© 2012 International Society of Nephrology

Open

A family-based association study after genome-wide linkage analysis identified two genetic loci for renal function in a Mongolian population

Hansoo Park[1,2,10], Hyun-Jin Kim[1,3,10], Seungbok Lee[1,3,10], Yun Joo Yoo[4], Young Seok Ju[1,5], Jung Eun Lee[6], Sung-Il Cho[7], Joohon Sung[7], Jong-Il Kim[1,3,8,9,11] and Jeong-Sun Seo[1,3,5,8,9,11]

[1]Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Republic of Korea; [2]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [3]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Republic of Korea; [4]Department of Mathematics Education, Seoul National University, Seoul, Republic of Korea; [5]Macrogen, Seoul, Republic of Korea; [6]Division of Nephrology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; [7]Department of Epidemiology and Institute of Environment and Health, School of Public Health, Seoul National University, Seoul, Korea; [8]Psoma Therapeutics, Seoul, Republic of Korea and [9]Department of Biochemistry and Molecular Biology, Genomic Medicine Institute, Seoul National University College of Medicine, Seoul, Republic of Korea

The estimated glomerular filtration rate is a well-known measure of renal function and is widely used to follow the course of disease. Although there have been several investigations establishing the genetic background contributing to renal function, Asian populations have rarely been used in these genome-wide studies. Here, we aimed to find candidate genetic determinants of renal function in 1007 individuals from 73 extended families of Mongolian origin. Linkage analysis found two suggestive regions near 9q21 (logarithm of odds (LOD) 2.82) and 15q15 (LOD 2.70). The subsequent family-based association study found 2 and 10 significant single-nucleotide polymorphisms (SNPs) in each region, respectively. The strongest SNPs on chromosome 9 and 15 were rs17400257 and rs1153831 with *P*-values of 7.21×10^{-9} and 2.47×10^{-11}, respectively. Genes located near these SNPs are considered candidates for determining renal function and include *FRMD3*, *GATM*, and *SPATA5L1*. Thus, we identified possible loci that determine renal function in an isolated Asian population. Consistent with previous reports, our study found genes linked and associated with renal function in other populations.

Kidney International advance online publication, 19 December 2012; doi:10.1038/ki.2012.389

KEYWORDS: family-based association test; genome-wide linkage study; isolated Mongolian population; renal function

Correspondence: *Jeong-Sun Seo, Department of Biochemistry and Molecular Biology, Genomic Medicine Institute, Seoul National University College of Medicine, Daehakro, Yeongun-dong, Jongno-gu, Seoul 110-799, Republic of Korea. E-mail: jeongsun@snu.ac.kr*

[10]These authors contributed equally to this work.

[11]These senior authors contributed equally to this work.

Received 24 January 2012; revised 31 August 2012; accepted 14 September 2012

The decline in glomerular filtration rate (GFR), which is an overall indicator of renal function, has been recognized as a global health problem, leading to an increased risk of cardiovascular events and mortality.[1,2] Previous studies have provided evidence for genetic factors affecting renal function, showing heritability ranging from 0.41 to 0.75 in populations with risk factors such as hypertension or diabetes[3,4] and from 0.21 to 0.33 in general populations.[5–7]

To date, there have been several linkage studies to identify genetic loci determining renal function in individuals with renal disease or in the normal population.[8–11] Puppala *et al.* suggested candidate regions including 2q36.3 and 9q21.31–q21.33 in the Mexican–American population, and Schelling *et al.* reported the 1q43, 7q36.1, 8q13.3, and 18q23.3 regions using multiethnic diabetic populations.[9,11] There are several additional regions that have been suggested from other linkage studies, including 7p15.3–p13, 12p12.2, and 16q12.2–16q23.1.[8,10] In recent years, researchers have tried to determine in more detail the genetic basis of the estimated GFR (eGFR) through genome-wide association studies (GWAS), and have begun suggesting some genes or variants as determinants of eGFR or chronic kidney disease.[12–14] These studies, based on a large number of samples, have identified several variants showing a high level of significance and reproducibility near or within genes including *UMOD*, *SHROOM3*, *GATM*, and *SPATA5L1*.[12–15]

Although a number of genetic loci have been implicated from the genome-wide linkage and association studies, few studies have been carried out in Asian populations. In addition, some studies were focused on samples from patients having a specific disease, and the results might not reflect the renal function of the general population.[9,11,16–18] This study was conducted as a part of the GENDISCAN project (GENe DIScovery for Complex traits in large isolated families of Asians of the Northeast), which was designed to

investigate genetic influences on complex traits in extended families in Mongolia.[19–22] This project has several unique features compared with other studies: (1) the study population is isolated in a rural area and has relatively little ethnic admixture; (2) the subjects mostly consist of large extended families; and (3) the sample selection was unbiased by health status, and thus the samples represent the community-based population. These points strengthen the power of the genetic studies, and enable the identification of causal genetic loci for each phenotype.[23]

Family studies have a long history in human genetics. In particular, linkage analysis using families was successful in mapping for human oligogenic traits. In the past few years, along with the advances in genotyping technology, a population-based GWAS has become a popular tool for gene mapping of common complex diseases. However, the inability of the common variants identified by GWAS to explain the heritability of diseases has again led to interest in family-based studies, such as association studies based on linkage information.[24]

In this study, we aimed to investigate the genetic background of eGFR in isolated Mongolian families. Subsequently, the candidate loci were compared with those identified in previous studies on different populations, and the reproducibility of the results determined.

