

Rare mutations associating with serum creatinine and chronic kidney disease

Gardar Sveinbjornsson¹, Evgenia Mikaelsdottir¹, Runolfur Palsson^{2,5}, Olafur S. Indridason⁵, Hilma Holm⁷, Aslaug Jonasdottir¹, Agnar Helgason^{1,3}, Snaevar Sigurdsson¹, Adalbjorg Jonasdottir¹, Asgeir Sigurdsson¹, Gudmundur Ingi Eyjolfsson⁸, Olof Sigurdardottir⁹, Olafur Th. Magnusson¹, Augustine Kong^{1,4}, Gisli Masson¹, Patrick Sulem¹, Isleifur Olafsson⁶, Unnur Thorsteinsdottir^{1,2}, Daniel F. Gudbjartsson^{1,4,*} and Kari Stefansson^{1,2,*}

¹deCODE Genetics, 101 Reykjavik, Iceland, ²Faculty of Medicine, ³Department of Anthropology, ⁴School of Engineering and Natural Sciences, University of Iceland, 101 Reykjavik, Iceland, ⁵Internal Medicine Services, ⁶Department of Clinical Biochemistry, Landspítali - The National University Hospital of Iceland, 101 Reykjavík, Iceland, ⁷Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA, ⁸Icelandic Medical Center (Laeknasetridd) Laboratory in Mjodd (RAM), 109 Reykjavik, Iceland and ⁹Department of Clinical Biochemistry, Akureyri Hospital, 600 Akureyri, Iceland

Received July 16, 2014; Revised and Accepted July 29, 2014

Chronic kidney disease (CKD) is a complex disorder with a strong genetic component. A number of common sequence variants have been found to associate with serum creatinine (SCr), estimated glomerular filtration rate (eGFR) and/or CKD. We imputed 24 million single-nucleotide polymorphisms and insertions/deletions identified by whole-genome sequencing of 2230 Icelanders into 81 656 chip-typed individuals and 112 630 relatives of genotyped individuals over the age of 18 with SCr measurements. The large set of sequenced individuals allowed accurate imputation of variants to a minor allele frequency (MAF) of 0.1%. We tested the imputed variants for association with SCr. In addition to replicating established loci, we discovered missense and loss-of-function variants associating with SCr in three solute carriers (*SLC6A19*, *SLC25A45* and *SLC47A1*) and two E3 ubiquitin ligases (*RNF186* and *RNF128*). All the variants are within coding sequences and all but one are rare (MAF <2%) with SCr effects between 0.085 and 0.129 standard deviations. These rare variants have a larger effect on SCr than previously reported common variants, explaining 0.5% of the variability of SCr in Icelanders in addition to the 1% already accounted for. We tested the five variants associating with SCr for association with CKD in an Icelandic sample of 15 594 cases and 291 428 controls. Three of the variants also associated with CKD. These variants may either affect kidney function or creatinine synthesis and excretion. Of note were four mutations in *SLC6A19* that associate with reduced SCr, three of which have been shown to cause Hartnup disease.

INTRODUCTION

Chronic kidney disease (CKD) is a common disorder that can progress to kidney failure and is associated with an increased risk of cardiovascular morbidity and death (1). Studies indicate a dramatic increase in the prevalence of CKD with advancing age (2,3). With increased lifespan, the burden of CKD has grown steadily in the Western world (4) and is recognized as a public health problem substantially impacting the health

service (5). The cause of CKD is not always known and frequently appears multifactorial, with hypertension and diabetes mellitus being the most important contributing factors (6–9). Genetic factors may be a direct cause of CKD, but they may also confer vulnerability to other risk factors such as diabetes, hypertension and old age. The glomerular filtration rate (GFR) is considered the best measure of kidney function. Direct measurements of GFR using exogenous or endogenous filtration markers are too cumbersome and expensive for use in routine clinical practice.

*To whom correspondence should be addressed at: deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland. Tel: +354 570 1900; Fax: +354 570 1903; Email: daniel.gudbjartsson@decode.is (D.F.G.); deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland. Tel: +354 570 1900; Fax: +354 570 1903; Email: kari.stefansson@decode.is (K.S.)

In recent years, estimates of GFR (eGFR) obtained using equations based on serum creatinine (SCr) have become the most commonly used indicator of kidney function. However, the relationship between the SCr level and GFR varies between people as creatinine is mostly generated in striated muscle and is therefore affected by muscle mass. Furthermore, creatinine is partly excreted by tubular secretion in the kidney, in addition to glomerular filtration. Current SCr-based equations to calculate eGFR account for age, sex and ethnicity as surrogates for muscle mass (10). CKD is generally defined as a GFR <60 ml/min/1.73 m² and/or other signs of kidney damage, lasting for 3 months or more (11).

Previous reports of genome-wide association studies (GWAS) of SCr, eGFR, based on SCr (eGFR_{crea}) and cystatin C (eGFR_{cys}), and/or CKD have listed several sequence variants that associate with these phenotypes (12–17). The variants associating with SCr levels affect kidney function or creatinine metabolism (12). These common variants (minor allele frequency, MAF >5%) explain <1% of the variability of SCr/eGFR, whereas the heritability of SCr and eGFR_{crea} has been estimated as 29 and 33%, respectively, implying the existence of undiscovered genetic factors (18).

RESULTS

To search for sequence variants associating with SCr, we performed a GWAS of 24 million single-nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs; Supplementary Material, Table S1) on a sample of 194 286 Icelanders with SCr measurements. The participants had a median of 4 (range 1–645) SCr measurements and an average (geometric) of 4.9. The SCr measurements were obtained from the three largest medical laboratories in Iceland (Supplementary Material, Table S2). The mean (arithmetic) of the quantile–quantile standardized SCr values, adjusted for sex and age at the time of measurement for each individual, was used in the analysis. The SNPs and INDELs were found by whole-genome sequencing of 2230 Icelanders. The variants were imputed into 81 656 chip-genotyped individuals using long-range phasing-based imputation (19). In addition, the genotypes of 112 630 relatives of chip-typed individuals with SCr measurements were inferred and used in the analysis.

Sequence variants predicted to cause loss of function (LoF) of a gene, such as stop-gain mutations and frameshift INDELs, have the largest impact of any known class of sequence variants.

