

OPEN

Fine mapping and identification of serum urate loci in American Indians: The Strong Heart Family Study

Geetha Chittoor^{1,2}, Karin Haack³, Poojitha Balakrishnan⁴, Christopher Bizon⁵, Sandra Laston⁶, Lyle G. Best⁷, Jean W. MacCluer³, Kari E. North⁸, Jason G. Umans⁹, Nora Franceschini⁸, Gauri Prasad^{10,11}, Luis Macias-Kauffer¹², Teresa Villarreal-Molina¹², Dwaipayan Bharadwaj^{11,14}, Samuel Canizales-Quinteros¹³, Ana Navas-Acien⁴, Shelley A. Cole³ & V. S. Voruganti^{1*}

While studies have reported genetic loci affecting serum urate (SU) concentrations, few studies have been conducted in minority populations. Our objective for this study was to identify genetic loci regulating SU in a multigenerational family-based cohort of American Indians, the Strong Heart Family Study (SHFS). We genotyped 162,718 single nucleotide polymorphisms (SNPs) in 2000 SHFS participants using an Illumina MetaboChip array. A genome-wide association analysis of SU was conducted using measured genotype analysis approach accounting for kinships in SOLAR, and meta-analysis in METAL. Our results showed strong association of SU with rs4481233, rs9998811, rs7696092 and rs13145758 (minor allele frequency (MAF) = 25–44%; $P < 3 \times 10^{-14}$) of solute carrier family 2, member 9 (*SLC2A9*) and rs41481455, rs2231142 and rs1481012 (MAF = 29%; $p < 3 \times 10^{-9}$) of ATP-binding cassette protein, subfamily G, member 2 (*ABCG2*). Carriers of G alleles of rs9998811, rs4148155 and rs1481012 and A alleles of rs4481233, rs7696092 and rs13145758 and rs2231142 had lower SU concentrations as compared to non-carriers. Genetic analysis of SU conditional on significant *SLC2A9* and *ABCG2* SNPs revealed new loci, nucleobindin 1 (*NUCB1*) and neuronal PAS domain protein 4 (*NPAS4*) ($p < 6 \times 10^{-6}$). To identify American Indian-specific SNPs, we conducted targeted sequencing of key regions of *SLC2A9*. A total of 233 SNPs were identified of which 89 were strongly associated with SU ($p < 7.1 \times 10^{-10}$) and 117 were American Indian specific. Analysis of key SNPs in cohorts of Mexican-mestizos, European, Indian and East Asian ancestries showed replication of common SNPs, including our lead SNPs. Our results demonstrate the association of SU with uric acid transporters in a minority population of American Indians and potential novel associations of SU with neuronal-related genes which warrant further investigation.

¹Department of Nutrition, and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA. ²Biomedical and Translational Informatics, Geisinger, Danville, PA, USA. ³Population Health Program, Texas Biomedical Research Institute, San Antonio, TX, USA. ⁴Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York, USA. ⁵Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC, USA. ⁶South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA. ⁷Missouri Breaks Industries Research Inc., Eagle Butte, SD, USA. ⁸Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁹Medstar Health Research Institute, Hyattsville, MD, USA. ¹⁰Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, 110 020, India. ¹¹Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi, 110 020, India. ¹²Laboratorio de Enfermedades Cardiovasculares, INMEGEN, Mexico City, 14610, Mexico. ¹³Unidad de Genomica de Poblaciones Aplicada a la Salud Facultad de Quimica, UNAM-Instituto Nacional de Medicina Genomica, Mexico City, Mexico. ¹⁴Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi, 110 067, India.

*email: saroja@unc.edu

Phenotype*	Normouricemia	Hyperuricemia**	P value
Anthropometric			
Body weight (kg)	88.08 (24.1)	97.79 (23.8)	<0.0001
Waist circumference (cm)	103.57 (18.8)	107.59 (18.0)	<0.0001
BMI (kg/m ²)	33.04 (8.2)	34.38 (7.3)	NS
Lipids			
Total cholesterol (mg/dl)	177.46 (34.6)	189.74 (42.3)	<0.0001
HDL cholesterol (mg/dl)	51.56 (14.6)	48.26 (14.1)	<0.0001
LDL cholesterol (mg/dl)	95.92 (28.1)	104.39 (31.8)	<0.0001
Triglycerides (mg/dl)	156.65 (129)	200.44 (254)	<0.0001
ECHO measures			
Heart rate (BPM)	66.99 (10.6)	67.38 (11.5)	NS
Left ventricular mass (g)	152.72 (39.0)	175.27 (41.7)	<0.0001
Relative wall thickness	0.298 (0.04)	0.301 (0.04)	0.030
Stroke volume (ml)	80.81 (14.5)	84.90 (15.1)	<0.0001
Ejection fraction	60.41 (5.7)	58.63 (5.8)	<0.0001
Cardiac output (ml/min)	5376.44 (1153)	5673.078 (1252)	<0.0001
Total peripheral resistance (dyne/cm/sec)	1410.02 (304)	1392.27 (316)	NS
Blood pressure			
Systolic (mmHg)	121.25 (17.2)	126.56 (15.8)	<0.0001
Diastolic (mmHg)	75.47 (11.0)	78.46 (11.6)	<0.0001
Carotid measures			
Left common carotid diastolic diameter (mm)	5.67 (0.7)	6.00 (0.7)	<0.0001
Right common carotid diastolic diameter (mm)	5.78 (0.7)	6.10 (0.7)	<0.0001
Left common carotid systolic diameter (mm)	6.40 (0.7)	6.78 (0.7)	<0.0001
Right common carotid systolic diameter (mm)	6.51 (0.7)	6.87 (0.8)	<0.0001
Left common carotid artery intimal medial thickness (mm)	0.66 (0.2)	0.69 (0.2)	<0.0001
Right common carotid artery intimal medial thickness (mm)	0.66 (0.2)	0.68 (0.2)	0.003

Table 1. Hyperuricemia and cardiovascular disease risk factors in SHFS. *Phenotypes are shown as means (SD); NS – Not significant. **Hyperuricemia was defined as serum urate concentrations > 6 mg/dl in women and > 7 mg/dl in men.

Hyperuricemia or elevated concentration of urate in serum (SU) is a risk factor for gout, hypertension, chronic kidney disease (CKD) and cardiovascular disease (CVD)^{1–4}. Uric acid is the final product of purine metabolism in humans, and urate homeostasis involves balancing its production with secretion and reabsorption in the proximal convoluted tubule of kidneys^{3,4}. The variation in SU concentration is under significant genetic influence and its pattern of inheritance suggests that many genes may influence it¹. Correspondingly, the renal transport of urate involves several genes including solute carrier family 2, member 9 (*SLC2A9*), ATP-binding cassette ABC, sub-family G, member 2 (*ABC2*), solute carrier family 22, members 11 and 12 (*SLC22A11* and *SLC22A12*), solute carrier family 17, members 1, 3 and 4 (*SLC17A1*, *SLC17A3* and *SLC17A4*), and solute carrier family 16, member 9 (*SLC16A9*). Most of these genes have been associated with hyperuricemia^{1–6}.

Both hyper and hypouricemia have been linked to increased risk for metabolic diseases. While hypouricemia has been linked to neurological disorders such as multiple sclerosis and Parkinson's disease^{7,8}, hyperuricemia is causal for gout and nephrolithiasis and seems to increase the risk for CKD and CVD^{1–4}. Originally thought to be just a marker, SU's role in development and progression of these diseases is being increasingly recognized³. While a recent review by Li *et al.*⁹ found no clear role for uric acid in metabolic diseases other than gout and nephrolithiasis, many studies including ours, have shown that gout patients and asymptomatic hyperuricemic individuals tend to be at high risk for CVD and CKD (Table 1)^{10,11}. Therefore, it is essential to understand the genetic and environmental factors that affect the variation in SU. Even though genome-wide association studies (GWAS) have identified many SU-related loci, the majority of these studies have been conducted in European, African American and Asian populations^{12–22}. To better understand genetic variation, biological significance and translation to human health, it is important to study ethnically diverse populations²³. Further, the linkage disequilibrium (LD) pattern differences in ethnically diverse populations may offer a unique perspective on fine mapping of genetic loci.

