



Pathway analysis of genome-wide association studies on uric acid concentrations

Young Ho Lee*, Gwan Gyu Song

Division of Rheumatology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, 126-1, Anam-dong 5-ga, Seongbuk-gu, Seoul 136-705, Republic of Korea

ARTICLE INFO

Article history:

Received 16 February 2012

Accepted 8 May 2012

Available online 17 May 2012

ABSTRACT

Objective: The aims of this study were to identify the candidate causal single nucleotide polymorphisms (SNPs) and candidate causal mechanisms influencing uric acid level and to generate hypotheses for the SNP to gene to pathways that influence uric acid concentrations.

Methods: Meta-analysis data of 954 SNPs with genome-wide significance in 14 genome-wide association studies (GWASs) comprising 28,141 individuals of European ancestry was subjected to ICSNPPathway (Identify candidate Causal SNPs and Pathways) analysis to establish associations between pathways and uric acid concentrations.

Results: ICSNPPathway analysis identified 14 candidate causal SNPs, five genes, and two candidate causal pathways, which provided two hypothetical biologic mechanisms: (1) rs2728121 (regulatory region) to polycystic kidney disease 2 (PKD2) to ion transmembrane transporter activity; (2) rs942377, rs3799346, rs3799344, rs2762353, rs13197601, rs3757131, rs1165215, rs1165196 to SLC17A1 to ion transmembrane transporter activity and secondary active transmembrane transporter activity. SLC17A1, SLC17A3, SLC17A4, SLC22A11, and SLC2A9 were involved in both pathways, and PKD2 and SLC16A9 in one pathway.

Conclusion: By applying ICSNPPathway analysis to GWAS data on uric acid levels, 14 SNPs, five genes, and two pathways involving the PKD2, SLC17A1, SLC17A3, SLC17A4, and SLC2A9 genes were identified that might contribute to the condition in Europeans.

© 2012 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

1. Introduction

Gout is a common inflammatory arthritis that results from the deposition of monosodium urate crystals in joints, and uric acid is the final catabolic purine derivative in man. In most mammals, uric acid is degraded by uricase to allantoin, which is highly soluble and readily excreted in urine [1]. However, due to a loss of hepatic uricase activity in man, uric acid is not metabolized further and is excreted unmodified. The main cause of gout is excess urate or hyperuricemia. Serum urate levels are highly heritable and hyperuricemia has been shown to have a genetic component [2].

Genome-wide association studies (GWASs) provide a powerful means of identifying genes that confer susceptibility to complex diseases, and the large number of GWASs reported have resulted in the discovery and replication of new disease genes in general [3]. Furthermore, the availability of large numbers of GWAS datasets offers powerful research opportunities. However, despite the large-scale GWASs conducted to identify genes associated with complex diseases, much of the genetic contribution to changes in uric acid levels remain unexplained.

Genes and genetic variants make small contributions to the overall risks of complex diseases by interacting with each other. One of the key challenges posed by GWAS data interpretation is the identification of causal single nucleotide polymorphisms (SNPs) and the provision of evidence and hypotheses regarding the mechanism through which they affect the trait being investigated [4]. Furthermore, the use of new methods to study existing GWAS data may provide additional biological insights and highlight new candidate genes [5]. ICSNPPathway (Identify candidate Causal SNPs and Pathways) was developed to identify candidate causal SNPs and their corresponding candidate causal pathways from GWAS datasets by integrating linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) [6].

In the present study, we applied ICSNPPathway analysis to the meta-analysis data of 14 GWAS datasets of uric acid concentrations to identify candidate causal SNPs and mechanisms, and to generate SNP to gene to pathway hypotheses.

2. Materials and methods

2.1. Study populations

We used the data generated by GWAS meta-analysis of uric acid concentrations [7]. The study participants were as follows: The MRC

* Corresponding author. Fax: +82 2 922 5974.

E-mail address: lyhghg@korea.ac.kr (Y.H. Lee).

