nature genetics

A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy

Xue-Qing Yu[1,2,15], Ming Li[1,2,15], Hong Zhang[3,15], Hui-Qi Low[4], Xin Wei[1,2], Jin-Quan Wang[5], Liang-Dan Sun[6,7], Kar-Seng Sim[4], Yi Li[4], Jia-Nee Foo[4], Wei Wang[1,2], Zhi-Jian Li[1,2], Xian-Yong Yin[6,7], Xue-Qing Tang[1,2], Li Fan[1,2], Jian Chen[8], Rong-Shan Li[9], Jian-Xin Wan[10], Zhang-Suo Liu[11], Tan-Qi Lou[12], Li Zhu[3], Xiao-Jun Huang[13], Xue-Jun Zhang[6,7], Zhi-Hong Liu[5] & Jian-Jun Liu[4,14]

We performed a two-stage genome-wide association study of IgA nephropathy (IgAN) in Han Chinese, with 1,434 affected individuals (cases) and 4,270 controls in the discovery phase and follow-up of the top 61 SNPs in an additional 2,703 cases and 3,464 controls. We identified associations at 17p13 (rs3803800, $P = 9.40 \times 10^{-11}$, OR = 1.21; rs4227, $P = 4.31 \times 10^{-10}$, OR = 1.23) and 8p23 (rs2738048, $P = 3.18 \times 10^{-14}$, OR = 0.79) that implicated the genes encoding tumor necrosis factor (*TNFSF13*) and α-defensin (*DEFA*) as susceptibility genes. In addition, we found multiple associations in the major histocompatibility complex (MHC) region (rs660895, $P = 4.13 \times 10^{-20}$, OR = 1.34; rs1794275, $P = 3.43 \times 10^{-13}$, OR = 1.30; rs2523946, $P = 1.74 \times 10^{-11}$, OR = 1.21) and confirmed a previously reported association at 22q12 (rs12537, $P = 1.17 \times 10^{-11}$, OR = 0.78). We also found that rs660895 was associated with clinical subtypes of IgAN ($P = 0.003$), proteinuria ($P = 0.025$) and IgA levels ($P = 0.047$). Our findings show that IgAN is associated with variants near genes involved in innate immunity and inflammation.

IgA nephropathy (IgAN) is the most common primary glomerulonephritis among individuals undergoing renal biopsy[1,2]. It is characterized by the deposition of IgA in the mesangial area of glomeruli, accompanied by various histopathological lesions including mesangial cell proliferation and accumulation of extracellular matrix[2,3]. The prevalence of IgAN is higher in Asia than in western countries[4,5]. Although most individuals with IgAN have a mild form of the disease, 15–40% of cases will progress to end-stage renal diseases within 20 years of disease onset[3,6].

The pathogenesis of IgAN is unclear, but both genetic and environmental factors contribute to its development[6,7]. Several lines of evidence support the importance of genetic risk factors, including differences in ethnic and geographical distributions, familial clustering and inter-individual variation in disease course and prognosis[8–10].

Both linkage and association studies have been performed to identify genetic risk factors for IgAN[11]. Three genome-wide linkage studies of familial IgAN[12–14] reported linkages at 6q22–23, 4q26–31 and 17q12–22, but no disease genes were identified. Studies of human leukocyte antigen (HLA) molecules have found associations with *HLA-DQ* and *HLA-DR* alleles[6,15–17]. Candidate genes encoding proteins involved in adaptive and innate immunity, glycosylation of IgA1 and the renin-angiotensin system have also been investigated[18], but most of these studies were limited by small sample sizes, insufficient methodologies and lack of validation in independent samples. Recently, a genome-wide association study (GWAS) of IgAN reported associations within the MHC, 1q32 and 22q12 regions[19].

Here we performed a two-stage GWAS of IgAN using three independent samples of Han Chinese consisting of a total of 4,137 cases and 7,734 controls (**Table 1**). In the discovery stage, we analyzed 444,882 common autosomal SNPs in 1,434 cases and 4,270 controls. Principal components analysis (PCA) confirmed all the samples to be Chinese but indicated moderate genetic mismatch between cases and controls (**Supplementary Fig. 1**), which caused an inflation of the genome-wide association results ($\lambda_{gc} = 1.36$). We used two different methods to minimize the adverse effects of population stratification. First, we performed the association analysis using PCA-based correction for population stratification, which reduced the λ_{gc} to 1.054. Second, we divided all cases and controls from the discovery sample

[1]Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China. [2]Key Laboratory of Nephrology, Ministry of Health, Guangzhou, Guangdong, China. [3]Renal Division, Peking University First Hospital, Peking University, Institute of Nephrology, Beijing, China. [4]Human Genetics, Genome Institute of Singapore, Singapore. [5]Department of Nephrology, Nanjing General Hospital of Nanjing Military Command, Nanjing, Jiangsu, China. [6]Institute of Dermatology and Department of Dermatology, No.1 Hospital, Anhui Medical University, Hefei, Anhui, China. [7]State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Hefei, Anhui, China. [8]Department of Nephrology, Fuzhou General Hospital of Nanjing Military Command, Fuzhou, Fujian, China. [9]Department of Nephrology, The Second Hospital of Shanxi Medical University, Taiyuan, Shanxi, China. [10]Department of Nephrology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian, China. [11]Department of Nephrology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China. [12]Department of Nephrology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China. [13]Department of Hematology, Peking University People's Hospital, Beijing, China. [14]School of Life Sciences, Anhui Medical University, Hefei, Anhui, China. [15]These authors contributed equally to the work. Correspondence should be addressed to X.-Q.Y. (yuxq@mail.sysu.edu.cn) or J.-J.L. (liuj3@gis.a-star.edu.sg).

