nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0504-x

Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels

Elevated serum urate levels cause gout and correlate with cardiometabolic diseases via poorly understood mechanisms. We performed a trans-ancestry genome-wide association study of serum urate in 457,690 individuals, identifying 183 loci (147 previously unknown) that improve the prediction of gout in an independent cohort of 334,880 individuals. Serum urate showed significant genetic correlations with many cardiometabolic traits, with genetic causality analyses supporting a substantial role for pleiotropy. Enrichment analysis, fine-mapping of urate-associated loci and colocalization with gene expression in 47 tissues implicated the kidney and liver as the main target organs and prioritized potentially causal genes and variants, including the transcriptional master regulators in the liver and kidney, *HNF1A* and *HNF4A*. Experimental validation showed that HNF4A transactivated the promoter of *ABCG2*, encoding a major urate transporter, in kidney cells, and that HNF4A p.Thr139Ile is a functional variant. Transcriptional coregulation within and across organs may be a general mechanism underlying the observed pleiotropy between urate and cardiometabolic traits.

Serum urate levels reflect a balance between uric acid production and its renal and intestinal excretion. Elevated serum urate levels define hyperuricemia, which is associated with metabolic, cardiovascular and kidney-related conditions. Hyperuricemia can cause kidney stones and gout, the most common inflammatory arthritis[1,2]. Gout attacks are a highly painful response to the deposition of urate crystals and are a significant cause of morbidity and related healthcare costs[3]. Although gout has become a major public health issue, it is undertreated because of low awareness, poor patient adherence[4] and inappropriate prescription practices of the most commonly used drug, allopurinol[5]. A better understanding of the mechanisms controlling serum urate may help to develop medications for gout treatment and prevention and provide insights into the regulatory mechanisms shared between urate and cardiometabolic traits.

Heritability of serum urate varies between 30 and 60% (refs. [6–11]). Candidate gene and genome-wide associations studies (GWAS) have identified three genes as major determinants of urate levels: *SLC2A9*, *ABCG2* and *SLC22A12* (refs. [7,12–18]). While *SLC2A9* and *ABCG2* harbor common variants of relatively large effect[15,19], *SLC22A12* contains many rare or low-frequency variants[18,20]. The largest GWAS meta-analyses performed to date have identified 28 loci among European ancestry[21] and 27 among Japanese individuals[22]. Many genes in the associated loci encode renal and intestinal urate transporters or their regulators, while others are relevant to glucose and lipid metabolism, functions of the liver, where uric acid is generated. With increased public availability of large annotation and gene expression datasets, fine-mapping of associated loci to prioritize target tissues, pathways and potentially causal genes and variants has become possible[23,24].

In this study, we perform a trans-ancestry meta-analysis of GWAS of serum urate in 457,690 individuals and identify 183 associated loci that improve gout risk prediction in an independent sample of 334,880 UK Biobank (UKBB) participants. We evaluate the genetic correlation of serum urate with hundreds of cardiometabolic traits and diseases and use a recently developed latent causal variable model to examine the contribution of causality versus pleiotropy. We prioritize target variants, genes, tissues and pathways that contribute to the complex regulation of urate levels through comprehensive data integration. Lastly, we conduct proof-of-principle experimental studies showing that HNF4A, a transcriptional master regulator in the liver and kidney proximal tubule, can regulate transcription of the gene encoding the major urate transporter ABCG2 in kidney cells and that the fine-mapped *HNF4A* variant p.Thr139Ile is functional. Transcriptional coregulation of processes linked to energy metabolism within and across organs may underlie the pleiotropy observed between urate levels and numerous cardiometabolic traits.

Results

Trans-ancestry meta-analysis identifies 183 urate-associated loci. Trans-ancestry meta-analyses were conducted to maximize the sample size for locus discovery and European ancestry-specific analyses were used where population-specific linkage disequilibrium (LD) was required to characterize loci (Supplementary Fig. 1). The primary trans-ancestry meta-analysis included 457,690 individuals (European ancestry, $n = 288{,}649$; East Asian ancestry, $n = 125{,}725$; African Americans, $n = 33{,}671$; South Asian ancestry, $n = 9{,}037$; and Hispanics, $n = 608$) from 74 studies. Mean urate levels ranged from 4.2 to 7.2 mg dl^{-1} (Supplementary Table 1). GWAS were performed based on genotypes imputed using the 1000 Genomes Project or Haplotype Reference Consortium reference panels (Methods and Supplementary Table 2). Results were combined through fixed-effect inverse-variance weighted meta-analysis after central study-specific quality control. There was no evidence of inflation due to unmodeled population structure (LD score regression intercept = 1.01; genomic control inflation factor $\lambda_{GC} = 1.04$). Post-meta-analysis variant filtering left 8,249,849 high-quality SNPs for downstream analyses (Methods).

We identified 183 loci that contained at least one genome-wide significant SNP ($P \leq 5 \times 10^{-8}$; Fig. 1 and Supplementary Table 3). Of these, 36 contained an index SNP reported in previous GWAS of serum urate[13,15,17,18,21,22,25,26]; 147 were previously unknown (Fig. 1). Allelic effects on serum urate ranged from 0.28 to 0.017 mg dl^{-1} (mean = 0.038 mg dl^{-1}, s.d. = 0.033). Regional association plots are shown in the Supplementary Dataset.

The index SNPs at all 183 loci explained 7.7% of the serum urate variance (Methods), compared to 5.3% explained by the index SNPs previously reported from GWAS in European ancestry populations[21]. In a large participating general population-based pedigree

A full list of authors and affiliations appears at the end of the paper.

study, the 183 index SNPs explained 17% of serum urate genetic heritability ($h^2 = 37.5\%$, 95% credible interval: 29.5, 45.3%), with 5% attributed to the index SNPs at *SLC2A9*, *ABCG2* and *SLC22A12* (Supplementary Fig. 2 and Methods).

Characterization of ancestry-related heterogeneity. For the 183 index SNPs, we observed no evidence of systematic between-study heterogeneity (median $I^2 = 2\%$, interquartile range 0–14%; Supplementary Table 3). Fourteen index SNPs showed significant evidence of ancestry-associated heterogeneity ($P_{\text{anc-het}} < 2.7 \times 10^{-4} = 0.05/183$) when tested using meta-regression (Methods), which is consistent with their higher measures of between-study heterogeneity ($I^2 > 25\%$; Fig. 1, Supplementary Table 3 and Supplementary Fig. 3). The most significant ancestry-associated heterogeneity was observed for rs3775947 at *SLC2A9* ($P_{\text{anc-het}} = 1.5 \times 10^{-127}$, allelic effect 0.34 (European ancestry), 0.26 (African Americans), 0.17 (East Asian ancestry), 0.41 (Hispanics) and 0.21 (South Asian ancestry) mg dl^{-1}), consistent with previous reports of population heterogeneity at this locus[27]. Nine genome-wide significant loci identified through meta-regression did not overlap with the 183 loci, including *SLC2A2* and *KCNQ1,* which were genome-wide significant in East Asian ancestry individuals (Supplementary Table 4). Ancestry-specific meta-analyses of European, African American, East Asians and South Asian populations are summarized in Supplementary Tables 5–8, respectively, and in the Supplementary Note.

Sex-stratified meta-analyses of serum urate GWAS. Mean serum urate levels and gout risk are higher in men than women[28]. Therefore, we tested whether the 183 urate-associated index SNPs showed sex-specific differences. Six SNPs showed significant effect differences ($P_{\text{diff}} < 2.7 \times 10^{-4} = 0.05/183$), at *SLC2A9*, *ABCG2*, *CAPN1*, *GCKR*, *IDH2* and *SLC22A12* (Supplementary Table 9). The genome-wide test for differences in genetic effects on urate levels between men and women identified only SNPs at *SLC2A9* and *ABCG2* ($P_{\text{diff}} < 5 \times 10^{-8}$; Methods and Supplementary Fig. 4), consistent with previous reports[7,14,15,21], and several suggestive loci ($P_{\text{diff}} < 1 \times 10^{-5}$; Supplementary Table 10).

Urate index SNPs are associated with gout. We next assessed the association of the 183 trans-ancestry urate index SNPs with gout in a trans-ancestry meta-analysis of 20 studies comprising 763,813 participants with 13,179 gout cases (Methods, Fig. 1 and Supplementary Table 1). Consistent with the causal role of hyperuricemia in gout, genetic effects were highly correlated (Spearman rank correlation coefficient = 0.87, Supplementary Fig. 5a; 0.82 for SNPs with urate association *P* values between 5×10^{-8} and 1×10^{-8}). Fifty-five SNPs were significantly associated with gout ($P < 2.7 \times 10^{-4} = 0.05/183$). In agreement with previous findings[21,29], the largest odds ratio (OR) for gout was observed at *ABCG2* (OR = 2.04, 95% confidence interval (CI) 1.96–2.12, $P = 7.7 \times 10^{-299}$). Genetic effects were generally larger in index SNPs with lower minor allele frequency, with the exception of a few common large-effect SNPs in the known major urate loci *SLC2A9*, *ABCG2* and *SLC22A12* (refs. [21,30]; Supplementary Fig. 5b).

A genetic risk score (GRS) for urate improves gout risk prediction. We evaluated whether a weighted urate GRS improved gout risk prediction when added to demographic information in a large, independent sample of 334,880 UKBB participants, including 4,908 gout cases (Methods). Across categories of the GRS, gout prevalence increased from 0.1 to 12.9% (Fig. 2a and Supplementary Table 11). Compared to the most common GRS category, the age- and sex-adjusted ORs of gout ranged from 0.09 (95% CI 0.02–0.37, $P = 7.8 \times 10^{-4}$) in the lowest to 13.6 (95% CI 7.2–25.8, $P = 1.4 \times 10^{-15}$) in the highest GRS category (Fig. 2b and Supplementary Table 11). The 3.5% of individuals in the three highest GRS categories had a greater than threefold increase in gout risk compared to individuals in the most common GRS category. This risk is comparable to a monogenic disease of modest effect size[31], but affects a higher proportion of the population.

We additionally constructed gout risk prediction models in the UKBB sample, which was not part of the discovery analysis of serum urate-associated variants. Gout status was regressed on the GRS alone (genetic model), on age and sex (demographic model) and on the GRS, age and sex (combined model) in a model development subset of 90% of the individuals to obtain precise estimates. These models were then used to predict gout status in the remaining 10%, the validation sample. The genetic model was a weaker predictor (area under the receiver operating characteristic curve (AUROC) = 0.67) than the demographic model (AUROC = 0.80). Addition of the GRS (combined model) significantly increased prediction accuracy (AUROC = 0.84, DeLong test $P < 2.2 \times 10^{-16}$; Fig. 2c) and achieved a sensitivity of 84% and specificity of 68%. Tenfold cross-validation of the regression models provided mean AUROCs of 0.67 (s.d. = 0.011), 0.78 (s.d. = 0.006) and 0.83 (s.d. = 0.008) for the genetic, demographic and combined models, respectively (Methods). The GRS represents a lifelong predisposition to higher urate levels and can be calculated at birth. Thus, the GRS may help to identify individuals with a high genetic predisposition for gout, allowing for compensatory lifestyle choices to reduce the risk of gout.

High genetic correlations of serum urate with cardiometabolic traits. Serum urate is positively correlated with many cardiometabolic risk factors and diseases[32]. We assessed genetic correlations between urate and 748 complex traits using cross-trait LD score regression (Methods). Serum urate levels were significantly ($P < 6.7 \times 10^{-5} = 0.05/748$) genetically correlated with 214 complex traits and diseases (Supplementary Table 12). The highest positive genetic correlation (r_g) was with gout ($r_g = 0.92$, $P = 3.3 \times 10^{-70}$), followed by traits representing components of the metabolic syndrome, such as the homeostatic model assessment for insulin resistance (HOMA-IR) ($r_g = 0.49$) and fasting insulin ($r_g = 0.45$; Fig. 3). The largest negative correlations were observed with high-density lipoprotein (HDL) cholesterol-related measurements (r_g up to −0.46) and with estimated glomerular filtration rate (eGFR) $r_g = -0.38$ and −0.26 for cystatin C and creatinine-based eGFR, respectively, consistent with the known role of the kidneys in urate excretion. Overall, the genetic correlations were consistent with observational associations from epidemiological studies[32].

To examine whether these genetic correlations reflect causal relationships or pleiotropy, we applied a recently developed latent causal variable (LCV) model to estimate the genetic causality proportion (GCP) for seven commonly studied cardiometabolic traits (Methods). As a positive control, we analyzed gout, confirming a genetically causal effect of urate on gout (GCP = 0.79; Supplementary Table 13), consistent with Mendelian randomization studies[33,34]. The seven cardiometabolic traits showed a GCP range consistent with mostly or partially genetically causal effects on serum urate. The largest GCP estimates were observed for adiposity-related traits (for example, GCP = −0.84 for waist circumference; Supplementary Table 13), where higher cell numbers should result in higher purine and consequently urate production. A bidirectional Mendelian randomization study reported a causal effect of adiposity on serum urate levels[35]. While the GCP and Mendelian randomization methods estimate different quantities to assess causality, the direction of effect can be compared and was consistent with a positive causal effect of obesity on serum urate. On the other hand, smaller GCP estimates for HDL cholesterol levels (GCP < 0.5; Supplementary Table 13)