BRIGHT study comprises 2500 hypertensive participants and 2000 normotensive controls of white European ancestry. The Cohorte Lausannoise (CoLaus) is a population-based study aimed at assessing the prevalence and molecular determinants of cardiovascular risk factors in the Caucasian population. Participants in the study (6188) were randomly selected from the population register of Lausanne in 2003 ($n = 56,694$). The CROATIA study includes 1031 unselected adult participants who were recruited in a population-based study during 2003 and 2004 in the villages of Vis and Komiza on the Dalmatian island of Vis. The population-based Health 2000 Survey was carried out in Finland between August 2000 and July 2001. Of the 8028 persons 6986 (87%) were interviewed in their home or in an institution, and 6354 (79%) participated a comprehensive health examination. GWA analyzed here was performed on a case control setting on metabolic syndrome ($n = 2212$). The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany (KORA F4, KORA S4). The MICROS (Micro-isolates in South Tyrol) study is an extensive survey carried out in the period 2001–2003 in South Tyrol, Italy. The 1086 participants presented in this study are those for whom both phenotypic and GWA data were available. The Swedish samples are part of the Northern Swedish Population Health Study (NSPHS) representing a family-based population study. DNA samples ($N = 700$) were genotyped. The Orkney Complex Disease Study (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Fast-ing blood samples were collected and over 200 health-related phenotypes and environmental exposures were measured in each individual. PROCARDIS is a European collaborative project studying genetic susceptibility to coronary artery disease. Uric acid measurements were available for 1203 unrelated survivors of myocardial infarction with European ancestry. The SardiNIA study is a population-based study, that recruited and phenotyped 6148 individuals, male and female, ages 14–102 years, from the Lanusei Valley of Sardinia. The Study of Health in Pomerania (SHIP) is a cross-sectional survey in West Pomerania, the north-east area of Germany. Seven thousand and eight subjects were sampled, with 292 persons of each gender in each of the 12 five-year age strata. SSAGA cohort is based on Australian twin pairs born before 1964 who participated in questionnaire- and interview-based studies on psychological and metabolic risk factors for psychiatric and cardiovascular disease. Seven thousand seven hundred and sixty four people were interviewed with the Semi-Structured Assessment for Genetics of Alcoholism instrument in 1992–1993 (SSAGA Study), and 3375 of them gave blood samples for DNA preparation and serum biochemistry in 1993–1996 (SSAGA Blood Study). The TwinsUK cohort is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population. A total of 2113 females with acid uric phenotype were included in the analysis. The meta-analysis combined data from 14 GWASs, which in total involved 28,141 individuals of European ancestry. Meta-analysis yielded 954 SNPs that exceeded a genome-wide significance threshold of 5×10^{-8} . We used the full list of 954 SNPs for the genome-wide significance threshold.

2.2. ICSNPathway using GWAS data

We applied ICSNPathway analysis to the meta-analysis data of 14 GWAS on uric acid concentrations. ICSNPathway analysis involves two stages [6]. The first stage involves the pre-selection of candidate causal SNPs by LD analysis and functional SNP annotation based on most significant SNPs, and the second stage involves the annotation of biological mechanisms to pre-selected candidate causal SNPs using a PBA algorithm named *i*-GSEA (improved-gene set enrichment analysis). A full list of 954 GWAS SNP *p*-values was entered into ICSNPathway analysis. One concept is in LD analysis, which searches