American Indians are such a population that is understudied and underrepresented in genetic databases. The prevalence of CVD and CKD is high in American Indians with heart disease being the leading cause of death [https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_aian.htm]. The Strong Heart Family Study (SHFS) is a multigenerational family-based study of CVD in American Indians. This cohort has high rates of obesity, diabetes, CKD and CVD^{24–26}. In addition, about 25% of individuals have hyperuricemia (SU > 6 mg/dl)²⁷. Thus, our aim in this study was to identify the genetic loci that regulate SU concentrations in American Indians. The

SNP	Minor Allele	MAF	P-value	β	SE	Effect Size	Gene	Gene loc	Chr	Chr pos hg18
Arizona										
rs13145758	A	0.43	1.1×10^{-10}	-0.35	0.05	0.06	SLC2A9	Intron	4	9981997
rs9998811	G	0.44	7.4×10^{-10}	-0.33	0.05	0.06	SLC2A9	Intron	4	9966477
rs7862063	G	0.41	4.5×10^{-8}	0.29	0.05	0.04	—	—	9	110002036
rs4481233	A	0.41	2.3×10^{-6}	-0.26	0.05	0.03	SLC2A9	Intron	4	9956079
rs6688009	A	0.01	9.1×10^{-6}	1.07	0.24	0.03	PPAP2B	Intron	1	24593576
Dakotas										
rs4481233	A	0.25	6.7×10^{-8}	-0.29	0.05	0.04	SLC2A9	Intron	4	9956079
rs7696092	C	0.19	1.6×10^{-6}	-0.29	0.06	0.03	SLC2A9	Intron	4	10025320
rs179409	G	0.45	3.4×10^{-6}	-0.23	0.05	0.03	KCNQ1	Intron	11	2483882
rs9998811	A	0.41	4.2×10^{-6}	-0.22	0.05	0.03	SLC2A9	Intron	4	9966477
rs13145758	G	0.40	5.1×10^{-6}	-0.22	0.05	0.03	SLC2A9	Intron	4	9981997
rs7947391	A	0.22	5.6×10^{-6}	-0.26	0.06	0.02	NPAS4	Intron	11	66186882
Oklahoma										
rs4481233	A	0.30	3.3×10^{-9}	-0.30	0.05	0.06	SLC2A9	Intron	4	9956079
rs9998811	A	0.46	5.8×10^{-7}	-0.23	0.05	0.04	SLC2A9	Intron	4	9966477
rs4148155	G	0.20	1.1×10^{-7}	0.29	0.06	0.03	ABCG2	Intron	4	89054667
rs2231142	A	0.20	1.1×10^{-6}	0.29	0.06	0.03	ABCG2	Missense	4	89052323
rs1481012	G	0.20	1.1×10^{-6}	0.29	0.06	0.03	ABCG2	Intron	4	89039082
rs7696092	C	0.29	2.1×10^{-6}	-0.25	0.05	0.04	SLC2A9	Intron	4	10025320
rs13145758	G	0.45	2.7×10^{-6}	-0.22	0.05	0.03	SLC2A9	Intron	4	9981997
rs746075	A	0.36	2.8×10^{-6}	0.22	0.05	0.04	NUCB1	Intron	19	49416936

Table 2. Genome-wide association analysis of serum urate stratified by center. *SNP: single nucleotide polymorphism; MAF: minor allele frequency; P-value: P-values from measured genotype analysis; β : beta coefficient of the SNP; SE: standard error; Effect Size: Proportion of the residual phenotypic variance that is explained by the minor allele of the SNP; Gene loc: Gene location; Chr pos: chromosome position in base pairs; SLC2A9: solute carrier family 2, member 9; PPAP2B: phosphatidic acid phosphatase type 2; KCNQ1: voltage gated KQT-like subfamily Q, member 1; ATP13A5: ATPase type 13A5; ABCG2: ATP-binding cassette family G, member 2; NUCB: Nucleobindin.

GWAS was first conducted in each of the three centers of the SHFS (Arizona, Oklahoma and Dakotas (North and South)), followed by a meta-analysis of all three centers. As a secondary aim, we aimed to identify American Indian-specific SNPs in *SLC2A9*, the gene most strongly associated with SU in this study.

Results

The current study included 3000 SHFS participants (1282 men and 1718 women) from three study centers, Arizona, North and South Dakota (Dakotas) and Oklahoma. The mean SU concentrations were 5.14 ± 1.5 mg/dl (4.6 ± 1.3 mg/dl in women and 6.0 ± 1.4 mg/dl in men); 4.9 ± 1.5 mg/dl, 5.2 ± 1.5 mg/dl, 5.3 ± 1.5 mg/dl in Arizona, Dakotas and Oklahoma respectively. Genetic analysis was conducted using rank-inverse-normal transformed SU concentrations, which were regressed on covariates such as age, sex, and their interactions, diabetes status, and medications^{27,28}.

MetaboChip data analysis. MetaboChip genotyping was conducted in a subset of 2000 SHFS (Arizona = 300, Dakotas = 850, Oklahoma = 850) participants who were free of diabetes at visit 1. The final data set included 162,718 autosomal SNPs. MetaboChip data analysis, conducted in each of the three SHFS centers, revealed significant associations of SU with *SLC2A9* SNPs ($P < 4 \times 10^{-7}$); rs13145758, rs9998811, rs7862063 in Arizona and rs4481233 in Dakotas and Oklahoma. The minor allele frequencies (MAFs) ranged between 25 and 44%, and the effect sizes (proportion of the residual phenotypic variance that is explained by the minor allele of the SNP) ranged between 4 and 6% (Table 2). The most significant SNP in Arizona was rs13145758 whereas it was rs4481233 in Oklahoma and the Dakotas. Several other SNPs showed associations at $P < 1 \times 10^{-5}$ including rs4148155, rs1481012 and rs2231142 of *ABCG2*.

Meta-analysis. As a follow-up, we conducted a meta-analysis of SNPs associated with SU concentrations in each of the three centers. The order of the MetaboChip-wide significantly associated SNPs remained similar after meta-analysis, but with increased statistical significance (most significant SNP - rs4481233 (A/G); $P = 9 \times 10^{-20}$) (Table 3). Individuals with G alleles of rs9998811, rs4148155 and rs1481012 and A alleles of rs7696092, rs4481233, rs13145758 and rs2231142 had lower SU concentrations as compared to other alleles. Table 3 shows these top seven SNPs that exhibited significant associations. Figure 1 shows the Manhattan plot with meta-analysis results showing strong association of *SLC2A9* and *ABCG2* variants on chromosome 4 with SU.

SNP	Allele1	Allele2	Zscore	P-value	Direction	Gene	Gene loc	Chr	Chr pos hg18
rs4481233	A	G	-9.28	1.7×10^{-20}	—	<i>SLC2A9</i>	Intron	4	9956079
rs9998811*	A	G	-9.02	1.8×10^{-19}	—	<i>SLC2A9</i>	Intron	4	9966477
rs7696092	A	C	7.78	7.2×10^{-19}	+++	<i>SLC2A9</i>	Intron	4	10025320
rs13145758*	A	G	8.98	2.8×10^{-14}	+++	<i>SLC2A9</i>	Intron	4	9981997
rs4148155*	A	G	-5.98	2.2×10^{-09}	—	<i>ABCG2</i>	Intron	4	89054667
rs2231142*	A	C	5.94	2.8×10^{-09}	+++	<i>ABCG2</i>	Missense	4	89052323
rs1481012*	A	G	-5.94	2.9×10^{-09}	—	<i>ABCG2</i>	Intron	4	89039082

Table 3. Variants associated with serum urate in meta-analysis of the three centers. *SNPs are in linkage disequilibrium (LD; $r^2 > 0.90$); SNP: single nucleotide polymorphism; P-value: P-values from meta-analysis; Gene loc: Gene location; Chr pos: chromosome position in base pairs; *SLC2A9*: solute carrier family 2, member 9; *ABCG2*: ATP-binding cassette family G, member 2.