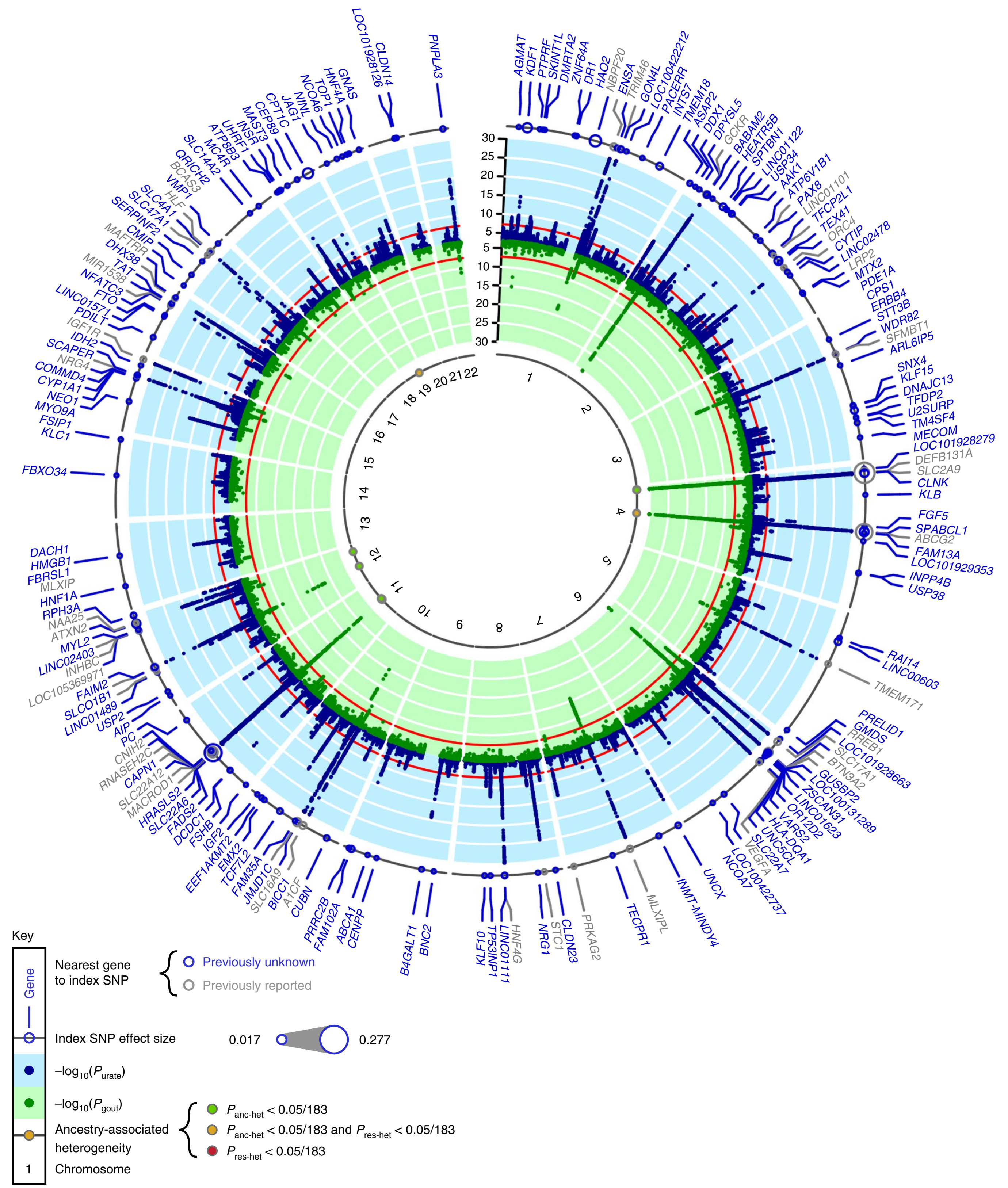

Fig. 1 | Trans-ancestry GWAS meta-analysis identifies 183 loci associated with serum urate. Outer ring: Dot size represents the genetic effect size of the index SNP at each labeled locus on serum urate. Blue band: $-\log_{10}$(two-sided meta-analysis P) for association with serum urate ($n = 457{,}690$) by chromosomal position (GRCh37 (hg19) reference build). The red line indicates genome-wide significance ($P = 5 \times 10^{-8}$). The blue gene labels indicate novel loci; the gray labels indicate loci reported in previous GWAS of serum urate. Green band: $-\log_{10}$(two-sided meta-analysis P) for association with gout ($n = 763{,}813$) by chromosomal position. The red line indicates genome-wide significance ($P = 5 \times 10^{-8}$). Association P values are truncated at 10^{-30}. Inner band: the dots represent the index SNPs with significant heterogeneity and are color-coded according to their source: green for ancestry-related heterogeneity ($P_{\text{anc-het}} < 2.7 \times 10^{-4}$ (0.05/183)); red for residual heterogeneity ($P_{\text{res-het}} < 2.7 \times 10^{-4}$); and yellow for both ($P_{\text{anc-het}}$ and $P_{\text{res-het}} < 2.7 \times 10^{-4}$). Loci are labeled with the gene closest to the index SNP. $P_{\text{anc-het}}$ and $P_{\text{res-het}}$ were generated using MR-MEGA (see Methods).

suggest the existence of a genetic process with a causal effect on both HDL cholesterol and serum urate, for example, coregulated metabolic processes in the liver. These processes may explain a large fraction of heritability for cholesterol levels and a modest fraction for urate, a type of asymmetry expected to produce a partially genetically causal relationship consistent with the one observed. Mendelian randomization studies did not support a causal relationship between cholesterol levels and serum urate[36].

Enriched tissues and pathways. To identify tissues and molecular mechanisms relevant for urate metabolism and handling, and to provide potential clues to the observed genetic correlations, we investigated which tissues, cell types and systems were significantly enriched for the expression of genes mapping to urate-associated loci (Methods). Based on all SNPs with $P<1\times10^{-5}$, we identified significant enrichment (false discovery rate (FDR) < 0.01) for 19 physiological systems, 3 tissues and 2 cell types (Supplementary Table 14). The strongest enrichment was observed for the kidney ($P=9.5\times10^{-9}$) and urinary tract ($P=9.9\times10^{-9}$), consistent with the kidney's prominent role in controlling urate levels. Additional significant enrichments were observed for endocrine and digestive systems, including the liver, the major site of urate production. Interestingly, a previously unknown significant enrichment was also observed in the musculoskeletal system, specifically for the synovial membrane, joint capsule and joints (Fig. 4a), the sites of gout attacks.

Next, we tested for cell type groups with evidence for enriched heritability based on cell type-specific functional genomic elements using stratified LD score regression (Methods). The strongest enrichment was observed for the kidney (11.5-fold), followed by the liver (5.39-fold; Supplementary Table 15).

Lastly, we tested whether any gene sets were enriched for variants associated with urate at $P<10^{-5}$ (Methods). Significant enrichment (FDR < 0.01) was observed for 383 reconstituted gene sets (Supplementary Table 16). Since many of these contained overlapping groups of genes, we used affinity propagation clustering to identify 57 metagene sets (Methods and Supplementary Table 17), including a prominent group of intercorrelated gene sets related to kidney and liver development, morphology and function (Fig. 4b). Together, these results underscore the prominent roles of the kidney and liver in regulating serum urate levels and implicate the kidney as a major target organ for lowering serum urate.

Prioritization via fine-mapping, functional annotation and gene expression. We established a workflow that combined fine-mapping of urate-associated loci with functional annotation and a systematic evaluation of tissue-specific differential gene expression to prioritize target SNPs and genes for translational research.

Statistical fine-mapping prioritizes candidate SNPs. Statistical fine-mapping was performed starting from the 123 genome-wide significant loci identified in the European ancestry-specific meta-analysis because the workflow included methods that used LD estimates from an ancestry-matched reference panel (Methods)[37]. After LD-based combination into 99 larger genomic regions, stepwise model selection in each region identified 114 independent SNPs ($r^2<0.01$; Methods). Overall, 87 regions contained 1 independent SNP, 10 contained 2 independent SNPs, the *ABCG2* locus contained 3 and the *SLC2A9* locus 4 independent SNPs (Supplementary Table 18). We computed 99% credible sets representing the smallest set of SNPs that collectively account for 99% posterior probability of containing the variant(s) driving the association signal (PPA)[38]. The 99% credible sets contained a median of 16 SNPs (Q1, Q3: 6, 57), and 6 of them only a single SNP, mapping in or near *INSR*, *RBM8A*, *MPPED2*, *HNF4A*, *CPT1C* and *SLC2A9* (Supplementary Table 18). Among 28 small credible sets (≤5 SNPs), several mapped in or near

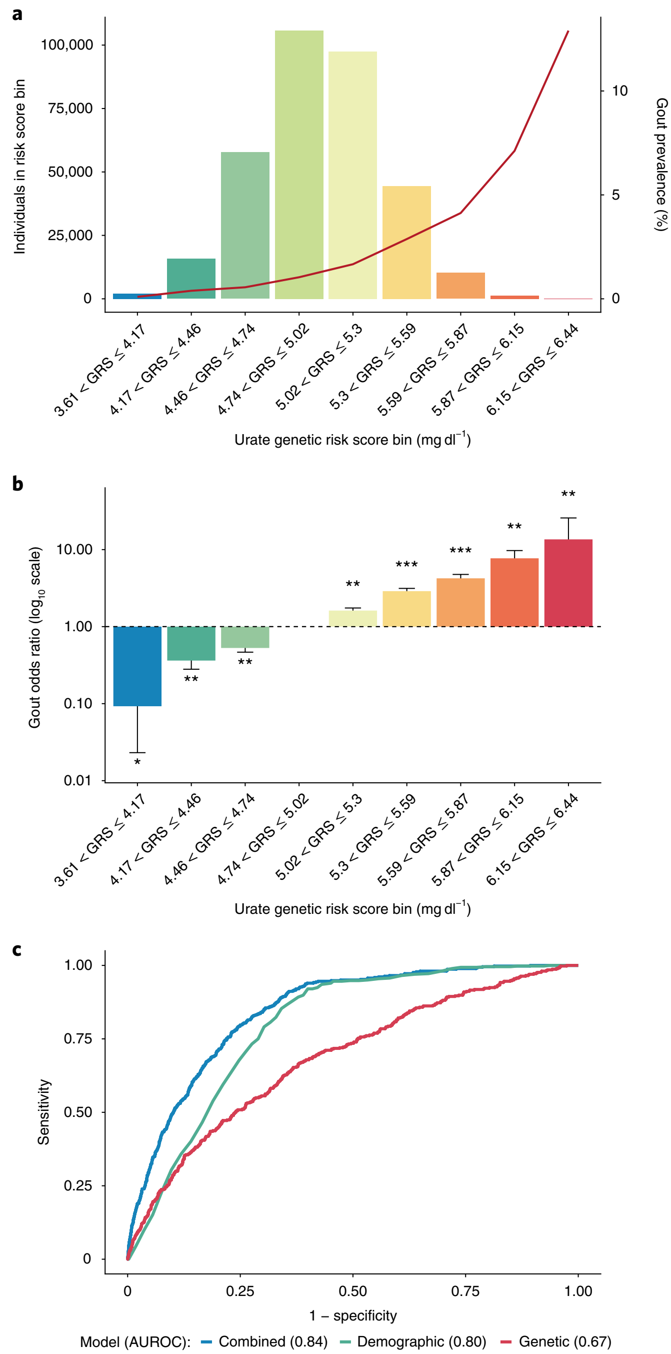

Fig. 2 | A GRS for serum urate improves gout risk prediction. **a**, Histogram of the urate GRS in 334,880 European ancestry participants of the UK Biobank. The *y* axes show the number of individuals (left) and the prevalence of gout (right); the *x* axis shows the categories of the urate GRS. The units on the *x* axis represent genetically predicted serum urate levels (mg dl^{-1}). **b**, Age- and sex-adjusted OR of gout (*y* axis) by GRS category (*x* axis) in 334,880 European ancestry participants of the UK Biobank, comparing each category to the most prevalent category ($4.74<GRS\leq5.02$) with the error bars representing the 95% CIs; *logistic regression two-sided $P<0.05$, **$P<5\times10^{-10}$ and ***$P<5\times10^{-100}$. **c**, Comparison of the ROC curves of different prediction models of gout: genetic (GRS only; red); demographic (age + sex; green), and combined (GRS + age + sex; blue). *y* axis: sensitivity; *x* axis: 1 – specificity. At the optimal cut points determined by the maximum of Youden's index, the sensitivity of the combined model was 84% and specificity was 68%.

Fig. 3 | Serum urate shows widespread genetic correlations with cardiometabolic risk factors and diseases. The Circos plot shows significant genome-wide genetic correlations between serum urate and 214 complex traits or diseases (genetic correlation $P < 6.7 \times 10^{-5} = 0.05/748$ traits tested), with the bar height proportional to the genetic correlation coefficient (r_g) estimate for each trait and colored according to its direction (dark blue, $r_g > 0$; light blue, $r_g < 0$). Traits and diseases are labeled on the outside of the plot and grouped into nine different categories. Each category is color-coded (inner ring, inset). The greatest genetic correlation was observed with gout ($r_g = 0.92$, $P = 3.3 \times 10^{-70}$). Genetic correlations with multiple cardiometabolic risk factors and diseases reflect their known directions from observational studies. The serum urate association statistics for estimating genetic correlations were from the European ancestry meta-analysis ($n = 288,649$). BMI, body mass index; CSE, certificate of secondary education; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; HOMA-B, homeostatic model assessment for β-cell function; MI, myocardial infarction; VLDL, very-low-density lipoprotein.

genes with an established role in urate handling such as *SLC2A9*, *PDZK1*, *ABCG2*, *SLC22A11* and *SLC16A9* (ref. [20]). These credible sets contain the most supported SNPs and greatly reduce the number of candidate variants for experimental follow-up.

Credible set SNPs were annotated for their functional consequence and regulatory potential (Methods). Missense SNPs with PPA > 50% or belonging to small credible sets were identified in *ABCG2*, *UNC5CL*, *HNF1A*, *HNF4A*, *CPS1* and *GCKR* (Fig. 5a

