Identification of CDC42BPG as a novel susceptibility locus for hyperuricemia in a Japanese population

CrossMark

ORIGINAL ARTICLE

Identification of *CDC42BPG* as a novel susceptibility locus for hyperuricemia in a Japanese population

Yoshiki Yasukochi[1,2] · Jun Sakuma[2,3,4] · Ichiro Takeuchi[2,4,5] · Kimihiko Kato[1,6] · Mitsutoshi Oguri[1,7] · Tetsuo Fujimaki[8] · Hideki Horibe[9] · Yoshiji Yamada[1,2]

Abstract Chronic kidney disease and hyperuricemia are serious global health problems. Recent genome-wide association studies have identified various genetic variants related to these disorders. However, most studies have been conducted in a cross-sectional manner. To identify novel susceptibility loci for chronic kidney disease or hyperuricemia, we performed longitudinal exome-wide association studies (EWASs), using ~ 244,000 genetic variants and clinical data of Japanese individuals who had undergone annual health checkups for several years. After establishing quality controls, the association of renal function-related traits in 5648 subjects (excluding patients with dialysis and population outliers) with 24,579 single nucleotide variants (SNVs) for three genetic models ($P < 3.39 \times 10^{-7}$) was tested using generalized estimating equation models. The longitudinal EWASs revealed novel relations of five SNVs to renal function-related traits. Cross-sectional data for renal function-related traits in 7699 Japanese subjects were examined in a replication study. Among the five SNVs, rs55975541 in *CDC42BPG* was significantly ($P < 4.90 \times 10^{-4}$) related to the serum concentration of uric acid in the replication cohort. We also examined the SNVs detected in our longitudinal EWASs with the information on P values in GKDGEN meta-analysis data. Four SNVs in *SLC15A2* were significantly associated with the estimated glomerular filtration rate in European ancestry populations, although these SNVs were related to the serum concentration of uric acid with borderline significance in our longitudinal EWASs. Our findings indicate that *CDC42BPG* may be a novel susceptibility locus for hyperuricemia.

Keywords Chronic kidney disease · Creatinine · Estimated glomerular filtration rate · Generalized estimating equation · Hyperuricemia · Uric acid

Communicated by S. Hohmann.

✉ Yoshiki Yasukochi
hyasukou@proof.ocn.ne.jp

1 Department of Human Functional Genomics, Advanced Science Research Promotion Center, Organization for the Promotion of Regional Innovation, Mie University, 1577 Kurima-machiya, Tsu, Mie 514–8507, Japan

2 CREST, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan

3 Computer Science Department, College of Information Science, University of Tsukuba, Tsukuba, Ibaraki, Japan

4 RIKEN Center for Advanced Intelligence Project, Chuo-ku, Tokyo, Japan

5 Department of Computer Science, Nagoya Institute of Technology, Nagoya, Aichi, Japan

6 Department of Internal Medicine, Meitoh Hospital, Nagoya, Aichi, Japan

7 Department of Cardiology, Kasugai Municipal Hospital, Kasugai, Aichi, Japan

8 Department of Cardiovascular Medicine, Inabe General Hospital, Inabe, Mie, Japan

9 Department of Cardiovascular Medicine, Gifu Prefectural Tajimi Hospital, Tajimi, Gifu, Japan

Introduction

Chronic kidney disease (CKD) and hyperuricemia are caused by renal function abnormalities and are multifactorial disorders resulting from interactions between genetic background and environmental factors. The goal of the present study was to identify novel susceptibility loci for CKD or hyperuricemia in Japanese individuals. CKD is a global health problem and an independent risk factor for cardiovascular diseases (Schiffrin et al. 2007; Hill et al. 2016; Webster et al. 2017). Recent genome-wide association studies (GWASs) have identified various genetic variants that confer susceptibility to CKD-related traits. A previous GWAS examined relations between genetic variants and the prevalence of CKD or estimated glomerular filtration rate (eGFR), based on serum creatinine and cystatin C levels in four population-based cohorts with European ancestry, and identified several genetic loci related to any of the clinical parameters examined (Köttgen et al. 2009). Several GWASs consistently identified a genetic variant at rs12917707 in the *UMOD* gene as a CKD susceptibility locus (Chambers et al. 2010; Böger et al. 2011; Pattaro et al. 2016). However, a recent GWAS for Icelanders did not detect this association (Sveinbjornsson et al. 2014). A meta-analysis of GWASs for 71,149 East Asian individuals, including those in an in silico replication study, identified another genetic variant, rs11864909, in *UMOD* and the linkage disequilibrium (LD) with rs12917707 was not strong ($r^2 = 0.02$), although rs12917707 was excluded from the GWAS meta-analysis because of low minor allele frequency (MAF < 0.01) (Okada et al. 2012). This analysis showed that 32 loci were significantly associated with four kidney function-related traits (blood urea nitrogen, serum creatinine, eGFRcrea, and uric acid).

Hyperuricemia is an important risk factor for gout (Lioté 2003). Overproduction of uric acid or decreased renal uric acid excretion may cause this disorder. Two meta-analyses of GWASs for population-based cohorts with European ancestry commonly identified relations between serum concentrations of uric acid and single nucleotide variants (SNVs) in six loci (*ABCG2, GCKR, PDZK1, SLC2A9, SLC17A1*, and *SLC22A11*) (Kolz et al. 2009; Yang et al. 2010). In previous studies, several solute carrier family (SLC) genes showed associations with hyperuricemia or gout in diverse ethnic groups (Matsuo et al. 2008; Köttgen et al. 2013; Phipps-Green et al. 2016; Nakayama et al. 2017). However, allele frequencies of SNVs associated with serum concentrations of uric acid varied across different ethnic groups (Köttgen et al. 2013). Therefore, disease-associated SNVs are different between ethnic groups owing to genetic background differences.