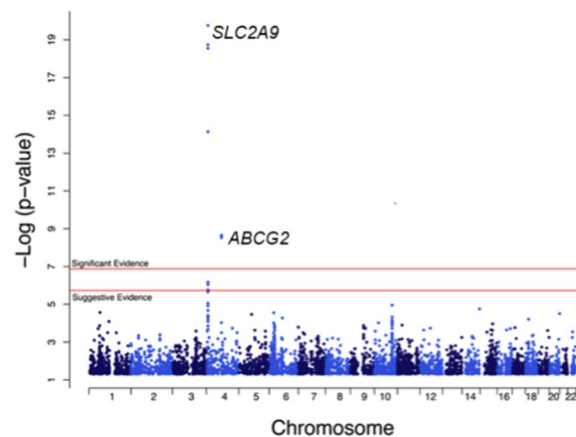


Figure 1. Genome-wide association analysis shows strong association of serum urate with *SLC2A9* and *ABCG2* SNPs.

SLC2A9 sequence variants. To identify American Indian-specific SNPs in our top hit, we conducted targeted sequencing of key regions of *SLC2A9* including all exons, 74Kb of introns and 10 kb of upstream and downstream regions of the gene in 902 SHFS founders (individuals with offspring in the SHFS but no parents). A total of 427 autosomal, polymorphic variants were identified in the 96 kb region of *SLC2A9* that was sequenced. These included 233 single nucleotide polymorphisms (SNPs; MAF $\geq 1\%$); 125 single nucleotide variants (SNVs; minor allele count ≥ 2); 26 singletons (variants found only in one of our samples); and 43 indels/triallelics (insertions or deletions or more than two alleles); 117 variants of these were novel based on comparison with dbSNP database (Supplementary Table 1). The 233 SNPs were then genotyped in all 3000 SHFS (Arizona = 586, Dakotas = 1208, Oklahoma = 1206) participants of all three centers. A total of 89 SNPs were associated at the significance level of $P < 2 \times 10^{-4}$ after adjustment for multiple tests. Table 4 lists the top 10 SNPs and their associations with SU concentrations. The rest of the SNPs and their association with SU are shown in Supplementary Table 2. Figure 2 shows LD patterns of these 10 SNPs in the SHFS and other ethnicities from the 1000 Genomes project²⁹.

Genetic analysis conditional on significant *SLC2A9* and *ABCG2* variants. To identify secondary signals for loci associated with SUA, we repeated GWAS of SU conditional on the significant *SLC2A9* and *ABCG2* variants in each center³⁰. We found suggestive evidence of association of SU concentration with SNPs in nucleobindin (*NUCB1-AS1*) (rs746075, $P = 2.0 \times 10^{-6}$) in Oklahoma and neuronal PAS-domain containing protein 4 (*NPAS4*) (rs7947391, $P = 3.2 \times 10^{-6}$) in the Dakotas. No significance or suggestive level of significance was found for SNPs in the Arizona center.

Replication of SU associations in independent cohorts. To ascertain whether SU-associated SNPs, particularly the novel ones, identified in American Indians are generalizable to other ethnicities, we conducted replications in four different cohorts: Europeans (publicly available data)¹⁵, Mexicans-mestizos³¹, Indians³² and East Asians (publicly available data)³³. Table 5 shows the SNPs identified in SHFS along with the association results in the other four ethnicities. SNPs rs62293300 and rs4385059 of *SLC2A9* were associated with SU with direction of the effect in American Indians being consistent with Mexican-mestizos. rs4385059, rs28592748 and rs7696092 of *SLC2A9* were associated with SU in all three populations. However, the direction of the effect was same in American Indians, Mexican-mestizos and Indians but different in Europeans and East Asians. The SU

SNP*	Chr-pos hg19	Major/Minor Allele	MAF	P-value	β	SE	Effect Size
rs4481233	9956079	G/A	0.30	4.2×10^{-24}	-0.35	0.03	0.05
rs11723439	9951819	G/A	0.28	1.5×10^{-23}	-0.35	0.03	0.05
rs28592748	9998605	G/A	0.25	2.7×10^{-22}	-0.35	0.04	0.05
rs7669607	9997801	G/A	0.25	8.2×10^{-22}	-0.35	0.04	0.05
rs13111638	9996890	G/A	0.23	1.4×10^{-21}	-0.36	0.04	0.05
rs10939650	9998440	A/G	0.35	4.1×10^{-21}	-0.31	0.03	0.04
rs4529048	9997112	A/C	0.36	5.0×10^{-21}	-0.31	0.03	0.04
rs3733588	9997303	A/G	0.35	7.3×10^{-21}	-0.31	0.03	0.04
rs1014290	10001861	A/G	0.35	7.6×10^{-21}	-0.31	0.03	0.04
rs4385059	9989233	A/G	0.23	3.6×10^{-20}	-0.34	0.04	0.04

Table 4. *SLC2A9* sequence variants associated with serum urate (Top 10 significant associations reported here). *SNPs after accounting for linkage disequilibrium (LD; $r^2 \geq 0.80$); SNP: single nucleotide polymorphism; Chr pos: chromosome position in base pairs; MAF: minor allele frequency; P-value: P-values from measured genotype analysis; β : beta coefficient of the SNP; SE: standard error; Effect Size: Proportion of the residual phenotypic variance that is explained by the minor allele of the SNP; Gene loc: Gene location.

associations with *NPAS4* SNP rs7947391 was significantly replicated in the cohort of East Asians with the direction of the effect being same as American Indians and Europeans in contrast to Mexican- Mestizos and Indians. rs746075 of *NUCB1* SNP was not associated with SU in any of the replication cohorts.

Discussion

Our extensive association analysis using genome-wide as well as candidate gene SNPs in American Indians of the SHFS showed that uric acid transporters *SLC2A9* and *ABCG2* are key genes regulating SU concentrations. Previously, significant heritability was obtained, and linkages were localized for SU concentrations in American Indian participants of the SHFS²⁷. Also, our previous candidate gene study replicated 7 *SLC2A9* gene polymorphisms in these participants in all centers combined and when stratified by recruitment center²⁸. However, so far no genome-wide analyses for SU have been reported in American Indians. In this regard, our MetaboChip data represents for the first time a detailed genome-wide investigation to identify genetic factors affecting the variation in SU in this population.

The top SNPs from our SU association analysis belonged to gene *SLC2A9* located on chromosome 4 confirming results from previous studies from our and other groups^{14–22}. The SNPs rs4481233, rs9998811, rs7696092 and rs3145758 were consistently associated with SU across centers. Our meta-analysis reproduced and strengthened our genome-wide MetaboChip results that were found for all the three study centers separately. In addition to *SLC2A9*, *ABCG2* variants also significantly affected SU concentrations, which further implicates uric acid transporters in the regulation of SU concentrations.

