

# Geographic Differences in Genetic Susceptibility to IgA Nephropathy: GWAS Replication Study and Geospatial Risk Analysis

Krzysztof Kiryluk<sup>1</sup>, Yifu Li<sup>1</sup>, Simone Sanna-Cherchi<sup>1,2</sup>, Mersedeh Rohanizadegan<sup>1</sup>, Hitoshi Suzuki<sup>3</sup>, Frank Eitner<sup>4</sup>, Holly J. Snyder<sup>1</sup>, Murim Choi<sup>5</sup>, Ping Hou<sup>6</sup>, Francesco Scolari<sup>7</sup>, Claudia Izzi<sup>7</sup>, Maddalena Gigante<sup>8</sup>, Loreto Gesualdo<sup>8</sup>, Silvana Savoldi<sup>9</sup>, Antonio Amoroso<sup>10</sup>, Daniele Cusi<sup>11</sup>, Pasquale Zamboli<sup>12</sup>, Bruce A. Julian<sup>13</sup>, Jan Novak<sup>13</sup>, Robert J. Wyatt<sup>14,15</sup>, Krzysztof Mucha<sup>16</sup>, Markus Perola<sup>17,18,19</sup>, Kati Kristiansson<sup>17</sup>, Alexander Viktorin<sup>20</sup>, Patrik K. Magnusson<sup>20</sup>, Gudmar Thorleifsson<sup>21,22</sup>, Unnur Thorsteinsdottir<sup>21,22</sup>, Kari Stefansson<sup>21,22</sup>, Anne Boland<sup>23</sup>, Marie Metzger<sup>24</sup>, Lise Thibaudin<sup>25</sup>, Christoph Wanner<sup>26</sup>, Kitty J. Jager<sup>27</sup>, Shin Goto<sup>28</sup>, Dita Maixnerova<sup>29</sup>, Hussein H. Karnib<sup>30</sup>, Judit Nagy<sup>31</sup>, Ulf Panzer<sup>32</sup>, Jingyuan Xie<sup>33</sup>, Nan Chen<sup>33</sup>, Vladimir Tesar<sup>29</sup>, Ichiei Narita<sup>28</sup>, Francois Berthoux<sup>25</sup>, Jürgen Floege<sup>4</sup>, Benedicte Stengel<sup>24</sup>, Hong Zhang<sup>6</sup>, Richard P. Lifton<sup>5</sup>, Ali G. Gharavi<sup>1\*</sup>

**1** Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America, **2** Department of Medicine, St. Luke's-Roosevelt Hospital Center, New York, New York, United States of America, **3** Division of Nephrology, Department of Internal Medicine, Juntendo University Faculty of Medicine, Tokyo, Japan, **4** Department of Nephrology, RWTH University of Aachen, Aachen, Germany, **5** Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, United States of America, **6** Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China, **7** University of Brescia and Second Division of Nephrology, Montichiari Hospital, Montichiari, Italy, **8** Section of Nephrology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy, **9** Nephrology and Dialysis Unit, Ciriè Hospital, Torino, Italy, **10** Department of Genetics, Biology, and Biochemistry, University of Torino, Torino, Italy, **11** Renal Division, DMCO, San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy, **12** Department of Nephrology, Second University of Naples, Naples, Italy, **13** Departments of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **14** Children's Foundation Research Center, Le Bonheur Children's Hospital, Memphis, Tennessee, United States of America, **15** Division of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, Tennessee, United States of America, **16** Department of Immunology, Transplantation, and Internal Medicine, Warsaw Medical University, Warsaw, Poland, **17** Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **18** Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland, **19** The Estonian Genome Center, University of Tartu, Tartu, Estonia, **20** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **21** deCODE Genetics, Reykjavik, Iceland, **22** Faculty of Medicine, University of Iceland, Reykjavik, Iceland, **23** Centre National de Génotypage, CEA, Institut de Génétique, Evry, France, **24** INSERM, Centre for Research in Epidemiology and Population Health, Villejuif, France, **25** Nephrology, Dialysis, and Renal Transplantation Department, University North Hospital, Saint Etienne, France, **26** Department of Medicine, Division of Nephrology, University Hospital, Würzburg, Germany, **27** ERA-EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **28** Division of Clinical Nephrology and Rheumatology, Niigata University, Niigata, Japan, **29** Department of Nephrology, 1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic, **30** Clinical Research Center and Division of Nephrology, Al Rassoul Al-Aazam Hospital, Beirut, Lebanon, **31** Nephrology Center and 2nd Department of Internal Medicine, Medical Faculty, University of Pécs, Pécs, Hungary, **32** III Medizinische Klinik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany, **33** Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

## Abstract

IgA nephropathy (IgAN), major cause of kidney failure worldwide, is common in Asians, moderately prevalent in Europeans, and rare in Africans. It is not known if these differences represent variation in genes, environment, or ascertainment. In a recent GWAS, we localized five IgAN susceptibility loci on Chr.6p21 (*HLA-DQB1/DRB1*, *PSMB9/TAP1*, and *DPA1/DPB2* loci), Chr.1q32 (*CFHR3/R1* locus), and Chr.22q12 (*HORMAD2* locus). These IgAN loci are associated with risk of other immune-mediated disorders such as type I diabetes, multiple sclerosis, or inflammatory bowel disease. We tested association of these loci in eight new independent cohorts of Asian, European, and African-American ancestry ( $N = 4,789$ ), followed by meta-analysis with risk-score modeling in 12 cohorts ( $N = 10,755$ ) and geospatial analysis in 85 world populations. Four susceptibility loci robustly replicated and all five loci were genome-wide significant in the combined cohort ( $P = 5 \times 10^{-32} - 3 \times 10^{-10}$ ), with heterogeneity detected only at the *PSMB9/TAP1* locus ( $I^2 = 0.60$ ). Conditional analyses identified two new independent risk alleles within the *HLA-DQB1/DRB1* locus, defining multiple risk and protective haplotypes within this interval. We also detected a significant genetic interaction, whereby the odds ratio for the *HORMAD2* protective allele was reversed in homozygotes for a *CFHR3/R1* deletion ( $P = 2.5 \times 10^{-4}$ ). A seven-SNP genetic risk score, which explained 4.7% of overall IgAN risk, increased sharply with Eastward and Northward distance from Africa ( $r = 0.30$ ,  $P = 3 \times 10^{-128}$ ). This model paralleled the known East–West gradient in disease risk. Moreover, the prediction of a South–North axis was confirmed by registry data showing that the prevalence of IgAN-attributable kidney failure is increased in Northern Europe, similar to multiple sclerosis and type I diabetes. Variation at IgAN susceptibility loci correlates with differences in disease prevalence among world populations. These findings inform genetic, biological, and epidemiological investigations of IgAN and permit cross-comparison with other complex traits that share genetic risk loci and geographic patterns with IgAN.

**Citation:** Kiryluk K, Li Y, Sanna-Cherchi S, Rohanizadegan M, Suzuki H, et al. (2012) Geographic Differences in Genetic Susceptibility to IgA Nephropathy: GWAS Replication Study and Geospatial Risk Analysis. *PLoS Genet* 8(6): e1002765. doi:10.1371/journal.pgen.1002765

**Editor:** Mark I. McCarthy, University of Oxford, United Kingdom

**Received:** December 5, 2011; **Accepted:** April 30, 2012; **Published:** June 21, 2012

**Copyright:** © 2012 Kiryluk et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by the Center for Glomerular Diseases at Columbia University and NIH/NIDDK grants K23DK090207 (K Kiryluk), RC1DK087445 (AG Gharavi, RP Lifton), and R01DK082753 (AG Gharavi, J Novak, BA Julian, RJ Wyatt). S Sanna-Cherchi is supported by the AHA Scientist Development Grant 0930151N and the ASN Carl W Gottschalk Research Scholar Grant. The GN-PROGRESS study was supported by grants from the Ministry of Health (PHRC AOM 00022), the Ministry of Environment (Décision d'aide EN00D08), the Ministry of Research (Décision d'aide 01P0513), and the Biomedicine Agency (AO Recherche et Greffes 2005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: ag2239@columbia.edu

## Introduction

IgA nephropathy (IgAN) is a common kidney disease with a complex genetic determination. This disorder is diagnosed based on detection of mesangial proliferation and glomerular deposits of IgA1. Most frequently, IgAN has a progressing course and 20–50% of cases develop end-stage renal disease (ESRD) within 20 years of follow-up [1]. The disease has been detected among all ethnicities worldwide, but displays a striking geographic variation. It is the most common cause of kidney failure in East Asian countries, has intermediate prevalence in European and US populations but is rarely reported in populations of African ancestry. The diagnosis of IgAN requires a kidney biopsy, complicating accurate determination of heritability and population prevalence of disease. Autopsy and donor biopsy series suggest a prevalence of up to 1.3% in Finland [2] and 3.7% in Japan [3]. Familial aggregation of IgAN has also been recognized throughout the world [4,5,6,7,8,9,10,11] and up to 14% of cases may be familial [8]. Moreover, family members frequently have aberrant glycosylation of the hinge region of circulating IgA1, a defect with an estimated heritability of 40–50% [12,13]. These data suggest a strong genetic contribution to disease.

Recently, we have completed a large-scale genome-wide association study (GWAS) involving a cohort of 3,144 sporadic IgAN cases [14]. The discovery phase samples (1,194 cases and 902 controls) were recruited in Beijing, China and were comprised of individuals of Han Chinese ancestry. The most associated SNPs were then followed up in additional cohorts of Han Chinese and Europeans (1,950 cases and 1,920 controls). In the combined analysis, we discovered 5 novel susceptibility loci with consistent effects across individual cohorts. These include 3 distinct intervals in the MHC-II region on chromosome 6p21, with the strongest signal encompassing the HLA *DQB1/DQA1/DRB1* locus (abbreviated as *DQB1/DRB1* hereafter). Imputation of classical alleles suggested that this signal was partially conveyed by a strong protective effect of the *DRB1\*1501-DQB1\*0602* haplotype. The second signal on Chr. 6p21 encompassed a ~100 Kb region containing *TAP2*, *TAP1*, *PSMB8*, and *PSMB9* genes (*TAP2/PSMB9* locus) and the third signal on Chr. 6p21 contained the HLA *DPA1/DPB1/DPB2* genes (*DPA1/DPB2* locus). Independence of these three regions on Chr. 6p21 was demonstrated by their localization within distinct LD blocks as well as genome-wide significant associations after rigorous conditional analyses. We also detected significant association within the *Complement factor H (CFH)* gene cluster on Chr. 1q32, where alleles tagging a common deletion in the *CFHR3* and *CFHR1* genes imparted a significant protective effect (*CFHR3/R1* locus). Finally, a fifth signal centered on the *HORMAD2* gene on Chr. 22q12 and containing multiple genes demonstrated significant association with risk of IgAN (*HORMAD2* locus). These five loci individually conferred a

moderate risk of disease (OR 1.25–1.59), but together explained 4–5% of the variation in risk across the populations examined.

To follow-up these studies and better assess the risk imparted by susceptibility alleles in diverse populations, we performed a replication study in eight independent case-control cohorts and performed a meta-analysis of all available genetic data including the original GWAS, totaling in 10,755 individuals. The expanded sample size allowed us to formally assess locus heterogeneity, identify new independent risk variants by conditional analyses and search for first-order genetic interactions. Finally, we refined a genetic risk score for IgAN and analyzed differences in the distributions of the IgAN susceptibility alleles among the major world populations.

## Results

### Replication Study

For replication we examined eight cohorts (five European, two East Asian, and one African-American cohort, totaling 2,228 cases and 2,561 controls, described in Table S1). While each individual cohort at best had 40–50% power to replicate original GWAS findings, the combined replication cohort (2,228 cases and 2,561 controls) provided essentially 100% power for replication across the range of allele frequencies and odds ratios initially observed (Table S2).

We genotyped the two top-scoring SNPs for the *CFHR3/R1*, *TAP2/PSMB9*, *DPA1/DPB2*, and *HORMAD2* loci, but four SNPs were included for the *DQB1/DRB1* locus to test for independent alleles at this interval by conditional analysis. After a standard assessment of genotype quality control, we performed association testing within each cohort using the standard Cochran-Armitage trend test (Table S3). We also tested for heterogeneity of associations and performed a meta-analysis under both fixed and random effects models (Table 1).