the SNPs in LD with the most significant SNPs of GWAS to ensure to capture more possible candidate causal SNPs based on the extended data set which includes HapMap data. The other concept is functional SNPs. ICSNPathway pre-selects candidate causal SNPs based on functional SNPs, which are important for understanding the underlying genetics of human health. Functional SNPs are defined as SNPs that may alter protein, gene expression or the role of protein in context of pathway. The functional SNPs include deleterious and non-deleterious non-synonymous SNPs, SNPs leading in gain or loss of stop codon, SNPs resulting in frame shift, SNPs in essential splice site and SNPs in regulatory region. ICSNPathway server implements a PBA algorithm, as named improved-gene set enrichment analysis (*i*-GSEA), on the full list of GWAS SNP *p*-values to detect pathways associated with traits. Briefly, (1) Each SNP is mapped to its nearest gene according to the SNP and gene localization in Ensembl 61 database (<http://www.wnseembl.org/biomart/martview>), and the maximum $t = -\log(p\text{-value})$ of the SNPs mapped to a gene is assigned to represent the gene. Then all the genes are raked by decreasing their representation value t . (2) For each pathway S , ES (enrichment score, i.e. a Kolmogor–Smirnov like running-sum statistics with weight (a)) which measures the tendency that genes of a pathway are located at the top of the ranked gene list, is calculated. (3) ES is converted to $SPES$ (significant proportion based enrichment score) by multiplying it to m_1/m_2 , where m_1 is the proportion of significant genes (defined as genes mapped with at least one of the top 5% most significant SNPs of all SNPs in GWAS) for pathways S and m_2 is the proportion of significant genes for all the genes in the GWAS. (4) SNP label permutation and normalization are employed to generate the distribution of $SPES$ and to correct gene variation (the bias due to different genes with different number of mapped SNPs). (5) Based on all the distribution of $SPES$ s generated by permutation, nominal *p*-value is calculated and false discovery rate (*FDR*) is computed for multiple testing correction ‘Within gene’ was selected as a parameter for ICSNPathway analysis, that is, only *p*-values of SNPs located within genes were utilized in PBA. To avoid stochastic bias and the testing of biological processes that are too general, pathways containing <5 or >100 genes were ignored. Although several options are available for pathway annotation, we chose four databases of pathways, namely, the Kyoto Encyclopedia of Genes and Genomes (KEGG) [8], BioCarta, Gene Ontology (GO) biological process [9], and GO molecular function, to ensure comprehensive coverage of pathways and high-quality information for well-defined pathways. The extensive nature of LD can confound the interpretation of an association signal as the true causal variant(s) can lie at considerable distance from the initial association signal. HapMap informs about genomic locations, alleles, and LD patterns for a large fraction of common variants in the human genome. When a candidate SNP was not present on a particular genotyping array, proxy SNPs in LD with that candidate SNP can be identified based on observed LD patterns in HapMap. SNAP was developed to identify and annotate nearby SNPs in LD (proxies) on HapMap. SNAP can (i) find proxy SNPs, (ii) determine if SNP proxies are in genes, (iii) resolve whether associations from multiple SNPs represent a similar association, (iv) plot regional views of associations or LD structure, and (v) retrieve annotations for SNPs. Thus, SNAP, a tool for identifying and annotating proxy SNPs using HapMap, was used to find proxy SNPs, plot regional views of LD structure, and retrieve annotations for SNPs [10].

3. Results

3.1. Candidate causal SNPs and pathways from the meta-analysis data of GWASs of uric acid concentrations

Utilizing the 954 GWAS SNPs *p*-values as input, ICSNPathway analysis identified 14 candidate causal SNPs, five genes, and two can-

Table 1

Candidate causal SNPs of uric acid concentrations.

Candidate causal SNP	Functional class	Gene	Candidate causal pathway ^a	$-\log_{10}(p)^b$	In LD with	r^2	D'	$-\log_{10}(p)^c$
rs2728121	Regulatory region	PKD2	1	9.623	rs2728121	–	–	9.623
rs942377	Regulatory region	SLC17A1	1 2	–	rs1165161	0.849	0.946	9.963
rs3799346	Regulatory region	SLC17A1	1 2	7.418	rs3799346	–	–	7.418
rs3799344	Regulatory region	SLC17A1	1 2	11.452	rs3799344	–	–	11.452
rs2762353	Regulatory region	SLC17A1	1 2	12.924	rs2762353	–	–	12.924
rs13197601	Regulatory region	SLC17A1	1 2	10.955	rs13197601	–	–	10.955
rs3757131	Regulatory region	SLC17A1	1 2	11.435	rs3757131	–	–	11.435
rs1165215	Regulatory region	SLC17A1	1 2	13.148	rs1165215	–	–	13.148
rs1165196	Non-synonymous coding (deleterious)	SLC17A1	1 2	12.600	rs1165196	–	–	12.600
rs1165165	Non-synonymous coding	SLC17A3	1 2	8.243	rs1165165	–	–	8.243
rs11754288	Non-synonymous coding	SLC17A4	1 2	11.359	rs11754288	–	–	11.359
rs16890979	Non-synonymous coding	SLC2A9	1 2	188.450	rs16890979	–	–	188.450
rs2280205	Non-synonymous coding	SLC2A9	1 2	22.294	rs2280205	–	–	22.294
rs13137343	Regulatory region	SLC2A9	1 2	66.833	rs13137343	–	–	66.833