This class is followed by the class of predicted missense variants. We weighed sequence variants according to their prior probability of affecting gene function by applying thresholds for genome-wide significance that depend on the variant class. We allocated the type I error rate of 0.05 equally between three classes of variants. We tested 6476 LoF variants, 100 502 missense variants and 23 854 999 other variants (Supplementary Material, Table S1), yielding class-specific Bonferroni genome-wide significance thresholds of 2.6×10^{-6} , 1.7×10^{-7} and 7.0×10^{-10} , respectively. Sequence variants at 19 loci associate with SCr (Supplementary Material, Figs S1–S3 and Table S3). SNPs at 15 of the 19 loci have already been reported to associate with kidney function (12–17), but 4 are novel. One locus has a genome-wide significant signal that is independent of previously reported variants at the locus (Table 1). All of these five newly discovered variants are coding variants and four of them are rare (MAF <2%). These variants have substantially larger effects than the previously reported common variants (Fig. 1A), and the effects are inversely proportional to MAF (Fig. 1B). The genes affected by the novel variants encode either solute carriers (*SLC25A45*, *SLC47A1* and *SLC6A19*) or E3 ubiquitin ligases (*RNF128* and *RNF186*).

We tested the variants for association with CKD in a sample of 15 594 cases and 291 428 controls from Iceland (Table 1). GFR was estimated from the SCr measurements using the MDRD Study equation, and individuals who persistently had eGFR_{crea} <60 ml/min/1.73 m² for more than 3 months were classified as having CKD (11). Three of the five novel variants associating with SCr also associate with CKD risk ($P < 0.05/5 = 0.01$). We additionally defined a severe (stage IV and V) CKD phenotype (11) as eGFR_{crea} <30 ml/min/1.73 m² and tested the five variants for association in a sample of 1716 cases (Supplementary Material, Table S8). While no significant association was detected, three variants showed a directionally consistent effect.

SLC47A1 harbors a missense mutation, p.A465V (rs111653425, MAF = 1.8%), that associates with SCr (effect = 0.105SD, $P = 9.5 \times 10^{-14}$) and with risk of CKD (OR = 1.24, $P = 4.1 \times 10^{-4}$). *SLC47A1* encodes an organic cation/H⁺ electroneutral exchanger involved in the final steps of organic cation extrusion in the kidney and liver (20,21), with creatinine being one of its substrates (22). Knockout of *slc47a1* in mice results in significantly increased creatinine and blood urea nitrogen (BUN) levels (23). p.A465V affects a conserved amino acid located within the long cytoplasmic loop between the 12th and 13th transmembrane domains (24), a known location for mutations that alter *SLC47A1* transporting properties (25). We found six other

Table 1. Sequence variants showing a genome-wide significant association with SCr which are independent of reported markers. SCr measurements were available for 81 656 chip-typed individuals and 112 630 relatives of chip-typed individuals. Their association with CKD in a set of 15 594 cases and 291 420 controls is also shown. For each sequence variant, the reference SNP ID number (rs#), chromosome (Chr), hg18 position, its effect on a gene (coding effect), minor allele frequency (MAF) and imputation information (Info) are provided in addition to the effect on SCr in SDs, the odds ratio (OR) for CKD and the corresponding *P*-values

rs#	Chr	Position (hg18)	Gene	Coding effect	MAF (%)	Info	SCr Effect	<i>P</i>	CKD OR	<i>P</i>
rs111653425	17	19 415 467	<i>SLC47A1</i>	p.A465V	1.775	0.97	0.105	9.5×10^{-14}	1.24	0.00041
rs56121637	X	105 902 924	<i>RNF128</i>	p.N204Y	1.988	0.99	0.085	4.4×10^{-13}	1.10	0.080
rs36095412	1	20 013 647	<i>RNF186</i>	p.R179X	0.776	0.97	0.129	8.6×10^{-10}	1.33	0.0016
rs142979576	5	1 270 062	<i>SLC6A19</i>	c.1173+2T>G	0.856	0.99	-0.115	1.3×10^{-8}	0.75	0.0019
rs34400381	11	64 900 468	<i>SLC25A45</i>	p.R285C/p.R243C	5.497	0.99	-0.048	3.3×10^{-9}	0.95	0.18

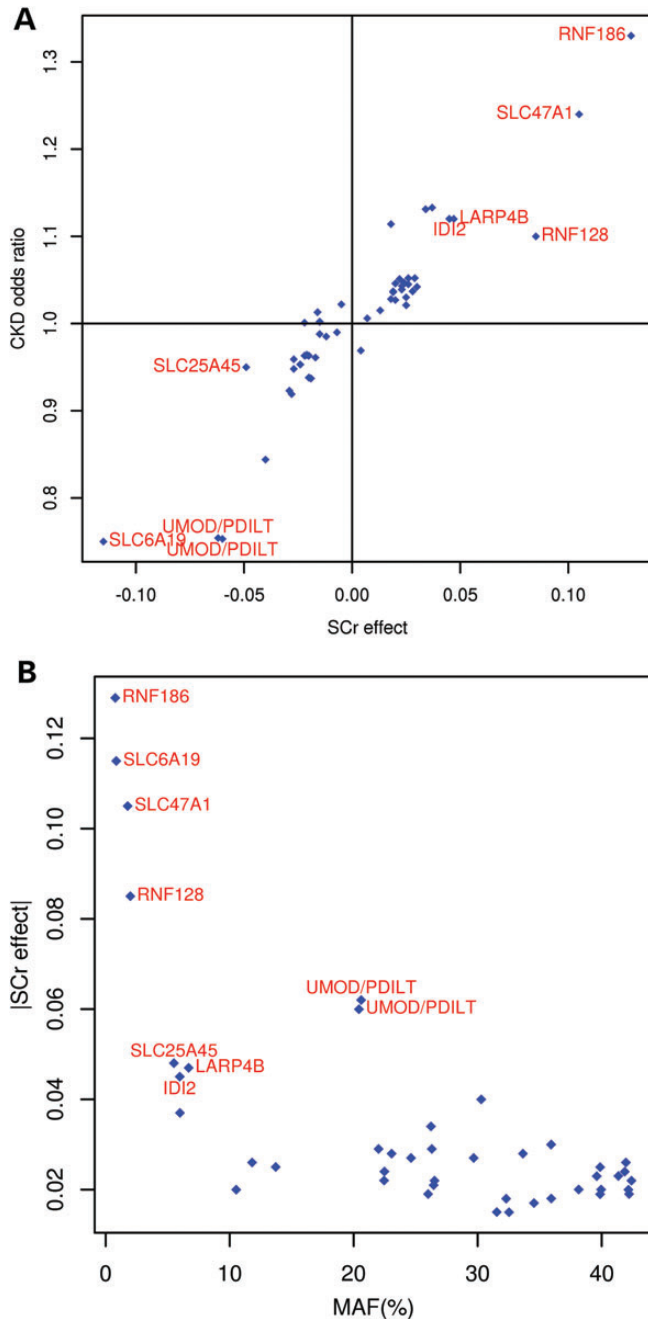


Figure 1. Scatter plot showing SCr effects versus CKD odds ratios (A) and MAF of variants versus the absolute values of SCr effects (B). SCr effects are shown in SDs. Exact values for all variants can be found in Supplementary Material, Table S6.