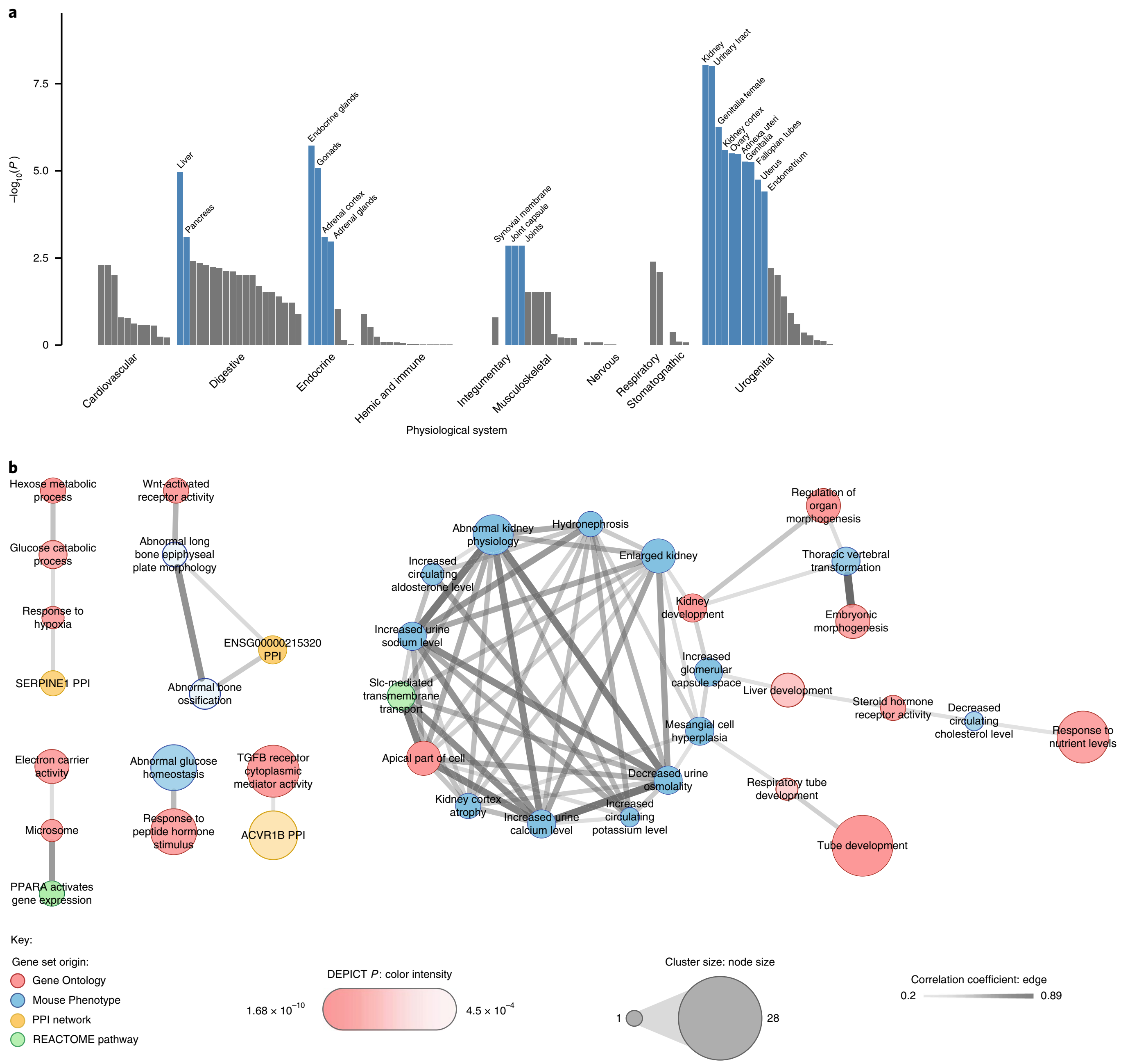

Fig. 4 | Genes expressed in urate-associated loci are enriched in kidney tissue and pathways. a, Grouped physiological systems (*x* axis) that were tested individually for enrichment of expression of genes in urate-associated loci among European ancestry individuals ($n = 288,649$) using DEPICT are shown as a barplot, with the $-\log_{10}$(enrichment P) on the y axis. Significantly enriched systems are labeled and highlighted in blue (enrichment FDR < 0.01). **b**, Correlated ($r > 0.2$) metagene sets that were strongly enriched (enrichment FDR < 0.01) for genes mapping to urate-associated loci among European ancestry individuals ($n = 288,649$). The thickness of the edges represents the magnitude of the correlation coefficient; node size, color and intensity represent the number of clustered gene sets, gene set origin and enrichment P, respectively. PPI, protein–protein interaction.

and Supplementary Table 19). All missense SNPs except the one in *GCKR* had a combined annotation-dependent depletion (CADD) score > 15, supporting them as potentially deleterious. Indeed, functional effects have already been demonstrated experimentally for rs2231142 (p.Gln141Lys, $r^2 = 1$ to the index SNP rs74904971) in *ABCG2*, rs742493 (p.Arg432Gly) in *UNC5CL* and rs1260326 (p.Leu446Pro) in *GCKR* (Table 1). Nonexonic variants with a PAA > 90% and mapping into open chromatin in enriched tissues were identified in *RBM8A*, *SLC2A9*, *INSR*, *HNF4A*, *PDZK1*, *NRG4*, *UNC5CL* and *AAK1* (Methods, Supplementary Fig. 6 and Supplementary Table 19). When complemented by evidence of gene expression colocalization, these SNPs may represent causal regulatory variants and highlight their potential effector genes.

We compared our fine-mapping workflow ('Wakefield'), established in previous studies[39,40], to an alternative approach implemented in FINEMAP v.1.3 (Methods)[41]. FINEMAP identified 152 credible sets (median of 7 SNPs). With regard to known causal variants in *ABCG2* (rs2231142), *GCKR* (rs1260326), *HNF4A* (rs1800961) and *PDZK1* (rs1967017), the Wakefield approach identified the causal variants in *ABCG2*, *GCKR* and *HNF4A* as credible set members, whereas FINEMAP found those in *ABCG2* and *HNF4A*. A comparison of all SNPs mapping into small credible sets

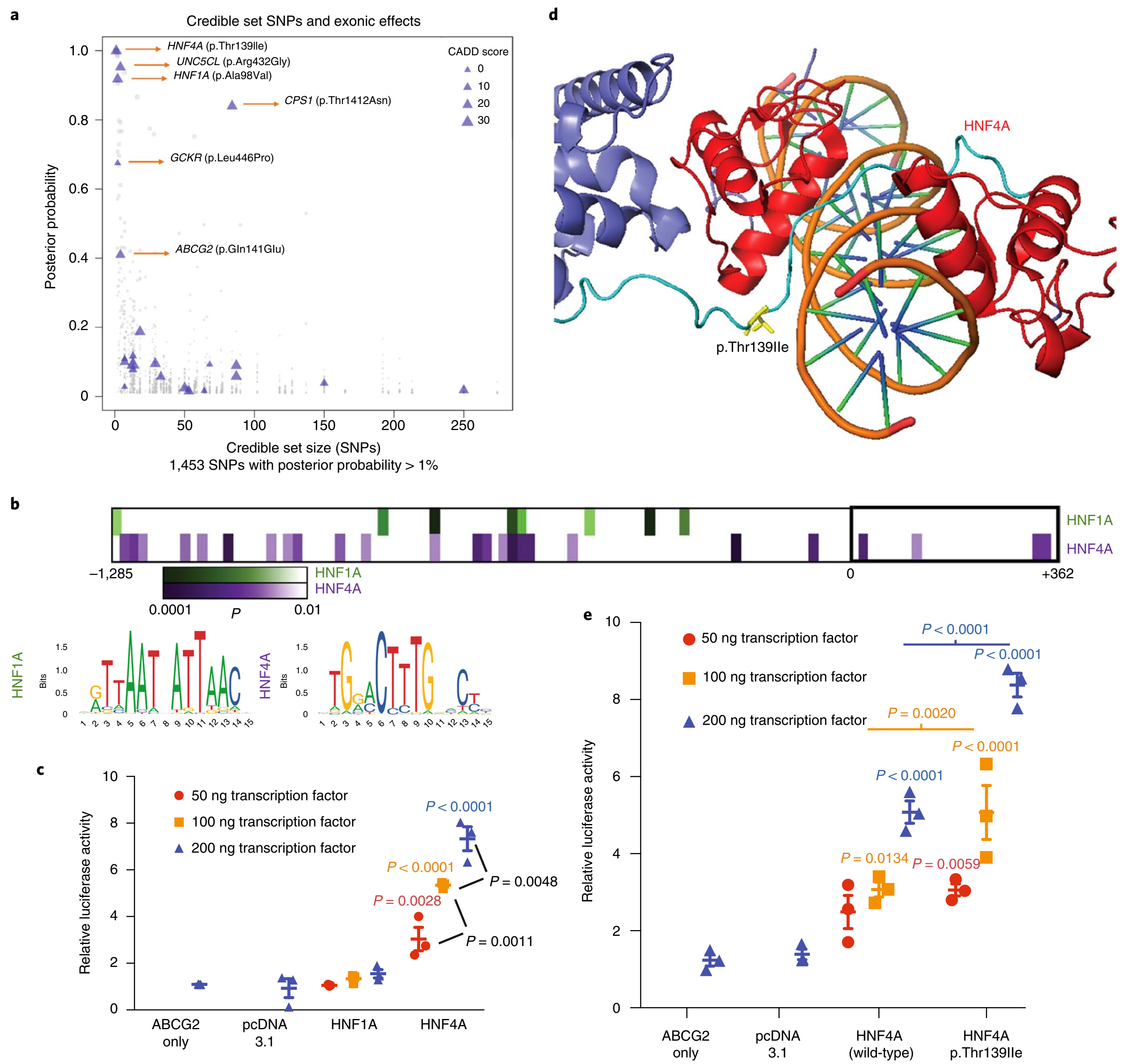

Fig. 5 | Prioritization of p.Thr139Ile at *HNF4A* and functional study of HNF4A regulation of *ABCG2* transcription. a, Graph shows credible set size (*x* axis) against the posterior probability of association (PPA, *y* axis) for each of 1,453 SNPs with a PPA > 1% in 114 99% credible sets. The triangles mark missense SNPs, with size proportional to their CADD score. The blue triangles indicate missense variants mapping into small (≤5 SNPs) credible sets or with high PPA (≥50%). **b**, Predicted HNF1A or HNF4A binding sites in the promoter region of *ABCG2* using LASAGNA 2.0, consensus affinity sequence and *P* of probable matches based on nucleotide position within a consensus transcription factor binding site (see Methods). **c**, Relative luciferase activity and transactivation of the *ABCG2* promoter in cells transfected with a variable amount of HNF1A or HNF4A constructs (mean (line) ± s.e.m. (whiskers), $n = 3$ independent experiments, *P* values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test). **d**, Position of p.Thr139Ile (T139I) in DNA binding domain/hinge region within HNF4A homodimer structure (PDB 4IQR). **e**, Relative luciferase activity and transactivation of the *ABCG2* promoter in cells transfected with variable amounts of constructs (nanograms of transfected DNA) of wild-type HNF4A (threonine) or isoleucine at position 139 (±s.e.m., $n = 3$ independent experiments). *P* values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test.

(≤5 SNPs) identified through both approaches found highly correlated posterior probabilities (Pearson correlation coefficient = 0.86).

Gene prioritization via gene expression colocalization analyses. The urate association signals were next tested for colocalization with expression quantitative trait loci (eQTLs) in *cis* across three kidney tissue resources and 44 Genotype-Tissue Expression (GTEx) tissues (Methods). High posterior probability of colocalization (H4 ≥ 0.8; Methods) supports a trait-associated variant acting through gene expression in the tissue where colocalization is identified. We identified colocalization with the expression of 13 genes in the kidney (Fig. 6), the organ with the strongest enrichment for urate-associated variants. Whereas colocalization of some genes was only observed in the kidney (*SLC17A4*, *BICC1*, *UMOD*, *GALNTL5*, *NCOA7*),

Table 1 | Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal

Gene	SNP	No. of SNPs in set	SNP posterior probability	Consequence	CADD[a]	DHS	Gout meta-analysis *P* (European ancestry)	Brief summary of literature and gene function
ABCG2	rs2231142	4	0.41	p.Gln141Lys (NP_004818.2)	18.2	ENCODE epithelial	1.21×10^{-290}	Encodes a xenobiotic and high-capacity urate membrane transporter expressed in the kidney, liver and gut. Causal variants have been reported for gout susceptibility (MIM 138900) and the Junior Jr(a−) blood group phenotype (MIM 614490). The locus was first identified in association with serum urate through GWAS[15] and confirmed in many studies since. The common causal variant encoding p.Gln141Lys has been experimentally confirmed[19] as a partial loss of function.
UNC5CL	rs742493	4	0.95	p.Arg432Gly (NP_775832.2) (within death domain)	21.0	ENCODE epithelial	2.73×10^{-1}	Encodes for the death domain-containing UNC-5 family C terminal-like membrane-bound protein. Suggested as a candidate gene for mucosal diseases, with a role in epithelial inflammation and immunity[61]. In human HEK 293 cells, UNC5CL can transduce proinflammatory programs via activation of NF-κB, with the p.Gly432 variant less potent than the p.Arg432 one[61].
HNF1A	rs1800574	2	0.92	p.Ala98Val (NP_000536.5)	23.4		1.83×10^{-2}	Encodes a transcription factor with strong expression in the liver, gut and kidney. Rare mutations cause autosomal-dominant MODY3 (MIM 600496). Locus found in GWAS of type 2 diabetes[62] and blood urea nitrogen[22]. Together with *HNF4A*, it was first recognized as the master regulator of hepatocyte and islet transcription. Knockout mice show proximal tubular dysfunction (Fanconi syndrome). HNF1A enhanced the promoter activity of PDZK1, URAT1, NPT4 and OAT4 in human renal proximal tubule cell-based assays[45], supporting a role in the coordinated expression of components of the urate 'transportosome'.
HNF4A	rs1800961	1	1.00	p.Thr139Ile (NP_000448.3)	24.7	ENCODE pancreas	7.43×10^{-3}	Encodes another nuclear receptor and transcription factor that controls expression of many genes, including *HNF1A* and other overlapping target genes. Rare mutations cause autosomal-dominant MODY1 (MIM 125850) and autosomal-dominant renal Fanconi syndrome 4 (MIM 616026). Shown to regulate the expression of SLC2A9 and other members of the ukrate 'transportosome' in cell-based assays[54,55]. The GWAS locus has been reported for multiple cardiometabolic traits and type 2 diabetes[63].
CPS1	rs1047891	84	0.84	p.Thr1412Asn (NP_001116105.1)	22.1		5.66×10^{-2}	Encodes mitochondrial carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle. Rare mutations cause autosomal-recessive carbamoylphosphate synthetase I deficiency (MIM 237300). In addition to hyperammonemia, this disease features increased synthesis of glutamine, a precursor of purines. Elevated uric acid excretion has been reported in patients with hyperammonemia[64]. GWAS locus for eGFR[65], homocysteine[66] and urinary glycine concentrations[67].

Continued

Table 1 | Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal (continued)

Gene	SNP	No. of SNPs in set	SNP posterior probability	Consequence	CADD[a]	DHS	Gout meta-analysis *P* (European ancestry)	Brief summary of literature and gene function
GCKR	rs1260326	2	0.67	p.Leu446Pro (NP_001477.2)	0.1	ENCODE kidney	4.09×10^{-41}	Encodes a regulatory protein prominently expressed in the liver that inhibits glucokinase. Identified in a previous GWAS of urate[21] and multiple other cardiometabolic traits. The p.Leu446 protein was shown to be less activated than p.Pro446 by physiological concentrations of fructose 6-phosphate, leading to reduced glucokinase inhibitory ability[68].

[a]CADD, combined annotation dependent depletion phred score. Gout meta-analysis *P* values were two-sided ($n = 763{,}813$). Posterior probabilities were estimated from statistical fine-mapping using the Wakefield approach (Methods). The DHS column contains an entry only when the SNP mapped into a DHS in a target tissue. Genes were included if they contained a missense variant with a posterior PPA > 50% or mapped into a small credible set (≤5 SNPs).

others showed colocalization in several tissues (for example, *ARL6IP5*). The direction of change in gene expression with higher urate levels could vary for the same gene across tissues. For instance, the allele associated with higher serum urate at *SLC16A9* was associated with higher gene expression in the kidney, which is consistent with a regulatory variant in a transporter mediating the reabsorption of urate. This same allele was associated with lower gene expression in other tissues, such as the aorta, pointing toward tissue-specific regulatory mechanisms[42]. Details of the 13 genes with evidence for colocalization with gene expression in the kidney are summarized in Supplementary Table 20. Significant colocalizations across all 47 tissues (Supplementary Fig. 7) revealed additional insights such as colocalization of the urate association signal with *NFAT5* expression in subcutaneous adipose tissue, emphasizing its role in adipogenesis[43], or *PDZK1* expression in the colon and ileum, important sites of urate excretion.

