $P>0.05$), the amount of extracellular HNP1-3 after stimulation with LPS (100ng/ml for 6 hours) was significantly lower in neutrophils isolated from patients than controls ($P=0.006$)(Fig.4). However, we did not detect any association between the *DEFA1A3* CNVs and the amount of extracellular HNP1-3 (Table S10).

Given that the formation of galactose-deficient IgA1-containing immune complexes is the key pathogenic factor contributing to the development of IgAN(*34*), we also investigated the effects of *DEFA1A3* CNVs on serum IgA1 levels and the proportion of galactose-deficient IgA1

in 96 IgAN patients and 62 controls. As expected, the serum IgA1 levels and proportion of galactose-deficient IgA1 were both significantly higher in IgAN patients than in healthy controls ($P<0.01$) (Fig.S7). Although there was no significant evidence for individual *DEFA1A3* CNVs, the multi-variant GS of these variants did show significant negative association with both the serum IgA1 levels ($P=0.02$) and the proportion of galactose-deficient IgA1 ($P=0.03$) (TableS11). These results were suggestive and need to be confirmed by further studies.

Discussion

Our current study has discovered three independent risk factors for IgAN within the *DEFA1A3* CNV locus, the copy numbers of *DEFA1A3*, *DEFA3* and *211bp,* by analyzing two independent Chinese IgAN cohorts and further confirming the association in a Caucasian IgAN cohort. Although each extra copy of these CNVs was only associated with moderate reduction of IgAN risk (by 12% to 25%), the overall impact of the whole *DEFA1A3* CNV locus was strong, because each person carried multiple copies of these variants. By explaining 4.96 % of risk variance, the *DEFA1A3* CNV locus is the strongest genetic susceptibility factor that has been discovered for IgAN so far. Interestingly, our study also revealed that the low copy number of *DEFA1A3* CNV was associated with poor renal function in IgAN patients as well as high serum IgA1 levels and the proportion of galactose-deficient IgA1, a known pathogenic risk factor for IgAN. Consistently, our study has also demonstrated that HNP1-3 expression was elevated during an inflammatory response as seen in the serum and urine samples of patients; neutrophils from healthy individuals produced more HNP1-3 than did those of IgAN patients when stimulated by LPS. Taken together, our genetic and expression analysis suggested that HNP1-3 may provide a protective effect against development of IgAN and the progression of renal dysfunction.

Interestingly, 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* 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 primary glomerulonephritis that 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*). Taken 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* copy number beyond the *DEFA1A3*copy number. *DEFA3* is a more recent human-specific variant that arose from*DEFA1* through a non-synonymous coding mutation in the third exon(*15*). Previous studies have suggested that HNP3 was less active than HNP1, although HNP3was expressed at about twice the level of HNP1(*33, 37, 38*). It is intriguing that the *211bp* copy number also showed an independent association effect. The *211bp* is a 4bp deletion variant (TATC) within the second intron of *DEFA1A3*. Its molecular function is unclear, but the analysis by Human Splicing Finder(*39*) suggests that this 4bp deletion may create a new splice acceptor site (CAG), raising

the possibility that this 4bp deletion may create new splice isoforms, which may in turn influence the expression and activities of the 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 as well as 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(*15, 22*) as well as plasma protein analysis in patients with severe sepsis did not show correlation (*40*), but a small study (8 subjects) showed the correlation of HNP1-3 amounts inside neutrophils with the copy numbers of *DEFA1A3*(*17*). The failure to show a correlation between the *DEFA1A3* copy numbers and expression of HNP1-3 in the previous and current studies is probably due to the fact that HNP1-3 are mainly produced by the bone marrow precursors of neutrophils, promyelocytes and early myelocytes(*33*). As demonstrated in other CNV loci (*41, 42*), a full understanding of the haplotype structure (allelic structure) of the whole *DEFA1A3* CNV locus may also be necessary for demonstrating the correlation.