Identification of genetic variants underlying complex traits in minority populations in the US is challenging as they are underrepresented in genetic association studies and databases²³, particularly Native Americans. Previous studies from our group involving minority populations such as Mexican Americans¹⁹, Zuni Indians²⁰ and Hispanic children²¹ have shown *SLC2A9* to be the key gene affecting SU concentrations. The same has been shown by others in Europeans^{14,15}, Asians^{16,17} and African Americans^{12,22}. However, the associated variants seem to differ by population. Three recent studies have found significant association of SU with rs2231142 of *ABCG2* and rs7678287 of *SLC2A9* in Mexican-mestizos³¹ and rs2231142 of *ABCG2* and rs3775948 of *SLC2A9* in Indians³² and rs7679724 of *SLC2A9* and rs4148155 of *ABCG2* in Japanese individuals³³. While rs2231142 and rs4148155 of *ABCG2* and rs3775948 were strongly associated with SUA in our study, rs7678287 and rs7679724 were not associated with SU in either individual centers or in the meta-analysis.

Our MetaboChip (center-specific and meta-analysis) association analysis has consistently shown rs4481233 of *SLC2A9* to be strongly associated with SU concentration, further confirmed by our sequence variant analysis. An intronic variant, rs4481233, has been shown to be strongly associated with urate and gout^{34–36}. The minor allele (A) of rs4481233 has been shown to be associated with lower concentrations of SU³⁴ replicated by our study where SU was decreased by 0.32 mg/dl with every allele of the minor allele. Similar results have also been reported in a GWAS of untargeted serum metabolomics in about 3000 individuals from two large population-based European cohorts where rs4481233 was found to be strongly associated with urate^{35,36}. Although intronic, rs4481233 is in high LD with missense variant rs16890979 and is part of a LD block that contains the polymorphic *Alu* elements³⁷ and regulatory motifs for Ets- and TCF12-family of transcription factors³⁸ with potential for affecting splicing and gene expression.

We also found significant associations of SU with SNPs rs2231142, rs4148155, and rs1481012 (all three in LD) belonging to *ABCG2*, yet another uric acid transporter gene. Notably, rs2231142 is a missense variant of *ABCG2* and is likely functional. This variant has been extensively reported to be associated with SU in several populations including Mexican-mestizos³¹, Mexican Americans^{13,39}, European Americans, African Americans, and Asian populations^{12–17}, although *SLC22A12* has been shown to be strongest gene to be associated with SU in East Asians¹⁸. *ABCG2* variants had not only been shown to be associated with SU concentrations, but also with fasting glucose⁴⁰, and different forms of cancer^{41,42}, especially rs1481012 associated with decreased risk of

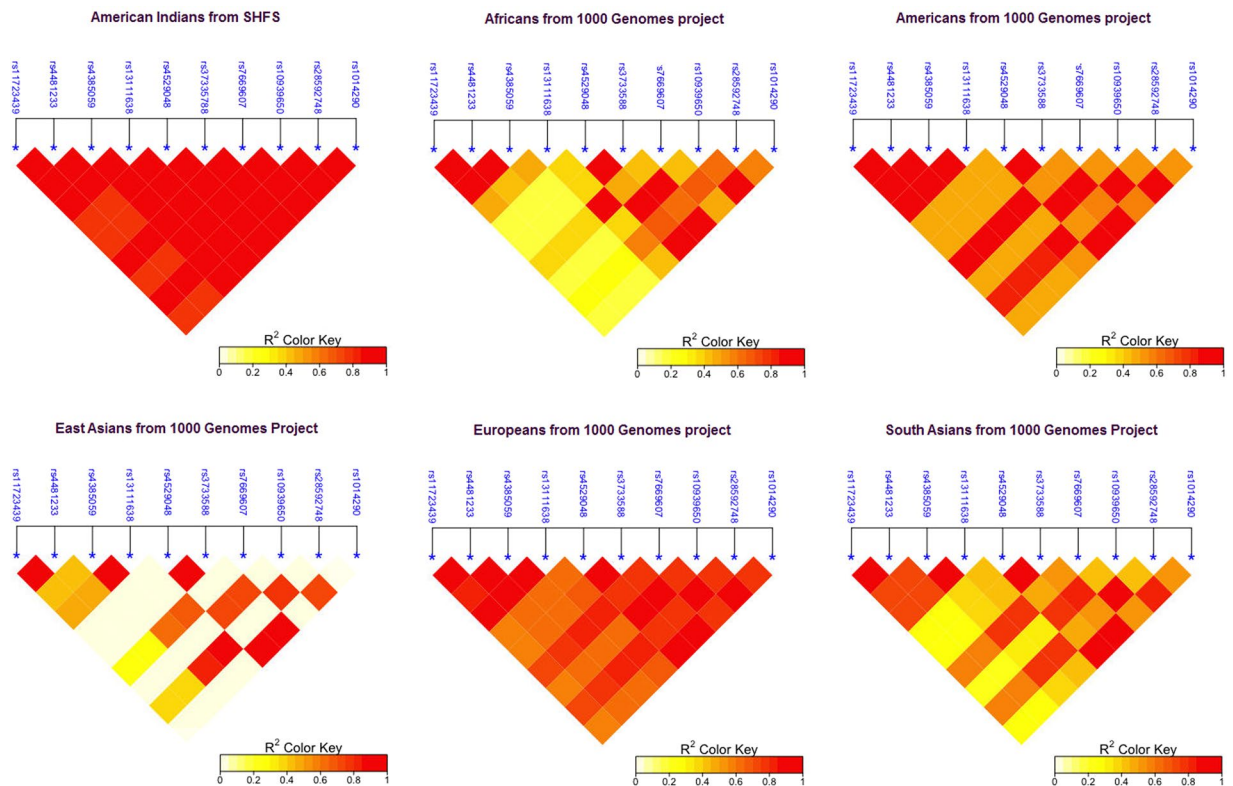


Figure 2. Comparison of LD patterns between ethnicities for top 10SU-associated *SLC2A9* sequence variants.

colorectal cancer⁴³. This gene, however, was not associated with SU in a cohort of Zuni Indians, another American Indian group²⁰.

Other SNPs that were linked to variation in SU concentrations were rs7862063, located on chromosome 9, and rs6688009 of phosphatidic acid phosphatase type 2 (*PPAP2B* or *PLPP3*) gene in the Arizona center. Studies have indicated a role for *PPAP2B* in adipogenesis⁴⁴ and vascular inflammation⁴⁵. It seems to have a protective role in endothelial dysfunction by negatively regulating inflammatory cytokines⁴⁵. However, this association was not shown by our other two centers or replication cohorts. In the Dakotas, we observed marginal association of SU with rs179409 of potassium channel, voltage gated KQT-like subfamily Q, member 1 (*KCNQ1*) gene, which has been shown to associate with gout and hyperuricemia^{46,47}. The potential mechanism seems to be through alterations in innate immunity^{47,48}, though, these variants and our SNP are about ~3 Mb apart and also not in LD with each other. Therefore, these associations need further investigations. *KCNQ1* has been widely reported to be associated with body mass index and other anthropometric measures, and also with type 2 diabetes mellitus in several populations^{49–51}.

Furthermore, our conditional analysis identified two novel genes, *NUCB1* and *NPAS4*, to be associated with SU. Conditional analysis is a tool that is used to identify secondary signals that may be otherwise masked by the strong effects of lead SNPs³⁰. The *NUCB1* and *NPAS4* SNPs were associated with lower SU concentrations and found only in the Oklahoma and Dakota centers, respectively. Center-specific statistics for these two SNPs show considerable differences in minor allele frequencies between them, rs7947391 of *NPAS4* – 0.05, 0.22, 0.13 and rs746075 of *NUCB1* – 0.31, 0.49, 0.36 in AZ, DK, OK, respectively. *NUCB1* is a Golgi-protein with potential role in calcium homeostasis and immunity⁵². It is believed to control protein unfolding in Alzheimer's disease⁵³ and stimulate insulin secretion⁵⁴. *NPAS4* is a neuronal transcription factor involved in the regulation of cognitive functions in the brains⁵⁵. Its association with SU in our study was replicated in Japanese individuals of East Asian ancestry³³. In addition, this study also found a significant expression quantitative trait locus (eQTL) for *NPAS4* in monocytes ($p = 6 \times 10^{-10}$). The common link between *NPAS4* and SU seems to be oxidative stress and inflammation-associated ischemia in the brain⁵⁶. This assumes significance considering the increasing importance of uric acid in cognitive, and neuronal function and its recognition as an important biomarker for Parkinson's disease^{57,58}.

