MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank

Xue Li,¹ Xiangrui Meng,¹ Athina Spiliopoulou,² Maria Timofeeva,³ Wei-Qi Wei,⁴ Aliya Gifford,⁴ Xia Shen,¹,⁶ Yazhou He,¹,⁶ Tim Varley,⁷ Paul McKeigue,² Ioanna Tzoulaki,⁸⁹¹⁰ Alan F Wright,¹¹ Peter Joshi,¹ Joshua C Denny,¹ Harry Campbell,¹ Evropi Theodoratou¹,¹²

ABSTRACT

Objectives  We aimed to investigate the role of serum uric acid (SUA) level in a broad spectrum of disease outcomes using data for 120 091 individuals from UK Biobank.

Methods  We performed a phenome-wide association study (PheWAS) to identify disease outcomes associated with SUA genetic risk loci. We then implemented conventional Mendelian randomisation (MR) analysis to investigate the causal relevance between SUA level and disease outcomes identified from PheWAS. We next applied MR Egger analysis to detect and account for potential pleiotropy, which conventional MR analysis might mistake for causality, and used the HEI (heterogeneity in dependent instruments) test to remove cross-phenotype associations that were likely due to genetic linkage.

Results  Our PheWAS identified 25 disease groups/ outcomes associated with SUA genetic risk loci after multiple testing correction (P<8.57e-05). Our conventional MR analysis implicated a causal role of SUA level in three disease groups: inflammatory polyarthropathies (OR=1.22, 95% CI 1.11 to 1.34), hypertensive disease (OR=1.08, 95% CI 1.03 to 1.14) and disorders of metabolism (OR=1.07, 95% CI 1.01 to 1.14); and four disease outcomes: gout (OR=4.88, 95% CI 3.91 to 6.09), essential hypertension (OR=1.08, 95% CI 1.03 to 1.14), myocardial infarction (OR=1.16, 95% CI 1.03 to 1.30) and colorectal cancer (OR=1.41, 95% CI 1.05 to 1.89). After balancing pleiotropic effects in MR Egger analysis, only gout and its encompassing disease group of inflammatory polyarthropathies were considered to be causally associated with SUA level. Our analysis highlighted a locus (ATXN2/S2HB3) that may influence SUA level and multiple cardiovascular and autoimmune diseases via pleiotropy.

Conclusions  Elevated SUA level is convincing to cause gout and inflammatory polyarthropathies, and might act as a marker for the wider range of diseases with which it associates. Our findings support further investigation on the clinical relevance of SUA level with cardiovascular, metabolic, autoimmune and respiratory diseases.

INTRODUCTION

Uric acid (UA) is the end product of the exogenous and endogenous purine metabolism, catalysed by the action of xanthine oxidase.¹ Due to the evolved loss of uricase enzyme, humans are unable to convert UA into highly soluble compounds, leaving urate circulating in the blood and resulting in a high basal level of serum uric acid (SUA).² The prevalence rate of hyperuricaemia (elevated SUA level >7.0 mg/dL) is in the range of 5%–25% across different countries.³ A progressively rising trend of hyperuricaemia prevalence has been observed worldwide.⁴ Concernedly, hyperuricaemia is thought to inflict multiple clinical consequences, which is believed to be causally related to gout and suggestively associated with a number of prevalent health conditions, such as cardiovascular and metabolic diseases.⁵–⁸

Our recently published umbrella review presented a comprehensive overview of the breadth of disease outcomes related to SUA level by incorporating evidence from multiple sources.⁹ A large number of disease outcomes were reported to be associated with SUA level in observational studies, covering a wide range of diseases, including cardiovascular disease, metabolic syndrome, diabetes, cancer and neurological disorders. However, evidence as to whether these associations are actually causal is not yet well developed, given that observational associations are susceptible to a variety of biases, confounding and/or reverse causality. Although results from randomised controlled trials (RCTs) have provided some evidence about the beneficial effects of SUA-lowering therapy on some intermediate traits or biomarkers (eg, blood pressure, endothelial function, serum creatinine), there remains a lack of RCTs focusing on the more important clinical disease endpoints.¹⁰–¹² A number of Mendelian randomisation (MR) studies, using the genetic variants influencing SUA level as instruments, provide alternative evidence to distinguish causal from non-causal associations. However, these MR studies examined a limited set of disease outcomes and were not able to detect moderate effect size due to limited power.¹³–¹⁵ Increasing sample size and the range of outcomes in an enlarged MR study thus offers the prospect of deeper and wider insight into the causal role of SUA.