DEFA1A3 CNVs could potentially influence IgAN development and renal disease progression in several ways. First, as *DEFA1A3* is primarily expressed in promyelocytes within bone marrow, it may have physiological effects on the bone marrow microenvironment (*43, 44*). Second, recent studies of IgAN pathogenesis have highlighted the important role of impaired immune regulation along the mucosa-bone marrow axis and have suggested that galactose-deficient IgA1 seems to be overproduced in bone marrow(*45-47*). *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 against infection. *DEFA1A3* CNVs may influence the dysregulated mucosal immune response in IgAN by enhancing inflammatory signals(*48*).

Our study has demonstrated that the *DEFA1A3* CNVs may be the primary driver of the previously reported association within the GWAS locus on 8p23.1(*4*). The *DEFA1A3* CNVs showed stronger associations than the previously reported SNP *rs2738048*, and conditioning on the effect of *DEFA1A3* CNVs fully abolished the association at *rs2738048*, but not vice versa. So, the association at *rs2738048*was secondaryto the association effect of *DEFA1A3* CNVs. In addition, our study has also demonstrated that the newly reported susceptibility signal *rs12716641*(*8*)was independent of*DEFA1A3* CNVs. A previous study reported a positive correlation between *DEFA1A3*copy number and *rs2738048,* but did not evaluate the association of *DEFA1A3* with IgAN(*49*). Recently, a newly published IgAN GWAS revealed another independent SNP *rs10086568* within the GWAS locus on 8p23.1 in a Caucasian population(*7*). 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 multi-allelic 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 discovered the association of the *DEFA1A3* CNV locus with IgAN risk and renal disease progression, and is one of the few cases where the association of a complex multi-allelic CNV has been demonstrated robustly. 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 primary glomerular diseases. By demonstrating its protective effect again IgAN development and renal disease progression, our study has suggested the possibility of HNP1-3 as a candidate therapeutic target and prognostic marker.

Materials and Methods

Study design

The study was designed to investigate the effect of *DEFA1A3* CNV locus on IgAN development by analyzing the association between the copy numbers of *DEFA1A3* CNVs with IgAN susceptibility and renal disease progression. First, two independent cohorts of Southern Chinese Han, consisting of 1189 IgAN cases and 1187 controls, were employed for the association analysis. Second, for replication, the three risk variants of the *DEFA1A3* CNV locus discovered in Chinese subjects with IgAN were further analyzed in additional samples from Caucasian IgAN patients and controls (531cases and 198 age-, gender-matched controls), Chinese subjects with diabetic kidney disease and controls(806 cases and 786 age-, gender-matched controls) and Chinese subjects with membranous nephropathy and controls (493 cases and 500 age-, gender-matched controls). Clinicopathologic information and follow-up data were collected from

Chinese IgAN patients to further assess the clinical impact of *DEFA1A3* CNVs on renal disease progression. We also investigated the relationship between the *DEFA1A3* CNV locus and the previously published GWAS locus on 8q23.1 by genotyping the reported SNPs (*rs2738048*, *rs12716641*) in the same Chinese cohorts. Lastly, we investigated the effects of *DEFA1A3* CNVs on the expression of HNP1-3 and the proportion of 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-Schönlein purpura nephritis, hepatitis B-associated glomerulonephritis, HIV infection and systemic lupus erythematosus. All cases of diabetic kidney disease were diagnosed as persistent albuminuria (>30mg/g) and/or low estimated glomerular filtration rate (eGFR<60m/min/1.73m^2) in patients with diabetes(*50*). Diabetes was defined as fasting plasma glucose≥7.0 mmol/L and/or 2-h postprandial plasma glucose ≥ 11.1mmol/L, by the use of hypoglycemic agents despite fasting plasma glucose, or any self-reported history of diabetes. All membranous nephropathy cases were adult biopsy-proven idiopathic membranous nephropathy . 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

Using the established paralogue ratio test-based methods(*16, 51*), we measured the copy numbers

of *DEFA1A3* (*DEFA1 + DEFA3*) (by *MLT1A0* and *DEFA4*PRT) and the copy numbers of

DEFA1 and *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 as part of PRT methodology which can

help to determine the copy number more accurately(*16*). Previous studies have suggested that

internal variants may influence the expression and functionality of HNP proteins(*15*) and showed

good correlation with major haplotype classes of *DEFA1A3* locus at least in Europeans(*51*). 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 previously described(*16*).