Sequence analysis of key regions of *SLC2A9* identified 384 SNPs/SNVs/singleton variants, 117 variants of these were novel based on comparison with dbSNP database. Several of those SNPs, including rs3775946, rs6826764, rs6823877, rs56239136, rs4697693, rs2240721, rs1107710, and rs7698858 have not been previously reported as affecting SU concentrations. One SNP (chr_pos: 4_10027969) was found to be novel. Although, most studies reported the SNPs in the *SLC2A9* locus to be significantly associated with SU in different populations, *SLC2A9* variants such as rs3733585, rs6855911, rs1014290 and rs12499857 associated with SU in our study have also been linked to Parkinson's disease^{59,60}, type 2 diabetes⁶¹, anxiety disorders⁶² and nonsyndromic cleft palate⁶³. Our results also showed that the minor alleles of most of these SNPs were associated with lower SU concentrations.

Gene ^a	SNP ^b	Chr. Chr. pos.hg19	Proxy SNP ^c	American Indian ancestry			Mexican-Mestizo ancestry			European ancestry ^d			Indian ancestry			East Asian ancestry ^e		
				Minor allele (frequency)	β (SE) ^f	P value	Minor allele (frequency)	β (SE)	P value	Minor allele (frequency)	β (SE)	P value	Minor allele (frequency)	β (SE)	P value	Minor allele (frequency)	β (SE)	P value
<i>PLPP3</i>	rs6688009	1: 24593576		A (0.01)	1.1 (0.2)	9.1×10^{-6}	A (0.07)	0.08 (0.06)	0.20	T (0.17)	0.004 (0.007)	0.58				T (0.01)	0.04 (0.02)	0.05
<i>SLC2A9</i>	rs62293300	4: 9890359		T (0.01)	0.44 (0.2)	0.052	A (0.03)	0.17 (0.09)	0.07									
<i>SLC2A9</i>	NOVEL	4: 9896970		C (0.05)	0.25 (0.12)	0.03												
<i>SLC2A9</i>	NOVEL	4: 9897372	rs13136962	A (0.34)	0.15 (0.05)	0.001							G (0.44)	0.07 (0.06)	0.20	C (0.32)	-0.03 (0.004)	7.3×10^{-11}
<i>SLC2A9</i>	NOVEL	4: 9923260		T (0.01)	-0.57 (0.22)	0.008												
<i>SLC2A9</i>	NOVEL	4: 9930721		T (0.001)	-1.47 (0.8)	0.049												
<i>SLC2A9</i>	NOVEL	4: 9931893		C (0.001)	2.2 (1.1)	0.039												
<i>SLC2A9</i>	NOVEL	4: 9986995		C (0.26)	-0.13 (0.06)	0.031												
<i>SLC2A9</i>	rs4385059	4: 9989233	rs6449213		-0.34 (0.04)	3.7×10^{-20}	G (0.22)	-0.48 (0.04)	3.1×10^{-41}	C (0.18)	0.39 (0.007)	0.000	G (0.18)	-0.17 (0.07)	0.019	C (0.01)	0.20 (0.02)	3.9×10^{-19}
<i>SLC2A9</i>	rs28592748	4: 9998605	rs6449213		-0.36 (0.04)	2.7×10^{-22}	A (0.24)	-0.48 (0.04)	8.6×10^{-35}	C (0.18)	0.39 (0.007)	0.000	G (0.18)	-0.17 (0.07)	0.019	T (0.01)	0.18 (0.02)	7.4×10^{-16}
<i>SLC2A9</i>	rs7696092	4: 10025320	rs4543113	C (0.29)	-0.31 (0.05)	1×10^{-10}	C (0.26)	-0.44 (0.03)	7.0×10^{-39}	A (0.40)	0.17 (0.005)	3.7×10^{-202}	G (0.25)	-0.13 (0.06)	0.034	C (0.01)	0.18 (0.02)	5.1×10^{-16}
<i>SLC2A9</i>	NOVEL	4: 10026105		A (0.001)	2.2 (1.1)	0.037												
<i>SLC2A9</i>	NOVEL	4: 10027969		T (0.26)	-0.32 (0.05)	5.4×10^{-10}												
<i>SLC2A9</i>	NOVEL	4: 10030657		G (0.001)	2.12 (1.0)	0.039												
<i>SLC2A9</i>	NOVEL	4: 10036877		C (0.0007)	1.12 (0.6)	0.066												
<i>SLC2A9</i>	NOVEL	4: 10039575		G (0.012)	0.57 (0.24)	0.02												
-	rs7862063	9: 110002036	rs10978764	G (0.41)	0.29 (0.05)	4.5×10^{-8}	A (0.18)	-0.008 (0.05)	0.86	C (0.40)	0.002 (0.006)	0.72	G (0.43)	0.015 (0.06)	0.79	C (0.35)	0.001 (0.005)	0.75
<i>KCNQ1</i>	11_2483882	11: 2483882		A (0.45)	-0.23 (0.05)	3.4×10^{-6}												
<i>NPAS4</i>	rs7947391	11: 66186882		A (0.22)	-0.26 (0.06)	5.6×10^{-6}	G (0.48)	0.011 (0.04)	0.77	A (0.19)	-0.0009 (0.006)	0.89	A (0.20)	0.057 (0.07)	0.39	A (0.23)	-0.01 (0.005)	0.003
<i>NUCB1</i>	rs746075	19: 49416936		A (0.41)	0.22 (0.05)	2.8×10^{-6}				T (0.42)	0.002 (0.008)	0.80	A (0.37)	-0.106 (0.06)	0.067	A (0.41)	-0.0003 (0.004)	0.95

Table 5. Replication of the associations observed in SHFS. ^a*PLPP3* or *PPAP2B*: Phospholipid phosphatase 3; *SLC2A9*: Solute carrier family 2 member 9; *KCNQ1*: Potassium voltage-gated channel subfamily Q member 1; *NPAS4*: Neuronal PAS domain protein 4; *NUCB1*: Nucleobindin 1; ^bSNP: Single nucleotide polymorphism (index SNP); ^cProxy SNP: SNP in high LD with the index SNP. Used by replication studies when the original SNP data were not available; ^d<http://metabolomics.helmholtz-muenchen.de/gugc/> and http://useast.ensembl.org/Homo_sapiens/Variation/Explore?r=4:9950721; ^eKanai *et al.* 2018; Nature Genet. ^f β (SE): Beta coefficient (standard error).

There are some limitations of the study. First, the MetaboChip may not be the best option for identifying SNP associations with SU in this population. Secondly, the sample size for GWAS may be only moderate. However, family-based studies, a key strength of this study, are favorable for detection of significant associations and/or gene discoveries as they are homogenous, are robust to the effects of population stratification and have increased power to detect novel associations due to reduced residual variance⁶⁴. Another strength of this study is inclusion of an understudied population with unique genetic and environmental background and availability of extensive covariate information.

In summary, our results replicated known associations of uric acid transporters with SU in a minority population of American Indians and demonstrated potential novel associations of SU with neuronal-related genes which need further investigation.