Received 6 July; accepted 1 December; published online 25 December 2011; doi:10.1038/ng.1047

Table 1 Summary of samples in the GWAS and replication studies

		Cases				Controls		
Cohort	Population	Sample size	Mean age	Mean age at diagnosis	Male/female (%)	Sample size	Mean age	Male/female (%)
GWAS	Southern Chinese	1,434	35.0	32.0	44.0/56.0	4,270	43.8	47.5/52.5
Replication 1	Northern Chinese	1,402	33.3	31.7	54.7/45.3	1,716	29.5	58.2/41.8
Replication 2	Southern Chinese	1,301	35.0	31.4	45.2/54.8	1,748	36.5	51.8/48.2
Total		4,137	34.5	31.7	48.3/51.7	7,734	40.0	53.7/46.3

into three genetically homogeneous clusters (using the *K*-means method) and then performed the association analysis, which reduced the λ_{gc} to 1.074. For both analyses, the quantile-quantile analysis of the observed *P* values showed a good overall fit with the null distribution (**Supplementary Fig. 2**), and the two analyses yielded consistent results (**Fig. 1**). Taken together, our results indicated that the final association results from our discovery stage showed minimal inflation effects due to population stratification.

Our discovery analysis identified extensive associations within the MHC region (**Supplementary Fig. 3**). Forward conditional analysis of the 2,644 SNPs within the MHC identified multiple independent associations at rs2523946, rs9264942, rs660895, rs1794275 and rs9277554 (**Supplementary Table 1**). Conditioning on these five SNPs abolished the extensive associations within the MHC (smallest $P_{conditional} > 0.0002$) (**Supplementary Fig. 3**). In addition, the discovery analysis also identified suggestive associations with SNPs on chromosomes 7, 8, 17 and 22 ($P < 10^{-5}$) (**Fig. 1**).

To validate the results from the discovery analysis, we selected and genotyped 61 SNPs (56 non-MHC SNPs and 5 MHC SNPs) in two independent replication samples of southern and northern Han Chinese, consisting of a total of 2,703 cases and 3,464 controls. Of the 56 non-MHC SNPs, four showed consistent odds ratios (ORs) across the three independent GWAS and replication samples, as well as significant association in the combined replication samples ($P < 0.05$ after correction for testing 61 SNPs). The associations at the four SNPs all surpassed genome-wide significance ($P < 5 \times 10^{-8}$) in the combined GWAS and replication samples (rs2738048 at 8p23.1, $P = 3.18 \times 10^{-14}$, OR = 0.79; rs3803800 at 17p13, $P = 9.40 \times 10^{-11}$, OR = 1.21; rs4227 at 17p13, $P = 4.31 \times 10^{-10}$, OR = 1.23; rs12537 at 22q12, $P = 1.17 \times 10^{-11}$, OR = 0.78) (**Table 2**). rs3803800 and rs4227 are 28 kb apart and in strong linkage disequilibrium (LD) ($D' = 0.94$, $r^2 = 0.40$), and conditioning on rs4227 abolished the association at rs3803800 ($P_{conditional} = 0.03$).

Of the five MHC SNPs, we confirmed three that surpassed genome-wide significance in the combined GWAS and replication samples (rs2523946, $P = 1.74 \times 10^{-11}$, OR = 1.21; rs660895, $P = 4.13 \times 10^{-20}$, OR = 1.34; rs1794275, $P = 3.43 \times 10^{-13}$, OR = 1.30) (**Table 2**). Conditional analysis showed that these associations at these three SNPs are independent (**Supplementary Table 2**). Furthermore, of the eight haplotypes comprising the three SNPs, four showed significant association (TAG, $P = 8.54 \times 10^{-21}$, OR = 0.75; TGA, $P = 7.41 \times 10^{-4}$, OR = 1.34; CGA, $P = 9.70 \times 10^{-12}$, OR = 1.47; CGG, $P = 4.46 \times 10^{-7}$, OR = 1.24) (**Supplementary Table 3**). These results clearly implicate multiple independent associations within the MHC region.

Adjustment for age and gender had a moderate impact on the associations at the seven validated SNPs, causing a few adjusted *P* values to fall just below genome-wide significance (**Supplementary Table 4**). Also, analysis of the seven SNPs using the 4,137 cases and 2,535 clinically verified healthy controls yielded similar OR estimates to those obtained during analysis of 4,137 cases and all 7,734 controls (including both the clinically verified and self-reported healthy controls) (**Supplementary Table 4**). The validation results of the other 54 SNPs are summarized in **Supplementary Table 5**.

We further investigated *HLA* allele associations in our discovery sample through imputation. We identified multiple associations, with the strongest one at *B*4001* ($P = 5.64 \times 10^{-7}$, OR = 1.34). Forward stepwise conditional analysis identified additional independent associations at *HLA-DQB*0302* ($P = 6.18 \times 10^{-5}$, OR = 1.42) and *HLA-A*1101* ($P = 4.01 \times 10^{-4}$, OR = 1.19) (**Supplementary Table 6**). Furthermore, our imputation analysis also identified two *HLA* haplotypes showing associations: *DRB*0901-DQB*0303* ($P = 6.01 \times 10^{-5}$, OR = 0.77) and *DQA*0201-DQB*0201* ($P = 3.84 \times 10^{-5}$, OR = 0.60) (**Supplementary Table 7**).