SNP: single nucleotide polymorphism, LD: linkage disequilibrium, GWAS: genome-wide association study.

^a The number indicates the index of pathways, which are ranked by their statistical significance (false discovery rate), 1. Ion transmembrane transporter activity, 2. Secondary active transmembrane transporter activity.^b $-\log_{10}(p)$ for candidate causal SNP in original GWAS. '–' denotes that this SNP is not represented in the original GWAS.^c $-\log_{10}(p)$ for the SNP (which candidate causal SNP is in LD with) in original GWAS.**Table 2**

Candidate causal pathways of uric acid concentrations.

Index	Candidate causal pathway	Description	Nominal p	FDR
1	Ion transmembrane transporter activity	GO: 0015075	<0.001	<0.001
2	Secondary active transmembrane transporter activity	GO: 0015291	<0.001	<0.001

FDR: false discovery rate, GO: gene ontology.

didate causal pathways (Tables 1 and 2, Fig. 1). Rs2728121, which was represented in the original GWAS ($-\log_{10}(p) = 9.623$), is not in LD with any SNP. SNP rs942377, which was not represented in the original GWAS, is in LD with rs1165161 ($r^2 = 0.849$). However, rs3799346, rs3799344, rs2762353, rs13197601, rs3757131, rs1165215, rs1165196, rs1165165, rs11754288, rs16890979, rs2280205, and rs13137343, which were represented in the original GWAS, are not in LD with any SNP.

These 14 candidate causal SNPs and two candidate causal pathways provide five hypothetical biological mechanisms of elevated uric acid concentrations: (1) rs2728121 (regulatory region) to polycystic kidney disease 2 (PKD2) to ion transmembrane transporter activity; (2) rs942377, rs3799346, rs3799344, rs2762353, rs13197601, rs3757131, rs1165215, rs1165196 to SLC17A1 to ion transmembrane transporter activity and secondary active transmembrane transporter activity; (3) rs1165165 to SLC17A3 to ion transmembrane transporter activity and secondary active transmembrane transporter activity; (4) rs11754288 to SLC17A4 to ion transmembrane transporter activity and secondary active transmembrane transporter activity; and (5) rs16890979, rs2280205, and rs13137343 to SLC2A9 to ion transmembrane transporter activity and secondary active transmembrane transporter activity.

We then investigated which genes play a role in the pathways revealed by PBA. A distinct clustering of genes involved in two candidate causal pathways was observed. Further examination of the gene contents of these two pathways revealed some overlap, that is, SLC17A1, SLC17A3, SLC17A4, SLC22A11, and SLC2A9 were involved in both pathways. On the other hand, PKD2, SLC16A9, PKD2, and SLC16A9 were involved in one pathway (Table 3). The genes without an obvious known immunological function are of interest and may lead to novel mechanisms. This analysis suggests roles for the pathways and genes listed in Tables 2 and 3.

4. Discussion

GWASs have been successfully used to identify novel common genetic variants that contribute to susceptibility to complex diseases [11]. However, individual GWASs have limited power to identify new loci, because a limited set of variants are genotyped, and the reported variant is unlikely to be the causal variant, but rather in LD with relevant variants [12]. Furthermore, multiple related genes may work together to confer disease susceptibility. Thus, PBA is required to identify further new loci that confer susceptibility to complex diseases [13,14]. It is well known that uric acid concentrations are caused by interactions between multiple genetic factors and environmental factors, and that a complex molecular network and cellular pathways play key roles in determining uric acid levels. If a specific pathway participates in the determination of uric acid concentration, association signals would be expected to be overrepresented among the SNPs of genes in the pathway since pathways represent combined genetic effects and candidate causal SNPs are supported to have higher confidence to be true [15]. Given the limited power of GWASs to detect single SNP associations, we adopted a pathway-based approach, which are more powerful, as they take into account biological interplay between genes, and attractive because they provide insight as to how multiple genes might contribute to the pathogenesis of a disease.