Four of the five original GWAS loci displayed significant replication with direction-consistent ORs and no heterogeneity comparable to the original findings (Table 1). The strongest replication was at the *DQB1/DRB1* locus and achieved genome-wide significance in the replication cohort (fixed effects OR 0.75, P-value  $4 \times 10^{-11}$ ). The *CFHR3/R1* locus on Chr.1q32, the *HORMAD2* locus on Chr.22q12, and the *DPA1/DPB2* locus on Chr.6p21 were also robustly replicated (fixed effects p-values  $3 \times 10^{-3}$ – $7 \times 10^{-7}$ ), with minimal between-cohort heterogeneity ( $I^2 < 25\%$ ). Accordingly, when combined with the four cohorts studied in the original GWAS, these four loci provided highly significant evidence of association (fixed effects p-values  $3 \times 10^{-10}$ – $5 \times 10^{-32}$ ).

In contrast, the *TAP2/PSMB9* locus on Chr. 6p21 displayed direction-consistent replication only in the Italian, German, Czech, and Japanese cohort but the full replication cohort did

## Author Summary

IgA nephropathy (IgAN) is the most common cause of kidney failure in Asia, has lower prevalence in Europe, and is very infrequent among populations of African ancestry. A long-standing question in the field is whether these differences represent variation in genes, environment, or ascertainment. In a recent genome-wide association study of 5,966 individuals, we identified five susceptibility loci for this trait. In this paper, we study the largest IgAN case-control cohort reported to date, composed of 10,775 individuals of European, Asian, and African-American ancestry. We confirm that all five loci are significant contributors to disease risk across this multi-ethnic cohort. In addition, we identify two novel independent susceptibility alleles within the HLA-DQB1/DRB1 locus and a new genetic interaction between loci on Chr.1p36 and Chr.22q22. We develop a seven-SNP genetic risk score that explains nearly 5% of variation in disease risk. In geospatial analysis of 85 world populations, the genetic risk score closely parallels worldwide patterns of disease prevalence. The genetic risk score also predicts an unsuspected Northward risk gradient in Europe. This genetic prediction is verified by examination of registry data demonstrating, similarly to other immune-mediated diseases such as multiple sclerosis and type I diabetes, a previously unrecognized increase in IgAN-attributable kidney failure in Northern European countries.

not support this association (Table 1, Table S3). However, when combined with the four cohorts from the original GWAS, this locus remained genome-wide significant (fixed effects p-values  $1 \times 10^{-8}$  and  $6 \times 10^{-10}$  for rs9357155 and rs2071543, respectively, Table 1). As expected,  $I^2$  and Q-tests provided evidence of heterogeneity and random effects meta-analysis, which explicitly models heterogeneity, was 1–3 orders of magnitude more significant than fixed effect meta-analysis at this interval (e.g. random effects p-value  $3 \times 10^{-11}$ ,  $I^2 = 61\%$  for rs9357155; Table 1). The heterogeneity was not attributable to differences in ethnicity or cohort size as the association results varied within Asian and European cohorts of differing size (Table S3).

## Conditional Analysis Reveals New Independent Risk Alleles within the HLA-DQB1/DRB1 Locus

The top signals in the original GWAS, represented by rs9275596 and located within the *DQB1/DRB1* locus, were mediated by a very strong protective effect of the *DRB1\*1501-DQB1\*602* haplotype [14]. However, the SNPs in this interval are in incomplete LD and conditional analyses in our GWAS [14] and in an independent study of Europeans [15] had indicated that additional independent haplotypes also contributed to the signal. Therefore, taking advantage of our expanded cohort size, we examined additional SNPs that were in partial LD with rs9275596 to detect potentially independent effects (rs9275224, rs2856717 and rs9275424, which had an  $r^2$  of 0.09 to 0.7 with rs9275596, Table S4).

After mutually conditioning each SNP on the remaining SNPs, three of the four SNPs in the *DQB1/DRB1* region exhibited a genome-wide significant independent effect (rs9275596, rs9275224 and rs2856717, conditioned p-values  $< 5 \times 10^{-8}$ , Table 2). Interestingly, the conditioned effect of the minor allele of rs2856717 was reversed compared to the crude effect estimate, suggesting that the adjustment for LD structure has uncovered a risk haplotype in this region (conditioned OR 1.61,  $p = 2 \times 10^{-10}$ ).

The above data indicated that there are multiple risk haplotypes within the *DQB1/DRB1* locus. To better define these findings, we next phased four-SNP haplotypes at this locus and tested associations with disease (Table 3). We confirmed a very strong protective effect of the ATAC haplotype (freq. 0.21) which, based on our previous imputation analysis, carries the *DRB1\*1501/DQB1\*602* classical alleles. In addition, we defined a new protective haplotype (ACAT, freq. 0.13) and a new risk haplotype (ATAT, freq. 0.05). The ATAC protective haplotype and the ATAT risk haplotype differ only by the rs9275596-C/T allele, explaining the reversal of OR for the rs2856717-T allele after conditioning for rs9275596 (Table 3). Additionally, the GCGT risk haplotype, tagged by the rs9275424-G allele, exhibited a weaker protective effect. These results were supported by both Asian and European cohorts (Table S5). Further support is provided by the global haplotype association test, which achieved a p-value of  $3 \times 10^{-43}$ . Based on these analyses, we concluded that there are at least three independent haplotypes conferring risk of IgAN within this region.

Nonetheless, these 3 independent haplotypes in *DQB1/DRB1* locus still did not explain associations in other Chr. 6p21 regions (*TAP2/PSMB9* and *DPA1/DPB2* loci, respectively represented by rs9357155 and rs1883414), and a fully adjusted model that included all independently associated SNPs continued to support the original GWAS findings of three discrete genome-wide significant intervals on Chr. 6p21 (Table 4).

## First-Order Interaction Screen Reveals Significant Interaction between *CFHR3/R1* and *HORMAD2* Loci

We tested the possibility of interaction between the 7 risk-contributing SNPs and therefore tested for all possible pairwise interactions (Table S6). We detected strong evidence for a multiplicative interaction (defined as departure from additivity on the log-odds scale) between the *CFHR3/R1* (rs6677604) and the *HORMAD2* loci (rs2412971). In this interaction, the rs2412971-A allele has a strong and consistent protective effect among all genotypic subgroups, but its effects are reversed among homozygotes for the rs6677604-A allele, which closely tags a *CFHR3/R1* deletion (Figure 1, Table S6). The significance of this interaction ( $p = 2.5 \times 10^{-4}$ ) exceeds a Bonferroni-corrected threshold for 21 tests, and is most discernable among the European cohorts ( $p = 1.4 \times 10^{-3}$ ), where both SNPs have higher minor allele frequencies. The 4-df genotypic interaction test was also significant for these two loci ( $p = 6.4 \times 10^{-3}$ ), but the 1-df multiplicative interaction model provided a better fit.

## Improved Prediction of Genetic Risk with a Refined Risk Score

The original IgAN risk score model was based on the genotypes of the top scoring SNPs at the 5 independent loci discovered in the GWAS [14]. We refined this risk score by incorporating the newly discovered independent effects of rs9275224 and rs2856717 and the interaction between the *CFHR3/R1* and the *HORMAD2* loci. A stepwise regression algorithm in the entire cohort defined a new risk score that retained the 7 SNPs exhibiting an independent effect as well as the rs6677604\* rs2412971 interaction term (Table 4). When compared with the original GWAS model, the newly refined score was more strongly associated with disease risk and explained a greater proportion of the disease variance in both the replication and the original GWAS dataset (Table 5). Moreover, the refined risk score was a highly significant predictor of disease in each individual replication cohort (Table S7). In all datasets combined, the new risk score explained 4.7% in disease

**Table 1.** Replication Study Results and Combined Meta-Analysis.

Chr	Location (kb)	SNP (minor allele)	Replication Study N = 4,789 across 8 cohorts (2,228 cases/2,561 controls)				Replication and GWAS N = 10,755 across 12 cohorts (5,372 cases/5,383 controls)				Annotation				
			OR	P-value	OR	P-value	OR	P-value	OR	P-value		I <sup>2</sup>	Q-test		
			Fixed Effects				Random Effects <sup>#</sup>								
1	194,918	<b>rs3766404 (C)</b>	0.78	2.5 × 10 <sup>-4</sup>	0.78	4.2 × 10 <sup>-4</sup>	0%	0.84 (NS)	0.78	7.9 × 10 <sup>-8</sup>	0.78	1.3 × 10 <sup>-7</sup>	6%	0.39 (NS)	CFHR3/R1 locus
1	194,953	<b>rs6677604 (A)</b>	0.78	3.1 × 10 <sup>-5</sup>	0.78	5.5 × 10 <sup>-5</sup>	0%	0.48 (NS)	0.74	2.1 × 10 <sup>-13</sup>	0.74	4.6 × 10 <sup>-13</sup>	21%	0.23 (NS)	
6	32,768	<b>rs9275224 (A)</b>	0.75	3.6 × 10 <sup>-11</sup>	0.75	7.1 × 10 <sup>-11</sup>	0%	0.67 (NS)	0.72	8.5 × 10 <sup>-30</sup>	0.72	2.8 × 10 <sup>-29</sup>	0%	0.69 (NS)	DQB1/DRB1 locus
6	32,778	<b>rs2856717 (T)</b>	0.86	1.1 × 10 <sup>-3</sup>	0.86	1.8 × 10 <sup>-3</sup>	0%	0.71 (NS)	0.77	6.6 × 10 <sup>-16</sup>	0.78	7.3 × 10 <sup>-16</sup>	29%	0.16 (NS)	
6	32,779	<b>rs9275424 (G)</b>	1.22	5.0 × 10 <sup>-5</sup>	1.22	8.7 × 10 <sup>-5</sup>	19%	0.27 (NS)	1.28	2.6 × 10 <sup>-14</sup>	1.26	4.6 × 10 <sup>-14</sup>	30%	0.14 (NS)	
6	32,789	<b>rs9275596 (C)</b>	0.75	5.3 × 10 <sup>-9</sup>	0.75	9.5 × 10 <sup>-9</sup>	0%	0.60 (NS)	0.67	5.0 × 10 <sup>-32</sup>	0.67	3.1 × 10 <sup>-32</sup>	43%	0.05 (NS)	
6	32,917	<b>rs9357155 (A)</b>	0.96	5.8 × 10 <sup>-1</sup>	0.97	9.4 × 10 <sup>-2</sup>	54%	0.025*	0.79	1.1 × 10 <sup>-8</sup>	0.87	2.6 × 10 <sup>-11</sup>	70%	1.0 × 10 <sup>-4**</sup>	TAP2/PSMB9 locus
6	32,919	<b>rs2071543 (A)</b>	0.91	1.7 × 10 <sup>-1</sup>	0.92	1.2 × 10 <sup>-1</sup>	43%	0.08 (NS)	0.78	5.7 × 10 <sup>-10</sup>	0.84	4.0 × 10 <sup>-11</sup>	61%	2.0 × 10 <sup>-3**</sup>	
6	33,194	<b>rs1883414 (T)</b>	0.87	3.1 × 10 <sup>-3</sup>	0.87	5.0 × 10 <sup>-3</sup>	0%	0.96 (NS)	0.82	3.0 × 10 <sup>-10</sup>	0.82	5.9 × 10 <sup>-10</sup>	0%	0.86 (NS)	DPA1/DPB2 locus
6	33,205	<b>rs3129269 (T)</b>	0.89	1.1 × 10 <sup>-2</sup>	0.89	1.7 × 10 <sup>-2</sup>	0%	0.75 (NS)	0.83	2.5 × 10 <sup>-9</sup>	0.83	4.6 × 10 <sup>-9</sup>	0%	0.51 (NS)	
22	28,824	<b>rs2412971 (A)</b>	0.81	1.1 × 10 <sup>-6</sup>	0.81	2.1 × 10 <sup>-6</sup>	24%	0.23 (NS)	0.80	4.0 × 10 <sup>-15</sup>	0.80	9.5 × 10 <sup>-15</sup>	12%	0.33 (NS)	HORMAD2 locus
22	28,859	<b>rs2412973 (A)</b>	0.81	6.9 × 10 <sup>-7</sup>	0.81	1.2 × 10 <sup>-6</sup>	29%	0.19 (NS)	0.80	9.9 × 10 <sup>-15</sup>	0.80	2.3 × 10 <sup>-14</sup>	16%	0.29 (NS)	

Combined association results for 12 SNPs, representing 5 independent regions that reached genome-wide significance in the original GWAS. The combined effect estimates (per allele odds ratios) in the replication cohorts were all direction-consistent with the ones in the original GWAS cohorts. Significant heterogeneity was noted only for the second HLA locus represented by rs9357155 and rs2071543.