Lastly, we investigated whether any trans-ancestry index SNPs or their proxies ($r^2 > 0.8$) were reproducibly associated with gene expression in *trans* in several large eQTL studies (Supplementary Table 21 and Supplementary Note). We identified interchromosomal associations between 5 index SNPs and 16 transcripts that were enriched in the cardiovascular-disease-related term 'heart disease' based on the Human Disease Ontology database (release datestamp 20150323, sub-version 2806, Supplementary Note and Supplementary Table 22).

HNF4A activates *ABCG2* transcription, and *HNF4A* p.Thr139Ile is a functional variant. The gene and variant prioritization workflow was validated using the identified candidates *HNF1A* and *HNF4A*. Coregulation of target genes by these transcriptional master regulators in the kidney proximal tubule and liver could potentially explain the observed genetic correlations[44].

We first tested whether HNF1A and HNF4A affect transcription of *ABCG2*, which encodes for a major human urate transporter and represents the locus with the highest gout risk in our screen. The *ABCG2* promoter region contains several predicted HNF1A and HNF4A binding sites (Fig. 5b). A luciferase reporter assay in the HEK 293 cell line was used to assess transactivation of the human *ABCG2* promoter by the HNF4A and HNF1A proteins (Methods and Supplementary Fig. 8a). Coexpression of HNF4A significantly increased the *ABCG2* promoter-driven luciferase activity in a transfection dose-dependent and HNF4A protein abundance-dependent manner (Fig. 5c and Supplementary Fig. 8b). No increase in luciferase activity occurred with the negative control vector devoid of the *ABCG2* promoter (Supplementary Fig. 8d,e). Results for HNF1A indicated that the observed association with serum urate probably does not occur via activation of *ABCG2* in kidney cells (Fig. 5c); however, *HNF1A* has been reported to activate the transcription of *PDZK1*, which encodes a regulatory protein for several other renal urate transporters[45,46] also identified in this study.

Next, we tested the functional relevance of the prioritized p.Thr139Ile allele in *HNF4A* (NM_178849.2, isoform 1; Methods). Its location within the hinge/DNA binding domain (Fig. 5d and Supplementary Fig. 8f) supports potentially altered interactions with targeted promoter regions. The isoleucine substitution at position 139 significantly increased the transactivation of the *ABCG2* promoter compared to the wild-type threonine (Fig. 5e), without altering HNF4A protein abundance (Supplementary Fig. 8c). Thus, HNF4A can activate *ABCG2* transcription in a kidney cell line and *HNFA4* p.Thr139Ile is a functional variant. Increased activation of the urate excretory protein ABCG2 by the allele encoding the isoleucine residue should result in lower serum urate levels, consistent with the observed negative association in our GWAS.

Discussion

This trans-ancestry GWAS meta-analysis of serum urate based on 457,690 individuals represents a fourfold increase in sample size over previous studies[21,22,47] and identified 183 urate-associated loci, 147 of which were previously unknown. A genetic urate risk score led to significant improvements in gout risk prediction in 334,880 UKBB participants: 3.5% had a risk of gout comparable to a Mendelian disease effect size. Genetic correlation and causality analyses confirmed the causal effect of urate on gout and were consistent with transcriptional coregulation as a source of pleiotropy in the widespread genetic correlations between serum urate and cardiometabolic traits. Tissue and cell type-specific enrichment analyses supported the notion that the kidney and liver, the sites of urate excretion and generation, are key target tissues. Comprehensive fine-mapping and colocalization analyses with gene expression across 47 tissues delivered an extensive list of target genes and SNPs for follow-up studies, of which we experimentally confirmed *HNF4A* p.Thr139Ile as a functional allele involved in transcriptional regulation of urate homeostasis.

Major challenges of GWAS are to pinpoint causal genes and variants and provide actionable insights into disease-relevant mechanisms. This study developed a comprehensive resource of urate-related candidate SNPs, genes, tissues and pathways that will enable a wide range of follow-up studies. Out of the many and biologically plausible findings, we highlight two instances where colocalization analyses provided new insights. First, colocalization helped to prioritize genes in association peaks that previous

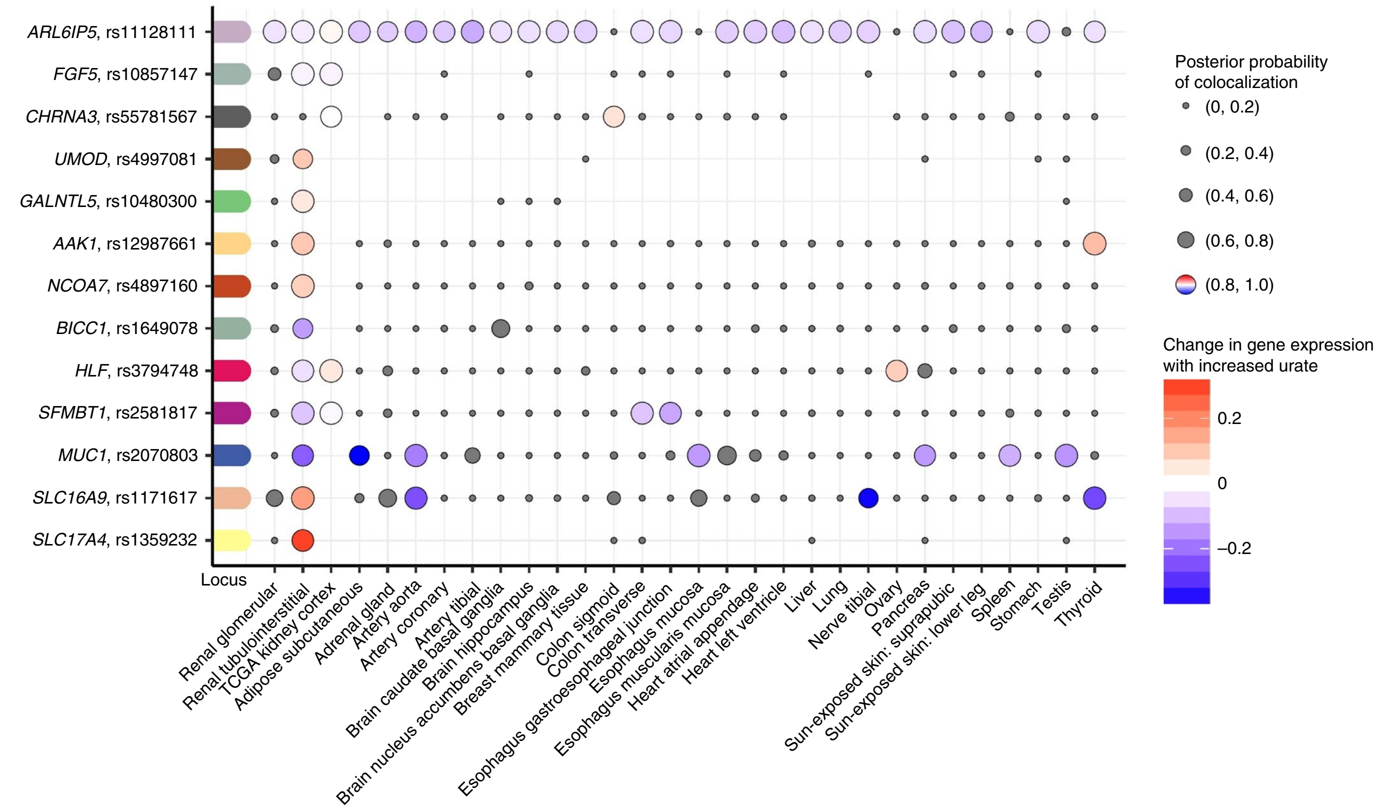

Fig. 6 | Colocalization of urate association signals with gene expression in *cis* in kidney tissues. The serum urate association signals identified in European ancestry individuals ($n = 288{,}649$) were tested for colocalization with all eQTLs where the eQTL *cis*-window overlapped (±100 kb) the index SNP. Genes with ≥1 positive colocalization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective index SNP and gene (*y* axis). Colocalizations across all tissues (*x* axis) are illustrated as dots, where the size of the dot indicates the posterior probability of the colocalization. Negative colocalizations (posterior probability of H4 <0.80) are marked in gray, while positive colocalizations are color-coded relative to the change in expression with a color gradient as indicated in the key to the figure.

GWAS could not resolve. For example, the locus at chromosome 6p22.2 contains genes encoding for four members of the SLC17 transporter family (*SLC17A1*–*SLC17A4*). Systematic testing of colocalization across genes and tissues identified evidence only for *SLC17A4* in the kidney, with higher expression associated with higher serum urate. Previous experimental studies have implicated SLC17A4 as a urate exporter in the intestine[48] and our data support its yet unappreciated role in renal urate transport. Second, colocalization with *MUC1*, *BICC1* and *UMOD* expression in the kidney suggests a shared biological mechanism. Rare mutations in all three genes underlie monogenic cystic kidney diseases[49–51].

Another noteworthy finding is the significant genetic correlations with many cardiometabolic traits, consistent with observational associations[52]. Many of these traits are influenced by liver metabolism. The estimated GCPs supported their genetic correlations to be partly driven by overlapping or coregulated metabolic pathways and not only by a fully causal effect of, for example, cholesterol or insulin levels on urate. Likewise, significant genetic correlations with kidney-related traits such as eGFR may reflect shared regulatory processes in the kidney. The observed pleiotropic effects of many urate-associated variants could thus be the potential manifestation of coregulation of processes that occur within and across tissues relevant to the implicated traits, a mechanism probably prevailing in metabolic but also other traits.

In the kidney, nuclear HNF4A is exclusively detected in the proximal tubule[53], where it has been reported to regulate the expression of SLC2A9 isoform 1 (ref. [54]) and PDZK1 (ref. [55]). Kidney-specific deletion of *Hnf4a* in mice phenocopies Fanconi renotubular syndrome[56]. Transcriptomic analyses support HNF4A to drive a proximal tubule signature cluster of 221 coexpressed genes, including many candidate genes for urate metabolism and transport[53]. In addition to *HNF4A*, *HNF4G* and *HNF1A*, ten genes in this cluster also map to the urate-associated loci we identified (*A1CF*, *CUBN*, *LRP2*, *PDZK1*, *SERPINF2*, *SLC2A9*, *SLC16A9*, *SLC17A1*, *SLC22A12* and *SLC47A1*). In addition, our study establishes that HNF4A can transactivate transcription of *ABCG2* in a kidney cell line, the key urate secretory transporter in gut and kidney epithelium[57]. The genetic variant encoding the p.Thr139Ile substitution is located in a region of the HNF4A protein harboring many causative mutations for monogenic maturity-onset diabetes of the young type 1 (MODY1)[58]. Yet, unlike the severe MODY1 missense mutations p.Arg127Trp, p.Asp126Tyr and p.Arg125Trp,[59] p.Thr139Ile has not been reported to cause MODY1. Instead, it has been reported to increase the risk of type 2 diabetes, possibly through a liver-specific loss of HNF4A phosphorylation at p.Thr139, and to associate with HDL cholesterol levels[58,60]. These data point to additional complexities when interpreting pleiotropic effects because there may be several tissue-specific mechanisms by which genetic variants in transcriptional regulators influence metabolic pathways and urate homeostasis.

Some limitations warrant mention. The numbers of individuals of ancestries other than European or East Asian were small and the generalizability of the gout prediction models should be assessed in future independent studies of non-European ancestry. Focusing on SNPs present in the majority of studies emphasizes those that may be of greatest importance globally over population-specific variants.

General limitations of the field include that statistical fine-mapping approaches based on meta-analysis summary statistics cannot clearly prioritize functional variants in regions of tight LD, and that they are influenced by the availability and imputation quality of SNPs in the contributing studies. Only few regulatory maps from important target tissues such as the synovial membrane and kidney are available, but we evaluated differential gene expression in three kidney datasets. Generating additional regulatory and expression datasets across disease states, developmental stages and additional cell types in the kidney and other metabolically active organs constitutes an important future research avenue. Lastly, a large independent sample for adequately powered replication testing was unavailable and represents a future endeavor. However, high correlations between genetic effects on serum urate and gout even for SNPs with the weakest significant urate associations, as well as no indication of significant heterogeneity, reduce concerns about false positives.

In summary, this large-scale study generated an atlas of candidate SNPs, genes, tissues and pathways involved in urate metabolism and its shared regulation with multiple cardiometabolic traits that will enable a wide range of follow-up studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0504-x.