Extensive large-scale GWASs for renal function-related traits have been conducted focusing on populations with European ancestry. Therefore, it is possible that genetic relations in non-European populations have not been definitively defined. Conventional GWASs have been conducted in a cross-sectional manner that measures traits at a single point in time. In the present study, we traced longitudinal changes of anthropometric and clinical data in 5648 Japanese individuals who had undergone annual health checkups for several years. On the basis of these data, we performed longitudinal exome-wide association studies (EWASs) with the use of exome-array-based genotyping methods to explore novel susceptibility loci for CKD or hyperuricemia in the Japanese cohort.

Methods

Study subjects

A total of 5648 community-dwelling individuals were recruited from those who visited the Health Care Center of Inabe General Hospital (Inabe, Mie, Japan) for an annual health check-up from April 2003 to March 2014. All participants had each undergone 1–11 medical examinations, and the average follow-up period was 5 years. We refer to this cohort as the "discovery cohort." Methods for the collection and storage of medical examination data and genomic DNA samples have been described previously (Yamada et al. 2015). Cross-sectional data for CKD- and hyperuricemia-related traits in 7699 Japanese subjects (Gifu Prefectural Tajimi Hospital, Tajimi; Gifu Prefectural General Medical Center, Gifu; Japanese Red Cross Nagoya First Hospital, Nagoya; Hirosaki University Hospital and Hirosaki Stroke Center, Hirosaki, Japan) were used for replication studies of candidate SNVs identified from our longitudinal EWASs. We refer to this cohort as the "replication cohort".

Glomerular filtration rate was estimated with the use of the simplified prediction equation derived from the modified version of that described in the Modification of Diet in Renal Disease (MDRD) Study, as proposed by the Japanese Society of Nephrology (Zhang et al. 2017): eGFR (mL min^{-1} 1.73 m^{-2}) = 194 × [age (years)]$^{-0.287}$ × [serum creatinine (mg/dL)]$^{-1.094}$ × [0.739 if female]. The National Kidney Foundation-Kidney Disease Outcomes Quality Initiative guidelines recommend a diagnosis of CKD if eGFR is < 60 mL min^{-1} 1.73 m^{-2} (Go et al. 2004). We thus adopted the criterion of an eGFR of < 60 mL min^{-1} 1.73 m^{-2} (actual range 12.2 to 59.9 mL min^{-1} 1.73 m^{-2}) for the diagnosis of CKD. The control individuals for the EWAS of CKD had an eGFR of ≥ 90 mL min^{-1} 1.73 m^{-2} (actual range 90.1 to 595.5 mL min^{-1} 1.73 m^{-2}) and did not appear to have functional or structural abnormalities of the kidneys or a history of renal disease. Subjects with hemodialysis or peritoneal dialysis were excluded from the study.

Hyperuricemia was defined by the serum uric acid concentration of > 7.0 mg/dL (> 416 µmol/L, actual range, 416.4 to 814.9 µmol/L), as proposed by the Japanese Society of Gout and Nucleic Acid Metabolism (Japanese Society of Gout and Nucleic Acid Metabolism 2012), or the taking of uric acid-lowering medication. Individuals taking drugs that may cause secondary hyperuricemia were excluded. The control individuals for the EWAS of hyperuricemia had a serum uric acid concentration of ≤ 416 µmol/L (actual range, 29.7 to 410.4 µmol/L) and had no history of hyperuricemia, gout, or uric acid-lowering medication.

The distributions of eGFR and serum concentrations of creatinine and uric acid are shown in Fig. 1. In the longitudinal EWASs for eGFR or the serum concentration of creatinine, a total of 5636 subjects were examined (a total of 25,371 examinations). A total of 5487 subjects were examined in the EWAS for the serum concentration of uric acid (a total of 24,873 examinations). In the EWASs for the prevalence of CKD or hyperuricemia, 1646 individuals (606 subjects with CKD and 1040 controls; a total of 7583 examinations) or 5487 individuals (847 subjects with hyperuricemia and 4640 controls; a total of 24,873 examinations) were examined, respectively. The replication cohort involved 2517 subjects with CKD and 885 controls in the study for CKD, or 972 subjects with hyperuricemia and 4584 controls in the study for hyperuricemia.

Longitudinal EWAS

A longitudinal EWAS for the discovery cohort was performed based on ~ 244,000 genetic variants and longitudinal data of medical examinations. Infinium HumanExome-12 ver 1.2 BeadChip and Infinium Exome-24 ver 1.0 BeadChip (Illumina, San Diego, CA, USA) were used for genotyping subjects in the discovery cohort. These arrays include putative functional exonic variants selected from $> 12,000$ individual exome and whole-genome sequences across diverse ethnic populations (Grove et al. 2013). We performed quality controls and discarded the following SNVs: (1) monomorphic sites, (2) SNVs with MAF of < 0.05, (3) SNVs in which the genotype distribution significantly deviated from Hardy–Weinberg equilibrium ($P < 0.001$) in controls, and (4) SNVs located on mitochondrial DNA or sex chromosomes. Among Inabe subjects, four were identified as population outliers by principal component analysis (PCA) of SNVs, using the EIGENSTRAT method (Price et al. 2006) via JMP Genomics version 6.0 (SAS Institute, Cary, NC, USA) to detect population stratification on a genome-wide scale. The four outliers were removed from our longitudinal EWAS.