Table 1 | Characteristics of study participants

Characteristics	Linkage study *N* (%) or mean (s.d.)	Association study *N* (%) or mean (s.d.)
Subject information		
No. of families	73	54
No. of participants	1007	722
Minimum no. of individuals per family	4	6
Median no. of individuals per family	18	20
Maximum no. of individuals per family	54	54
Risk factor of renal function		
Age (years)	33 (16.3)	32 (16.0)
Female (%)	537 (53.3)	392 (54.3)
BMI (kg/m^2)	23.4 (4.2)	23.3 (4.1)
Smoking (yes)	180 (17.9)	117 (16.2)
Under antihypertensive treatment (yes)	122 (12.1)	82 (11.4)
Fasting glucose (mg/dl)	95.4 (18.6)	95.3 (19.9)
SBP (mm Hg)	115 (16.1)	114.2 (15.7)
Renal function		
eGFR[a] (ml/min per 1.73 m^2)		
Male	99.8 (26.5)	101.2 (26.6)
Female	96.4 (22.9)	96.5 (22.6)
Total	98.0 (24.7)	98.6 (24.6)
No. of CKD	23 (2.3)	13 (1.8)

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; SBP, systolic blood pressure.
[a]eGFR was estimated by the MDRD-6 (Modification of Diet in Renal Disease).[25]

RESULTS

Descriptive characteristics of study subjects

The descriptive characteristics of the study subjects are shown in Table 1. This study consists of two steps of analysis: (1) genome-wide linkage analysis followed by (2) a family-based association study. The study population used for the linkage analysis includes 73 families comprising 1007 individuals. The number of individuals per family ranged from 4 to 54. Of these, 722 individuals from 54 families were genotyped with a single-nucleotide polymorphism (SNP) microarray and chosen for the subsequent association study. The minimum and maximum numbers of individuals per family were 6 and 54, respectively. As shown in Table 1, the distribution of each trait in samples for the association study shows no difference to that in samples for the linkage study. The eGFR, which represents the renal function of each subject, was calculated according to the MDRD-6 (Modification of Diet in Renal Disease) equation.[25]

Genetic evidence of eGFR from familial correlation and heritability analyses

To identify the evidence of genetic factors for eGFR levels, we calculated familial correlation coefficients in familial pairs and estimated the heritability, which is a useful concept to evaluate the amount of genetic contribution to total phenotypic variation (Supplementary Table S1 online). Overall, there were 760 parent–offspring pairs, 623 sibling pairs, 725 avuncular pairs, 520 cousin pairs, and 94 spouse pairs. In the age- and sex-adjusted model (Model 1), sibling correlation was significant ($r = 0.20$, $P < 0.01$), and among all the subtype pairs the highest correlation was shown in brother–brother pairs ($r = 0.37$, $P < 0.01$). The narrow-sense heritability for eGFR was 0.27 ($P < 0.01$). In the multivariable-adjusted model (Model 2), parent–offspring and sibling correlations were estimated to be 0.10 ($P = 0.02$) and 0.18 ($P < 0.01$), respectively. Similar to the age- and sex-adjusted model, the highest correlation was shown in brother–brother pairs ($r = 0.35$, $P < 0.01$). Among subtypes of parent–offspring pairs, mother–daughter pairs had a significant familial correlation ($r = 0.14$, $P = 0.03$). As shown in Supplementary Table S1 online, the narrow-sense heritability in the multivariable-adjusted model was slightly higher than that of the age- and sex-adjusted model ($h^2 = 0.29$, $P < 0.01$). As a result, in both models of analyses, the overall familial correlations for genetically related pairs were significant. However, the correlations for spouse pairs, which indicate shared environmental or assortative mating effect, were not significant. This correlation pattern may suggest the importance of genetic components for eGFR, along with the significant heritability.

Genome-wide linkage scan for eGFR

Results for suggestive linkages with a logarithm of odds (LOD) score greater than 1.9 are reported in Figure 1 and Table 2.[26] The two suggestive linkage peaks in this study were detected on chromosomes 9 and 15 (Figure 1a). The linkage