MR analysis is typically hypothesis-driven based on prior knowledge to specify the outcome to be examined in relation to the exposure of
interest. Traditionally, only one (or a limited number) association between the exposure and one (or a few) predefined outcome(s) is tested in an MR study. Recently, phenome-wide Mendelian randomisation (MR-PheWAS) analysis has been proposed by integrating the phenome-wide association study (PheWAS) and MR method to build a hypothesis-searching approach, which aims to explore potential causal relationships between an exposure (using genetic instruments as proxies) and a range of phenome-wide disease outcomes in a high-throughput manner. This approach is effective in evaluating or replicating the associations reported in observational studies, as well as discovering new relationships and generating new hypotheses on the genetic architecture shared by the related phenotypes. With its wealth of genotypic and phenotypic data collected in very large numbers, the UK Biobank study provides an excellent opportunity to explore the causal role of SUA level across a broad spectrum of disease outcomes. In this study, we performed an MR-PheWAS in UK Biobank database to discover disease outcomes related to genetic variations of SUA level and to investigate if any association is causal.

METHODS

UK Biobank data

The UK Biobank is a large-scale, population-based, prospective cohort that enrolled over 500,000 participants aged 40–69 years. The recruited participants provided a wide range of self-reported baseline information. Blood samples were collected for biochemical tests and genotyping. Their national health records have been linked with the baseline and genotypic data for longitudinal follow-up. Genotypic and phenotypic data used in this study were obtained from UK Biobank under an approved data request application (application ID: 10775).

Genotyping and quality control

Genotyping, quality control and genetic imputation were performed by the UK Biobank team prior to the interim release of genotypic data for 150,000 participants. The procedure of genotyping and quality control is presented in detail at https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_qc.pdf. We used the field variables made available by the UK Biobank for quality control to exclude the samples that had high missingness or heterozygosity, outlying short runs of homozygosity, and sex mismatch (see online supplementary table S1). We constrained our analyses to participants who were self-reported British and confirmed to be Caucasians based on the genetic principal component analysis performed by the UK Biobank. The quality control process generated a genotypic data set output with 120,091 individuals included in the current analysis.

Phenotyping and mapping ICD-10 or ICD-9 to phecode

We focused on phenotypes in relation to diagnostic disease outcomes. We analysed two phenotypic data sets (inpatient hospital episode records and cancer registry data) in the UK Biobank using the phecode schema (see online supplementary text for phenotyping and mapping process). The coding for clinical diagnoses in these data sets followed the WHO’s International Classification of Diseases (ICD) coding systems, but used different ICD versions (ICD-10 or ICD-9) according to the date of record. We included both ICD-10 and ICD-9 codes to define the case and control groups. Since cancer registry data overlapped with the cancer diagnosis in inpatient hospital records, we pooled the cancer registry data into the hospital episode data as a complement to the cancer diagnosis.

Statistical analysis

The statistical analysis included three main steps: first, we performed a PheWAS to identify disease outcomes that were associated with genetic risk loci of SUA level; second, we performed MR analysis by using both the inverse-variance-weighted (IVW) method and MR Egger approach to explore causal relationship for identified PheWAS associations; and third, we applied HEIDI (heterogeneity in dependent instruments) test to exclude the cross-phenotype associations caused by genetic linkage.

Genetic instruments

We selected 31 SUA-associated single nucleotide polymorphisms (SNPs) as genetic instruments (see online supplementary table S2), which were previously reported to be independently associated with SUA level in genome-wide association studies (GWAS). We obtained the SNP effect on SUA level from the largest GWAS performed in European population. The overall proportion of variance ($R^2$) of SUA level explained by the selected genetic instruments) was estimated to be close to 7.0%.

Phenome-wide association analysis

In phenome-wide analysis, we used 31 SUA-associated SNPs as genetic instruments individually to scan across a wide range of disease outcomes defined by the phecode system. With the PheWAS algorithm, a series of case-control tests was performed: (1) the case group was generated by including patients with the tested phecode; (2) participants were assigned to the control group based on the absence of both the tested phecode and related phecodes (patients who had the parent, child or sibling phecodes of the tested phecode were excluded from the control group); and (3) to ensure statistical power, analysis was only performed for phecode with no less than 200 cases. This minimum number of cases was suggested based on a simulation of power estimates for PheWAS analysis. We used logistic regression to test the associations between 31 individual genetic instruments (assuming an additive genetic model) and each phecode (number of cases $\geq200$) after adjusting for multiple covariates, including sex, body mass index (BMI), age, assessment centre and the principal components. Considering many phecodes were not independent, we used the false discovery rate (FDR) method to account for multiple testing.