mutations in *SLC47A1* (Supplementary Material, Table S5), including p.Pro489SerfsX10 (MAF = 0.20%) that associates with SCr (effect = 0.144SD, $P = 8.0 \times 10^{-4}$). Previous studies reported an association of the intronic SNP rs2453580 with eGFR-crea (14). This variant is independent of our discovery SNP and p.Pro489SerfsX10 ($r^2 = 0.020$ and 0.0015 , respectively).

A splice variant, c.1173+2T->G (rs142979576, MAF = 0.86%), in *SLC6A19* associates with lower SCr (effect = -0.115 SD, $P = 1.3 \times 10^{-8}$). Mutations in *SLC6A19*, including

c.1173+2T->G, have been linked to Hartnup disease (26,27), an autosomal recessive disease characterized by impaired reabsorption of neutral amino acids (28). We detected seven additional mutations in *SLC6A19*, three of which associate with lower SCr (Supplementary Material, Table S4 and Table 2). The missense mutations p.D173N (MAF = 0.32%) and p.P579L (MAF = 0.36%) have also been linked to Hartnup disease (27,29), whereas p.F360S has not been observed outside Iceland and is not known to associate with the disease. Among the 4298 individuals with available data in the Exome Sequencing Project (ESP) (30), the splice variant, c.1173+2T->G, is only seen in two individuals (ESP MAF = 0.023%) and the missense mutation P.579L is not observed at all. The missense mutation, p.D173M (ESP MAF = 0.34%), is by far the most common disease causing mutation in Europeans (31) and has a similar frequency in Iceland, although there the splice variant, c.1173+2T->G, is most common. These four mutations are pairwise independent ($r^2 < 0.00010$)—no sequenced individual carries more than one mutation on the same chromosome—and all associate with lower SCr and have similar effect sizes (Table 2). The splice variant associates with a decreased risk of CKD (OR = 0.75, $P = 0.0019$) and the three other mutations show trends in the same direction (Table 2). Collectively, the three additional mutations have a frequency of 1.0% and associate with a decreased risk of CKD (OR = 0.82, $P = 0.011$).

A missense mutation that affects isoforms a and b of *SLC25A45* (rs34400381, MAF = 5.5%), named p.R285C and p.R243C, respectively, associates with lower SCr levels (effect = -0.048 SD, $P = 3.3 \times 10^{-9}$). *SLC25A45* belongs to a family of mitochondrial transporters that translocate several solutes across the mitochondrial membrane and participate in essential metabolic pathways (32). Analysis of *SLC25A45* expression in human tissues revealed high expression in small intestine (Supplementary Material, Fig. S4A), consistent with results of expression studies in the rat (33). Another SNP, rs4014195, located near *RNASEH2C* and 200 kb upstream of rs34400381 ($r^2 = 0.015$), has been reported to associate with kidney function (12) (reported $P = 1 \times 10^{-7}$), but did not replicate in their follow-up (reported $P = 0.14$). In our samples, rs4014195 associates with SCr and CKD ($P = 4.1 \times 10^{-5}$ and 6.7×10^{-10} , respectively; Supplementary Material, Table S6).

A stop-gain mutation, p.R179X (rs36095412, MAF = 0.78%), in *RNF186* associates with higher SCr (effect = 0.129SD, $P = 8.6 \times 10^{-10}$). *RNF186* has recently been shown to possess an E3 ubiquitin ligase activity and localize to the endoplasmic reticulum (34). Expression analysis revealed *RNF186* expression in the small intestine, kidney and colon (Supplementary Material, Fig. S4B). Furthermore, p.R179X associates with an increased risk of CKD (OR = 1.33, $P = 1.6 \times 10^{-3}$).

A missense mutation, p.N204Y (rs56121637, MAF = 2.0%), in *RNF128* associates with higher SCr (effect = 0.085SD, $P = 4.4 \times 10^{-13}$). *RNF128* encodes an E3 ubiquitin ligase that localizes to the endocytic pathway and promotes polyubiquitination and proteasomal degradation of its substrates (35,36). A high expression of *RNF128* was observed in the small intestine as well as in the liver, kidney and muscle (Supplementary Material, Fig. S4C).

Table 2. Coding variants in *SLC6A19* and *SLC47A1* associating with SCr and CKD. SCr measurements were available for 81 656 chip-typed individuals and 112 630 relatives of chip-typed individuals. Their association with CKD in a set of 15 594 cases and 291 420 controls is also given. For each sequence variant, the reference SNP ID number (rs#), its coding effect on a gene, minor allele frequency (MAF) and imputation information (info) are given in addition to the effect on SCr in SDs, the odds ratio (OR) for CKD and the corresponding *P*-values

Gene	Coding effect	rs#	MAF (%)	Info	SCr Effect	<i>P</i> -values	CKD OR	<i>P</i> -values
<i>SLC6A19</i>	c.1173+2T>G ^a	rs142979576	0.880	0.99	-0.115	1.2×10^{-8}	0.75	0.0019
	p.F360S	-	0.321	0.95	-0.101	0.0017	0.77	0.084
	p.D173N ^a	rs121434346	0.317	0.99	-0.098	0.0022	0.86	0.34
	p.P579L ^a	-	0.363	0.99	-0.091	0.0038	0.85	0.26
	p.F360S, p.D173N or p.P579L	-	1.001	-	-0.097	1.4×10^{-7}	0.82	0.011
<i>SLC47A1</i>	p.A465V	rs111653425	1.748	0.97	0.105	9.5×10^{-14}	1.24	4.1×10^{-4}
	p.Pro489SerfsX10	-	0.188	0.96	0.144	8.0×10^{-4}	1.25	0.23

^aThese mutations have been associated with Hartnup disease.

