A genome-wide association analysis reveals new pathogenic pathways in gout

Major, T.J.; Takei, R.; Matsuo, H.; Leask, M.P.; Sumpter, N.A.; Topless, R.K.; ... ; 23andme Res Team

nature genetics

Article

https://doi.org/10.1038/s41588-024-01921-5

A genome-wide association analysis reveals new pathogenic pathways in gout

Received: 17 January 2023

Accepted: 21 August 2024

Published online: 15 October 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Gout is a chronic disease that is caused by an innate immune response to deposited monosodium urate crystals in the setting of hyperuricemia. Here, we provide insights into the molecular mechanism of the poorly understood inflammatory component of gout from a genome-wide association study (GWAS) of 2.6 million people, including 120,295 people with prevalent gout. We detected 377 loci and 410 genetically independent signals (149 previously unreported loci in urate and gout). An additional 65 loci with signals in urate (from a GWAS of 630,117 individuals) but not gout were identified. A prioritization scheme identified candidate genes in the inflammatory process of gout, including genes involved in epigenetic remodeling, cell osmolarity and regulation of NOD-like receptor protein 3 (NLRP3) inflammasome activity. Mendelian randomization analysis provided evidence for a causal role of clonal hematopoiesis of indeterminate potential in gout. Our study identifies candidate genes and molecular processes in the inflammatory pathogenesis of gout suitable for follow-up studies.

Gout is an important public health issue of increasing prevalence[1] and burden[2] exacerbated by comorbidity with cardiometabolic and renal disease[3]. Gout typically presents as recurrent self-limiting flares of extremely painful acute inflammatory arthritis. It is a chronic disease of monosodium urate (MSU) crystal deposition mainly in joints that occurs in the setting of hyperuricemia[4]. Flares occur when resident macrophages interact with MSU crystals to form and activate the NLRP3 inflammasome, resulting in production of mature interleukin (IL)-1β and IL-18 (ref. 5). Soluble urate trains the immune system to have an enhanced response to MSU crystals through durable epigenetic modifications[6].

GWAS have provided insight into molecular mechanisms of urate control, with loci encoding renal and intestinal urate transporters prominent[7]. There have been fewer GWAS in gout[7–11]; the largest studies (13,179 and 37,105 people with gout[7,11]) identified 27 and 52 gout-associated loci, respectively, the considerable majority of which also associate with urate levels. Most individuals with hyperuricemia do not develop gout[12]; many who have not experienced a gout flare have MSU crystal deposits[13] that associate with subclinical inflammation, indicating that other factors must be involved in developing symptomatic gout[14]. Genetic variants in inflammatory genes have been associated with gout, but none are widely replicated[14]. Here, we present findings from the largest gout GWAS to date that prioritize previously unidentified candidate pathways for the inflammatory process in gout.

Results

Gout loci, genetic risk prediction and genetic fine-mapping

Study participants were sourced from 13 cohorts (Methods and Supplementary Table 1), with 74% from 23andMe. The GWAS was in four ancestral groups (African, 80,997 participants, 24.6 million SNPs; East Asian, 93,546 participants, 10.2 million SNPs; European, 2,206,883 participants, 13.4 million SNPs; Latinx, 241,956 participants, 19.0 million SNPs) and a trans-ancestry meta-analysis (TAMA), each including sex-specific analyses (Supplementary Table 1 and Fig. 1a). We carried out fixed-effect inverse-variance-weighted GWAS meta-analyses for gout in each of the four ancestral groups. While there was inflation of test statistics (λ_{GC} = 1.050 (African), 1.121 (East Asian), 1.611 (European), 1.105 (Latinx)), linkage disequilibrium (LD) score regression (LDSC) analysis revealed no evidence of inflation due to confounding factors (LD score intercept = 1.025 (African), 1.049 (East Asian), 1.026 (European), 1.048 (Latinx)). TAMA was conducted with 6,386,511 variants that passed quality-control (QC) metrics of minor allele frequency ≥ 0.1% and presence in ≥2 sample sets and/or data from ≥80% of participants (Methods).

Two hundred ninety-five genome-wide significant ($\log_{10}$ (Bayes factor) ≥ 6) loci were detected in the TAMA, 20 with more than one genetically independent signal, resulting in 318 independent genome-wide significant signals. In the single-ancestry GWAS, two genome-wide significant loci were detected in African, ten in East

e-mail: tony.merriman@otago.ac.nz

https://doi.org/10.1038/s41588-024-01921-5

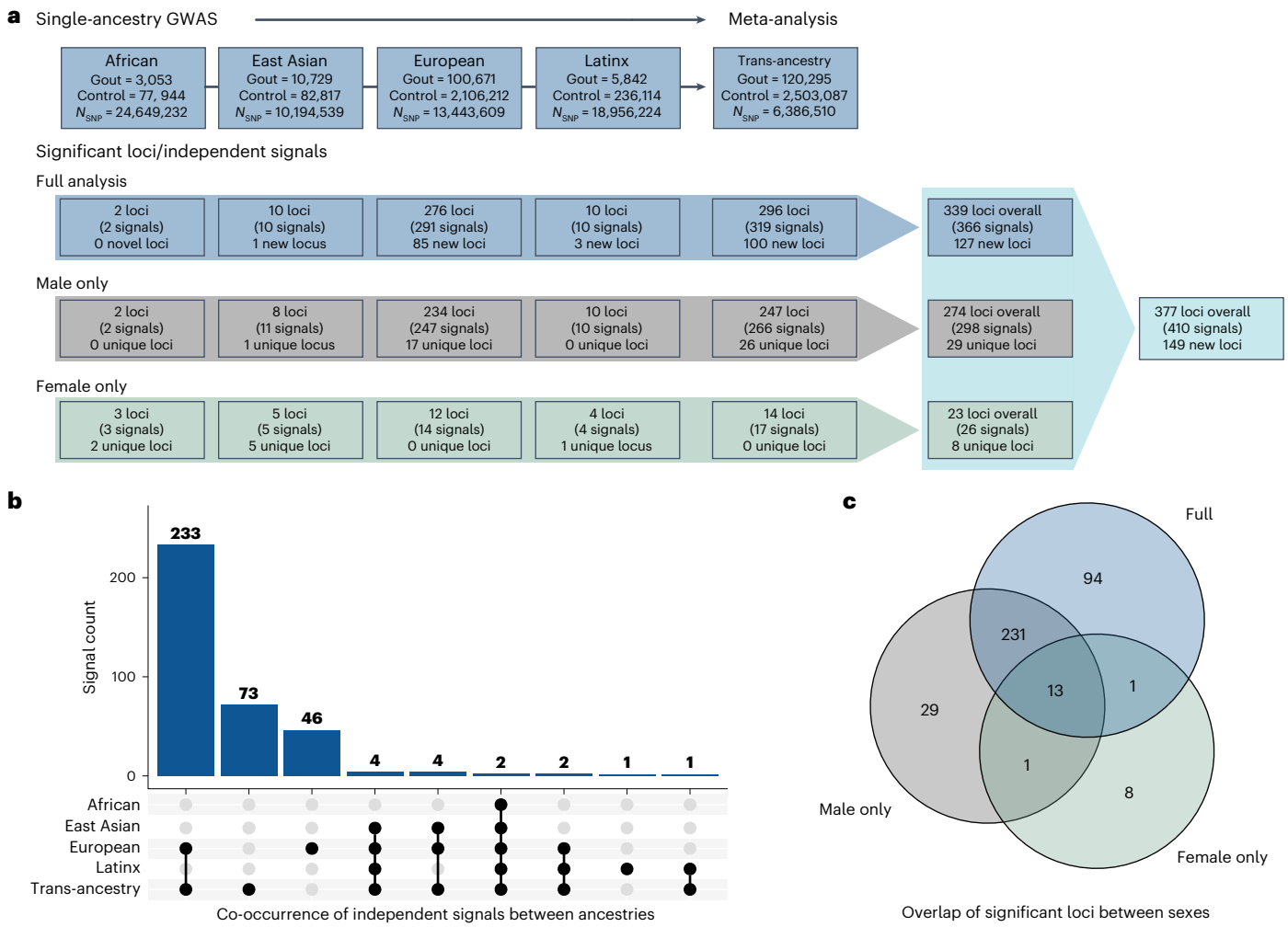


Fig. 1 | Gout associates with 377 genetic loci, representing 410 independent signals, across the ancestry-specific and trans-ancestry analyses. There were 31 male-only and eight female-only independent signals. **a**, Summaries of the cohorts used, the GWAS done and loci detected. Male:female case numbers are 2,009:1,044 for African, 9,779:950 for East Asian, 74,503:22,112 for European (excluding two datasets for which there were no sex data) and 4,272:1,570 for Latinx groups. **b**, Upset plot of the overlap between the 366 significant independent signals in the single-ancestry and trans-ancestry analyses for the full cohort. The black lines and dots beneath the bar graph indicate the ancestry combination of interest, while the bar graph values indicate how many independent signals were identified in the ancestry combination of interest. **c**, Venn diagram of the overlap between the 377 significant loci across the amalgamated full (blue), male-only (gray) and female-only (green) GWAS.

Asian, 276 in European (291 independent lead SNPs) and ten in Latinx datasets (Fig. 1a,b, Extended Data Figs. 1–5, Supplementary Data 1 and Supplementary Tables 2–4). We also performed the TAMA using variants that only passed QC metrics in two or three ancestral groups, detecting an additional locus (chromosome (chr)X:66.15–66.44 Mb). There were 339 non-overlapping loci and 366 genetically independent signals across all analyses. A total of 139 loci had different lead SNPs across the TAMA and the four ancestral-specific GWAS. Of these, the distinct SNPs at 93 loci exhibited strong LD ($r^2 \geq 0.8$; Supplementary Table 3) in the ancestral groups in which they were detected.

The prevalence of gout is 3–4-fold greater in men than in women[1]; therefore, we conducted a sex-stratified GWAS, identifying 247 and 14 loci in the male-only and female-only TAMA, respectively (Fig. 1 and Supplementary Table 3). There were 90,563 men and 25,676 women with gout in these analyses. Across all ancestries, there were 298 and 26 independent signals in the male-only and female-only analyses, respectively; 29 of the male-only loci (36 signals) and eight of the female-only loci were not detected in the full (combined sexes) analyses (Supplementary Table 4). One locus (chr5:134.22–140.17 Mb) was detected in each sex-stratified analysis but not in the full analysis (Fig. 1c) due to heterogeneity between the sex-specific signals (Supplementary Table 5). Of the total 409 genetically independent signals identified across all three analysis sets (male only, female only and combined sexes), 44 (representing 36 loci) had significant heterogeneity ($P < 1 \times 10^{-6}$) between sexes, with effects greater in men for all 36 loci (Supplementary Table 5 and Supplementary Fig. 1). Thirty-two (89%) had previously been associated with urate[7], including ten loci with the strongest effects on urate[15] (*SLC2A9*, *ABCG2*, *SLC17A1*, *SLC16A9*, *GCKR*, *SLC22A11*, *INHBC*, *RREB1*, *PDZK1*, *SLC22A12*), consistent with urate control and subsequent risk of gout having a stronger genetic component in men (Supplementary Note). We identified 497 unique genes mapping to the 44 loci (Supplementary Table 6) that had heterogeneous expression by sex in one or more Genotype–Tissue Expression (GTEx) database tissues[16], and pathway enrichment analysis identified the urate metabolic process as the top GO pathway (Supplementary Table 7).

Of the 377 unique loci identified across all analyses (339 + 29 + 8 + 1), 149 (including four with two independent signals) have not previously had any variant within the locus boundaries associated with urate or gout (Supplementary Table 8). Of the remaining 229 loci, 144 had the same lead SNP or were in high LD ($r^2 \geq 0.8$) with a SNP previously associated with urate or gout, 64 had low to moderate LD ($0.1 \leq r^2 < 0.8$) and 21 had weak LD ($r^2 < 0.1$) (Supplementary Table 8). Of the 30 loci with

more than one independent signal, 17 loci had LD of $r^2 < 0.1$ between one of the lead SNPs and the previously reported SNP(s) at the locus. Therefore, 190 (148 + 4 + 21 + 17) of the independent signals have not previously been reported.

A gout polygenic risk score (PRS) was applied using effect sizes from the European ancestry cohort after removing the UK Biobank cohort (Supplementary Note). UK Biobank participants were binned into ten groups based on their PRS values, and gout prevalence increased across these bins from 0.0% to 22.6% (Fig. 2). Compared to the bin that included the most individuals, age- and sex-adjusted odds ratios ranged from 0.14 (0.02, 1.00) in the lowest bin to 22.1 (10.9, 44.9) in the highest bin. For comparison, male sex conferred a gout risk of 13.4 (12.3, 14.7) (Supplementary Table 9). The PRS improved a model including age and sex gout risk prediction in Europeans (area under the receiver operating characteristic curve (AUROC)) from 0.77 to 0.83, similar to an analysis using 123 lead urate-associated SNPs in the UK Biobank (0.80 to 0.84)[7]. Equivalent PRS were generated using SNPs from the European sex-specific GWAS. The PRS improved AUROC prediction over an age-only model from 0.59 to 0.71 in men and from 0.68 to 0.73 in women.

Next, we conducted fine-mapping to identify candidate causal variants from all 377 gout loci using three approaches (Bayes factor, FINEMAP[17], PAINTOR[18]; Methods) in the European dataset. We considered 99% credible sets from any of the approaches with ≥1 variant with posterior inclusion probability (PIP) ≥ 0.5, identifying a credible set in 285 loci (75.6%) (Supplementary Table 10). Of the 114 credible sets previously identified in urate control[7], we reported PIP ≥ 0.5 for ≥1 of the top three variants in 41 (36%) of our sets (Supplementary Table 10). Fine-mapping of complex phenotype loci is compromised when summary statistics are derived from meta-analysis of cohorts using different genotyping arrays and/or imputation panels and reference panel mismatch[19]: we identified 70 compromised loci (24.6%) using SLALOM[19] (Supplementary Tables 10 and 11).

To gain insights into candidate causal genes, we selected 608 candidate causal SNPs that had high PIP (≥0.5) in at least one fine-mapping approach and were within a 99% credible set with ≤5 SNPs (Supplementary Table 12 and Supplementary Note). We combined these SNPs with the lead SNPs across all GWAS analyses and identified, for the purpose of fine-mapping, conditionally associated European SNPs (Methods) to make a pool of 1,465 candidate causal variants (Supplementary Table 13). From this pool, we identified missense variants within 47 genes, where the missense variants were either the lead variant or in very strong LD ($r^2 \geq 0.98$) with the lead variant (Supplementary Note, Tables 1 and 2 and Supplementary Table 14). Of the 16 genes with information on functional impact of the variant, seven were not previously reported as implicated in gout (*ABCA6*, p.Cys1359Arg; *GLS2*, p.Leu581Pro; *MC4R*, p.Val103Ile; *PNPLA3*, p.Ile148Met; *SH2B3*, p.Trp262Arg; *SLC39A8*, p.Ala391Thr; *SLCO1B1*, p.Val174Ala) (Table 1). We identified nine very strong candidate causal variants with Bayes factor PIP ≥ 0.95, PAINTOR PIP ≥ 0.95 and FINEMAP ≥ 0.95 (Supplementary Table 15). These comprised missense variants in *ABCG2*, *CPS1* and *FGF21* (Table 1) and candidate regulatory variants at six other loci. Of the latter, those at *TMEM174*, *VEGFA* and *HLF* mapped within confirmed protein-binding sites (Supplementary Table 15). Eight have been fine-mapped[7] in urate control, with six of our variants (rs1260326, rs2231142, rs17632159, rs10851885, rs7224610, rs1047891) among the top three credible variants at the respective loci[7].

Relationship of genetic control of gout and urate

Urate is causal for gout, and these two traits share biological pathways. We used LDSC to investigate genetic correlation between the European gout GWAS and 934 traits in the UK Biobank (Methods) with the purpose of identifying correlations not previously seen with urate GWAS data[7]. There was a genetic correlation between gout and 348 phenotypes (strongest with urate; $r_g \geq 0.89$), which spanned 25 of 27 broad phenotype categories (Fig. 3, Supplementary Fig. 2 and Supplementary Table 16). Outside of cardiovascular disease and its body composition-related risk factors, two sex-related blood biomarkers (testosterone ($r_g = -0.17$) and sex hormone-binding globulin ($r_g = -0.31$)) showed negative correlation, consistent with a recent report of greater genetically predicted levels of each hormone associating with reduced gout risk[20]. There was a positive genetic correlation between gout and blood counts of leukocytes, lymphocytes, neutrophils, eosinophils and reticulocytes. Genetic correlations with urate levels in European UK Biobank participants (Methods) were similar, including in the sex-specific analysis (Fig. 3b and Supplementary Tables 16 and 17), with 300 of 348 phenotypes correlated with gout also significantly correlated with urate in a consistent direction. There were no notable genetic correlations in the gout analysis that were not observed in the urate analysis (Supplementary Tables 16 and 17). Mendelian randomization to test for causal relationship(s) between gout and the 348 phenotypes was not done owing to vertical pleiotropy in Europeans when using 289 gout SNPs as a potential instrumental variable (Supplementary Note and Supplementary Table 18).