MR IVW, MR Egger and HEIDI test

We then explored the identified PheWAS associations in three possible scenarios (see online supplementary figure S1): (1) causality: the observed association was causal (through the SUA pathway); (2) pleiotropy: the observed association was due to pleiotropic effect of one causal variant (ie, linked to SUA level and the particular disease outcome through pleiotropy); and (3) genetic linkage: the observed association was caused by the linkage disequilibrium (LD) between two distinct causal variants, with one affecting SUA level and the other affecting the disease outcome.

MR IVW

To explore if there was any causal effect on identified disease outcomes, we performed the conventional MR analysis by pooling the individual effect of each SNP using the IVW method.
to estimate the overall causal effect (see online supplementary text). 30

MR Egger
We then performed MR Egger to attempt to correct for any potential pleiotropic effect in the causal estimates. This approach is applied to balance the pleiotropic effects derived from multiple genetic instruments (see online supplementary text). 21

HEIDI test
We calculated HEIDI statistics for the SUA genetic loci that were associated with more than one disease outcome. This test was to examine if the cross-phenotype association was due to genetic linkage (see online supplementary text). 24

Sex stratification analysis
To account for any sex difference, we performed PheWAS and MR analyses in men and women separately. The sex-specific effects of SNPs on SUA level (see online supplementary table S2) were taken from the summary-level GWAS data provided by Köttgen et al. 22

RESULTS
A total of 120,091 UK Biobank participants were included in the analysis, consisting of 56,845 men and 63,246 women with a mean age of 64.86 years in 2016 (SD of 7.95) (see online supplementary table S3). Within phenotypic data sets, we identified 684,324 hospital episodes and 23,174 cancer registration records, which included 7990 unique ICD-10 codes and 1998 unique ICD-9 codes. After mapping diagnostic ICD-10 or ICD-9 codes to phenotypes, the phenotypic data consisted of 1807 distinct phenotypes. After filtering out disease outcomes with low prevalence (number of cases <200), 568 phenotypes (median number of cases=694 (range: 200–39,777) were included in PheWAS analysis. These 568 phenotypes were classified into 17 broadly related disease categories (table 1). We noted that the distribution of phenotypes examined was skewed across the different disease categories (see online supplementary figure S2), in which a large number of disease phenotypes were included in digestive, circulatory, endocrine and metabolic systems, but some disease categories, for example congenital anomalies, were not well represented in the study population.

Phenome-wide association analysis
The PheWAS analysis performed 17,608 case–control tests, leading to an adjusted significance threshold of P<8.57e-05 corresponding to an FDR of q<0.05 to account for the multiple testing. A total of 27 pairs of genotype–phenotype associations passed the significance threshold of FDR correction (P<8.57e-05) in the overall PheWAS analysis with adjustment for covariates (table 2). Results of PheWAS without adjustment for BMI are shown in online supplementary table S4. The sex-stratified PheWAS analysis identified 10 pairs of genotype–phenotype association in men and 10 pairs of genotype–phenotype association in women (see online supplementary table S5). When compared with the overall PheWAS analysis, five new pairs of association were identified from the sex-stratified PheWAS analysis (see online supplementary table S5).

These identified genotype–phenotype associations were distributed across 15 SUA genetic loci, of which 5 loci were associated with more than one disease outcome: rs653178 in ATXN2/SH2B3 locus (number of disease outcomes: noutcomes = 10), rs1165151 in SLC17A3 locus (noutcomes = 3), rs1260326 in GCKR locus (noutcomes = 5), rs2231142 in ABCG2 locus (noutcomes = 4) and rs2079742 in BCAS3 locus (noutcomes = 2). Of note, six disease outcomes shared genetic associations with SUA level at more than one locus: gout (number of loci: nloci = 3), inflammatory polyarthropathies (nloci = 2), disorders of iron metabolism (nloci = 2), coeliac disease (nloci = 2), hypertensive disease (nloci = 2) and essential hypertension (nloci = 2).