Next, we converted the genotyping data of the subjects into numeric data with dominant, additive, and recessive models, resulting in 24,579 SNVs. The dominant and recessive models were defined as "0, AA; 1, AB + BB" and
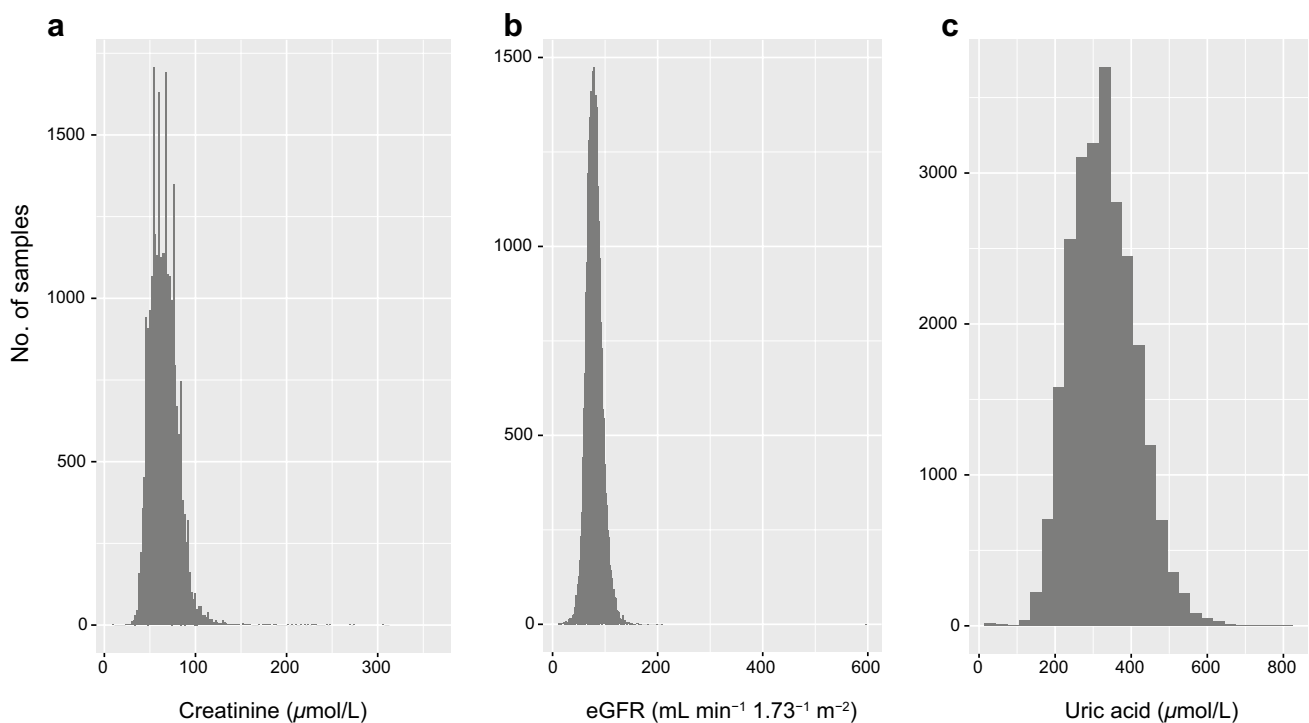
Fig. 1 Count distributions for longitudinal data of serum concentration of creatinine (**a**), eGFR (**b**), and serum concentration of uric acid (**c**) in the discovery cohort. *eGFR* estimated glomerular filtration rate

"0, AA + AB; 1, BB" (A, major allele; B, minor allele), respectively, whereas the additive model was defined as "0, AA; 1, AB; 2, BB". Quantile–quantile plots for the *P* values in the three genetic models are shown in Online Resources 1–3. The genomic inflation factor (λ) of *P* values ranged from 1.05 to 1.21.

Statistical analyses

To examine associations between SNVs and longitudinal changes in CKD-related traits (the prevalence of CKD, eGFR, and serum concentration of creatinine), we used the generalized estimating equation (GEE) model (Liang and Zeger 1986; Hanley et al. 2003) with adjustments for age, gender, and the prevalence of hypertension and type 2 diabetes mellitus using the R package 'geepack' (Halekoh et al. 2006). In longitudinal EWASs for hyperuricemia-related traits (the prevalence of hyperuricemia and serum concentration of uric acid), the association with genetic variants was tested using the GEE model with adjustments for age and gender. The wave argument was used to specify the ordering of repeated measurements within individuals. The statistical significance of the association was $P < 3.39 \times 10^{-7}$ (0.05/24,579 SNVs × 6) for each genetic model, after Bonferroni's correction to compensate for multiple comparisons of genotypes with the clinical parameters. Sitlani et al. (Sitlani et al. 2015) reported that a small effective sample size can increase the chances of generating type I errors. They recommended the use of 'approxdf', a scale of small effective sample size, and an approxdf of ≥ 10 could reduce type I errors. Therefore, we estimated the approxdf by the R package 'bosswithdf' (Voorman et al. 2012; Sitlani et al. 2015). To solve the issue of association with small effective sample sizes, we set to strict approxdf threshold, and discarded SNVs with approxdf of < 30.

The association of candidate SNVs identified from our longitudinal EWASs was tested in the replication cohort using the Chi-square test (for categorical data) or linear regression analysis (for quantitative data). In addition, we also examined relations of the candidate SNVs identified, using information on *P* values in datasets of four meta-analysis studies (Köttgen et al. 2010; Pattaro et al. 2016; Gorski et al. 2017; Li et al. 2017) from GKDGEN meta-analysis data which are publicly available (https://www.nhlbi.nih.gov/research/intramural/researchers/ckdgen).