Kolz et al. conducted a meta-analysis of genome-wide association scans from 14 studies totaling 28,141 participants of European descent, resulting in identification of 954 SNPs distributed across nine loci [7]. The common variants associated with serum uric acid levels were found in the nine genes such as SLC2A9, ABCG2, SLC17A1, SLC22A11, SLC22A12, SLC16A9, GCKR, LRRC16A, and near PDZK1, supporting that transport proteins are key in regulating serum uric acid levels. In the present study, 14 candidate causal SNPs, five genes, and two candidate causal pathways were found to have significant nominal p -values by ICSNPPathway analysis, and these candidate causal SNPs and pathways provided five hypothetical biological mechanisms. This result is consistent with the well-known effects of PKD2, SLC17A1, SLC17A3, SLC17A4, and SLC2A9 on uric acid levels [16,17]. The most significant two SNPs to genes to hypotheses were as follows: first, rs2728121 (regulatory region) to PKD2 to ion transmembrane transporter activity; and second, rs942377, rs3799346, rs3799344, rs2762353, rs13197601, rs3757131, rs1165215, rs1165196 to SLC17A1 to ion transmembrane transporter activity and secondary active transmembrane transporter activity.

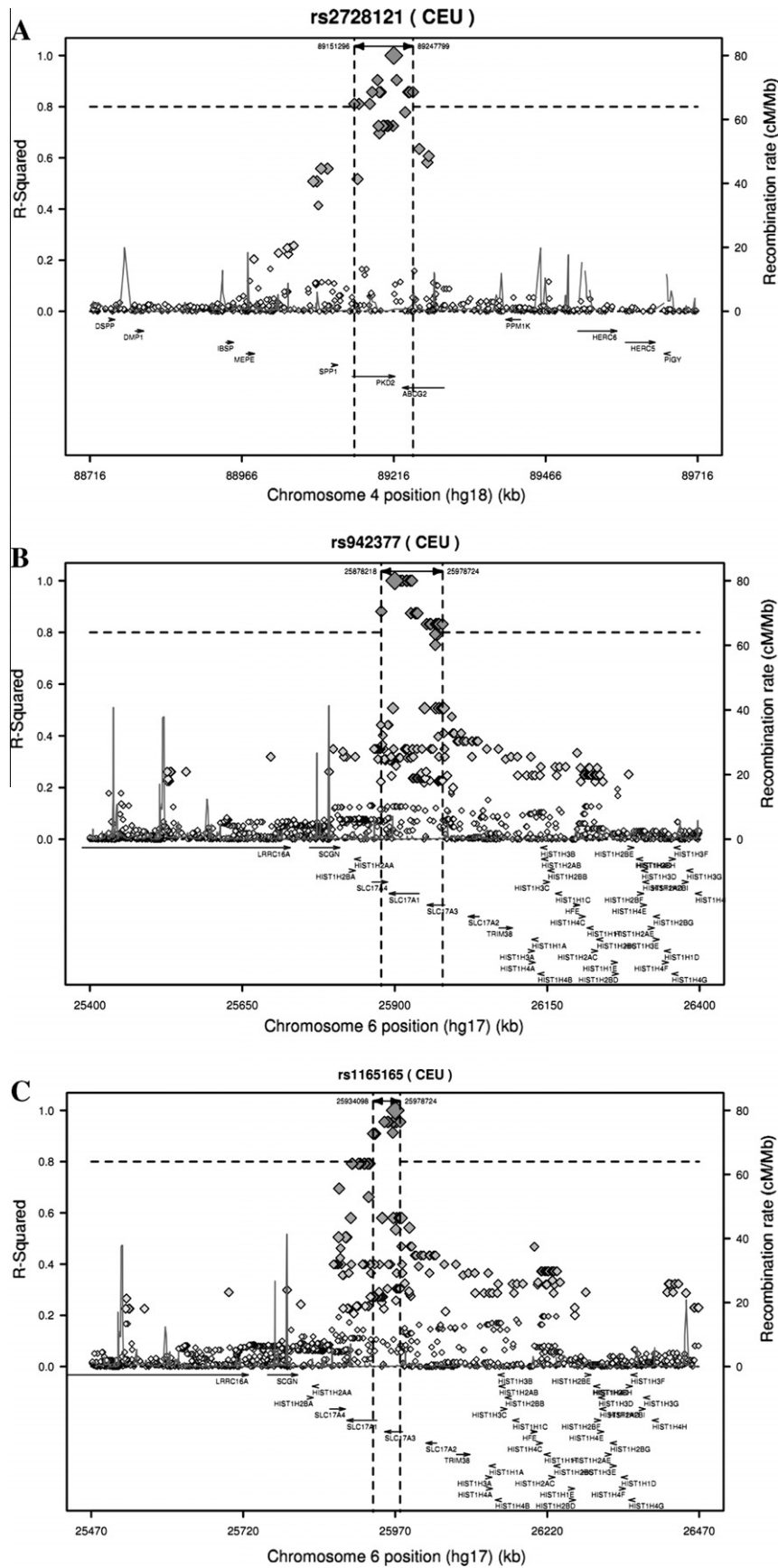


Fig. 1. Regional LD plots for SNPs rs2728121 (PKD2) (A), rs942377 (SLC17A1) (B), rs1165165 (SLC17A3) (C), rs11754288 (SLC17A4) (D), and rs16890979 (SLC2A9) (E). SNPs are plotted along with their proxies (based on HapMap CEU) as a function of genomic location, annotated by the recombination rate across the locus (light-blue line). On the y-axis, the pairwise r^2 is given for each proxy SNPs using color shading. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