REPLICATION OF REPORTED VARIANTS

We evaluated 45 SNPs that have been associated with SCr and/or CKD (Supplementary Material, Table S6). The association of 41 SNPs was confirmed ($P < 0.05$). The association of the variant rs16864170 at *SOX11* with CKD did not replicate ($P = 0.55$); this association also failed to replicate in a previous attempt (12). The association of rs3828890, near *MHC* with SCr reported in an East Asian population (13), did not replicate in our data ($P = 0.49$). Finally, rs6431731, near the *DDX1* gene (14) and rs6465825, near *TMEM60* (12) did not associate with SCr ($P = 0.41$ and 0.072 , respectively).

CONDITIONAL ANALYSIS

We performed conditional analysis to search for secondary signals at the 19 SCr loci with a genome-wide significant association with SCr in our data. Conditioning on the most significant sequence variant at each locus, we searched for signals within a 200 kb window. On average, 0.9 LoF variants, 11.3 missense variants and 1682.5 other variants were tested at each locus, giving class-specific Bonferroni significance thresholds of 1.9×10^{-2} , 1.5×10^{-3} and 9.9×10^{-6} , respectively. In addition to the mutations in *SLC6A19* and *SLC47A1* described above, we found a second significant sequence variant in three regions (in or near *SLC34A1*, *GATM* and *UMOD/PDILT*) within the 200 kb window around the most significant SNP (Table 3).

In *SLC34A1*, a rare (MAF = 0.46%) missense mutation, p.Y489C, within 20 kb of the SNP showing the most significant association in the region (rs35716097), associates with SCr when conditioning on rs35716097. We replicated the known association of rs35716097 in *SLC34A1* with kidney stone disease (12) and demonstrated the association of the p.Y489C mutation with the disorder (Supplementary Material, Table S7).

For the known loci, where our most significantly associated SNP differed from the reported SNP, we performed a conditional analysis in order to determine if our most significant SNP represents a refinement of the published signal. Several correlated sequence variants were more significantly associated with SCr than the previously reported rs10794720 in *WDR37* on chromosome 10p15 (reported $P = 1 \times 10^{-8}$, MAF = 5.99%, our $P = 1.9 \times 10^{-6}$, effect = 0.037SD) (12). The most significant

association is with rs35772020, located in *LARP4B* (MAF = 6.7%, effect = 0.047SD, $P = 4.1 \times 10^{-10}$), which is correlated with rs10794720 ($r^2 = 0.40$). A correlated ($r^2 = 0.78$) stop-gain mutation, rs1044261, in the *ID12* gene also showed a highly significant association with SCr (MAF = 5.97%, effect = 0.45SD, $P = 7.1 \times 10^{-9}$). Conditional on the effect of the reported SNP, rs10794720 in *WDR37*, the association of these SNPs with SCr remained significant ($P = 9.2 \times 10^{-7}$ and 4.3×10^{-8}).

The most significantly associated sequence variant on chromosome 16, rs77924615, was in *PDILT* on 16p11, located adjacent to the *UMOD* gene on chromosome 16p12, which has been associated with CKD (15,16). After conditioning on the previously reported rs4293393, rs77924615 ($r^2 = 0.38$) remained significant ($P_{\text{adj}} = 2.4 \times 10^{-14}$). When conditioning on rs77924615, the most significant association within the region was rs60136849, a close correlate of rs4293393 ($r^2 = 0.98$). Hence, not all of the association with SCr at the *UMOD*-*PDILT* locus is captured by the strongest variant in this region.

DISCUSSION

Through whole-genome sequencing and imputation into a large number of individuals with several SCr measurements each, we discovered novel sequence variants associating with SCr and CKD. Because of the large sample of sequenced individuals used to inform imputation of sequence variants, we were able to discover rare variants with relatively large effect on SCr. Taken together, these novel variants add 0.5% to the explained variability of SCr in the Icelandic population.

While the observed association with SCr is convincing, we cannot distinguish between effects of the variants on kidney function and non-GFR determinants of SCr. Accordingly, the implications for CKD are uncertain due to the limitations associated with using SCr and creatinine-based GFR estimates for defining the disorder. Hence, the associations with CKD might not reflect true alterations in GFR. Our study is therefore limited by the lack of other filtration markers such as cystatin C, which would enable us to differentiate between effects on creatinine metabolism and kidney function. Since eGFR data using another filtration marker were not available, a severe (stage IV and V) CKD phenotype was defined in order to eliminate the effects of sequence variants causing altered creatinine metabolism. Limiting the phenotype reduces power to detect

Table 3. Secondary and refinement signals identified through conditional analysis of the reported loci. SCr measurements were available for 81 656 chip-typed individuals and 112 630 relatives of chip-typed individuals. For the secondary and refinement variants, the effect on SCr in standard units conditional on the primary SNP is given, along with the corresponding P -value and correlation measured by r^2 . Their association with CKD in a set of 15 594 cases and 291 420 controls is also given. For each sequence variant, the reference SNP ID number (rs#), chromosome (Chr), hg18 position, its effect on a gene, minor allele frequency (MAF) and imputation information (info) are given in addition to the effect on SCr in SDs, the odds ratio (OR) for CKD and the corresponding P -values