For the newly-characterized 4bp deletion (TATC)variant within the second intron of *DEFA1A3*,

the *Del4* assay was developed in this study. The PCR for the *Del4* assay was performed by

using1uM 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*, 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 of these reference samples had been determined

unequivocally in previous studies using both segregation and read depth analyses (*16, 51*).The

total *DEFA1A3*copy numbers 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 copy numbers of two variants.

Copy number measurements were carried out by the Nottingham group (Chinese IgAN cohort I),

the Guangzhou group (Chinese IgAN cohort II, Chinese DKD and MN cohort) and the Leicester

group (UK IgAN cohort) using the same methods(Table S1).

Genotyping analysis of *rs2738048* and *rs12716641*

Genotyping was performed using the TaqMan assay from ABI and established methods (see Supplement Materials).

Measurement of HNP1-3 in serum and urine

The serum and urine levels 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 HNP1–3was calculated according to the standard curve.

Isolation, culture and stimulation of human neutrophils

Whole blood (5mL) was collected using EDTA-coated vacutainer and processed immediatelyfor neutrophil purification. Neutrophils were isolated by single-step centrifugation of whole blood onto Polymorphprep (Axis-Shield, Norway) according to the manufacturer's recommendation. Neutrophils were resuspended in RPMI 1640 medium supplemented with 10% fetal bovine serum at a final concentration of $1x10^6$/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% $CO_2$and then stimulated with 100ng/mL of Ec-LPS (Sigma,cat.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 HNP 1-3 in culture supernatants was measured with the HNP1-3 ELISA kit (Hycult Biotech, Uden, The Netherlands) as above description. The ratio for HNP1-3 level in stimulated supernatants normalized with that of negative controls was used in cell culture experiments.

Serum IgA1and galactose-deficient IgA1

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 by 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 600ng/ml to 4.68ng/ml. Rabbit anti-human IgA F(ab')$_2$ and HRP-labeled F(ab')$_2$ fragment of goat IgG anti-human IgA1 were purchased from Dako and SouthernBiotech, 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 IgA1was calculated according to the standard curve. The relative galactose-deficient IgA1 level (proportion of serum IgA1) was calculated as the A_{490} value of HPA over the A_{490} level of IgA1.

Statistical analysis

Analysis of *DEFA1A3* CNVs and SNPs. Wilcoxon rank sum tests and Chi-square tests were used to determine the distribution differences in *DEFA1A3* CNVs and SNPs between cases and controls. Spearman correlation analyses were used to test the correlations between *DEFA1A3* CNVs and SNPs. Association analyses were performed using PROC LOGISTIC in SAS 9.2 software (SAS Institute), and the fixed-effects meta-analysis was performed in PLINK(*52*). The

degree of heterogeneity was estimated using the heterogeneity index (I^2) and Cochran's Q test. Conditional association analyses were performed using a stepwise logistic regression model. A multivariate association analysis was conducted to confirm the independence of associations. Interaction analysis was performed by adding the product term of two CNVs into the logistic regression analysis. A composite genetic score (GS) of the three risk-associated variants (*DEFA1A3, DEFA3* and *211bp*) was built by calculating the standardized weighted sum of 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). The percentage of the total variance explained was estimated by calculating Nagelkerke's pseudo R^2 using the fmsb package where the diploid copies of independent *DEFA1A3* CNVs and the affection status were entered into the glm function in R (v 2.15.1).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.

Analysis of clinical phenotypes and *DEFA1A3* CNVs. A linear regression model was applied for association analysis of quantitative phenotypes, while a logistic regression model was used for analyzing categorical variables and an ordinal logistic regression model was used for ordinal variables. Survival analysis was performed using both univariate (for the GS only) and multivariate (GS and known prognostic factors) Cox regression analysis and Kaplan-Meier method. Hazard ratios (HR) were reported to evaluate the prognostic effects.