Methods

Study population: strong heart family study (SHFS). The Strong Heart Family Study (SHFS) is a genetic study of CVD risk in American Indians. Description of the phenotypes, SU measurement techniques, and other related analytical approaches have been detailed elsewhere^{27,28}. In short, the SHFS is a multi-center family-based genetic study in American Indian communities from Arizona, the Dakotas, and Oklahoma, which are experiencing extraordinarily high rates of progressive chronic kidney disease (CKD), obesity, diabetes, CVD, and

diabetic nephropathy. Approximately 3000 members (including 902 founders) belonging to multigenerational families of Arizona, North and South Dakota (Dakotas), and Oklahoma participated in the study. These individuals aged 14 to 93 years were recruited without regard to disease status in 1998^{65,66}. The Indian Health Service Institutional Review Board and the Institutional Review Boards from Texas Biomedical Research Institute and the University of North Carolina at Chapel Hill approved the SHFS protocol and all subjects gave informed consent. For participants under the age of 18 years, informed consent was obtained from their parent or legal guardian. Study design and methods of the SHFS are in accordance with institutional guidelines and have been described previously^{65,66}.

Phenotyping. Blood was collected after an overnight fast. Uric acid concentrations in serum were assayed in the SHFS central laboratory by the uricase and peroxidase method⁶⁷.

Genotyping. *MetaboChip data.* Blood collected from individuals who were free of diabetes at baseline visit ($n = 2000$) was used for this study. Cardio-Metabo DNA Analysis BeadChip (Illumina catalog# WG-310-1001 or WG-310-1002) was used for genotyping. The MetaboChip contained 196,725 markers. The original annotation file for the Cardio-Metabo BeadChip is *MetaboChip_Gene_Annotation*. Simwalk2 was used to remove genotyping inconsistencies⁶⁸. Participants were excluded if genotyping call rate was $<95\%$ ($n = 3$). SNPs were excluded if the call rate $<98\%$ ($n = 0$), not autosomal ($n = 250$), no data after imputation ($n = 33,599$) or Hardy-Weinberg equilibrium $P < 1 \times 10^{-5}$ ($n = 20,067$). Pairwise correlations (r^2) between markers were calculated to estimate linkage disequilibrium (LD). The final cleaned, imputed data set includes 162,718 autosomal marker information available for 2000 American Indian participants.

Sequencing of SLC2A9 gene. We sequenced 96 kb of the *SLC2A9* gene, using a Illumina's TruSeq Custom Amplicon kit and MiSeq Sequencer, in 902 founders of multigenerational families. The target regions contained all exons, 2.2 kb, 74 kb of introns, and 10 kb of upstream and downstream region of the gene. Illumina generated sequence data (BAM files) were aligned to the Human Genome Reference Sequence version 37.1 (hg19). Variants were called, recalibrated and QC'd using the Genome Analysis Toolkits (GATK v.3.3) Haplotype Caller⁶⁹. Pairwise correlations (r^2) between markers were calculated to estimate linkage disequilibrium (LD). We identified 427 autosomal, non-monomorphic variants, 384 of which affected a single base (233 single nucleotide polymorphisms (SNPs; $MAF \geq 1\%$); 125 single nucleotide variants (SNVs; minor allele count ≥ 2); 26 singletons (variants found only in one of our samples)); and 43 were indels/triallelic (insertions or deletions or more than two alleles). Out of the forty-three indel/triallelic variants, 14 variants were listed in dbSNP (rs140391260, rs5856025, rs34839464, rs137899691, rs35950306, rs139025036, rs35614040, rs58702202, rs112058434, rs66622652, rs60841869, rs142713311, rs3834235, and rs66943961). The MAFs of all variants ranged between 0.1 and 49%, except for indels/triallelic.

Statistical analysis. *Genotype cleaning and population stratification assessment.* Genotype frequencies for each SNP were estimated and tested for departures from Hardy-Weinberg equilibrium in the software package, Sequential Oligogenic Linkage Analysis Routines (SOLAR)⁷⁰. Also, we used principal component (PC) scores to model differences in ancestral contributions among study participants for MetaboChip data. PCs were calculated using the unrelated SHFS founders and a subset of 15,158 selected SNPs ($r^2 < 0.1$; $MAF > 0.05$). PCA was performed on a matrix of "doses" (copies of minor allele) for the selected SNPs, using "prcomp" in R. The PC scores were then predicted for all genotyped individuals using the PCA model fit to the founder data^{71,72}. While no PC accounted for a large percentage of total variance in genotypes scores, the first four PCs account for substantially more than the rest and were, therefore, included as additional covariates in association analyses.

Measured genotype analysis (MGA). The association of SNPs with SU was estimated using a measured genotype analysis (MGA)⁷³ executed in SOLAR after accounting for family relationships based on variance components approach. This approach allows us to account for the non-independence among family members. To minimize the problem of non-normality, the SU data were inverse-normal-transformed using SOLAR. All analyses involved adjustment for the covariate effects (see results). The appropriate significance level was determined to be $P < 4 \times 10^{-7}$ for MetaboChip data, and $P < 2 \times 10^{-4}$ for sequence data after correcting for multiple tests.

METAL. METAL⁷⁴ software was used to perform meta-analysis of GWAS results taken from the three study centers, each study containing individual genome-wide MetaboChip association results for multiple markers.

Conditional analysis and functional annotation of significant variants. To identify additional independent loci that are associated with SU concentrations, we performed association analysis conditioned on significant *SLC2A9* and *ABCG2* SNPs³⁰.

Replication studies. *Mexican-mestizos studies.* Seven of the SNPs that reached genome-wide suggestive significance in the discovery phase were tested for replication in an independent cohort of Mexican Mestizo individuals (1,061 children and 1,101 adults). Population characteristics, biochemical measurements and genotyping have been previously described. Briefly, genotypes of six SNPs were obtained from a Multi-Ethnic Genotyping Array (MEGA, Illumina, San Diego, CA, USA), while rs7947391 genotypes were imputed using 1000 Genomes Project and Native Mexican individuals as ref.³¹. Associations with SUA were tested separately in children and adults in a linear mixed model that considered the genetic relatedness matrix as a random effect, while genotype, age, sex and BMI percentile or body mass index (kg/m^2) were included as fixed effects. Results were meta-analyzed with the inverse variance method⁷⁵.

Indian diabetes consortium. The study participants included the members of the INdian DIabetes Consortium (INDICO)⁷⁶. Details of the study recruitment and phenotype measurements are given in Giri *et al.*³². In short, samples were enrolled in the study by conducting diabetes awareness camp organized in various parts of North India. Prior informed written consent was obtained from the study participants. The study was approved by the Human Ethics Committee of the CSIR-Institute of Genomics and Integrative Biology and the All India Institute of Medical Sciences research Ethics Committee. The study was conducted in accordance with the principles of the Helsinki Declaration. Genotyping was conducted using the Illumina Human 610-quad bead chip array. Association with SU concentrations was tested using linear regression models in PLINK⁷⁷. Sex, age, BMI and first three principal components of genotypes were used as covariates in the model.

European studies. Replication analysis in Europeans was conducted with publicly available data from Kottgen *et al.*¹⁵. We used the summary statistics from meta-analysis of serum urate from the article by Kottgen *et al.* The meta-analyses comprised of 14 studies totaling 2,115 cases and 67,259 controls [<http://metabolomics.helmholtz-muenchen.de/gugc/>, http://useast.ensembl.org/Homo_sapiens/Variation/Explore?r=4:9950721-9951721;source=dbSNP;v=rs1079128;vdb=variation;vf=250145324].

East asian studies. Replication analysis in East Asians was conducted with publicly available data from Kanai *et al.*³³. We used the summary statistics from a genome-wide association analysis of serum uric acid in Japanese individuals from Kanai *et al.* The GWAS was conducted for 58 quantitative traits, including serum uric acid, in 162,255 individuals [<https://www.ncbi.nlm.nih.gov/pubmed/?term=Kanai+M%2C+2018%2C+Nature+Genetics>]⁷⁸ This research was supported by the Tailor-Made Medical Treatment Program (the BioBank Japan Project) of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and the Japan Agency for Medical Research and Development (AMED)⁷⁸.