Received: 17 January 2019; Accepted: 27 August 2019;
Published online: 2 October 2019

References

1. Kuo, C. F., Grainge, M. J., Zhang, W. & Doherty, M. Global epidemiology of gout: prevalence, incidence and risk factors. *Nat. Rev. Rheumatol.* **11**, 649–662 (2015).
2. Li, X. et al. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* **357**, j2376 (2017).
3. Jinno, S., Hasegawa, K., Neogi, T., Goto, T. & Dubreuil, M. Trends in emergency department visits and charges for gout in the United States between 2006 and 2012. *J. Rheumatol.* **43**, 1589–1592 (2016).
4. Kuo, C. F., Grainge, M. J., Mallen, C., Zhang, W. & Doherty, M. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann. Rheum. Dis.* **74**, 661–667 (2015).
5. Mikuls, T. R., Farrar, J. T., Bilker, W. B., Fernandes, S. & Saag, K. G. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology (Oxford)* **44**, 1038–1042 (2005).
6. Yang, Q. et al. Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism* **54**, 1435–1441 (2005).
7. Vitart, V. et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat. Genet.* **40**, 437–442 (2008).
8. Pilia, G. et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2**, e132 (2006).
9. Wang, W. et al. Heritability and genome-wide association analyses of serum uric acid in middle and old-aged Chinese twins. *Front. Endocrinol. (Lausanne)* **9**, 75 (2018).
10. MacCluer, J. W. et al. Heritability of measures of kidney disease among Zuni Indians: the Zuni Kidney Project. *Am. J. Kidney Dis.* **56**, 289–302 (2010).
11. Rule, A. D. et al. Genome-wide linkage analysis for uric acid in families enriched for hypertension. *Nephrol. Dial. Transplant.* **24**, 2414–2420 (2009).
12. Enomoto, A. et al. Molecular identification of a renal urate anion exchanger that regulates blood urate levels. *Nature* **417**, 447–452 (2002).
13. Li, S. et al. The *GLUT9* gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet* **3**, e194 (2007).
14. Döring, A. et al. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.* **40**, 430–436 (2008).
15. Dehghan, A. et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* **372**, 1953–1961 (2008).
16. Kolz, M. et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* **5**, e1000504 (2009).
17. Yang, Q. et al. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ. Cardiovasc. Genet.* **3**, 523–530 (2010).
18. Tin, A. et al. Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum. Mol. Genet.* **20**, 4056–4068 (2011).
19. Woodward, O. M. et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl Acad. Sci. USA* **106**, 10338–10342 (2009).
20. Major, T. J., Dalbeth, N., Stahl, E. A. & Merriman, T. R. An update on the genetics of hyperuricaemia and gout. *Nat. Rev. Rheumatol.* **14**, 341–353 (2018).
21. Köttgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
22. Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
23. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
24. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
25. Kamatani, Y. et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).
26. Okada, Y. et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.* **44**, 904–909 (2012).
27. Merriman, T. R. Population heterogeneity in the genetic control of serum urate. *Semin. Nephrol.* **31**, 420–425 (2011).
28. Roddy, E. & Choi, H. K. Epidemiology of gout. *Rheum. Dis. Clin. North Am.* **40**, 155–175 (2014).
29. Phipps-Green, A.J. et al. Twenty-eight loci that influence serum urate levels: analysis of association with gout. *Ann. Rheum. Dis.* **75**, 124–130 (2016).
30. George, R.L. & Keenan, R.T. Genetics of hyperuricemia and gout: implications for the present and future. *Curr. Rheumatol. Rep.* **15**, 309 (2013).
31. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
32. Feig, D. I., Kang, D. H. & Johnson, R. J. Uric acid and cardiovascular risk. *N. Engl. J. Med.* **359**, 1811–1821 (2008).
33. Keenan, T. et al. Causal assessment of serum urate levels in cardiometabolic diseases through a Mendelian randomization study. *J. Am. Coll. Cardiol.* **67**, 407–416 (2016).
34. Jordan, D. M. et al. No causal effects of serum urate levels on the risk of chronic kidney disease: a Mendelian randomization study. *PLoS Med.* **16**, e1002725 (2019).
35. Lyngdoh, T. et al. Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. *PLoS ONE* **7**, e39321 (2012).
36. White, J. et al. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* **4**, 327–336 (2016).
37. Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
38. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
39. Gaulton, K. J. et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
40. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
41. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
42. Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
43. Asano, T. et al. The role of *N*-glycosylation of GLUT1 for glucose transport activity. *J. Biol. Chem.* **266**, 24632–24636 (1991).
44. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
45. Prestin, K. et al. Regulation of PDZ domain-containing 1 (PDZK1) expression by hepatocyte nuclear factor-1α (HNF1α) in human kidney. *Am. J. Physiol. Renal Physiol.* **313**, F973–F983 (2017).

46. Maher, J. M. et al. Alterations in transporter expression in liver, kidney, and duodenum after targeted disruption of the transcription factor HNF1α. *Biochem. Pharmacol.* **72**, 512–522 (2006).
47. Sulem, P. et al. Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat. Genet.* **43**, 1127–1130 (2011).
48. Togawa, N., Miyaji, T., Izawa, S., Omote, H. & Moriyama, Y. A Na^+-phosphate cotransporter homologue (SLC17A4 protein) is an intestinal organic anion exporter. *Am. J. Physiol. Cell Physiol.* **302**, C1652–C1660 (2012).
49. Kirby, A. et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in *MUC1* missed by massively parallel sequencing. *Nat. Genet.* **45**, 299–303 (2013).
50. Kraus, M. R. et al. Two mutations in human *BICC1* resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. *Hum. Mutat.* **33**, 86–90 (2012).
51. Hart, T. C. et al. Mutations of the *UMOD* gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J. Med. Genet.* **39**, 882–892 (2002).
52. Sodini, S. M., Kemper, K. E., Wray, N. R. & Trzaskowski, M. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics* **209**, 941–948 (2018).
53. Lindgren, D. et al. Cell-type-specific gene programs of the normal human nephron define kidney cancer subtypes. *Cell Rep.* **20**, 1476–1489 (2017).
54. Prestin, K. et al. Transcriptional regulation of urate transportosome member SLC2A9 by nuclear receptor HNF4α. *Am. J. Physiol. Renal Physiol.* **307**, F1041–F1051 (2014).
55. Ketharnathan, S. et al. A non-coding genetic variant maximally associated with serum urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Hum. Mol. Genet.* **27**, 3964–3973 (2018).
56. Marable, S. S., Chung, E., Adam, M., Potter, S. S. & Park, J. S. *Hnf4a* deletion in the mouse kidney phenocopies Fanconi renotubular syndrome. *JCI Insight* **3**, 97497 (2018).
57. Matsuo, H. et al. ABCG2 dysfunction causes hyperuricemia due to both renal urate underexcretion and renal urate overload. *Sci. Rep.* **4**, 3755 (2014).
58. Daigo, K. et al. Proteomic analysis of native hepatocyte nuclear factor-4α (HNF4α) isoforms, phosphorylation status, and interactive cofactors. *J. Biol. Chem.* **286**, 674–686 (2011).
59. Chandra, V. et al. Multidomain integration in the structure of the HNF-4α nuclear receptor complex. *Nature* **495**, 394–398 (2013).
60. Zhu, Q. et al. T130I mutation in *HNF-4α* gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**, 567–573 (2003).
61. Heinz, L. X. et al. The death domain-containing protein Unc5CL is a novel MyD88-independent activator of the pro-inflammatory IRAK signaling cascade. *Cell Death Differ.* **19**, 722–731 (2012).
62. Saxena, R. et al. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *Am. J. Hum. Genet.* **90**, 410–425 (2012).
63. Kooner, J. S. et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* **43**, 984–989 (2011).
64. Van Gennip, A. H., Van Bree-Blom, E. J., Grift, J., DeBree, P. K. & Wadman, S. K. Urinary purines and pyrimidines in patients with hyperammonemia of various origins. *Clin. Chim. Acta* **104**, 227–239 (1980).
65. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
66. van Meurs, J. B. et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *Am. J. Clin. Nutr.* **98**, 668–676 (2013).
67. Raffler, J. et al. Genome-wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality. *PLoS Genet.* **11**, e1005487 (2015).
68. Beer, N. L. et al. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.* **18**, 4081–4088 (2009).

Acknowledgements

We thank D. Di Domizio (Eurac Research) and J. Knaus (University of Freiburg) for IT assistance, and T. Johnson (GlaxoSmithKline) for sharing his code and for the discussion about fine-mapping a credible set and the colocalization analysis. This research was conducted using the UKBB Resource (application no. 20272). Study-specific acknowledgements and funding sources are listed in the Supplementary Note. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health (NIH) or the US Department of Health and Human Services.

Author contributions

A.Tin, J.M., V.L.H.K., Y.L., M.Wuttke, H.K., K.B.S., C.Q., M.Gorski, M.Scholz, A.M.H., A.Teumer, C.P., O.M.W., V.V. and A.Köttgen wrote the manuscript. A.Tin, J.M., M.Wuttke, M.Gorski, C.F., A.Teumer, C.P., O.M.W., V.V. and A.Köttgen designed the study. A.S.B., A.M.H., A.Teumer, A.P., A.P.J.D.V., A.B.Z., A.D.G., A.Metspalu, A.A.H., A.Tönjes, A.Köttgen, A.P., A.Krajčoviechová, A.R., A.C., B.Ponte, B.K.K., B.J., B.W.J.H.P., B.M.P., C.H., C.A.B., C.N.S., C.G., C.J.O.D., C.M.V.D., C.P., D.T., D.C., D.M., E.B., E.I., F.K., G.S., G.E., G.W., G.B., G.N.N., G.W.M., H.S., H.Schmidt, I.R., J.M.G., J.F.W., J.G.W., J.S.K., J.O.C., J.Thiery, J.Tremblay, J.B.W., J.C.C., J.C., K.-U.E., K.L.M., K.Stefansson, K.Ho, K.M., K.Strauch, M.A.I., M.E.K., M.C., M.P., M.L., M.Scholz, M.H.D.B., M.Waldenberger, M.Stumvoll, M.K.E., M.K., M. Kähönen, M.B., M.R., N.G.M., O.D., O.T.R., O.P., P.G., P.P.P., P.V., P.V.D.H., Q.Y., R.C., R.Rettig, R.S., R.D.M., R.J.C., R.T.G., R.J.F.L., S.A.P., S.H.W., S.J.L.B., T.B.H., T.L., T.Perls, T.J.R., U.V., V.Giedraitis, V.Gudnason, W.Z., W.Kiess, W.M., W.Koenig, Y. L. and Y.M. managed an individual contributing study. A.S.B., A.M.H., A.Tin, A.P., A.P.J.D.V., A.B.Z., A.V.S., A.Teumer, A.G.U., A.Tönjes, A.Köttgen, A.P., A.H., A.Körner, A.Krajčoviechová, A.R., A.Mahajan, A.Y.C., A.G., B.K.K., B.J., B.N., B.Prins, B.W.J.H.P., B.K., B.M.P., C.H., C.A.B., C.N.S., C.Q., C.M., C.F., C.G., C.J.O.D., C.P., D.F.G., D.R., D.M., D.O.M.-K., E.I., E.P.B., E.Catamo, F.K., G.S., G.B., G.G., G.N.N., G.D., G.W.M., H.S., H.C., H.J., H.H., I.R., I.M.N., I.J., I.M.H., J.F.W., J.G.W., J.J., J.Tremblay, J.B.W., J.M., J.C., K.-U.E., K.L.M., K.E., K.B.S., K.Susztak, K.M.R., K.Ho, K.N., K.Strauch, L.M.R., L.-P.L., L.A.L., L.J.O.C., M.L., M.E.K., M.Ciullo, M.L., M.Scholz, M.H.D.B., M.La Bianca, M.M.-N., M.L.Biggs, M.Gorski, M.N., M.Wuttke, M.Waldenberger, M.H.P., M.K.E., M.Kähönen, M.A.N., M.R., N.G.M., N.V., N.H.-K., N.B., O.D., O.T.R., O.D.W., O.P., P.S., P.H., P.K.J., P.V.D.H., Q.Y., R.C., R.Rettig, R.M.L., R.N., R.D.M., R.J.F.L., S.G., S.C.F., S.M.T., S.S., S.A.P., S.H.W., S.D.G., S.-J.H., S.M.K., S.J.L.B., T.B.H., T.N., T.L., T.S.B., T.M., T.L.E., T.J.R., U.T., U.V., V.V., W.H., W.M., W.Koenig, Y.L. and Z.Y. critically reviewed the manuscript. A.V.S., A.Teumer, A. Tin, A.Köttgen, A.H., A.Mahajan, A.Y.C., A.D., A.G., B.J., B.N., B.Prins, B.K., C.A.B., C.N.S., C.Q., C.H.L.T., C.F., C.P., D.N., D.F.G., E.H., E.S., F.R., G.S., G.B., G.D., H.K., I.M.N., I.M.H., J.J., J.Tremblay, J.M., J.L., K.B.S., K.Susztak, K.A.R., K.Horn, K.M.R., L.M.R., L.-P.L., L.A.L., M.L., M.Brumat, M.E.K., M.P.C., M.Scholz, M.Gögele, M.L.Biggs, M.Kanai, M.A., M.Cocca, M.Gorski, M.N., M.Wuttke, M.H.P., M.A.N., M.R., N.S.J., N.P., N.V., N.M., P.P.M., P.H., P.J.V.D.M., P.K.J., P.V.D.H., Q.Y., R.N., R.Rueedi, R.J.C., S.G., S.M.T., S.S., S.A.P., S.-J.H., T.C., T.N., T.S.B., T.L.E., T.H., V.V., W.Z., W.M., Y.S., Y.X., Y.K., Y.L. and Y.O. carried out the statistical methods and analysis. A.P.J.D.V., A.B.Z., A.Teren, A.Metspalu, A.Tönjes, A.K., A.C., B.Ponte, B.J., B.H.S., B.W.J.H.P., C.A.B., C.M., C.P., D.R., D.C., D.J.P., E.P.B., F.K., G.W., H.C., H.J., I.R., I.O., J.F.W., J.G.W., J.S.K., J.Ä., J.Tremblay, J.B.W., J.C.C., K.D., K.N., K.M., M.Ciullo, M.K.E., M.K., M.Kähönen, M.R., N.G.M., N.H.-K., O.T.R., O.P., P.S., P.V., R.S., R.D.M., R.T.G., S.A., S.P., S.A.P., S.H.W., S.V., T.Poulain, T.L., T.J.R., V.Gudnason, W.H. and W.M. recruited the participants. A.V.S., A.K., A.H., A.Y.C., A.G., B.L., B.Prins, C.A.B., C.N.S., C.Q., C.M.S., D.B., D.O.M.-K., E.H., E.Campana, E.S., F.R., G.E., G.P., H.K., I.M.H., J.F.W., J.J., J.Tremblay, J.M., K.L.M., K.B.S., K.Susztak, K.Horn, L.-P.L., M.L., M.E.K., M.P.C., M.Scholz, M.Cocca, M.Gorski, M.Wuttke, M.H.P., N.S.J., N.P., P.P.M., P.H., P.J.V.D.M., R.N., R.M., R.Rueedi, R.J.C., S.G., S.S., S.A.P., S.D.G., S.B., T.C., T.N., W.Z., W.M., Y.S., Y.X., Y.L., Y. M. and Z.Y. carried out the bioinformatics analysis. A.Tin, A.Teumer, A.G.U., A.K., A.G., B.J., C.A.B., C.N.S., C.Q., C.G., C.J.O.D., C.P., H.J., H.K., I.M.H., J.Tremblay, J.M., K.L.M., K.E., K.B.S., K.D., K.Susztak, K.Horn, K.Ho, L.J.O.C., M.L., M.Scholz, M.Gorski, M.Wuttke, M.R., N.V., O.M.W., P.H, P.V.D.H., S.G., S.S., S.A.P., S.-J.H., V.V., V.L.H.K., W.H., W.Koenig, Y.X. and Y.L. interpreted the results. A.B.Z., A.Teumer, A.G.U., A.Köttgen, A.C., A.D., B.H.S., B.W.J.H.P., C.H., C.A.B., C.N.S., C.F., C.M.V.D., D.B., D.R., D.T., D.J.P., D.O.M.-K., E.I., E.S., F.R., F.K., G.E., G.W.M., H.C., J.F.W., J.G.W., J.S.K., J.Ä., J.Tremblay, J.C.C., K.L.M., L.-P.L., L.A.L., M.E.K., M.Waldenberger, M.H.P., M.K.E., M.K., M.Kähönen, M.A.N., N.A., N.M., O.T.R., P.S., P.H., P.K., P.V.D.H., R.B., R.T.G., S.M.T., S.P., S.D.G., S.V., T.L., T.M., U.V., W.H., W.M., W.Koenig, and Y.M. carried out the genotyping. V.H.K., R.L. and O.M.W. carried out the functional study.