In summary, the PheWAS analyses identified 25 unique disease groups/outcomes (corresponding to 25 unique phenocodes) that shared genetic risk loci with SUA level, which included 9 disease groups (inflammatory polyarthropathies, hypertensive disease, circulatory disease, disorders of metabolism, disorders of thyroid, other diseases of respiratory system, disorder of skin and subcutaneous tissue, benign neoplasm of digestive system, and complications of labour and delivery) and 16 specific disease

<table>
<thead>
<tr>
<th>Disease categories</th>
<th>Phenotypes (n)</th>
<th>Cases (n)</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circulatory system</td>
<td>61</td>
<td>221</td>
<td>665</td>
<td>2937</td>
<td>39142</td>
<td></td>
</tr>
<tr>
<td>Congenital anomalies</td>
<td>6</td>
<td>206</td>
<td>265</td>
<td>302</td>
<td>522</td>
<td></td>
</tr>
<tr>
<td>Dermatological diseases</td>
<td>24</td>
<td>201</td>
<td>706</td>
<td>2736</td>
<td>32738</td>
<td></td>
</tr>
<tr>
<td>Diseases in sense organs</td>
<td>34</td>
<td>201</td>
<td>425</td>
<td>1216</td>
<td>11306</td>
<td></td>
</tr>
<tr>
<td>Infectious diseases</td>
<td>73</td>
<td>201</td>
<td>949</td>
<td>2176</td>
<td>23129</td>
<td></td>
</tr>
<tr>
<td>Neoplasms</td>
<td>59</td>
<td>203</td>
<td>763</td>
<td>1916</td>
<td>30101</td>
<td></td>
</tr>
<tr>
<td>Endocrine and metabolic diseases</td>
<td>25</td>
<td>229</td>
<td>492</td>
<td>2304</td>
<td>13592</td>
<td></td>
</tr>
<tr>
<td>Haematopoietic diseases</td>
<td>10</td>
<td>205</td>
<td>1187</td>
<td>1600</td>
<td>3669</td>
<td></td>
</tr>
<tr>
<td>Neurological diseases</td>
<td>21</td>
<td>229</td>
<td>452</td>
<td>1282</td>
<td>10828</td>
<td></td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>38</td>
<td>219</td>
<td>712</td>
<td>1713</td>
<td>19238</td>
<td></td>
</tr>
<tr>
<td>Mental disorders</td>
<td>18</td>
<td>205</td>
<td>673</td>
<td>1926</td>
<td>8942</td>
<td></td>
</tr>
<tr>
<td>Genitourinary diseases</td>
<td>77</td>
<td>200</td>
<td>666</td>
<td>1606</td>
<td>29859</td>
<td></td>
</tr>
<tr>
<td>Pregnancy complications</td>
<td>11</td>
<td>227</td>
<td>360</td>
<td>707</td>
<td>2531</td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal diseases</td>
<td>44</td>
<td>263</td>
<td>1076</td>
<td>2482</td>
<td>21822</td>
<td></td>
</tr>
<tr>
<td>Clinical symptoms</td>
<td>14</td>
<td>267</td>
<td>1237</td>
<td>2570</td>
<td>12287</td>
<td></td>
</tr>
<tr>
<td>Injuries and poisonings</td>
<td>37</td>
<td>211</td>
<td>589</td>
<td>911</td>
<td>4842</td>
<td></td>
</tr>
</tbody>
</table>
Clinical and epidemiological research

Table 2 Genotype–phenotype associations identified from PheWAS after correcting multiple testing by FDR (P<8.57e-05)