Data availability

The Strong Heart Study is conducted as a partnership between the American Indian Tribes that are part of the study and the study investigators. All the intellectual property and data generated by this project is administered according to policies from the Tribal Nations, research organizations that are involved in the study, and the NIH. The data is owned by the Tribal Nations, not the study investigators. The study investigators accessed the data used in this manuscript through a formal request for data after a paper proposal was approved by the Strong Heart Study Publication and Presentation committee and following all the procedures that have been approved by the Tribal Nations. The protocols for paper proposal and data access requests can be found on the SHS website: <http://strongheart.ouhsc.edu/>. The authors confirm that interested researchers may apply for access to these data in the manner described.

Received: 23 January 2019; Accepted: 12 October 2019;

Published online: 29 November 2019

References

- Merriman, T. R. & Dalbeth, N. The genetic basis of hyperuricaemia and gout. *Joint Bone Spine*. **78**, 35–40 (2011).
- Nakagawa, T. *et al.* Unearthing uric acid: An ancient factor with recently found significance in renal and cardiovascular disease. *Kidney Int.* **69**, 1722–1725 (2006).
- Kutzing, M. K. & Firestein, B. L. Altered uric acid levels and disease states. *J. Pharmacol. Exp. Ther.* **324**, 1–7 (2008).
- Nakagawa, T., Tuttle, K. R., Short, R. A. & Johnson, R. J. Hypothesis: Fructose-induced hyperuricemia as a causal mechanism for the epidemic of the metabolic syndrome. *Nat. Clin. Pract. Nephrol.* **1**, 80–6 (2005).
- Caulfield, M. J. *et al.* SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med.* **5**, e197 (2008).
- Anzai, N. & Endou, H. Urate transporters: an evolving field. *Semin. Nephrol.* **31**, 400–409 (2011).
- Paganoni, S. & Schwarzschild, M. A. Urate as a marker of risk and progression of neurodegenerative disease. *Neurotherapeutics*. **14**, 148–153 (2017).
- Moccia, M. *et al.* Uric acid in relapsing-remitting multiple sclerosis: a 2-year longitudinal study. *J. Neurol.* **262**, 961–967 (2015).
- Li, X. *et al.* Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomized controlled trials and Mendelian randomization studies. *BMJ*. **357**, j2376 (2017).
- Clarson, L. E. *et al.* Increased cardiovascular mortality associated with gout: a systematic review and meta-analysis. *Eur J Prev Cardiol.* **22**, 335–343 (2015).
- Roughley, M. *et al.* Risk of chronic kidney disease in patients with gout and the impact of urate lowering therapy: a population-based cohort study. *Arthritis Res Ther.* **30**, 243 (2018).
- Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum. Mol. Genet.* **20**, 4056–68 (2011).
- Zhang, L. *et al.* Association of functional polymorphism rs2231142 (Q141K) in the ABCG2 gene with serum uric acid and gout in 4 US populations: The PAGE study. *Am. J. Epidemiol.* **177**, 923–32 (2013).
- Dehghan, A. *et al.* Association of three genetic loci with uric acid concentration and risk of gout: A genome-wide association study. *Lancet*. **372**, 1953–61 (2008).
- Köttgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
- Yang, B. *et al.* A genome-wide association study identifies common variants influencing serum uric acid concentrations in a Chinese population. *BMC Med. Genomics*. **7**, 10 (2014).
- Matsuo, H. *et al.* Common defects of ABCG2, a high-capacity urate exporter, cause gout: A function-based genetic analysis in a Japanese population. *Sci. Transl. Med.* **1**, 5–11 (2009).
- Nakayama, A. *et al.* GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include transporter genes. *Ann. Rheum. Dis.* **76**, 869–877 (2017).
- Voruganti, V. S. *et al.* Genome-wide association analysis confirms and extends the association of SLC2A9 with serum uric acid levels to Mexican Americans. *Front. Genet.* **4**, 279 (2013).
- Laston, S. L. *et al.* Genetics of kidney disease and related cardiometabolic phenotypes in Zuni Indians: the Zuni Kidney Project. *Front. Genet.* **6**, 6 (2015).