Estimates of linkage disequilibrium

We estimated the LD among SNVs, using Haploview version 4.2 program (Barrett et al. 2005). In addition, we surveyed the LD between a candidate SNV detected from our longitudinal EWASs and other SNVs that were not included in the exome arrays used, employing LDlink web-based tools [https://analysistools.nci.nih.gov/LDlink/, (Machiela and Chanock 2015)]. The regulatory potential of each SNV was based on RegulomeDB scores [http://www.regulomedb.org/, (Boyle et al. 2012)]. Relations of candidate SNVs to disorders reported in previous studies were investigated using DisGeNET [http://www.disgenet.org/web/DisGeNET/, (Piñero et al. 2015)], GWAS Catalogue [https://www.ebi.ac.uk/gwas/, (MacArthur et al. 2017)], and GWAS Central [http://www.gwascentral.org/, (Beck et al. 2014)] databases.

Results

Characteristics of subjects in the discovery cohort

Characteristics of 5648 subjects in the discovery cohort, including 606 subjects with CKD and 1040 controls, and 847 subjects with hyperuricemia and 4640 controls, are shown in Table 1. The prevalence of both CKD and hyperuricemia was higher in males than in females; the proportion of males was 64.7% in subjects with CKD and 94.3% in subjects with hyperuricemia. The prevalence of hypertension, type 2 diabetes mellitus, and dyslipidemia was higher in patients with CKD or hyperuricemia than in corresponding controls. Weight, body mass index, waist circumference, systolic and diastolic blood pressures, fasting plasma glucose, blood hemoglobin A_{1c}, and serum concentrations of triglycerides, low-density lipoprotein cholesterol, creatinine, and uric acid, but not height or eGFR, were also greater in patients with CKD or hyperuricemia than in corresponding controls.

SNVs associated with CKD and hyperuricemia in longitudinal EWASs

The GEE model with adjustments for age, gender, and the prevalence of hypertension and type 2 diabetes mellitus in the dominant and additive models showed significant relations ($P < 3.39 \times 10^{-7}$) of rs11543349 in *OGFR* to serum creatinine concentration and eGFR (Table 2 and Online Resource 4). Approxdf values were 262 and 71 in the dominant and additive models, respectively. The association of rs11543349 with the serum concentration of creatinine showed the significance level with $P = 1.2 \times 10^{-8}$ and 2.7×10^{-8} in dominant and additive models, respectively. According to DisGeNET, GWAS Catalogue, and GWAS Central databases, the relation of this SNV to renal function has not been reported.

The GEE model with adjustments for age and gender showed that 16 SNVs were significantly ($P < 3.39 \times 10^{-7}$) associated with the prevalence of hyperuricemia or serum concentration of uric acid in the three genetic models, and their approxdf values were > 30 (Online Resource

Table 1 Characteristics of study subjects in the discovery cohort without dialysis patients or population outliers identified by the PCA method

Characteristic	Control[a]	CKD[a]	Control[a]	Hyperuricemia[a]
No. of subjects[b]	1040	606	4640	847
Sex (male/female, %)[b]	50.4/49.6	64.7/35.3	49.4/50.6	94.3/5.7
Age (years)	46.6±0.14 (5608)	64.4±0.21 (1975)	53.2±0.08 (20,775)	52.1±0.18 (4103)
Height (cm)	162.8±0.12 (5581)	161.2±0.21 (1923)	161.5±0.06 (20,686)	168.6±0.11 (4064)
Weight (kg)	60.7±0.17 (5581)	61.7±0.27 (1923)	59.6±0.08 (20,686)	69.9±0.19 (4064)
Body mass index (kg/m^2)	22.8±0.05 (5581)	23.6±0.07 (1923)	22.7±0.02 (20,686)	24.5±0.06 (4064)
Waist circumference (cm)	80.2±0.15 (4130)	83.1±0.24 (1446)	80.0±0.07 (16,140)	85.6±0.16 (3182)
Smoking (%)	41.4 (5608)	33.0 (1975)	35.2 (20,775)	65.8 (4103)
Hypertension (%)	23.3 (5608)	63.5 (1975)	31.4 (20,775)	48.2 (4103)
Systolic blood pressure (mmHg)	118.0±0.21 (5581)	126.2±0.38 (1921)	119.8±0.11 (20,684)	125.2±0.25 (4062)
Diastolic blood pressure (mmHg)	73.1±0.16 (5581)	77.0±0.28 (1921)	73.9±0.08 (20,684)	80.2±0.19 (4062)
Type 2 diabetes mellitus (%)	10.5 (5608)	21.5 (1975)	11.6 (20,775)	12.5 (4103)
Fasting plasma glucose (mmol/L)	5.62±0.018 (5601)	5.78±0.028 (1957)	5.57±0.008 (20,750)	5.74±0.015 (4094)
Blood hemoglobin A$_{1c}$ (%)	5.65±0.012 (4000)	5.88±0.015 (1711)	5.70±0.005 (16,245)	5.72±0.010 (2854)
Dyslipidemia (%)	49.4 (5608)	73.7 (1975)	57.5 (20,775)	74.7 (4103)
Serum triglycerides (mmol/L)	1.23±0.015 (5599)	1.40±0.020 (1967)	1.18±0.005 (20,760)	1.73±0.020 (4100)
Serum HDL-cholesterol (mmol/L)	1.62±0.006 (5598)	1.52±0.010 (1963)	1.65±0.003 (20,757)	1.44±0.006 (4098)
Serum LDL-cholesterol (mmol/L)	3.05±0.011 (5393)	3.17±0.018 (1854)	3.18±0.005 (20,040)	3.28±0.013 (3938)
Serum creatinine (μmol/L)	52.7±0.12 (5608)	92.2±0.58 (1975)	63.2±0.10 (20,701)	78.2±0.28 (4085)
eGFR (mL min^{-1} 1.73^{-1} m^{-2})	100.7±0.16 (5608)	52.9±0.17 (1975)	80.6±0.11 (20,701)	74.3±0.24 (4085)
Serum uric acid (μmol/L)	303.5±1.11 (5448)	374.7±1.99 (1926)	300.7±0.44 (20,775)	462.7±0.70 (4103)