Competing interests

D.O.M.-K. works as a part-time clinical research consultant for METABOLON. B.W.J.H.P. has received research funding (unrelated to the work reported in this article) from Jansen Research and Boehringer Ingelheim. K.B.S is full-time employee of GlaxoSmithKline. G.S., D.F.G., I.J., H.H., P.S., U.T. and K.Stefansson are full time employees of deCODE genetics and Amgen. A.Y.C. is an employee of Merck & Co. W.Koenig received modest consultation fees for advisory board meetings from Amgen, DalCor Pharmaceuticals, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi. D.C. is scientific consultant for Bio4Dreams. W.M. is employed by Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. M.A.N.'s participation in this project is supported by a consulting contract between Data Tecnica International, the National Institute on Aging and NIH, and consults or has consulted during this time for Lysosomal Therapeutics, Neuron23, Illumina, the Michael J. Fox Foundation and the University of California Healthcare. O.P. is an owner of Gen-info. K.Ho has disclosed a research and financial relationship with Sanofi Genzyme. B.M.P. serves on the Data and Safety Monitoring Board of a clinical trial funded by the manufacturer (Zoll Lifecor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. A.S.B. received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. F.R. is a scientific consultant for ePhood. M.Scholz consults for and received grant support not related to this project from Merck Serono. A.Köttgen received grant support not related to this project from Grünenthal. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0504-x.

Correspondence and requests for materials should be addressed to A.T. or A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Adrienne Tin[1,2,195*], Jonathan Marten[3,195], Victoria L. Halperin Kuhns[4,195], Yong Li[5,195], Matthias Wuttke[5,195], Holger Kirsten[6,7,195], Karsten B. Sieber[8], Chengxiang Qiu[9], Mathias Gorski[10,11], Zhi Yu[1,12], Ayush Giri[13,14], Gardar Sveinbjornsson[15], Man Li[16], Audrey Y. Chu[17], Anselm Hoppmann[5], Luke J. O'Connor[18], Bram Prins[19], Teresa Nutile[20], Damia Noce[21], Masato Akiyama[22,23], Massimiliano Cocca[24], Sahar Ghasemi[25,26], Peter J. van der Most[27], Katrin Horn[6,7], Yizhe Xu[16], Christian Fuchsberger[21], Sanaz Sedaghat[28], Saima Afaq[29,30], Najaf Amin[28], Johan Ärnlöv[31,32], Stephan J. L. Bakker[33], Nisha Bansal[34,35], Daniela Baptista[36], Sven Bergmann[37,38,39], Mary L. Biggs[40,41], Ginevra Biino[42], Eric Boerwinkle[43], Erwin P. Bottinger[44], Thibaud S. Boutin[3], Marco Brumat[45], Ralph Burkhardt[7,46,47], Eric Campana[45], Archie Campbell[48], Harry Campbell[49], Robert J. Carroll[50], Eulalia Catamo[24], John C. Chambers[29,51,52,53,54], Marina Ciullo[20,55], Maria Pina Concas[24], Josef Coresh[1], Tanguy Corre[37,38,56], Daniele Cusi[57,58], Sala Cinzia Felicita[59], Martin H. de Borst[33], Alessandro De Grandi[21], Renée de Mutsert[60], Aiko P. J. de Vries[61], Graciela Delgado[62], Ayşe Demirkan[28,63], Olivier Devuyst[64], Katalin Dittrich[65,66], Kai-Uwe Eckardt[67,68], Georg Ehret[36], Karlhans Endlich[26,69], Michele K. Evans[70], Ron T. Gansevoort[33], Paolo Gasparini[24,45], Vilmantas Giedraitis[71], Christian Gieger[72,73,74], Giorgia Girotto[24,45], Martin Gögele[21], Scott D. Gordon[75], Daniel F. Gudbjartsson[15], Vilmundur Gudnason[76,77], German Chronic Kidney Disease Study[78], Toomas Haller[79], Pavel Hamet[80,81], Tamara B. Harris[82], Caroline Hayward[3], Andrew A. Hicks[21], Edith Hofer[83,84], Hilma Holm[15], Wei Huang[85,86], Nina Hutri-Kähönen[87,88], Shih-Jen Hwang[89,90], M. Arfan Ikram[28], Raychel M. Lewis[91], Erik Ingelsson[92,93,94,95], Johanna Jakobsdottir[76,96], Ingileif Jonsdottir[15], Helgi Jonsson[97,98], Peter K. Joshi[49], Navya Shilpa Josyula[99], Bettina Jung[10], Mika Kähönen[100], Yoichiro Kamatani[22,101], Masahiro Kanai[22,102], Shona M. Kerr[3], Wieland Kiess[7,65,66], Marcus E. Kleber[62], Wolfgang Koenig[103,104,105], Jaspal S. Kooner[52,53,106,107], Antje Körner[7,65,66], Peter Kovacs[108], Bernhard K. Krämer[62], Florian Kronenberg[109], Michiaki Kubo[110], Brigitte Kühnel[72], Martina La Bianca[24], Leslie A. Lange[111], Benjamin Lehne[29], Terho Lehtimäki[87], Lifelines Cohort Study[78], Jun Liu[28,112], Markus Loeffler[6,7], Ruth J. F. Loos[113,114], Leo-Pekka Lyytikäinen[87], Reedik Magi[79], Anubha Mahajan[115,116], Nicholas G. Martin[75], Winfried März[62,117,118], Deborah Mascalzoni[21], Koichi Matsuda[119], Christa Meisinger[120,121], Thomas Meitinger[104,122,123], Andres Metspalu[79], Yuri Milaneschi[124], V. A. Million Veteran Program[78], Christopher J. O'Donnell[125,126], Otis D. Wilson[127], J. Michael Gaziano[126,128], Pashupati P. Mishra[87], Karen L. Mohlke[129], Nina Mononen[87], Grant W. Montgomery[130], Dennis O. Mook-Kanamori[60,131], Martina Müller-Nurasyid[104,132,133,134], Girish N. Nadkarni[113,135], Mike A. Nalls[136,137], Matthias Nauck[26,138], Kjell Nikus[139,140], Boting Ning[141], Ilja M. Nolte[27], Raymond Noordam[142], Jeffrey R. O'Connell[143], Isleifur Olafsson[144], Sandosh Padmanabhan[145], Brenda W. J. H. Penninx[124], Thomas Perls[146], Annette Peters[73,74,104], Mario Pirastu[147], Nicola Pirastu[49], Giorgio Pistis[148], Ozren Polasek[149,150], Belen Ponte[151], David J. Porteous[48,152], Tanja Poulain[7], Michael H. Preuss[113], Ton J. Rabelink[61,153], Laura M. Raffield[129], Olli T. Raitakari[154,155,156], Rainer Rettig[157], Myriam Rheinberger[10], Kenneth M. Rice[41], Federica Rizzi[158,159], Antonietta Robino[24], Igor Rudan[49], Alena Krajcoviechova[160,161], Renata Cifkova[160,162], Rico Rueedi[37,38], Daniela Ruggiero[20,55],

Kathleen A. Ryan[163], Yasaman Saba[164], Erika Salvi[158,165], Helena Schmidt[166], Reinhold Schmidt[83], Christian M. Shaffer[50], Albert V. Smith[77], Blair H. Smith[167], Cassandra N. Spracklen[129], Konstantin Strauch[132,133], Michael Stumvoll[168], Patrick Sulem[15], Salman M. Tajuddin[70], Andrej Teren[7,169], Joachim Thiery[7,46], Chris H. L. Thio[27], Unnur Thorsteinsdottir[15], Daniela Toniolo[59], Anke Tönjes[170], Johanne Tremblay[80,171], André G. Uitterlinden[172], Simona Vaccargiu[147], Pim van der Harst[173,174,175], Cornelia M. van Duijn[28,112,176], Niek Verweij[173,177], Uwe Völker[26,178], Peter Vollenweider[179], Gerard Waeber[179], Melanie Waldenberger[72,73,104], John B. Whitfield[75], Sarah H. Wild[180], James F. Wilson[3,49], Qiong Yang[141], Weihua Zhang[29,52], Alan B. Zonderman[70], Murielle Bochud[56], James G. Wilson[181], Sarah A. Pendergrass[182], Kevin Ho[183,184], Afshin Parsa[185,186], Peter P. Pramstaller[21], Bruce M. Psaty[187,188], Carsten A. Böger[10,189], Harold Snieder[27], Adam S. Butterworth[190], Yukinori Okada[191,192], Todd L. Edwards[193,194], Kari Stefansson[15], Katalin Susztak[9], Markus Scholz[6,7], Iris M. Heid[11], Adriana M. Hung[127,194,196], Alexander Teumer[25,26,196], Cristian Pattaro[21,196], Owen M. Woodward[4,196], Veronique Vitart[3,196] and Anna Köttgen[1,5,196]*

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [2]Welch Centre for Prevention, Epidemiology and Clinical Research, Baltimore, MD, USA. [3]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [4]Department of Physiology, University of Maryland School of Medicine, Baltimore, MD, USA. [5]Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center—University of Freiburg, Freiburg, Germany. [6]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [7]LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany. [8]Target Sciences—Genetics, GlaxoSmithKline, Collegeville, PA, USA. [9]Department of Medicine and Genetics, University of Pennsylvania, Philadelphia, PA, USA. [10]Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. [11]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [12]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [13]Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA. [14]Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. [15]deCODE Genetics, Amgen Inc., Reykjavik, Iceland. [16]Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, UT, USA. [17]Genetics, Merck & Co., Inc, Kenilworth, NJ, USA. [18]Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [19]Strangeways Research Laboratory, University of Cambridge, Cambridge, UK. [20]Institute of Genetics and Biophysics Adriano Buzzati-Traverso—CNR, Naples, Italy. [21]Eurac Research, Institute for Biomedicine, Bolzano, Italy. [22]Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences, Yokohama (Kanagawa), Japan. [23]Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [24]Institute for Maternal and Child Health—IRCCS Burlo Garofolo, Trieste, Italy. [25]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [26]DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany. [27]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [28]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [29]Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK. [30]Institute of Public Health & Social Sciences, Khyber Medical University, Peshawar, Pakistan. [31]Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden. [32]School of Health and Social Studies, Dalarna University, Falun, Sweden. [33]Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [34]Division of Nephrology, University of Washington, Seattle, WA, USA. [35]Kidney Research Institute, University of Washington, Seattle, WA, USA. [36]Cardiology, Geneva University Hospitals, Geneva, Switzerland. [37]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [38]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [39]Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa. [40]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [41]Department of Biostatistics, University of Washington, Seattle, WA, USA. [42]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. [43]Human Genetics Centre, University of Texas Health Science Centre, Houston, TX, USA. [44]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [45]University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy. [46]Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany. [47]Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. [48]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [49]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [50]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [51]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [52]Department of Cardiology, Ealing Hospital, London, UK. [53]Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [54]MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [55]IRCCS Neuromed, Pozzilli, Italy. [56]Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland. [57]Institute of Biomedical Technologies, Italy National Research Council, Milano, Italy. [58]Bio4Dreams, Milano, Italy. [59]San Raffaele Research Institute, Milano, Italy. [60]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, the Netherlands. [61]Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, the Netherlands. [62]Fifth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [63]Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands. [64]Institute of Physiology, University of Zurich, Zurich, Switzerland. [65]Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany. [66]Centre for Pediatric Research, University of Leipzig, Leipzig, Germany. [67]Department of Nephrology and Medical Intensive Care, Charité—Universitätsmedizin Berlin, Berlin, Germany. [68]Department of Nephrology and

Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany. [69]Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany. [70]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore, MD, USA. [71]Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden. [72]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Centre for Environmental Health, Neuherberg, Germany. [73]Institute of Epidemiology, Helmholtz Zentrum München—German Research Centre for Environmental Health, Neuherberg, Germany. [74]German Center for Diabetes Research, Neuherberg, Germany. [75]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [76]Icelandic Heart Association, Kópavogur, Iceland. [77]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [78]A list of members and affiliations appears in the Supplementary Note. [79]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [80]Montreal University Hospital Research Centre, Centre Hospitalier de lUniversité de Montréal, Montreal, Quebec, Canada. [81]Medpharmgene, Montreal, Quebec, Canada. [82]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, USA. [83]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. [84]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. [85]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China. [86]Shanghai Industrial Technology Institute, Shanghai, China. [87]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [88]Department of Pediatrics, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [89]National Heart, Lung, and Blood Institute Framingham Heart Study, Framingham, MA, USA. [90]The Centre for Population Studies, National Heart, Lung, and Blood Institute, Framingham, MA, USA. [91]Department of Physiology, University of Maryland School of Medicine, Baltimore, MD, USA. [92]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA. [93]Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. [94]Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [95]Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA. [96]The Centre of Public Health Sciences, University of Iceland, Reykjavik, Iceland. [97]Landspitalinn University Hospital, Reykjavík, Iceland. [98]University of Iceland, Reykjavík, Iceland. [99]Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA. [100]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [101]Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [102]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [103]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [104]German Centre for Cardiovascular Research, Partner Site Munich Heart Alliance, Munich, Germany. [105]Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany. [106]MRC-PHE Centre for Environment and Health, 323 School of Public Health, Imperial College London, London, UK. [107]National Heart and Lung Institute, Imperial College London, London, UK. [108]Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany. [109]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. [110]RIKEN Centre for Integrative Medical Sciences, Yokohama (Kanagawa), Japan. [111]Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver—Anschutz Medical Campus, Aurora, CO, USA. [112]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [113]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [114]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [115]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [116]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [117]Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany. [118]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. [119]Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [120]Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany. [121]Ludwig-Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany. [122]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. [123]Institute of Human Genetics, Technische Universität München, Munich, Germany. [124]Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, the Netherlands. [125]VA Boston Healthcare System, Boston, MA, USA. [126]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [127]Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA. [128]Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA. [129]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [130]University of Queensland, St Lucia, Queensland, Australia. [131]Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, the Netherlands. [132]Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Centre for Environmental Health, Neuherberg, Germany. [133]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [134]Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University Munich, Munich, Germany. [135]Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [136]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [137]Data Tecnica International, Glen Echo, MD, USA. [138]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [139]Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland. [140]Department of Cardiology, Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [141]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [142]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, the Netherlands. [143]University of Maryland School of Medicine, Baltimore, MD, USA. [144]Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland. [145]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [146]Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA. [147]Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Sassari, Italy. [148]Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland. [149]Faculty of Medicine, University of Split, Split, Croatia. [150]Gen-info Ltd, Zagreb, Croatia. [151]Nephrology Service, Department of Specialties in Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland. [152]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [153]Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, the Netherlands. [154]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [155]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [156]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. [157]Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany. [158]Department of Health Sciences, University of Milan, Milano, Italy. [159]ePhood Scientific Unit, ePhood SRL, Milano, Italy. [160]Center for Cardiovascular Prevention, First Faculty of Medicine, Charles University and Thomayer Hospital, Prague, Czech Republic. [161]Thomayer Hospital, Prague, Czech Republic. [162]Department of Medicine II, First Faculty of Medicine, Charles University, Prague, Czech Republic. [163]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. [164]Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria. [165]Neurology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy. [166]Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria. [167]Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [168]Division of Endocrinology, Nephrology and Rheumatology, University of