21. Voruganti, V. S. *et al.* Serum uric acid concentrations and SLC2A9 genetic variation in Hispanic children: The Viva La Familia study. *Am. J. Clin. Nutr.* **101**, 725–32 (2015).
22. Charles, B. A. *et al.* A genome-wide association study of serum uric acid in African Americans. *BMC Med. Genomics.* **4**, 17 (2011).
23. Bentley, A. R., Callier, S. & Rotimi, C. N. Diversity and inclusion in genomic research: Why the uneven progress? *J. Community Genet.* **8**, 255–66 (2017).
24. Deen, J. F. *et al.* Cardiovascular disease in American Indian and Alaska Native youth: Unique risk factors and areas of scholarly need. *J. Am. Heart Assoc.* **6**, e007576 (2017).
25. Franceschini, N. *et al.* The association of genetic variants of type 2 diabetes with kidney function. *Kidney Int.* **82**, 220–225 (2012).
26. North, K. E. *et al.* Evidence for joint action of genes on diabetes status and CVD risk factors in American Indians: the strong heart family study. *Int. J. Obes. Relat. Metab. Disord.* **27**, 491–497 (2003).
27. Voruganti, V. S. *et al.* Genetic influence on variation in serum uric acid in American Indians: The strong heart family study. *Hum. Genet.* **126**, 667–676 (2009).
28. Voruganti, V. S. *et al.* Replication of the effect of SLC2A9 genetic variation on serum uric acid levels in American Indians. *Eur. J. Hum. Genet.* **22**, 938–943 (2014).
29. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* **526**, 68–74 (2015).
30. Yang, J., Ferreira, T., Morris, A. P. & Medland, S. E. Genetic investigation of anthropometric traits (GIANT) consortium, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **201**, 369–373
31. Macias-Kaufer, L. R. *et al.* Genetic contributors to serum uric acid levels in Mexicans and their effect on premature coronary artery disease. *Int. J. Cardiol.* **279**, 168–173 (2019).
32. Giri, A. K. *et al.* Genome-wide association study of uric acid in Indian population and interaction of identified variants with type 2 diabetes. *Sci. Rep.* **6**, 21440 (2016).
33. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
34. Bhatnagar, V. *et al.* Analysis of ABCG2 and other urate transporters in uric acid homeostasis in chronic kidney disease: potential role of remote sensing and signaling. *Clin. Kidney J.* **9**, 444–453 (2016).
35. Albrecht, E. *et al.* Metabolite profiling reveals new insights into the regulation of serum urate in humans. *Metabolomics.* **10**, 141–51 (2014).
36. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature.* **477**, 7362 (2011).
37. Payer, L. M. *et al.* *Alu* insertion variants alter mRNA splicing. *Nucleic Acids Res.* **47**, 421–431 (2019).
38. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
39. Hall, M. A. *et al.* Detection of pleiotropy through a phenome-wide association study (PheWAS) of epidemiologic data as part of the environmental architecture for genes linked to environment (EAGLE) study. *PLoS Genet.* **10**, e1004678 (2014).
40. Sun, X. *et al.* Common variants related to serum uric acid concentrations are associated with glucose metabolism and insulin secretion in a Chinese population. *PLoS One.* **10**, e0116714 (2015).
41. Xie, Z. Y., Lv, K., Xiong, Y. & Guo, W. H. ABCG2-mediated multidrug resistance and tumor-initiating capacity of side population cells from colon cancer. *Oncol. Res. Treat.* **37**, 666–668 (2014).
42. Wu, Y. *et al.* Cholesterol reduces the sensitivity to platinum-based chemotherapy via upregulating ABCG2 in lung adenocarcinoma. *Biochem. Biophys. Res. Commun.* **457**, 614–620 (2015).
43. Campa, D. *et al.* A gene-wide investigation on polymorphisms in the ABCG2/BRCP transporter and susceptibility to colorectal cancer. *Mutat. Res.* **645**, 56–60 (2008).
44. Bianchini, L. *et al.* Identification of PPAP2B as a novel recurrent translocation partner gene of HMGA2 in lipomas. *Genes Chromosomes Cancer.* **52**, 580–590 (2013).
45. Touat-Hamici, Z. *et al.* Role of lipid phosphate phosphatase 3 in human aortic endothelial cell function. *Cardiovasc Res.* **112**, 702–713 (2016).
46. Li, C. *et al.* Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nat Commun.* **6**, 7041 (2015).
47. Sakiyama, M. *et al.* Common variant of BCAS3 is associated with gout risk in Japanese population: the first replication study after gout GWAS in Han Chinese. *BMC Med Genet.* **19**, 96 (2018).
48. Zhu, W., Deng, Y. & Zhou, X. Multiple membrane transporters and some immune regulatory genes are major genetic factors to gout. *Open Rheumatol.* **12**, 94–113 (2018).
49. Bazzi, M. D. *et al.* Association between FTO, MC4R, SLC30A8, and KCNQ1 gene variants and type 2 diabetes in Saudi population. *Genet. Mol. Res.* **13**, 10194–10203 (2014).
50. Wang, J. *et al.* Association of KCNQ1 and KLF14 polymorphisms and risk of type 2 diabetes mellitus: A global meta-analysis. *Hum. Immunol.* **75**, 342–347 (2014).
51. Wen, W. *et al.* Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–504 (2014).
52. Kanuru, M. & Aradhya, G. K. Chaperone-like activity of calnexin prevents amyloid aggregation. *Biochemistry.* **56**, 149–59 (2017).
53. Bonito-Oliva, A., Barbash, S., Sakmar, T. P. & Graham, W. V. Nucleobindin 1 binds to multiple types of pre-fibrillar amyloid and inhibits fibrillization. *Sci. Rep.* **7**, 42880 (2017).
54. Ramesh, N., Mohan, H. & Unniappan, S. Nucleobindin-1 encodes a nesfatin-1-like peptide that stimulated insulin secretion. *Gen. Comp. Endocrinol.* **216**, 182–9 (2015).
55. Coutellier, L. *et al.* NPAS4: A neuronal transcription factor with a key role in social and cognitive functions relevant to developmental disorders. *PLOS One.* **9**, e46604 (2012).
56. Choy, F. C. *et al.* Reduction of the neuroprotective transcription factor Npas4 results in increased neuronal necrosis, inflammation and brain lesion size following ischemia. *J. Cerebral Blood Flow & Metabolism.* **36**, 1449–1463 (2016).
57. Ascherio, A. *et al.* Urate as a predictor of the rate of clinical decline in Parkinson disease. *Arch. Neurol.* **66**, 1460–68 (2009).
58. Simon, K. C. *et al.* Mendelian randomization of serum urate and Parkinson's disease progression. *Ann. Neurol.* **76**, 862–68 (2014).
59. Gao, J., Xu, H., Huang, X. & Chen, H. Short communication: genetic variations of SLC2A9 in relation to Parkinson's disease. *Translational Neurodegeneration.* **2**, 5 (2013).
60. Facheris, M. F. *et al.* Variation in the uric acid transporter gene SLC2A9 and its association with AAO of Parkinson's disease. *J. Mol. Neurosci.* **43**, 246–250 (2011).
61. Liu, W. C. *et al.* The rs1014290 polymorphisms of the SLC2A9 gene is associated with type 2 diabetes in Han Chinese. *Exp. Diabetes Res.* **2011**, e527520 (2011).
62. Lyngdoh, T. *et al.* Association of serum uric acid and SLC2A9 variant with depressive and anxiety disorders: a population-based study. *PLOS One.* **8**, e76336 (2013).
63. Wu, T. *et al.* Evidence of gene-environment interaction for two genes on chromosome 4 and environmental tobacco smoke in controlling the risk of nonsyndromic cleft palate. *PLoS One.* **9**, e88088 (2014).
64. Borecki, I. B. & Province, M. A. Genetic and genomic discovery using family studies. *Circulation.* **118**, 1057–1063 (2008).

65. Lee, E. T. *et al.* The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am. J. Epidemiol.* **132**, 1141–55 (1990).
66. North, K. E. *et al.* Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am. J. Epidemiol.* **157**, 303–14 (2003).
67. Domagk, G. F. & Schlicke, H. H. A colorimetric method using uricase and peroxidase for the determination of uric acid. *Anal. Biochem.* **22**, 219–24 (1968).
68. Sobel, E., Papp, J. C. & Lange, K. Detection and Integration of genotyping errors in statistical genetics. *Am. J. Hum. Genet.* **70**, 496–508 (2002).
69. McKenna, A. *et al.* The Genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research.* **20**, 1297–1303 (2010).
70. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
71. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
72. Burdick, J. T., Chen, W.-M., Abecasis, G. R. & Cheung, V. G. *In silico* methods for inferring genotypes in pedigrees. *Nat. Genet.* **38**, 1002–1004 (2006).
73. Boerwinkle, E., Chakroborty, R. & Sing, C. F. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann. Hum. Genet.* **50**, 181–194 (1986).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics.* **26**, 2190–191 (2010).
75. Rucker, G., Schwarzer, G., Carpenter, J. R., Binder, H. & Schumacher, M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics.* **12**, 122–142 (2011).
76. Indian Diabetes Consortium, INDICO: the development of a resource for epigenomic study of Indians undergoing socioeconomic transition. *The HUGO Journal.* **5**, 65–69 (2011).
77. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* **81**, 559–575 (2007).
78. Nagai, A. *et al.* Overview of the Biobank Japan Project: Study design and profile. *J Epidemiol.* **27**(3 suppl), S2–S8 (2017).

Acknowledgements

We thank the study participants, Indian Health Service facilities, and participating tribal communities for their extraordinary cooperation and involvement, which has been critical to the success of the Strong Heart Family Study. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Indian Health Service or NIH. This work was supported by cooperative agreements HL65520, HL41642, HL41652, HL41654 and HL65521, and TR000101 and NIH grants R01DK092238 and R01ES021367. Development of SOLAR was supported by NIH grant MH59490. Replication studies: This work was supported by grants 113861, Fronteras 2015-2 1093, and PEI-230129 from Mexican National Council for Science and Technology (CONACyT). Luis R. Macías Kauffer was supported by graduate scholarship 131414 from CONACyT INDICO: This study was supported by grants provided by Council of Scientific and Industrial Research (CSIR), Government of India through Centre for Cardiovascular and Metabolic Disease Research (CARDIOMED) project [Grant No: BSC0122] provided to CSIR-Institute of Genomics and Integrative Biology. This study was also funded by the Department of Science and Technology, Government of India through PURSE II CDST/SR/PURSE PHASE II/11 given to Jawaharlal Nehru University.

Author contributions

G.C. and V.S.V. conceived and designed the experiments and performed or supervised all aspects of statistical analyses; Additional help with statistical analyses and replications were provided by P.B., C.B., G.P., L.M.K., T.V.M., D.B. and S.C.Q.; K.N., L.G.B. and S.L. were involved with S.H.F.S. recruitment and data collection; and the SNP genotyping and SLC2A9 sequencing were provided by S.A.C. and K.H.; Advice on the interpretation of results and preparation/editing of manuscript was provided by L.G.B., J.W.M., K.E.N., J.G.U., N.F., A.N.A. and S.A.C. and the manuscript was written by G.C. and V.S.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52924-w>.

Correspondence and requests for materials should be addressed to V.S.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019