Description	rs#	Chr	Position (hg18)	Info	Gene	Coding effect	MAF (%)	SCr Effect	P -values	CKD OR	P -values	Conditional SCr Effect	P -values	r^2
<i>SLC34A1</i> region														
Primary	rs35716097	5	176 739 242	0.99	Near <i>SLC34A1</i>	—	33.61	0.030	4.0×10^{-14}	1.07	2.0×10^{-4}	—	—	—
Secondary	—	5	176 757 439	0.99	<i>SLC34A1</i>	p.Y489C	0.46	0.116	3.2×10^{-5}	1.091	0.49	0.13	5.8×10^{-6}	0.0022
<i>GATM</i> region														
Primary	rs1288775	15	43 448 970	0.98	<i>GATM</i>	—	25.43	0.034	6.6×10^{-15}	1.05	0.0098	—	—	—
Secondary	rs4580098	15	43 384 201	0.98	Near <i>SLC28A2</i>	—	22.94	0.023	2.9×10^{-7}	1.03	0.16	-0.061	1.1×10^{-7}	0.85
<i>UMOD/PDILT</i> region														
Primary	rs77924615	16	20 299 833	0.95	<i>PDILT</i>	—	19.82	-0.068	5.9×10^{-45}	0.73	2.5×10^{-44}	—	—	—
Secondary	rs60136849	16	20 261 316	0.95	<i>UMOD</i>	—	20.51	-0.063	9.7×10^{-40}	0.75	2.5×10^{-38}	-0.035	4.4×10^{-9}	0.37
Refinement of the 10p15 locus														
Primary	rs10794720	10	1 146 165	0.99	<i>WDR37</i>	—	5.99	0.037	1.9×10^{-6}	1.13	3.5×10^{-4}	—	—	—
Refinement	rs35772020	10	853 482	0.99	<i>LARP4B</i>	—	6.69	0.047	4.1×10^{-10}	1.12	8.7×10^{-4}	0.043	4.3×10^{-8}	0.40
Refinement	rs1044261	10	1 055 710	0.99	<i>ID12</i>	p.W144X	5.97	0.045	7.1×10^{-9}	1.12	0.0019	0.041	9.2×10^{-7}	0.54

significant associations, but a directionally consistent trend would be suggestive of an effect on CKD. Variants in *RNF186*, *SLC6A19* and *SLC25A45* showed a directionally consistent effect. Sequence variants that only affect creatinine metabolism, but do not predispose to CKD, may improve the accuracy of equations used for estimating GFR.

All the variants alter a protein-coding sequence and affect the function of either solute carriers or E3 ubiquitin ligases. Four of the variants we describe are rare with SCr effects between 0.085 and 0.129SD which are greater than the maximum effect of the previously reported common variants, which all have effects below 0.069SD.

The SLC superfamily encodes membrane-bound transporters that translocate diverse substances across the cell membrane (37). Variants in or near SLC genes have been associated with creatinine production and/or secretion (12) as well as with renal function and CKD (12,17). We found several pairwise independent coding variants associating with decreased SCr in *SLC6A19*. *SLC6A19* is responsible for reabsorption of neutral amino acids in the kidneys and small intestine (28). Homozygotes and compound heterozygotes for mutations in *SLC6A19* have Hartnup disease (26,27), an autosomal recessive condition characterized by pellagra-like light-sensitive rash, cerebellar ataxia, emotional instability and aminoaciduria (28). Although SCr is affected primarily by the GFR, there are additional factors that can affect SCr levels without a change in GFR, such as altered creatinine generation or renal tubular secretion. While these non-GFR determinants of SCr may account for the association of *SLC6A19* and SCr/CKD, this finding still raises the intriguing possibility that the *SLC6A19* mutations confer a heterozygote advantage (38). Thus, it would be important to use other filtration markers or a direct measurement of GFR to establish that the variants do indeed confer protection against CKD.

A common variant in an intron of *SLC47A1* has been associated with kidney function (14). We found a rare missense mutation that associates with both increased SCr and an increased risk of CKD. As *SLC47A1* encodes a multidrug and toxin extruder, one can speculate that the inability to excrete toxic organic cations may eventually lead to kidney damage and CKD. This notion is supported by a knockout mouse model, in which a concomitant rise in BUN suggests a reduction in GFR (22). However, creatinine may also be one of the many substrates of this transporter and, therefore, a missense mutation might cause impairment of renal tubular secretion, resulting in higher SCr.

Finally, we found a missense mutation in *SLC25A45*, an orphan mitochondrial transporter that associates significantly with renal quantitative traits but not with CKD. Based on our data and previous reports, we postulate that *SLC25A45*, which is related to the ornithine/citrulline transporters *SLC25A2* and *SLC25A15* (32), may play a role in biosynthesis of arginine, one of the three amino acids involved in the synthesis of creatine, the precursor of creatinine.

E3 ubiquitin ligases play an important role in regulation of diverse cellular processes (39–41). *RNF186*, which encodes an E3 ubiquitin ligase localizing to the ER and promotes Lys29- and Lys63-linked polyubiquitination (34), harbors a rare stop-gain mutation associating with both elevated SCr and a risk of CKD. *RNF128* is an E3 ubiquitin ligase in the endocytic

pathway that reportedly plays a role in glucose and lipid metabolism in the mouse and in T-cell anergy (35,42,43). A missense mutation in *RNF128* associates with elevated SCr but not with CKD. The expression profile of *RNF128* indicates that it has a role in the small intestine, liver, kidney and skeletal muscle, which are sites relevant for creatinine production and/or excretion. Additional studies are required to search for the substrates of RNF186 and RNF128 and to determine whether these ligases are involved in molecular pathways relevant to kidney function and pathogenesis of kidney disease.

In conclusion, we have discovered several rare protein-coding variants that associate with SCr and CKD. These variants add 0.5% to the explained variability of SCr in the Icelandic population. Future studies will require more accurate characterization of CKD, using additional filtration markers for estimation of GFR and alternative markers of kidney disease, particularly proteinuria.

MATERIALS AND METHODS

Study population

Landspítali – The National University Hospital of Iceland (LUH) is a regional hospital for the greater Reykjavik area and a tertiary referral center for the entire Icelandic nation. The nation's only nephrology clinic is located at LUH, and all laboratory tests for the primary care clinics of the greater Reykjavik area are performed in the hospital's laboratories. We obtained results of all serum creatinine, uric acid, potassium, calcium and phosphate measurements performed at LUH during the period 1993–2012 (Supplementary Material, Table S7). We also obtained all measurements of these traits from The Icelandic Medical Center (Laeknasetríð) Laboratory in Mjódd (RAM), a private outpatient clinic, during the period from 2004 to 2012 and from the Department of Clinical Biochemistry, Akureyri Hospital, the regional hospital of the north of Iceland, during the period from 2004 to 2010 (Supplementary Material, Table S2). We did not observe a difference between laboratories in the distribution of age or sex, but SCr measurements from LUH, the only tertiary care hospital among the three, were on average higher than those from the other two institutions. All quantitative measurements were quantile–quantile standardized to a standard normal distribution and then adjusted for sex and the age at measurement. For individuals with multiple measurements, the mean adjusted measurement was used. Only SCr measurements of individuals that had reached the age of 18 years were used in the analysis.