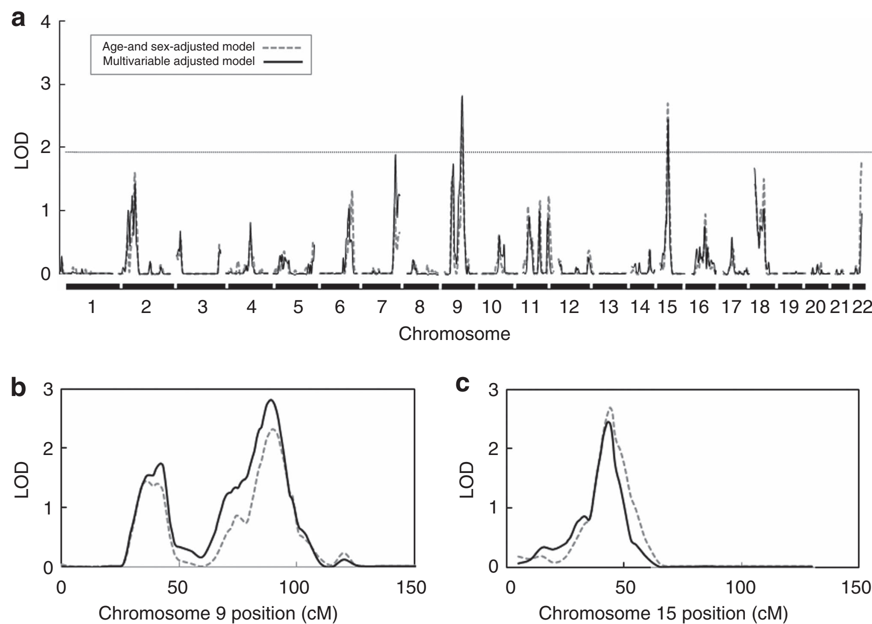

Figure 1 | Multipoint logarithm of odds (LOD) results of the genome-wide linkage scan for estimated glomerular filtration rate. (**a**) The linkage results of autosomal regions. (**b**) The linkage peak on chromosome 9, which shows the highest LOD score in this study (LOD = 2.8). (**c**) The linkage peak on chromosome 15, reaching the threshold level for suggestive linkage (LOD > 1.9).

Table 2 | Chromosomal regions from genome-wide linkage scan for eGFR (LOD > 1.9[a])

Chromosome (location, cM)	Nearest marker	Cytogenetic region[b]	LOD-1 interval (cM)	LOD score[c] (empirical *P*-value[d]) Model 1	Model 2	Previous linkage evidence
9 (90)	D9S307	9q21.33	81–96	2.32 (0.0005)	2.82 (0.0003)	Puppala *et al.*[9] and Arar *et al.*[5]
15 (44)	D15S214	15q15.1	38–50	2.70 (0.0003)	2.45 (0.0006)	Pattaro *et al.*[27]

Abbreviations: cM, centimorgan; eGFR, estimated glomerular filtration rate; LOD, logarithm of odds.
[a]Suggestive linkage regions under Lander and Kruglyak's guidelines.[26]
[b]NCBI Build 36, region where the nearest marker is located.
[c]LOD scores in Model 1 and Model 2 were estimated under age- and sex-adjusted model and multivariable-adjusted model, respectively; see the Materials and Methods section for details.
[d]Empirical *P*-value was estimated by 10,000 simulations.

peak with the highest LOD score was observed on chromosome 9q21 (LOD = 2.8, Model 2), with the nearest marker being D9S307, and the linkage interval encompassing a maximum 1-LOD ranged from 81 to 96 centimorgan (cM) (Figure 1b). The empirical *P*-value after 10,000 simulation replicates was 0.0003. The second highest peak was found on chromosome 15q15 (LOD = 2.70, Model 1), with the nearest marker being D15S214. The linkage interval of this peak ranged from 38 to 50 cM (Figure 1c), and the empirical *P*-value was 0.0003. We also explored the consistency of our linkage results with those in other studies (Table 2).[5,9,27]

Family-based association test for fine mapping of the suggestive linkage peaks

We identified two potential linkage peaks on chromosomes 9 and 15. For the additional association study, we focused on interval regions under these peaks to carry out the family-based association test (FBAT) for eGFR. In Table 3, family-based association results with *P*-values $<1.0\times10^{-5}$ corresponding to a Bonferroni correction are reported, and the nearby *RefSeq* genes are explored within 150 kb upstream or downstream of each SNP. The list of additional significant SNPs, which were in strong linkage disequilibrium (LD) ($r^2 \geqslant 0.8$) with the SNPs in Table 3, is shown in Supplementary Table S2 online. The two SNPs reaching the stringent level of significance were identified at chromosome 9q21.32. The more strongly associated SNP was rs17400257 ($P = 7.21\times10^{-9}$, Model 2), which was in low LD with the other, rs6559725 ($P = 9.12\times10^{-7}$, Model 2) ($r^2 = 0.00$). The closest gene to both SNPs was *FRMD3*, which is known as a diabetic nephropathy susceptibility gene (Figure 2a).[16–18] The SNP rs6559725 is located in the intronic region of *FRMD3*, whereas rs17400257 is 45 kb downstream of this gene. At chromosome 15q, 10 independent SNPs reached the threshold level of significance, as shown in Table 3. Of these, the strongest association was found for rs1153831, an intergenic SNP near *SLC30A4*, *GATM*, and *SPATA5L1* with a *P*-value of 2.47×10^{-11} in Model 1. Two SNPs, rs1153829

Table 3 | Family-based association results for eGFR under supportive linkage peaks (P-value < 1.0×10^{-5})