Q-test: P-value for the Cochran's Q statistic for heterogeneity, NS: heterogeneity test not significant.

\*heterogeneity P < 0.05.

\*\*heterogeneity P < 0.01.

I<sup>2</sup>: Heterogeneity Index (0–100%), where <25% corresponds to low, 50%–75% to medium, and >75% to high level of heterogeneity.

OR: Additive (per-allele) Odds Ratio.

<sup>#</sup>Han and Eskin random effects model.

doi:10.1371/journal.pgen.1002765.t001

**Table 2.** Conditional analysis of the *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1* locus.

	Replication Study N = 4,789 across 8 cohorts (2,228 cases/2,561 controls)				Replication and GWAS N = 10,755 across 12 cohorts (5,372 cases/5,383 controls)				
	UNADJUSTED		CONDITIONED		UNADJUSTED		CONDITIONED		CONDITIONING SNPs
	OR	P-value	OR	P-value	OR	P-value	OR	P-value	
<b>rs9275224</b>	0.75	$4 \times 10^{-11}$	0.71	$2 \times 10^{-6}$	0.72	$9 \times 10^{-30}$	0.75	$7 \times 10^{-10}$	rs2856717, rs9275424, rs9275596
<b>rs2856717</b>	0.86	$1 \times 10^{-3}$	1.72	$1 \times 10^{-6}$	0.77	$7 \times 10^{-16}$	1.61	$2 \times 10^{-10}$	rs9275224, rs9275424, rs9275596
<b>rs9275424</b>	1.22	$5 \times 10^{-5}$	1.06	$3 \times 10^{-1}$	1.28	$3 \times 10^{-14}$	1.11	$7 \times 10^{-3}$	rs9275224, rs2856717, rs9275596
<b>rs9275596</b>	0.75	$5 \times 10^{-9}$	0.64	$2 \times 10^{-6}$	0.67	$5 \times 10^{-32}$	0.58	$3 \times 10^{-16}$	rs9275224, rs2856717, rs9275424

doi:10.1371/journal.pgen.1002765.t002

variance and was 13 orders of magnitude more significant than the original score. In this model, one standard deviation increase in the score was associated with nearly 50% increase in the odds of disease (OR = 1.47, 95% CI: 1.42–1.54,  $P = 1.2 \times 10^{-72}$ ). This translates into nearly a 5-fold increase in risk between individuals from the opposing extremes of the risk score distribution (with tails defined by  $\geq 2$  standard deviations from the mean).

### Geospatial Modeling of Genetic Risk Reveals New Geographic Patterns in Disease Prevalence

Similar to the GWAS study, we detected pronounced differences in the distributions of risk alleles among the three different ethnicities studied: for each of these seven risk loci, the frequency of the risk alleles was highest in East Asians and lowest in African-Americans (Figure S1). These differences were also reflected in highly significant disparities in the risk score distributions by ethnicity (Figure 2). Motivated by these observations, we examined global geographic variation in the genetic risk for IgAN by applying the newly refined IgAN risk score in 6,319 healthy individuals across 85 worldwide populations. We observed marked differences in the genetic risk across the world. Overall, the mean standardized risk score was lowest for Africans, intermediate for Middle Easterners and Europeans, and highest for East Asians and Native Americans (Figure 3 and Figure S2). Accordingly, the risk increased sharply with eastward distance from the prime meridian (Pearson's  $r = 0.27$ ,  $p = 3.5 \times 10^{-108}$ ). The same geospatial pattern were detected if we included only native populations of HGDP and HapMap-III (Figure S3), demonstrating that the findings are not biased by inclusion of control populations from the genetic

association study. These data are consistent with the known East-West gradient in prevalence of IgAN, suggesting that genetic risk predicts prevalence.

Unexpectedly, higher resolution analysis of the European continent revealed an additional increase in the risk from South to North (Pearson's  $r = 0.11$ ,  $p = 1.3 \times 10^{-9}$ ). For example, northwestern Russians and northern inhabitants of Orkney Islands (Scotland) have the highest risk scores when compared with the rest of the European continent (Tables S8 and S9). To confirm these finding and test whether North-South variation in genetic risk is also reflected in differences in IgAN occurrence, we obtained genetic data from additional European populations (Belgian, British, Finnish, Swedish and Icelandic) and compared genetic risk scores with the incidence and point prevalence of IgAN among end-stage renal disease (IgAN-ESRD) populations across Europe (Table S10). As predicted by the genetic risk score, our analysis confirmed a strong North-South cline of both incidence and prevalence across the European continent (Figure 4). Notably, this analysis includes only patients with end-stage IgAN, on dialysis or after kidney transplantation, thus it underestimates the true incidence and population prevalence of IgAN. Because the point prevalence of IgAN-ESRD (Figure 4b) can be confounded by differential survival on renal replacement therapy and differences in kidney biopsy practice by country, we also examined IgAN-ESRD prevalence expressed as a percentage of all ESRD (Figure 4c), and ESRD due to biopsy-diagnosed primary glomerulonephritis (Figure 4d). Regardless of the metric used to quantify differences in IgAN occurrence, regression of the genetic risk score and the prevalence data on the average latitude resulted in positive correlations and parallel trends.

**Table 3.** Haplotype analysis of rs9275224, rs2856717, rs9275424, and rs9275596 at the *HLA-DQB1/DRB1* locus.

All Cohorts: N = 10,755 (5,372 cases/5,383 controls)						
	Freq. Overall	Freq. Cases	Freq. Controls	OR	95%CI	P-global
<b>G CAT</b>	0.352	0.365	0.338	-reference-	-reference-	$3 \times 10^{-43}$
<b>A T A C</b>	0.213	0.180	0.245	0.69	0.64–0.74	
<b>A C A T</b>	0.130	0.119	0.141	0.78	0.71–0.85	
<b>A T A T</b>	0.050	0.058	0.043	1.25	1.10–1.42	
<b>G C G T</b>	0.246	0.270	0.222	1.12	1.04–1.20	

The most common haplotype of 4 major alleles (GCAT) is used as a reference to derive odds ratios for all other haplotypes. Only common haplotypes (frequency > 1%) are tested for association.

doi:10.1371/journal.pgen.1002765.t003

**Table 4.** The best predictive model for IgAN based on all the genotyped SNPs and their pairwise interaction terms.

Predictor (Reference Allele)	Best Predictive Model			Chr.	Annotation of Genes in the Region
	Coefficient ( $\beta$ )	OR (95%CI)	P-value		
rs6677604 (A)	-0.49371	0.61 (0.53–0.71)	$2.2 \times 10^{-11}$	1q32	<i>CFH, CFHR1, CFHR3</i>
rs9275224 (A)	-0.31307	0.73 (0.67–0.80)	$2.5 \times 10^{-11}$	6p21	<i>HLA-DQB1, -DQA1, -DRB1 (variant 1)</i>
rs2856717 (T)	0.42265	1.53 (1.31–1.78)	$8.2 \times 10^{-8}$	6p21	<i>HLA-DQB1, -DQA1, -DRB1 (variant 2)</i>
rs9275596 (C)	-0.51157	0.60 (0.52–0.69)	$5.9 \times 10^{-13}$	6p21	<i>HLA-DQB1, -DQA1, -DRB1 (variant 3)</i>
rs9357155 (A)	-0.28621	0.75 (0.69–0.82)	$3.8 \times 10^{-10}$	6p21	<i>HLA-DOB, PSMB8, PSMB9, TAP1, TAP2</i>
rs1883414 (T)	-0.1805	0.83 (0.78–0.90)	$4.8 \times 10^{-7}$	6p21	<i>HLA-DPB2, -DPB1, -DPA1</i>
rs2412971 (A)	-0.28592	0.75 (0.70–0.81)	$2.3 \times 10^{-15}$	22q12	<i>HORMAD2, MTMR3, LIF, OSM, GATSL3, SF3A1</i>
rs6677604 (A)* rs2412971 (A)	0.23171	1.26 (1.12–1.43)	$2.2 \times 10^{-4}$	–	<i>1q32 by 22q12 interaction term</i>

This model represents the solution of a stepwise logistic regression algorithm (BIC-based stepwise model selection). The coefficients from this model are used to refine the risk score for IgAN.

doi:10.1371/journal.pgen.1002765.t004

The co-variation in genetic risk score and IgAN-ESRD occurrence among world populations may also be in part influenced by differences in environment, or by other factors such as local medical guidelines for screening and treatment. To better distinguish these possibilities, we examined native populations that live under a uniform environment yet show variation in IgAN risk. In the densely sampled North Italian populations, the Alpine villagers of the Valtrompia region have a 3.5-fold higher prevalence of ESRD attributable to IgAN and primary glomerulonephritis when compared to the national average [16]. Consistent with this prevalence data, the median standardized risk score in this population was comparable to some of the Northern European countries and ranked as number one among the 17 Italian populations sampled in our study (Figure 5, Table S8).

Conversely, we compared the genetic risk score and IgAN-ESRD prevalence in populations in the United States, where diverse ethnicities live under different environments and health care systems compared to the ancestral populations. The analysis of the USRDS dataset confirmed the striking ethnic differences in IgAN-ESRD prevalence (Table S11): the percentage of ESRD attributable to IgAN was 5-fold greater for Caucasian and 15-fold greater for Asian Americans compared to African-Americans. This increased IgAN-ESRD occurrence in Asian- compared to African-Americans far exceeds the 50% increase in risk predicted by genetic risk-score (one standard deviation difference), suggesting the presence of additional unaccounted genetic and environmental factors (Figure 6).