CAD coronary artery disease, *MI* myocardial infarction, *CI* cerebral infarction, *HDL* high-density lipoprotein, *LDL* low-density lipoprotein, *eGFR* estimated glomerular filtration rate

[a] Values in parentheses indicate the numbers of measurements taken. Quantitative data are means and standard error

[b] The numbers are based on data examined in the latest year

4). Of these SNVs, nine are located at seven gene loci (*DDX39B, NFKBIL1, CDC42BPG, CDC63, BRAP, ACAD10*, and *PCNX3*) that might be novel susceptibility loci for hyperuricemia. However, five candidate SNVs at 12q24.1 exhibited moderate or strong LD ($0.43 \leq r^2 \leq 0.99$) with rs671 of *ALDH2* that has been reported to be associated with renal function-related traits (Levey et al. 2005; Sakiyama et al. 2016, 2017) (Online Resource 5). Thus, we did not consider the candidate SNVs at 12q24.1 as novel susceptibility loci for hyperuricemia. It has been reported that rs505802 of *SLC22A12* and rs504915 of *NRXN2* are related to the serum concentration of uric acid (Kolz et al. 2009; Okada et al. 2012). These SNVs are located relatively close (~ 240 kb) to rs55975541 of *CDC42BPG* at 11q13.1. However, these SNVs were not in LD (rs505802, $r^2 = 0.06$, $D' = 0.26$; rs504915, $r^2 = 0.10$, $D' = 0.39$) (Fig. 2). Therefore, it is possible that rs55975541 of *CDC42BPG* independently affects the serum concentration of uric acid. Consequently, four genes (*DDX39B, NFKBIL1, CDC42BPG*, and *PCNX3*) were identified as candidates for novel susceptibility loci that might be related to the serum concentration of uric acid. Of these

candidates, the association of rs55975541 in *CDC42BPG* showed the significance level with $P = 3.7 \times 10^{-12}$ in the dominant model (Online Resource 4).

Replication studies for candidate SNVs related to CKD or hyperuricemia

We examined the relation of 17 candidate SNVs identified from our longitudinal EWASs to CKD or hyperuricemia using the Chi-square test or linear regression analysis, using cross-sectional data for CKD and hyperuricemia-related traits in 7699 Japanese subjects of the replication cohort (Online Resource 6). In this replication study, rs55975541 in *CDC42BPG* was significantly [$P < 4.90 \times 10^{-4}$ (0.05/17 SNVs × 6)] related to the serum concentration of uric acid. We also assessed relations of the candidate SNVs detected in the discovery cohort with CKD or hyperuricemia, using GKDGEN meta-analysis data on European or African ancestry populations (Online Resource 7). The in silico replication study did not show significant association with the five candidate SNVs that were newly identified in the discovery cohort. However, four SNVs (rs2049330, rs2257212,

Table 2 Newly identified SNVs associated with CKD, hyperuricemia, serum concentrations of creatinine or uric acid, or eGFR in the discovery cohort

Trait/disease	RefSNP ID	Location[a]	Genotype	CKD[b]	Controls[b]	Hyperuricemia[b]	Controls[b]	Creatinine	eGFR	Uric acid
Creatinine	rs11543349	20: 62,813,587	GG	691 (35.0%)	2549 (45.5%)	1658 (40.4%)	8668 (41.7%)	75.8 ± 0.97	79.6 ± 0.17	327.1 ± 0.85
	OGFR		GC	969 (49.1%)	2476 (44.2%)	1932 (47.1%)	9543 (45.9%)	77.5 ± 0.92	78.1 ± 0.16	330.5 ± 0.80
			CC	315 (15.9%)	583 (10.4%)	513 (12.5%)	2564 (12.3%)	81.7 ± 2.20	76.6 ± 0.31	331.3 ± 1.56
eGFR	rs2239709	6: 31,539,670	CC	1303 (66.0%)	3474 (61.9%)	2665 (65.0%)	13,082 (63.0%)	76.3 ± 0.76	78.4 ± 0.14	331.6 ± 0.68
	DDX39B		CT	621 (31.4%)	1905 (34.0%)	1326 (32.3%)	6875 (33.1%)	80.5 ± 1.30	78.6 ± 0.20	327.1 ± 0.97
			TT	51 (2.6%)	229 (4.1%)	112 (2.7%)	818 (3.9%)	66.2 ± 1.59	80.2 ± 0.48	306.2 ± 2.86
Uric acid	rs2071593	6: 31,545,022	GG	1306 (66.1%)	3482 (62.1%)	2674 (65.2%)	13,104 (63.1%)	76.3 ± 0.75	78.4 ± 0.14	331.7 ± 0.68
	NFKBIL1		GA	618 (31.3%)	1897 (33.8%)	1319 (32.1%)	6853 (33.0%)	80.5 ± 1.31	78.6 ± 0.20	327.0 ± 0.97
			AA	51 (2.6%)	229 (4.1%)	110 (2.7%)	818 (3.9%)	66.1 ± 1.59	80.2 ± 0.48	305.8 ± 2.85
	rs55975541	11: 64,829,729	GG	1375 (69.6%)	3934 (70.1%)	2956 (72.0%)	14,152 (68.1%)	74.8 ± 0.66	78.8 ± 0.13	333.6 ± 0.65
	CDC42BPG		GA	553 (28.0%)	1551 (27.7%)	1058 (25.8%)	6048 (29.1%)	82.5 ± 1.53	78.1 ± 0.21	320.9 ± 1.04
			AA	47 (2.4%)	123 (2.2%)	89 (2.2%)	575 (2.8%)	86.8 ± 4.85	76.0 ± 0.68	305.6 ± 3.90
	rs12801636	11: 65,623,846	AA	509 (25.8%)	1405 (25.1%)	997 (24.3%)	5297 (25.5%)	79.4 ± 1.41	78.2 ± 0.23	324.1 ± 1.10
	PCNX3		AG	979 (49.6%)	2846 (50.7%)	1967 (47.9%)	10,631 (51.2%)	78.7 ± 0.95	78.5 ± 0.16	328.3 ± 0.76
			GG	487 (24.7%)	1357 (24.2%)	1139 (27.8%)	4847 (23.3%)	72.1 ± 0.99	79.0 ± 0.21	336.4 ± 1.14