Association analysis of HNP1-3 and IgA1 expression with *DEFA1A3* **CNVs.** Linear regression analyses were applied to detect associations between *DEFA1A3* CNVs and expression level of HNP1-3 and IgA1 in samples adjusted by disease status. We assigned normal transformed values of HNP1-3 or IgA1 as independent variables, while three*DEFA1A3* CNVs (*DEFA1A3*, *DEFA3* and *211bp,* or the combined genetic score) were treated as dependent variables. Standardized *β* and *P* value were reported to evaluate the associations.

All the statistical analyses of clinical phenotypes and HNP1-3 and IgA1 expression were performed using SPSS 16.0 software (SPSS, Chicago, IL) and SAS 9.2 software (SAS Institute).

Supplementary Materials

Supplementary Methods

Fig. S1. Schematic map of the *DEFA1A3* CNV

Fig.S2. Distribution of *DEFA1A3* CNVs in the combined Chinese IgAN cohort

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

Fig.S4. Distribution of genetic scores in the combined Chinese IgAN cohort

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

Fig.S6. Distribution of three risk variants of *DEFA1A3* CNVs in the Caucasian IgAN cohort (531 cases/198 controls) and Chinese IgAN cohorts (1189 cases/1187 controls)

Fig.S7. Serum IgA1 and proportion of galactose-deficient IgA1

Table S1. Summary of the study cohorts for the *DEFA1A3* CNVs association analysis

Table S2.Copy number distribution of *DEFA1A3* CNVs in the combined IgAN cohorts (1189 cases/1187 controls)

Table S3.Spearman correlation analysis between *DEFA1A3* CNVs and SNPs in the combined Chinese IgAN cohorts(1189 cases/1187 controls)

Table S4.Logistic regression analysis of *DEFA1A3* CNVs and SNPs in the two independent Chinese IgAN cohorts

Table S5. Interaction analysis of three *DEFA1A3* CNVs in the combined Chinese IgAN cohorts (1189 cases/1187 controls)

Table S6. Comparison of the copy numbers of *DEFA1A3* CNVs between cases and controls in the combined Chinese IgAN cohorts

Table S7. The clinical and pathological features at the time of diagnosis for 1189 Chinese IgAN patients

Table S8. Association analysis between the genetic scores of *DEFA1A3* CNVs and clinicopathologic features in 1189 Chinese IgAN patients

Table S9. Logistic regression analysis of *rs2738048*, *rs12716641*and *DEFA1A3* CNVs in the combined Chinese IgAN cohort(1189 cases/1187 controls)

Table S10. Multivariate linear regression analysis between serum HNP1-3, urine HNP1-3, neutrophil extracellular HNP1-3 (LPS stimulated) and *DEFA1A3* CNVs

Table S11. Linear regression analysis of serum IgA1 and proportion of galactose-deficient IgA1 with the copy numbers of *DEFA1A3* CNVs in 158 subjects

References and Notes

1.   J. Barratt, J. Feehally, IgA nephropathy. *Journal of the American Society of Nephrology : JASN***16**, 2088-2097 (2005).

2.   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).

3.   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).

4.   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).

5.   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).

6.   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. *Nature genetics***43**, 321-327 (2011).

7. 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. Magistroni, 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).

8. M. Li, J. N. Foo, J. Q. Wang, H. Q. Low, X. Q. Tang, K. Y. Toh, P. R. Yin, C. C. Khor, Y. F. Goh, I. D. Irwan, R. C. Xu, A. K. Andiappan, J. X. Bei, O. Rotzschke, M. H. Chen, C. Y. Cheng, L. D. Sun, G. R. Jiang, T. Y. Wong, H. L. Lin, T. Aung, Y. H. Liao, S. M. Saw, K. Ye, R. P. Ebstein, Q. K. Chen, W. Shi, S. H. Chew, J. Chen, F. R. Zhang, S. P.

Li, G. Xu, E. S. Tai, L. Wang, N. Chen, X. J. Zhang, Y. X. Zeng, H. Zhang, Z. H. Liu, X. Q. Yu, J. J. Liu, Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nature communications* **6**, 7270 (2015).

9.   Wellcome Trust Case Control Consortium, 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, J. 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).

10. Y. C. Tang, A. Amon, Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**, 394-405 (2013).

11. 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. Tavare, 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).

12. 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).