Chr	SNP ID	Locus	Position[a]	Genotype (RA)	MAF[b]	Model[c]		Nearby gene(s)[d]
						P-value (Model 1)	*P*-value (Model 2)	
9	rs17400257	9q21.32	85,002,773	A/C (C)	0.04	4.37×10^{-6}	7.21×10^{-9}	***FRMD3***, *RASEF*
	rs6559725	9q21.32	85,252,250	A/G (G)	0.04	2.60×10^{-4}	9.12×10^{-7}	***FRMD3***
15	rs1153831	15q21.1	43,559,740	A/G (G)	0.07	2.47×10^{-11}	4.11×10^{-9}	*SLC30A4*, ***GATM***, ***SPATA5L1***, *C15orf48*, *MIR147B*, *C15orf21*, *PLDN*
	rs12908295	15q21.1	47,576,537	A/G (G)	0.23	9.80×10^{-7}	7.93×10^{-9}	*C15orf33*, *FGF7*, *DTWWD1*
	rs2305707	15q21.2	49,356,702	A/G (A)	0.16	1.87×10^{-6}	1.13×10^{-7}	*CYP19A1*, *GLDN*
	rs8040312	15q14	37,067,395	A/G (A)	0.35	1.98×10^{-4}	2.90×10^{-7}	—
	rs10163098	15q21.2	49,460,922	A/G (A)	0.06	5.05×10^{-5}	3.34×10^{-6}	*GLDN*, *CYP19A1*, *DMXL2*
	rs8037395	15q15.3	42,367,146	A/C (C)	0.11	3.93×10^{-5}	4.63×10^{-6}	*CASC4*, *FRMD5*
	rs8042458	15q21.2	47,951,441	A/G (A)	0.10	3.29×10^{-5}	5.60×10^{-6}	*ATP8B4*
	rs11854805	15q21.1	46,463,166	C/T (T)	0.04	3.87×10^{-1}	6.88×10^{-6}	*FBN1*, *DUT*, *SLC12A1*
	rs16968439	15q14	37,290,484	C/T (C)	0.23	1.69×10^{-3}	9.21×10^{-6}	*C15orf54*
	rs17718330	15q15.1	37,910,641	C/T (T)	0.05	2.38×10^{-6}	9.38×10^{-6}	*GRP176*, *FSIP1*, *EIF2AK4*

Abbreviations: Chr, chromosome; eGFR, estimated glomerular filtration rate; MAF, minor allele frequency; RA, reference allele; SNP, single-nucleotide polymorphism.
[a]SNP positions are based on NCBI Build 36.
[b]MAF was estimated from parental alleles.
[c]Model 1 and Model 2 represents age- and sex-adjusted model and multivariable-adjusted model, respectively; see the Materials and Methods section for details.
[d]Nearby gene(s) are based on *RefSeq* genes (NCBI Build 36) and within 150 kb upstream or downstream. The gene(s) closest to the SNP are listed first and are in boldface if the gene(s) are identified by previous genome-wide association studies.[13–18,28]

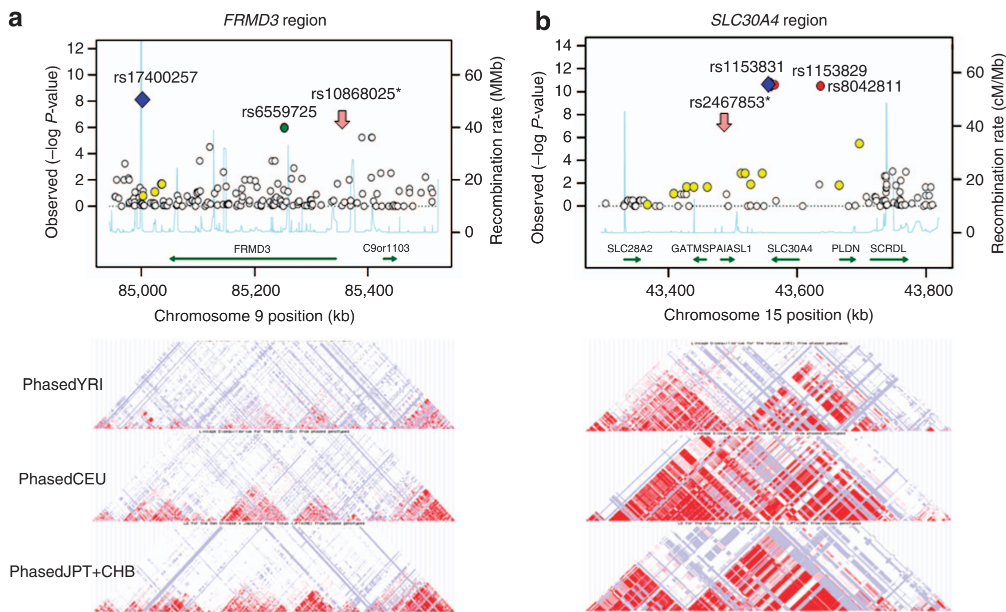

Figure 2 | Regional association plots and linkage disequilibrium (LD) structures in different ethnicities (YRI: African, CEU: European, and JPT + CHB: Japanese and Chinese) based on HapMap data. The blue diamond indicates the most significant single-nucleotide polymorphism (SNP) of each region, and nearby SNPs are color-coded to show their LD relationships with the top SNP ($r^2<0.2$; white, $0.2\leqslant r^2<0.4$; yellow, $0.4\leqslant r^2<0.8$; orange, $r^2\geqslant0.8$; red). Asterisk indicates previously reported SNPs to be associated with estimated glomerular filtration rate.[12,13,16] Local LD is reflected by the recombination rates from HapMap data. (**a**) The association plot near 9q21.32. Green dot is another candidate SNP in *FRMD3* that has no LD with our top SNP on chromosome 9. (**b**) The association plot near 15q21.1.

and rs8042811, were in strong LD with rs1153831 (Supplementary Table S2 online); rs1153829 is located in the 3′ untranslated region of *SLC30A4*, and rs8042811 is a nonsynonymous SNP of *C15orf21*. Interestingly, *GATM* and *SPATA5L1*, which have been identified in previous GWAS, are within 200 kb of these three SNPs (Figure 2b).[12–15] The second most significant SNP, rs12908295, was located within an intron of *C15orf33*, and we identified eight additional significant SNPs associated with eGFR, which were distributed over three different genetic loci (15q14, 15q15.1–3, and 15q21.1–2).