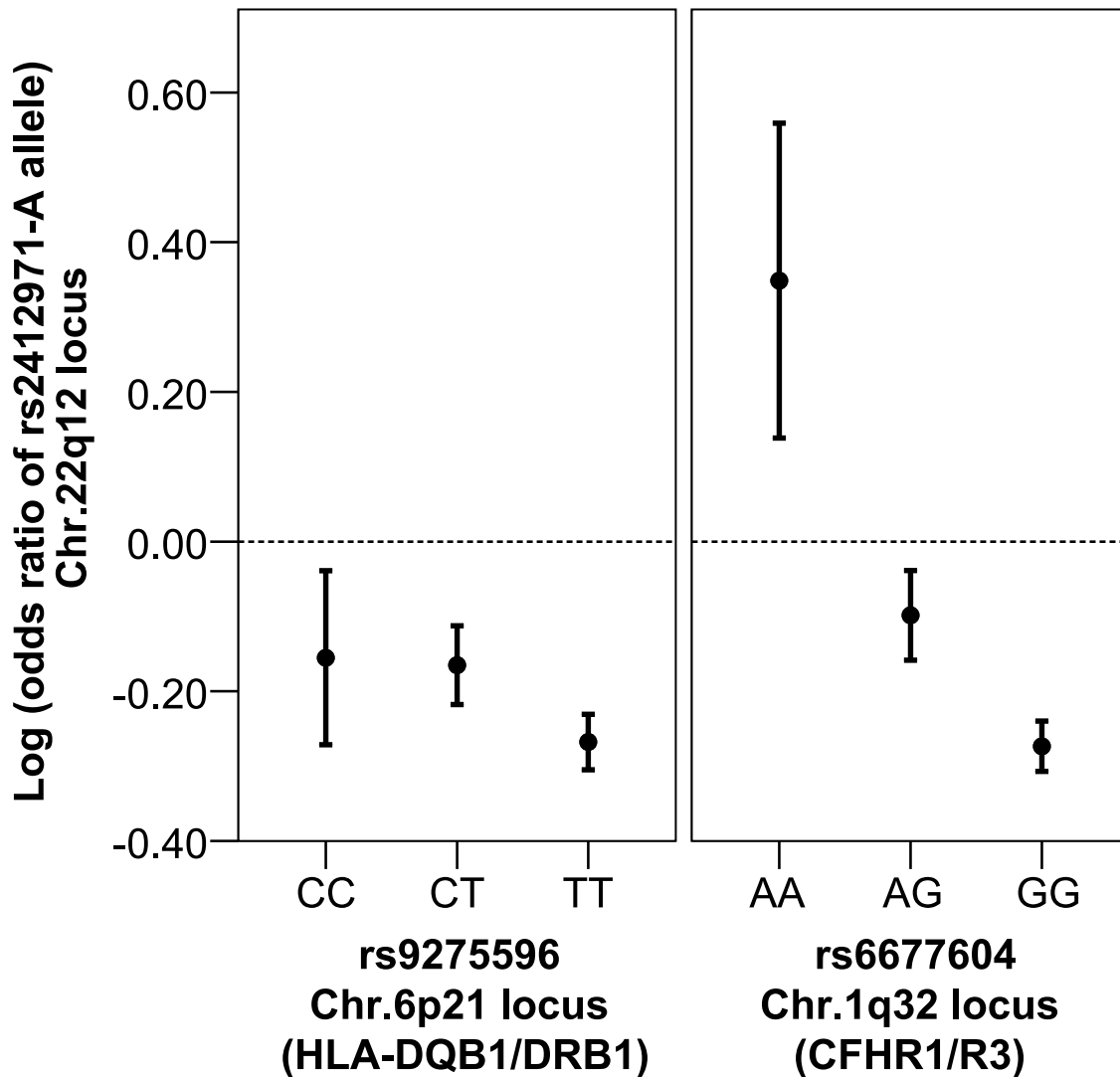
## Discussion

In this study, we examined the largest IgAN case-control cohorts reported to date. We first verified the five top signals identified in a recent GWAS for IgAN in independent cohorts and demonstrated robust replication of four loci, and heterogeneity at one locus. Using combined dataset of 10,755 individuals, we also identified novel risk alleles for IgAN in the *DQB1/DRB1* locus and detected a significant interaction between the *CFHR3/R1* and the *HORMAD2* loci. We also defined a more powerful genetic risk score that explained 4.7% in disease variance across all cohorts. Finally, in examination of 85 world populations, the genetic risk score paralleled the prevalence of IgAN, confirming the known East-West cline but also led to the detection of an association of IgAN-ESRD prevalence with latitude in Europe.

While ten of twelve tested SNPs (four susceptibility loci) were robustly replicated with direction-consistent ORs across all cohorts, the *TAP2/PSMB9* locus demonstrated moderately high level of heterogeneity. This locus remained genome-wide significant in the combined analyses under both fixed and random effects model. Family-based studies [17,18], sperm typing experiments [19] and HapMap data have identified a recombination hotspot directly centered over the *TAP2* gene (22 cM/Mb, 5.5-kb centromeric from the 2 SNPs selected for replication). We can therefore hypothesize that high heterogeneity at this locus is due to the unusually high rates of recombination in this region, which perturbs LD patterns between tag-SNPs and causal variants; this situation has been shown to cause a “flip-flop” phenomenon in association results [20]. Therefore, higher density of SNP coverage on either side of the recombination hotspot will be needed to guide future replication and fine mapping efforts.

In addition to the independent replication of GWAS data, we identified two new signals in the *DQB1/DRB1* region that exhibit independent genome-wide significant effect in conditional analyses, providing support for multiple causal variants at this locus. These findings are consistent with previous studies of IgAN [15,21] and other autoimmune diseases [22,23,24,25], highlighting the complexity of associations in the MHC region. In our study, the strongest association signal originates in a protective haplotype tagged by rs9275596-C that carries *HLA-DRB1\*1501* and *DQB1\*602*, also associated with protection against type I diabetes [24]. The causal variants underlying the other haplotypes remain obscure and their discovery will likely require comprehensive re-sequencing to define classical alleles.

Genetic interactions have been seldom described in association studies [26]. We detected a multiplicative interaction between the *CFHR3/R1* and the *HORMAD2* loci, which was most evident in the European cohorts, likely because the frequencies of both protective variants are considerably higher in this population. While this interaction was robust to multiple-testing correction for 7 SNPs, it will require confirmation in additional independent cohorts or via functional studies that examine whether these two loci are involved in a common biological pathway. Because the rs6677604-A allele tags a deletion in the *CFHR3/CFHR1* genes, this finding suggests that the absence of these proteins abrogates the benefit imparted by *HORMAD2* protective alleles. It is thus noteworthy that the *HORMAD2* locus encodes several cytokines (LIF, OSM) that can interact with complement factors [27].



**Figure 1. Multiplicative interaction between Chr. 22q12 (rs2412971) and Chr. 1q32 (rs6677604) loci.** The allelic effects of rs2412971-A by genotype class of rs9275596 (top signal in the HLA, no interaction) and rs6677604 (top signal in at *CFHR1/R3* locus on Chr. 1q32, significant interaction). The protective effect of rs2412971-A allele is reversed in homozygotes for the rs6677604-A allele, which tags a deletion in *CFHR3/R1*. The allelic effects are expressed on the log-odds scale and correspond to beta coefficients of the logistic regression model. Error bars correspond to 95% confidence intervals.

doi:10.1371/journal.pgen.1002765.g001

**Table 5.** The comparison of the original and the newly refined IgAN risk score.

Cohort:	N <sup>#</sup>	Original Risk Score				Newly Refined Risk Score			
		R <sup>2*</sup>	C <sup>**</sup>	OR <sup>***</sup>	P-value <sup>****</sup>	R <sup>2*</sup>	C <sup>**</sup>	OR <sup>***</sup>	P-value <sup>****</sup>
<b>Original GWAS Cohorts</b>	5,631	5.0%	0.61	1.51	$3.1 \times 10^{-46}$	5.7%	0.62	1.56	$4.1 \times 10^{-52}$
<b>Replication Cohorts</b>	4,422	2.2%	0.58	1.29	$5.4 \times 10^{-17}$	3.2%	0.59	1.36	$3.3 \times 10^{-24}$
<b>Asian Cohorts Combined</b>	4,582	4.5%	0.60	1.53	$3.0 \times 10^{-34}$	5.0%	0.61	1.52	$2.6 \times 10^{-38}$
<b>European Cohorts Combined</b>	5,386	2.6%	0.58	1.34	$3.7 \times 10^{-24}$	3.6%	0.59	1.42	$6.7 \times 10^{-33}$
<b>All Cohorts Combined</b>	10,053	3.8%	0.60	1.42	$6.2 \times 10^{-63}$	4.7%	0.61	1.47	$1.2 \times 10^{-76}$

The expanded version of this table can be found in supplemental material (Table S7).

<sup>#</sup>Number of analyzed individuals with 100% non-missing genotypes across all 7 scored loci.

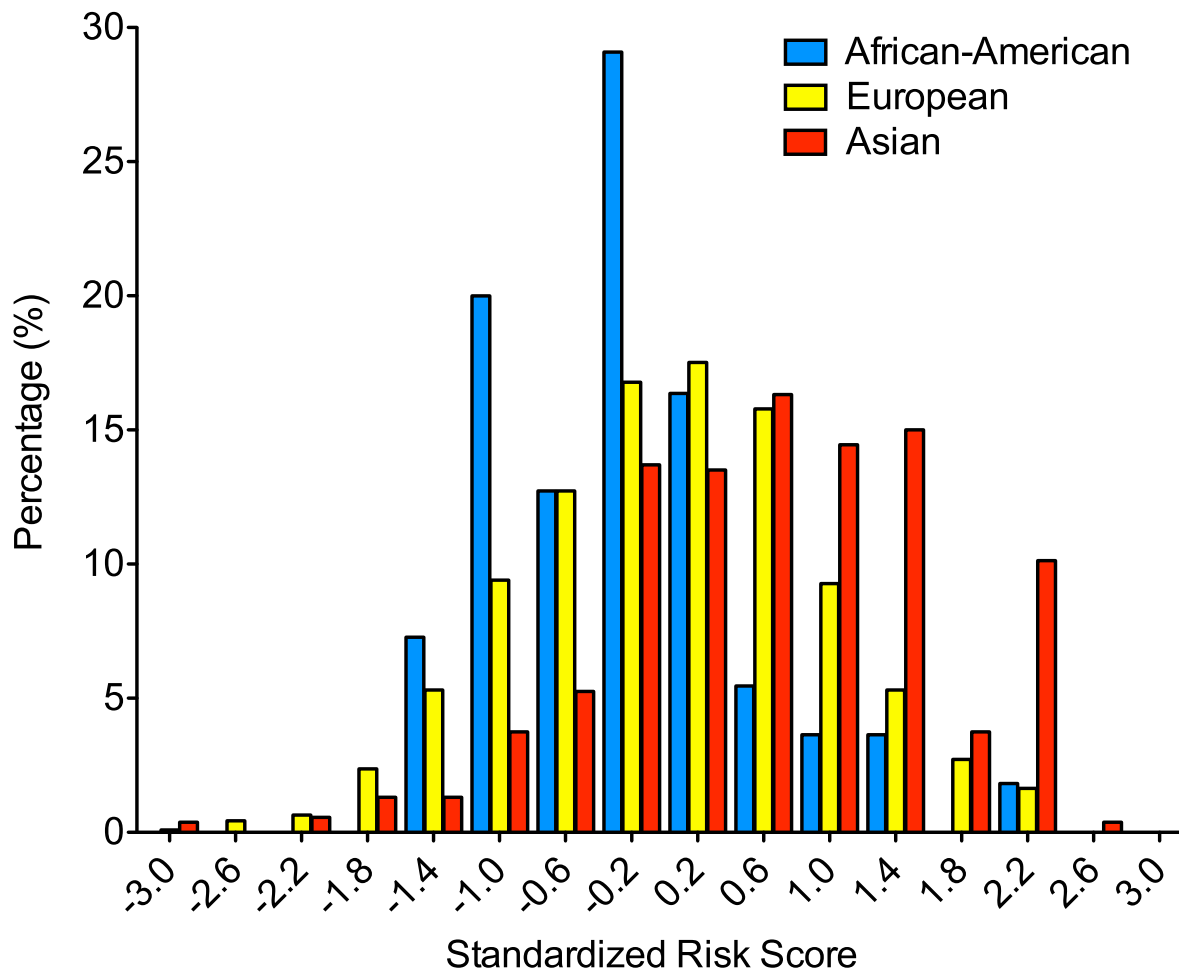
<sup>\*</sup>2: Nagelkerke R square (expressed as percentage).

<sup>\*\*</sup>C-statistic: area under the ROC curve.

<sup>\*\*\*</sup>Odds ratio per one standard deviation of the standardized risk score.

<sup>\*\*\*\*</sup>Wald's test for risk score as a quantitative predictor of disease status.

doi:10.1371/journal.pgen.1002765.t005



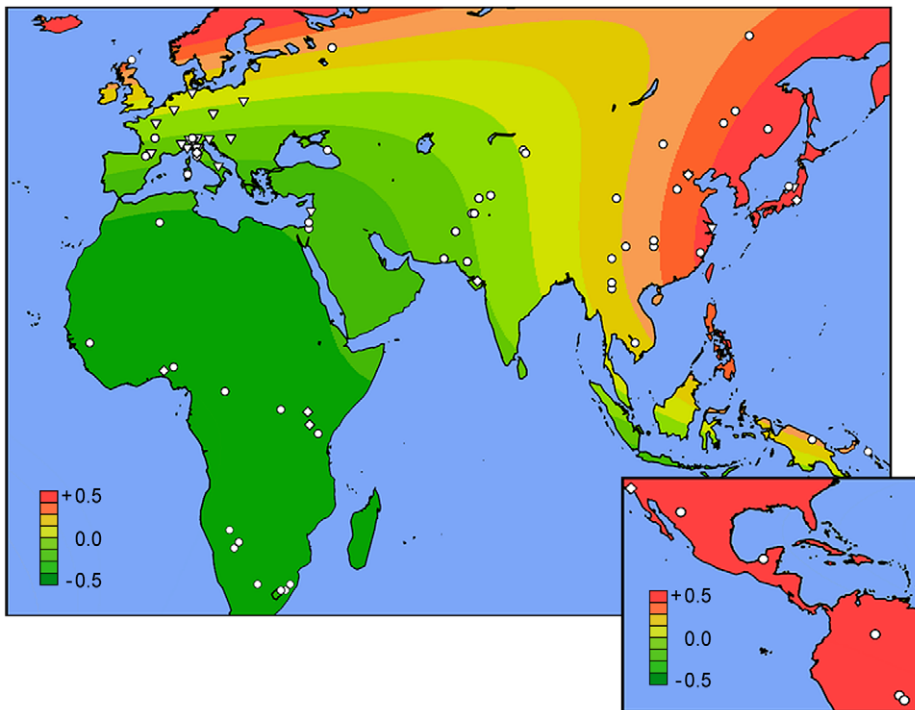
**Figure 2. Differences in the distribution of the 7-SNP genetic risk score by ethnicity.** Only healthy control participants of the replication studies that were fully genotyped at all 7 loci were used in this analysis. Similar to the GWAS study, the risk score distributions were significantly different by ethnicity (ANOVA  $p = 2.1 \times 10^{-38}$ ). The corresponding differences in the distribution of risk alleles are depicted in Figure S1. doi:10.1371/journal.pgen.1002765.g002