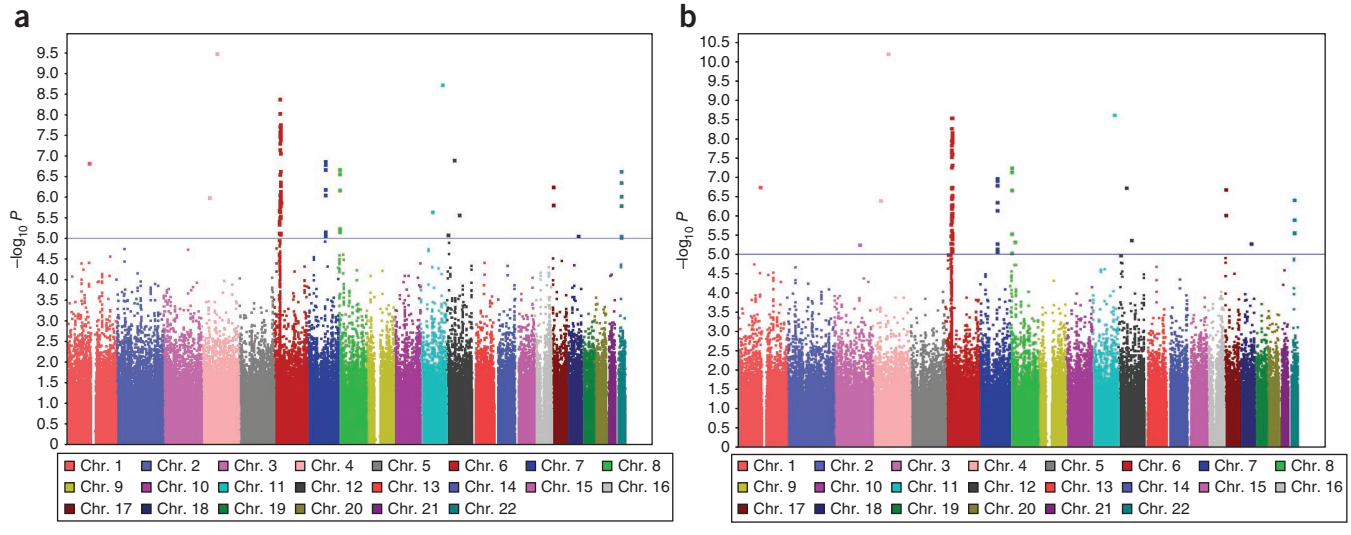

Figure 1 Manhattan plots of the *P* values of association obtained by analysis of the 444,882 polymorphic SNPs in the discovery sample of 1,434 IgAN cases and 4,270 controls. (**a**) Plot obtained following adjustment for the top five principal components from the PCA. (**b**) Plot obtained after subdividing samples into three genetically homogeneous clusters. The $-\log_{10}$ *P* values (*y* axis) are presented against their chromosomal positions (*x* axis). The blue horizontal line represents a *P*-value threshold of 10^{-5} for suggestive significance.

Table 2 Association results of the seven validated SNPs in the GWAS, replication and combined samples

dbSNP locus	Alleles (risk allele)	Location (bp)	Coding or noncoding	Gene	Stage[a]	Cases		Controls					
						n	RAF	*n*	RAF	*P*	OR (95% CI)[c]	I^2	Q^b
rs2738048	G/A	6810195	Noncoding	*DEFAs*	GWAS	1,434	0.27	4,270	0.32	2.04E–07	0.77 (0.70–0.85)		
8p23.1	G				Replication 1	1,402	0.27	1,716	0.32	1.80E–05	0.78 (0.70–0.88)		
					Replication 2	1,301	0.29	1,748	0.33	3.48E–04	0.82 (0.74–0.91)		
					Overall	4,137	0.27	7,734	0.32	3.18E–14	0.79 (0.74–0.84)	0	0.729
rs3803800	A/G	7403693	Coding	*TNFSF13*	GWAS	1,434	0.42	4,270	0.36	1.50E–06	1.25 (1.14–1.36)		
17p13	A				Replication 1	1,402	0.35	1,716	0.31	6.75E–05	1.24 (1.12–1.38)		
					Replication 2	1,301	0.38	1,748	0.35	2.62E–02	1.13 (1.02–1.26)		
					Overall	4,137	0.38	7,734	0.35	9.40E–11	1.21 (1.14–1.28)	0.082	0.336
rs4227	G/T	7431901	Coding	*MPDU1*	GWAS	1,434	0.27	4,270	0.22	5.52E–07	1.29 (1.17–1.42)		
17p13	G				Replication 1	1,402	0.21	1,716	0.17	1.77E–03	1.22 (1.08–1.39)		
					Replication 2	1,301	0.24	1,748	0.21	1.55E–02	1.16 (1.03–1.31)		
					Overall	4,137	0.24	7,734	0.21	4.31E–10	1.23 (1.16–1.32)	0	0.428
rs12537	T/C	28753460	Coding	*MTMR3*	GWAS	1,434	0.14	4,270	0.18	2.30E–07	0.72 (0.64–0.82)		
22q12	T				Replication 1	1,402	0.21	1,716	0.24	1.17E–03	0.82 (0.73–0.92)		
					Replication 2	1,301	0.15	1,748	0.18	1.19E–03	0.80 (0.69–0.91)		
					Overall	4,137	0.16	7,734	0.19	1.17E–11	0.78 (0.72–0.84)	0.076	0.339
rs2523946	C/T	30049922	Noncoding	*HLA-A*	GWAS	1,434	0.58	4,270	0.51	4.05E–09	1.30 (1.19–1.42)		
6p21.3	C				Replication 1	1,402	0.48	1,716	0.45	2.82E–02	1.12 (1.01–1.24)		
					Replication 2	1,301	0.55	1,748	0.51	1.15E–03	1.18 (1.07–1.31)		
					Overall	4,137	0.54	7,734	0.50	1.74E–11	1.21 (1.15–1.28)	0	0.996
rs660895	G/A	32685358	Noncoding	*HLA-DRB1*	GWAS	1,434	0.31	4,270	0.26	8.19E–08	1.29 (1.18–1.42)		
6p21.3	G				Replication 1	1,402	0.26	1,716	0.19	1.08E–10	1.49 (1.32–1.68)		
					Replication 2	1,301	0.29	1,748	0.24	2.29E–05	1.28 (1.14–1.44)		
					Overall	4,137	0.28	7,734	0.24	4.13E–20	1.34 (1.26–1.42)	0	1
rs1794275	A/G	32779226	Noncoding	*HLA-DQA/B*	GWAS	1,434	0.18	4,270	0.14	3.28E–08	1.38 (1.23–1.55)		
6p21.3	A				Replication 1	1,402	0.20	1,716	0.16	3.81E–04	1.25 (1.10–1.41)		
					Replication 2	1,301	0.19	1,748	0.15	5.74E–04	1.25 (1.10–1.42)		
					Overall	4,137	0.19	7,734	0.15	3.43E–13	1.30 (1.21–1.39)	0	0.997

"Validated" means (i) showing consistent OR across the three independent GWAS and replication samples; (ii) showing significant association in the combined replication sample ($P < 0.05$ after correction for testing 61 SNPs in the replication study); and (iii) reaching genome-wide significance ($P < 5 \times 10^{-8}$) in the combined GWAS and replication samples. [a]Replication 1, northern samples; replication 2, southern samples. [b]Q, *P* value for the Cochran's *Q* statistic; all the *P* values are unadjusted. [c]95% CI, 95% confidence interval.