We estimated the GFR using the MDRD Study equation: $eGFR \text{ (ml/min/1.73 m}^2\text{)} = 175 \times (S_{cr})^{-1.154} \times (\text{Age})^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if black})$. CKD was defined as $eGFR < 60 \text{ ml/min/1.73 m}^2$ as for at least 3 months. When defining CKD, all SCr measurements obtained during an episode of acute kidney injury were excluded and individuals who had $eGFR < 60 \text{ ml/min/1.73 m}^2$ for <3 months duration were excluded from the CKD sample set. A total of 15 594 CKD cases were identified using this method. Other chip-typed individuals who had not been diagnosed with CKD, and first- and second-degree relatives of chip-typed individuals were used as controls.

The 6736 Icelandic kidney stone disease cases consisted of patients with confirmed radiopaque kidney stones from the Icelandic Kidney Stone Disease Registry at LUH.

The study was approved by the Icelandic Data Protection Authority and the National Bioethics Committee. All participating subjects that donated blood also signed informed consent. Personal identities of the patients and biological samples were encrypted by a third party system provided by the Icelandic Data Protection Authority.

The full GWAS results provide specific information on the genetic characteristics of the whole Icelandic nation, precluding full disclosure of the data. Association results for all variants with P -value of $< 10^{-4}$, MAF > 0.01 and imputation information score higher than 0.95 are shown in Supplementary Material, Table S9.

Illumina chip genotyping

Icelandic chip-typed samples were assayed using the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. SNPs were excluded if they had (i) yield $< 95\%$, (ii) MAF $< 1\%$ in the population or (iii) significant deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.001$), (iv) if they produced an excessive inheritance error rate (over 0.001) or (v) if there was substantial difference in allele frequency between chip types (from just a single chip if that resolved all differences, but from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. For the HumanHap series of chips, 304 937 SNPs were used for long-range phasing, whereas for the Omni series of chips 564 196 SNPs were included. The final set of SNPs used for long-range phasing was composed of 707 525 SNPs.

Whole-genome sequencing

Whole-genome sequencing was performed for 2230 Icelanders, selected for various conditions. All the individuals were sequenced at a depth of at least $10\times$ (average sequencing depth = $22\times$).

Template DNA fragments were hybridized to the surface of flow cells [GA PE cluster kit (v2) or HiSeq PE cluster kits (v2.5 or v3)] and amplified to form clusters using the Illumina cBot. In brief, DNA (2.512 pM) was denatured, followed by hybridization to grafted adaptors on the flow cell. Isothermal bridge amplification using Phusion polymerase was then followed by linearization of the bridged DNA, denaturation, blocking of 3' ends, and hybridization of the sequencing primer. Sequencing-by-synthesis (SBS) was performed on Illumina GAIIX and/or HiSeq 2000 instruments. Paired-end libraries were sequenced at 2×101 (HiSeq) or 2×120 (GAIIX) cycles of incorporation and imaging using the appropriate TruSeq™ SBS kits. Each library or sample was initially run on a single GAIIX lane for QC validation followed by further sequencing on either GAIIX (3 lanes) or HiSeq (3 lanes) with targeted raw cluster densities of 5 00 800K/mm², depending on the version of the data imaging and analysis packages (SCS2.6-2-9/RTA1.6-1.9, HCS1.3.8-1.4.8/RTA1.10.36-1.12.4.2). Real-time analysis involved the conversion of image data to base-calling in real time.

Generation of whole-genome genotype data

SNP and INDEL calling from the whole-genome sequence data of the 2230 Icelanders and generation of imputed genotypes have been described previously (44). Briefly, the genotypes identified were imputed into 81 656 chip-genotyped and long-range-phased Icelanders with SCr measurements. Probabilities of genotypes were furthermore predicted for 112 630 relatives of chip-typed individuals with SCr measurements.

Association testing

A generalized form of linear regression that accounts for the relatedness between individuals was used to test for the association of quantitative traits, such as SCr and serum uric acid, with sequence variants (44) (Supplementary Material, Methods). Conditional analysis was performed by including the sequence variant being conditioned on as a covariate in the model under the null and the alternative in the generalized linear regression.

Logistic regression was used to test for association between sequence variants and disease (CKD and kidney stone disease), treating disease status as the response and genotype counts as covariates. Other available individual characteristics that correlate with disease status were also included in the model as nuisance variables. These characteristics were: sex, county of birth, current age or age at death (first- and second-order terms included), blood sample availability for the individual and an indicator function for the overlap of the lifetime of the individual with the timespan of phenotype collection (Supplementary Material, Methods).

The genomic inflation factor for SCr was $\lambda_{GC} = 1.91$. This large inflation factor is certainly, in part, due to the large sample size and the polygenic nature of the genetic architecture that affects SCr (45). The correction may therefore be overly conservative.

Genotype imputation information

The informativeness of genotype imputation was estimated by the ratio of the variance of imputed expected allele counts and the variance of the actual allele counts:

$$\frac{\text{Var}[E(\theta|\text{chip data})]}{\text{Var}(\theta)},$$

where $\theta \in \{0, 1\}$ is the allele count. $\text{Var}[E(\theta|\text{chip data})]$ was estimated by the observed variance of the imputed expected counts and $\text{Var}(\theta)$ was estimated by $p(1 - p)$, where p is the allele frequency. Sequence variants with imputation information below 0.8 were excluded from the analysis.

Gene and variant annotation

For the annotation of the exome data, coordinates of variants were converted between hg18 and hg19 using the liftOver tool from UCSC (46). Variants in hg19 coordinates were annotated with information from Ensembl release 70 using the Variant Effect Predictor (VEP) version 2.8 (47). Only protein-coding transcripts from RefSeq Release 56 (48) were considered.

Correction for relatedness of the Icelandic subjects and genomic control

Individuals in both the Icelandic case and control groups are related, causing the χ^2 test statistic to have a mean of >1 and median of >0.675 . We estimated the inflation factor λ_g based on a subset of about 300 000 common variants and the P -values adjusted by dividing the corresponding χ^2 values by this factor to adjust for both relatedness and potential population stratification (49).