To verify whether our top SNP in each chromosome was within the same LD block with previously reported SNPs, we checked the LD structures near top SNPs from HapMap

(YRI: African, CEU: European, and JPT + CHB: Japanese and Chinese) (Figure 2). Whereas rs17400257 and rs6559725 in 9q21.32 were located in a different LD block to rs10868025, which had been reported by Pezzolesi *et al.*,[16] candidate SNPs in 15q21.1 (rs1153831, rs1153829, and rs8042811) were within the same LD block as rs2467853, a known susceptibility SNP of renal function,[12,13] in HapMap JPT + CHB.

DISCUSSION

This study was conducted to explore the genetic basis of eGFR in large extended families of Mongolian origin.[19–22] We estimated the renal function of each individual using the MDRD-6 equation, and applied linkage analysis and a subsequent FBAT to our study population. We identified suggestive linkage regions on chromosomes 9 and 15, which have been previously reported to be associated with renal function or disease in other populations. A subsequent FBAT under linkage regions revealed that 2 and 10 independent SNPs on chromosomes 9 and 15 were significantly related to the eGFR, and the strongest association signal in each chromosome included the candidate genes identified by previous GWAS (*FRMD3*, *GATM*, and *SPATA5L1*). In addition, we discovered new susceptibility loci for renal function at 15q21, 15q14, and 15q15 (*SLC30A4*, *C15orf33*, *FGF7*, *CYP19A1*, *GLDN*, *CASC4*, *ATP8B4*, *FBN1*, *C15orf54*, and *GRP176*).

There have been several previous linkage studies of renal function in diverse populations.[5,8–11,27] However, the renal function of Asian populations, especially Northeast Asians, has been rarely studied. Here, we conducted genome-wide linkage analysis on an isolated Mongolian population and found suggestive linkage peaks near 9q21 and 15q15. In previous linkage studies with Mexican–American subjects, the peak on chromosome 9q21 was reported as a candidate region for renal function, and these studies also used a similar method to measure the renal function of each individual.[5,9] The region on chromosome 15 was also previously identified in a linkage study of Caucasian populations.[27] As both of the suggestive linkage regions we identified were also identified in studies in other populations, we decided to focus and emphasize the targeting of these linkage regions in the subsequent family-based association study.

Our association results revealed that genes near the strongest association from each chromosome were consistent with candidates identified in other population-based GWAS. On chromosome 9, both the significant SNPs are located near the *FRMD3* gene. As there is very low LD between the two SNPs ($r^2 = 0.00$), each SNP might separately affect the eGFR value, making it more reliable that *FRMD3* might be a key gene in determining renal function in the general population. This *FRMD3* gene, which encodes the protein FERM domain containing 3, is a well-known susceptibility gene for diabetic nephropathy. It was first reported in European–American subjects with type 1 diabetes, and replicated in Japanese and African–American patients with type 2 diabetes.[16–18] As *FRMD3* was identified in patients with both types of diabetes having nephropathy or end-stage renal disease, we can predict that this gene might have a role in renal function. Freedman *et al.*[18] reported that variants in *FRMD3* were associated with type 2 diabetic nephropathy, but not with type 2 diabetes *per se*.

Several GWAS studies based on a large number of samples have suggested some loci near *GATM* and *SPATA5L1* genes as determinants of renal function in Caucasian populations, and another locus of *GATM* was also found to be significant in a study of patients with African ancestry.[13–15,28] In this study, the strongest candidate locus on chromosome 15 is also located near these genes (Table 3), and the LD block of this region partly overlaps with that reported in the previous study (Figure 2).[13] As the previous GWAS studies used a method similar to ours to estimate renal function, this overlapping result might be more important.

For both suggestive linkage regions, we could replicate candidate loci identified by previous GWAS on renal function at the gene level. However, our candidate SNPs were not consistent with well-known susceptibility SNPs for eGFR. Apart from the SNPs reported by Freedman *et al.*, the candidate SNPs identified in other studies described above were not included in our platform, and the SNPs reported by Freedman *et al.* were not significant in our study. However, as they conducted gene–gene interaction analysis and used different statistical methods for analysis, direct comparison would not be appropriate.