13. T. J. Aitman, R. Dong, T. J. Vyse, P. J. Norsworthy, M. D. Johnson, J. Smith, J. Mangion, C. Roberton-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).

14. S. Cantsilieris, S. J. White, Correlating multiallelic copy number polymorphisms with disease susceptibility. *Human mutation* **34**, 1-13 (2013).

15. 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).

16. 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).

17. 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).

18. K. N. Lai, Future directions in the treatment of IgA nephropathy. *Nephron* **92**, 263-270 (2002).

19. 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).

20. H. Jiang, L. Liang, J. Qin, Y. Lu, B. Li, Y. Wang, C. Lin, Q. Zhou, S. Feng, S. H. Yip, F. Xu, E. Lai, J. Wang, J. Chen, Functional networks of aging markers in the glomeruli of IgA nephropathy: a new therapeutic opportunity. *Oncotarget*, (2016).

21. K. Kiryluk, J. Novak, The genetics and immunobiology of IgA nephropathy. *The Journal of clinical investigation***124**, 2325-2332 (2014).

22. 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).

23. T. W. Groeneveld, T. H. Ramwadhdoebe, 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).

24. 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).

25. 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).

26. 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).

27. S. Walker, S. Janyakhantikul, J. A. Armour, Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics***93**, 98-103 (2009).

28. 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).

29. 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).

30. 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).

31. R. J. Wyatt, B. A. Julian, IgA nephropathy. *The New England journal of medicine***368**, 2402-2414 (2013).

32. 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).

33. T. Ganz, Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol***3**, 710-720 (2003).

34. 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).

35. K. Reidy, H. M. Kang, T. Hostetter, K. Susztak, Molecular mechanisms of diabetic kidney disease. *The Journal of clinical investigation***124**, 2333-2340 (2014).

36. H. C. Stanescu, M. Arcos-Burgos, A. Medlar, D. Bockenhauer, A. Kottgen, L. Dragomirescu, C. Voinescu, N. Patel, K. Pearce, M. Hubank, H. A. Stephens, V. Laundy, S. Padmanabhan, A. Zawadzka, J. M. Hofstra, M. J. Coenen, M. den Heijer, L. A. Kiemeney, D. Bacq-Daian, B. Stengel, S. H. Powis, P. Brenchley, J. Feehally, A. J. Rees, H. Debiec, J. F. Wetzels, P. Ronco, P. W. Mathieson, R. Kleta, Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *The New England journal of medicine***364**, 616-626 (2011).

37. M. E. Klotman, T. L. Chang, Defensins in innate antiviral immunity. *Nat Rev Immunol***6**, 447-456 (2006).

38. 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).

39. 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).

40. 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).

41. 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).

42. 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).

43. 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).

44. 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).

45. K. N. Lai, Pathogenesis of IgA nephropathy. *Nature reviews. Nephrology***8**, 275-283 (2012).

46. 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).

47. 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).

48. 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).

49. Y. Y. Qi, X. J. Zhou, F. J. Cheng, P. Hou, L. Zhu, S. F. Shi, L. J. Liu, J. C. Lv, H. Zhang, DEFA gene variants associated with IgA nephropathy in a Chinese population. *Genes and immunity***16**, 231-237 (2015).

50. 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).

51. H. A. Black, F. F. Khan, J. Tyson, J. A. Armour, Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. *BMC genomics***15**, 614 (2014).

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).

Table1.Logistic regression analysis of *DEFA1A3* CNVs and SNPs in the combined Chinese IgAN cohort (1189 cases/1187 controls)