<table>
<thead>
<tr>
<th>Phencode</th>
<th>Description</th>
<th>SNP_risk allele*</th>
<th>Allele frequency</th>
<th>Total (n)</th>
<th>Cases (n)</th>
<th>OR (95% CI)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>274.1</td>
<td>Gout</td>
<td>rs2231142</td>
<td>0.11</td>
<td>119555</td>
<td>1003</td>
<td>1.89 (1.69 to 2.12)</td>
<td>5.41e-28</td>
</tr>
<tr>
<td>275.1</td>
<td>Disorders of iron metabolism</td>
<td>rs1165151_G</td>
<td>0.45</td>
<td>119063</td>
<td>205</td>
<td>3.56 (2.78 to 4.56)</td>
<td>1.41e-23</td>
</tr>
<tr>
<td>244.4</td>
<td>Hypothyroidism</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>118821</td>
<td>4146</td>
<td>1.21 (1.16 to 1.27)</td>
<td>3.90e-17</td>
</tr>
<tr>
<td>246</td>
<td>Disorders of thyroid</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119601</td>
<td>4926</td>
<td>1.18 (1.14 to 1.23)</td>
<td>8.82e-16</td>
</tr>
<tr>
<td>274.1</td>
<td>Gout</td>
<td>rs12498742_A</td>
<td>0.23</td>
<td>119860</td>
<td>1002</td>
<td>1.54 (1.37 to 1.74)</td>
<td>7.94e-13</td>
</tr>
<tr>
<td>275.1</td>
<td>Disorders of iron metabolism</td>
<td>rs742132_A</td>
<td>0.29</td>
<td>119271</td>
<td>205</td>
<td>2.80 (2.10 to 3.74)</td>
<td>3.13e-12</td>
</tr>
<tr>
<td>401</td>
<td>Hypertensive disease</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119762</td>
<td>23634</td>
<td>1.06 (1.04 to 1.09)</td>
<td>1.68e-08</td>
</tr>
<tr>
<td>401.1</td>
<td>Essential hypertension</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119688</td>
<td>23560</td>
<td>1.06 (1.04 to 1.09)</td>
<td>2.00e-08</td>
</tr>
<tr>
<td>411.4</td>
<td>Coronary atherosclerosis</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119460</td>
<td>9526</td>
<td>1.09 (1.05 to 1.12)</td>
<td>1.27e-07</td>
</tr>
<tr>
<td>411</td>
<td>Ischaemic heart disease</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119401</td>
<td>9467</td>
<td>1.09 (1.05 to 1.12)</td>
<td>1.33e-07</td>
</tr>
<tr>
<td>211</td>
<td>Benign neoplasm of digestive system</td>
<td>rs11624341_C</td>
<td>0.43</td>
<td>117030</td>
<td>1504</td>
<td>0.83 (0.77 to 0.89)</td>
<td>2.41e-07</td>
</tr>
<tr>
<td>274.1</td>
<td>Gout</td>
<td>rs1260326_T</td>
<td>0.39</td>
<td>119555</td>
<td>1003</td>
<td>1.26 (1.15 to 1.38)</td>
<td>3.86e-07</td>
</tr>
<tr>
<td>459.9</td>
<td>Circulatory disease</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>119677</td>
<td>39142</td>
<td>1.05 (1.03 to 1.06)</td>
<td>2.24e-06</td>
</tr>
<tr>
<td>411.2</td>
<td>Myocardial infarction</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>113559</td>
<td>3625</td>
<td>1.12 (1.07 to 1.18)</td>
<td>2.80e-06</td>
</tr>
<tr>
<td>557.1</td>
<td>Coeliac disease</td>
<td>rs1165151_G</td>
<td>0.45</td>
<td>99782</td>
<td>549</td>
<td>1.33 (1.18 to 1.51)</td>
<td>4.30e-06</td>
</tr>
<tr>
<td>557.1</td>
<td>Coeliac disease</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>99965</td>
<td>550</td>
<td>1.31 (1.16 to 1.48)</td>
<td>9.28e-06</td>
</tr>
<tr>
<td>427.2</td>
<td>Atrial fibrillation and flutter</td>
<td>rs6598541_A</td>
<td>0.35</td>
<td>113261</td>
<td>4333</td>
<td>1.11 (1.06 to 1.16)</td>
<td>9.92e-06</td>
</tr>
<tr>
<td>960</td>
<td>Poisoning by antibiotics</td>
<td>rs1165151_G</td>
<td>0.45</td>
<td>112343</td>
<td>1027</td>
<td>0.82 (0.75 to 0.90)</td>
<td>1.22e-05</td>
</tr>
<tr>
<td>535</td>
<td>Gastritis and duodenitis</td>
<td>rs478607_G</td>
<td>0.15</td>
<td>115386</td>
<td>5233</td>
<td>1.12 (1.07 to 1.19)</td>
<td>1.34e-05</td>
</tr>
<tr>
<td>411.3</td>
<td>Angina pectoris</td>
<td>rs653178_C</td>
<td>0.48</td>
<td>114967</td>
<td>5033</td>
<td>1.09 (1.05 to 1.14)</td>
<td>3.01e-05</td>
</tr>
<tr>
<td>669</td>
<td>Complications of labour and delivery</td>
<td>rs729761_G</td>
<td>0.28</td>
<td>113240</td>
<td>2376</td>
<td>1.17 (1.09 to 1.26)</td>
<td>3.78e-05</td>
</tr>
<tr>
<td>272.11</td>
<td>Hypercholesterolaemia</td>
<td>rs1260326_T</td>
<td>0.39</td>
<td>118921</td>
<td>10201</td>
<td>1.07 (1.03 to 1.10)</td>
<td>3.82e-05</td>
</tr>
<tr>
<td>366</td>
<td>Cataract</td>
<td>rs6770152_G</td>
<td>0.43</td>
<td>116218</td>
<td>4567</td>
<td>1.09 (1.05 to 1.14)</td>
<td>4.14e-05</td>
</tr>
<tr>
<td>471</td>
<td>Nasal polyps</td>
<td>rs10821905_A</td>
<td>0.17</td>
<td>112745</td>
<td>983</td>
<td>1.26 (1.13 to 1.40)</td>
<td>4.61e-05</td>
</tr>
<tr>
<td>454.1</td>
<td>Varicose veins of lower extremity</td>
<td>rs2231142_T</td>
<td>0.11</td>
<td>111390</td>
<td>3204</td>
<td>0.84 (0.78 to 0.92)</td>
<td>5.79e-05</td>
</tr>
<tr>
<td>401</td>
<td>Hypertensive disease</td>
<td>rs2079742</td>
<td>0.13</td>
<td>115659</td>
<td>22832</td>
<td>1.07 (1.03 to 1.10)</td>
<td>7.00e-05</td>
</tr>
<tr>
<td>401.1</td>
<td>Essential hypertension</td>
<td>rs2079742_T</td>
<td>0.13</td>
<td>115588</td>
<td>22761</td>
<td>1.07 (1.03 to 1.10)</td>
<td>7.02e-05</td>
</tr>
</tbody>
</table>