CKD chronic kidney disease, *eGFR* estimated glomerular filtration rate, *Std.err* standard error, *MAF* minor allele frequency

[a] Location in NCBI build GRC38

[b] Values indicate the numbers of measurements taken, with the percentages in parentheses

Fig. 2 Linkage disequilibrium (LD) of 12 biallelic sites across approximately 475 kb genomic region at 11q13.1 using single nucleotide variations (SNVs) data used in the present study. The diagram was depicted by the Haploview version 4.2. The SNVs with the minor allele frequency of <0.01 were removed from the analysis. The hyperuricemia-associated SNVs (rs505802 of *SLC22A12* and rs55975541 of *CDC42BPG*) identified by the generalized estimating equation model are shown in bold
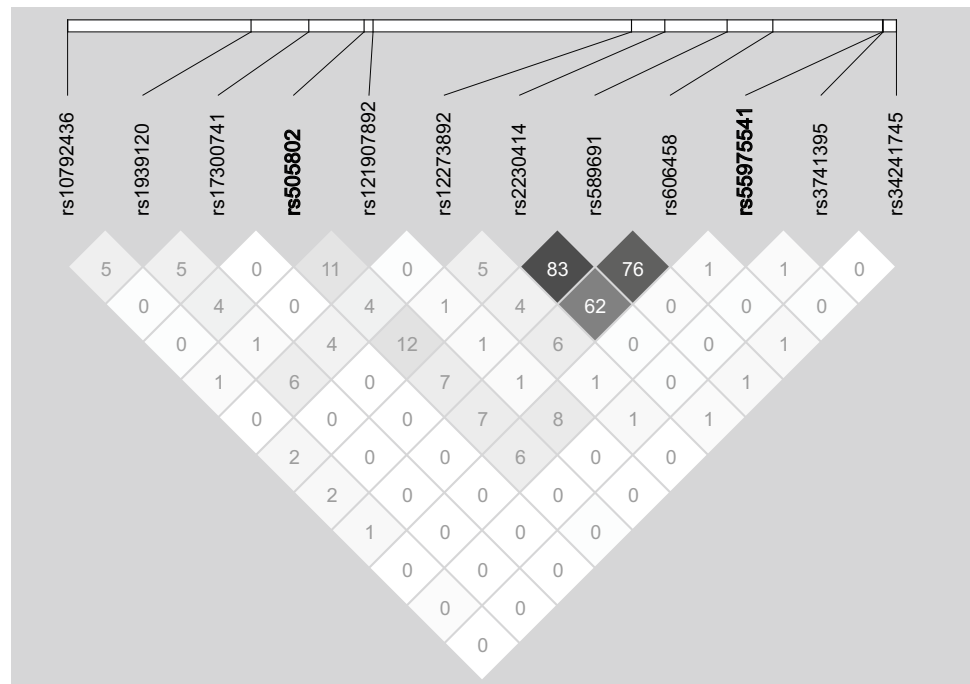
rs1143671, and rs1143672) in *SLC15A2* were significantly ($P < 4.90 \times 10^{-4}$) associated with eGFR in European ancestry populations, although these SNVs were related to the serum concentration of uric acid with borderline significance ($P = 5.50 \times 10^{-7}$ to 6.40×10^{-7}) in our longitudinal EWASs.

Survey of LD with disease-associated SNVs that were not examined in the longitudinal EWAS

We further investigated whether the allele frequency of the replicated target SNV in *CDC42BPG* was affected by other previously reported SNVs that were not included in the exome arrays used and which exhibited strong LD. LDproxy, an LDlink application, indicated rs55975541 in *CDC42BPG* was in LD ($r^2 \geq 0.8$, $P < 0.0001$) with rs80094634 only in JPT (Japanese in Tokyo, Japan) from the 1000 Genomes Project database [http://www.internationalgenome.org, (The 1000 Genomes Project Consortium 2010)] (Online Resource 8). The nucleotide substitution at the non-genotyped SNV was silent (intronic variant), whereas the variant at the candidate SNV of *CDC42BPG* can alter an amino acid (nonsynonymous substitution). According to the Ensembl human database (http://www.ensembl.org/Homo_sapiens/), it is possible that the nonsynonymous substitution has an effect on the protein function because the amino acid substitution was predicted to be "probably damaging" with scores of zero by SIFT prediction and 0.998 by PolyPhen analysis. Therefore, rs55975541 is more likely to be related to hyperuricemia than, the non-genotyped SNV, rs80094634. The LDproxy analysis also showed that rs55975541 in

CDC42BPG was not in LD with any genetic variants in the established *SLC22A11, SLC22A12*, and *NRXN2* loci (Online Resource 8).

