Title: Variants in urate transporters, ADH1B, GCKR, and MEPE genes associate with transition from asymptomatic hyperuricaemia to gout: results of the first gout vs. asymptomatic hyperuricaemia GWAS in Caucasians using data from the UK Biobank.

Authors: Miss Gabriela Sandoval-Plata[1,2,3], Professor Kevin Morgan[3], Professor Abhishek Abhishek[1,2].

Affiliations: [1]Academic Rheumatology, University of Nottingham, Nottingham, UK; [2]Nottingham NIHR-BRC, Nottingham, UK; [3] Human Genetics, School of Life Sciences University of Nottingham, Nottingham, UK

Corresponding author: Gabriela Sandoval-Plata

A27, Academic Rheumatology. Clinical Sciences Building. University of Nottingham.

City Hospital Nottingham. Hucknall Road. Nottingham. NG5 1PB

Phone: +44 07955102573

Email: gabriela.sandovalplata@nottingham.ac.uk

ORCiD iD: 0000-0002-7809-7864

Word count: 3000

Abstract: 250

Key Messages

What is already known about this subject?

- Previous GWAS identified loci in inflammatory genes (CNTN-5, ZNF724 and MIR302F) as risk factors for transition from asymptomatic hyperuricaemia (AH) to gout, and was conducted in Japanese population.

What does this study add?

- This is the largest GWAS of gout cases and AH controls, and the first in Caucasian population.
- 13 variants in urate transporters and metabolic genes, but none in inflammatory genes associated with transition from AH to gout. A novel GWAS-significant gout risk locus was identified in ADH1B gene.
- Genetic and demographic factors performed moderately well in predicting gout status in AH.

How might this impact on clinical practice?

- Adults with AH should be advised lifestyle and dietary interventions that lower their serum urate levels in order to reduce their risk of gout.

ABSTRACT

Objectives: To perform a genome-wide association study (GWAS) of gout cases vs. asymptomatic hyperuricaemia (AH) controls, and gout cases vs. normouricaemia controls, and to generate a polygenic risk score (PRS) to determine gout-case vs. AH-control status.

Methods: Gout cases and AH controls (serum urate (SU) ≥6mg/dL) from the UK Biobank were divided into discovery (4,934 cases, 56,948 controls) and replication (2,115 cases, 24,406 controls) cohorts. GWAS was conducted and PRS generated using summary statistics in discovery cohort as the base dataset and the replication cohort as the target dataset. The predictive ability of the model was evaluated. GWAS were performed to identify variants associated with gout compared to normouricaemic controls using SU <6.0 mg/dL and <7.0 mg/dL thresholds respectively.

Results: Thirteen independent SNPs in ABCG2, SLC2A9, SLC22A11, GCKR, MEPE, PPM1K-DT, LOC105377323, and ADH1B reached genome-wide significance and replicated as predictors of AH to gout transition. Twelve of 13 associations were novel for this transition, and rs1229984 (ADH1B) was identified as GWAS locus for gout for the first time. The best PRS model was generated from association data of 17 SNPs; and had predictive ability of 58.5% that increased to 69.2% upon including demographic factors. Two novel SNPs rs760077(MTX1) and rs3800307(PRSS16) achieved GWAS significance for association with gout compared with normouricaemic controls using both SU thresholds.

Conclusion: The association of urate transporters with gout supports the central role of hyperuricaemia in its pathogenesis. Larger GWAS are required to identify if variants in inflammatory pathways contribute to progression from AH to gout.

Key words: Crystal arthropathies, gout, arthritis.

INTRODUCTION Gout is a common form of inflammatory arthritis caused by the deposition of monosodium urate (MSU) crystals. Elevated serum urate (SU) concentration is the precursor to MSU crystal deposition, and the onset of gout.[1] However, the majority of people with hyperuricemia do not develop gout. For instance, in the USA, the prevalence of hyperuricaemia (defined as SU >7.0 mg/dL) is 20%, while that of gout is 3.9%.[2] The reason(s) why only some people with hyperuricaemia develop gout is poorly understood. Genome wide association studies (GWAS) have improved the understanding of the pathophysiology of hyperuricaemia and gout over the last 10-15 years. For instance, genetic variants located in urate transporters such as the ABCG2, SLC2A9, and SLC22A11 genes have been identified as risk loci for both hyperuricemia and gout.[3-6] Additional genetic variants such as GCKR and ALDH2 that play important roles in carbohydrate and alcohol metabolism respectively have been associated with both phenotypes.[5 7-9] However, the genetic variants associated with progression from hyperuricaemia to gout remain poorly understood. To date, only a single relatively small GWAS (n=6,009 Japanese adults, 2,860 with gout) has examined this and revealed two novel loci: CNTN5 and MIR302F, that participate in immune and inflammatory responses.[10] However, the identified polymorphism in CNTN5 is intronic while the SNP near MIR302F is intergenic. Further analyses in independent populations and larger sample sizes are needed to improve the understanding of the molecular mechanisms involved in transitioning from asymptomatic hyperuricaemia (AH) to gout.