We also investigated the previously reported associations within the MHC, 1q32 and 22q12 regions[17] in our discovery sample. The reported rs9275596 in *HLA-DR-DQ* showed strong association in our discovery sample ($P = 1.65 \times 10^{-8}$, OR = 0.68) and is in strong LD with rs660895 and rs1794275 ($D' = 1.00$). Conditioning on the haplotypes of our three validated MHC SNPs greatly reduced the strength of the association at rs9275596 ($P = 2.06 \times 10^{-3}$, OR = 0.78), but the TAG haplotype of the three SNPs remained significant genome-wide ($P = 1.89 \times 10^{-12}$, OR = 0.70) after conditioning on rs9275596 (**Supplementary Table 8**). Our discovery sample also yielded moderate evidence for the other two reported MHC SNPs (rs9357155, $P = 2.07 \times 10^{-3}$, OR = 0.82; rs1883414, $P = 1.80 \times 10^{-2}$, OR = 0.88).

The reported rs2412971 at 22q12 also showed significant association in our discovery sample ($P = 4.31 \times 10^{-7}$, OR = 0.78). rs2412971 and rs12537 are in strong LD ($D' = 0.98$, $r^2 = 0.50$) (**Supplementary Fig. 4**), and conditioning on rs12537 abolished the association at rs2412971 ($P_{conditional} = 0.06$, $OR_{conditional} = 0.87$). Our discovery sample did not identify significant associations for the two reported SNPs at 1q32 (**Supplementary Fig. 5**), but it did yield consistent OR estimates (being protective) with the previous study, although the ORs from our study were much more moderate than the reported ones (**Supplementary Table 9**). Notably, imputation analyses identified suggestive associations in the region adjacent to the 1q32 locus that were independent of the reported SNPs ($D' < 0.05$) (**Supplementary Fig. 5**). The previous GWAS was performed using samples of northern Chinese[19], whereas our study was done in southern Chinese. The disparity of association between the two studies may be a result of chance (owing to the low frequency of these SNPs in the Chinese population), but it could also indicate potential genetic heterogeneity between different geographical populations of China. Further studies will be needed to investigate this possibility.

We further investigated the associations of the seven validated susceptibility SNPs with the clinical phenotypes of IgAN in 3,394 cases in which clinical information were collected (**Supplementary Table 10**). We found that the minor, or risk, allele (G) of rs660895 was associated with a mild subtype (asymptomatic hematuria or proteinuria) of IgAN ($P = 0.003$), mild proteinuria ($P = 0.025$) but higher IgA level ($P = 0.047$) in cases (**Table 3**). Causal variant(s) underlying the

Table 3 Association between rs660895 and clinical phenotypes

	AA	AG	GG	OR (95% CI)[a]	*P*
Asymptomatic hematuria or proteinuria	0.27 (268)	0.31 (249)	0.34 (62)	1	
Nephritic syndrome	0.61 (592)	0.59 (467)	0.60 (110)	0.88 (0.76–1.02)	
Nephrotic syndrome	0.12 (119)	0.10 (78)	0.06 (11)	0.67 (0.52–0.86)	0.003
Mean of proteinuria (95% CI)	1.67 (1.56–1.78)	1.56 (1.40–1.61)	1.30 (1.11–1.48)		0.025
Number of observations	1,487	1,176	276		
Mean of IgA level (95% CI)	2.95 (2.88–3.02)	2.96 (2.89–3.03)	3.10 (2.95–3.24)		
Number of observations	1,546	1,201	277		0.047

[a]Estimated OR of nephritic or nephrotic syndrome subtype per G allele (with asymptomatic subtype as reference group).

association at rs660895 may influence development of IgAN by affecting homeostasis of urine proteins and IgA. This idea is consistent with previous findings that tubular HLA-DR expression was higher in IgAN cases with proteinuria than without proteinuria[20], and proteinuria was associated with higher percentages of HLA-DR-positive cells in HIV-infected cases[21]. In addition, we also observed associations of rs12537 with severe proteinuria ($P = 0.002$), rs3803800 with higher IgA level ($P = 0.014$) and rs2523946 with higher incidence of gross hematuria ($P = 0.022$) and lower glomerular filtration rate ($P = 0.031$) (**Supplementary Tables 11** and **12**).

We also compared the allele frequencies of the newly identified SNPs among different ethnic populations. Whereas the protective allele of rs2738038 at 8p23 is more common in Asian than in non-Asian populations, the risk alleles of rs3803800 and rs4227 at 17p13 are similar between Asian and European ancestry populations but much higher in African populations (**Supplementary Table 13**). The allele frequency differences at these loci, however, cannot explain the higher prevalence of IgAN in Asian than in non-Asian populations. We further examined the fixation index (F_{st}), integrated haplotype score (iHS) and cross-population extended haplotype homozygosity (XP-EHH) score around the two newly discovered loci using the HGDP Selection Browser (see URLs), but did not find any evidence of natural selection (**Supplementary Fig. 6**).

The association at 8p23.1 is within an LD block where several members of the *DEFA* family reside (**Fig. 2**). Imputation analysis of the region did not identify any SNPs showing stronger association than rs2738048, a noncoding variant without any indication of regulatory function (according to the information at the UCSC Genome Browser). *DEFA* genes encode α-defensins, which can act as endogenous antibiotics and are involved in the inflammation response to infection[22]. Defensins are also immune modulators that chemoattract native T cells, immature dendritic cells and monocytes and induce the release of chemokines and cytokines, including interleukin-8 (IL-8) and monocyte chemoattractant MCP-1 (ref. 23), whose urinary levels were shown to be elevated in individuals with IgAN[24,25]. Defensins are also involved in kidney diseases. Elevated serum α-defensin at one to three time normal concentrations were observed in type 1 diabetic nephropathy[26] and lupus nephritis[27].