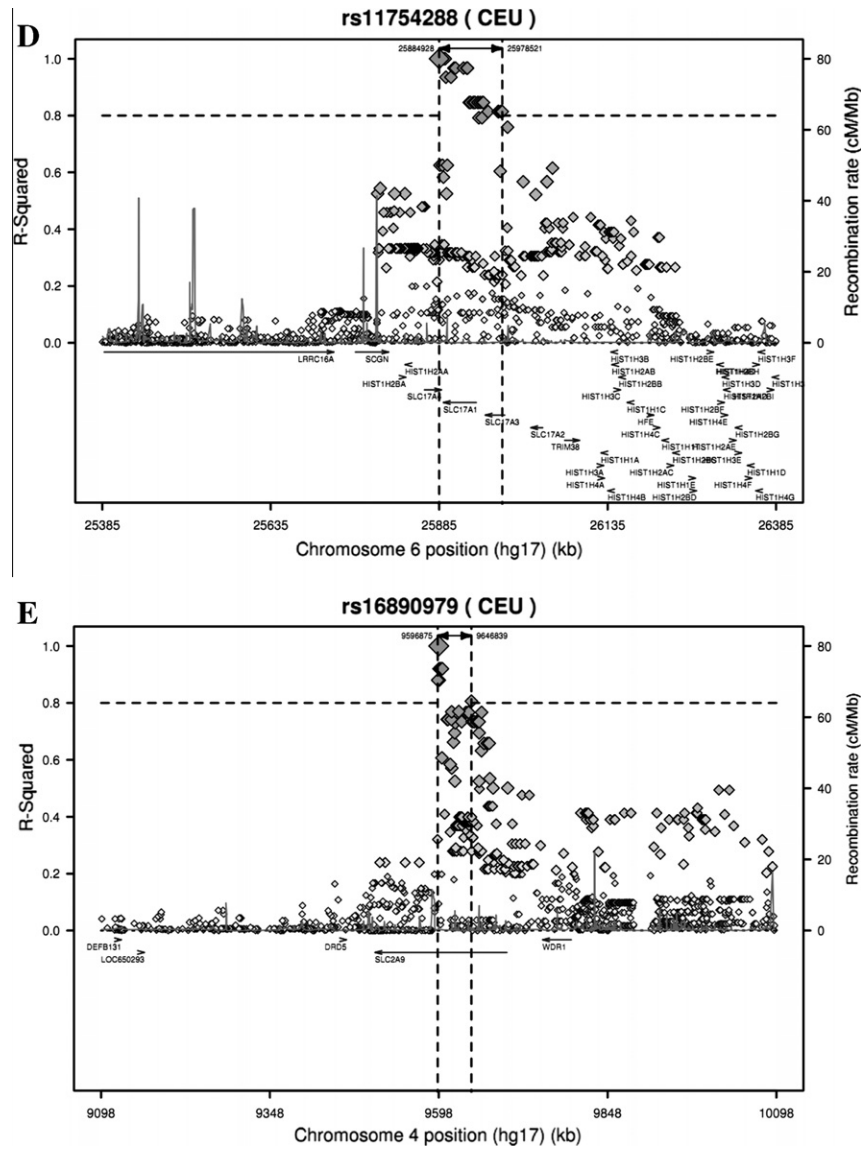


Fig. 1 (continued)

PKD2 encodes a member of the polycystin protein family. Its protein product may be an integral membrane protein involved in cell–cell/matrix interactions, function in renal tubular development, morphology, and function, and may modulate intracellular calcium homeostasis and other signal transduction pathways. PKD2 functions as a calcium permeable cation channel, but it has not been associated with serum uric acid levels. Thus, the role played by PKD2 in serum uric acid level requires confirmation and further functional analysis. SLC17A1, which is located directly downstream of SLC17A3 (NPT4), encodes NPT1 (renal sodium phosphate transporter protein 1). NPT1 is localized at the apical membrane of renal proximal tubules and acts as a renal sodium-dependent urate transporter. SLC2A9 codes GLUT9, which is known to be strongly associated with hyperuricemia and gout and to act as a high capacity urate transporter [18]. Furthermore, SLC2A9 is frequently associated with serum uric acid levels in GWASs [17]. SLC2A9 may alter serum uric acid levels by modulating the renal excretion of uric acid. Variants of SLC2A9 have been reported to exchange uric acid with fructose and glucose in proximal tubule lumen across the apical and basolateral membranes, leading to a net reabsorption of uric acid. Variation in

Table 3

Genes in the pathway and mapped with variant.

Gene name	Variant	Functional class	$-\log_{10}(p\text{-value})$
<i>A. Ion transmembrane transporter activity</i>			
SLC2A9	rs734553	–	200.282
SLC17A1	rs1183201	–	13.517
SLC17A3	rs1165182	–	13.423
SLC22A11	rs17300741	–	13.175
PKD2	rs2725215	–	12.697
SLC17A4	rs11754288	Nonsynonymous coding	11.359
<i>B. Secondary active transmembrane transporter activity</i>			
SLC2A9	rs734553	–	200.282
SLC17A1	rs1183201	–	13.517
SLC17A3	rs1165182	–	13.423
SLC22A11	rs17300741	–	13.175
SLC17A4	rs11754288	Nonsynonymous coding	11.359
SLC16A9	rs12356193	–	7.971

–: unknown.

SLC2A9 was found to explain up to 5% of the variation in serum urate concentrations [19].