Thus, the purpose of this study was to examine the genetic variants associated with transition from hyperuricaemia to gout. In order to meet this objective, we performed a GWAS using gout cases and [1] AH controls (SU ≥6.0 mg/dL), [2] normouricaemia controls with SU <6.0 mg/dL, and [3] normouricaemia controls with SU <7.0 mg/dL derived from the UK Biobank resource. Genotype data was used to develop a polygenic risk score (PRS) to predict gout-case and AH-control status. We chose a threshold of ≥6.0 mg/dL to define AH as the risk of incident gout increases above this SU level.[11]

METHODS

Data source: This study was conducted using data from the UK Biobank resource (Project ID 45987). Briefly, the UK Biobank is a prospective study of ~500,000 participants, aged 40-60 years and recruited across England, Wales and Scotland between the years 2006 and 2010. Data were collected on lifestyle and sociodemographic information, cognitive function, health status, and family medical history. Participants had standard physical and functional measurements, and provided blood samples for genetic analyses. Details about recruitment and samples processing for genotyping are described elsewhere.[12 13]

Subjects: For this research, three phenotypes were derived from the UK Biobank cohort.

Gout cases: Gout was defined as present if any of the following criteria were met:[14] self-reported physician diagnosed gout; urate lowering therapy (ULT) prescription without a hospital diagnosis of lymphoma or leukaemia (ICD-10 codes C81-C96); or a primary or secondary diagnosis of gout in hospital discharge letters using the ICD-10 codes M10, M100-M14, and M109. Participants with self-reported physician diagnosed gout were excluded if their SU was <6.0 mg/dL and they did not report prescription of ULT at the UK Biobank visit.

AH controls: Participants with SU ≥6.0 mg/dL and not classified as gout. A threshold of ≥6.0 mg/dl was chosen as it associates with incident gout in prospective studies.[11]

Normouricaemia controls: Participants with SU <6.0 mg/dL and not classified as gout[14] were considered as normouricaemia controls. Given the uncertainty around definition of normal SU, for example SU <6.0 mg/dL being the treatment threshold for treat-to-target ULT while epidemiological studies use a cut-off of <7.0 mg/dl, another group of normouricaemia controls was ascertained with SU <7.0 mg/dl and not classified as gout[14].

Genotyping and quality control: UK Biobank samples were genotyped by Affymetrix using two arrays: The UK BiLEVE Axiom array (n=49,950; 807,411 markers), and the UK Biobank Axiom array (n=438,427; 825,927 markers). These arrays shared 95% of content, resulting in >805,000 genotyped variants for 488,288 participants. For this study, participants with non-European ancestry were excluded to avoid population stratification. Thus, genotyping data for

409,629 European descendants were available following UK Biobank centrally performed quality control (QC) procedures. Detailed information about genotyping and QC have been described previously.[13 15] Further stringent QC filters were applied using PLINK 1.9.[16] Individuals with a kinship coefficient equivalent to second degree (or greater) relatives were excluded. Individuals were also excluded if they had a heterozygosity ±3 standard deviations (SD) from the mean, a call rate <90%, or were identified as gender mismatches. Markers with a call rate <95%, or those deviating from Hardy-Weinberg equilibrium (Bonferroni corrected p value threshold $p=6.82x10^{-8}$) were removed from the dataset.

Gout vs. AH: A sample of 354,825 individuals with 717,091 genotyped SNPs were included in this analysis, from which the phenotypes of interest were derived. The cohort comprised of 7,049 cases and 81,354 controls that were divided into two datasets: 70% (n=61,882) was used as the discovery dataset, and the remaining 30% (n=26,521) was used as the replication dataset (Figure 1).

Gout vs. normouricaemia: Two separate GWAS were conducted, using gout cases and 64,424 controls with SU <6.0 mg/dL, and 79,531 controls with SU <7.0 mg/dL respectively.

Statistical analyses: Baseline data were summarised using mean (SD) for continuous variables, and number (%) for categorical variables. Independent sample t-test and chi-square test were used to compare continuous and categorical data respectively.

Gout vs. AH GWAS: Discovery and replication association tests were performed using PLINK v1.9. Odds ratios (OR) and 95% confidence intervals (CI) were computed using additive logistic regression. We adjusted for sex, age at recruitment, and 10 principal components (PCs). To determine the number of independent loci from the GWAS analysis, linkage disequilibrium (LD) clumping was performed using PLINK v1.9. SNPs with a p value $<1x10^{-5}$, r^2 >0.1, and a 500kb window from the index SNPs were assigned to the clump. Annotation of lead SNPs was conducted using the SNP2GENE tool of the Functional Mapping and Annotation of Genome Wide Association Studies (FUMA-GWAS).[17] Pairwise LD patterns from

SNPs identified as independent, located in the same gene or less than 500kb apart, were further analysed using the R package LDlinkR [18 19] which uses the 1000 Genomes Project data as the reference panel. HaploView was used to generate the LD plot.[20] For the discovery analysis, genome wide significance was set at $p=5x10^{-8}$.

For replication analysis, the thirteen variants that reached genome wide significance in the discovery analysis were tested for association with gout in the replication cohort. Logistic regression was adjusted for sex, age, and 10 PCs. A Bonferroni corrected p value of <0.004 (0.05/13) was used to determine significant associations in the replication analysis. The results from the discovery GWAS and the replication analysis were combined by meta-analysis using PLINK. The fixed-effects model was used to estimate pooled ORs and 95% CI, and Cochran's Q test *p* values and I^2 values were used to assess heterogeneity.