The genetic correlation between gout and urate was not absolute (Fig. 3; $r_g = 0.89$), suggesting the presence of loci conferring risk to gout but not urate and/or vice versa. The former would indicate loci that influence the risk of gout after hyperuricemia is established. To identify these loci, we first created a large GWAS of serum urate levels by meta-analyzing the independent datasets of Tin et al.[7] and our UK Biobank urate GWAS (Methods) and tested for colocalization (posterior probability (PP) ≥ 0.8) of 291 European gout lead SNPs with the urate GWAS summary statistics (Supplementary Table 19). As expected, a large number ($n = 162$) of gout loci colocalized with urate loci (hypothesis 4 (H_4)). There were nine loci with a signal in gout but not urate (H_1), with little statistical support for association of the gout lead SNP with urate ($P_{urate} > 0.01$; Supplementary Table 20). There were an additional 11 loci with different signals between urate and gout (H_3; Supplementary Table 20). Conditional analysis of 31 gout loci with two genetically independent signals followed by colocalization revealed two further loci associated with gout but not urate (Supplementary Table 21). In total, there were 22 loci with genome-wide significant signals of association in gout but no signal in urate ($P_{urate} > 0.01$; Supplementary Tables 20 and 21 and Supplementary Data 2), directly implicating these loci in gout inflammation. Notably, the *IL1R1* and *IL1RN* loci (encoding the IL-1 receptor 1 and the IL-1 receptor antagonist) were among the H_3 loci with both a signal in gout but not urate and a signal in urate but not gout. *IL6R* had a signal in gout but not urate. These findings directly implicate the various cytokine and cytokine receptor genes at these loci in gout inflammation.

There were 11 H_3 loci with signals of association with gout but not urate that also included separate signals of association with urate but not gout without evidence for a shared underlying variant (Supplementary Table 20). We further evaluated this phenomenon by testing for colocalization of the 422 urate loci with the European gout GWAS summary statistics (Supplementary Table 22). One hundred eighty-three urate lead variants colocalized with gout (H_4; Supplementary Table 22). However, there were also urate association signals (18 for H_2 and 20 for H_3) with little evidence of association with gout ($P_{gout} > 0.01$; Supplementary Table 23). Considering the possibility that some of the urate loci may not have strong association signal and therefore lower evidence of colocalization with gout, we also identified 23 loci that had PP < 0.8 in each of H_2 (urate only) and H_4 (both urate and gout) but together had PP($H_2 + H_4$) ≥ 0.8 and no evidence for association with gout ($P_{gout} > 0.01$; Supplementary Table 23). Conditional analysis of 37 urate loci with two genetically independent signals (Supplementary Note) followed by colocalization analysis identified four further loci associated with urate but not gout (Supplementary Table 24). In total, there were 65 signals associated with urate but not gout (Supplementary Data 3). Intuitively, these findings are biologically implausible

https://doi.org/10.1038/s41588-024-01921-5

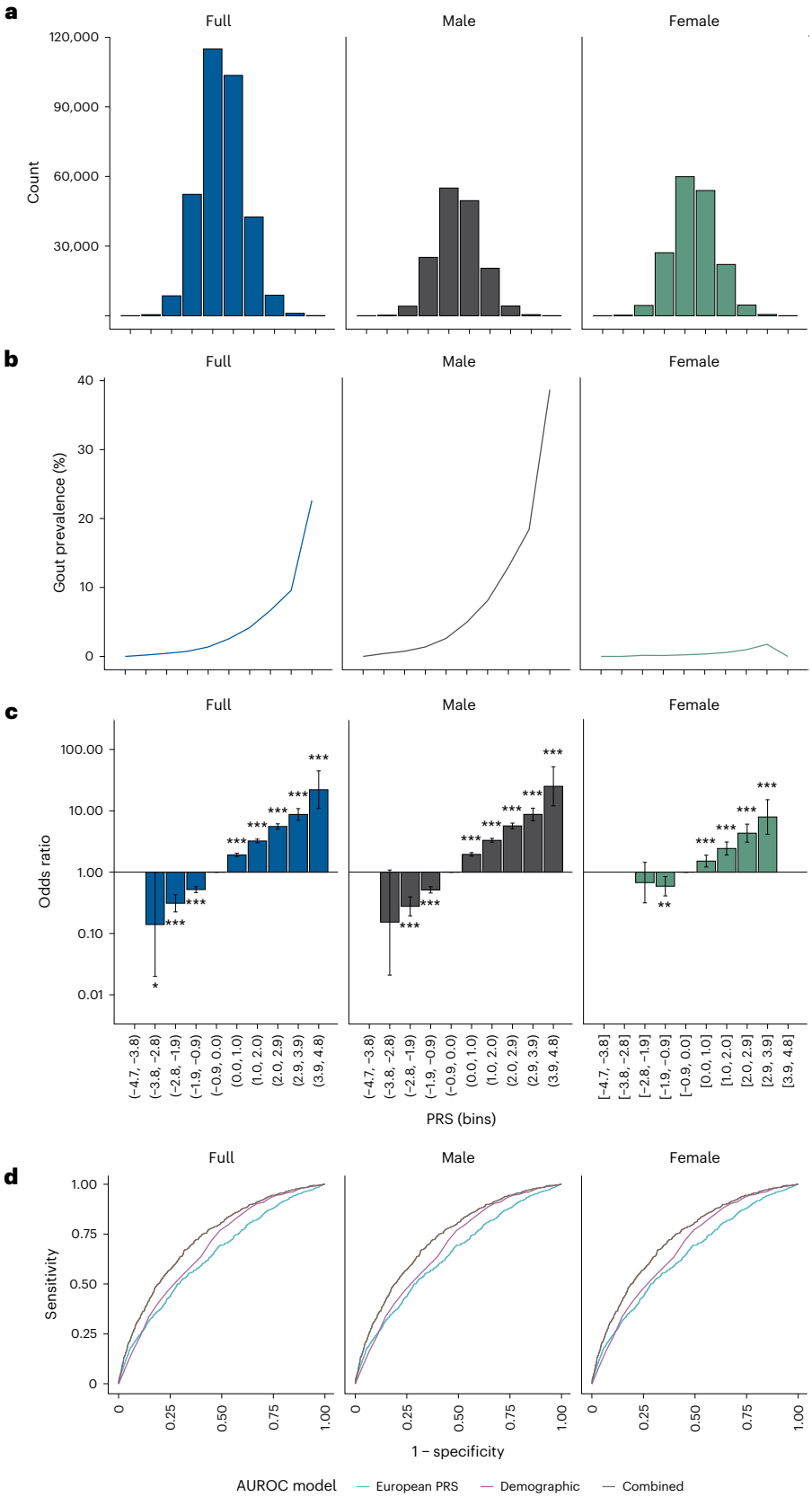


Fig. 2 | Association of PRS with gout in European participants of the UK Biobank in combined sexes (7,131 cases, 325,239 controls; left), men (6,584 cases, 152,777 controls; middle) and women (547 cases, 172,462 controls; right). a, PRS category distribution. **b**, Gout prevalence in different PRS categories. **c**, The odds ratio for gout for each different risk score bin compared to the most common bin (visualized in **a**)). Logistic regression (age- and sex-adjusted in combined sexes, age-adjusted in sex-specific data), two-sided $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Whiskers represent 95% confidence intervals. **d**, Comparison of AUROC graphs in combined sexes, men and women for different prediction models of gout: genetic (PRS only; blue), demographic (age and sex in combined and age-only in sex-specific data; purple) and combined (PRS and demographic; brown).

https://doi.org/10.1038/s41588-024-01921-5

Table 1 | Missense lead variants or variants that are in high LD with lead variants: those with known functional consequences

Gene	Variant		Lead or LD with lead[a]	Risk allele	Frequency[b]	CADD score[c]	PIP[d]	Variant functional consequence
	AA change	rsID						
ABCA6	p.Cys1359Arg	rs77542162	Lead	Cysteine	0.99	28.7	1.0	ATP-binding cassette subfamily member A6. May play a role in macrophage lipid transport and homeostasis. Arginine allele disrupts protein stability[58].
ABCG2	p.Gln141Lys	rs2231142	Lead	Lysine	0.09	16.4		ATP-binding cassette subfamily member G2. p.Gln141Lys causes internalization of ABCG2 (ref. 59), preventing secretion of urate[60]. Lysine allele associates with impaired response to allopurinol[61]. p.Val12Met does not change urate transport activity[60].
	p.Val12Met	rs2231137	Lead (COJO)	Valine	0.94	2.4		
ADH1B	p.His48Arg	rs1229984	Lead	Histidine	0.03	13.1	1.0	Alcohol dehydrogenase 1B. The histidine allele causes rapid oxidation of alcohol[62], predicted to increase depletion of the ATP pool and increase urate levels by ADP catabolism.
ALDH2	p.Glu504Lys	rs671	1.0 (EAS)	Glutamate	0.83 (EAS)	32.0		Aldehyde dehydrogenase 2. Glutamate allele causes increased ALDH2 activity, consistent with increased hypoxanthine and urate levels after ingestion of alcohol[63,64].
CPS1	p.Thr1412Asn	rs1047891	Lead	Threonine	0.70	18.2	1.0	Carbamoyl phosphate synthase. Threonine allele associates with lower levels of blood arginine[43].
GCKR	p.Leu446Pro	rs1260326	Lead	Leucine	0.41	13.2		Glucokinase regulatory protein. Leucine allele causes relaxation of inhibition of glucokinase[65], predicted to increase glucose phosphorylation, deplete the ATP pool and increase urate levels by ADP catabolism.
GLS2	p.Leu581Pro	rs2657879	1.0	Proline	0.21	15.2		Glutaminase 2. Regulated by METTL3. Proline allele associates with lower plasma glutamine levels[44].
HNF4A	p.Thr139Ile	rs1800961	Lead (COJO)	Threonine	0.96	21.4	1.0	Hepatocyte nuclear factor 4A. Threonine variant reduces transactivation of the *ABCG2* promoter, predicted to reduce ABCG2 levels and urate secretion[7].
MC4R	p.Val103Ile	rs2229616	1.0	Valine	0.99	18.5		Melanocortin 4 receptor. Isoleucine variant protective of obesity. Gain-of-function variant[66]. Likely reflects causal effect of obesity on gout.
PNPLA3	p.Ile148Met	rs738409	1.0	Isoleucine	0.78	13.7	1.0	Patatin-like phospholipase domain-containing protein 3 (adiponutrin). Methionine allele is loss of function and reduces hepatic VLDL secretion[67].
SH2B3	p.Trp262Arg	rs3184504	1.0 (EAS/ AFR)	Tryptophan	0.46	10.7		SH2B adaptor protein 3. Negative regulator of cytokine signaling. Deficiency promotes monocyte proliferation upon stimulation[68]. Tryptophan allele is loss of function[69].
SLC17A1	p.Thr269Ile	rs1165196	Lead	Isoleucine	0.56	8.5		Solute carrier family member 17A1, encodes NPT1. Renal secretory urate transporter. Isoleucine allele reduces urate transport activity[70].
SLC17A3	p.Gly279Arg	rs56027330	0.99 (COJO)	Arginine	0.14	17.4		Solute carrier family member 17A3, encodes NPT4. Renal secretory urate transporter. Arginine allele reduces urate transport activity[71].
SLC2A9	p.Gly25Arg	rs2276961	0.98 (COJO)	Arginine	0.53	2.3		Solute carrier family member 2A9, encodes GLUT9 (ref. 72). Reuptake urate transporter[72]. The variant does not influence urate transport activity[73].
SLC39A8	p.Ala391Thr	rs13107325	Lead	Alanine	0.92	22.0	1.0	Encodes a solute carrier gene, a zinc–cadmium–manganese transporter. Increases NF-κβ signaling in macrophages[74]. Threonine allele slightly reduces zinc transport activity[75].
SLCO1B1	p.Val174Ala	rs4149056	Lead	Valine	0.84	25.1	1.0	Solute carrier organic anion transporter family member 1B1. Uptake transporter. Alanine allele reduces transporter activity[76].

Candidate causal variants were drawn from Supplementary Table 14 as either a lead variant (including from the conditional COJO analysis) or in very strong LD ($r^2 \geq 0.98$) with a lead variant (including from the conditional COJO analysis). AA, amino acid; AFR, African; CADD, combined annotation dependent depletion; COJO, conditional and joint; EAS, East Asian; VLDL, very-low-density lipoprotein. [a]In European, unless otherwise stated. [b]Risk allele in European, unless otherwise stated. [c]From Ensembl. [d]Highest PIP of FINEMAP, PAINTOR and BayesFactor (Supplementary Table 10). Variants without a PIP value either had a PIP < 0.50 or were within a locus identified as compromised by SLALOM or were identified by conditional analysis.

because genetic variants that influence urate levels should also influence the risk of gout. There was sufficient power to detect association with gout; using gout effect size estimates derived by regressing gout effect size against urate effect size (Supplementary Note and Supplementary Fig. 3), the power to detect the association of all 65 variants at $P < 0.01$ in our European gout cohort was >99% (Supplementary Fig. 3).

Table 2 | Missense lead variants or variants that are in high LD with lead variants: those without known functional consequences

Gene	Variant		Lead or LD with lead[a]	Risk allele	Frequency[b]	CADD score[c]	PIP[d]	Gene function
	AA change	rsID						
ADO	p.Gly25Trp	rs2236295	Lead	Glycine	0.63	24.1	0.81	2-aminoethanethiol dioxygenase. Forms hypotaurine.
AP4E1	p.Cys88Arg	rs2306331	1.0 (EAS)	Cysteine	0.53	17.3		Adaptor-related protein complex 4 subunit ε1. Role in mediating vesicle formation in secretion and endocytosis.
AQP10	p.His123Tyr	rs6685323	1.0	Histidine	0.68	5.6		Aquaporin 10. Water-selective integral membrane channel.
BSCL2	p.Lys268Arg	rs6856	0.99	Lysine	0.81	5.7		Encodes endoplasmic reticulum transmembrane protein seipin. Important for lipid droplet morphology.
CRIP3	p.Ile188Thr	rs2242416	0.98	Isoleucine	0.42	21.1		Cysteine-rich protein 3.
CUBN	p.Ile2984Val	rs1801239	1.0	Valine	0.09	15.0		Cubilin. Intrinsic factor–vitamin B12 receptor. Range of physiological functions. Associates with megalin (LRP2) in the proximal tubule of the kidney to resorb filtered proteins[77].
	p.Glu3002Gly	rs1801240		Glycine		19.3	0.70	
CTAGE9	p.Leu398Val	rs202051647	1.0 (EAS)	Leucine	0.92	17.8		CTAGE family member 9. Predicted to be involved in vesicle transport from the endoplasmic reticulum to the Golgi.
DTL	p.Ala394Val	rs3135474	1.0	Alanine	0.04	7.6		Denticleless E3 ubiquitin protein ligase homolog. Involved in protein ubiquitination.
DDIT4L	p.Lys180Arg	rs201713115	1.0	Arginine	0.09	22.9		DNA damage-inducible transcript 4 like. Promotes autophagy via mTOR[78].
EPB41	p.Val214Ile	rs111642750	Lead	Isoleucine	0.04	22.5	1.0	Erythrocyte membrane protein band 4.1. Involved in erythrocyte shape.
EVI5	p.Ile336Val	rs2391199	1.0	Isoleucine	0.91	4.3		Ecotropic viral integration site 5. Involved in endosome-to-Golgi transport.
FAM35A	p.Ser550Cys	rs11202365	0.99	Cysteine	0.36	22.8		Encodes shieldin complex subunit 2. Shieldin complex involved in DNA repair.
FGF21	p.Leu174Pro	rs739320	Lead	Leucine	0.37	5.9		Fibroblast growth factor 21. Hepatokine that regulates sugar intake and metabolism.
FRK	p.Gly122Arg	rs3756772	0.99	Arginine	0.42	19.3		Fyn-related Src family tyrosine kinase.
GLIS3	p.Pro456Gln	rs6415788	Lead	Proline	0.37	14.5		GLI-similar zinc finger protein family. Has a role in islet β cell development[79].
INHBC	p.Arg322Gln	rs2229357	0.99	Arginine	0.81	19.6		Inhibin subunit βC. Member of the TGF-β superfamily. Inhibition of activin A signaling.
HNF1A	p.Ala98Val	rs1800574	Lead	Alanine	0.97	22.6	1.0	Hepatocyte nuclear factor 1A. The valine variant is implicated in MODY3.
JMJD1C	p.Glu2535Asp	rs1935	Lead	Aspartate	0.51	11.2	1.0	Jumonji domain-containing protein 1C or KDM3C. A histone demethylase.
KIAA0100	p.Val1373Gly	rs12602520	1.0	Glycine	0.07	15.9		Bridge-like lipid transfer protein family member 2 (BLTP2).
LRP2	p.Lys4094Glu	rs2075252	0.99	Glutamate	0.24	13.1		LDL receptor-related protein 2 (megalin). Associates with cubilin in the kidney proximal tubule to resorb filtered proteins[77].
MLXIPL	p.Ala191Val	rs35332062	0.99	Alanine	0.88	19.4		MLX-interacting protein like. Transcription factor, activates motifs of triglyceride synthesis genes.
	p.Gln306His	rs3812316	Lead	Histidine	0.88	17.1	0.97	
MFSD12	p.Tyr182His	rs2240751	Lead	Tyrosine	0.99	27.4		Major facilitator superfamily domain-containing protein 12. A mediator of cysteine transport[80].
NPHS2	p.Arg229Gln	rs61747728	Lead	Arginine	0.96	22.8	1.0	Encodes podocin, regulates glomerular permeability.
POM121	p.Pro995Ser	rs9955	Lead	Proline	0.87	1.1	1.0	Encodes a transmembrane nucleoporin. Inhibits the macrophage LPS response by blocking NF-κβ translocation to the nucleus[81].
SH2B1	p.Thr171Ala	rs7498665	1.0	Alanine	0.33	12.8		SH2B adaptor protein 1. Involved in kinase activation and signaling.