A seven-SNP genetic risk score explained nearly 5% of IgAN variance and demonstrated co-variation with IgAN prevalence across multiple settings. The major limitations of geospatial modeling include variable sampling density and inadequate coverage of certain geographic regions. Using the most comprehensive resources presently available for geo-genetic analyses, we found that the genetic risk score strongly paralleled the well-known East-West gradient in IgAN prevalence [3,28,29,30,31,32]. For each of these seven risk loci, the frequency of the risk alleles was highest in East Asians, lowest in African-Americans and intermediate in European populations. Accordingly, we detected co-variation of genetic risk with IgAN-ESRD incidence and prevalence among Asian-, White- and African-Americans, which share genetic background but not environment with their ancestral populations. Representative genetic data for U.S. Native Americans was not available from HGDP nor HapMap projects, precluding a direct comparison of their risk score with prevalence. However, the USRDS data and other reports indicate a high prevalence of IgAN-ESRD in US Native Americans [33,34,35,36,37], consistent with their ancestral origin from an Asian subpopulation that migrated across the Bering land bridge over 15,000 years ago [38]. In the more homogeneous population of Northern Italy, the median risk score in the Valtrompia valley

was the highest among Northern Italian populations and comparable with the Northern European scores, consistent with Valtrompia's 3.5-fold higher prevalence of ESRD, which is largely attributable to IgAN [16]. Taken together, these data strongly suggested that variation in genetic risk partly explains the variation in geo-epidemiology of disease.

Because the genetic score captured general trends in IgAN epidemiology, we also tested whether the Northward gradient in genetic risk in Europe is mirrored by higher prevalence of kidney failure from IgAN. The ERA-EDTA data, which are the most unbiased source of information available, demonstrate that Nordic countries have over 2-fold higher incidence and prevalence of IgAN-ESRD compared to the Southern European countries. Although higher risk of IgAN in Northern Europe has not been previously appreciated, similar latitudinal risk gradients in prevalence and incidence have been well established for several other immune-mediated diseases, including type 1 diabetes [39,40], multiple sclerosis [41,42], and inflammatory bowel disease [43]. Interestingly, these disorders share risk alleles with IgAN, suggesting that variation in common genetic risk factors may mediate variation in prevalence of autoimmune disorders. Since our analysis was limited to prevalent IgAN-ESRD in countries with epidemiological data available and only a portion of IgAN





**Figure 3. Worldwide geospatial risk analysis.** Surface interpolation of the standardized risk score over Africa and Euroasia (main), and Americas (inset). Symbols represent the locations of sampled populations: HGDP (circles), HapMap-III (diamonds), and healthy controls from this study (triangles).

doi:10.1371/journal.pgen.1002765.g003

cases progresses to ESRD, studies that can better estimate the population prevalence of all IgAN can confirm these findings and better delineate epidemiological connections to other immune mediated disorders.

The genetic and environmental factors leading to the observed geospatial pattern of genetic risk and disease prevalence are not clear. The pre-modern history of IgAN is not known because this disease was only first described in 1968 [44], shortly after the discovery and application of immunofluorescence in the analysis of kidney tissue. It is well known that mucosal infections can exacerbate disease, but specific environmental factors influencing the development of IgAN are not known. Based on a recently proposed pathogenesis model, the IgAN risk loci participate in sequential processes leading to the initiation and exacerbation of IgAN [45]. This may further explain the correlation of the genetic risk score with disease epidemiology. Interestingly, many of the IgAN loci are known to exhibit opposing effects on other autoimmune conditions [14]; for example, the *HLA-DQB1* and *HORMAD2* risk alleles are respectively protective for systemic lupus erythematosus, and inflammatory bowel disease. Thus balancing selection, in conjunction with local environmental factors, may be responsible for maintenance of risk alleles in different populations.

The current IgAN risk score captures a greater proportion of the disease variance compared to other GWAS for kidney functions, such as a recent study of 60,000 individual that reported 13 loci explaining only 1.4% of the variance for estimated glomerular filtration rate [46]. Nonetheless, the fraction of the IgAN variation explained remains modest. For example, the one standard deviation risk-score difference between Asian- and African-Americans predicts a 50% increase in risk, yet there is over 10-fold difference IgAN-ESRD occurrence between these

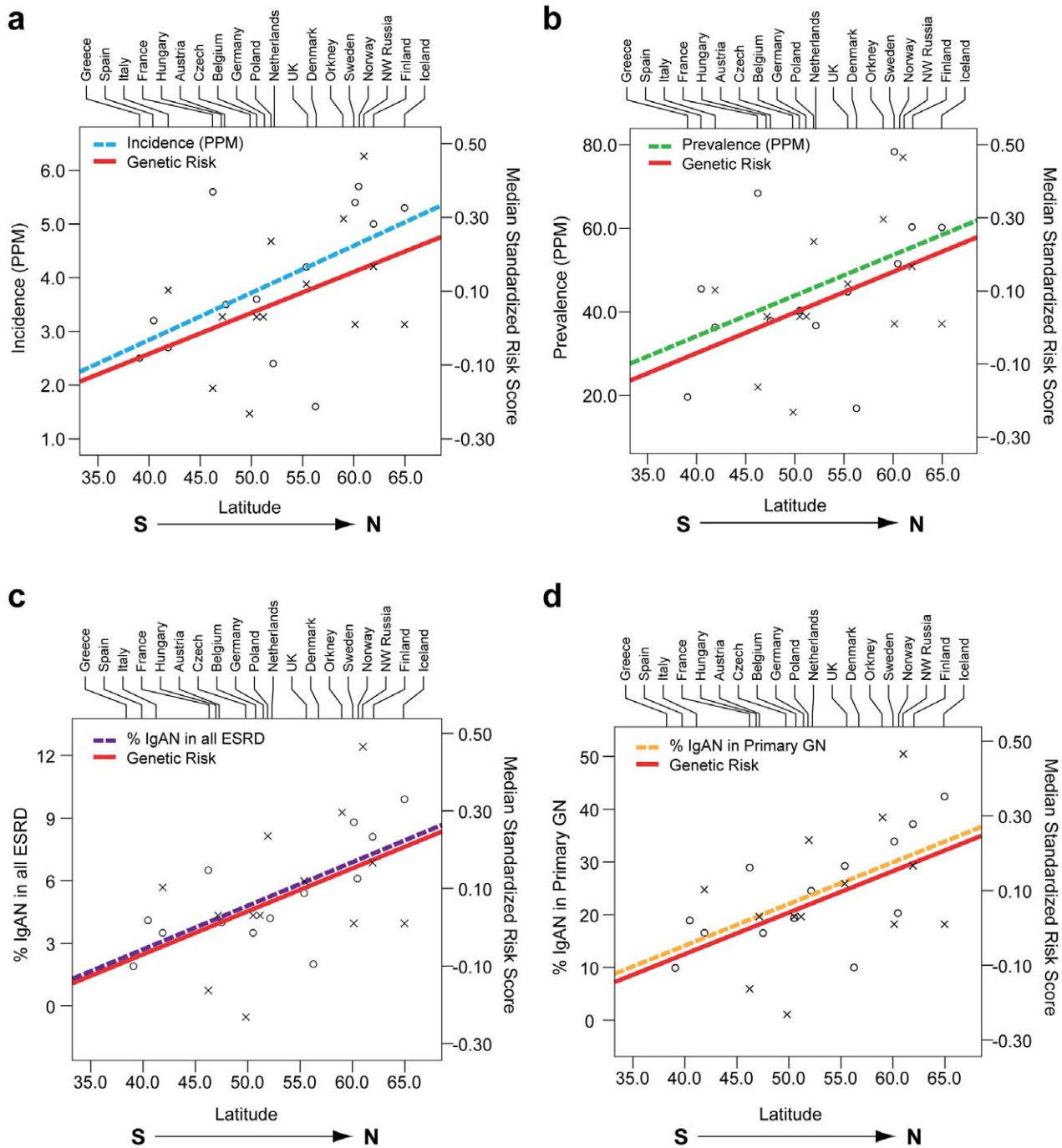
two groups. These data suggest that additional genetic and environmental factors influence risk. Based on the effect sizes and allelic frequencies of the discovered SNPs, we estimate that doubling the GWAS sample size is likely to find up to 7 additional loci, while tripling the sample size would identify up to 11 additional loci at genome-wide significant  $p$ -values  $< 10^{-8}$  (calculation performed as proposed by Park et al. [47]). Conditional analyses and higher-level interaction screens of more risk loci are likely to explain additional fraction of the missing heritability and better explain differences in population prevalence of this disease.

In summary, we report results of the largest collaborative genetic study of IgAN. We confirm that the IgAN risk loci discovered in GWAS explain a significant proportion of the disease risk worldwide and likely contribute to the geographic variation in disease prevalence. Our geospatial model suggests previously unrecognized northward risk gradient in Europe, which will require further confirmation by alternative sources of prevalence data, such as country specific biopsy-registry data or kidney donor-biopsy series. The approach presented in this study may serve as a blueprint for geo-genetic modeling of other complex traits that exhibit marked geographic differences in prevalence.

## Materials and Methods

### Ethics Statement

This investigation was conducted according to the principles expressed in the Declaration of Helsinki. All subjects provided informed consent to participate in genetic studies and the Institutional Review Board of Columbia University as well as local ethic review committees for each of the individual cohorts approved our study protocol.

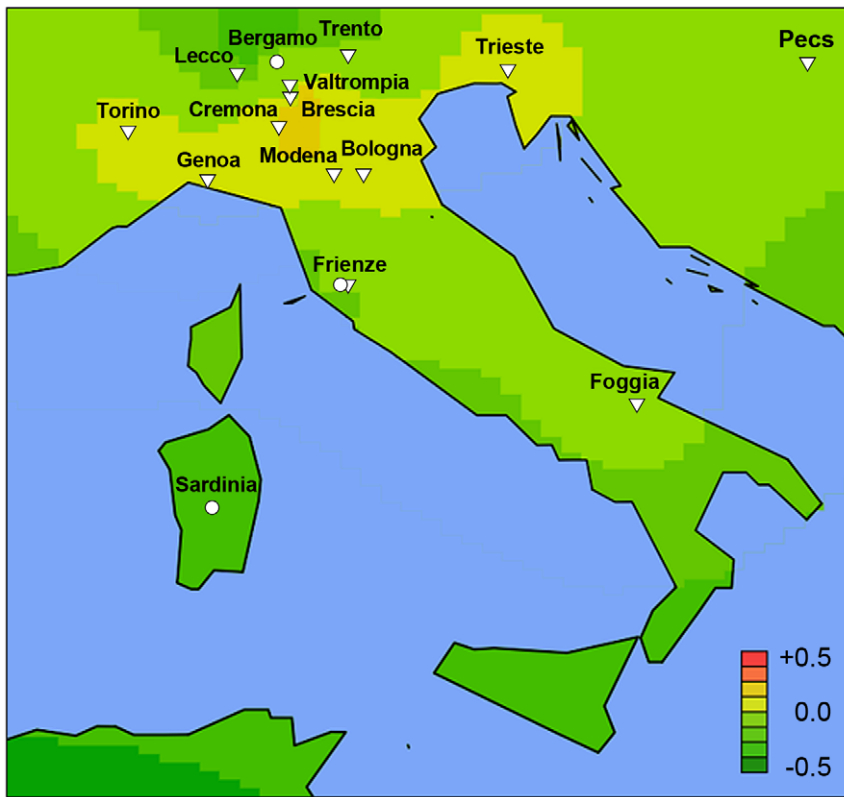


**Figure 4. Correlation of average country latitude with country-specific genetic risk and IgAN-attributable ESRD across the European continent.** The South to North latitude is indicated on the X-axis. The median genetic risk (x) is indicated on the right Y-axis. The following incidence and prevalence metrics (o) are indicated on the left Y-axis: (panel a) the incidence of ESRD due to IgAN per million population (correlation with latitude:  $r = 0.54$ ,  $p = 0.05$ ); (panel b) the prevalence of ESRD due to IgAN per million population (correlation with latitude:  $r = 0.47$ ,  $p = 0.10$ ); (panel c) the percent of IgAN patients among all ESRD cases (correlation with latitude:  $r = 0.67$ ,  $p = 0.01$ ); and (panel d) among ESRD cases due to primary glomerular disease (correlation with latitude:  $r = 0.71$ ,  $p = 0.006$ ). All p-values are derived based on a two-sided hypothesis test. doi:10.1371/journal.pgen.1002765.g004