Leipzig, Leipzig, Germany. [169]Heart Centre Leipzig, Leipzig, Germany. [170]Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany. [171]Centre de Recherche du CHUM, Montreal, Quebec, Canada. [172]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [173]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [174]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [175]Durrer Centre for Cardiovascular Research, the Netherlands Heart Institute, Utrecht, the Netherlands. [176]Leiden Academic Centre for Drug Research, Leiden University, Leiden, the Netherlands. [177]Genomics plc, Oxford, UK. [178]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [179]Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland. [180]Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [181]Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson, MS, USA. [182]Geisinger Research, Biomedical and Translational Informatics Institute, Danville, PA, USA. [183]Kidney Health Research Institute, Geisinger, Danville, PA, USA. [184]Department of Nephrology, Geisinger, Danville, PA, USA. [185]Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [186]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [187]Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle, WA, USA. [188]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [189]Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Traunstein, Germany. [190]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [191]Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences, Osaka, Japan. [192]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [193]Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA. [194]Department of Veterans Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. [195]These authors contributed equally: Adrienne Tin, Jonathan Marten, Victoria L. Halperin Kuhns, Yong Li, Matthias Wuttke, Holger Kirsten. [196]These authors jointly supervised this work: Adriana M. Hung, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen.
*e-mail: atin1@jhu.edu; anna.koettgen@uniklinik-freiburg.de

Methods

Phenotype definition, genotyping and imputation in participating studies. The primary study outcome was serum urate in mg dl^{-1}. The laboratory methods for measuring serum urate in each study are reported in Supplementary Table 1. Prevalent gout was analyzed as a secondary outcome to examine whether urate-associated SNPs conferred gout risk. Gout cases were ascertained based on self-report, intake of urate-lowering medications or International Statistical Classification of Diseases and Related Health Problems (ICD) codes for gout (Supplementary Table 1). The participants of all studies provided written informed consent. Each study had its research protocol approved by the corresponding local ethics committee.

Each study performed genotyping separately and imputed the genotypes to the reference panels of the Haplotype Reference Consortium v.1.1 (ref. [69]), 1000 Genomes Project phase 3 v.5 ALL or the 1000 Genomes Project phase 1 v.3 ALL[70]. Study-specific quality filters and software used for phasing and imputation are provided in Supplementary Table 2 and the Supplementary Note. Variants were annotated using NCBI b37 (hg19).

Study-specific association analysis. Phenotype generation was standardized across studies using a common script; study-specific association analyses followed a centrally developed analysis plan. GWAS summary statistics were checked centrally using GWAtoolbox v.2.2.4-7 (ref. [71]) and custom scripts (Supplementary Note). Each study performed ancestry-specific association analysis of serum urate by generating age- and sex-adjusted residuals of serum urate and regressing the residuals on SNP dosage levels, adjusting for study-specific covariates, such as study centers and genetic principal components, assuming an additive genetic model. Gout was analyzed as a binary outcome adjusting for age, sex, genetic principal components and study-specific covariates. The software programs used for these regression analyses were: EPACTS (q.emmax for family-based studies and q.linear otherwise; https://genome.sph.umich.edu/wiki/EPACTS); SNPTEST[72]; RegScan[73]; RVTESTS[74]; PLINK 1.90 (ref. [75]); ProbABEL[76]; GWAF[77]; GEMMA[78]; mach2qtl[79]; and R. Family-based studies used methods that accounted for relatedness.

Trans-ancestry, ancestry-specific and sex-stratified meta-analyses. GWAS results from each study were prefiltered to retain biallelic SNPs with an imputation quality score >0.6 and minor allele count (MAC) > 10 before inclusion into the meta-analysis. Fixed-effect inverse-variance weighted meta-analysis was performed using METAL[80] (released 25 March 2011) with modifications to output higher precision (six decimal places). Genomic control was applied for each study. The genomic control inflation factor λ_{GC} (ref. [81]) was calculated to assess inflation of the test statistics. For each meta-analysis result (trans-ancestry, ancestry-specific and sex-specific), we excluded SNPs that were present in <50% of the studies and with a total MAC < 400. For ancestry-specific meta-analysis, we additionally excluded SNPs with a heterogeneity I^2-statistic[82] >95%. Genome-wide significance was defined as $P < 5\times10^{-8}$. The LD score regression intercept was calculated to assess the evidence for associations driven by population structure[83]. For downstream characterization, 8,249,849 and 8,217,339 autosomal SNPs were retained in the trans-ancestry and European ancestry meta-analysis, respectively. Ancestry-specific meta-analyses were conducted for European, African American, East Asian and South Asian ancestries using the same methods and variant filters as the trans-ancestry meta-analysis.

Secondary meta-analyses were performed separately in men and women, using the same analytical approaches. To test for significant difference of association between men and women, we used a two-sample t-test:

$$t = \frac{\beta_M - \beta_F}{\sqrt{SE_M^2 + SE_F^2}}$$

where β_M and β_F are the β coefficients in men and women, respectively, and SE_M and SE_F were the standard errors of the β coefficients in men and women, respectively.

Initial determination and annotation of genome-wide significant loci. For each meta-analysis result, the SNP with the lowest P per chromosome was selected as an initial index SNP; along with the ±500 kilobases (kb) surrounding, it was defined as one 1-Mb locus. This procedure was repeated with the SNP with the lowest P not yet assigned to a locus, until no genome-wide significant SNPs outside the 1-Mb loci remained. To visualize the loci, the genomic region ±500 kb around each index SNP was plotted and could contain two index SNPs when the index SNPs were >500 kb but <1 Mb apart. An ancestry-specific locus was defined as a genome-wide significant locus in an ancestry-specific meta-analysis where the index SNP did not map to within the ±500 kb intervals of any genome-wide significant loci in the trans-ancestry meta-analysis. Index SNPs were annotated using their position and the nearest gene based on hg19, RefSeq genes and dbSNP database (build 147) downloaded from ftp://hgdownload.soe.ucsc.edu/mysql/hg19/ on 23 March 2017.

Proportion of phenotypic variance explained and estimated heritability. The proportion of phenotypic variance explained by index SNPs was calculated as the sum of the variance explained by each index SNP based on this formula: $\beta^2\left(\frac{2p(1-p)}{\mathrm{var}}\right)$, where β is the β coefficient, p is the minor allele frequency of the SNP and var is the phenotypic variance. For this study, we used the variance of the age- and sex-adjusted residuals of serum urate in European ancestry participants of the Atherosclerosis Risk in Communities (ARIC) study as the estimate of the phenotypic variance (var = 1.767).

The genetic heritability of age- and sex-adjusted urate levels was estimated using the R package MCMCglmm v.2.26 (ref. [84]) in the Cooperative Health Research In South Tyrol (CHRIS) study[85], a participating European ancestry study with 4,373 individuals split into 186 up to 5 generation pedigrees[86]. Genetic heritability was estimated overall, after accounting for the index SNPs of the three major urate loci (*SLC2A9*, *ABCG2* and *SLC22A12*), and after accounting for the index SNPs of all genome-wide significant loci for both the trans-ancestry and European ancestry-specific meta-analyses. Estimates were obtained by running 1,000,000 Markov chain Monte Carlo iterations (burn in = 500,000) based on previously described settings[86]. The difference between overall heritability and heritability excluding the index SNPs represents the heritability explained by the identified loci.

Trans-ancestry meta-regression. Before conducting the trans-ancestry meta-regression, we applied the same study-specific SNP filters as those applied to the fixed effects trans-ancestry meta-analysis (imputation quality score > 0.6 and MAC > 10). An additional filter for minor allele frequency > 0.0025 was also applied to reduce the influence of rare SNPs that passed the MAC filter in very large studies. Trans-ancestry meta-regression was conducted using the MR-MEGA software package v.0.1.2 (ref. [87]), which models ancestry-associated heterogeneity in the allelic effect as a function of principal components generated from a matrix of mean pairwise allele frequency differences between studies. Three principal components generated from a matrix of mean pairwise allele frequency differences between studies were sufficient to separate the self-reported ancestry groups. Due to software requirements, the minimum number of cohorts for each SNP had to be greater than the number of principal components plus two, resulting in the exclusion of SNPs present in five or fewer cohorts. In addition to genome-wide SNP associations with urate, MR-MEGA reports ancestry-associated ($P_{\text{anc-het}}$) and residual ($P_{\text{res-het}}$) heterogeneity. Index SNPs from the fixed effects meta-analysis with $P_{\text{anc-het}} < 2.7\times10^{-4}$ (0.05/183) in MR-MEGA were considered to have significant ancestry-associated heterogeneity.

Effect of urate-associated index SNPs on gout and risk prediction for gout. To evaluate the association of the trans-ancestry urate-associated index SNPs with gout, we conducted a trans-ancestry meta-analysis of gout with the same study-specific filtering criteria as for the urate trans-ancestry meta-analysis.

The association between a genetic urate risk score constructed from the 114 independent serum urate-associated SNPs identified in European individuals (see Statistical fine-mapping of genome-wide significant loci in European ancestry section) and gout was assessed in a large, independent sample from the UKBB (project nos. 19655 and 20272)[88]. We selected 334,880 unrelated individuals (pairwise kinship coefficient < 0.0313) of white British ancestry with sex chromosome euploidy and concordance of phenotypic and genotypic sex, including 4,908 with gout identified by self-report at the inclusion visit. Individuals with an ICD-10 for gout (M10) in hospital admissions who did not self-report gout were excluded from the analysis. A GRS was constructed as the sum of the imputed dosage of the allele associated with higher urate levels (risk alleles) over all SNPs, multiplied by the genetic effect of the risk allele on serum urate levels. The GRS distribution was divided into ten evenly spaced categories and individuals were assigned to a category based on their GRS. The category with the lowest GRS did not contain any gout cases and was thus combined with its adjacent category. Gout status was regressed on GRS category in a logistic model, including age and sex as covariates, with the category containing the largest number of individuals (genetically predicted mean urate levels 4.74–5.02 mg dl^{-1} higher compared to individuals without any urate-increasing alleles) as the reference group.

The performance of the GRS for risk prediction of gout was first evaluated in a randomly selected model development sample comprising 90% of the participants to obtain precise estimates and tested in a validation sample of the remaining 10%. Logistic regression was used to regress gout on the GRS alone (genetic model), age and sex (demographic model) and GRS with age and sex (combined model) in the model development sample. Each of these models was then used to predict gout status in the validation sample. Model performance was assessed by comparing predicted and true gout status using the AUROC curve. A cutoff of the AUROC curve to report the sensitivity and specificity of a combined GRS-based diagnostic test was determined by the maximum of the Youden's index (sensitivity + specificity − 1). Tenfold cross-validation of the models was performed by randomly dividing the UKBB sample into ten equally sized groups. Each group in turn was used as the validation sample for the estimates developed on the remaining data. The AUROC curve was calculated for each of the three models for all ten validation samples and the means and s.d. are reported.

Genetic correlation. To assess the genetic correlation between serum urate and other traits in the European ancestry group, we conducted cross-trait LD score regression[89] using LD Hub v.1.3 (ref. [90]) with the European ancestry-specific urate meta-analysis results as input. Genetic correlation estimates with 746 traits were obtained from LD Hub, excluding two previous serum urate GWAS results. For presentation, the 212 significantly correlated traits ($P < 6.7 \times 10^{-5} = 0.05/746$) were grouped into 9 categories based on the trait names and labels and presented in a circos plot.

To determine whether the observed genetic correlations between serum urate and cardiometabolic traits represented causal relationships, we used the recently developed LCV method to estimate the GCP between serum urate and another trait[91]. Compared to multiple regression, the LCV method produces fewer false positive results in the setting of high genetic correlation and large sample sizes, a situation applicable to our analysis[91]. The GCP describes what proportion of the genetic component of one trait also affects the other trait; a positive GCP value indicates that a proportion of the genetic component of urate affects the other trait, and vice versa for a negative GCP value. LCV produces posterior mean and s.d. estimates of the GCP using mixed fourth moments of the bivariate effect size distribution, based on GWAS summary statistics and LD scores. When using the summary statistics of cardiometabolic traits generated from the UKBB, we assumed nonoverlapping populations, and overlapping populations otherwise. We selected six unique continuous cardiometabolic traits commonly examined in epidemiological studies with high genetic correlation with serum urate ($r_g > 0.35$). We additionally included gout as a positive control and creatinine-based GFR. European ancestry-specific GWAS summary statistics were used as input to match the ancestry of the LD scores used with the method (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2).

Functional enrichment. To assess gene set and tissue enrichment, we used the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) analysis v.1 release 194 (ref. [92]), which performed gene set enrichment analysis by testing whether genes in 14,461 reconstituted gene sets were enriched for urate-associated SNPs ($P < 1 \times 10^{-5}$) from the trans-ancestry meta-analysis results. Affinity propagation clustering[93], implemented in the R package APCluster v.1.4.5 (ref. [94]), was applied to all urate-associated reconstituted gene sets with an FDR-corrected enrichment $P < 0.01$ to cluster gene sets containing similar combinations of genes. More details on the methods of DEPICT and affinity propagation clustering are provided in the Supplementary Note. The methods for using stratified LD score regression[89] based on cell type-specific genomic annotations to identify cell type- and tissue-specific enrichments of serum urate heritability are reported in the Supplementary Note.

Statistical fine-mapping of genome-wide significant loci in European ancestry. Statistical fine-mapping to identify potentially causal variants was performed for the genome-wide significant loci from the European ancestry-specific meta-analysis. LD was estimated based on 16,969,363 SNPs from 13,558 unrelated UKBB participants after quality control (Supplementary Note). The analyses were based on a previously described workflow[39,40,95] using genome-wide complex trait analysis (cojo-slct option) to identify independent index SNPs in each region, followed by using genome-wide complex trait analysis (cojo-cond option) to obtain conditional β and s.e.m. for regions with >1 independent signal. Next, approximate Bayes factors were calculated using Wakefield's formula[38], as implemented in the R package gtx v.2.1.0. (https://github.com/tobyjohnson/gtx). The posterior probability for a variant being the driver of the association signal was calculated as the approximate Bayes factors of the variant divided by the sum of the approximate Bayes factors in the region. The 99% credible sets of a region were derived by summing the posterior probabilities in descending order until the cumulative posterior probability was >99%. We prioritized variants in credible sets containing ≤5 SNPs or SNPs with posterior probabilities >0.5. More details on statistical fine-mapping are provided in the Supplementary Note.