Table 2 (continued) | Missense lead variants or variants that are in high LD with lead variants: those without known functional consequences

Gene	Variant		Lead or LD with lead[a]	Risk allele	Frequency[b]	CADD score[c]	PIP[d]	Gene function
	AA change	rsID						
SLC25A5	p.Leu111Arg	rs371749	Lead	Arginine	0.75	22.3	0.76	Solute carrier family member 25A5. Encodes ADP–ATP translocase 2 (ANT2), which mediates transport of ADP into the mitochondrial matrix. Drives proinflammatory macrophage activity[82].
SLC5A1	p.Asn51Ser	rs17683011	0.98	Serine	0.06	20.2		Solute carrier family member 5A1. Encodes sodium–glucose cotransporter 1 (SGLT1). A haplotype of missense variants.
	p.Ala411Thr	rs17683430		Alanine		15.7		
	p.His615Gln	rs33954001		Glutamine		5.4		
SLC5A9	p.Ala600Val	rs78427303	1.0 (COJO)	Alanine	0.91	0.0		Encodes sodium–glucose transporter SGLT4.
SOS2	pPro191Arg	rs72681869	Lead	Arginine	0.99	22.5	0.89	SOS Ras/Rho guanine nucleotide exchange factor 2.
TSPAN6	p.Ala108Thr	rs1802288	Lead	Alanine	0.81	26.0	1.0	Tetraspanin 6. Implicated in production of extracellular vesicles[83].
UPF3A	p.Arg64Lys	rs3752105	0.99	Lysine	0.26	15.2		UPF3A, regulator of nonsense-mediated mRNA decay.

Candidate causal variants were drawn from Supplementary Table 14 as either a lead variant (including from the conditional COJO analysis) or in very strong LD ($r^2 \geq 0.98$) with a lead variant (including from the conditional COJO analysis). FAM35A, familial sequence similarity 35A; KIAA0100, Kazusa DNA Research Institute AA; LDL, low-density lipoprotein; LPS, lipopolysaccharide; MODY3, maturity-onset diabetes of the young type 3; mTOR, mammalian target of rapamycin; TGF, transforming growth factor. [a]In European, unless otherwise stated. [b]Risk allele in European, unless otherwise stated. [c]From Ensembl. [d]Highest PIP of FINEMAP, PAINTOR or BayesFactor (Supplementary Table 10). Variants without a PIP value either had a PIP < 0.50 or were within a locus identified as compromised by SLALOM or were identified by conditional analysis.

Candidate genes, tissues and pathways

We tested for enriched tissues and specific cell types in single-ancestry data using covariate-adjusted LDSC (Methods)[21]. There was significant enrichment (Bonferroni-corrected $P_{coefficient} \leq 0.05/40$) in the kidney (10.8-fold enrichment, $P_{coefficient} = 2.3 \times 10^{-4}$) in the East Asian dataset and in the kidney (11.6-fold enrichment, $P_{coefficient} = 8.6 \times 10^{-7}$) and liver (5.1-fold enrichment, $P_{coefficient} = 8.3 \times 10^{-6}$) in the European dataset (Supplementary Table 25 and Extended Data Fig. 6). We observed cell type enrichments (false discovery rate (FDR) ≤ 0.05) in the small and large intestine, the colonic mucosa, the duodenum mucosa, the rectal mucosa, the pancreas, skeletal muscle, kidneys and liver in the European dataset (Supplementary Table 26 and Extended Data Fig. 7). There were three kidney cell histone marks enriched in the East Asian dataset that were also found in the European dataset (histone H3 lysine 4 trimethylation, histone H3 lysine 9 acetylation, histone H3 lysine 27 acetylation; Supplementary Table 26). Given the absence of urogenital system tissues in the LDSC dataset, we applied the European combined-sex data to DEPICT[22], with significant enrichment in the urinary tract, female genitalia, Fallopian tubes, the myometrium, ovaries and adnexa uteri (Supplementary Table 27).

To identify candidate causal genes possibly controlled through regulation of gene expression, we focused on gout association signals that colocalized with *cis* and *trans* expression quantitative trait loci (eQTL) across 49 tissues (Methods). Of the 340 trans-ancestry and European lead SNPs (or a proxy variant; $r^2 \geq 0.8$) present in GTEx, 273 (80.3%) had ≥1 colocalized eQTL (PP of colocalization ≥0.8), representing 704 genes. There were 1,715 *cis*-eQTL and 292 *trans*-eQTL colocalized over multiple tissues for 424 and 286 genes, respectively (Supplementary Table 28 and Supplementary Fig. 4), and six genes had both *cis*-eQTL and *trans*-eQTL (*JAZF1*, *SEPTIN9*, *RNF157*, *SLC16A9*, *TANC1*, *TUBB8P7*). Only four colocalized kidney cortex eQTL were observed in GTEx; therefore, we leveraged other kidney eQTL data[23] to identify autosomal *cis*-eQTL for 29 genes (17 in the tubule, 21 in the glomerulus; nine overlapping genes) that colocalize with gout association signals (Supplementary Table 29). The genes included *SLC16A9*, which encodes monocarboxylate transporter 9 (MCT9), previously implicated in gout driven by a combination of reduced gut excretion and overproduction of urate (renal overload[24]). However, there were no established urate transporters in the kidney eQTL colocalization.

Reasoning that colocalized eQTL restricted to a single tissue may be more tractable to specific biological insights, we identified 496 tissue-specific colocalized eQTL genes (Supplementary Note and Supplementary Table 30). Of specific interest are the eQTL in the testis for *HNF4G* and *PRPS2*, genes with roles in urate homeostasis[25,26], suggesting male-specific mechanisms of urate control, and there were 17 whole-blood-specific and 42 kidney-specific eQTL (Supplementary Note and Supplementary Table 30).

Epigenetic reprogramming ('trained immunity') of myeloid cells by soluble urate is implicated in gout[6,27,28]. To evaluate the possibility that gout-associated loci control DNA methylation status in blood, we conducted colocalization analysis with whole-blood DNA methylation QTL (meQTL) in the Genetics of DNA Methylation Consortium (GoDMC) dataset[29], using 323 independent signals encompassing 444 unique lead SNPs in the full and sex-specific European GWAS. We identified 520 colocalized CpG sites within 1 Mb of a lead SNP (PP ≥ 0.8; Supplementary Table 31) and determined that the CpG sites were enriched with binding sites of 27 transcription factors confirmed by chromatin immunoprecipitation followed by sequencing (ChIP–seq) (Supplementary Note and Supplementary Table 32). This enrichment included IRF1, which regulates the interferon-stimulated transcriptional response in monocytes, is within a gout-associated locus and was highly prioritized as a candidate gene in gouty inflammation (see next section).

Enrichment analyses were carried out using 1,764 unique genes implicated by candidate causal missense variants ($n = 47$; Tables 1 and 2), colocalized GTEx eQTL ($n = 704$; Supplementary Table 28), kidney eQTL ($n = 29$; Supplementary Table 29) and MAGMA genes ($n = 1,255$; Supplementary Table 33) (Extended Data Fig. 8 and Supplementary Table 34). Enrichment was observed for 104 pathways including the urate metabolic process (Gene Ontology (GO)), chromatin organization and chromatin-modifying enzymes (Reactome) and type II diabetes mellitus (Kyoto Encyclopedia of Genes and Genomes (KEGG); Extended Data Fig. 8 and Fig. 4). The 49 genes that contributed to chromatin-related Reactome terms included *KMT2A*, *SETD1A* and *SETD2* (encoding histone methyltransferases), *KDM3A*, *KDM4C* and *KDM6B* (encoding lysine demethylases) and histone genes. Twenty histone genes at the chr6:25.07–32.18 Mb locus that included conditionally associated SNPs (Supplementary Tables 3 and 13), were essential

https://doi.org/10.1038/s41588-024-01921-5

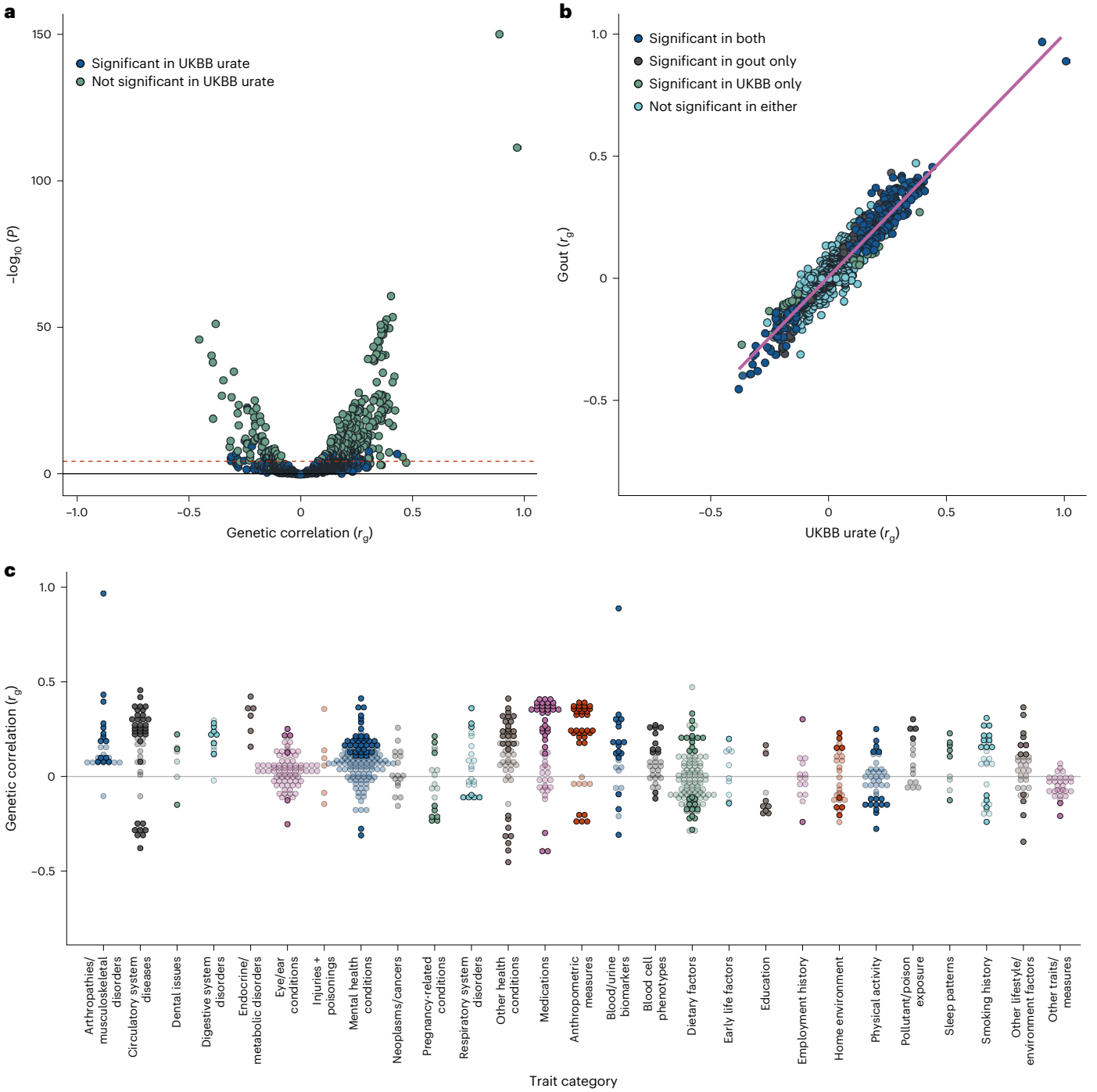


Fig. 3 | Genetic correlation between the European gout GWAS and 934 UK Biobank GWAS traits. a, Volcano plot of genetic correlation. Points are colored based on whether the equivalent correlation between the UK Biobank (UKBB) serum urate GWAS and the trait of interest was significant ($P < 5.35 \times 10^{-5}$) as output by the LD Score regression software[86]. Exact *P* values are in Supplementary Table 16. **b**, Linear relationship between the genetic correlation results of the European gout GWAS and the UK Biobank serum urate GWAS. Points are colored based on whether they were significant in both, one or neither of the genetic correlation analyses. **c**, Genetic correlation values (r_g) across 27 trait categories. In all plots, each point represents the result of a genetic correlation analysis between our European gout GWAS and one of the 934 UK Biobank GWAS traits. Nonsignificant results are shown as transparent plot points.

to 90 of the 104 pathways (Supplementary Table 34), including the chromatin-related Reactome terms.

Gout-associated inflammatory mechanisms

To prioritize genes that may be involved in the inflammatory response in gout, we developed a set of criteria that we applied to 5,358 unique genes within the boundaries of all identified GWAS loci and *cis*-eQTL and *trans*-eQTL genes outside the boundaries (Supplementary Table 35 and Supplementary Note). One hundred eight genes at 58 loci were prioritized (Fig. 5), including *FADS1*, *DGAT2* and the top-ranked *FADS2*, all involved in lipid metabolism, and *IRF5*, *IL1R1*, *TRAF4*, *IL6R* and *MAST3*, all implicated in the inflammatory response. Of the 108 genes, four have drugs approved by the Food and Drug Administration: IL-1R1, anakinra; IL-6R, for example, tocilizumab; the sodium/potassium transporting ATPase subunit ATP1B3, for example, deslanoside; and the ribosomal protein RPL23A, ataluren. We also took a

https://doi.org/10.1038/s41588-024-01921-5

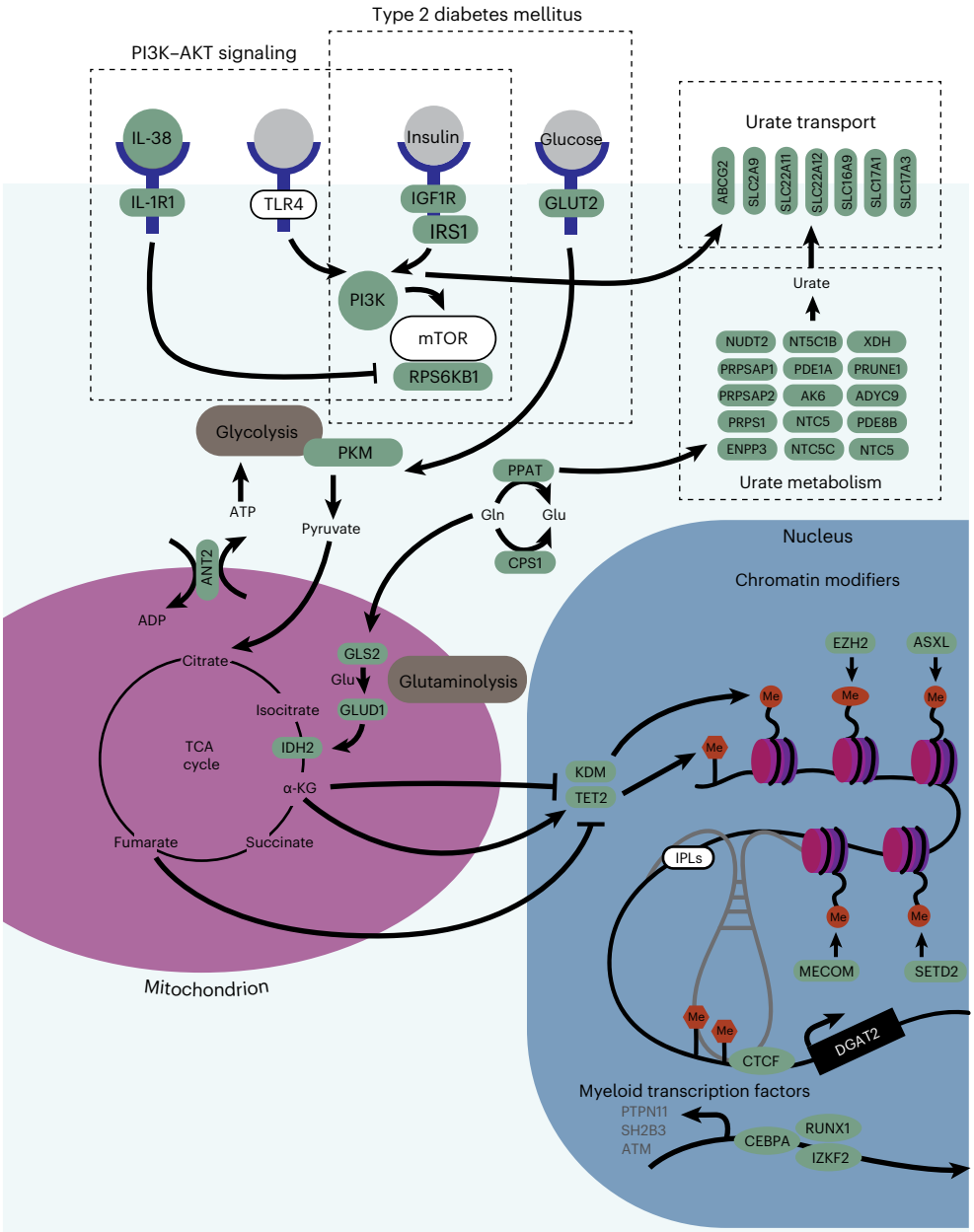


Fig. 4 | Linking of insulin signaling, the tricarboxylic acid cycle and epigenomic reprogramming cellular pathways with candidate gout-associated genes. ASXL1, ASXL transcriptional regulator 1; CEBP, CCAAT enhancer-binding protein; CTCF, CCCTC-binding factor; EZH2, enhancer of zeste 2 polycomb repressive complex 2 subunit; GLS, glutaminase; GLUD1, glutamate dehydrogenase 1; GLUT2, glucose transporter 2; IDH2, isocitrate dehydrogenase 2; IGF1R, insulin-like growth factor 1 receptor; IRS1, insulin receptor substrate 1; IZKF2, IKAROS family zinc finger 2; KDM, lysine demethylase; MECOM, MDS1 and EVI1 complex locus; PI3K, phosphoinositide 3-kinase; PKM, pyruvate kinase; PPAT, phosphoribosyl pyrophosphate amidotransferase; RPS6KB1, ribosomal protein S6 kinase B1; RUNX1, runt-related transcription factor 1;TCA, tricarboxylic acid; TLR4, Toll-like receptor 4; α-KG, α-ketoglutarate.