Linear regression was used to examine the effect of GWAS hits on SU levels. This was performed using the full cohort, and adjusted for sex and age at recruitment. Beta coefficients and standard errors (SE), and adjusted Beta coefficients and SE were calculated. As previous GWAS [10 21] have used a cut-off of 7.0 mg/dl to define hyperuricaemia, a sensitivity analysis was conducted to evaluate if the association of GWAS hits and gout remained significant if controls had a SU ≥7.0 mg/dL.

PRS: PRS was calculated using PRSice-2.[22] The discovery GWAS summary statistics were used as the base dataset, while the replication cohort genotype-phenotype data were used as the target dataset. Clumping parameters in PRSice were set to an r^2 >0.1 and a 500kb window from the index SNPs, which generated a final number of 266,754 SNPs available for PRS calculation. Odds ratios (OR) and *p* values from the GWAS summary statistics were used to calculate the best PRS model, which was generated from testing different *p* value thresholds. The best-fit model was defined by the largest Nagelkerke's R^2 value. Logistic regression was used to estimate the effect of the demographic variables for inclusion into the predictive models using SPSS Statistics 24. The area under the receiving operatic characteristic curve

(AUROC) was used to evaluate the predictive ability of the PRS, demographic characteristics (age, sex and body mass index (BMI)), and combined models.

Gout vs. normouricaemia GWAS: Two GWAS were conducted. Prior to conducting these analyses, both datasets underwent the same genotyping QC filters as described earlier. The association tests were performed with PLINK v1.9, using age, sex and the first 10 PCs as covariates.

RESULTS

Demographic characteristics: Following genotype QC filters, data for 7,049 gout cases and 81,354 AH controls were included. The entire cohort comprised of 80.77% men, and their mean (SD) age, BMI, and SU were 57.87 (7.77) years, 29.63 (4.81) kg/m^2, and 6.92 (0.88) mg/dL respectively. This cohort was divided into the discovery and replication datasets (Table 1, Figure 1). The two datasets had comparable disease and demographic characteristics.

Gout vs. AH:

GWAS: An additive logistic regression was performed to test the association between gout and 710,030 variants. Thirty-four SNPs reached genome wide significance and after filtering for tight LD (r^2<2.0), 13 SNPs were identified as independent associations (Figure 2). These lead SNPs were selected for the replication study, where they were tested for association with gout in the remaining 30% of the dataset. Successful replication was defined if the p value was < 0.004. Summary results for both the discovery and the replication analyses are shown in Table 2. The SNP with the greatest effect was rs2231142 in ABCG2 gene with OR=1.66 ($2.05x10^{-78}$) in the discovery stage, and OR=1.64 ($1.17x10^{-32}$) in the replication stage. This was followed by a novel locus: rs1229984 in ADH1B gene (OR=1.51, $p=5.00x10^{-12}$; OR=1.44, $p=4.77x10^{-5}$). The remaining SNPs were located in or near GCKR, PPM1K-DT, SLC2A9, MEPE, LOC105377323, and SLC22A11. Pairwise LD parameters were evaluated for SNPs located within the same gene or in genes <500kb apart and had r^2<0.2 (Figure S1).

Table 1. Demographic, life-style and comorbidities for gout cases and asymptomatic hyperuricaemia controls of the UK Biobank