### Study Cohorts

The case-control cohorts analyzed in this study were contributed by clinical nephrology centers across Europe, Asia, and North America (Table S1). All cases carried a biopsy diagnosis of IgAN defined by typical light microscopy features and predominant IgA staining on kidney tissue immunofluorescence, in the absence of liver disease or other autoimmune conditions. Each individual

cohort of cases was accompanied by a control cohort of similar size, matched based on self-reported ethnicity and recruited from the same clinical center. The French cohort was composed of two sub-cohorts: the St. Etienne cohort recruited in the University North Hospital of Saint Etienne (289 cases and 244 controls), and the GN-Progress cohort recruited from the nephrology departments of the Paris region (207 cases and 159 controls). The Italian



**Figure 5. High-resolution geospatial risk analysis for Italy.** A well defined region of higher genetic risk was uncovered in Northern Italy that centers on Valtrompia, Brescia, and Cremona (median standardized risk scores 0.31, 0.24 and 0.24, respectively). The healthy individuals from Valtrompia had the highest risk scores when compared to 16 other Italian populations sampled. doi:10.1371/journal.pgen.1002765.g005

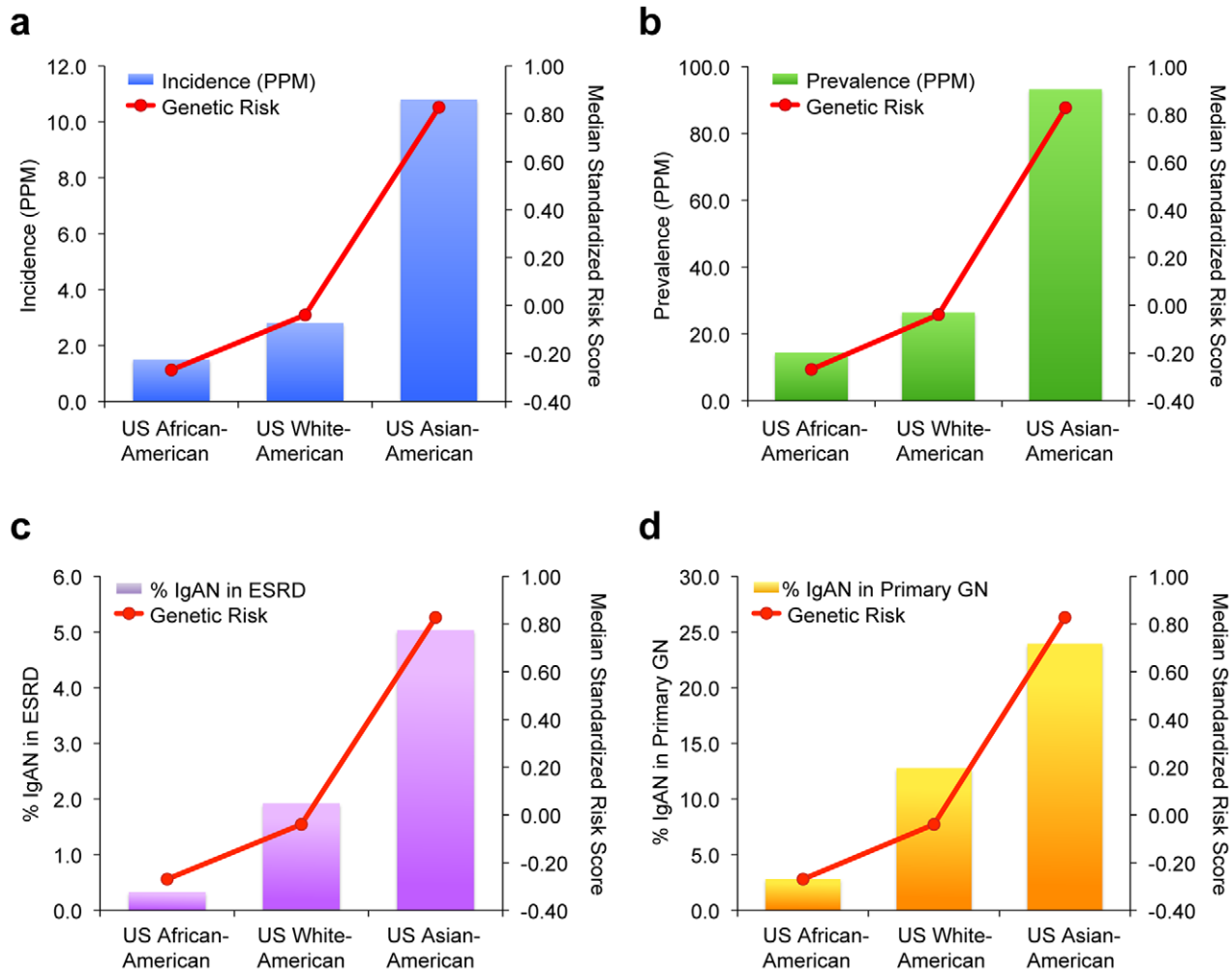
cohort was also composed of two sub-cohorts: the North Italian cohort recruited in the clinical centers of Genova, Torino, Brescia, Trento, Modena, Bologna, and Trieste (410 cases and 524 controls), and the South Italian cohort recruited in Foggia (81 cases and 80 controls). The German cohorts also represent two recruitment sites: the Stop-IgAN cohort recruited among the participants of the Stop-IgAN clinical trial based in Aachen (150 cases and 293 controls), and the Hamburg-Eppendorf cohort from northern Germany (101 cases and 80 controls). The Czech and the Hungarian cohorts were recruited through the Department of Nephrology, 1st Faculty of Medicine and General University Hospital, Charles University in Prague (245 cases and 223 controls) and the Nephrology Department of the University of Pécs (139 cases and 305 controls), respectively. The Japanese participants (264 cases and 294 controls) were recruited by the nephrologists of Niigata University. The Beijing cohort (333 cases and 289 controls) was recruited by the Renal Division of the Peking University First Hospital. Finally, our African-American cohort (34 cases and 60 controls) was recruited at Columbia University (New York, NY) and at the University of Alabama (Birmingham, AL). This smaller cohort is unique, as IgAN is exceedingly rare among individuals of African ancestry. In total, 2,253 cases and 2,621 controls were available for genotyping in the replication study. The composition and recruitment of the GWAS cohorts have been discussed in detail elsewhere [14].

#### Genotyping and Genotype Quality Control

The genotyping was performed by KBiosciences (Hoddeston, England), and genotype calls were determined using an automated

clustering algorithm the (SNP Viewer v.1.99, KBiosciences, 2008). The genotype clusters were also examined visually across all plates, to assure lack of technical artifacts. The overall genotyping rate across all samples was 98.2%. For quality control we calculated minor allele frequencies, as well as per-SNP and per-individual rates of missingness within each case-control cohort separately. Additionally, we tested for Hardy-Weinberg equilibrium among the control groups from each cohort to assure lack of bias due to genotyping artifacts or population stratification. All SNPs included in the final analyses had minor allele frequency greater than 1%, per-SNP missingness rate less than 5%, and all passed the HWE test in controls ( $p > 1 \times 10^{-2}$ ). Individuals with more than 2 missing genotypes out of the 12 loci were also excluded from the analysis.

The participants of the smaller GN-Progress study (207 cases and 159 controls) were genotyped using the Illumina HumanCNV370-duo chip at the Centre National de Génotypage (CEA, Evry, France). The analysis of intensity clusters and genotype calls were performed using the Illumina Genome Studio software. Of 366 genotyped individuals, two cases and 1.8% of SNPs were excluded based on low call rates ( $< 95\%$ ). The overall genotyping rate was 99.6%. In total, 6 of 12 SNPs analyzed for replication were also present on the Illumina HumanCNV370-duo chip. The genotypes at the reminder loci were imputed using the phased HapMap-III CEU reference dataset (see Web Resources). The imputation was performed simultaneously for cases and controls, using MACH 1.0 software (see Web Resources). We used a standard single-step imputation approach, with 60 rounds of Markov Chain iterations to estimate the crossover maps, error rate maps, and all missing genotypes across each analyzed locus. The



**Figure 6. Genetic risk and IgAN-attributable ESRD among major US ethnicities.** The relationship between IgAN risk scores (red line) and IgAN incidence and prevalence (bars) among US ethnicities are shown. The following metrics of IgAN occurrence are depicted: (panel a) the incidence of ESRD due to IgAN per million population by ethnicity, (panel b) the prevalence of ESRD due to IgAN per million population by ethnicity, (panel c) percent of IgAN among the total ESRD population by ethnicity; and (panel d) percent of IgAN among ESRD due to glomerular disease by ethnicity. doi:10.1371/journal.pgen.1002765.g006

imputed SNPs had an excellent imputation quality, with an average estimated correlation between imputed genotypes and experimental genotypes of 0.98 (range 0.94–1.0). Consequently, association analyses using either the allelic dosage approach that accounts for imputation uncertainty, or the most likely genotype approach yielded similar results. Therefore, the most probable genotype calls were used in the downstream analyses. In the final quality control step, we compared the allelic frequencies and effect estimates between the two French cohorts (GN-Progress and St. Etienne). For each locus, we observed nearly identical frequencies among cases and controls and the odds ratios were homogenous between the two cohorts. The formal heterogeneity tests were not statistically significant for any of the tested loci (Cochrane's Q-test  $P > 0.05$ , average  $I^2 = 0$ ). Therefore, these two cohorts were combined into a single cohort of 493 cases and 402 controls.

Similarly to the French cohorts, there was no significant heterogeneity at any of the loci for the two smaller German cohorts (STOP-IgAN and Hamburg-Eppendorf), and these were also combined into a single cohort of 249 cases and 372 controls. Analysis of the Northern and Southern Italian cohorts suggested

some heterogeneity at 3 out of 12 SNPs ( $I^2 = 40$ –50%). Although these observations were not statistically significant (Q-test  $P > 0.05$ ), we used a conservative stratified approach for all downstream analyses for these two cohorts. The final summary of all study cohorts before and after quality control is provided in Table S1.

### Power Calculation

We performed a power calculation for the final replication cohort size of 4,789 individuals (2,228 cases/2,561 controls) as a function of disease allele frequency and genotype relative risk (Table S2). The power was calculated in reference to a protective allele, with the range of allelic frequencies and effects comparable to the ones observed in the original GWAS. Assumptions included disease prevalence of 1%, log-additive model, no heterogeneity, and  $\alpha = 0.01$  (Bonferroni-adjusted considering five independent loci tested). This analysis confirmed that our study had ample power (nearly 100% for most loci) to replicate the associations observed in the initial GWAS. The power calculations were performed using QUANTO v.1.2 software [48].

## Association Analyses

The primary association analyses were performed using PLINK version 1.07 [49]. Similar to GWAS, we selected a standard 1-df Cochran-Armitage trend test as the primary association test. We also estimated the per-allele odds ratios and 95% confidence intervals for all tested SNPs within each individual cohort. The results across multiple cohorts were combined using an inverse variance-weighted method under a fixed-effects model (PLINK), as well as using a random effects model as proposed by Han and Eskin (METASOFT) [50]. We also tested for heterogeneity across cohorts by performing a formal Cochran's Q heterogeneity test as well as by estimating the heterogeneity index ( $I^2$ ) [51].

## Conditional Analyses

The conditional association tests of the HLA loci were performed after controlling for the genotypes of the conditioning SNPs within each cohort using logistic regression (PLINK). The adjusted (conditioned) effect estimates were then combined across cohorts using a fixed effect meta-analysis considering no significant heterogeneity across these loci. For the purpose of validation of this approach, we also combined the results by adding cohort information as an additional covariate in the stratified analysis within the logistic regression framework. As expected, the results of both approaches were similar.