Thresholds for genome-wide significance

We weighed sequence variants according to their prior probability of affecting gene function by applying thresholds for genome-wide significance that depend on the variant class. The Bonferroni correction for multiple testing can be adjusted to account for prior importance of sequence variants. We performed a weighted Holm–Bonferroni correction based on giving equal weight to the classes of LoF, MSSNS and other variants (50) (Supplementary Material, Table S1). For example, sequence variants in the LoF class get weight $\frac{1}{3 \times 6476}$. The sum of these weights over all the variants in the genome is 1, and the Bonferroni threshold for significance within a class containing m sequence variants will be $\frac{0.05}{3m}$.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors thank all the individuals who participated in these studies and whose contribution made this work possible. We also thank our valued colleagues who contributed to the data collection and phenotypic characterization of clinical samples, as well as genotyping and analysis of the genome-wide association data.

Conflict of Interest statement. The authors who are affiliated with deCODE are all employees of deCODE genetics/AMGEN.

FUNDING

Funding was provided by deCODE Genetics/AMGEN.

REFERENCES

- Eckardt, K.U., Coresh, J., Devuyst, O., Johnson, R.J., Kottgen, A., Levey, A.S. and Levin, A. (2013) Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet*, **382**, 158–169.
- Viktorsdottir, O., Palsson, R., Andresdottir, M.B., Aspelund, T., Gudnason, V. and Indridason, O.S. (2005) Prevalence of chronic kidney disease based on estimated glomerular filtration rate and proteinuria in Icelandic adults. *Nephrol. Dial. Transplant.*, **20**, 1799–1807.
- Coresh, J., Selvin, E., Stevens, L.A., Manzi, J., Kusek, J.W., Eggers, P., Van Lente, F. and Levey, A.S. (2007) Prevalence of chronic kidney disease in the United States. *JAMA*, **298**, 2038–2047.