		Discovery GWAS		Replication stage	
	All	Gout cases	Controls	Gout cases	Controls
	(88,403)	(4,934)	(56,948)	2,115	24,406
Age at recruitment, years, mean (SD)	57.87 (7.77)	60.11 (6.88)	57.67 (7.82)	60.06 (6.81)	57.69 (7.80)
Male sex, n (%)	71,401 (80.77)	4,529 (91.79)	45,506 (79.91)	1,950 (92.20)	19,416 (79.55)
BMI, kg/m^2, mean (SD)	29.63 (4.81)	30.79 (4.96)	29.53 (4.78)	30.62 (4.93)	29.53 (4.78)
Waist circumference, cm, mean (SD)	99.56 (11.74)	103.76 (12.14)	99.23 (11.64)	103.49 (12.12)	99.16 (11.64)
SU, mg/dL, mean (SD)	6.92 (0.88)	6.74 (1.78)	6.94 (0.76)	6.77 (1.74)	6.93 (0.75)
Alcohol intake, n (%)*					
Never	4,397 (4.97)	221 (4.48)	2,822 (4.96)	85 (4.02)	1,269 (5.20)
Special occasions	6,698 (7.58)	272 (5.51)	4,405 (7.74)	115 (5.44)	1,906 (7.81)
<1/week	7,588 (8.58)	278 (5.63)	5,006 (8.79)	128 (6.05)	2,176 (8.92)
1-2/week	21,908 (24.78)	1,093 (22.15)	14,190 (24.92)	496 (23.45)	6,129 (25.11)
3-4/week	23,370 (26.44)	1,351 (27.38)	15,035 (26.40)	589 (27.85)	6,395 (26.20)
Daily or almost daily	24,365 (27.56)	1,710 (34.66)	15,448 (27.13)	701 (33.14)	6,506 (26.66)
Smoking status, n (%)*					
Non-smoker	41,778 (47.26)	1,998 (40.49)	27,243 (47.84)	862 (40.76)	11,675 (47.84)
Ex-smoker	37,705 (42.65)	2,481 (50.28)	23,871 (41.92)	1,065 (50.35)	10,288 (42.15)
Current smoker	8,590 (9.72)	434 (8.80)	5,613 (9.86)	183 (8.65)	2,360 (9.67)
Comorbidities, n (%)					
Metabolic syndrome, n (%)	40,049 (45.30)	2,999 (60.78)	24,900 (43.72)	1,325 (62.65)	10,825 (44.35)
Diabetes Mellitus	5,389 (6.10)	589 (11.94)	3,205 (5.63)	245 (11.58)	1,350 (5.53)
Hypertension	35,776 (40.47)	2,824 (57.24)	22,206 (38.99)	1,241 (58.68)	9,505 (38.95)
Hypercholesterolemia	15,831 (17.91)	1,368 (27.73)	9,737 (17.10)	585 (27.66)	4,141 (16.97)
Ischaemic Heart Disease	6,817 (7.71)	666 (13.50)	4,051 (7.11)	290 (13.71)	1,810 (7.42)
Cardiac failure	133 (0.15)	28 (0.57)	62 (0.11)	18 (0.85)	25 (0.10)
Chronic Kidney Disease stages*					
G1 (>90 ml/min	39,800 (45.02)	1,962 (39.76)	25,849(45.39)	855 (40.43)	11,134 (45.62)
G2 (60-90 ml/min)	43,774 (49.52)	2,434 (49.33)	28,260 (49.62)	1,026 (48.51)	12,054 (49.39)
G3a (45-59 ml/min)	3,722 (4.21)	343 (6.95)	2,258 (3.97)	159 (7.52)	962 (3.94)
G3b (30-44 ml/min)	792 (0.89)	138 (2.80)	419 (0.74)	40 (1.89)	195 (0.80)
G4 (15-29 ml/min)	213 (0.24)	41 (0.83)	106 (0.19)	22 (1.04)	44 (0.18)
G5 (<15 ml/min)	51 (0.06)	13 (0.26)	19 (0.03)	12 (0.57)	7 (0.03)

*The following data were missing: alcohol intake for 0.09%, smoking satus for 0.73%, and CKD information for 0.06%.

Diabetes, hypertension, hypercholesterolemia,ischaemic heart disease and cardiac failure were defined as present if they were self-reported as diagnosed by a doctor. Chronic Kidney Disease stages were defined as per the National Institute of Heath and Care Excellence (NICE) guidelines CG182.[23] Metabolic syndrome was calculated as recommended by the International Diabetes Federation in 2006.[24]

Table 2. Summary of GWAS and replication analysis of 13 lead SNPs in gout cases and AH controls (SU >= 6 mg/dL).

						Discovery GWAS		Replication stage		Meta-Analysis[1]			
SNP	**Chr**	**bp**	**Gene**	**A1**	**Freq**	**aOR (95% CI)[2]**	**P value**	**aOR (95% CI)[2]**	**P value**	**aOR (95% CI)[2]**	**P value**	**Cochrane's Q p value**	**I^2**
rs1260326	2	27730940	GCKR	T	0.41	1.13 (1.08-1.17)	$3.54x10^{-08}$	1.15 (1.08-1.23)	$2.54x10^{-05}$	1.14 (1.10-1.18)	$5.10x10^{-12}$	0.61	0
rs2231142	4	89052323	ABCG2	T	0.14	1.66 (1.58-1.76)	$2.05x10^{-78}$	1.64 (1.51-1.78)	$1.17x10^{-32}$	1.65 (1.58-1.73)	$3.33x10^{-109}$	0.75	0
rs13120400	4	89033527	ABCG2	C	0.28	0.82 (0.78-0.86)	$1.56x10^{-16}$	0.84 (0.78-0.91)	$3.43x10^{-06}$	0.83 (0.79-0.86)	$3.66x10^{-21}$	0.50	0
rs7672194	4	89126647	ABCG2	T	0.48	1.16 (1.11-1.21)	$3.21x10^{-12}$	1.15 (1.08-1.23)	$1.13x10^{-05}$	1.16 (1.12-1.20)	$1.58x10^{-16}$	0.88	0
rs4693211	4	89249061	PPM1K-DT	C	0.07	1.41 (1.31-1.52)	$6.97x10^{-20}$	1.26 (1.12-1.42)	$1.21x10^{-04}$	1.37 (1.28-1.45)	$1.55x10^{-22}$	0.11	61.62
rs28793136	4	89216768	PPM1K-DT	C	0.08	1.35 (1.26-1.45)	$8.19x10^{-17}$	1.26 (1.12-1.40)	$6.29x10^{-05}$	1.32 (1.25-1.40)	$4.79x10^{-20}$	0.27	19.60
rs1545207	4	89239492	PPM1K-DT	A	0.28	1.14 (1.09-1.20)	$6.82x10^{-09}$	1.12 (1.05-1.20)	$1.07x10^{-03}$	1.14 (1.09-1.18)	$3.38x10^{-11}$	0.66	0
rs16890979	4	9922167	SLC2A9	T	0.17	0.79 (0.74-0.83)	$3.19x10^{-16}$	0.73 (0.67-0.80)	$1.05x10^{-11}$	0.77 (0.74-0.81)	$5.45x10^{-26}$	0.19	42.09
rs16891234	4	9946163	SLC2A9	C	0.24	1.16 (1.11-1.22)	$6.06x10^{-10}$	1.13 (1.05-1.22)	$8.86x10^{-04}$	1.15 (1.11-1.20)	$2.72x10^{-12}$	0.57	0
rs1229984	4	100239319	ADH1B	T	0.03	1.51 (1.34-1.69)	$5.00x10^{-12}$	1.44 (1.21-1.72)	$4.77x10^{-05}$	1.49 (1.35-1.64)	$1.15x10^{-15}$	0.68	0
rs114791459	4	88591554	LOC105377323	A	0.02	1.42 (1.26-1.60)	$7.99x10^{-09}$	1.47 (1.22-1.77)	$5.47x10^{-05}$	1.43 (1.30-1.59)	$2.01x10^{-12}$	0.76	0
rs114580333	4	88790118	MEPE	A	0.02	1.44 (1.26-1.63)	$3.01x10^{-08}$	1.39 (1.15-1.69)	$9.18x10^{-04}$	1.42 (1.28-1.59)	$1.10x10^{-10}$	0.79	0
rs2078267	11	64334114	SLC22A11	C	0.47	1.16 (1.11-1.21)	$1.72x10^{-12}$	1.14 (1.07-1.22)	$6.27x10^{-05}$	1.15 (1.11-1.20)	$6.65x10^{-16}$	0.62	0