*Effect allele was harmonised to be the SUA-raising allele defined by Köttgen et al.*

†Significance threshold of P<8.57e-05 corresponds to an FDR of q<0.05 after correcting the multiple testing.

FDR, false discovery rate; PheWAS, phenome-wide association study; SUA, serum uric acid.

To explore and correct for any possible pleiotropic effect of multiple instruments, we then conducted the MR Egger analysis (table 3). After balancing out the potential pleiotropic effects, the putative causal link of SUA level with gout (OR=4.58, 95% CI 2.72 to 7.72, P_effect=1.76e-06) and its umbrella disease group, inflammatory polyarthropathies (OR=1.15, 95% CI 1.01 to 1.31, P_effect=0.03), remained statistically significant and there was no indication of unbalanced pleiotropy (P_pleiotropy=0.73 and P_pleiotropy=0.23, respectively). The putative causal effect of SUA level on the other five disease groups/outcomes was not statistically significant in the MR Egger model. The causal effects of each individual SNPs on these seven disease groups/outcomes are shown in online supplementary figures S3–S9. Unbalanced pleiotropy was observed for essential hypertension (P_pleiotropy=0.001) and its umbrella disease group, hypertensive disease (P_pleiotropy=0.001). For myocardial infarction, coeliac disease and disorders of metabolism, the putative causal effect was not statistically significant in the MR Egger model (P_effect=0.75, P_effect=0.41 and P_effect=0.80, respectively), although there was no evidence of unbalanced pleiotropy (P_pleiotropy=0.13, P_pleiotropy=0.75 and P_pleiotropy=0.18, respectively). The results of the sex-stratified MR IVW are presented in online supplementary table S7.

Finally, to distinguish the genotype–phenotype association of pleiotropy from LD, the HEIDI test was performed for the five genetic loci (rs653178 in ATXN2/SH2B3 locus, rs1165151...
in SLC17A3 locus, rs1260326 in GCKR locus, rs2231142 in ABCG2 locus and rs2079742 in BCAS3 locus) that were associated with multiple disease outcomes in the PhEwas analysis (see online supplementary figures S10–S14). Based on the HEIDI test, we identified 14 disease outcomes that were associated with the SUA genetic risk loci due to pleiotropy (with \( P_{\text{HEIDI}} > 0.05 \)). The strongest pleiotropic locus was the ATXN2/SH2B3, where three SNPs (rs653178, rs4766578 and rs3184504) in near-complete LD (\( r^2 = 0.99 \)) were tagged as the lead SNPs associated with multiple disease outcomes in the PhEwas analysis (see online supplementary figure S13). Similarly, for the associations between the SLC17A3 locus (rs1165151) and coeliac disease (\( P_{\text{HEIDI}} = 6.51 \times 10^{-16} \)) (see online supplementary figure S13), and the GCKR locus (rs1260326) and hypercholesterolaemia (\( P_{\text{HEIDI}} = 3.27 \times 10^{-11} \)) (see online supplementary figure S14), the pattern of shared genetic association was more consistent with a genetic linkage model, and the SNP with the smallest \( P \) value was tagged as an index of the distinct causal variant affecting the examined disease outcome.