The association at 17p13 is within a 20-kb LD block containing a number of genes, including *TNFSF13*, *MPDU1*, *EIF4A1*, *CD68*, *TP53* and *SOX15* (**Fig. 2**). Imputation identified additional SNPs showing similar or slightly stronger association than did rs4227 (**Fig. 2**), but none of them is a coding variant (data not shown). *TNFSF13* (also called *APRIL*) encodes a member of the tumor necrosis factor (TNF) ligand family, which is important for B-cell development[28]. The serum IgA level and IgA response to mucosal antigen exposure were both reduced in *Tnfsf13*$^{-/-}$ mice[29] compared to wild-type mice, suggesting that TNFSF13 has an important role in T cell–dependent antibody responses. TNFSF13 also seems to be important for IgA class switch recombination[30]. Local expression of TNFSF13 was found in glomeruli and tubulointerstitium of human proliferative lupus nephritis, accompanied by a prominent accumulation of CD68-positive macrophages in glomeruli[31]. Interventions in the TNFSF13-TNFSF13B system ameliorated proteinuria and renal inflammation in a lupus mouse model[32]. In addition, infiltration of CD68-positive macrophages in glomeruli has been considered a useful marker for evaluating the activity of IgAN[33].

The associations of the *DQB*0302* allele and *DRB*0901-DQB*0303* and *DQA*0201-DQB*0201* haplotypes are consistent with the previously reported association of *DQB*[16,17]. Furthermore, *DRB*0901-DQB*0303* has been identified as a common haplotype in the Chinese

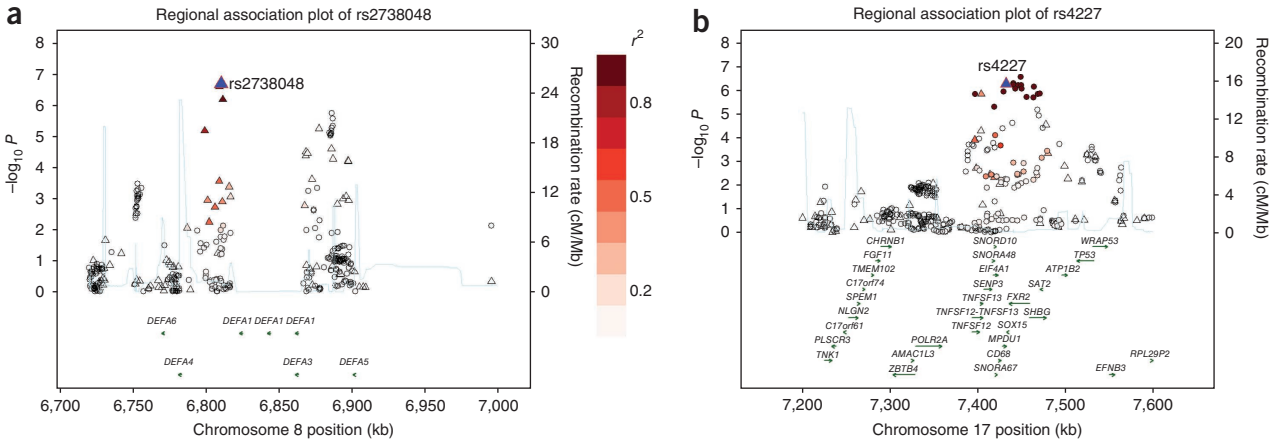

Figure 2 Regional plots of the association results from the discovery sample and recombination rates within the two new susceptibility loci. (**a**) 8p23 region. (**b**) 17p13 region. The –log10 *P* values (*y* axis) of both genotyped (triangles) and imputed (circles) SNPs are presented against their chromosomal positions (*x* axis). The most significant SNP within the region is indicated by a blue triangle, and its LD (r^2) values with nearby SNPs are indicated by different gradients of red. The recombination rate (cM/Mb) (estimated using the HapMap CHB+JPT samples) is shown with blue lines, and the genes within the region are shown as green arrows.

population, and its imputed frequency in this study (15%) is similar to that (14%) previously determined by HLA typing[34,35]. Notably, *DRB*0901-DQB*0303* has been reported to be associated with type 1 diabetes[36] and microscopic polyangiitis[37]. These two diseases are both associated with renal complications, and end-stage renal disease can occur in up to 20% of individuals with Wegener's granulomatosis and microscopic polyangiitis[38]. The association of *HLA-B* has been reported, but the susceptibility allele *B*4001* identified in this study is different from the previously reported *Bw35* allele identified by HLA serotyping[39,40]. Imputation of HLA alleles in the Chinese population remains challenging, and further study by HLA type is required to confirm the current findings based on imputation.

We have discovered two new associations at 17p13 and 8p23 and confirmed previously reported associations within the MHC and 22q12 regions, and we have further demonstrated that genetic polymorphisms involved in immunity and inflammation can influence both susceptibility to and clinical manifestation of IgAN.

URLs. HGDP Selection Browser, http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/; R statistical software, http://www.R-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to all the subjects and healthy volunteers who participated in this work. We thank the staff in The First Affiliated Hospital of Sun Yat-sen University for help with sample collection, DNA extraction and sample storage; W.Y. Meah, X. Chen, H.B. Toh, C.H. Wong, P.L. Ng and I.D. Irwan for performing the genotyping; Z.X. Zhang and Z. B. Hu for providing the control samples in the validation study and R. Johnson for reviewing the manuscript and providing constructive suggestions. This work was funded by the Major State Basic Research Development Program of China (973 program) (2012CB517700-2012CB517706), Key Program of National Natural Science Foundation of China (81130012), National Natural Science Foundation of China for Distinguished Young Scholars (30925019) and the Agency for Science & Technology and Research of Singapore (A*STAR) (to J.-J.L.).