[1]Fixed-effects meta-analysis of the discovery GWAS and the replication analysis.

[2]Adjusted for age, sex, and 10 first PCs

A1, allele 1/effect allele; Bp, base pair position; Chr, chromosome; Freq, frequency; OR, Odds Ratios; SNP, single nucleotide polymorphism

Genetic variants and SU: All lead SNPs associated with SU, with rs2231142 (ABCG2) and rs16890979 (SLC2A9) showing the greatest effects: adjusted β=0.107 and $p=1.21x10^{-80}$, and adjusted β=-0.055 and $p=1.67x10^{-43}$ respectively (Table S1). On sensitivity analysis examining the association between 13 lead SNPs and gout, excluding AH controls with SU <7.0 mg/dL, the ORs diminished in magnitude but remained significant (Table S2).

PRS model: A PRS for all cases and controls was constructed with PRSice using the replication cohort as the test dataset. The best-fit p value threshold that gave the highest Nagelkerke's R^2 (0.016) was $4.0x10^{-6}$, and included 17 SNPs (Table S4). The mean (±SD) PRS for cases was 0.018 (±0.017), and 0.013 (±0.016) for controls (p <0.0001). The predictive ability of this PRS model was evaluated using the AUROC curve, and compared to the demographics model (age, sex and BMI) and combined model (age, sex, BMI and PRS). The AUC for each model was 58.5%, 66.7%, and 69.2% respectively (Figure 3).

Gout vs. normouricaemia: We conducted two GWAS of gout vs normouricaemia using SU cut-off values <6.0 mg/dL and <7.0 mg/dL respectively. The first GWAS identified 52 lead SNPs, while the second identified 46 lead SNPs (Table S3). Three novel SNPs (rs760077, rs3800307, and rs11227299 in MTX1, PRSS16, and AP5B1 genes respectively) associated with gout compared to SU <6mg/dL with GWAS significance. Two (rs760077 and rs3800307) remained GWAS significant when a higher SU threshold of <7 mg/dL was used to define normouricaemia. We then plotted the OR of each lead SNP in the gout vs AH-control GWAS and the gout vs SU <6mg/dL GWAS (Figure 4A). To be consistent with our sensitivity analysis, we plotted the OR of the thirteen lead SNPs upon excluding AH controls with SU 6-7mg/dL, with the gout vs SU <7mg/dL GWAS (Figure 4B). Same loci were responsible for transition from AH and normouricaemia to gout.

DISCUSSION This is the largest GWAS to date and the first in Caucasians to examine the SNPs associated with transition from AH to gout. Using UK Biobank data, it identified 13 independent SNPs from eight loci that reached genome wide significance for association with gout vs. AH, and replicated. These loci include urate transporters, metabolic pathway genes (e.g. GCKR, ADH1B), and *MEPE* gene that regulates renal phosphate handling and skeletal mineralization.[25] The latter may promote progression to gout via pro-mineralising osteopontin like function or via low phosphate levels that associates with incident hyperuricaemia.[26] The identified loci in PPM1K-DT and LOC105377323 were in non-coding regions and their molecular mechanism is unclear.