systematic approach to identifying drug targets across all identified GWAS loci using the Genome for Repositioning pipeline (Supplementary Fig. 5)[30], with nominally significant categories being neoplasms, blood biochemistry and metabolic disorders, driven by *CASP9*, *CCND1*, *CHEK2*, *DRD5*, *IGF1R*, *INSR*, *PPARG* and *DGAT2* (Supplementary Note and Supplementary Table 36).

Clonal hematopoiesis of indeterminate potential (CHIP) results from white blood cell somatic mutation in a suite of 71 genes[31], predominantly in the epigenomic modifiers encoded by *TET2* and *DNMT3A*. CHIP has been associated with incident[32] and prevalent gout[31], and our pathway analysis also revealed enrichment in chromatin organization and chromatin-modifying enzymes. Therefore, we investigated whether germline genetic variation in loci harboring CHIP genes could play a role in gout. CHIP genes were over-represented in the 2,650 protein-coding genes of the 5,358 genes for prioritization (Supplementary Table 35), where 19 of 71 (Supplementary Table 37) were present in the 2,650 genes (of 21,509 protein-coding genes genome wide) (χ^2 $P = 4.76 \times 10^{-4}$). We performed two-sample Mendelian randomization to test for a causal relationship of CHIP for gout using GWAS data of association of CHIP with germline variation[32] (Supplementary Note and Supplementary Table 38). The weighted median method and Mendelian randomization robust adjusted profile score (MR-RAPS) methods showed some evidence of positive causality of all CHIP and *DNMT3A* CHIP for gout ($0.001 < P < 0.025$) but not for *TET2* CHIP ($P > 0.05$).

Immune-priming long noncoding (lnc)RNA species (IPLs) prime immune genes for robust transcription by facilitating transcriptional

https://doi.org/10.1038/s41588-024-01921-5

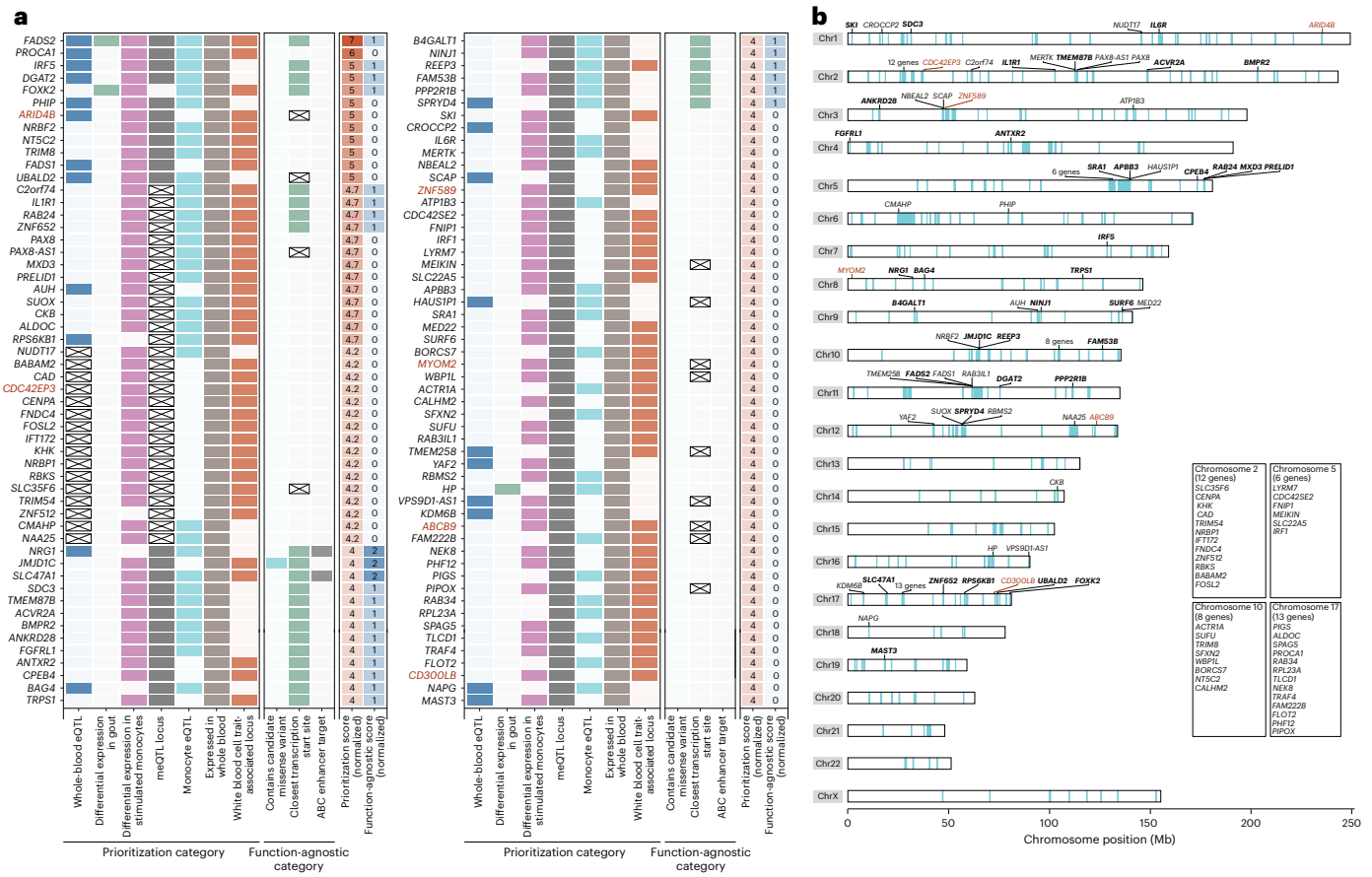


Fig. 5 | Genes prioritized for a role in gouty inflammation. a, One hundred eight genes with a normalized prioritization score ≥ 4 are ranked from highest to lowest score. The seven prioritization categories (left), three function-agnostic categories (middle) and the normalized scores (right) are given for each gene. Cells are colored if the gene gained a point in the prioritization and/or function-agnostic scores based on the criteria of that category and are crossed if category data were unavailable for that gene. Red gene labels represent those that were identified as a trans-eQTL. **b**, Ideogram showing the genomic location of the 108 genes with prioritization score ≥ 4. Bolded gene labels represent those that had a function-agnostic score ≥ 1, and red gene labels represent those that were identified as a trans-eQTL. Light blue highlighting within the ideogram chromosomes indicates the genomic location of all significant loci identified, amalgamated across all ancestry-specific, trans-ancestry and sex-specific analyses.

machinery assembly (for example, the prototypical UMLILO lncRNA[33]). Indeed, two gout-associated loci containing *CSF1* and *IRF1* were previously identified as being connected by an IPL[33]. To identify candidate IPLs, we identified 61 *cis*-eQTL for 76 lncRNA species that colocalized with a gout GWAS signal, representing 18% (76 of 424) of all colocalized *cis*-eQTL genes. Of the 170 *cis*-eQTL variants that controlled the expression of ≥1 protein-coding gene, 48 (28%) also had a cis-eQTL for a lncRNA. We then selected loci based on the protein-coding gene having an eQTL in whole blood and a high prioritization score (≥4; Fig. 5), identifying five candidate IPL loci (Table 3). Two IPLs had an eQTL in whole blood (*DGAT2*, RP11-535A19.2 and *BAG4*, RP11-350N15.5). Inspection of the *DGAT2* locus suggests that this region is under immune cell-specific regulatory control mediated by DNA methylation and expression of RP11-535A19.2 (Fig. 4, Supplementary Note and Extended Data Fig. 9).

An enhancer physically interacts with a promoter through activity by contact (ABC)[34] (Methods). Of the 273 lead variants with colocalized *cis*- and/or *trans*-eQTL, 13 were within an ABC enhancer for 81 genes (ABC score ≥ 0.015; Supplementary Table 39) in stimulated innate immune cells, with loci containing *CSF1* (rs2938616) and *CSF1R* (rs2282804) notable. Connections at the *CSF1* locus were between a 544-bp segment of DNA and the promoters of 23 genes including the neighboring *CSF1* gene and at the *CSF1R* locus between a 915-bp segment of DNA within *CSF1R* and 37 genes including *CSF1R* (for comparison, the other 11 loci had ABC enhancer connections with ≤6 genes, and ABC enhancers are predicted to regulate 2.8 genes on average[34]). The connections at both loci were exclusively restricted to stimulated cells and not observed in unstimulated cells or other immune cells (Supplementary Table 39). The lead SNP at each locus (rs2938616 at *CSF1* and rs2282804 at *CSF1R*) disrupts five and eight transcription factor-binding motifs, respectively, including nuclear factor (NF)-κβ at *CSF1R* and the repressor element 1 silencing transcription factor (REST) at both loci (Haploview).

Discussion

Our study substantially increased the number of known gout loci to 377 (of which 149 had not previously been reported). Sex-specific GWAS led to the discovery of 37 loci not detected in the main analysis. Overall, our results emphasize the importance of molecular control of urate levels in the pathogenesis of gout. We focused on identifying molecular mechanisms of inflammation in gout, for which we generated a comprehensive list of candidate genes.

We prioritized genes predicted to control the innate immune response to MSU crystals, focusing, by necessity, on the gene per se, and did not incorporate aspects of previously proposed functionally agnostic variant-to-gene criteria[35,36]. The adjacent genes *FADS1* and *FADS2* were top ranked and encode fatty acid desaturases involved in the synthesis of unsaturated omega-3 fatty acids that inhibit NLRP3 inflammasome activation[37]. Increased *FADS2* expression correlates with reduced circulating arachidonic acid and increased circulating IL-1β, tumor necrosis factor (TNF) and IL-6 levels[38]. Furthermore,

Table 3 | Candidate immune-priming lncRNA species

Gene	lncRNA	Lead variant		Normalized prioritization score		Gene function
		rsID	Chr:position	Gene	lncRNA	
AUH	RP11-305L7.7	rs2387099	9:93,985,824	4.7	1.2	3-methlyglutaconyl-CoA hydratase. In the leucine degradation pathway. Located in mitochondrial membrane. Binds to AU-rich elements in 3′-UTRs of rapidly decaying RNA.
BAG4	RP11-350N15.4 RP11-350N15.5	rs4537271	8:38,040,503	4	2	BCL2-associated athanogene 4. Negative regulator of apoptosis. *BAG4* and RP11-350N15.5 have whole-blood eQTL.
DGAT2	RP11-535A19.2	rs10899113	11:75,468,464	5	2	Diacylglycerol acyl transferase 2. Connects diacylglycerol to long-chain fatty acyl-CoA in triglyceride synthesis[84]. *DGAT2* and RP11-535A19.2 both have whole-blood eQTL.
NAPG	RP11-883A18.3	rs489837	18: 10,485,357	4	1	*N*-ethylmaleimide-sensitive factor attachment protein γ or γ-SNAP. Regulates endosomal trafficking of transferrin[85].
YAF2	RP11-328C8.2 RP11-351C21.2	rs11181466	12:42,765,181	4	1 2	YY1-associated factor 2. Is a component of polycomb repressive complex 1 (PRC1).

Candidate immune-priming lncRNA species were extracted from Supplementary Table 28 based on the lead variant having a cis-eQTL for both a lncRNA and a protein-coding gene with a high prioritization score in Supplementary Table 35. SNP positions are noted in human genome build GRCh37 coordinates. CoA, coenzyme A; UTR, untranslated region.

FADS2-knockdown reduces innate immune response to *Candida albicans*, and FADS2 inhibition reduces proliferation of innate immune cells[39]. Consistent with a role in gout, the gout risk allele at *FADS2* associates with increased expression of *FADS2* in blood. *IL1R1* and *IL6R* are also notable (Supplementary Note).

We identified 47 genes with candidate causal missense variants. *AQP10* is a colocalized eQTL in whole blood, and the lead variant is in strong LD ($r^2 \geq 0.98$) with two missense variants (rs6668968 (p.Arg15Gln) and rs6685323 (p.His123Tyr)). AQP10 transports water and glycerol, playing a role in lipolysis in adipocytes[40]. Regulation of cell volume induces ionic changes that affect NLRP3 inflammasome activation[41], and inhibition of aquaporins reduces MSU crystal-induced production of IL-1β by THP-1 monocytes[42], suggesting that cellular osmolarity could be a causal mechanism in gout. Gout risk alleles of two enzymes, encoded by *CPS1* and *GLS2*, are associated with lower arginine[43] and glutamine levels[44], respectively. These missense variants may contribute to gout risk via immunometabolic epigenetic reprogramming through histone modifications[45].

We did not detect genome-wide evidence for association at genes encoding well-characterized components of the NLRP3 inflammasome or accessory molecules and innate immune molecules such as Toll-like receptors (Supplementary Note). Rather, genes and pathways discovered in this study point toward cytokines and their receptors, immune priming and regulation of the NLRP3 inflammasome response to MSU crystals. Highly prioritized genes include *TMEM176B* (encoding a transmembrane calcium channel that inhibits NLRP3 inflammasome activation)[46], *SCAP* (the SCAP–SREBP2 complex activates the inflammasome)[47], *INSIG2* (encodes a binding partner of SCAP)[48] and *NINJ1* (its product regulates the NLPR3 inflammasome during the MSU crystal response)[49]. We also implicated genes encoding epigenomic regulators, which may mediate training of the innate immune system to be more responsive to MSU crystals. The findings from the ABC enhancer–gene connection analysis strongly implicate stimulus-responsive chromatin conformational changes in innate immune cells at each of the *CSF1* and *CSF1R* loci, influencing the expression of many proteins, including colony-stimulating factor 1 (CSF1) and its receptor. CSF1 controls monocyte and macrophage differentiation and proliferation[50]. Therefore, CSF1 and its receptor are strong candidates for a role in gouty inflammation. The NRSF or REST transcription factor, implicated at both the *CSF1* and *CSF1R* loci, silences target genes via epigenomic remodeling[51]. Notably, the co-repressor RCOR1, which binds to the C-terminal domain of REST, is genetically associated with gout in a Chinese cohort, and suppression of *REST* expression reduces production of pro-IL-1β in THP-1 cells after stimulation with MSU crystal[52]. In the European GWAS, the *RCOR1* locus associated with urate but not gout (Supplementary Fig. 5), and the lead urate-associated SNP (rs146091133) also associates with expression of *RCOR1* in whole blood (GTEx). Additionally, *PPARGC1B*, which encodes PGC-1β, which coactivates peroxisome proliferator-activated receptor γ (PPAR-γ) in maintaining mitochondrial biogenesis, is also contacted by the ABC enhancer at the *CSFR1* locus. Expression of *PPARGC1B* is induced by MSU crystals[53], and it is possible that this is mediated via the stimulus-responsive enhancer at *CSF1R*.