AUTHOR CONTRIBUTIONS

H.Z., Z.-H.L., M.L., X.W., J.-Q.W., J.C., R.-S.L., J.-X.W., Z.-S.L., L.Z., T.-Q.L. and X.-J.H. performed clinical characterization and recruitment of subjects and contributed samples; X.W., X.-Q.T., Z.-J.L., L.F. and W.W. prepared DNA; M.L. and X.-Y.Y. performed experiments. J.-J.L., H.-Q.L., K.-S.S. and M.L. analyzed data; Y.L., J.-N.F. and L.-D.S. contributed to analytical support and discussion; X.-Q.Y. and X.-J.Z. provided the platform and organized the study; J.-J.L., X.-Q.Y. and M.L. prepared the manuscript; J.-J.L. and X.-Q.Y. conceived and supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. D'Amico, G. The commonest glomerulonephritis in the world: IgA nephropathy. *Q. J. Med.* **64**, 709–727 (1987).
2. Barratt, J. & Feehally, J. IgA nephropathy. *J. Am. Soc. Nephrol.* **16**, 2088–2097 (2005).
3. Donadio, J.V. & Grande, J.P. IgA nephropathy. *N. Engl. J. Med.* **347**, 738–748 (2002).
4. Levy, M. & Berger, J. Worldwide perspective of IgA nephropathy. *Am. J. Kidney Dis.* **12**, 340–347 (1988).
5. Tsukamoto, Y. *et al.* Report of the Asian Forum of Chronic Kidney Disease Initiative (AFCKDI) 2007. "Current status and perspective of CKD in Asia": diversity and specificity among Asian countries. *Clin. Exp. Nephrol.* **13**, 249–256 (2009).
6. Hsu, S.I., Ramirez, S.B., Winn, M.P., Bonventre, J.V. & Owen, W.F. Evidence for genetic factors in the development and progression of IgA nephropathy. *Kidney Int.* **57**, 1818–1835 (2000).
7. Hsu, S.I. Racial and genetic factors in IgA nephropathy. *Semin. Nephrol.* **28**, 48–57 (2008).
8. Beerman, I., Novak, J., Wyatt, R.J., Julian, B.A. & Gharavi, A.G. The genetics of IgA nephropathy. *Nat. Clin. Pract. Nephrol.* **3**, 325–338 (2007).
9. Izzi, C. *et al.* Familial aggregation of primary glomerulonephritis in an Italian population isolate: Valtrompia study. *Kidney Int.* **69**, 1033–1040 (2006).
10. Scolari, F. Inherited forms of IgA nephropathy. *J. Nephrol.* **16**, 317–320 (2003).
11. Kiryluk, K. *et al.* Genetic studies of IgA nephropathy: past, present, and future. *Pediatr. Nephrol.* **25**, 2257–2268 (2010).
12. Bisceglia, L. *et al.* Genetic heterogeneity in Italian families with IgA nephropathy: suggestive linkage for two novel IgA nephropathy loci. *Am. J. Hum. Genet.* **79**, 1130–1134 (2006).
13. Gharavi, A.G. *et al.* IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22–23. *Nat. Genet.* **26**, 354–357 (2000).
14. Paterson, A.D. *et al.* Genome-wide linkage scan of a large family with IgA nephropathy localizes a novel susceptibility locus to chromosome 2q36. *J. Am. Soc. Nephrol.* **18**, 2408–2415 (2007).
15. Cao, H.X. *et al.* Human leukocyte antigen DRB1 alleles predict risk and disease progression of immunoglobulin A nephropathy in Han Chinese. *Am. J. Nephrol.* **28**, 684–691 (2008).
16. Fennessy, M. *et al.* HLA-DQ gene polymorphism in primary IgA nephropathy in three European populations. *Kidney Int.* **49**, 477–480 (1996).
17. Raguénès, O., Mercier, B., Cledes, J., Whebe, B. & Ferec, C. HLA class II typing and idiopathic IgA nephropathy (IgAN): DQB1*0301, a possible marker of unfavorable outcome. *Tissue Antigens* **45**, 246–249 (1995).
18. Yu, H.H. *et al.* Genetics and immunopathogenesis of IgA nephropathy. *Clin. Rev. Allergy Immunol.* **41**, 198–213 (2011).
19. Gharavi, A.G. *et al.* Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat. Genet.* **43**, 321–327 (2011).
20. Alexopoulos, E., Apostolos, K. & Papadimitriou, M. Increased glomerular and interstitial LFA-1 expression in proteinuric immunoglobulin A nephropathy. *Am. J. Kidney Dis.* **27**, 327–333 (1996).
21. Gupta, S.K. *et al.* Proteinuria, creatinine clearance, and immune activation in antiretroviral-naive HIV-infected subjects. *J. Infect. Dis.* **200**, 614–618 (2009).
22. Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* **3**, 710–720 (2003).
23. Rehaume, L.M. & Hancock, R.E. Neutrophil-derived defensins as modulators of innate immune function. *Crit. Rev. Immunol.* **28**, 185–200 (2008).
24. Yokoyama, H. *et al.* Urinary levels of chemokines (MCAF/MCP-1, IL-8) reflect distinct disease activities and phases of human IgA nephropathy. *J. Leukoc. Biol.* **63**, 493–499 (1998).
25. Stangou, M. *et al.* Urinary levels of epidermal growth factor, interleukin-6 and monocyte chemoattractant protein-1 may act as predictor markers of renal function outcome in immunoglobulin A nephropathy. *Nephrology (Carlton)* **14**, 613–620 (2009).
26. Saraheimo, M. *et al.* Increased levels of alpha-defensin (-1, -2 and -3) in type 1 diabetic patients with nephropathy. *Nephrol. Dial. Transplant.* **23**, 914–918 (2008).
27. Vordenbäumen, S. *et al.* Elevated levels of human beta-defensin 2 and human neutrophil peptides in systemic lupus erythematosus. *Lupus* **19**, 1648–1653 (2010).
28. Stein, J.V. *et al.* APRIL modulates B and T cell immunity. *J. Clin. Invest.* **109**, 1587–1598 (2002).
29. Castigli, E. *et al.* Impaired IgA class switching in APRIL-deficient mice. *Proc. Natl. Acad. Sci. USA* **101**, 3903–3908 (2004).
30. He, B. *et al.* Intestinal bacteria trigger T cell-independent immunoglobulin A_2 class switching by inducing epithelial-cell secretion of the cytokine APRIL. *Immunity* **26**, 812–826 (2007).
31. Neusser, M.A. *et al.* Intrarenal production of B-cell survival factors in human lupus nephritis. *Mod. Pathol.* **24**, 98–107 (2011).
32. Ramanujam, M. *et al.* Similarities and differences between selective and nonselective BAFF blockade in murine SLE. *J. Clin. Invest.* **116**, 724–734 (2006).
33. Maruhashi, Y. *et al.* Analysis of macrophages in urine sediments in children with IgA nephropathy. *Clin. Nephrol.* **62**, 336–343 (2004).
34. Hei, A.L. *et al.* Analysis of high-resolution HLA-A, -B, -Cw, -DRB1, and -DQB1 alleles and haplotypes in 718 Chinese marrow donors based on donor-recipient confirmatory typings. *Int. J. Immunogenet.* **36**, 275–282 (2009).
35. Trachtenberg, E. *et al.* HLA class I (A, B, C) and class II (DRB1, DQA1, DQB1, DPB1) alleles and haplotypes in the Han from southern China. *Tissue Antigens* **70**, 455–463 (2007).
36. Kawabata, Y. *et al.* Differential association of HLA with three subtypes of type 1 diabetes: fulminant, slowly progressive and acute-onset. *Diabetologia* **52**, 2513–2521 (2009).
37. Tsuchiya, N., Kobayashi, S., Hashimoto, H., Ozaki, S. & Tokunaga, K. Association of HLA-DRB1*0901–DQB1*0303 haplotype with microscopic polyangiitis in Japanese. *Genes Immun.* **7**, 81–84 (2006).
38. Geetha, D. & Seo, P. Renal transplantation in the ANCA-associated vasculitides. *Am. J. Transplant.* **7**, 2657–2662 (2007).
39. Berthoux, F.C. *et al.* HLA-Bw35 and mesangial IgA glomerulonephritis. *N. Engl. J. Med.* **298**, 1034–1035 (1978).
40. Berthoux, F.C., Alamartine, E., Pommier, G. & Lepetit, J.C. HLA and IgA nephritis revisited 10 years later: HLA-B35 antigen as a prognostic factor. *N. Engl. J. Med.* **319**, 1609–1610 (1988).