**DISCUSSION**

In PhEwas analysis by using SUA-associated SNPs as genetic instruments, we replicated the findings of the largest GWAS performed by Kötting and the findings of the most recent candidate gene-based association study conducted in UK Biobank, which indicated that two SUA-related SNPs (rs12498742 in SLC2A9 locus and rs2231142 in ABCG2 locus) are significantly associated with gout at GWAS \( P \) value threshold (\( P < 5.0 \times 10^{-8} \)).

We conducted a conventional MR analysis (using the IVW method) and an MR Egger analysis, which accounts for potential pleiotropic effects, to investigate potential causal links with SUA level. These both confirmed potential causal effects of SUA level on gout and inflammatory polyarthropathies. The latter category represents the disease group term that includes gout, and thus this finding may just reflect the causal role of SUA in gout. However, this study cannot exclude a causal association between SUA and other inflammatory polyarthropathies, and this may be worth further study. Given that many comorbidities...
are commonly reported in patients with gout, it is of interest to consider the evidence for SUA sharing genetic risk loci with some of these diseases, such as cardiovascular/metabolic diseases and autoimmune disorders, and the evidence for a possible causal role for SUA in these conditions.

Overall, we identified 32 pairs of genotype–phenotype associations, which covered a wide range of phenotypic categories including endocrine/metabolic diseases, cardiovascular diseases and autoimmune disorders. Our PheWAS analysis replicated 14 pairs of previously known genotype–phenotype (or closely related phenotypic groups) associations reported in the GWAS Catalog (see online supplementary table S2 and table 2). For example, rs653178 (ATXN2/S2HB3 locus) was previously reported to be associated with diastolic blood pressure, myocardial infarction, peripheral artery disease, coeliac disease and serum thyroid peroxidase antibody levels. In our PheWAS, this SNP was statistically significantly associated with the same phenotypes (ie, coeliac disease, myocardial infarction) or similar phenotypic groups (ie, hypertension, circulatory and heart diseases, hypothyroidism and other disorders of thyroid). We also identified 18 novel genotype–phenotype associations (at the PheWAS threshold of P<8.57e-05), of which the association between rs1165151 (SLC17A3 locus) and disorders of iron metabolism had the smallest P value (P=1.23e-19).

We performed conventional MR analysis, using the IVW method, to investigate whether there was a potential causal link between SUA level and the 25 unique disease groups/outcomes identified from PheWAS. The results of MR IVW analysis suggested a potential causal effect of SUA level on three disease groups, including inflammatory polyarthropathies (as noted above), essential hypertension, myocardial infarction and peripheral artery disease, although the MR Egger analysis showed a negative (null) effect on essential hypertension and heart disease, which reversed the sign of the overall MR findings were concordant with results from observational studies, which indicated that the relationship between SUA level and cardiovascular disease was particularly strong in women, especially for heart disease. Although these putative causal associations specific to women were not verified by MR Egger, this may be due to the decreased statistical power of MR Egger (and a higher risk of type 2 error). The biological mechanism that can lead the association of SUA level with cardiovascular disease to be more pronounced in women than in men remains a matter for further investigation.

We also found that several PheWAS associations were likely driven by LD. For instance, the outstanding PheWAS association between disorders of iron metabolism and the SNP rs1165151 in SLC17A3 locus was not consistent with a pleiotropic model, and further examination found the SUA-associated SNP rs1165151 was located in LD (r²=0.24), with the rs17342717 variant in SLC17A1 locus, which was strongly associated with disorders of iron metabolism (P=1.69e-129). This SNP (rs17342717) is also associated with red blood cell traits and serum iron levels in previous GWAS. We suggest that the implications of these findings have wider relevance for PheWAS studies. Typically, associations of a single SNP with multiple phenotypes were claimed to be due to pleiotropy in previous PheWAS. However, as PheWAS focused on single variant without considering the correlations between SNPs, we would suggest that an additional examination of LD is necessary when we identify pleiotropic links.