We identified a phenomenon by which loci that associated with urate did not associate with gout. This is counter to the causal role of increased urate levels in gout. We hypothesize that the lead variants at these loci have pleiotropic effects, in which the urate-increasing allele also affects the expression of gene(s) such that the risk of gout is reduced. The directionality of effect at the *IL33* and *IL1R1* loci supports this hypothesis. Lead urate SNPs rs10122305 and rs11465727 associate with the expression of *IL33* and *IL1RL1*, respectively, in GTEx[54]. IL-33 ameliorates MSU crystal-linked inflammation[55] and signals through IL-1RL1. The urate-increasing alleles at each locus associate with increased *IL33* and *IL1RL1* expression, which can be predicted to protect from gout, resulting in no observed gout effect. However, our hypothesis is difficult to test genetically in the absence of a genetic effect on gout and will be more readily evaluated by mechanistic dissection of the loci. If our hypothesis were correct, it would lead to the identification of further genes involved in gouty inflammation.

A major limitation of our study is the dominance of participants of European ancestry (84.1%), with only 9.2%, 3.6% and 3.1% for Latinx, East Asian and African ancestries, respectively, contributed to by inequity in participation of non-European populations in genetic research (Supplementary Note). While our study design did not include replication, the high correlation between gout and urate effects reduces concern about false positives. Finally, some methods of gout diagnosis (self-report and medication data) are not as direct as a clinical diagnosis. However, we have previously reported that these survey definitions of gout have 82% sensitivity and 72% specificity[56], similar to a definition from the Study for Updated Gout Classification Criteria, which used five weighted items that included a clinical measure (hyperuricemia)[57].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01921-5.

References

1. Kuo, C. F. et al. Global epidemiology of gout: prevalence, incidence and risk factors. *Nat. Rev. Rheumatol.* **11**, 649–662 (2015).
2. Safiri, S. et al. Prevalence, incidence, and years lived with disability due to gout and its attributable risk factors for 195 countries and territories 1990–2017: a systematic analysis of the Global Burden of Disease Study 2017. *Arthritis Rheumatol.* **72**, 1916–1927 (2020).
3. Zhu, Y. et al. Comorbidities of gout and hyperuricemia in the US general population: NHANES 2007–2008. *Am. J. Med.* **125**, 679–687 (2012).
4. Dalbeth, N. et al. Gout. *Lancet* **388**, 2039–2052 (2016).
5. Martinon, F. et al. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
6. Cabău, G. et al. Urate-induced immune programming: consequences for gouty arthritis and hyperuricemia. *Immunol. Rev.* **294**, 92–105 (2020).
7. Tin, A. et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* **51**, 1459–1474 (2019).
8. Nakayama, A. et al. Subtype-specific gout susceptibility loci and enrichment of selection pressure on *ABCG2* and *ALDH2* identified by subtype genome-wide meta-analyses of clinically defined gout patients. *Ann. Rheum. Dis.* **79**, 657–665 (2020).
9. Li, C. et al. Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nat. Commun.* **6**, 7041 (2015).
10. Sulem, P. et al. Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat. Genet.* **43**, 1127–1130 (2011).
11. Zhou, W. et al. Global Biobank Meta-analysis Initiative: powering genetic discovery across human disease. *Cell Genom.* **2**, 100192 (2022).
12. Dalbeth, N. et al. Relationship between serum urate concentration and clinically evident incident gout: an individual participant data analysis. *Ann. Rheum. Dis.* **77**, 1048–1052 (2018).
13. Dalbeth, N. et al. Urate crystal deposition in asymptomatic hyperuricaemia and symptomatic gout: a dual energy CT study. *Ann. Rheum. Dis.* **74**, 908–911 (2015).
14. Major, T. J. et al. An update on the genetics of hyperuricaemia and gout. *Nat. Rev. Rheumatol.* **14**, 341–353 (2018).
15. Yang, Q. O. 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).
16. Oliva, M. et al. The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
17. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
18. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
19. Kanai, M. et al. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genom.* **12**, 100210 (2022).
20. Yuan, S. et al. Genetically predicted sex hormone levels and health outcomes: phenome-wide Mendelian randomization investigation. *Int. J. Epidemiol.* **51**, 1931–1942 (2022).
21. Luo, Y. et al. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum. Mol. Genet.* **30**, 1521–1534 (2021).
22. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
23. Qiu, C. et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat. Med.* **24**, 1721–1731 (2018).
24. Nakayama, A. et al. Common missense variant of monocarboxylate transporter 9 (*MCT9*/*SLC16A9*) gene is associated with renal overload gout, but not with all gout susceptibility. *Hum. Cell* **26**, 133–136 (2013).
25. Ahmed, M. et al. Accelerated transcription of PRPS1 in X-linked overactivity of normal human phosphoribosylpyrophosphate synthetase.*J. Biol. Chem.* **274**, 7284–7488 (1999).
26. Halperin Kuhns, V. L. et al. Differential expression of renal urate transporters in male and female mice. *FASEB J.* **34**, S1 (2020).
27. Badii, M. et al. Urate-induced epigenetic modifications in myeloid cells. *Arthritis Res. Ther.* **23**, 202 (2021).
28. Wang, Z. et al. Differential DNA methylation of networked signaling, transcriptional, innate and adaptive immunity, and osteoclastogenesis genes and pathways in gout. *Arthritis Rheumatol.* **72**, 802–814 (2020).
29. Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**, 1311–1321 (2021).
30. Sakaue, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics* **35**, 3821–3823 (2019).
31. Agrawal, A. et al. *TET2*-mutant clonal hematopoiesis and risk of gout. *Blood* **140**, 1094–1103 (2022).
32. Kessler, M. D. et al. Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature* **612**, 301–309 (2022).
33. Fanucchi, S. et al. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.* **51**, 138–150 (2019).
34. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
35. Gazal, S. et al. Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.* **54**, 827–836 (2022).
36. Mountjoy, E. et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
37. Legrand-Poels, S. et al. Free fatty acids as modulators of the NLRP3 inflammasome in obesity/type 2 diabetes. *Biochem. Pharmacol.* **92**, 131–141 (2014).
38. Chu, X. et al. Integration of metabolomics, genomics, and immune phenotypes reveals the causal roles of metabolites in disease. *Genome Biol.* **22**, 198 (2021).
39. Sibbons, C. M. et al. Polyunsaturated fatty acid biosynthesis involving Δ8 desaturation and differential DNA methylation of *FADS2* regulates proliferation of human peripheral blood mononuclear cells. *Front. Immunol.* **9**, 432 (2018).
40. Gotfryd, K. et al. Human adipose glycerol flux is regulated by a pH gate in AQP10. *Nat. Commun.* **9**, 4749 (2018).
41. Compan, V. et al. Cell volume regulation modulates NLRP3 inflammasome activation. *Immunity* **37**, 487–500 (2012).
42. Chirayath, T. et al. The inflammation induced by monosodium urate and calcium pyrophosphate crystals depends on osmolarity and aquaporin channels. *Arthritis Rheumatol.* **74**, S9 (2022).
43. Pearson, D. L. et al. Neonatal pulmonary hypertension: urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. *N. Engl. J. Med.* **344**, 1832–1838 (2001).
44. Kettunen, J. et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
45. Riksen, N. P. & Netea, M. G. Immunometabolic control of trained immunity. *Mol. Asp. Med.* **77**, 100897 (2021).
46. Segovia, M. et al. Targeting TMEM176B enhances antitumor immunity and augments the efficacy of immune checkpoint blockers by unleashing inflammasome activation. *Cancer Cell* **35**, 767–781 (2019).

47. Guo, C. et al. Cholesterol homeostatic regulator SCAP–SREBP2 integrates NLRP3 inflammasome activation and cholesterol biosynthetic signaling in macrophages. *Immunity* **49**, 842–856 (2018).
48. Yan, R. et al. A structure of human Scap bound to Insig-2 suggests how their interaction is regulated by sterols. *Science* **371**, eabb2224 (2021).
49. Zhang, H. et al. Role of NINJ1 in gout flare and potential as a drug target. *J. Inflamm. Res.* **15**, 5611–5620 (2022).
50. Sehgal, A., Irvine, K. M. & Hume, D. A. Functions of macrophage colony-stimulating factor (CSF1) in development, homeostasis and tissue repair. *Semin. Immunol.* **54**, 101509 (2021).
51. Hwang, J.-Y. & Zukin, R. S. REST, a master transcriptional factor in neurodegenerative disease. *Curr. Opin. Neurobiol.* **48**, 193–200 (2018).
52. Ji, A. et al. Novel genetic loci in early-onset gout derived from whole-genome sequencing of an adolescent gout cohort. *Arthritis Rheumatol.* https://doi.org/10.1002/art.42969 (2024).
53. Chang, W.-C. et al. Genetic variants of PPAR-γ coactivator 1B augment NLRP3-mediated inflammation in gouty arthritis. *Rheumatology* **56**, 457–466 (2017).
54. Castel, S. E. et al. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* **21**, 234 (2020).
55. Shang, K. et al. IL-33 ameliorates the development of MSU-induced inflammation through expanding MDSCs-like cells. *Front. Endocrinol.* **10**, 36 (2019).
56. Cadzow, M. et al. Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. *Arthritis Res. Ther.* **19**, 181 (2017).
57. Dalbeth, N. et al. Survey definitions of gout for epidemiologic studies: comparison with crystal identification as the gold standard.*Arthritis Care Res.* **68**, 1894–1898 (2016).
58. He, B. et al. Hypercholesterolemia risk associated Abca6 does not regulate lipoprotein metabolism in mice or hamster. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **1866**, 159006 (2021).
59. Basseville, A. et al. Histone deacetylase inhibitors influence chemotherapy transport by modulating expression and trafficking of a common polymorphic variant of the ABCG2 efflux transporter. *Cancer Res.* **72**, 3642–3651 (2012).
60. Matsuo, H. et al. Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. *Sci. Transl. Med.* **1**, 5ra11 (2009).
61. Wallace, M. C. et al. Association between ABCG2 rs2231142 and poor response to allopurinol: replication and meta-analysis. *Rheumatology* **57**, 656–660 (2018).
62. Takeshita, T. et al. The contribution of polymorphism in the alcohol dehydrogenase β subunit to alcohol sensitivity in a Japanese population. *Hum. Genet.* **97**, 409–413 (1996).
63. Farrés, J. et al. Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde dehydrogenase. A model to study human (Oriental type) class 2 aldehyde dehydrogenase. *J. Biol. Chem.* **269**, 13854–13860 (1994).
64. Yamanaka, H. et al. Analysis of the genotypes for aldehyde dehydrogenase 2 in Japanese patients with primary gout. *Adv. Exp. Med. Biol.* **370**, 53–56 (1994).
65. Rees, M. et al. Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* **55**, 114–122 (2012).
66. Xiang, Z. et al. Pharmacological characterization of 40 human melanocortin-4 receptor polymorphisms with the endogenous proopiomelanocortin-derived agonists and the agouti-related protein (AGRP) antagonist. *Biochemistry* **45**, 7277–7288 (2006).
67. Pirazzi, C. et al. Patatin-like phospholipase domain-containing 3 (PNPLA3) I148M (rs738409) affects hepatic VLDL secretion in humans and in vitro. *J. Hepatol.* **57**, 1276–1282 (2012).
68. Allenspach, E. J. et al. The autoimmune risk R262W variant of the adaptor SH2B3 improves survival in sepsis. *J. Immunol.* **207**, 2710–2719 (2021).
69. Wang, W. et al. LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis. *Circ. Res.* **119**, e91–e103 (2016).
70. Chiba, T. et al. NPT1/SLC17A1 is a renal urate exporter in humans and its common gain-of-function variant decreases the risk of renal underexcretion gout. *Arthritis Rheumatol.* **67**, 281–287 (2015).
71. Jutabha, P. et al. Functional analysis of human sodium–phosphate transporter 4 (NPT4/*SLC17A3*) polymorphisms. *J. Pharmacol. Sci.* **115**, 249–253 (2011).
72. Matsuo, H. et al. Mutations in glucose transporter 9 gene *SLC2A9* cause renal hypouricemia. *Am. J. Hum. Genet.* **83**, 744–751 (2008).
73. Hurba, O. et al. Complex analysis of urate transporters SLC2A9, SLC22A12 and functional characterization of non-synonymous allelic variants of GLUT9 in the Czech population: no evidence of effect on hyperuricemia and gout. *PLoS ONE* **9**, e107902 (2014).
74. Hall, S. C. et al. Critical role of zinc transporter (ZIP8) in myeloid innate immune cell function and the host response against bacterial pneumonia. *J. Immunol.* **207**, 1357–1370 (2021).
75. Fujishiro, H. et al. Effects of individual amino acid mutations of zinc transporter ZIP8 on manganese- and cadmium-transporting activity. *Biochem. Biophys. Res. Commun.* **616**, 26–32 (2022).
76. Türkmen, D. et al. Statin treatment effectiveness and the *SLCO1B1**5 reduced function genotype: long-term outcomes in women and men. *Br. J. Clin. Pharmacol.* **88**, 3230–3240 (2022).
77. Zhao, B. et al. Identification of potential megalin/cubilin substrates using extensive proteomics quantification from kidney megalin-knockdown mice. *AAPS J.* **24**, 109 (2022).
78. Simonson, B. et al. DDiT4L promotes autophagy and inhibits pathological cardiac hypertrophy in response to stress. *Sci. Signal.* **10**, eaaf5967 (2017).
79. Scoville, D. W. & Jetten, A. M. GLIS3: a critical transcription factor in islet β-cell generation. *Cells* **10**, 3471 (2021).
80. Adelmann, C. H. et al. MFSD12 mediates the import of cysteine into melanosomes and lysosomes. *Nature* **588**, 699–704 (2020).
81. Ge, W. et al. POM121 inhibits the macrophage inflammatory response by impacting NF-κB P65 nuclear accumulation. *Exp. Cell Res.* **377**, 17–23 (2019).
82. Moon, J.-S. et al. ANT2 drives proinflammatory macrophage activation in obesity. *JCI Insight* **6**, e147033 (2021).
83. Ghossoub, R. et al. Tetraspanin-6 negatively regulates exosome production. *Proc. Natl Acad. Sci. USA* **117**, 5913–5922 (2020).
84. Bhatt-Wessel, B. et al. Role of DGAT enzymes in triacylglycerol metabolism. *Arch. Biochem. Biophys.* **655**, 1–11 (2018).
85. Inoue, H. et al. γ-SNAP stimulates disassembly of endosomal SNARE complexes and regulates endocytic trafficking pathways. *J. Cell Sci.* **128**, 2781–2794 (2015).
86. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

Tanya J. Major[1,77], Riku Takei[2,77], Hirotaka Matsuo[3,4,77], Megan P. Leask[2,77], Nicholas A. Sumpter[2,77], Ruth K. Topless[1,77], Yuya Shirai[5,77], Wei Wang[6,77], Murray J. Cadzow[1], Amanda J. Phipps-Green[1], Zhiqiang Li[7], Aichang Ji[8,9], Marilyn E. Merriman[2], Emily Morice[2], Eric E. Kelley[10], Wen-Hua Wei[1,11], Sally P. A. McCormick[1], Matthew J. Bixley[1], Richard J. Reynolds[2], Kenneth G. Saag[2], Tayaza Fadason[12], Evgenia Golovina[12], Justin M. O'Sullivan[12,13,14,15], Lisa K. Stamp[16], Nicola Dalbeth[17], Abhishek Abhishek[18], Michael Doherty[18], Edward Roddy[19,20], Lennart T. H. Jacobsson[21], Meliha C. Kapetanovic[22], Olle Melander[23,24], Mariano Andrés[25,26], Fernando Pérez-Ruiz[27], Rosa J. Torres[28,29], Timothy Radstake[30], Timothy L. Jansen[31], Matthijs Janssen[31], Leo A. B. Joosten[32,33], Ruiqi Liu[32], Orsolya I. Gaal[33], Tania O. Crişan[33], Simona Rednic[34], Fina Kurreeman[35], Tom W. J. Huizinga[35], René Toes[35], Frédéric Lioté[36,37], Pascal Richette[37], Thomas Bardin[37], Hang Korng Ea[37], Tristan Pascart[38], Geraldine M. McCarthy[39], Laura Helbert[39], Blanka Stibůrková[40,41], Anne-K. Tausche[42], Till Uhlig[43], Véronique Vitart[44], Thibaud S. Boutin[44], Caroline Hayward[44], Philip L. Riches[45], Stuart H. Ralston[45], Archie Campbell[45], Thomas M. MacDonald[46], FAST Study Group*, Akiyoshi Nakayama[3], Tappei Takada[47], Masahiro Nakatochi[48], Seiko Shimizu[3], Yusuke Kawamura[2,3,49], Yu Toyoda[3], Hirofumi Nakaoka[49], Ken Yamamoto[50], Keitaro Matsuo[51,52,53], Nariyoshi Shinomiya[3], Kimiyoshi Ichida[54], Japan Gout Genomics Consortium*, Chaeyoung Lee[55], Asia Pacific Gout Consortium*, Linda A. Bradbury[56], Matthew A. Brown[56], Philip C. Robinson[57], Russell R. C. Buchanan[58], Catherine L. Hill[59,60], Susan Lester[59,60], Malcolm D. Smith[61], Maureen Rischmueller[59,60], Hyon K. Choi[62], Eli A. Stahl[63], Jeff N. Miner[64], Daniel H. Solomon[65], Jing Cui[65], Kathleen M. Giacomini[66], Deanna J. Brackman[66], Eric M. Jorgenson[67], GlobalGout Genetics Consortium*, Hongbo Liu[68,69], Katalin Susztak[68,69], 23andMe Research Team*, Suyash Shringarpure[6], Alexander So[70,71], Yukinori Okada[5,72,73,77], Changgui Li[8,9,77], Yongyong Shi[74,75,77] & Tony R. Merriman[2,9,76] ✉