Of the eight loci, ABCG2, SLC2A9, SLAC22A11, PPM1K-DT, GCKR, and MEPE have previously been associated with gout or SU levels in different populations but never in the transition from AH to gout in a GWAS.[5 21 27 28] This is the first such report. In a previous study, Tin *et al* generated a genetic risk score using variants associated with SU and examined their ability to predict gout cases in 334,800 UK Biobank participants not specifically selected for high SU levels. Ours is the first study to attempt to generate a PRS for predicting gout status in an AH population i.e. those with SU ≥6.0 mg/dL, and reports an AUC of 58.5% for genetic factors alone, that increased to just under 70% when demographic factors were added. This is lower than the AUC of 67.2% from genetic factors alone in the study by Tin *et al* and is likely to be due to lower genetic variance due to selection of a high SU control group.[27] A smaller study using candidate gene hypothesis reported nominal association for ABCG2 polymorphism and gout vs. hyperuricaemia.[29] The only previous gout vs. hyperuricaemia GWAS was conducted in a Japanese population and reported rs7927466 in CNTN5, rs9952962 in MIR302F, and a suggestive locus rs12980365 in ZNF724 that do not affect SU.[10] Although rs7927466 is not included in the UK Biobank genotype platform, it is covered by its proxy SNP rs7942264 (r^2=1) that did not show an association with gout; neither did rs12980365. MIR302F was not included in UK Biobank and further research on this gene is needed.

ADH1B was identified as a risk variant for gout vs. AH. It has never previously been associated with gout in a GWAS – even when compared to general population. ADH1B, mediates the oxidation of ethanol into acetaldehyde.[30] The SNP rs1229984 in ADH1B causes a change of an arginine to histidine, increases ethanol clearance in liver, facilitates its' conversion to highly reactive acetaldehyde[31], increases the NADH/NAD ratio that results in high lactic acid levels and increased urate reabsorption via URAT1.[32] The risk allele of rs1229984 also promotes a "flush response" to alcohol and reduces the amount of alcohol consumed[33]. Thus, the association between this polymorphism and gout may be due to increased production and reabsorption of urate from per unit alcohol consumed. This is consistent with the observation by Yokoyama *et al* in which rs1229984 associated with SU≥7 mg/dL (OR (95%CI) 2.04 (1.58–2.65)), while the daily alcohol intake was comparable across variants.[34] In agreement with our study, Sakiyama *et al* (n=1,048 gout cases and 1,334 male controls) evaluated the effect of rs1229984 in ADH1B gene on gout. They, reported an increased risk for gout with OR of 1.69 and 1.80 for His/Arg and His/His genotypes respectively, that remained significant after correcting for alcohol consumption.[35] However, in their study patients with gout and rare variants of the SNP had greater alcohol consumption, suggesting an additional role for the latter.

Urate transporters ABCG2, SLC2A9 and SLC22A11 play essential roles in pathogenesis of hyperuricaemia.[36 37] SLC2A9 has the strongest effect on SU, accounting for 2-3% of variance, followed by ABCG2 that explains 1% of SU variation.[37] Although both loci have also been associated to gout, GWAS of gout cases vs controls, have shown a greater effect of ABCG2 than SLC2A9.[7] In this study, rs2231142 in ABCG2 had larger effect size on gout status compared to AH-control than that of rs16890979 in SLC2A9 (which is in tight LD with the GWAS hit rs12498742) and also twice as much effect on SU than the latter. This supports the hypothesis that ABCG2, plays a causal role in transition from hyperuricaemia to gout via its effect on SU. However, additional mechanisms such as defects in ABCG2 causing deficient autophagy may also operate.[29 38]

Our GWAS comparing gout cases with normouricaemic controls did not identify any inflammatory genes. A large number of lead SNPs were identified at genome-wide significance level. Most have been associated with gout or SU previously.[5 7 27 28 39] However, we identified three novel SNPs associated with gout compared to SU <6.0 mg/dL. Of these, rs11227299 (AP5B1) is associated with reduced eGFR and may cause gout by resulting hyperuricemia.[40] The variants in MTX1 and PRSS16 genes associated with gout compared to SU <6.0 mg/dL, also associate negatively with Parkinson's Disease and Schizophrenia.[41 42] This is consistent with the negative associations between Parkinson's Disease and gout, and Schizophrenia and elevated SU.[43 44]

This is the first GWAS to examine transition from AH to gout in Caucasians. Other strengths include a large sample size, and assessment of transition from AH or normouricaemia to gout in the same source population. However, there are several caveats to this study. Firstly, gout definition was not based on ACR/EULAR classification criteria, but was ascertained via self-report of physician diagnosis, hospital diagnoses and ULT prescriptions. However, the UK Biobank data collection predates the classification criteria. In addition, the classification of AH controls was based on a single SU measurement which could have been affected by diet during the previous days. Additionally, the use of non-imputed data limited the discovery power of the GWAS and PRS.

In conclusion, this study identified 13 GWAS significant risk loci, 12 of which have never previously been associated with the transition from AH to gout at GWAS level. The preponderance of urate transporters and metabolic genes that affect SU levels support the central role of hyperuricaemia in the pathogenesis of gout. Larger GWAS are required to identify if variants in inflammatory pathways also contribute to this transition.

Funding: Miss Sandoval-Plata's work is supported by a PhD scholarship from The Mexican National Council for Science and Technology (CONACYT) [Grant number 472298].

Acknowledgements: This research was conducted using the UK Biobank resource (Project ID 45987). The authors would like to thank all participants.

Contributors: Study concept and design: Abhishek, Sandoval-Plata, Morgan. Data analysis: Sandoval-Plata. Manuscript preparation: Sandoval-Plata, Abhishek, Morgan. Manuscript revision for important intellectual content: all authors.