In contrast, the pattern of shared regional genetic associations of SUA level with multiple disease outcomes at ATXN2/S2HB3 locus was more consistent with a pleiotropic model, where we interpreted this locus influenced a cluster of cardiovascular diseases and autoimmune disorders. However within the ATXN2/S2HB3 locus, there are three leading SNPs (rs653178, rs4766578 and rs3184504) in high LD (r²=0.99). In this case, we were unable to provide an indication of whether the observed associations are due to pleiotropy or genetic linkage, as it was difficult to infer the causal variant. Although SNP rs653178 was reported as the lead variant influencing SUA level at this locus in GWAS, the potential biological mechanism underlying this effect is unclear. Furthermore, although the implication of the rs653178 on the regulation of blood pressure, cardiovascular diseases and coeliac disease has been suggested by a few GWAS, a clear biological explanation for this role could not be demonstrated. Evidence from the functional follow-up of the S2HB3 gene indicated that rs3184504 may be the causal variant, as the S2HB3 gene encodes one of the S2HB family proteins, which have a diverse physiological roles on haematopoiesis, immune response and signalling, and variation in rs3184504 may introduce a new phosphorylation site affecting the function of the S2HB protein. We believe that further uncovering of the biological functions of this pleiotropic locus (eg, gene function follow-up, expression quantitative trait loci analysis) might be helpful to understand the complex underlying relationship of SUA level with cardiovascular and autoimmune diseases.

The sex-stratified MR IVW analysis identified that unspecified diseases in respiratory system were potentially causally linked with SUA level in women (with the MR Egger analysis showing a consistent causal effect). This finding is consistent with recently published experimental studies, which demonstrated that human airway epithelial cells and lung tissue expressed a functional UA production/secretion system and UA was crucial in mediating the development of allergic airway diseases and regulating the antigen-specific T cell proliferation. It was also speculated that fine, inhaled particulate matter can induce increased UA production in the human airway, which may contribute to
 allergic sensitisation and asthma pathogenesis. Evidence from other epidemiological studies suggested that high SUA level was associated with low lung function and high risk of respiratory symptoms and chronic obstructive pulmonary disease, but the direct causal relationship has not been established. Further investigation may be worth to explore the clinical relevance of SUA level in lung health and respiratory diseases.

Key strengths of our study included its potential to make novel discoveries in genotype–phenotype associations and to identify novel cross-phenotype associations, possibly reflecting common aetiology or causal mechanisms. Unlike the genome, for which genetic structure can be measured by reliable biological techniques, the definition of phenotype varies across studies. Current published PheWAS have been limited primarily to billing ICD-9-clinical modification (CM) to phenotype system, and the method for aggregating ICD-9-CM codes into phecodes has proven to be valuable in previous PheWAS studies. Our work broadened the utility of phenotype system and illustrated the process of adopting phenotype system in the updated ICD-10 version to define the phenotype framework. Our mapping process revealed some potential shortcomings of the current phenotype system (eg, the ICD-10 codes involving the personal or family history were missing elements in the phenotype system), which should be improved as a future undertaking. Recent methodological applications (eg, tree-structured phenotypic model (TreeWAS)) can be applied in future PheWAS analyses. As we were preparing the manuscript for submission, a web resource within UK Biobank, the GeneATLAS, was released in the bioRxiv (prior to peer review). We checked our PheWAS findings in this database, but only 10 of the 31 SUA-related SNPs were included in their database (and associations with some disease outcomes were replicated for these SNPs). We focused on the causal relationships between SUA level and binary disease outcomes in MR analyses, and these findings were complementary to MR estimates of urate archived in the MR-Base database (http://evm.rbase.org/), which mainly focused on quantitative traits.

On the other hand, our analysis was limited to phenotypes with no less than 200 cases; therefore, diseases with relatively low prevalence were not analysed. As the UK Biobank grows, we expect to perform PheWAS and MR analyses for more low prevalence were not analysed. As the UK Biobank grows, relationships with SUA level are much controversial, such as estimates of urate archived in the MR-Base database (http://evm.rbase.org/), which mainly focused on quantitative traits.

Acknowledgements This research has been conducted using the UK Biobank resource (approval number: 10775). We wish to kindly thank all participants from the UK Biobank.

Contributors ET and HC conceived the study, and XL contributed to the study design. XL performed the data analysis. XL, XM, W-QW, AG, JCD and TV contributed to the mapping of ICD-10/9 codes to phecode. XL wrote the manuscript. All authors critically reviewed the manuscript and contributed important intellectual content. All authors have read and approved the final manuscript as submitted.

Funding XL and XM are supported by the China Scholarship Council. ET is supported by a CRUK Career Development Fellowship. W-QW is supported by the NIH grant R01 HL133786.

Competing interests None declared.

Author affiliations
1Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
2Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
3Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
4Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
5Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
6West China School of Medicine, West China Hospital, Sichuan University, Sichuan, China
7Public Health and Intelligence, NHS National Services Scotland, Edinburgh, UK
8Department Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
9MRc-PHE Centre for Environment, School of Public Health, Imperial College London, London, UK
10Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
11Medical Research Council Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
12Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
REFERENCES