Other than *GATM* and *SPATA5L1*, the gene nearest to the most significant SNP on chromosome 15 is *SLC30A4*. Unlike the genes described above, there has been no evidence that this is associated with renal function or renal disease. Several studies have suggested that solute carrier gene families might have some roles in renal function.[14,28] In 2010, Köttgen *et al.*[14] reported new loci for renal function, located near several solute carrier genes such as *SLC22A2*, *SLC6A13*, *SLC7A9*, and *SLC34A1*. Recently, a replication study with known renal loci was performed in African Americans, and variants near *SLC22A2* and *SLC6A13* were shown to be replicated.[28] Although *SLC30A4* is expressed weakly in the kidney (four transcripts per million in the expressed sequence tag profile of UniGene, NCBI), it might be another candidate gene for renal function, taking the accumulative evidences on solute carrier genes for renal function. In addition, rs8042811, a nonsynonymous SNP of *C15orf21*, was shown to have a perfect LD with the top SNP of chromosome 15. Even if the function of this gene has rarely been explored so far, it needs to be regarded as one of the candidates for renal function, considering that none of the SNPs suggested in Table 3 changes an amino acid of the corresponding protein. The BLOSUM score for predicted change (C91R) is '-3', which is the second lowest score among all possible amino-acid exchanges.[29]

We identified the additional candidate genes associated with eGFR such as *C15orf33*, *FGF7*, *CYP19A1*, *GLDN*, *CASC4*, *ATP8B4*, *FBN1*, *C15orf54*, and *GRP176* on

chromosome 15. Of these, an interesting gene with respect to renal function is *FGF7*, which is a member of the fibroblast growth factor family. It has been previously reported that FGF-7 levels modulate the extent of urteric bud growth during development, as well as the number of nephrons, which might determine the GFR of each individual.[30] Although the association of this gene with renal function cannot be concluded, its functional impact on kidney development needs to be considered in further studies on renal function or disease.

The sample size of our study is rather smaller than those of other previous association studies. However, several factors in our study design may enable us to detect similar candidate loci and replicate previous results at the gene level.[23] First, we used large extended families in an isolated population for gene-mapping studies. An isolated population is highly suitable for genetic research because of environmental and phenotypic homogeneity, decreased genetic heterogeneity, restricted geographical distribution, and good genealogical records.[31] In particular, extended multigeneration pedigrees with a small number of founders are known to enhance the genetic power.[32] Second, we used a two-stage strategy including genome-wide linkage and family-based association analyses. This approach might be distinct from previous reports, such as linkage analysis alone and GWAS in a population-based design. Our strategy, which performs fine mapping in the presence of linkage, facilitates the detection of more refined loci than linkage analysis alone.[33] Although GWAS has considerably contributed to understanding the genetic basis of complex traits, one of the issues in GWAS is to reduce the false-positive rate in multiple testing. Combining the linkage information may provide enhanced power to detect true associations. In addition, FBAT is robust against population stratification compared with population-based GWAS.[23,34]

In conclusion, this study aimed at revealing the genetic background determining renal function in the Northeast Asian population of Mongolia, and has identified some loci described in previous reports in other populations. The candidate loci that were reproducibly found in our study might have significant roles as determinants of renal function regardless of ethnicity.

MATERIALS AND METHODS

Study subjects and genomic DNA extraction

The subjects used for this study were recruited from the GENDISCAN project, which was initiated to discover the susceptibility loci for common traits in Asian populations.[19–22] In 2006, this project recruited 2008 participants residing in Dashbalbar, in Dornod Province of Mongolia, which is a geographically isolated region in Northeast Asia. Of the total sample, we selected 1007 subjects (about 50.1%) from 73 families for a linkage analysis, who have appropriate pedigree structures and phenotypic information. The pedigree structure of this study population is highly complex, with both multiple generations and a large number of siblings. Family relationships identified from personal interviews were further validated genetically by checking Mendelian inheritance using PREST (Version 3.02).[35] For the subsequent association analysis in the presence of linkage, we chose 722 individuals from 54 families from linkage samples, which were not only available for SNP genotyping but also comprised large pedigrees. Peripheral venous blood samples from study subjects were collected for DNA extraction, and the genomic DNA of each sample was extracted according to standard protocols. We obtained informed consent from all study subjects, and the study protocols were approved by the institutional review board of Seoul National University (approval number, H-0307-105-002).

Phenotype measurement

The serum creatinine level of each sample was measured with HITACHI 7180 (Hitachi, Tokyo, Japan) by the Jaffe method.[36] The eGFR, which is a value representing renal function, was calculated using the MDRD-6 equation: eGFR (ml/min per 1.73 m^2) $= 170 \times \mathrm{SCr(mg/dl)}^{-0.999} \times \mathrm{age}^{-0.176} \times \mathrm{BUN(mg/dl)}^{-0.170} \times \mathrm{Alb(g/dl)}^{0.318} \times$ (0.762 if female)$\times$(1.18 if black), where SCr is the serum creatinine concentration, BUN is the blood urea nitrogen concentration, and Alb is the serum albumin level.[25] To meet the normality assumptions for variance component analyses, we tested the distribution of our phenotype before analyses. Because of the non-normal distribution of eGFR values, we normalized this trait with inverse normal transformation, which can reduce deviations from normality and the effect of outliers. Normalized eGFR was first regressed on age and sex (age- and sex-adjusted model; Model 1), and then body mass index, glucose, hypertension treatment, systolic blood pressure, and smoking status were also included as covariates for an additional model (multivariable-adjusted model; Model 2). The phenotypic residual values (observed–expected probability) were obtained from two different models, and they were used in all the following analyses including linkage and association.