Relation of previously reported SNVs to renal function-related traits in longitudinal EWASs for the discovery cohort

We comprehensively assessed significance levels of 291 SNVs at 55 previously known CKD- or hyperuricemia-associated loci (Online Resource 9). Of the SNVs examined, 103 at 37 loci, including *ABCG2, GCKR*, and *UMOD*, were significantly ($P < 3.39 \times 10^{-7}$) associated with renal function-related traits in any of the three genetic models. In the discovery cohort, MAFs of some previously reported SNVs were less than 5%. For instance, the SNV rs121907892 of *SLC22A12* was significantly associated with the prevalence of hyperuricemia and serum concentration of uric acid in all the genetic models ($P < 2.0 \times 10^{-16}$). However, rs121907892 was excluded from our longitudinal EWAS because of low minor allele frequency (MAF < 0.03). The differences in significance levels between the present and previous studies might be due to sample sizes, statistical methods, or differences in genetic background among the populations examined.

Discussion

In the present study, 17 SNVs showed significant relations with CKD- or hyperuricemia-related traits. Five of these

SNVs have not been reported in previous studies. However, the relation of these SNVs to CKD or hyperuricemia could not be found in the replication study, except for rs55975541 in *CDC42BPG*. In the longitudinal EWAS for the discovery cohort, the SNV in *CDC42BPG* was significantly ($P = 3.7 \times 10^{-12}$) associated with the serum concentration of uric acid. The association of rs55975541 was determined using cross-sectional data for the serum concentration of uric acid of 7699 Japanese subjects in the replication cohort. The LDproxy analysis suggests that this SNV independently affects the incidence of hyperuricemia.

MRCKγ encoded by *CDC42BPG* is ubiquitously expressed in a variety of tissues and organs including the kidney, according to The Human Protein Atlas database (http://www.proteinatlas.org/). This protein may act as a downstream effector of CDC42 in the regulation of cytoskeletal reorganization (Ng et al. 2004). The kidney-specific Cdc42 knockout mice exhibited early postnatal death due to renal failure (Choi et al. 2013). The nucleotide substitution at rs55975541 in *CDC42BPG* alters an amino acid residue at position 1237 (Arg1237Trp). The amino acid change may have an effect on the protein function, as predicted by SIFT and PolyPhen. Although *CDC42BPG* may be a susceptibility locus for renal function, the functional relevance of the candidate SNV to the serum concentration of uric acid remains unclear.

Despite many studies of associations between several *SLC* genes and hyperuricemia-related traits in diverse ethnic groups (Matsuo et al. 2008; Köttgen et al. 2013; Phipps-Green et al. 2016; Nakayama et al. 2017), association with *SLC15A2* has not been reported, according to the public databases. It is noteworthy that three nonsynonymous substitutions (at rs2257212, rs1143671, and rs1143672) in *SLC15A2* were significantly associated with eGFR in European ancestry populations from GKDGEN meta-analysis data, whereas these SNVs were related to the serum concentration of uric acid with a borderline significance in our longitudinal EWASs.

There were some limitations to the present study. First, the longitudinal EWAS was conducted in a local Japanese population only. Although we examined relations of candidate SNVs to CKD- or hyperuricemia-related traits in the replication studies, these studies were conducted in a cross-sectional manner. Therefore, replication of longitudinal EWASs in other Japanese populations or other ethnic groups is required to verify the relations of the identified SNVs to the relevant diseases. Second, the functional relevance of the candidate SNV to the pathogenesis of the diseases examined remains unclear. Therefore, further functional analysis is required to clarify the results of this study.

In conclusion, our findings indicated that rs55975541 in *CDC42BPG* was significantly associated with the serum concentration of uric acid in our EWASs. Genotyping for the SNV detected in our longitudinal EWAS may be informative for assessment of genetic risk for hyperuricemia.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study protocol was approved by the Committees on the Ethics of Human Research of Mie University Graduate School of Medicine and Inabe General Hospital.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. https://doi.org/10.1093/bioinformatics/bth457

Beck T, Hastings RK, Gollapudi S et al (2014) GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet 22:949–952. https://doi.org/10.1038/ejhg.2013.274

Böger CA, Gorski M, Li M et al (2011) Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. PLoS Genet 7:e1002292. https://doi.org/10.1371/journal.pgen.1002292

Boyle AP, Hong EL, Hariharan M et al (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22:1790–1797. https://doi.org/10.1101/gr.137323.112

Chambers JC, Zhang W, Lord GM et al (2010) Genetic loci influencing kidney function and chronic kidney disease. Nat Genet 42:373–375. https://doi.org/10.1038/ng.566

Choi SY, Chacon-Heszele MF, Huang L et al (2013) Cdc42 deficiency causes ciliary abnormalities and cystic kidneys. J Am Soc Nephrol 24:1435–1450. https://doi.org/10.1681/ASN.2012121236

Go AS, Chertow GM, Fan D et al (2004) Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 351:1296–1305. https://doi.org/10.1056/NEJMoa041031

Gorski M, van der Most PJ, Teumer A et al (2017) 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. Sci Rep 7:45040. https://doi.org/10.1038/srep45040

Grove ML, Yu B, Cochran BJ et al (2013) Best practices and joint calling of the humanexome beadchip: the charge consortium. PLoS One 8:e68095. https://doi.org/10.1371/journal.pone.0068095

Halekoh U, Højsgaard S, Yan J (2006) The R package geepack for generalized estimating equations. J Stat Softw 15:1–11

Hanley JA, Negassa A, Edwardes MD, deB Forrester JE (2003) Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 157:364–375. https://doi.org/10.1093/aje/kwf215