[1]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [2]Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA. [3]Department of Integrative Physiology and Bio-Nano Medicine, National Defense Medical College, Saitama, Japan. [4]Department of Biomedical Information Management, National Defense Medical College Research Institute, National Defense Medical College, Saitama, Japan. [5]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. [6]Genomics R&D, 23andMe, Inc, Sunnyvale, CA, USA. [7]The Biomedical Sciences Institute and The Affiliated Hospital of Qingdao University, Qingdao University, Qingdao, Shandong, China. [8]Shandong Provincial Key Laboratory of Metabolic Diseases, Shandong Provincial Clinical Research Center for Immune Diseases and Gout, the Affiliated Hospital of Qingdao University, Qingdao, Shandong, China. [9]The Institute of Metabolic Diseases, Qingdao University, Qingdao, Shandong, China. [10]Department of Physiology and Pharmacology, West Virginia University, Morgantown, WV, USA. [11]Department of Women's and Children's Health, University of Otago, Dunedin, New Zealand. [12]Liggins Institute, University of Auckland, Auckland, New Zealand. [13]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom. [14]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, Singapore, Singapore. [15]Australian Parkinsons Mission, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. [16]Department of Medicine, University of Otago, Christchurch, Christchurch, New Zealand. [17]Department of Medicine, University of Auckland, Auckland, New Zealand. [18]Academic Rheumatology, School of Medicine, University of Nottingham, Nottingham, United Kingdom. [19]School of Medicine, Keele University, Keele, Staffordshire, United Kingdom. [20]Haywood Academic Rheumatology Centre, Midlands Partnership University NHS Foundation Trust, Stoke-on-Trent, UK. [21]Department of Rheumatology and Inflammation Research, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [22]Department of Clinical Sciences Lund, Section of Rheumatology, Lund University and Skåne University Hospital, Lund, Sweden. [23]Department of Clinical Sciences, Lund University, Malmö, Sweden. [24]Department of Emergency and Internal Medicine, Skåne University Hospital, Malmö, Sweden. [25]Rheumatology Department, Dr Balmis General University Hospital-ISABIAL, Alicante, Spain. [26]Department of Clinical Medicine, Miguel Hernandez University, Alicante, Spain. [27]Osakidetza, OSI-EE-Cruces, BIOBizkaia Health Research Institute and Medicine Department of Medicine and Nursery School, University of the Basque Country, Biskay, Spain. [28]Department of Biochemistry, Hospital La Paz Institute for Health Research (IdiPaz), Madrid, Spain. [29]Center for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain. [30]Department of Rheumatology and Clinical Immunology, University Medical Center, Utrecht, The Netherlands. [31]Department of Rheumatology, VieCuri Medical Centre, Venlo, The Netherlands. [32]Department of Internal Medicine and Radboud Institute of Molecular Life Science, Radboud University Medical Center, Nijmegen, The Netherlands. [33]Department of Medical Genetics, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania. [34]Department of Rheumatology, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Cluj, Romania. [35]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands. [36]Rheumatology Department, Feel'Gout, GH Paris Saint Joseph, Paris, France. [37]Rheumatology Department, INSERM U1132, BIOSCAR, University Paris Cité, Lariboisière Hospital, Paris, France. [38]Department of Rheumatology, Hopital Saint-Philibert, Lille Catholic University, Lille, France. [39]Department of Rheumatology, Mater Misericordiae University Hospital and School of Medicine, University College, Dublin, Ireland. [40]Department of Pediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic. [41]Institute of Rheumatology, Prague, Czech Republic. [42]Department of Rheumatology, University Clinic 'Carl Gustav Carus' at the Technical University, Dresden, Germany. [43]Center for Treatment of Rheumatic and Musculoskeletal Diseases, Diakonhjemmet Hospital, Oslo, Norway. [44]Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom. [45]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [46]MEMO Research, Division of Molecular and Clinical Medicine, University of Dundee Medical School, Ninewells Hospital, Dundee, United Kingdom. [47]Department of Pharmacy, The University of Tokyo Hospital, Tokyo, Japan. [48]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Aichi, Japan. [49]Department of Cancer Genome Research, Sasaki Institute, Sasaki Foundation, Tokyo, Japan. [50]Department of Medical Biochemistry, Kurume University School of Medicine, Fukuoka, Japan. [51]Division of Cancer Epidemiology & Prevention, Aichi Cancer Center, Aichi, Japan. [52]Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Aichi, Japan. [53]The Japan Multi-Institutional Collaborative Cohort (J-MICC) Study, Tokyo, Japan. [54]Department of Pathophysiology, Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan. [55]Department of Bioinformatics and Life Science, Soongsil University, Seoul, South Korea. [56]Institute of Health and Biomedical Innovation, Translational Research Institute, Queensland University of Technology, Brisbane, Australia. [57]School of

Clinical Medicine, Faculty of Medicine, University of Queensland, Brisbane, Australia. [58]Department of Rheumatology, Austin Hospital, Melbourne, Victoria, Australia. [59]Rheumatology Department, The Queen Elizabeth Hospital, Woodville South, South Australia, Australia. [60]Discipline of Medicine, University of Adelaide, Adelaide, Australia. [61]Flinders University, Adelaide, South Australia, Australia. [62]Division of Rheumatology, Allergy, and Immunology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [63]Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [64]Viscient Biosciences, 5752 Oberlin Dr., Suite 111, San Diego, CA 92121, USA. [65]Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [66]Department of Bioengineering and Therapeutic Sciences and Institute for Human Genetics, University of California, San Francisco, CA, USA. [67]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [68]Penn / The Children's Hospital of Pennsylvania Kidney Innovation Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19101, USA. [69]Renal Electrolyte and Hypertension Division, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19101, USA. [70]Service of Rheumatology, Center Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [71]University of Lausanne, Lausanne, Switzerland. [72]Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [73]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [74]Affiliated Hospital of Qingdao University and Biomedical Sciences Institute of Qingdao University (Qingdao Branch of SJTU Bio-X Institutes), Qingdao University, Qingdao, China. [75]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Collaborative Innovation Center for Brain Science, Shanghai Jiao Tong University, Shanghai, China. [76]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. [77]These authors contributed equally: Tanya J. Major, Riku Takei, Hirotaka Matsuo, Megan P. Leask, Nicholas A. Sumpter, Ruth K. Topless, Yuya Shirai, Wei Wang, Yukinori Okada, Changgui Li, Yongyong Shi. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: tony.merriman@otago.ac.nz

FAST Study Group

Fernando Pérez-Ruiz[27], Stuart H. Ralston[45] & Thomas M. MacDonald[46]

Japan Gout Genomics Consortium

Yuya Shirai[5,77], Akiyoshi Nakayama[3], Tappei Takada[47], Masahiro Nakatochi[48], Seiko Shimizu[3], Yusuke Kawamura[2,3,49], Yu Toyoda[3], Hirofumi Nakaoka[49], Ken Yamamoto[50], Keitaro Matsuo[51,52,53], Nariyoshi Shinomiya[3], Kimiyoshi Ichida[54], Yukinori Okada[5,72,73,77] & Hirotaka Matsuo[3,4,77]

Asia Pacific Gout Consortium

Tanya J. Major[1,77], Riku Takei[2,77], Hirotaka Matsuo[3,4,77], Yuya Shirai[5,77], Aichang Ji[8,9], Wen-Hua Wei[1,11], Lisa K. Stamp[16], Nicola Dalbeth[17], Thomas Bardin[37], Blanka Stibůrková[40,41], Akiyoshi Nakayama[3], Tappei Takada[47], Masahiro Nakatochi[48], Seiko Shimizu[3], Yusuke Kawamura[2,3,49], Yu Toyoda[3], Nariyoshi Shinomiya[3], Kimiyoshi Ichida[54], Chaeyoung Lee[55], Yukinori Okada[5,72,73,77], Changgui Li[8,9,77], Yongyong Shi[74,75,77] & Tony R. Merriman[2,9,76]

GlobalGout Genetics Consortium

Tanya J. Major[1,77], Riku Takei[2,77], Hirotaka Matsuo[3,4,77], Megan P. Leask[2,77], Sally P. A. McCormick[1], Lisa K. Stamp[16], Nicola Dalbeth[17], Abhishek Abhishek[18], Michael Doherty[18], Edward Roddy[19,20], Lennart T. H. Jacobsson[21], Meliha C. Kapetanovic[22], Olle Melander[23,24], Mariano Andrés[25,26], Fernando Pérez-Ruiz[27], Rosa J. Torres[28,29], Timothy Radstake[30], Timothy L. Jansen[31], Matthijs Janssen[31], Leo A. B. Joosten[32,33], Tania O. Crişan[33], Simona Rednic[34], Fina Kurreeman[35], Tom W. J. Huizinga[35], René Toes[35], Frédéric Lioté[36,37], Pascal Richette[37], Thomas Bardin[37], Hang Korng Ea[37], Tristan Pascart[38], Geraldine M. McCarthy[39], Blanka Stibůrková[40,41], Akiyoshi Nakayama[3], Yusuke Kawamura[2,3,49], Anne-K. Tausche[42], Till Uhlig[43], Philip L. Riches[45], Linda A. Bradbury[56], Matthew A. Brown[56], Philip C. Robinson[57], Russell R. C. Buchanan[58], Catherine L. Hill[59,60], Susan Lester[59,60], Malcolm D. Smith[61], Maureen Rischmueller[59,60], Hyon K. Choi[62], Eli A. Stahl[63], Daniel H. Solomon[65], Alexander So[70,71] & Tony R. Merriman[2,9,76]

23andMe Research Team

Wei Wang[6,77] & Suyash Shringapure[6]

Methods

Ethics statement

For the 23andMe cohort, participants provided informed consent and participated in the research online under a protocol approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services (https://www.eandireview.com). Ethical and Independent Review Services was recently acquired, and its new name as of July 2022 is Salus IRB (https://www.versiticlinicaltrials.org/salusirb). The UK Biobank study was undertaken with ethical approval from the North West Multi-Centre Research Ethics Committee of the UK. GlobalGout obtained ethical approval from the following committees: the New Zealand Multiregional Ethics Committee (MEC05/10/130); the Research and Ethics Committee, Repatriation General Hospital, South Australia (32/08); the Research Ethics Committee, University of New South Wales; the Ethikkommission, Technische Universität Dresden (EK 8012012); the South East Scotland Research Ethics Committee (04/S1102/41); the Commission Cantonale (VD) d'Éthique de la Recherche sur l'Être Humain, Université de Lausanne; the Commissie Mensgebonden Onderzoek regio Arnhem–Nijmegen; and the Partners Health Care System Institutional Review Board. The institutional review board of the Kaiser Foundation Research Institute provided ethical approval for the Kaiser Permanente sample set. The FinnGen study was approved by the Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa. The ethics review board at the Affiliated Hospital of Qingdao University approved the study in China. In Japan, ethical approvals were provided by the institutional ethical committees of the National Defense Medical College, Nagoya University and RIKEN. The Korean Association Resource was approved by the institutional review board of the Korea National Institute of Health. The FAST study and Generation Scotland were ethically approved by the UK Multi-Centre Research Ethics Committee (reference number 2011-001883-23) and the NHS Tayside Committee on Medical Research Ethics (REC reference number 05/S1401/89).

Study participants and phenotype definition

Study participants were sourced from 13 cohorts[8,9,87–94] (Supplementary Table 1). Gout was defined according to self-reported diagnosis and/or use of urate-lowering medication, clinical diagnosis and/or the American College of Rheumatology gout classification criteria[95], dependent on the information available in each source cohort. Excluding the 2,115 participants from ref. 90, 77.3% of people with gout were ascertained by self-report, 18.1% by clinical diagnosis and 4.6% using administrative data. Participants from these cohorts were divided into four ancestral groups (African, East Asian, European and Latinx) based on their self-reported ethnicity or genetic ancestry as outlined in Supplementary Table 1. This resulted in two African, five East Asian, ten European and two Latinx case–control study sets for analysis. Written informed consent was acquired from all participants, and each study had ethical approval from the appropriate local ethics committee. The UK Biobank[96] was also used for genetic risk prediction, genetic correlation analysis with other phenotypes, Mendelian randomization and fine-mapping. Participants in the 23andMe cohort[97] provided informed consent and participated in the research online and were included in the analysis based on consent status as checked at the time when data analyses were initiated.

Genotyping and imputation

Genotyping was performed separately for each of the 13 source cohorts. Cohort-specific genotyping and imputation platforms and post-imputation QC filters are detailed in Supplementary Table 1. Genotypes were imputed to one of the following reference panels: UK10K[98] + 1000 Genomes Project Phase 3 (all ancestries)[99], the Haplotype Reference Consortium panel version 1.1 (ref. 100), 1000 Genomes Phase 3 (all ancestries), the combined Haplotype Reference Consortium panel version 1.1 and UK10K + 1000 Genomes Project Phase 3 (all ancestries), HapMap phase 2 (ref. 101) (build 36) or the population-specific Sequencing Initiative Suomi project version 3 (build 38)[102]. The genome coordinates of the build 36- and build 38-imputed genotypes were converted to their equivalent build 37 (hg19) coordinates using SNPTracker version 1.0 (released December 2014)[103] or GATK Liftover[104], respectively. Post-imputation QC filters per cohort included filters for Hardy–Weinberg equilibrium, genotype missingness, minor allele frequency and imputation quality (Supplementary Table 1). Only biallelic SNPs were used for association analyses.

Study-specific association analysis

A genome-wide logistic regression analysis was conducted for each of the 19 case–control study sets separately with gout as a binary outcome. The regression model was adjusted for sex, age, genetic principal-component vectors and any necessary cohort-specific variables (for example, genotyping platform) (Supplementary Table 1). Regression analyses were repeated in male-only and female-only subsets for each case–control study set when possible, including the same adjusting variables except sex. QC steps applied to the per-study regression summary statistics included the removal of multi-allelic variants within and between study sets, the exclusion of results for SNPs present in less than 5% of participants and the removal of any results outside the statistical bounds of each output (for example, standard error = infinity). Variants with a minor allele frequency less than 0.1% or an absolute effect size (log (odds ratio)) greater than 10 were also removed from the study-specific summary statistics.

Ancestry-specific and trans-ancestry meta-analysis

An inverse-variance-weighted meta-analysis of the regression results for the full, male-only and female-only subsets was conducted for each of the four ancestral groups using METAL[105]. LDSC analysis revealed no evidence of inflation due to factors such as population structure; therefore, no genomic control correction was applied during this meta-analysis. SNPs with a heterogeneity I^2 statistic > 95% were excluded from any further analysis. Variants analyzed in <50% of the case–control study sets contributing to the meta-analysis (when there were more than two contributing cohorts), and/or variants that were analyzed in <20% of all contributing samples were also excluded from any further analysis.

The ancestry-specific meta-analysis results were then used as an input for the TAMA with MANTRA version 1 (ref. 106). Only SNPs with an effect size present in at least two ancestral groups were included in the TAMA. Genetic variants with meta-analysis $P < 5 \times 10^{-8}$ or $\log_{10}$ (Bayes factor) ≥ 6.0 were considered statistically significant in the ancestry-specific and TAMA analyses, respectively (Supplementary Note). Within the TAMA, MANTRA was also used to calculate the PP of heterogeneity, which estimates the homogeneity in allelic effects across ancestral groups. A PP > 0.90 indicates strong evidence for heterogeneity. In addition, loci with heterogeneous effect sizes between sexes were identified by performing an inverse-variance-weighted meta-analysis of the male-only and female-only ancestry-specific GWAS or TAMA effect estimates. A SNP was considered significantly heterogeneous between sexes if the heterogeneity Q statistic had $P < 1 \times 10^{-6}$.

We calculated the effective symmetric size that yielded equivalent power to the included cases and controls for each ancestry to be 5,874 cases and 5,874 controls for African (2.6% of effective N), 18,997 cases and 18,997 controls for East Asian (8.3% of effective N), 192,138 cases and 192,138 cases for European (84.1% of effective N) and 11,398 cases and 11,398 controls for Latinx (5.0% of effective N) groups.