Estimation of familial correlation and heritability

To explore the genetic background of renal function, we estimated heritability and familial correlations between family pairs. Familial correlation was calculated using the *FCOR* option in the Statistical Analysis for Genetic Epidemiology (S.A.G.E.) version 6.0 software.[37] *FCOR* estimates the intrafamilial correlations between all familial pairs including parent–offspring, sibling, avuncular, cousin, and spousal pairs. In addition, narrow-sense heritability (i.e., the proportion of phenotype variance attributable to additive genetic variance) was estimated by the variance components approach using Sequential Oligogenic Linkage Analysis Routines (SOLAR) version 4.2.7.[38]

Genome-wide linkage scan

We performed a genome-wide multipoint linkage scan to identify genetic loci associated with renal function in isolated Mongolian families. Seventy-three families comprising 1007 family members were genotyped for 1039 short tandem repeat markers. The detailed methods for genotyping error detection and correction are described in previous studies.[19–22] For multipoint linkage scan, multipoint identity-by-descent was calculated at each 1 cM distance using the LOKI package.[39] SOLAR version 4.2.7[38] was used for the genome-wide linkage scan. Empirical *P*-value of LOD scores was also obtained by the '*lodadj*' option implementing 10,000 permutations, and we determined suggestive linkage as an LOD score > 1.9.[26]

FBAT under linkage peaks

We tested the family-based association under two regions identified from the linkage analysis. Of the samples used for the linkage study, 54 families comprising 722 family members were genotyped by Illumina 610K Quad Beadchip (San Diego, CA). Before the FBAT, we checked the quality of the genotype data by several steps. The details for genotyping error correction are described in previous GENDISCAN studies.[21,22] After the genotype error correction, call rate and error rate of SNPs were assessed, and genotypes with a call rate $<99\%$ or an error rate $>1\%$ were excluded from analysis. In addition, we also removed genotypes with Hardy–Weinberg equilibrium *P*-values $<1.0\times10^{-6}$ or minor allele frequency $<1\%$, and, finally, 2467 SNPs on chromosome 9 and 2592 SNPs on chromosome 15 were used for analysis. The FBAT was performed using the PBAT tool in HelixTree software version 6.4 (GoldenHelix, Bozeman, MT).[40]

FBAT is an extension of the transmission disequilibrium test, which is an absolutely nonparametric method without assumptions about the model and distribution of disease. It incorporates additional conditions such as general pedigree, missing founders, and complex disease. The general FBAT statistic is as follows:

$$U = \sum T_{ij}(X_{ij} - E(X_{ij} \mid S_i))$$

where i is pedigree index and j is jth non-founders in the pedigree; T_{ij} is a phenotypic residual; and $E(X_{ij}|S_i)$ is the expected marker score under the null hypothesis. For a large sample size, U can be normalized as below to yield variable Z with a distribution approximating the normal $N(0,1)$:[34]

$$Z = U/\sqrt{(\mathrm{var}(U))}$$

We used the 'linkage and no association (sandwich variance)' null hypothesis to test the association in the presence of linkage. This hypothesis is valid when the same sample set was used in both linkage and association studies. In addition, our extended families are composed of many generations and multiple offspring. When families with multiple offspring are used, an association test that independently treats multiple offspring would not be appropriate because of the pattern of identity-by-descent. In this case, the empirical variance can be used and it requires correlation patterns between all family members.[34] To estimate a more robust variance in large extended pedigrees, we used the 'sandwich variance' method. This method can test family-based association without inflation of false-positive errors arising from ignoring correlations between family members.[41]

We used the generalized estimating equation for FBAT (FBAT-GEE), and the association results were generated under an additive genetic model. A significance level of 1.0×10^{-5} was used corresponding to a Bonferroni correction. For additional LD information, r^2 values were estimated among significant SNPs using the Haploview software version 4.1 (Broad Institute of Harvard and MIT, Cambridge, MA).

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

This work was supported by the Korean Ministry of Education, Science and Technology (Grant No. 2003-2001558).

SUPPLEMENTARY MATERIAL

Table S1. Familial correlations between family pairs for eGFR.
Table S2. The significant SNPs ($P<1.0\times10^{-5}$) which were in strong LD ($r^2 \geqslant 0.8$) with listed SNPs in Table 3.

Supplementary material is linked to the online version of the paper at http://www.nature.com/ki