Hill NR, Fatoba ST, Oke JL et al (2016) Global prevalence of chronic kidney disease—a systematic review and meta-analysis. PLoS One 11:e0158765. https://doi.org/10.1371/journal.pone.0158765

Japanese Society of Gout and Nucleic Acid Metabolism (2012) Revised Guideline for the Management of Hyperuricemia and Goat. Medical Review Co., Ltd, Osaka (in Japanese)

Kolz M, Johnson T, Sanna S et al (2009) Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS Genet 5:e1000504. https://doi.org/10.1371/journal.pgen.1000504

Köttgen A, Glazer NL, Dehghan A et al (2009) Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet 41:712–717. https://doi.org/10.1038/ng.377

Köttgen A, Pattaro C, Böger CA et al (2010) New loci associated with kidney function and chronic kidney disease. Nat Genet 42:376–384. https://doi.org/10.1038/ng.568

Köttgen A, Albrecht E, Teumer A et al (2013) Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 45:145–154. https://doi.org/10.1038/ng.2500

Levey AS, Eckardt K-U, Tsukamoto Y et al (2005) Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). Kidney Int 67:2089–2100. https://doi.org/10.1111/j.1523-1755.2005.00365.x

Li M, Li Y, Weeks O et al (2017) SOS2 and ACP1 loci identified through large-scale exome chip analysis regulate kidney development and function. J Am Soc Nephrol 28:981–994. https://doi.org/10.1681/ASN.2016020131

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22. https://doi.org/10.1093/biomet/73.1.13

Lioté F (2003) Hyperuricemia and gout. Curr Rheumatol Rep 5:227–234. https://doi.org/10.1007/s11926-003-0072-y

MacArthur J, Bowler E, Cerezo M et al (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45:D896–D901. https://doi.org/10.1093/nar/gkw1133

Machiela MJ, Chanock SJ (2015) LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31:3555–3557. https://doi.org/10.1093/bioinformatics/btv402

Matsuo H, Chiba T, Nagamori S et al (2008) Mutations in glucose transporter 9 gene SLC2A9 cause renal hypouricemia. Am J Hum Genet 83:744–751. https://doi.org/10.1016/j.ajhg.2008.11.001

Nakayama A, Nakaoka H, Yamamoto K et al (2017) GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include urate transporter genes. Ann Rheum Dis 76:869–877. https://doi.org/10.1136/annrheumdis-2016-209632

Ng Y, Tan I, Lim L, Leung T (2004) Expression of the human myotonic dystrophy kinase-related Cdc42-binding kinase γ Is regulated by promoter DNA methylation and Sp1 binding. J Biol Chem 279:34156–34164. https://doi.org/10.1074/jbc.M405252200

Okada Y, Sim X, Go MJ et al (2012) Meta-analysis identifies multiple loci associated with kidney function–related traits in east Asian populations. Nat Genet 44:904–909. https://doi.org/10.1038/ng.2352

Pattaro C, Teumer A, Gorski M et al (2016) Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun 7:10023. https://doi.org/10.1038/ncomms10023

Phipps-Green AJ, Merriman ME, Topless R et al (2016) Twenty-eight loci that influence serum urate levels: analysis of association with gout. Ann Rheum Dis 75:124–130. https://doi.org/10.1136/annrheumdis-2014-205877

Piñero J, Queralt-Rosinach N, Bravo À et al (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford) 2015:bav028. https://doi.org/10.1093/database/bav028

Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. https://doi.org/10.1038/ng1847

Sakiyama M, Matsuo H, Nakaoka H et al (2016) Identification of rs671, a common variant of ALDH2, as a gout susceptibility locus. Sci Rep 6:25360. https://doi.org/10.1038/srep25360

Sakiyama M, Matsuo H, Akashi A et al (2017) Independent effects of ADH1B and ALDH2 common dysfunctional variants on gout risk. Sci Rep 7:2500. https://doi.org/10.1038/s41598-017-02528-z

Schiffrin EL, Lipman ML, Mann JFE (2007) Chronic kidney disease: effects on the cardiovascular system. Circulation 116:85–97. https://doi.org/10.1161/CIRCULATIONAHA.106.678342

Sitlani CM, Rice KM, Lumley T et al (2015) Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. Stat Med 34:118–130. https://doi.org/10.1002/sim.6323

Sveinbjornsson G, Mikaelsdottir E, Palsson R et al (2014) Rare mutations associating with serum creatinine and chronic kidney disease. Hum Mol Genet 23:6935–6943. https://doi.org/10.1093/hmg/ddu399

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073. https://doi.org/10.1038/nature09534

Voorman A, Rice K, Lumley T (2012) Fast computation for genome-wide association studies using boosted one-step statistics. Bioinformatics 28:1818–1822. https://doi.org/10.1093/bioinformatics/bts291

Webster AC, Nagler EV, Morton RL, Masson P (2017) Chronic kidney disease. The Lancet 389:1238–1252. https://doi.org/10.1016/S0140-6736(16)32064-5

Yamada Y, Matsui K, Takeuchi I et al (2015) Association of genetic variants with hypertension in a longitudinal population-based genetic epidemiological study. Int J Mol Med 35:1189–1198. https://doi.org/10.3892/ijmm.2015.2151

Yang Q, Köttgen A, Dehghan A et al (2010) Multiple genetic loci influence serum urate and their relationship with gout and cardiovascular disease risk factors. Circ Cardiovasc Genet 3:523–530. https://doi.org/10.1161/CIRCGENETICS.109.934455

Zhang C, Lu Y, Feng Q et al (2017) Differentiated demographic histories and local adaptations between Sherpas and Tibetans. Genome Biol 18:115. https://doi.org/10.1186/s13059-017-1242-y