Correspondence on 'Variants in urate transporters, *ADH1B*, *GCKR* and *MEPE* genes associated with transition from asymptomatic hyperuricaemia to gout: results of the first gout versus asymptomatic hyperuricaemia GWAS in Caucasians using data from the UK Biobank'

Sandoval-Plata *et al*[1] report 13 independently associated genetic variants in the *ABCG2, SLC2A9, SLC22A11, GCKR, MEPE, PPM1K-DT, LOC105377323* and *ADH1B* genes in a genome-wide association study (GWAS) of gout using people with asymptomatic hyperuricaemia as the comparison group. Eight of the genetic variants (representing four of the genes: *ABCG2, MEPE, PPM1K-DT, LOC105377323*) mapped to a 658 kb region on Chr4. This locus is the strongest for gout with the missense *ABCG2* variant (*rs2231142*, p.Gln141Lys), an extremely likely causal variant.[2 3] This variant was the lead-associated variant at the *ABCG2* locus (OR=1.65 (95% CI 1.58 to 1.73), $p=3.3\times10^{-109}$) in the Sandoval-Plata *et al* study.[1]

The reporting of eight independent genetic effects at the *ABCG2* locus stood out. Given the extremely strong statistical evidence for association of *rs2231142*, it is necessary to exclude the possibility that residual linkage disequilibrium (LD) is responsible for the apparently independent association of the seven additional variants. In the GWAS setting, a filtering step can be applied using 'LD clumping' in the PLINK software,[4] whereby an LD threshold within a defined region is set to determine the genetic variant(s) representing the genetic association. If independent genetic effects were observed, the LD clumping is followed by testing putative independent genetic variants for association conditioned on (adjusted for) genotype at the lead genetic variant. In GWAS, putative independent genetic variants should retain statistically significant evidence for association ($p<5\times10^{-8}$) after conditioning. Sandoval-Plata *et al* used LD clumping with an $r^2<0.1$ LD threshold to conclude independent effects at *ABCG2* without follow-up testing by conditional analysis. Selection of the r^2 threshold is important when effect size of a locus is strong and/or when sample size is large, as weak residual LD from a lead variant can cause apparently independent genome-wide significant associations using the LD clumping method if the clumping threshold is set too high.

We performed a GWAS in individuals of European ancestry comparing gout (n=7131) with people with asymptomatic hyperuricaemia (≥7 mg/dL serum urate; n=27 018) in the UK Biobank using an LD clumping cut-off of $r^2<0.1$, a window size of 500 kb and adjusting by age, sex and principle components 1–40. This was followed by conditional analysis. Our analysis differed from Sandoval-Plata *et al* as follows: (1) we used the entire UK Biobank sample set rather than divided for discovery and replication; (2) we used imputed genotypes rather than directly genotyped single-nucleotide polymorphisms (SNPs); and (3) we used a threshold for asymptomatic hyperuricaemia consistent with the definition of pathological hyperuricaemia of serum urate concentration over the saturation threshold[5] rather than >6 mg/dL as used by Sandoval-Plata *et al*. Intermarker LD was quantitated for LD clumping using the same UK Biobank samples as used in our GWAS. Sandoval-Plata *et al* did not specify the reference panel that they used. LD clumping was performed in PLINKv1.9b4.9[4] with a significance threshold $p<5\times10^{-8}$. There were 13 separate SNP effects (three at *ABCG2*, three at *SLC2A9*, two at *SLC22A11/A12* and one at each of *ADH1B, GCKR, MLXIPL, CNBD1, PDX1*). Conditional analysis at each of the *ABCG2* and *SLC2A9* loci, adjusting the logistic regression analysis by the lead SNPs, *rs2231142* and *rs16890979*, respectively (figure 1), revealed evidence for an independent signal at *SLC2A9* approaching genome-wide significance, but not at *ABCG2*. At *ABCG2*, of the seven variants reported as independent signals by Sandoval-Plata *et al*, only *rs13120400* retained nominal significance ($p<0.01$) after conditioning on *rs2231142*, with p value reducing from 1.3×10^{-12} to 8.8×10^{-4}. This variant maps within *ABCG2* ~19 kb from *rs2231142*. Conversely, *rs2231142* retained strong evidence for association with gout ($p<1\times10^{-40}$) after conditioning by each of the seven other *ABCG2* locus variants reported by Sandoval-Plata *et al* (not shown). We therefore conclude that there are no genome-wide significant independent effects associated with gout using asymptomatic controls with hyperuricaemia at the genes *MEPE, PPM1K-DT* and *LOC105377323* within the *ABCG2* locus.