## Haplotype-Based Association Tests

These analyses were carried out in PLINK v1.07 [49]. Haplotypes were first phased using EM algorithm across the *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1* region. The haplotype frequencies were estimated in the cases and controls separately, as well as jointly in the entire cohort. Only common haplotypes with overall frequency >1% were included in the association tests. Global haplotype association test was performed using a  $\chi^2$  test with  $n-1$  degrees of freedom for  $n$  common haplotype groups. The ORs and the corresponding 95% confidence intervals were estimated in reference to the most common haplotype (GCAT, frequency ~35%).

## First-Order Interaction Analyses

To explore the possibility of interactions between the 7 independent risk variants, we screened all possible pairwise interaction terms for association with disease within the framework of logistic regression models (R version 2.10). As a screening test, we used 1-df LRT to compare two nested models: one with main effects only and one with main effects and a multiplicative (logit-additive) interaction term. We included cohort membership as a fixed covariate in both of these models. For this analysis we selected a Bonferroni-adjusted significance of  $2.4 \times 10^{-3}$ , a conservative threshold that accounts for all 21 pairwise interaction terms tested. Significant interactions from this analysis were also tested using a 4-df genotypic interaction test. In this test, we compared a model with allelic effects, dominant effects, and their interaction terms with a reduced model with no interaction terms. We followed the coding proposed by Cordell and Clayton: for each SNP  $i$  we modeled its allelic effect  $x_{ia}$  by coding the genotypes AA, AB, and BB as  $x_{ia} = -1, 0, 1$ ; we modeled dominance effects as  $x_{id} = -0.5, 0.5, -0.5$  for the genotypes AA, AB, and BB, respectively [52].

## Distributions of Protective Alleles and Risk Score Analyses

Each study participant was scored for the number of risk alleles and the distributions of protective alleles were compared between cohorts of different ethnicity. Only individuals with complete

genotype information at the 7 scored loci (14 alleles) were included in this analysis. The distributions were analyzed separately for cases and controls. A  $\chi^2$  goodness-of-fit test was used to derive p-values for comparison of distributions. Because of a relatively small number of individuals at the tails of the distributions, for the purpose of statistical testing the tails of the distributions were binned into single-bin categories to achieve expected cell counts >5.

To confirm the results of conditional analyses and refine the genetic risk score proposed in the original GWAS, we subjected the genotype data from the entire cohort to a stepwise regression algorithm that selects significant covariates for the best predictive regression model based on Bayesian Information Criterion (the *step* function, R version 2.10). At model entry, we included all 12 genotyped SNPs, all 21 tested interactions, as well as cohort membership as a fixed covariate. Consistent with the results of our conditional analysis, the stepwise algorithm retained only the 7 SNPs exhibiting an independent effect along with the rs6677604\*rs2412971 interaction term. All other terms were automatically dropped from the regression model.

The risk score was calculated as a weighted sum of the number of protective alleles at each locus multiplied by the log of the OR for each of the individual loci from the final fully adjusted model. Only individuals with non-missing genotypes for all 14 alleles were included in this analysis. The risk score was standardized across all populations using a z-score transformation, thus the standardized score represented the distance between the raw score and the population mean in units of standard deviation. The percentage of the total variance in disease state explained by the risk score was estimated by Nagelkerke's pseudo  $R^2$  from the logistic regression model with the risk score as a quantitative predictor and disease state as an outcome. The C-statistic was estimated as an area under the receiver operating characteristic curve provided by the above logistic model. These analyses were carried out with SPSS Statistics version 19.0.

## Geospatial Analyses

For this purpose, we used publicly available genotype data of the Human Genome Diversity Panel (HGDP; 1,050 individuals representative of 52 worldwide populations), HapMap III (1,184 individuals representative of 11 populations), along with healthy controls genotyped as part of this study (4,547 individuals representative of 25 recruitment sites). The HGDP individuals have been previously genotyped for 660,918 markers using Illumina 650Y arrays (Stanford University). First, SNPs with genotyping rate <95% and samples with an overall call rate <98.5% were removed from the genome-wide data. Only 1,042 individuals with all 14 non-missing alleles at the 7 analyzed risk score loci were included in the final analysis. The geographic coordinates for the HGDP populations were downloaded from the CEPH website (see Web Resources). The HapMap III genotype data have been generated using two platforms: the Illumina Human1M (Wellcome Trust Sanger Institute) and the Affymetrix SNP 6.0 (Broad Institute). These files were merged into a single dataset of 1,440,616 markers, from which we removed (1) SNPs with genotyping rate <95%, (2) samples with an overall call rate <98.5%, (3) all non-founders from mother-father-child trios, and (4) individuals with missing genotypes at any of the 7 SNP loci used for risk scoring. In the global geospatial analyses, we excluded US-recruited individuals of African American (ASW), European (CEU), and Asian (CHD) ancestry considering non-specific geographic origin of these populations. However, the population of Guajarti Indians recruited in Houston (GIH) was mapped to the northwestern part of the Indian subcontinent, as these individuals

reported having at least three out of four Gujarati grandparents, speak the Gujarati language, and trace their ancestry to the region of Gujarat. In total, 730 HapMap III individuals representative of 8 populations met our selection criteria and were included in the final analysis.

Because many European populations are underrepresented in HGDP and HapMap III datasets, we also included a total of 4,462 healthy controls from the GWAS and replication studies that were collected across 25 recruitment centers participating in our studies. Similar to the above criteria, only individuals with non-missing genotypes at all 7 scored SNPs were included in this analysis. The geographic coordinates for our populations were based on the location of recruitment centers and determined with Google Earth (see Web Resources). This resulted in a final dataset of 6,319 individuals sampled across 85 worldwide populations for geospatial analysis.

We fitted a 3<sup>rd</sup> degree polynomial trend surface based on the latitude, longitude, and median standardized risk score for each of the 85 populations using least squares approach (Spatial package version 7.3-2, R version 2.10). For higher resolution maps, we used kriging technique and accounted for the possibility of spatial correlation of errors among more densely sampled populations by modeling the covariance function in an exponential form. The estimated risk surfaces were projected over the major continents using Maps package version 2.1–6 (R version 2.10).

### Analysis of Prevalence and Incidence Data

We obtained case counts of prevalent and incident ESRD stratified by primary renal diagnosis and by ethnicity from the United States Renal Data Systems (2011 USRDS Data Atlas, see Web Resources). For Europe, we obtained prevalent and incident ESRD case counts from the European Renal Association and European Dialysis and Transplant Association (ERA-EDTA Renal Registry, see Web Resources). Comprehensive data were available for a total of 13 European countries participating in this registry. We calculated the prevalence of ESRD due to IgAN using three definitions: (1) proportion of all ESRD cases attributable to IgAN, (2) proportion of all ESRD cases from primary glomerulonephritis attributable to IgAN, and (3) total number of ESRD cases due to IgAN per million population (PMP). The prevalence data for both USRDS and ERA-EDTA datasets were calculated for the same timepoint of December 31<sup>st</sup>, 2009. The incidence of ESRD due to IgAN was estimated using all the available data over a 3-year period for the ERA-EDTA registry (2007–2009), and a 5-year period for the USRDS registry (2005–2009). For correlation of genetic risk score with disease prevalence in the US, we scored representative samples of the three major US ethnic groups: 303 US Caucasians (CEU founders from HapMap-3 and healthy US controls from our original GWAS), 103 African-Americans (ASW founders from HapMap-3 and healthy controls from this study), and 74 Asian-Americans (CHD founders from HapMap-3). For correlation of genetic risk with disease prevalence in Europe, we calculated median standardized risk scores at a country level for 13 European countries for which we obtained genotype data. We confirmed the South-North disease gradient by regressing the prevalence and risk score data against each country's average latitude. The correlation and regression analyses were conducted in SPSS Statistics version 19.0.

### Web Resources

HAPMAP PHASE III Data: [http://hapmap.ncbi.nlm.nih.gov/downloads/phasing/2009-02\\_phaseIII](http://hapmap.ncbi.nlm.nih.gov/downloads/phasing/2009-02_phaseIII)  
 HGDP Genotype Data: <http://hagsc.org/hgdp>  
 HGDP Population Data: <http://www.cephb.fr/en/hgdp>

MACH: <http://www.sph.umich.edu/csg/abecasis/MaCH>  
 PLINK: <http://pngu.mgh.harvard.edu/~purcell/plink>  
 METASOFT: <http://genetics.cs.ucla.edu/meta>  
 CRAN: <http://cran.r-project.org>  
 GOOGLE EARTH: <http://www.google.com/earth>  
 SPATIAL: <http://cran.r-project.org/web/packages/spatial>  
 MAPS: <http://cran.r-project.org/web/packages/maps>  
 USRDS Data Atlas 2011: <http://www.usrds.org/atlas.aspx>  
 ERA-EDTA Registry Annual Report 2009: <http://www.era-edta-reg.org>

### Supporting Information

**Figure S1** Differences in the distributions of risk alleles at the 7 susceptibility loci among major ethnicities in the replication cohorts. Similar to the GWAS study, the distribution of the risk alleles differed by ethnicity: Asian controls carry more risk alleles compared to healthy Europeans or African-Americans ( $p = 3 \times 10^{-55}$  and  $p = 5 \times 10^{-7}$ , respectively); European controls have more risk alleles compared to African-Americans ( $p = 6 \times 10^{-3}$ ).

(PDF)

**Figure S2** Inter-continental differences in the genetic risk score based on 85 worldwide populations used for geospatial analysis.

(PDF)

**Figure S3** Geospatial risk model for native populations. Surface interpolation of the standardized risk score for HGDP (circles) and HapMap-III (diamonds) datasets. The risk increases globally with the distance from the prime meridian (Pearson's  $r = 0.31$ ,  $p < 2.2 \times 10^{-16}$ ) and northward within Europe (Pearson's  $r = 0.13$ ,  $p = 6.6 \times 10^{-4}$ ).

(PDF)

**Table S1** Summary of the case-control replication cohorts before and after quality control measures.

(PDF)

**Table S2** Power calculation. Study power for the replication cohort of 4,789 individuals (2,228 cases/2,561 controls) as a function of disease allele frequency and genotype relative risk. The power was calculated in reference to a protective allele; the range of allelic frequencies and effects was based on the results of the original GWAS. Assumptions include: disease prevalence of 1%, log-additive model, no heterogeneity, and  $\alpha = 0.01$  (Bonferroni-adjusted considering five independent loci tested).

(PDF)

**Table S3** Case-control association results for the individual replication cohorts.

(PDF)

**Table S4** Pairwise LD between the SNPs of the HLA region:  $r^2$  (top right half) and  $D'$  (bottom left half) for all cohorts (top), Europeans (middle) and Asians (bottom).

(PDF)

**Table S5** Haplotype analysis of rs9275224, rs2856717, rs9275424, and rs9275596 at the *HLA-DQB1/DRB1* locus. The most common haplotype of 4 major alleles (GCAT) is used as a reference to derive odds ratios for all other haplotypes. Only common haplotypes (frequency > 1%) are tested for association.

(PDF)

**Table S6** All possible 1<sup>st</sup> order multiplicative interactions between the 7 SNPs with independent effects on disease risk. Statistical significance is assessed using a Bonferroni-corrected threshold,  $\alpha = 0.05/21 = 2.4 \times 10^{-3}$ .