Identification of genome-wide significant signals and loci

A locus was defined as a genomic segment with ≥1 lead variant(s) that define a genetically independent signal. All nominally significant SNPs were identified ($P < 1 \times 10^{-7}$ for ancestry-specific GWAS or $\log_{10}$ (Bayes factor) > 5.0 for the TAMA), and a ±50-kb window of padding was added to the SNP chromosome position. The genomic overlap of these padded

chromosome positions was then identified, and the outer boundaries of these regions were created using GenomicRanges version 3.12 (ref. 107). After defining these locus boundaries, any regions with ≤1 SNP classified as genome-wide significant ($P < 5 \times 10^{-8}$ or $\log_{10}$ (Bayes factor) > 6.0 (ref. 108)) were disregarded in further locus definitions or analyses due to the potential unreliability of single-variant associations. Lead SNP(s) within each of the significant loci defined above were identified by conducting LD-based clumping using PLINK version 1.9b4 (ref. 109) with an ancestry-matched 1000 Genomes Project reference panel (Supplementary Note) and an r^2 threshold of 0.01. Significant loci were designated based on their chromosome and start and end positions (in Mb) of the significant region (for example, chr5:129.52 Mb–131.88 Mb).

A regional association plot (LocusZoom) was created for each locus (Supplementary Data 1) using the R script 'LocusZoom-like plots' (ref. 110) with the GENCODE human genome build 37 (release 38) gene annotation file to define gene start and end positions and gene names (HGNC gene symbol) in the gene track information[111]. The full gene annotation file was narrowed to unique gene names only and filtered to exclude pseudogenes. Gene track information was colored based on MAGMA[112] gene-based association analysis *P* values (Supplementary Note). After creating these regional association plots, each locus with multiple lead SNPs was visually inspected to determine whether the labeled SNPs represented a single association signal or multiple independent spatially separated association signals. For loci with extremely strong association signals (for example, *SLC2A9* and *ABCG2*), only a single lead SNP was included to represent the locus, even if multiple lead SNPs were identified by LD-based clumping (Supplementary Table 2). To identify conditional associations, European GWAS loci were analyzed with GCTA-COJO[113] by the stepwise method using the UK Biobank cohort as a reference panel (Supplementary Table 3). Due to a lack of adequately sized reference panels for the other ancestries, we did not conduct conditional analysis on the other three ancestral groups.

Previously unreported genome-wide significant loci and signals

To identify newly detected loci and signals, we searched the literature for all previously reported SNPs associated with serum urate or gout. A search of the GWAS Catalog[114] was performed on 19 July 2022, using the keywords 'urate', 'uric acid', 'hyperuric[a]emia' and 'gout'. This identified 68 studies published over the period from November 2007 to November 2021. All significant SNPs identified in these 68 studies were downloaded. Using build 38 genomic locations (locus boundaries were converted to build 38 genomic locations using Liftover[104]), we identified any previous GWAS signal that fell within our significant locus boundaries defined above. LD between the lead SNP and all previously identified SNPs in the locus was calculated using 1000 Genomes reference panels in LDlinkR[115] with ancestry matching to the lead SNP. LD values were categorized into high ($r^2 \geq 0.8$), moderate ($0.8 > r^2 \geq 0.5$), low ($0.5 > r^2 \geq 0.1$) or weak LD ($r^2 < 0.1$) groups to assess how similar the previously reported association signal was to the lead SNP. A locus was only considered as previously unreported if none of the 68 urate and gout GWAS had previously reported a significant SNP within the locus boundaries or there was only weak LD between the lead SNP and any previously reported SNP.

Testing of gout-associated loci for association with urate

In total, 332,370 unrelated participants of European ancestry were selected from the UK Biobank using self-reported and genetic (principal-component-determined) ancestry information (Supplementary Note). Participants with evidence of gout ($n = 7{,}131$) based on hospital-diagnosed gout, self-reported gout and use of urate-lowering medication and those with no serum urate measurement ($n = 15{,}531$) were excluded, resulting in 309,708 participants in the association analysis for serum urate levels. Variants with INFO < 0.3, MAF < 0.0001 and HWE $P < 1.0 \times 10^{-6}$ were removed. Association analysis was carried out using PLINK version 1.9b6.10 with hard-called imputed genotype data, adjusted by age, sex and principal-component vectors 1–40. Summary statistics from Tin et al.[7] for the European ancestry group were meta-analyzed together with the UK Biobank urate summary statistics and used to test for colocalization of gout loci with urate and vice versa. Both Tin et al. and UK Biobank summary statistics were used to group variants into urate-associated or non-urate-associated categories for ex vivo analysis of the IL-1β response to MSU crystal stimulation of PBMCs in the Human Functional Genomics Project (500FG) cohort[116] (Supplementary Note).

LD score regression with UK Biobank traits and diseases

LDSC[86] was used to run genetic correlation analysis of UK Biobank phenotypes with European gout GWAS summary statistics. Summary statistics for 934 primary phenotypes and baseline version 1.1 reference LD scores for European ancestry were used. Because there was complete sample overlap of the FinnGen cohort in our European GWAS meta-analysis, we excluded all FinnGen phenotypes to avoid sample overlap. We considered any genetic correlation with our European gout GWAS to be significant if the *P* value was less than a Bonferroni-corrected significance threshold of $P < 5.4 \times 10^{-5}$ (0.05/934). The resulting significantly correlated GWAS were then tested for horizontal pleiotropy using the two-sample Mendelian randomization package TwoSampleMR[117].

Covariate-adjusted LD score regression

LD score intercepts, tissue or cell type group and cell type-specific enrichment analysis for each ancestry (African, East Asian, European and Latinx) were determined using covariate-adjusted LDSC (cov-LDSC)[86]. Baseline LD scores for each ancestry were generated using ancestry-specific principal-component vectors from the 1000 Genomes Project (Supplementary Note), annotations from baseline version 1.1 and an LD window set to 20 cM. Baseline LD scores for the cell type group and cell type-specific analyses were generated similarly using the cell type group and cell type-specific annotations. The significance threshold for the coefficient *P* value for the cell type group analysis was set to 0.005 for a standard cutoff (for ten cell type groups) and 0.00125 for a conservative cutoff (for ten tissue or cell type groups and four ancestries). The threshold for cell type-specific analysis was set to 2.3×10^{-4} for a standard cutoff (for 220 cell types) and 5.7×10^{-5} for a conservative cutoff (for 220 cell types and four ancestries).

Statistical fine-mapping and credible set construction

Loci were fine-mapped using PAINTOR[118] and FINEMAP[17] in the European dataset and the Bayes factor approach in the TAMA dataset using variants that passed QC metrics in each of the four ancestral groups[106,119]. The region to be fine-mapped was set to the locus boundaries described in Supplementary Table 2 unless the region was larger than 1 Mb; in such cases, the fine-mapping region was defined as the 1-Mb region around the lead SNP but in such a way that the region to be fine-mapped did not exceed the locus boundary if the lead SNP was <500 kb from the boundary. For PAINTOR and FINEMAP, LD information from the UK Biobank was used. All variants were aligned to the same allele as in the UK Biobank and any variant with MAF that differed by >5% was excluded from the fine-mapping analysis. Prior standard deviation for FINEMAP was set to 0.058 (based on equation (8) in Wakefield[120], using the 95% confidence interval of the effect size (0.1134) in the European GWAS data), and a shotgun stochastic search algorithm was used (–sss option) with the maximum number of causal variants in a locus set to five (default). Five annotations were used for PAINTOR: four annotations were chosen based on the relevance of the annotation with the European gout GWAS summary statistics, and one annotation was included to give more weight to missense variants. PAINTOR was run with full enumeration (–enum option), assuming single causal variants except for those loci with evidence for multiple signals, in which case, a maximum of three causal variants was assumed. For the Bayes factor approach, Bayes factors of all variants were summed within

the fine-mapping region and the PP of each variant being causal in the region was calculated by dividing the variant's Bayes factor by the total Bayes factor of the region. The 99% credible set was determined by including the variant with the highest PP in the region until the cumulative PP of the set was greater than 0.99.

Identification of compromised loci with SLALOM

To mitigate fine-mapping miscalibration due to different genotyping arrays and imputation panels being used in the meta-analyzed cohorts (Supplementary Table 1), the summary statistics used for fine-mapping were checked with SLALOM[19]. The same prior standard deviation value from FINEMAP was used for SLALOM. Conversion-unstable positions[121] were downloaded, and reference LD from gnomAD (build 37) was calculated in the non-Finnish European population for the European ancestry cohort. A locus was identified as possibly miscalibrated if the locus contained at least one variant with DENTIST-S[122] $P < 1 \times 10^{-4}$ and $r^2 > 0.6$ with the lead variant (in the relevant reference population).

Candidate missense and noncoding variants

A pool of 1,465 candidate variants was gathered from lead variants, conditionally associated variants and variants identified from fine-mapping. LD with noncoding variants was calculated using 1000 Genomes Project data, and direct lead or high-LD ($r^2 \geq 0.8$) missense and nonsense variants were identified using information from dbSNP (build 155). FATHMM scores for the noncoding variants were queried, and the variants were overlaid for ABC enhancers.

Colocalization and investigation of genetic association signals of gout and urate

Lead SNPs of independent signals from the meta-analyzed Tin et al.[7] and UK Biobank serum urate GWAS in the European group (422 variants; Methods and Supplementary Note) and the European full gout GWAS (291 variants) were used for the colocalization analysis using the 'coloc' R package[123]. Colocalization analysis was performed using a 1-Mb (±500-kb) region around the lead variant, and PP ≥ 0.8 was considered to be strong evidence for a given hypothesis (Supplementary Note). The 'coloc' priors P_1 (probability of association with trait 1), P_2 (probability of association with trait 2) and P_{12} (probability of association with trait 2 given association with trait 1) were initially set to 1×10^{-4}, 1×10^{-4} and 1×10^{-5}, respectively. Additional loci with PP of H_3 or $H_2 + H_4 \geq 0.8$ were detected by adjusting P_{12} to 1×10^{-4} and 1×10^{-6}. For the loci with evidence of gout-specific effect (PP $H_1 \geq 0.8$) or pleiotropic effect (PP of H_2, H_3 or $H_2 + H_4 \geq 0.8$), the P value of the lead variant of the locus was checked for association in the other GWAS, with $P > 0.01$ indicating no significant association. To ensure that variants that showed evidence of association with urate but not gout were of sound quality, variants with MAF < 1% or with a >10% allele frequency difference between gout and urate GWAS or those not present in the large 23andMe European gout cohort were removed. The lead variant of 31 signals in the gout GWAS and that of 37 signals in the urate GWAS of loci, identified by the GCTA-COJO stepwise method as having two genetically independent signals, were used to perform conditional association analysis using GCTA-COJO, and the resultant summary statistics from loci that still exhibited a genome-wide significant signal (three from the conditioned analysis of the urate GWAS were not significant) were used for colocalization analysis. Detection power in the gout GWAS, given a signal in the urate GWAS, was calculated at a variety of alpha values[124]. For all power calculations, we used the N value for cases and controls in the European GWAS (~100,000 cases, ~2.1 million controls) and input the MAF of each variant along with its expected gout effect size based on its urate effect size (derived from the regression line in Supplementary Fig. 3a).

Colocalization of signals of genetic association with eQTL

Lead SNPs from the European and trans-ancestry analyses or a proxy if the lead variant was not present in the GTEx database were used to query GTEx data[54] for identification of *cis*-eQTL. For all SNPs with a significant eQTL, a 1-Mb (±500-kb) region around the lead variant was used for colocalization analysis using the 'coloc' R package[123,125]. PP ≥ 0.8 for H_4 was taken as evidence for colocalization. This hypothesis (H_4) indicated that the association between gout and gene expression was due to the same functional variant(s) (Supplementary Note). We also carried out *trans*-eQTL analysis, combining GTEx and publicly available Hi-C data integrated using the CoDeS3D (Contextualize Development SNPs using 3D information) algorithm[126], which prescreens and identifies spatial non-*cis* SNP–gene interactions using Hi-C data (Supplementary Note). CoDeS3D was used to identify long-distance regulatory relationships for all European and trans-ancestry lead variants or proxies. Variants were mapped onto Hi-C restriction fragments, and the genes that physically connected to these restriction fragments were identified and collated (variant–gene spatial pairs). Variant–gene pairs were screened using GTEx to identify spatially connected eQTL. The FDR was calculated using a stepwise Benjamini–Hochberg correction procedure and incorporated the number of tests and the eQTL value list. FDR < 0.05 was considered statistically significant.

Colocalization of signals of genetic association with methylation QTL

The GoDMC dataset[29] was used to first identify all significant DNA meQTL within a ±1-Mb window of the 444 unique lead European GWAS variants in the full and sex-specific analyses. For each pair of variant–CpG sites, a 1-Mb (±500-kb) region around the lead GWAS variant was defined and tested for colocalization with the meQTL signal using the 'coloc' R package. Regions with <100 variants with meQTL summary statistics for the CpG sites were excluded, and regions with PP ≥ 0.8 for H_4 were considered colocalized.

Enrichment analysis of transcription factor binding at meQTL CpG sites

A total of 1,544 transcription factor ChIP–seq experiment datasets (344 TFs and 221 cell lines)[127] were downloaded. Eight transcription factor ChIP–seq experiment datasets were excluded to ensure that all the ChIP–seq experiments were from human cell lines, leaving 338 unique transcription factors for enrichment analysis. Positions for 232,477 independent CpG sites used in GoDMC[29] were extracted. For each transcription factor, the total number of CpG sites the transcription factor overlapped (±50 bp) within the CpG sites was determined, and 520 colocalized meQTL CpG sites were identified. In total, 338 two-by-two contingency tables were constructed, and Fisher's exact test was used to determine whether there was significant enrichment of transcription factor binding to the meQTL CpG sites. Fisher's exact $P < 1.48 \times 10^{-4}$ (0.05/338) was considered statistically significant.

Overlap of variants and CpG sites with activity-by-contact enhancer region

Thirty ABC enhancer–gene connection datasets with ABC scores ≥ 0.015 (ref. 34) for stimulated primary innate immune cells and cell lines ($CD14^+$ monocytes treated with β-glucan or LPS, dendritic cells treated with LPS, PMA-differentiated THP-1 cell line treated with LPS, U937 cell line treated with LPS) were downloaded. ABC enhancer–gene connections at the 12 loci with connections in the stimulated immune cells were also tested in 27 other immune cell datasets. The positions of candidate noncoding SNPs, lead SNPs that had colocalized eQTL and CpG sites that had colocalized meQTL were queried in the ABC data to determine whether the SNP and/or the CpG site was within an ABC enhancer region.

Reporting summary

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

Data availability

The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Datasets will be made available at no cost for academic use (https://research.23andme.com/collaborate/#dataset-access/). Supplementary Table 40 contains summary statistics from 9,980 independently associated SNPs. For the full European GWAS, we took all SNPs at $P < 1 \times 10^{-4}$ and used the UK Biobank as the LD reference to clump at $R^2 < 0.01$ within windows of 5 Mb from each lead SNP. This resulted in 2,480 SNPs. For each of the African, East Asian and Latinx GWAS, we removed the *SLC2A9* and *ABCG2* regions (chr4:9.32 Mb–11.21 Mb and chr4:86.79 Mb–90.23 Mb, respectively) and then took the 2,500 most significant remaining SNPs for each. Summary statistics from meta-analyses without the 23andMe dataset (the four ancestries, each for men only, women only and combined sexes) are available at the GWAS Catalog (https://www.ebi.ac.uk, dataset accessions GCST90428594–GCST90428605; https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90428001-GCST90429000). The 1000 Genomes Project Phase 3 data were downloaded from http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/; information on the UK Biobank cohort can be viewed at https://www.ukbiobank.ac.uk/; the GWAS Catalog is available at https://www.ebi.ac.uk/gwas/; GTEx data were downloaded from https://gtexportal.org/home/datasets; variant information from dbSNP was downloaded from https://ftp.ncbi.nih.gov/snp/latest_release/VCF/GCF_000001405.25.gz; conversion-unstable positions used for SLALOM were downloaded from https://github.com/cathaloruaidh/genomeBuildConversion; GoDMC meQTL data were downloaded from http://mqtldb.godmc.org.uk/downloads; RELI transcription factor ChIP–seq data were from https://tf.cchmc.org/external/RELI/RELI_public_data.tar.bz2; baseline LD score version 1.1, cell type-specific and cell type group annotations were downloaded from https://alkesgroup.broadinstitute.org/LDSCORE/; functional annotations for PAINTOR were downloaded from https://ucla.box.com/s/x47apvgv51au1rlmuat8m4zdjhcniv2d; LDSC summary statistics of UK Biobank traits from the round 2 analysis were downloaded from https://nealelab.github.io/UKBB_ldsc/downloads.html; ABC enhancer–gene connection data were downloaded from ftp://ftp.broadinstitute.org/outgoing/lincRNA/ABC/AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt.gz; FATHMM scores for noncoding variants were obtained from http://fathmm.biocompute.org.uk/; ImmuNexUT eQTL data were downloaded from https://humandbs.biosciencedbc.jp/en/hum0214-v6; OneK1K eQTL data were downloaded from https://onek1k.s3.ap-southeast-2.amazonaws.com/onek1k_eqtl_dataset.zip; CHIP summary statistics were downloaded from https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90102001-GCST90103000/; the Susztak Kidney Biobank is available at https://susztaklab.com/; a list of differentially expressed genes in stimulated monocytes can be obtained from Table S2 in the original paper[128]; GWAS data for white blood cell traits used in the gene prioritization analysis are available at https://www.ebi.ac.uk/gwas/; FANTOM5 TSS data were downloaded from https://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/TSS_classifier/; HaploReg is available at https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php; GeneHancer tracks were accessed through USCS; KEGG, Reactome and GO pathway analyses were conducted at https://david.ncifcrf.gov/; and transcription factor-binding site enrichment was carried out at https://maayanlab.cloud/Enrichr/.