Annotation of the variants in the credible sets. We annotated SNPs in the credible sets for their exonic effect, CADD score and mapping to DNaseI hypersensitive sites (DHS) from the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics Consortium projects[96,97]. The exonic effect and CADD score were obtained using SNiPA v.3.2 (March 2017) (ref. [98]). SNiPA presented the CADD score as Phred-like transformation of the C score, which was based on CADD v.1.3 downloaded from http://cadd.gs.washington.edu/download. A CADD score of 15 is used to distinguish potentially deleterious variants from background noise in clinical genetics and represents the median value of all nonsynonymous variants in CADD v.1.0 (refs. [99,100]). As opposed to posterior probabilities causing the association signal, CADD scores represent an integrative measure of predicted deleteriousness based on an ensemble of variant annotations derived by contrasting common variants that survived natural selection with simulated mutations. Based on known pathogenic variants in the ClinVar database, the performance of the CADD score had an AUROC of 0.88 (ref. [101]).

Colocalization analysis of *cis*-eQTL and urate-associated loci. Colocalization analysis of urate-associated loci with gene expression was conducted using the European ancestry meta-analysis results, *cis*-eQTL results from microdissected human glomerular and tubulointerstitial kidney portions from 187 individuals in the Nephrotic Syndrome Study Network (NEPTUNE) study[102], as well as from 44 tissues in the GTEx Project (version 6p)[103]. For each urate locus, we identified all transcripts and all tissue transcript pairs with reported eQTLs within ±100 kb of each GWAS index SNP. The region for each colocalization test was defined as the eQTL *cis*-window in the underlying studies[102,103]. We used the default parameters and prior definitions set in the coloc.fast function from the R package gtx v.2.1.0, which is an adapted implementation of the colocalization method outlined by Giambartolomei et al.[24]. Evidence for colocalization was defined as an H4 ≥ 0.8, which represents the posterior probability that the association with serum urate and gene expression is due to the same underlying variant. In addition, colocalization of urate-associated loci was also performed with gene expression quantified using RNA sequencing of the healthy tissue portion of 99 kidney cortex samples from The Cancer Genome Atlas (TCGA)[104]. First, all transcripts that shared eQTL variants with urate index SNPs within ±100 kb were extracted. Then, the posterior probability of colocalization was calculated including eQTLs within the *cis*-window (±1 Mb from the transcription start site) for each gene using the R coloc package v.3.1 (ref. [24]) with default values for the three prior probabilities. The methods for *trans*-eQTL annotation are reported in the Supplementary Note.

Experimental study. *Promoter binding site predictions.* For promoter binding site predictions, we used the JASPAR 2018 database[105,106]. The frequency matrices were downloaded for transcription factor binding sites of both vertebrate and human sequences (HNF1A: MA0046.1 and MA0046.2; HNF4A: MA0114.1 and MA0114.2). These matrices were then used to query the promoter region of *ABCG2* (−1285/+362, or base pairs upstream of the transcription start site and/or downstream after the transcription start site)[107] using the LASAGNA 2.0 transcription factor binding site search tool with default parameters and a P cutoff of 0.01 (ref. [108]).

Site-directed mutagenesis. HNF1A and HNF4A clones were purchased from GeneCopoeia (catalog nos. EX-A7792-M02 and EX-Z5283-M02, respectively) and were mutagenized using the QuikChange Lightning Site-Directed Mutagenesis Kit (catalog no. 210518; Agilent Technologies) according to the manufacturer's instructions using polyacrylamide gel electrophoresis-purified primers, which are reported in the Supplementary Note.

Luciferase assay. HEK 293T cells were seeded in white-walled 96-well plates coated with Poly-L-lysine (Sigma P1274; 5 mg per 50 ml in water) at roughly 12,500 cells per well. Cells were transfected 18 h later with either the *ABCG2* promoter (−1285/+362) upstream of a firefly luciferase in the pGL4.14 vector (a generous gift from D.D. Ross) or the pGL4.14 vector (catalog no. E699A; Promega Corporation) without promoter construct, as well as a green fluorescent protein (GFP)-expressing vector used as an internal negative control (pEGFP-C1; Clontech)[109] using the X-tremeGENE 9 DNA Transfection Reagent (catalog no. 6365787001; Roche). Transfection cocktails were prepared according to the manufacturer's specifications either with or without transcription factor using the following ratio: 0.6 µg promoter construct; 0.2, 0.1 or 0.05 µg transcription factor; and 0.05 µg GFP. When no transcription factor was used, pcDNA3.1 was substituted. Approximately 48 h after transfection, cells were rinsed with 1× PBS, then lysed using Passive Lysis Buffer (catalog no. E194A; Promega Corporation) for 15 min. During this incubation, GFP measurements were taken using a CLARIOstar Plus microplate reader (BMG Labtech). Next, 30 µl of luciferase reagent (catalog no. E297A&B; Promega Corporation) were added to each well and the plate was incubated for an additional 20 min at room temperature. Finally, luciferase activity was measured using the CLARIOstar Plus microplate reader taking the average over 6 s. To evaluate the significance of transactivation of the *ABCG2* promoter, we compared cells expressing transcription factors to those transfected with the empty vector (pcDAN3.1); to evaluate transcription factor dose responses or differences in transcription factor variants, all experimental conditions from one plate were compared using an ordinary one-way analysis of variance (ANOVA), accounting for multiple comparisons with a Tukey's multiple comparison test. Statistical analysis was performed using Prism 7 (GraphPad Software).

Immunoblots. Equal volumes of deoxycholate-radioimmunoprecipitation assay buffer were added to wells containing the desired lysates following the luciferase assay; plates were then incubated at 4 °C overnight. Equal volumes of sample + 5× SDS loading dye + 10% β-mercaptoethanol were then loaded into 10% Mini-PROTEAN TGX Stain-Free Protein Gels (catalog no. 4568033; Bio-Rad Laboratories) and run according to the manufacturer's specifications. Gels were then cross-linked for 45 s and imaged to reveal the total protein load, which was used as the loading control for each lane. Representative images of these protein gels are found in Supplementary Fig. 8. Gels were then transferred onto nitrocellulose membranes using the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories), blocked for 2 h at room temperature in 5% milk in tris-buffered saline and Tween 20 (TBST) and incubated overnight at 4 °C with primary antibody. Membranes were then washed 3 times with TBST, incubated at room temperature for 1 h with goat anti-rabbit secondary antibody (catalog no. 111-035-144; Jackson ImmunoResearch) diluted 1:5,000 in 2.5% milk in TBST. Membranes were then washed again and developed using SuperSignal West Pico

PLUS Chemiluminescent Substrate (catalog no. 34577; Thermo Fisher Scientific) and imaged on the ChemiDoc MP Imaging System (Bio-Rad Laboratories). All primary antibodies were diluted 1:1,000 in 2.5% milk in TBST. Antibodies used included HNF4A (catalog no. 3113S; Cell Signaling Technology) and HNF1A (catalog no. 89670S; Cell Signaling Technology). Antibodies were validated using lysates of overexpressing HEK 293T cells transfected with either HNF construct, demonstrating bands at the appropriate sizes (Supplementary Fig. 8).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome-wide summary statistics for this study are available at the CKDGen Consortium (http://ckdgen.imbi.uni-freiburg.de) and will be made publicly available through the database of Genotypes and Phenotypes accession no. phs000930.v6.p1.

References

69. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
70. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
71. Fuchsberger, C., Taliun, D., Pramstaller, P. P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444–445 (2012).
72. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
73. Haller, T., Kals, M., Esko, T., Mägi, R. & Fischer, K. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Brief. Bioinform.* **16**, 39–44 (2015).
74. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
75. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
76. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
77. Chen, M. H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581 (2010).
78. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
79. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
80. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
81. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* **60**, 155–166 (2001).
82. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
83. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
84. Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *J. Stat. Softw.* **33**, 1–22 (2010).
85. Pattaro, C. et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J. Transl. Med.* **13**, 348 (2015).
86. Noce, D. et al. Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum. Genet.* **136**, 743–757 (2017).
87. Mägi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
88. Sudlow, C. 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 Med.* **12**, e1001779 (2015).
89. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
90. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
91. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
92. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
93. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
94. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464 (2011).
95. Wuttke, M. et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
96. Sheffield, N. C. et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* **23**, 777–788 (2013).
97. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
98. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).
99. Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
100. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
101. Li, J. et al. Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res.* **46**, 7793–7804 (2018).
102. Gillies, C. E. et al. An eQTL landscape of kidney tissue in human nephrotic syndrome. *Am. J. Hum. Genet.* **103**, 232–244 (2018).
103. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
104. Ko, Y. A. et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am. J. Hum. Genet.* **100**, 940–953 (2017).
105. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
106. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
107. Xie, Y. et al. Functional cyclic AMP response element in the breast cancer resistance protein (BCRP/ABCG2) promoter modulates epidermal growth factor receptor pathway- or androgen withdrawal-mediated BCRP/ABCG2 transcription in human cancer cells. *Biochim. Biophys. Acta* **1849**, 317–327 (2015).
108. Lee, C. & Huang, C. H. LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser. *Bioinformatics* **30**, 1923–1925 (2014).
109. Vesuna, F., Winnard, P. Jr. & Raman, V. Enhanced green fluorescent protein as an alternative control reporter to *Renilla* luciferase. *Anal. Biochem.* **342**, 345–347 (2005).

nature research

Corresponding author(s): Adrienne Tin
Anna Köttgen

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	To maximize phenotype standardization across studies, an analysis plan and a command line script (https://github.com/genepi-freiburg/ckdgen-pheno) were created centrally and provided to all participating studies. Data processing, analysis and troubleshooting instructions were distributed to all studies via a Wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were checked centrally. Upon phenotype approval, studies run their GWAS and uploaded results and imputation quality (IQ) information to a common calculation server. GWAS QC was performed using GWAtoolbox and custom (R, Bash) scripts to assess ancestry-matched allele frequencies and variant positions.
Data analysis	• Software tools for imputation: Michigan imputation server, Minimac 3 or 4, IMPUTE2 v2.3.2 or v4, Mach v1. 0.16, Sanger server (reported in Supplementary Table 2) • Software tools for association analysis: probABEL, MMAP, SNPTEST v2, EPACTS v3.2.6, v2.5.2, mach2dat/mach2qtl, custom R code, RegScan v.02, PLINK 1.90, rvtest, GWAF, PLATO v0.0.1, BOLT-LMM v.2.3 (reported in Supplementary Table 2). • QC tool: GWAtoolbox - https://github.com/cran/GWAtoolbox • Meta-analysis: METAL - http://csg.sph.umich.edu/abecasis/metal/download/ • Meta-regression: MrMega - https://www.geenivaramu.ee/en/tools/mr-mega • Genetic heritability: MCMCglmm - https://cran.r-project.org/web/packages/MCMCglmm/index.html • LD Score regression: ldsc - http://ldsc.broadinstitute.org • Functional enrichment: DEPICT - https://data.broadinstitute.org/mpg/depict/documentation.html • Pathway visualization: Cytoscape - http://cytoscape.org • Enrichment by heritability: LDHub - http://ldsc.broadinstitute.org/ldhub/

• Conditional analyses: GCTA - http://cnsgenomics.com/software/gcta/
• Colocalization: coloc.fast - https://github.com/tobyjohnson/gtx
• Genetic causal proportion: https://github.com/lukejoconnor/LCV
• Functional annotation of variants in credible sets: SNiPA v3.2 (http://snipa.helmholtz-muenchen.de/snipa3/), CADD v1.3 (http://cadd.gs.washington.edu/download)
• Promoter Binding Site Predictions: JASPAR 2018 database - jaspar.genereg.net/
• Promoter Binding Site Predictions: LASAGNA 2.0 - https://biogrid-lasagna.engr.uconn.edu/lasagna_search/
• Luciferase analysis of variance: Prism 7 (GraphPad Software Inc, USA)
• Miscellanea of software: R, Perl, Bash

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The following data availability statement is included:
The meta-analysis summary statistics are made publicly available at https://ckdgen.imbi.uni-freiburg.de/ and will be submitted to dbGaP accession number phs000930.v6.p1

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The study sample size was of 457,690 individuals for trans-ethnic serum urate meta-analysis, 334,880 for gout prediction, and 763,813 for gout trans-ethnic meta-analysis. We sought to include as many individuals as possible to increase the potential for discovering new loci. With a larger sample size than the previous meta-analyses of GWAS of serum urate, we identified new loci and confirmed previous loci.
Data exclusions	Pre-established exclusion criteria were set as: (1) for quantitative trait analysis, studies with <100 samples were excluded from analysis; (2) for binary trait analysis, studies with <100 cases or controls were excluded from analysis; (3) studies with no genome-wide SNP array data imputed based on the Haplotype Reference Consortium version 1.1 or the 1000 Genomes Project phase 3 version 5 ALL or phase 1 version 3 ALL panels were excluded from analysis (e.g.: metabochip-based studies). The rationale for exclusion criteria were sufficient sample sizes for reliable estimation of test statistics, and to obtain similar coverage of genome-wide common genetic variants across studies.
Replication	The results are based on a meta-analysis and hence already on combined evidence from 74 studies with low evidence of heterogeneity.
Randomization	Not relevant to this study because this is an observational study
Blinding	Not relevant to this study because this is an observational study

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	HNF4a (C11F12; lot 2) and HNF1a (D7Z2Q; lot 1); both from Cell Signaling. GAPDH (3E8AD9; A21994) from Invitrogen.
Validation	Mock transfected lysates and transfected (either HNF1a or HNF4a) lysates were compared. Both antibodies only detected a signal of any size in the transfected lysates (presented in Supplementary Figure 9).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T, GenHunter Corp, USA
Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See ICLAC register)	N/A

Human research participants

Policy information about studies involving human research participants

Population characteristics	All but two cohorts were adult studies with mean age ranged from 37.6 years to 76.4 years. The mean age of the children cohorts were 12.9 and 15.4 years. One cohort was exclusively female, and 4 cohorts were exclusively male. The proportion of males in the other cohorts ranged from 33% to 93% with a median of 45%. Across studies, there were 288,649 participants of European-ancestry, 125,725 of East Asian ancestry, 33,671 African Americans, 9,037 of South Asian ancestry; and 608 Hispanics More details are prpvided in Supplementary Table 1.
Recruitment	The types of studies were: population-based cohort studies (n=43), studies of isolated populations (n=8), case-control studies (n=5), patient populations (n=5), convenience samples (n=2), family-based studies (n=2), veteran population (n=1). Given that most cohorts were population-based and we observed little heterogeneity among the identified loci of serum urate, we expect that the sample selection did not introduce major biases in our main findings. All studies were approved by their respective institutional review boards. Participants provided informed consent. More details are provided in Supplementary Table 2.