(PDF)

**Table S7** The comparison of the original and the newly refined genetic risk score. (PDF)

**Table S8** African, Middle Eastern, and European populations included in the geospatial risk analysis. The populations were grouped by their continental origin and sorted based on the median genetic risk score. (PDF)

**Table S9** Asian, Oceanian, and American populations included in the geospatial risk analysis. The populations were grouped by their continental origin and sorted based on the median genetic risk score. (PDF)

**Table S10** Prevalence and Incidence of ESRD due to IgAN in Europe. Primary data obtained from the ERA-EDTA Registry. (PDF)

**Table S11** Prevalence and Incidence of ESRD due to IgAN in the US. Primary data obtained from the USRDS Annual Report, 2011. (PDF)

## Acknowledgments

We are grateful to all study participants for their contribution to this work. We also thank the patients and staff of all the dialysis and transplant units who have contributed data to the ERA-EDTA and USRDS Registries. Furthermore, we gratefully acknowledge the following European registries and persons for their participation in the prevalence data collection: R Kramar (Austria); H Augustijn, B De Moor, and J De Meester (Belgium, Dutch-speaking); JM des Grottes and F Collart (Belgium, French-speaking); J Heaf (Denmark); P Finne and C Grönhagen-Riska (Finland); M Lassalle and C Couchoud (France); GA Ioannidis (Greece); R Palsson (Iceland); A Rustici, M Nichelatti, A Limido, A Molino, M Salomone, G Cappelli, E Arosio, F Antonucci, A Santoro, E Mancini, A Rosati, GM Frascà, G Gaffi, M Standoli, M Bonomini, L DiLiberato, S DiGiulio, A DiNapoli, D Torres, F Casino, C Zoccali, M Postorino, and AM Pinna (Italy); T Leivestad (Norway); P Castro de la Nuez and MA Pérez Valdivia (Spain,

Andalusia); R Alonso de la Torre, Á Rocas, and E Sánchez (Spain, Asturias); A Magaz, J Aranzabal, I Lampreabe, J Arrieta, M Rodrigo, and I Moina (Spain, Basque country); J González Cotorruelo and O García Ruíz (Spain, Cantabria); R González and C García-Renedo (Spain, Castile and Leon); G Gutiérrez Ávila and I Moreno Alía (Spain, Castile-La Mancha); E Arcos, J Comas, R Deulofeu, and J Twose (Spain, Catalonia); JM Ramos Aceitero and MA García Bazaga (Spain, Extremadura); E Bouzas-Caamaño and J Sánchez-Ibáñez (Spain, Galicia); O Zurriaga Llorens and M Ferrer Alamar (Spain, Valencian region); KG Prütz, L Bäckman, S Schön, A Seeberger, and B Rippe (Sweden); A Hoitsma and A Hemke (The Netherlands); All the staff of the UK Renal Registry and of the renal units submitting data (United Kingdom, England, Wales and Northern Ireland); and all of the Scottish renal units (United Kingdom, Scotland). Additionally, we acknowledge individuals from the French GN-PROGRESS study, including F Martínez, F Vrtovnik, and D Droz for adjudicating all IgA nephropathy cases, as well as participating center investigators X. Belenfant (Hôpital A Grégoire, Montreuil); B. Charpentier, A Durrbach (AP-HP Hôpital Bicêtre, Kremlin-Bicêtre); G Rostoker (Hôpital C Galien, Quincy/Senart); J Rossert, C Jacquot (AP-HP, Hôpital Européen G Pompidou, Paris); P Lang, P Remy (AP-HP Hôpital H. Mondor, Créteil); O. Kourilsky (Hôpital L. Michel, Evry); J-P. Grünfeld, D. Chauveau (AP-HP Hôpital Necker, Paris); G Deray, H Izzedine (AP-HP Hôpital Pitié Salpêtrière, Paris); C Legendre, F Martínez (AP-HP Hôpital Saint-Louis, Paris); and P Ronco, E Rondeau (AP-HP Hôpital Tenon, Paris).

## Author Contributions

Conceived and designed the experiments: K Kiryluk, AG Gharavi. Performed the experiments: K Kiryluk, Y Li, M Rohanizadegan, AG Gharavi. Analyzed the data: K Kiryluk, AG Gharavi. Contributed reagents/materials/analysis tools: M Choi, M Perola, K Kristiansson, A Viktorin, PK Magnusson, G Thorleifsson, U Thorsteinsdottir, K Stefansson, A Boland, C Wanner, KJ Jager, RP Lifton. Wrote the paper: K Kiryluk, AG Gharavi. DNA preparation: Y Li, M Rohanizadegan. Data management: Y Li, K Kiryluk. Subject recruitment and clinical characterization: S Sanna-Cherchi, H Suzuki, F Eitner, HJ Snyder, P Hou, F Scolari, C Izzì, M Gigante, L Gesualdo, S Savoldi, A Amoroso, D Cusi, P Zamboli, BA Julian, J Novak, RJ Wyatt, K Mucha, M Metzger, L Thibaudin, S Goto, D Maixnerova, HH Karnib, J Nagy, U Panzer, J Xie, N Chen, V Tesar, I Narita, F Berthoux, J Floege, B Stengel, H Zhang.

## References

- Schena FP (1990) A retrospective analysis of the natural history of primary IgA nephropathy worldwide. *Am J Med* 89: 209–215.
- Varis J, Rantala I, Pasternack A, Oksa H, Jantti M, et al. (1993) Immunoglobulin and complement deposition in glomeruli of 756 subjects who had committed suicide or met with a violent death. *J Clin Pathol* 46: 607–610.
- Suzuki K, Honda K, Tanabe K, Toma H, Nihei H, et al. (2003) Incidence of latent mesangial IgA deposition in renal allograft donors in Japan. *Kidney Int* 63: 2286–2294.
- Julian BA, Quiggins PA, Thompson JS, Woodford SY, Gleason K, et al. (1985) Familial IgA nephropathy. Evidence of an inherited mechanism of disease. *N Engl J Med* 312: 202–208.
- O'Connell PJ, Ibels LS, Thomas MA, Harris M, Eckstein RP (1987) Familial IgA nephropathy: a study of renal disease in an Australian aboriginal family. *Aust N Z J Med* 17: 27–33.
- Levy M (1989) Familial cases of Berger's disease and anaphylactoid purpura: more frequent than previously thought. *Am J Med* 87: 246–248.
- Li PK, Burns AP, So AK, Lai KN, Rees AJ (1992) Familial IgA nephropathy: a study of HLA class II allelotypes in a Chinese kindred. *Am J Kidney Dis* 20: 458–462.
- Scolari F, Amoroso A, Savoldi S, Mazzola G, Prati E, et al. (1999) Familial clustering of IgA nephropathy: further evidence in an Italian population. *Am J Kidney Dis* 33: 857–865.
- Gharavi AG, Yan Y, Scolari F, Schena FP, Frasca GM, et al. (2000) IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. *Nat Genet* 26: 354–357.
- Bisceglia L, Cerullo G, Forabosco P, Torres DD, Scolari F, et al. (2006) Genetic Heterogeneity in Italian Families with IgA Nephropathy: Suggestive Linkage for Two Novel IgA Nephropathy Loci. *Am J Hum Genet* 79: 1130–1134.
- Paterson AD, Liu XQ, Wang K, Magistroni R, Song X, et al. (2007) 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.
- Kiryluk K, Moldoveanu Z, Sanders JT, Eison TM, Suzuki H, et al. (2011) Aberrant glycosylation of IgA1 is inherited in both pediatric IgA nephropathy and Henoch-Schönlein purpura nephritis. *Kidney Int* 80: 79–87.
- Gharavi AG, Moldoveanu Z, Wyatt RJ, Barker CV, Woodford SY, et al. (2008) Aberrant IgA1 glycosylation is inherited in familial and sporadic IgA nephropathy. *J Am Soc Nephrol* 19: 1008–1014.
- Gharavi AG, Kiryluk K, Choi M, Li Y, Hou P, et al. (2011) Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* 43: 321–7.
- Fechally J, Farrall M, Boland A, Gale DP, Gut I, et al. (2010) HLA has strongest association with IgA nephropathy in genome-wide analysis. *J Am Soc Nephrol* 21: 1791–1797.
- Izzi C, Sanna-Cherchi S, Prati E, Belleri R, Remedio A, et al. (2006) Familial aggregation of primary glomerulonephritis in an Italian population isolate: Valtrompia study. *Kidney Int* 69: 1033–1040.
- Cullen M, Erlich H, Klitz W, Carrington M (1995) Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. *Am J Hum Genet* 56: 1350–1358.
- Cullen M, Noble J, Erlich H, Thorpe K, Beck S, et al. (1997) Characterization of recombination in the HLA class II region. *Am J Hum Genet* 60: 397–407.
- Jeffreys AJ, Ritchie A, Neumann R (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 9: 725–733.
- Lin PI, Vance JM, Pericak-Vance MA, Martin ER (2007) No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80: 531–538.
- Berthoux FC, Gagne A, Sabatier JC, Ducret F, Le Petit JC, et al. (1978) HLA-Bw35 and mesangial IgA glomerulonephritis. *N Engl J Med* 298: 1034–1035.
- Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, et al. (2009) High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS Genet* 5: e1000696. doi:10.1371/journal.pgen.1000696.

23. Nejentsev S, Howson JM, Walker NM, Szesko J, Field SF, et al. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 450: 887–892.
24. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, et al. (2008) HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* 57: 1084–1092.
25. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, et al. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476: 214–219.
26. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, et al. (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 43: 761–767.
27. Kastl SP, Speidl WS, Kaun C, Katsaros KM, Rega G, et al. (2008) In human macrophages the complement component C5a induces the expression of oncostatin M via AP-1 activation. *Arterioscler Thromb Vasc Biol* 28: 498–503.
28. Waldherr R, Rambauck M, Duncker WD, Ritz E (1989) Frequency of mesangial IgA deposits in a non-selected autopsy series. *Nephrol Dial Transplant* 4: 943–946.
29. Li L (1996) End-stage renal disease in China. *Kidney Int* 49: 287–301.
30. Koyama A, Igarashi M, Kobayashi M (1997) Natural history and risk factors for immunoglobulin A nephropathy in Japan. Research Group on Progressive Renal Diseases. *Am J Kidney Dis* 29: 526–532.
31. Nair R, Walker PD (2006) Is IgA nephropathy the commonest primary glomerulopathy among young adults in the USA? *Kidney Int* 69: 1455–1458.
32. Okpechi I, Swanepoel C, Duffield M, Mahala B, Wearne N, et al. (2011) Patterns of renal disease in Cape Town South Africa: a 10-year review of a single-centre renal biopsy database. *Nephrol Dial Transplant* 26: 1853–1861.
33. Hughson MD, Megill DM, Smith SM, Tung KS, Miller G, et al. (1989) Mesangiopathic glomerulonephritis in Zuni (New Mexico) Indians. *Arch Pathol Lab Med* 113: 148–157.
34. Fischer EG, Harris AA, Carmichael B, Lathrop SL, Cerilli LA (2009) IgA nephropathy in the triethnic population of New Mexico. *Clin Nephrol* 72: 163–169.
35. Smith SM, Tung KS (1985) Incidence of IgA-related nephritides in American Indians in New Mexico. *Hum Pathol* 16: 181–184.
36. Smith SM, Harford AM (1995) IgA nephropathy in renal allografts: increased frequency in Native American patients. *Ren Fail* 17: 449–456.
37. Casiro OG, Stanwick RS, Walker RD (1988) The prevalence of IgA nephropathy in Manitoba Native Indian children. *Can J Public Health* 79: 308–310.
38. O'Rourke DH, Raff JA (2010) The human genetic history of the Americas: the final frontier. *Curr Biol* 20: R202–207.
39. Borchers AT, Uibo R, Gershwin ME (2010) The geoepidemiology of type 1 diabetes. *Autoimmun Rev* 9: A355–365.
40. Green A, Gale EA, Patterson CC (1992) Incidence of childhood-onset insulin-dependent diabetes mellitus: the EURODIAB ACE Study. *Lancet* 339: 905–909.
41. Milo R, Kahana E (2010) Multiple sclerosis: geoepidemiology, genetics and the environment. *Autoimmun Rev* 9: A387–394.
42. Handel AE, Handunnetthi L, Giovannoni G, Ebers GC, Ramagopalan SV (2010) Genetic and environmental factors and the distribution of multiple sclerosis in Europe. *Eur J Neurol* 17: 1210–1214.
43. Shapira Y, Agmon-Levin N, Shoenfeld Y (2010) Defining and analyzing geoepidemiology and human autoimmunity. *J Autoimmun* 34: J168–177.
44. Berger J, Hinglais N (1968) Les dépôts intercapillaires d'IgA-IgG. *J Urol Nephrol (Paris)* 74: 694–695.
45. Suzuki H, Kiryluk K, Novak J, Moldoveanu Z, Herr AB, et al. (2011) The pathophysiology of IgA nephropathy. *J Am Soc Nephrol* 22: 1795–1803.
46. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, et al. (2010) New loci associated with kidney function and chronic kidney disease. *Nat Genet* 42: 376–384.
47. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, et al. (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 42: 570–575.
48. Gauderman WJ (2002) Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 155: 478–484.
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
50. Han B, Eskin E (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 88: 586–598.
51. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J (2006) Assessing heterogeneity in meta-analysis: Q statistic or I<sup>2</sup> index? *Psychol Methods* 11: 193–206.
52. Cordell HJ, Clayton DG (2002) A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 70: 124–141.