ONLINE METHODS

Study subjects. The genome-wide discovery analysis involved 1,523 cases (from southern China) and 4,276 controls (972 controls from southern China, 1,228 controls from northern China and 2,076 Chinese controls from Singapore who share the same ancestral origin as the other Chinese controls[41]). For the validation study, two independent case-control samples were recruited from northern (replication 1: 1,402 cases and 1,716 controls) and southern (replication 2: 1,301 cases and 1,748 controls) China.

All cases were histopathologically diagnosed by 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[42], and (ii) excluding individuals with cirrhosis, Henoch-Schönlein purpura nephritis, hepatitis B–associated glomerulonephritis, HIV infection and systemic lupus erythematosus[2]. Clinical information was collected from 3,394 individuals at diagnosis, including serum IgA level, Lee's pathological grade, history of gross hematuria, degree of microhematuria, proteinuria, eGFR (estimated glomerular filtration rate using the MDRD formula) and chronic kidney disease (CKD) stage. On the basis of the collected clinical information, 1,950 IgAN cases were divided into three clinical subtypes: (i) asymptomatic hematuria or proteinuria: urine protein < 3.5 g in 24 h, Scr < 1.3 mg dl^{-1} and without edema or hypertension; (ii) nephritic syndrome: urine protein < 3.5 g in 24 h, with or without Scr > 1.3 mg dl^{-1}, edema or hypertension; and (iii) nephrotic syndrome: urine protein ≥ 3.5 g in 24 h. All controls were either self-reported (as free of major kidney diseases) or clinically verified healthy controls. The 2,535 clinically verified healthy controls, including 1,763 Singaporean Chinese controls and 772 Chinese controls, had normal urinalysis (without red blood cells and protein in urine), normal serum creatinine levels and eGFR > 60 ml per minute per 1.73 m^2. Gender and age information were collected from both cases and controls through questionnaires, and the cases and controls were matched for geographical origin and ethnicity. The baseline characteristics of the case and control samples are summarized in **Table 1**.

The study was approved by the Institutional Review Board at The First Affiliated Hospital of Sun Yat-sen University. Informed consent was obtained from all of the participants.

Sample preparation, genotyping and quality controls in GWAS. Genomic DNA was isolated from whole blood using a commercial DNA extraction kit (Qiagen) and quantified using Picogreen reagent (Invitrogen). Genotyping analysis of the discovery samples was conducted using Human660-Quad (1,523 cases), Human610-Quad (972 southern and 1,228 northern Chinese controls and the 1,146 Singaporean Chinese controls) and Human1M-Duo (930 Singaporean Chinese controls) BeadChips (Illumina).

All the SNPs on the X, Y and mitochondrial chromosomes as well as the copy number variation–related probes were excluded from analysis. In addition, two SNPs with bad clusters were removed, and 27,843 SNPs were excluded because they had either a call rate <90%, a minor allele frequency (MAF) <1% or significant deviation from the Hardy-Weinberg equilibrium (HWE) in the controls ($P < 10^{-6}$). Consequently, 444,882 common autosomal SNPs remained for association analysis. Furthermore, 44 cases with a SNP call rate <98% were removed. We examined the genetic relatedness of all the samples using pairwise identity by state (IBS)-based analysis in PLINK (v1.07)[43]. We identified 43 first- or second-degree relative pairs (all cases), and for each pair, the sample with lower genotype call rate was removed. We then used a PCA-based method[44] to detect population outliers and stratification. All the SNPs within the five regions of long-range LD were excluded from the PCA analysis as described[41], including the HLA region on chromosome 6, inversions on chromosomes 8 and 5, and two regions on chromosome 11. For the initial PCA analysis, all 5,712 samples (1,436 cases and 4,276 controls) were analyzed together with the 194 reference samples (CHB, 43; JPT, 42; CEU, 56; YRI, 53) from the HapMap Project. We identified and excluded eight population outliers (two cases and six controls) showing more than 3 s.d. away from the means of PC1 and PC2. The remaining 5,704 samples of Chinese origin were then used in the second PCA analysis to detect population stratification (**Supplementary Fig. 1**). Finally, we retained 444,882 common autosomal SNPs (442,238 SNPs outside the MHC region) in 1,434 cases and 4,270 controls for genome-wide association analysis.

To generate genetically homogeneous clusters, we performed *K*-means clustering[45] to partition 5,704 samples into three clusters. On the basis of the samples' PC statistics (PC1-10), three initial means were randomly selected from the data set and, subsequently, three clusters were created by associating every observation with the nearest mean.

SNP selection for validation. The SNP selection for validation was based on the results of the genome-wide association analysis with PCA-based (PC1-5) correction for population stratification. The genome-wide discovery analysis identified 557 SNPs whose *P* values of association were $<5 \times 10^{-4}$. The cluster plots of all the 557 SNPs were visually inspected to verify the quality of genotype calling. SNPs with ambiguous genotype scatter plots were excluded from further analysis. The remaining SNPs were then grouped into distinct genomic loci on the basis of their physical locations and regional patterns of LD. A candidate locus was defined as an LD region in which at least one SNP has $P < 5 \times 10^{-4}$. Then, the top one or two SNPs with the lowest *P* value were selected from each of the identified loci. In total, we selected 75 SNPs (70 non-MHC SNPs and 5 MHC SNPs) for the validation analysis.