4. Lindeman, R.D., Tobin, J. and Shock, N.W. (1985) Longitudinal studies on the rate of decline in renal function with age. *J. Am. Geriatr. Soc.*, **33**, 278–285.
5. Smith, D.H., Gullion, C.M., Nichols, G., Keith, D.S. and Brown, J.B. (2004) Cost of medical care for chronic kidney disease and comorbidity among enrollees in a large HMO population. *J. Am. Soc. Nephrol.*, **15**, 1300–1306.
6. Fliser, D., Franek, E., Joest, M., Block, S., Mutschler, E. and Ritz, E. (1997) Renal function in the elderly: impact of hypertension and cardiac function. *Kidney Int.*, **51**, 1196–1204.
7. Fliser, D. and Ritz, E. (1998) Relationship between hypertension and renal function and its therapeutic implications in the elderly. *Gerontology*, **44**, 123–131.
8. Lindeman, R.D., Tobin, J.D. and Shock, N.W. (1984) Association between blood-pressure and the rate of decline in renal-function with age. *Kidney Int.*, **26**, 861–868.
9. Ribstein, J., Du Cailar, G. and Mimran, A. (2001) Glucose tolerance and age-associated decline in renal function of hypertensive patients. *J. Hypertens.*, **19**, 2257–2264.
10. Levey, A.S., Coresh, J., Greene, T., Marsh, J., Stevens, L.A., Kusek, J.W. and Van Lente, F. (2007) Expressing the modification of diet in renal disease study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clin. Chem.*, **53**, 766–772.
11. Levey, A.S., de Jong, P.E., Coresh, J., El Nahas, M., Astor, B.C., Matsushita, K., Gansevoort, R.T., Kasiske, B.L. and Eckardt, K.U. (2011) The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int.*, **80**, 17–28.
12. Kottgen, A., Pattaro, C., Boger, C.A., Fuchsberger, C., Olden, M., Glazer, N.L., Parsa, A., Gao, X., Yang, Q., Smith, A.V. *et al.* (2010) New loci associated with kidney function and chronic kidney disease. *Nat. Genet.*, **42**, 376–384.
13. Okada, Y., Sim, X., Go, M.J., Wu, J.Y., Gu, D., Takeuchi, F., Takahashi, A., Maeda, S., Tsunoda, T., Chen, P. *et al.* (2012) Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.*, **44**, 904–909.
14. Pattaro, C., Kottgen, A., Teumer, A., Garnaas, M., Boger, C.A., Fuchsberger, C., Olden, M., Chen, M.H., Tin, A., Taliun, D. *et al.* (2012) Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.*, **8**, e1002584.
15. Kottgen, A., Glazer, N.L., Dehghan, A., Hwang, S.J., Katz, R., Li, M., Yang, Q., Gudnason, V., Launer, L.J., Harris, T.B. *et al.* (2009) Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.*, **41**, 712–717.
16. Gudbjartsson, D.F., Holm, H., Indridason, O.S., Thorleifsson, G., Edvardsson, V., Sulem, P., de Vegt, F., d'Ancona, F.C., den Heijer, M., Wetzels, J.F. *et al.* (2010) Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet.*, **6**, e1001039.
17. Chambers, J.C., Zhang, W., Lord, G.M., van der Harst, P., Lawlor, D.A., Sehmi, J.S., Gale, D.P., Wass, M.N., Ahmadi, K.R., Bakker, S.J. *et al.* (2010) Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.*, **42**, 373–375.
18. Fox, C.S., Yang, Q., Cupples, L.A., Guo, C.Y., Larson, M.G., Leip, E.P., Wilson, P.W. and Levy, D. (2004) Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J. Am. Soc. Nephrol.*, **15**, 2457–2461.
19. Kong, A., Masson, G., Frigge, M.L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P.I., Ingason, A., Steinberg, S., Rafnar, T. *et al.* (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.*, **40**, 1068–1075.
20. Otsuka, M., Matsumoto, T., Morimoto, R., Arioka, S., Omote, H. and Moriyama, Y. (2005) A human transporter protein that mediates the final excretion step for toxic organic cations. *Proc. Natl. Acad. Sci. USA*, **102**, 17923–17928.
21. Tsuda, M., Terada, T., Asaka, J., Ueba, M., Katsura, T. and Inui, K. (2007) Oppositely directed H⁺ gradient functions as a driving force of rat H⁺/organic cation antiporter MATE1. *Am. J. Physiol. Renal Physiol.*, **292**, F593–F598.
22. Ito, S., Kusuhara, H., Kuroiwa, Y., Wu, C., Moriyama, Y., Inoue, K., Kondo, T., Yuasa, H., Nakayama, H., Horita, S. *et al.* (2010) Potent and specific inhibition of mMate1-mediated efflux of type I organic cations in the liver and kidney by pyrimethamine. *J. Pharmacol. Exp. Ther.*, **333**, 341–350.
23. Tsuda, M., Terada, T., Mizuno, T., Katsura, T., Shimakura, J. and Inui, K. (2009) Targeted disruption of the multidrug and toxin extrusion 1 (mate1) gene in mice reduces renal secretion of metformin. *Mol. Pharmacol.*, **75**, 1280–1286.
24. Zhang, X., He, X., Baker, J., Tama, F., Chang, G. and Wright, S.H. (2012) Twelve transmembrane helices form the functional core of mammalian MATE1 (multidrug and toxin extruder 1) protein. *J. Biol. Chem.*, **287**, 27971–27982.
25. Chen, Y., Teranishi, K., Li, S., Yee, S.W., Hesselson, S., Stryke, D., Johns, S.J., Ferrin, T.E., Kwok, P. and Giacomini, K.M. (2009) Genetic variants in multidrug and toxic compound extrusion-1, hMATE1, alter transport function. *Pharmacogenomics J.*, **9**, 127–136.
26. Kleta, R., Romeo, E., Ristic, Z., Ohura, T., Stuart, C., Arcos-Burgos, M., Dave, M.H., Wagner, C.A., Camargo, S.R., Inoue, S. *et al.* (2004) Mutations in SLC6A19, encoding B0AT1, cause Hartnup disorder. *Nat. Genet.*, **36**, 999–1002.
27. Seow, H.F., Broer, S., Broer, A., Bailey, C.G., Potter, S.J., Cavanaugh, J.A. and Rasko, J.E.J. (2004) Hartnup disorder is caused by mutations in the gene encoding the neutral amino acid transporter SLC6A19. *Nat. Genet.*, **36**, 1003–1007.
28. Broer, S. (2009) The role of the neutral amino acid transporter B0AT1 (SLC6A19) in Hartnup disorder and protein nutrition. *IUBMB Life*, **61**, 591–599.
29. Camargo, S.M., Singer, D., Makrides, V., Huggel, K., Pos, K.M., Wagner, C.A., Kuba, K., Danilczyk, U., Skovby, F., Kleta, R. *et al.* (2009) Tissue-specific amino acid transporter partners ACE2 and collectrin differentially interact with Hartnup mutations. *Gastroenterology*, **136**, 872–882.
30. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA; <http://eversusgs.washington.edu/EVS/>; October, 2013.
31. Azmanov, D.N., Rodgers, H., Auray-Blais, C., Giguere, R., Bailey, C., Broer, S., Rasko, J.E. and Cavanaugh, J.A. (2007) Persistence of the common Hartnup disease D173N allele in populations of European origin. *Ann. Hum. Genet.*, **71**, 755–761.
32. Palmieri, F. (2013) The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol. Asp. Med.*, **34**, 465–484.
33. Haitina, T., Lindblom, J., Renstrom, T. and Fredriksson, R. (2006) Fourteen novel human members of mitochondrial solute carrier family 25 (SLC25) widely expressed in the central nervous system. *Genomics*, **88**, 779–790.
34. Wang, P., Wu, Y., Li, Y., Zheng, J. and Tang, J. (2013) A novel RING finger E3 ligase RNF186 regulate ER stress-mediated apoptosis through interaction with BNip1. *Cell. Signal.*, **25**, 2320–2333.
35. Anandasabapathy, N., Ford, G.S., Bloom, D., Holness, C., Paragas, V., Seroogy, C., Skrenta, H., Hollenhorst, M., Fathman, C.G. and Soares, L. (2003) GRAIL: an E3 ubiquitin ligase that inhibits cytokine gene transcription is expressed in anergic CD4⁺ T cells. *Immunity*, **18**, 535–547.
36. Lineberry, N., Su, L., Soares, L. and Fathman, C.G. (2008) The single subunit transmembrane E3 ligase gene related to anergy in lymphocytes (GRAIL) captures and then ubiquitinates transmembrane proteins across the cell membrane. *J. Biol. Chem.*, **283**, 28497–28505.
37. Hediger, M.A., Clemenccon, B., Burrier, R.E. and Bruford, E.A. (2013) The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol. Asp. Med.*, **34**, 95–107.
38. Hedrick, P.W. (2012) What is the evidence for heterozygote advantage selection? *Trends Ecol. Evol.*, **27**, 698–704.
39. Komander, D. and Rape, M. (2012) The ubiquitin code. *Annu. Rev. Biochem.*, **81**, 203–229.
40. Chasapis, C.T. and Spyroulias, G.A. (2009) RING finger E3 ubiquitin ligases: structure and drug discovery. *Curr. Pharm. Des.*, **15**, 3716–3731.
41. Tu, Y., Chen, C., Pan, J., Xu, J., Zhou, Z.G. and Wang, C.Y. (2012) The ubiquitin proteasome pathway (UPP) in the regulation of cell cycle control and DNA damage repair and its implication in tumorigenesis. *Int. J. Clin. Exp. Pathol.*, **5**, 726–738.
42. Nakamichi, S., Senga, Y., Inoue, H., Emi, A., Matsuki, Y., Watanabe, E., Hiramatsu, R., Ogawa, W. and Kasuga, M. (2009) Role of the E3 ubiquitin ligase gene related to anergy in lymphocytes in glucose and lipid metabolism in the liver. *J. Mol. Endocrinol.*, **42**, 161–169.
43. Seroogy, C.M., Soares, L., Ranheim, E.A., Su, L., Holness, C., Bloom, D. and Fathman, C.G. (2004) The gene related to anergy in lymphocytes, an E3 ubiquitin ligase, is necessary for anergy induction in CD4T cells. *J. Immunol.*, **173**, 79–85.
44. Styrkarsdottir, U., Thorleifsson, G., Sulem, P., Gudbjartsson, D.F., Sigurdsson, A., Jonasdottir, A., Oddsson, A., Helgason, A., Magnusson,

- O.T., Walters, G.B. *et al.* (2013) Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature*, **497**, 517–520.
45. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M. *et al.* (2011) Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.*, **19**, 807–812.
46. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
47. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.
48. Pruitt, K.D., Tatusova, T., Brown, G.R. and Maglott, D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.
49. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
50. Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.