Variable	Unconditional[b]		Condition on 211bp		Condition on DEFA1A3 &211bp		Multivariate analysis	
	P	OR(95%CI)[a]	P	OR (95%CI)[a]	P	OR (95%CI)[a]	P	OR (95%CI)[a]
DEFA1A3	3.99×10^{-9}	0.88(0.84,0.92)	3.07×10^{-3}	0.93(0.89,0.98)			2.20×10^{-3},	0.93(0.88, 0.97)
DEFA1	6.71×10^{-5}	0.92(0.88,0.96)	NS		NS			
DEFA3	6.55×10^{-5}	0.82(0.76,0.88)	1.35×10^{-3}	0.85(0.77,0.94)	3.14×10^{-3}	0.86(0.78,0.95)	3.33×10^{-3}	0.86(0.77,0.95)
129bp	1.37×10^{-2}	0.88(0.80,0.99)	NS		NS			
124bp	2.89×10^{-8}	0.87(0.82,0.91)	3.07×10^{-2}	0.94(0.89,0.99)	NS			
275bp	2.51×10^{-7}	0.89(0.83,0.91)	NS		NS			
211bp	3.50×10^{-16}	0.75(0.70,0.80)					1.71×10^{-9}	0.78(0.72,0.84)
rs2738048	9.58×10^{-4}	0.80(0.70,0.92)	NS		NS			
rs12716641	6.99×10^{-6}	0.72(0.62,0.83)	2.73×10^{-5}	0.74(0.64,0.85)	7.36×10^{-5}	0.75(0.65,0.86)		

[a]OR per copy of CNV [b]from the meta-analysis of the two independent cohorts

NS, not significant, $P\geq0.05$.

Table2. Survival analysis of the Genetic Score of the *DEFA1A3* CNVs in Chinese IgAN patients (n=382)

	Univariate		Multivariate[a]	
	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

[a]multivariate analysis adjusted for age, gender, proteinuria, hypertension and serum creatinine at biopsy

Table3. Logistic regression analysis of three *DEFA1A3* CNVs in Chinese and Caucasian IgAN cohorts and combined samples

Variant	Chinese IgAN cohort (1189cases/1187controls)		Caucasian IgAN cohort (531cases/198controls)		Meta-analysis (fixed effects)			
	P	OR^a(95%CI)	P	OR^a(95%CI)	P	OR^a	Q test	I^2
DEFA1A3	3.99×10^{-9}	0.88(0.84,0.92)	4.42×10^{-4}	0.82(0.74,0.92)	4.28×10^{-11}	0.87	0.27	0
DEFA3	6.55×10^{-5}	0.82(0.76,0.88)	4.30×10^{-3}	0.74(0.60,0.91)	5.81×10^{-6}	0.80	0.34	18
211bp	3.50×10^{-16}	0.75(0.70,0.80)	7.56×10^{-2}	0.79(0.61,1.02)	4.92×10^{-16}	0.75	0.76	0

[a]OR per copy of CNV

Table4. Logistic regression analysis of *DEFA1A3* CNVs in Chinese patients with diabetic kidney disease (DKD) or membranous nephropathy (MN)

Variant	Chinese DKD cohort (806 cases/786 controls)		Chinese MN cohort (493 cases/500 controls)	
	P	OR (95%CI)[a]	P	OR (95%CI)[a]
DEFA1A3	4.48×10^{-1}	0.98(0.94,1.03)	8.76×10^{-1}	1.00(0.94,1.07)
DEFA3	7.78×10^{-1}	0.99(0.89,1.09)	6.50×10^{-2}	1.16(0.99,1.35)
211bp	1.58×10^{-1}	0.94(0.87,1.02)	1.11×10^{-7}	0.74(0.67,0.83)

[a]OR per copy of CNV .

Figure Legends

Fig.1. Copy number distribution of the three variants of the *DEFA1A3* CNV locus. The copy number distribution of the three variants of the *DEFA1A3* CNV locus was examined in 1189 Chinese patients with IgAN 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 in controls: (**A**) *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 patient cohorts. The percentage of Chinese population carrying a particular copy number of CNV variant was shown on Y axis. The overall risk of IgAN decreased steadily with the increase in copy number for each variant: (**A**) *211bp* (**B**) *DEFA1A3* (**C**) *DEFA3*.

Fig.3. Survival analyses 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 (1st quintile), Q2 (2nd quintile), Q3 (3rd quintile), Q4 (4th quintile) and Q5 (5th quintile). A significant difference in survival of renal function was observed among the five groups of patients with different GS (Log Rank P=0.0008).