Genotyping and quality controls in the validation study. Genotyping analysis of the 75 SNPs selected for validation was performed by using either the MassArray system from Sequenom (for the non-MHC SNPs) or the TaqMan SNP genotyping assay from ABI (for the SNPs within the MHC region). Locus-specific PCR and detection primers were designed using the MassArray Assay Design 3.0 software (Sequenom). TaqMan assays (predesigned or customized) were obtained from Applied Biosystems. TaqMan reactions were carried out in 5-μl volumes containing 10–20 ng DNA according to the manufacturer's protocols. Fluorescence data were obtained in the ABI PRISM 7900HT, and SDS 2.4 software (Applied Biosystems) was used to call genotypes. The same SNP-filtering criteria as in the genome-wide analysis was applied, and all individuals with a SNP call rate of <90% were removed from further analysis. After quality control, 61 (56 non-MHC and 5 MHC) SNPs were left for further analysis. For all 61 SNPs, we examined the clustering patterns of genotypes from TaqMan and Sequenom assays and confirmed that the genotypes were of good quality.

Association analysis. Genome-wide association analysis was performed using the trend test and two different approaches for controlling population stratification. First, the analysis was performed in PLINK by using a logistic regression model, where the first five principal components were included as covariates to adjust for population stratification. Second, the analysis was performed by including three clusters as covariates in the logistic regression model. The Manhattan plot of $-\log_{10} P$ was generated using Haploview (v4.2)[46]. The quantile-quantile plot was generated using R[47] to evaluate the overall significance of the genome-wide associations and the potential impact of population stratification.

For the replication study, we analyzed 61 SNPs in each of the two replication samples by performing the same trend test in a logistic regression model. In the joint analysis of the replication 1 and 2 samples, a study indicator (for replication 1 and 2) was added into the model as a covariate. To combine the association evidences from the GWAS and the two replication samples, we treated the GWAS sample and the two independent validation samples as independent studies, and used the logistic regression model, which included the study indicator (GWAS, replication 1 and 2) and the first five principal components (the five principal components were set to zero for the two validation samples) as covariates. For all the validated SNPs, the associations were further analyzed with adjustment for age and gender (as covariates in logistic regression analysis).

OR values were measured as OR per allele and presented for the minor allele of a SNP unless otherwise stated. An independence test of association was carried out in a conditional logistic regression analysis implemented in R[47].

Haplotype analysis. Haplotypes of the top independent SNPs within the MHC were constructed using Beagle[48]. The most likely haplotype pair was taken for each subject. Haplotype-based association analysis was carried out using the trend test in the logistic regression model, which included haplotypes as variables, and the study indicator, age, gender and PC1-5 (adjustment for population stratification) as covariates.

doi:10.1038/ng.1047

Analysis of clinical phenotypes. The seven validated susceptibility SNPs were studied for association with clinical phenotypes by using different regression methods. Linear regression was applied for the correlation analysis of quantitative phenotypes including natural log–transformed serum IgA, proteinuria and eGFR levels. Binary logistic regression was carried out for the correlation analysis of the history of gross hematuria. Ordinal logistic regression was performed for the correlation analysis of clinical subtype, degree of microhematuria, CKD stage and Lee's biopsy grade at the time of diagnosis by biopsy. The analyses of clinical subtype and CKD stage were also performed using a χ^2 test, and the results remained the same (data not shown). The means and 95% confidence intervals of serum IgA, proteinuria and eGFR level were estimated for the patient groups carrying different genotypes of SNPs. All analysis was performed using R[47].

Imputation analysis. We combined the CHB and JPT phased haplotypes and the HLA allele data set from ref. 49 as the reference panel to impute HLA-A, HLA-B, HLA-C, DQA, DQB and DRB alleles at four-digit resolution. We adopted a similar approach[50] to directly impute the HLA allele probability and dosage using the BEAGLE program[48]. For imputation of untyped SNPs within the MHC (Chr6: 28,000–34,000 kb), we used IMPUTE version 2 (refs. 51,52). We imputed 5,704 samples by using 1,881 common genotyped SNPs (among all the case and control samples of GWAS) whose genotypes all passed the quality control thresholds (call rate >90%, MAF >1%, HWE $P > 10^{-6}$ in controls) in the individual data sets of the cases and four groups controls. Imputation was performed using the HapMap reference data (HapMap phase II + phase III, CHB+JPT data)[53] for the HLA region. Imputed genotypes with probability <90% were excluded, and SNPs with imputation information <80%, MAF <5% and missing rate >10% of genotypes were dropped from further analysis. Finally, 5,899 imputed SNPs passed quality control and were retained for association analysis. Imputation analysis of the two new loci 17p13 and 8p23 as well as the reported locus 1q32 was performed using the IMPUTE version 2 and the 1000 Genome[54] + HapMap phase III reference data (CHB+JPT data). Quality controls of imputed SNPs were carried out as described above.

Association testing was carried out using a logistic regression model, with the first five principal components included as covariates to adjust for population stratification. All the locus regional plots were generated in R using the recombination rate information (Build 35) from the HapMap project (http://hapmap.ncbi.nlm.nih.gov/downloads/recombination/).

41. Chen, J., Niaudet, P. & Levy, M. Schonlein-Henoch purpura and IgA nephropathy. in *Renal Pathology* (eds. C. Tisher & B. Brenner) 507–508 (Lippincott, Philadelphia, 1994).
42. Niaudet, P. & Levy, M. Schonlein-Henoch purpura and IgA nephropathy. in *Renal Pathology* (Tisher, C. & Brenner, B., eds.) 507–508 (Lippincott, Philadelphia, 1994).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
45. Steinhaus, H. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci.* (in French) **4**, 801–804 (1957).
46. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
47. R Development Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* (Vienna, 2008).
48. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
49. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
50. Pereyra, F. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551–1557 (2010).
51. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
52. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
53. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
54. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).