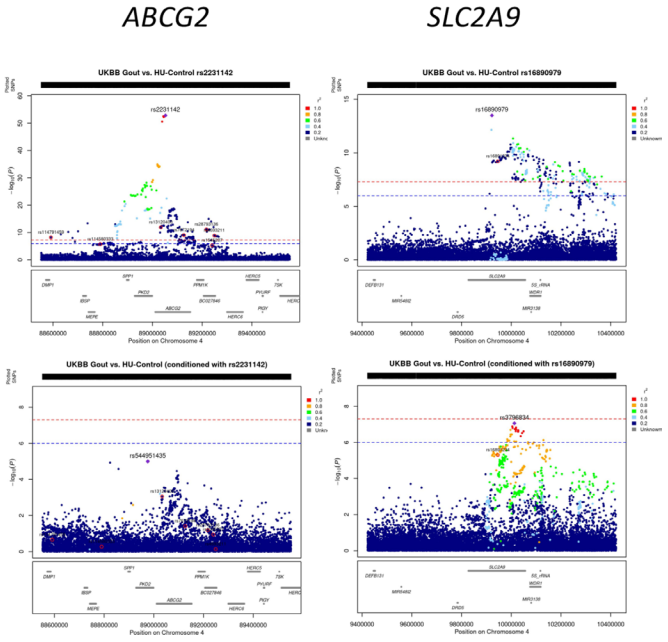

Figure 1 LocusZoom plots of unconditioned (top) and conditioned (bottom) analyses at the *ABCG2* and *SLC2A9* loci. In both instances, conditioning by the lead-associated SNP (*rs2231142* and *rs16890979*, respectively) resulted in the removal of all genome-wide significant signal. Each dot represents an individual SNP with the colour representing the LD with the most associated (lead) SNP in the panel. Variants reported as independent effects by Sandoval-Plata *et al* are indicated. The y-axis is the $-\log_{10}(P)$ for association with gout using hyperuricaemic (HU) controls. LD, linkage disequilibrium; SNP, single-nucleotide polymorphism; UKBB, UK Biobank.

Repeating the LD clumping using a more stringent threshold of $r^2<0.01$, we detected 11 separate SNP effects: two at *ABCG2* (*rs2231142* and *rs148356273*), three at *SLC2A9* (*rs16890979, rs3796834*—$r^2=0.78$ with *rs16891234* reported by Sandoval-Plata *et al* - and *rs6833292*), one at *GCKR* (*rs780093*—$r^2=1.00$ with *rs1260326* reported by Sandoval-Plata *et al*), one at *ADH1B* (*rs1229984*, reported by Sandoval-Plata *et al*), one at *SLC22A11* (*rs7943154*—$r^2=0.93$ with *rs2078267* reported by Sandoval-Plata *et al*), one at *PDX1* (*rs182200427*, not reported by Sandoval-Plata *et al*), one at *MLXIPL* (*rs7805504*, not reported by Sandoval-Plata *et al*), and one downstream from *CNBD1* (*rs181604403*, not reported by Sandoval-Plata *et al*).

We detected all genetic loci detected by Sandoval-Plata *et al* and three additional loci (*PDX1*, *MLXIPL*, *CNBD1*).

Finally, we note that the Sandoval-Plata *et al* study reports three apparently novel genes from their GWAS comparing gout with normouricaemic controls (<6.0 mg/dL)—*MTX1*, *PRSS16* and *AP5B1*. However, both *MTX1* and *AP5B1* are within established serum urate GWAS loci ('*TRIM46*' and '*OVOL1*', respectively). It is important also when determining candidate causal genes not to assume that the closest gene is necessarily causal. At these loci, colocalisation of GWAS signals with expression quantitative trait loci implicates *MUC1*, *GBAP1* and *FAM189B* genes as candidate causal genes at the *TRIM46* locus and *OVOL1-AS1* at the *OVOL1* locus.[6]

Riku Takei,[1] Nicholas A Sumpter,[1] Amanda Phipps-Green,[2] Murray Cadzow,[2] Ruth K Topless,[2] Richard J Reynolds,[1] Tony R Merriman [1]

[1]Division of Clinical Immunology and Rheumatology, The University of Alabama at Birmingham, Birmingham, Alabama, USA
[2]Department of Biochemistry, University of Otago, Dunedin, New Zealand

Correspondence to Dr Tony R Merriman, Division of Clinical Immunology and Rheumatology, The University of Alabama at Birmingham, Birmingham AL 35294, USA; trmerriman@uabmc.edu

Twitter Tony R Merriman @tonymerriman2

Contributors All authors made substantial contributions to the conception or design of the work, or the acquisition, analysis or interpretation of data. All authors were involved in drafting the work or revising it critically for important intellectual content. All authors approved the final approval of the version published.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

To cite Takei R, Sumpter NA, Phipps-Green A, *et al*. *Ann Rheum Dis* 2023;**82**:e174.

Received 12 May 2021
Accepted 14 May 2021
Published Online First 10 June 2021

Linked

► http://dx.doi.org/10.1136/annrheumdis-2021-220785

Ann Rheum Dis 2023;**82**:e174. doi:10.1136/annrheumdis-2021-220769

ORCID iD
Tony R Merriman http://orcid.org/0000-0003-0844-8726

REFERENCES

1 Sandoval-Plata G, Morgan K, Abhishek A. Variants in urate transporters, ADH1B, GCKR and MEPE genes associate with transition from asymptomatic hyperuricaemia to gout: results of the first gout versus asymptomatic hyperuricaemia GWAS in Caucasians using data from the UK Biobank. *Ann Rheum Dis* 2021;80:1220–6.
2 Cleophas MC, Joosten LA, Stamp LK, *et al*. ABCG2 polymorphisms in gout: insights into disease susceptibility and treatment approaches. *Pharmgenomics Pers Med* 2017;10:129–42.
3 Hoque KM, Dixon EE, Lewis RM, *et al*. The ABCG2 Q141K hyperuricemia and gout associated variant illuminates the physiology of human urate excretion. *Nat Commun* 2020;11:2767.
4 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
5 Bursill D, Taylor WJ, Terkeltaub R, *et al*. Gout, hyperuricaemia and Crystal-Associated disease network (G-CAN) consensus statement regarding labels and definitions of disease states of gout. *Ann Rheum Dis* 2019;78:1592–600.
6 Boocock J, Leask M, Okada Y, *et al*. Genomic dissection of 43 serum urate-associated loci provides multiple insights into molecular mechanisms of urate control. *Hum Mol Genet* 2020;29:923–43.