Fig.4. Amounts of HNP1-3 in serum and urine. Shown are the amounts of HNP1-3 in (A) serum (B) urine and (C) produced by neutrophils after stimulation with LPS. Wilcoxon rank sum tests were used to determine the significance of the expression differences. Serum and urine HNP1-3 were both significantly higher in IgAN patients (P<0.001), whereas HNP1-3 produced by neutrophils after stimulation with LPS in vitro was significantly lower in IgAN patients compared to controls (P<0.01).
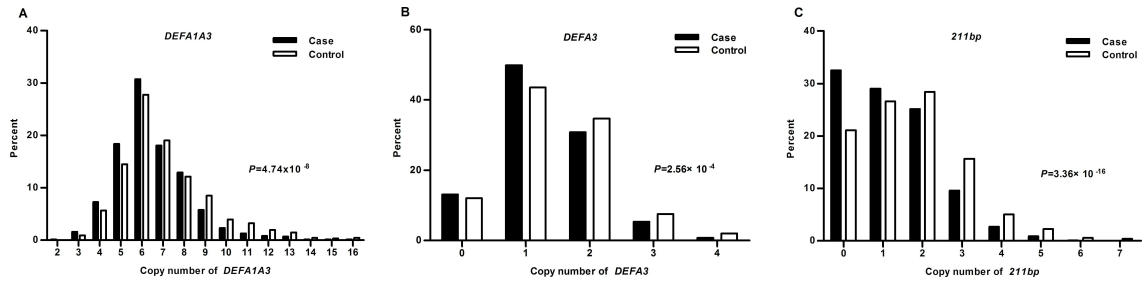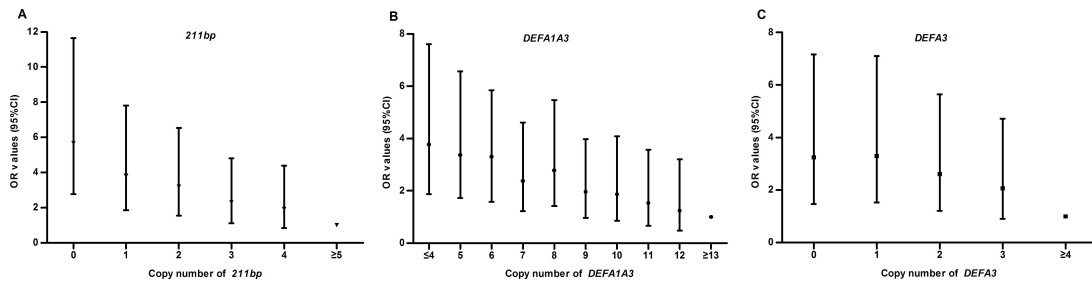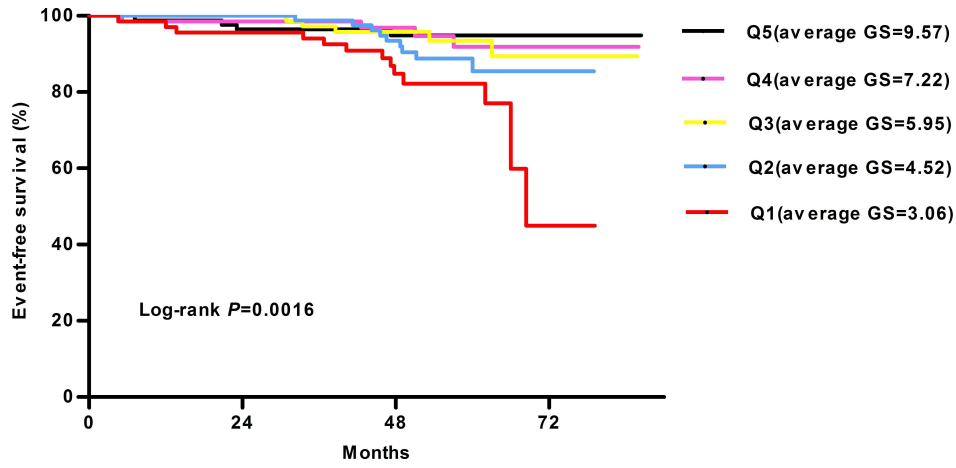
Fig.1.

Fig.2.

Fig.3.

Fig.4.