Pathway-based approaches, such as, ICSNPathway analysis, were developed to meet the challenge posed by GWAS data interpretation, but no method has been devised to compare various pathway analysis methods [6]. The incomplete annotation of the human genome is an important limitation of the pathway-based approach to GWAS analysis. ICSNPathway analysis does not purport to predict true causal SNPs and pathways because of limited understanding of their genetic bases in complex diseases. Furthermore, a proportion of human genes has not been characterized and thus, cannot be mapped to predicted pathways. In addition, pathway analysis of GWAS data is still undergoing development, and thus, results should be interpreted with caution [20]. Additional studies are needed to confirm the identities of causal SNPs and of genes underlying the associations between pathways and uric acid levels. Nevertheless, pathway-based approaches are set to play an important role in the identification of novel genes that confer disease susceptibility [4]. The goal of pathway-based approaches is not to replace conventional single-maker analysis but to play a complementary part in identifying novel gene or sets of genes that confer disease susceptibility. By taking into account prior biological knowledge about genes and pathways, we may have a better chance to identify the genes and mechanisms that are involved in disease pathogenesis. An important application of the ICSNPathway results is to allow investigators to test a hypothesis concerning pathways by using candidate causal SNPs. Thus, the results of our pathway association approach using ICSNPathway analysis, 14 SNPs, five genes, and two pathways involving the PKD2, SLC17A1, SLC17A3, SLC17A4, and SLC2A9 genes, may add new information to the previously well conducted study by Kolz et al. [7] and may result in the formulation of new hypotheses for additional validations.

Present study has some limitations that require consideration. First, pathway analysis is most effective when SNPs achieving nominal significance are entered into the analysis. This can uncover relationships not seen because of the stringency of genome-wide significance. For the pathway-based analysis component of ICSNPathway, all GWAS SNPs should be used. In this analysis, ICSNPathway method was applied to only the 954 'significant' SNPs from a GWAS meta-analysis, because we could not access to the full set of GWAS *p*-values. This leads to limitation for a GSEA-type analysis, which relies on ranking of all *p*-values, to allow for comparison of ranks for SNPs within a pathway with ranks for all SNPs outside of the pathway. Second, larger genes have a propensity to have false positives. However, there was no correction for gene size in the SNP significance, because the data extracted from the GWAS dataset did not present the correction.

Summarizing, we analyzed the data of the GWAS meta-analysis of uric acid concentrations by using ICSNPathway analysis to identify genetic associations with uric acid levels at the SNP and pathway levels, because pathway-based approaches may complement

single-marker analyses of GWAS datasets. In the present study, we identified 14 candidate causal SNPs, five genes, and two pathways involving the PKD2, SLC17A1, SLC17A3, SLC17A4, and SLC2A9 genes. Further studies are needed to confirm and explore genetic variations of the pathways that influence uric acid levels.

Conflict of interest statement

The authors have no financial or non-financial conflict of interest to declare.

References

- [1] Riches PL, Wright AF, Ralston SH. Recent insights into the pathogenesis of hyperuricaemia and gout. *Hum Mol Genet* 2009;18:R177–84.
- [2] Richette P, Bardin T. Gout. *Lancet* 2010;375:318–28.
- [3] Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. *BMC Med Genet* 2009;10:6.
- [4] Yaspan BL, Veatch OJ. Strategies for pathway analysis from GWAS data. *Curr Protoc Hum Genet* 2011;71:1–15. [chapter 1: Unit 1.20].
- [5] Elbers CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet Epidemiol* 2009;33:419–31.
- [6] Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res* 2011;39:437–43.
- [7] Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* 2009;5:e1000504.
- [8] Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010;38:355–60.
- [9] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
- [10] Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24:2938–9.
- [11] Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–76.
- [12] Hong MG, Pawitan Y, Magnusson PK, Prince JA. Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum Genet* 2009;126:289–301.
- [13] Pedrosa I, Breen G. Gene set analysis and network analysis for genome-wide association studies. *Cold Spring Harb Protoc* 2011;9:1071–9.
- [14] Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* 2011;19:837–43.
- [15] Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278–83.
- [16] Merriman TR, Dalbeth N. The genetic basis of hyperuricaemia and gout. *Joint Bone Spine* 2011;78:35–40.
- [17] Choi HK, Zhu Y, Mount DB. Genetics of gout. *Curr Opin Rheumatol* 2010;22:144–51.
- [18] Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* 2008;40:437–42.
- [19] Doring A, Gieger C, Mehta D, Gohlke H, Prokisch H, Coassin S, et al. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet* 2008;40:430–6.
- [20] Jia P, Wang L, Meltzer HY, Zhao Z. Pathway-based analysis of GWAS datasets: effective but caution required. *Int J Neuropsychopharmacol* 2011;14:567–72.