REFERENCES

1. Zhang R, Zheng LQ, Sun ZQ *et al.* Decreased glomerular filtration rate is associated with mortality and cardiovascular events in patients with hypertension: a prospective study. *PloS One* 2011; **6.**
2. Stevens LA, Coresh J, Greene T *et al.* Medical progress—assessing kidney function—measured and estimated glomerular filtration rate. *New Engl J Med* 2006; **354**: 2473–2483.
3. Bochud M, Elston RC, Maillard M *et al.* Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean). *Kidney Int* 2005; **67**: 61–69.
4. Langefeld CD, Beck SR, Bowden DW *et al.* Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am J Kidney Dis* 2004; **43**: 796–800.
5. Arar NH, Voruganti VS, Nath SD *et al.* A genome-wide search for linkage to chronic kidney disease in a community-based sample: the SAFHS. *Nephrol Dial Transplant* 2008; **23**: 3184–3191.
6. MacCluer JW, Scavini M, Shah VO *et al.* Heritability of measures of kidney disease among Zuni Indians: The Zuni Kidney Project. *Am J Kidney Dis* 2010; **56**: 289–302.
7. Fox CS, Yang Q, Cupples LA *et al.* Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J Am Soc Nephrol* 2004; **15**: 2457–2461.
8. Mottl AK, Vupputuri S, Cole SA *et al.* Linkage analysis of glomerular filtration rate in American Indians. *Kidney Int* 2008; **74**: 1185–1191.
9. Puppala S, Arya R, Thameem F *et al.* Genotype by diabetes interaction effects on the detection of linkage of glomerular filtration rate to a region on chromosome 2q in Mexican Americans. *Diabetes* 2007; **56**: 2818–2828.
10. Rao M, Mottl AK, Cole SA *et al.* Meta-analysis of genome-wide linkage scans for renal function traits. *Nephrol Dial Transplant* 2012; **27**: 647–656.
11. Schelling JR, Abboud HE, Nicholas SB *et al.* Genome-wide scan for estimated glomerular filtration rate in multi-ethnic diabetic populations: the Family Investigation of Nephropathy and Diabetes (FIND). *Diabetes* 2008; **57**: 235–243.
12. Chambers JC, Zhang W, Lord GM *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* 2010; **42**: 373–375.
13. Kottgen A, Glazer NL, Dehghan A *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* 2009; **41**: 712–717.
14. Kottgen A, Pattaro C, Boger CA *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* 2010; **42**: 376–384.
15. Pattaro C, De Grandi A, Vitart V *et al.* A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level. *BMC Med Genet* 2010; **11**: 41.
16. Pezzolesi MG, Poznik GD, Mychaleckyj JC *et al.* Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 2009; **58**: 1403–1410.
17. Maeda S, Araki S, Babazono T *et al.* Replication study for the association between four loci identified by a genome-wide association study on European American subjects with type 1 diabetes and susceptibility to diabetic nephropathy in Japanese subjects with type 2 diabetes. *Diabetes* 2010; **59**: 2075–2079.
18. Freedman BI, Langefeld CD, Lu L *et al.* Differential effects of MYH9 and APOL1 risk variants on FRMD3 association with diabetic ESRD in African Americans. *PLoS Genet* 2011; **7**: e1002150.
19. Ju YS, Park H, Lee MK *et al.* A genome-wide Asian genetic map and ethnic comparison: the GENDISCAN study. *BMC Genomics* 2008; **9**: 554.
20. Lee MK, Woo SJ, Kim JI *et al.* Replication of a glaucoma candidate gene on 5q22.1 for intraocular pressure in mongolian populations: the GENDISCAN Project. *Invest Ophthalmol Vis Sci* 2010; **51**: 1335–1340.
21. Paik SH, Kim HJ, Lee S *et al.* Linkage and association scan for tanning ability in an isolated Mongolian population. *BMB Rep* 2011; **44**: 741–746.
22. Paik SH, Kim HJ, Son HY *et al.* Gene mapping study for constitutive skin color in an isolated Mongolian population. *Exp Mol Med* 2011.
23. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. *Nat Rev Genet* 2011; **12**: 465–474.
24. Manolio TA, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
25. Levey AS, Bosch JP, Lewis JB *et al.* A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction

equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999; **130**: 461–470.
26. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
27. Pattaro C, Aulchenko YS, Isaacs A *et al*. Genome-wide linkage analysis of serum creatinine in three isolated European populations. *Kidney Int* 2009; **76**: 297–306.
28. Liu CT, Garnaas MK, Tin A *et al*. Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. *PLoS Genet* 2011; **7**: e1002264.
29. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 1992; **89**: 10915–10919.
30. Qiao J, Uzzo R, Obara-Ishihara T *et al*. FGF-7 modulates ureteric bud growth and nephron number in the developing kidney. *Development* 1999; **126**: 547–554.
31. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. *Genome Biol* 2008; **9**: 109.
32. Arcos-Burgos M, Muenke M. Genetics of population isolates. *Clin Genet* 2002; **61**: 233–247.
33. Lake SL, Blacker D, Laird NM. Family-based tests of association in the presence of linkage. *Am J Hum Genet* 2000; **67**: 1515–1525.
34. Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* 2006; **7**: 385–394.
35. McPeek MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet* 2000; **66**: 1076–1094.
36. Junge W, Wilke B, Halabi A *et al*. Determination of reference intervals for serum creatinine, creatinine excretion and creatinine clearance with an enzymatic and a modified Jaffe method. *Clin Chim Acta* 2004; **344**: 137–148.
37. Keen KJ, Elston RC. Robust asymptotic sampling theory for correlations in pedigrees. *Stat Med* 2003; **22**: 3229–3247.
38. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**: 1198–1211.
39. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997; **61**: 748–760.
40. Lange C, DeMeo D, Silverman EK *et al*. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; **74**: 367–369.
41. Borecki IB, Province MA. Genetic and genomic discovery using family studies. *Circulation* 2008; **118**: 1057–1063.

This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/