Ethics approval: UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (REC reference 16/NW/0274). This study did not involve recontacting participants and no separate ethics approval was required.

Data availability statement: Raw data used for this study are available from the UK Biobank resource.

Competing interests: Prof A Abhishek has received departmental research grants from AstraZeneca and Oxford Immunotec, speaker bureau fees from Menarini, scientific meeting support from Pfizer, author royalties from UpToDate and Springer, and has consulted for Inflazome unrelated to this work.

REFERENCES

1. Dalbeth N, Merriman TR, Stamp LK. Gout. *The Lancet* 2016;388(10055):2039-52. doi: 10.1016/s0140-6736(16)00346-9
2. Chen-Xu M, Yokose C, Rai SK, et al. Contemporary Prevalence of Gout and Hyperuricemia in the United States and Decadal Trends: The National Health and Nutrition Examination Survey, 2007-2016. 2019(2326-5205 (Electronic))
3. Li S, Sanna S, Maschio A, et al. The GLUT9 Gene Is Associated with Serum Uric Acid Levels in Sardinia and Chianti Cohorts. *PLOS Genetics* 2007;3(11):e194. doi: 10.1371/journal.pgen.0030194
4. Dehghan A, Köttgen A, Yang Q, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *The Lancet* 2008;372(9654):1953-61. doi: 10.1016/s0140-6736(08)61343-4
5. Kottgen A, Albrecht E, Teumer A, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* 2013;45(2):145-54. doi: 10.1038/ng.2500
6. Kolz M, Johnson T, Sanna S, et al. Meta-Analysis of 28,141 Individuals Identifies Common Variants within Five New Loci That Influence Uric Acid Concentrations. *PLOS Genetics* 2009;5(6):e1000504. doi: 10.1371/journal.pgen.1000504
7. Phipps-Green AJ, Merriman ME, Topless R, et al. Twenty-eight loci that influence serum urate levels: analysis of association with gout. *Ann Rheum Dis* 2016;75(1):124-30. doi: 10.1136/annrheumdis-2014-205877
8. Matsuo H, Yamamoto K, Nakaoka H, et al. Genome-wide association study of clinically defined gout identifies multiple risk loci and its association with clinical subtypes. *Ann Rheum Dis* 2016;75(4):652-9. doi: 10.1136/annrheumdis-2014-206191
9. Li C, Li Z, Liu S, et al. Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nat Commun* 2015;6:7041. doi: 10.1038/ncomms8041
10. Kawamura Y, Nakaoka H, Nakayama A, et al. Genome-wide association study revealed novel loci which aggravate asymptomatic hyperuricaemia into gout. *Annals of the Rheumatic Diseases* 2019;78(10):1430. doi: 10.1136/annrheumdis-2019-215521
11. Dalbeth N, Phipps-Green A, Frampton C, et al. Relationship between serum urate concentration and clinically evident incident gout: an individual participant data analysis. *Ann Rheum Dis* 2018;77(7):1048-52. doi: 10.1136/annrheumdis-2017-212288 [published Online First: 2018/02/22]
12. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* 2015;12(3):e1001779. doi: 10.1371/journal.pmed.1001779
13. Welsh S. Genotyping of 500,000 UK Biobank participants: Description of sample processing workflow and preparation of DNA for genotyping. 2017
14. Cadzow M, Merriman TR, Dalbeth N. Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. *Arthritis Res Ther* 2017;19(1):181. doi: 10.1186/s13075-017-1390-1
15. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 2017:166298. doi: 10.1101/166298
16. Chang CC, Chow CC, Tellier LCAM, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015;4(1) doi: 10.1186/s13742-015-0047-8
17. Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017;8(1):1826. doi: 10.1038/s41467-017-01261-5
18. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31(21):3555-57. doi: 10.1093/bioinformatics/btv402

19. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Frontiers in Genetics* 2020;11(157) doi: 10.3389/fgene.2020.00157
20. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2004;21(2):263-65. doi: 10.1093/bioinformatics/bth457
21. Nakayama A, Nakaoka H, Yamamoto K, et al. GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include urate transporter genes. *Annals of the Rheumatic Diseases* 2017;76(5):869.
22. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* 2019;8(7) doi: 10.1093/gigascience/giz082
23. NICE. Chronic kidney disease in adults: assessment and management 2015 [Available from: https://www.nice.org.uk/guidance/cg182 accessed 28 Septemeber 2020.
24. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabetic Medicine* 2006;23(5):469-80. doi: https://doi.org/10.1111/j.1464-5491.2006.01858.x
25. Rowe PSN. The wrickkened pathways of FGF23, MEPE and PHEX. *Crit Rev Oral Biol Med* 2004;15(5):264-81. doi: 10.1177/154411130401500503
26. Cao J, Zhang J, Li Q, et al. Serum Phosphate and the Risk of New-Onset Hyperuricemia in Hypertensive Patients. *Hypertension* 2019;74(1):102-10. doi: 10.1161/HYPERTENSIONAHA.119.12633
27. Tin A, Marten J, Halperin Kuhns VL, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nature Genetics* 2019;51(10):1459-74. doi: 10.1038/s41588-019-0504-x
28. Chen C-J, Tseng C-C, Yen J-H, et al. ABCG2 contributes to the development of gout and hyperuricemia in a genome-wide association study. *Scientific Reports* 2018;8(1):3137. doi: 10.1038/s41598-018-21425-7
29. Wrigley R, Phipps-Green AJ, Topless RK, et al. Pleiotropic effect of the ABCG2 gene in gout: involvement in serum urate levels and progression from hyperuricemia to gout. *Arthritis Research & Therapy* 2020;22(1):45. doi: 10.1186/s13075-020-2136-z
30. Polimanti R, Gelernter J. ADH1B: From alcoholism, natural selection, and cancer to the human phenome. *Am J Med Genet B Neuropsychiatr Genet* 2018;177(2):113-25. doi: 10.1002/ajmg.b.32523 [published Online First: 03/27]
31. Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 2007;30(1):5-13.
32. Lieber CS, Jones DP, Losowsky MS, et al. Interrelation of uric acid and ethanol metabolism in man. *The Journal of clinical investigation* 1962;41(10):1863-70. doi: 10.1172/JCI104643
33. Macgregor S, Lind PA, Bucholz KK, et al. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Human molecular genetics* 2009;18(3):580-93. doi: 10.1093/hmg/ddn372 [published Online First: 2008/11/11]
34. Yokoyama A, Yokoyama T, Mizukami T, et al. Alcohol Dehydrogenase-1B (rs1229984) and Aldehyde Dehydrogenase-2 (rs671) Genotypes and Alcoholic Ketosis Are Associated with the Serum Uric Acid Level in Japanese Alcoholic Men. *Alcohol and alcoholism (Oxford, Oxfordshire)* 2016;51(3):268-74. doi: 10.1093/alcalc/agv123 [published Online First: 2015/11/07]
35. Sakiyama M, Matsuo H, Akashi A, et al. Independent effects of ADH1B and ALDH2 common dysfunctional variants on gout risk. *Sci Rep* 2017;7(1):2500. doi: 10.1038/s41598-017-02528-z
36. Woodward OM, Köttgen A, Coresh J, et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proceedings of the National Academy of Sciences* 2009;106(25):10338.
37. Major TJ, Dalbeth N, Stahl EA, et al. An update on the genetics of hyperuricaemia and gout. *Nature Reviews Rheumatology* 2018;14(6):341-53. doi: 10.1038/s41584-018-0004-x

38. Schorn C, Janko C, Krenn V, et al. Bonding the foe - NETting neutrophils immobilize the pro-inflammatory monosodium urate crystals. *Front Immunol* 2012;3:376. doi: 10.3389/fimmu.2012.00376
39. Matsuo H, Ichida K, Takada T, et al. Common dysfunctional variants in ABCG2 are a major cause of early-onset gout. *Sci Rep* 2013;3:2014. doi: 10.1038/srep02014
40. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature Communications* 2016;7(1):10023. doi: 10.1038/ncomms10023
41. Gan-Or Z, Bar-Shira A, Gurevich T, et al. Homozygosity for the MTX1 c.184T>A (p.S63T) alteration modifies the age of onset in GBA-associated Parkinson's disease. *neurogenetics* 2011;12(4):325-32. doi: 10.1007/s10048-011-0293-6
42. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009;460(7256):753-57. doi: 10.1038/nature08192 [published Online First: 07/01]
43. Alonso A, Rodríguez LAG, Logroscino G, et al. Gout and risk of Parkinson disease. *Neurology* 2007;69(17):1696. doi: 10.1212/01.wnl.0000279518.10072.df
44. He Q, You Y, Yu L, et al. Uric acid levels in subjects with schizophrenia: A systematic review and meta-analysis. *Psychiatry Research* 2020;292:113305. doi: https://doi.org/10.1016/j.psychres.2020.113305

Figure 1 Study design. Workflow for the discovery and replication analyses. QC, quality control; SD, standard deviation; HWE, Hardy-Weinberg equilibrium; SNPs, single nucleotide polymorphisms; PCs, principal components.

Figure 2. Manhattan plot of the discovery GWAS of gout vs. AH (SU ≥6.0 mg/dL) controls. The y-axis shows –log10 P values ordered by chromosomal position on the x-axis. The horizontal dashed-line represents genome wide significance threshold ($5.0x10^{-8}$).

Figure 3. Area under the receiver operating characteristics (AUROC) curve for the PRS model, demographics model, and combined (demographics + PRS) model.

Figure 4. Scatter plot. A) Comparison of the ORs of lead SNPs for both GWAS: gout vs asymptomatic hyperuricaemia (SU ≥6.0 mg/dL) and gout vs SU <6.0 mg/dL. Black dots represent ORs of the common risk loci of both GWAS, while grey circles represent ORs of additional lead SNPs of the gout vs SU<6.0 mg/dL that were not significant at GWAS level in the gout vs asymptomatic hyperuricaemia GWAS. B) Comparison of the ORs of the thirteen lead SNPs for gout vs asymptomatic hyperuricaemia (defined as SU ≥7.0 mg/dL), compared to the ORs in the GWAS for gout vs SU <7.0 mg/dL. Where several SNPs were present in the same gene, only that with the smallest p value was plotted in this graph. See Supplementary material for names of genes not annotated in the figure.