Code availability

Code for the main analyses is available at https://github.com/MerrimanLab/Gout_GWAS_Code, permanently deposited at Zenodo (https://doi.org/10.5281/zenodo.13350995)[129].

References

87. Mackenzie, I. S. et al. Long-term cardiovascular safety of febuxostat compared with allopurinol in patients with gout (FAST): a multicentre, prospective, randomised, open-label, non-inferiority trial. *Lancet* **396**, 1745–1757 (2020).
88. Kvale, M. N. et al. Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* **200**, 1051–1060 (2015).
89. Toyoda, Y. et al. SNP-based heritability estimates of gout and its subtypes determined by genome-wide association studies of clinically defined gout. *Rheumatol.* **62**, e144–e146 (2023).
90. Köttgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
91. Lindström, S. et al. A comprehensive survey of genetic variation in 20,691 subjects from four large cohorts. *PLoS ONE* **12**, e0173997 (2017).
92. Boutin, N. T. et al. The evolution of a large biobank at Mass General Brigham. *J. Pers. Med.* **12**, 1323 (2022).
93. Shin, J. et al. Genetic architecture for susceptibility to gout in the KARE cohort study. *J. Hum. Genet.* **57**, 379–384 (2012).
94. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
95. Neogi, T. et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheumatol.* **67**, 2557–2568 (2015).
96. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
97. Eriksson, N. et al. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
98. Walter, K. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
99. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
100. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
101. Frazer, K. A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
102. Lim, E. T. et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
103. Deng, J.-E. et al. SNPTracker: a swift tool for comprehensive tracking and unifying dbSNP rs IDs and genomic coordinates of massive sequence variants. *G3* **6**, 205–207 (2016).
104. Broad Institute GitHub Repository. *Picard Toolkit* broadinstitute.github.io/picard/ (2019).
105. Willer, C. J. et al. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
106. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
107. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comp. Biol.* **9**, e1003118 (2013).
108. Wang, X. et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **22**, 2303–2311 (2013).
109. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
110. Major, T. J. & Takei, R. LocusZoom-like plots for GWAS results (v2.1). *Zenodo* https://doi.org/10.5281/zenodo.5154379 (2021).
111. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

112. de Leeuw et al. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comp. Biol.* **11**, e1004219 (2015).
113. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
114. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
115. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
116. ter Horst, R. et al. Host and environmental factors influencing individual human cytokine responses. *Cell* **167**, 1111–1124 (2016).
117. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
118. Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
119. Mahajan, A. et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
120. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
121. Ormond, C. et al. Converting single nucleotide variants between genome builds: from cautionary tale to solution. *Brief. Bioinform.* **22**, bbab069 (2021).
122. Chen, W. et al. Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. *Nat. Commun.* **12**, 7117 (2021).
123. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
124. Pirinen, M. *GWAS 3: Statistical Power* www.mv.helsinki.fi/home/mjxpirin/GWAS_course/material/GWAS3.html (2023).
125. Boocock, J. et al. Genomic dissection of 43 serum urate-associated loci provides multiple insights into molecular mechanisms of urate control. *Hum. Mol. Genet.* **29**, 923–943 (2020).
126. Fadason, T. et al. Physical interactions and expression quantitative traits loci identify regulatory connections for obesity and type 2 diabetes associated SNPs. *Front. Genet.* **8**, 150 (2017).
127. Harley, J. B. et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
128. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
129. rikutakei. MerrimanLab/Gout_GWAS_Code: Gout_GWAS_code. *Zenodo* https://doi.org/10.5281/zenodo.13350995 (2024).

Acknowledgements

This article is dedicated to Philip C. Robinson. We thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe Research Team contributed to this study: S. Aslibekyan, A. Auton, E. Babalola, R.K. Bell, J. Bielenberg, K. Bryc, E. Bullis, D. Coker, G. Cuellar Partida, D. Dhamija, S. Das, S.L. Elson, N. Eriksson, T. Filshtein, A. Fitch, K. Fletez-Brant, P. Fontanillas, W. Freyman, J.M. Granka, K. Heilbron, A. Hernandez, B. Hicks, D.A. Hinds, E.M. Jewett, Y. Jiang, K. Kukar, A. Kwong, K.-H. Lin, B.A. Llamas, M. Lowe, J.C. McCreight, M.H. McIntyre, S.J. Micheletti, M.E. Moreno, P. Nandakumar, D.T. Nguyen, E.S. Noblin, J. O'Connell, A.A. Petrakovitz, G.D. Poznik, A. Reynoso, M. Schumacher, A.J. Shastri, J.F. Shelton, J. Shi, S. Shringarpure, Q.J. Su, S.A. Tat, C.T. Tchakouté, V. Tran, J.Y. Tung, X. Wang, W.W., C.H. Weldon, P. Wilton and C.D. Wong. We thank the participants and staff of the UK Biobank study for their important contribution. This research has been conducted using the UK Biobank resource under application number 12611. We acknowledge the participants and investigators of the FinnGen study. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by the NCI, the NHGRI, the NHLBI, the NIDA, the NIMH and the NINDS. This study was supported by grants from the Health Research Council of New Zealand to T.R.M. (08/075, 11/1075, 14/527), Arthritis New Zealand to T.R.M., Lottery Health New Zealand to T.R.M., JSPS KAKENHI (grant numbers 20H00566, 21KK0173, 17H04128, 22H00476, 20K23152, 21H03350, 17015018, 221S0001, 2221S0002, 16H06279 (PAGS), 16H06277, 22H04923 (CoBiA), 20H00568), AMED (JP21gm4010006, JP22km0405211, JP22ek0410075, JP22km0405217, JP22ek0109594), JST Moonshot R&D (JPMJMS2021, JPMJMS2024), the Takeda Science Foundation, the Bioinformatics Initiative of Osaka University Graduate School of Medicine, the Gout and Uric Acid Foundation of Japan and the Ministry of Defense of Japan, the National Natural Science Foundation of China (82220108015) and the National Key Research and Development Program of China (2022YFE0107600, 2022YFC2503300). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006) and is currently supported by the Wellcome Trust (216767/Z/19/Z). Genotyping of the GS:SFHS samples was funded by the Medical Research Council UK, the Wellcome Trust (Wellcome Trust Strategic Award 'Stratifying Resilience and Depression Longitudinally' (STRADL), reference 104036/Z/14/Z), and analysis was supported by MRC University Unit core grant MC_UU_00007/10 (QTL in Health and Disease program). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

T.J.M., R. Takei, M.P.L., N.A.S., R.K.T. and T.R.M. wrote the manuscript. T.J.M., R.T., H.M., M.P.L., N.A.S., R.K.T., W.-H.W., H.K.C., E.A.S., A.S., Y.O., C. Li, Y. Shi and T.R.M. designed the study. H.M., T.M.M., W.W., H.K.C., J.N.M., D.H.S., E.M.J., C. Lee, C. Li and T.R.M. managed an individual contributing study. H.M., M.P.L., W.W., M.J.C., A.J.P.-G., E.M., J.M.O'S., L.K.S., N.D., A.A., M.D., E.R., L.T.H.J., M.A., R.J.T., T.R., T.L.J., M.J., L.A.B.J., R. Toes, F.L., H.K.E., T.P., G.M.M., B.S., T.U., V.V., P.L.R., S.H.R., C. Lee, P.C.R., J.C., H.L., K.S. and S. Shringarpure critically reviewed the manuscript. T.J.M., R. Takei, M.P.L., N.A.S., R.K.T., Y. Shirai, W.W., M.J.C., A.J.P.-G., Z.L., A.J., M.E.M., E.M., E.E.K., W.-H.W., M.J.B., R.J.R., T.F., E.G., J.M.O'S., L.A.B.J., R.L., O.I.G., T.O.C., V.V., T.S.B., C.H., S.H.R., A.C., A.N., Y.K., Y.T., C. Lee, M.A.B., E.A.S., J.C., K.M.G., D.J.B., E.M.J., H.L., K.S., S. Shringarpure, Y.O., Y. Shi and T.R.M. carried out data acquisition, statistical methods and bioinformatic analyses. H.M., M.E.M., S.P.A.M., R.J.R., K.G.S., L.K.S., N.D., A.A., M.D., E.R., L.T.H.J., M.C.K., O.M., M.A., F.P.-R., R.J.T., T.R., T.L.J., M.J., T.O.C., S.R., F.K., T.W.J.H., R. Toes, F.L., P.R., T.B., H.K.E., T.P., G.M.M., L.H., B.S., A.-K.T., T.U., S.H.R., T.M.M., T.T., M.N., S. Shimizu, H.N., K.Y., K.M., N.S., K.I., C. Lee, L.A.B., M.A.B., P.C.R., R.R.C.B., C.L.H., S.L., M.D.S., M.R., J.N.M., D.H.S., E.M.J., K.S., A.S., Y.O., C. Li and T.R.M. recruited participants for the study.

Competing interests

W.W. and S. Shringarpure are employed by and hold stock or stock options in 23andMe. All other authors declare no competing interests.

Additional information

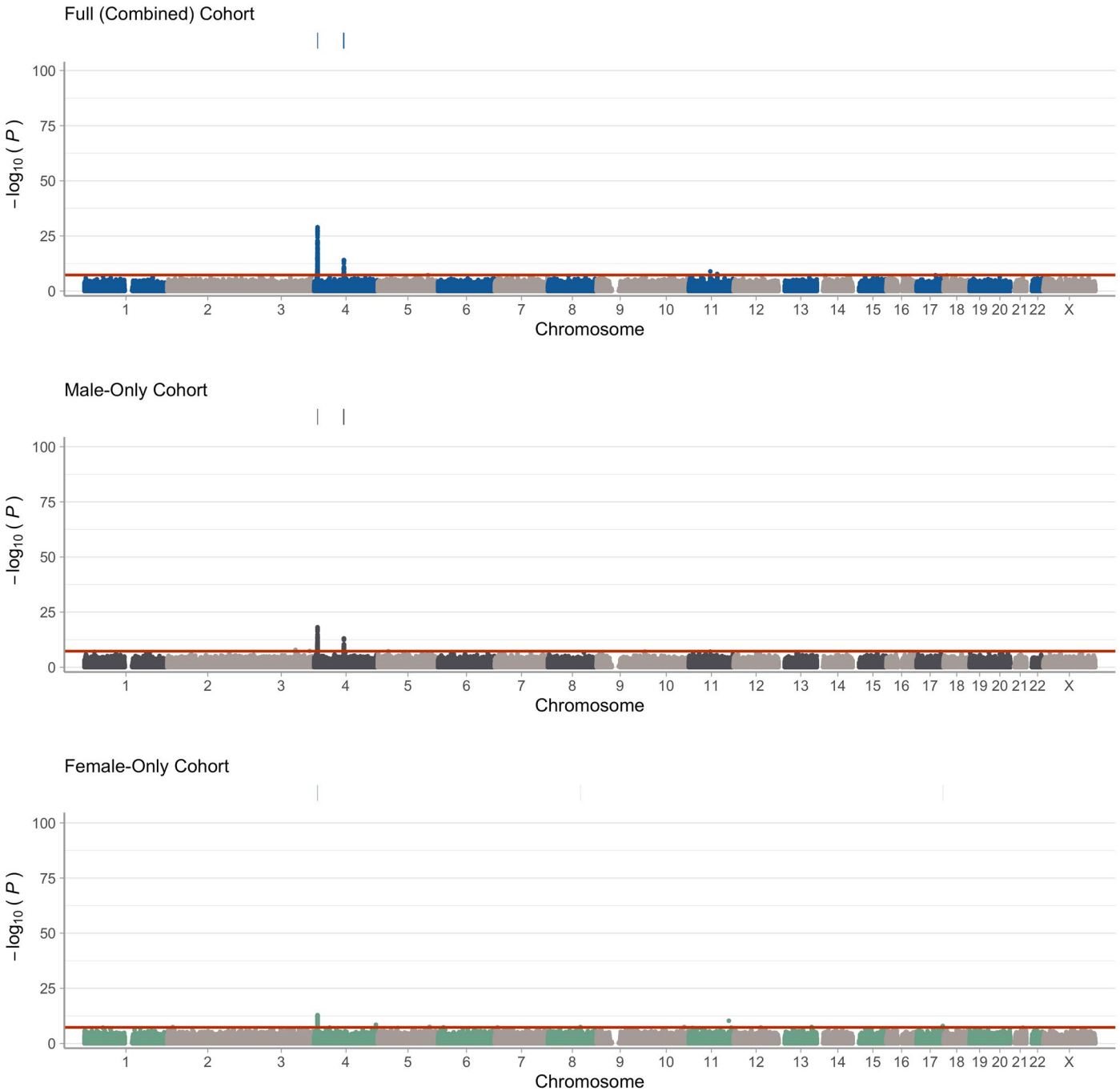
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-01921-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01921-5.

Correspondence and requests for materials should be addressed to Tony R. Merriman.

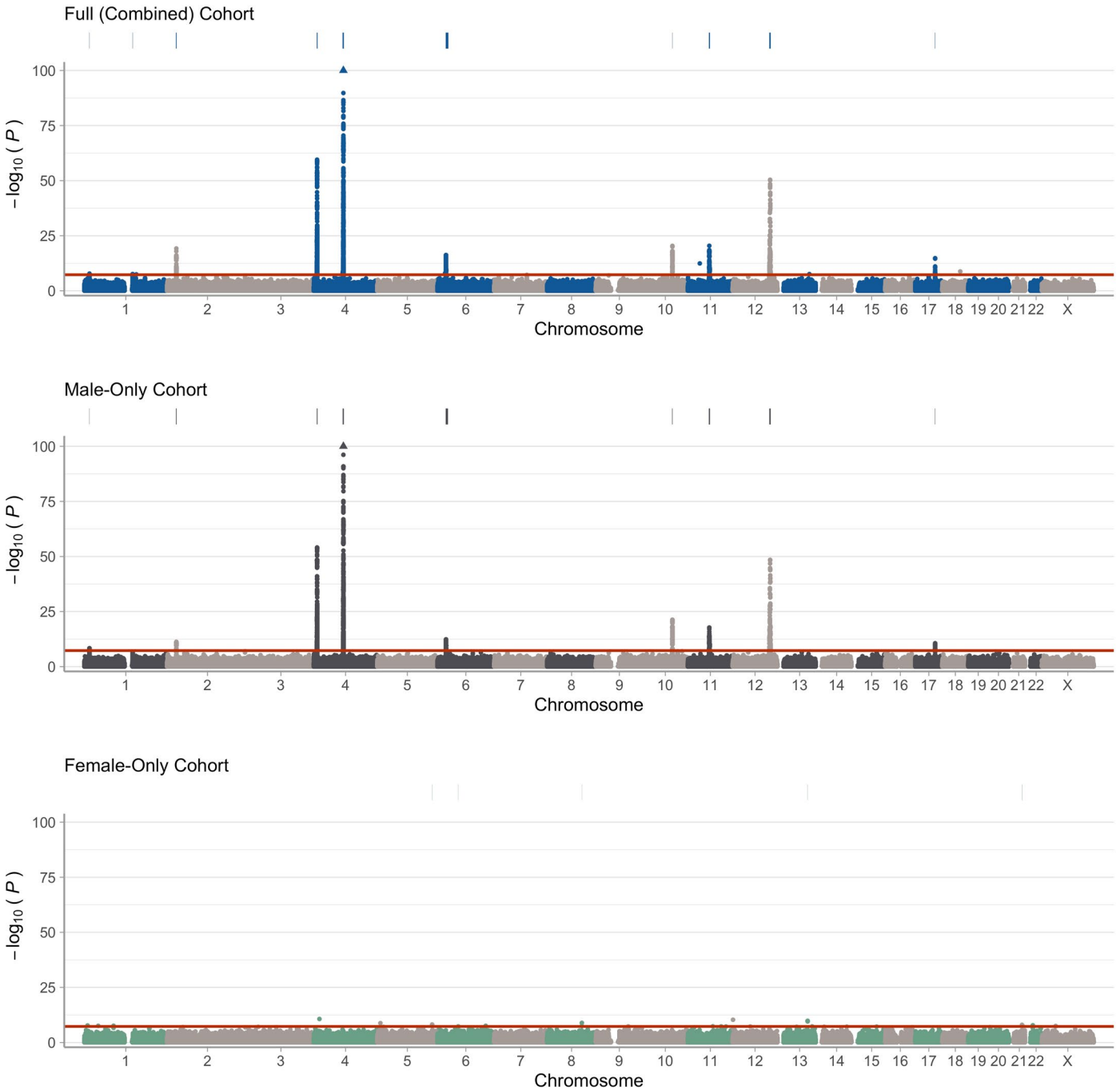
Peer review information *Nature Genetics* thanks Patrick Sulem and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-024-01921-5



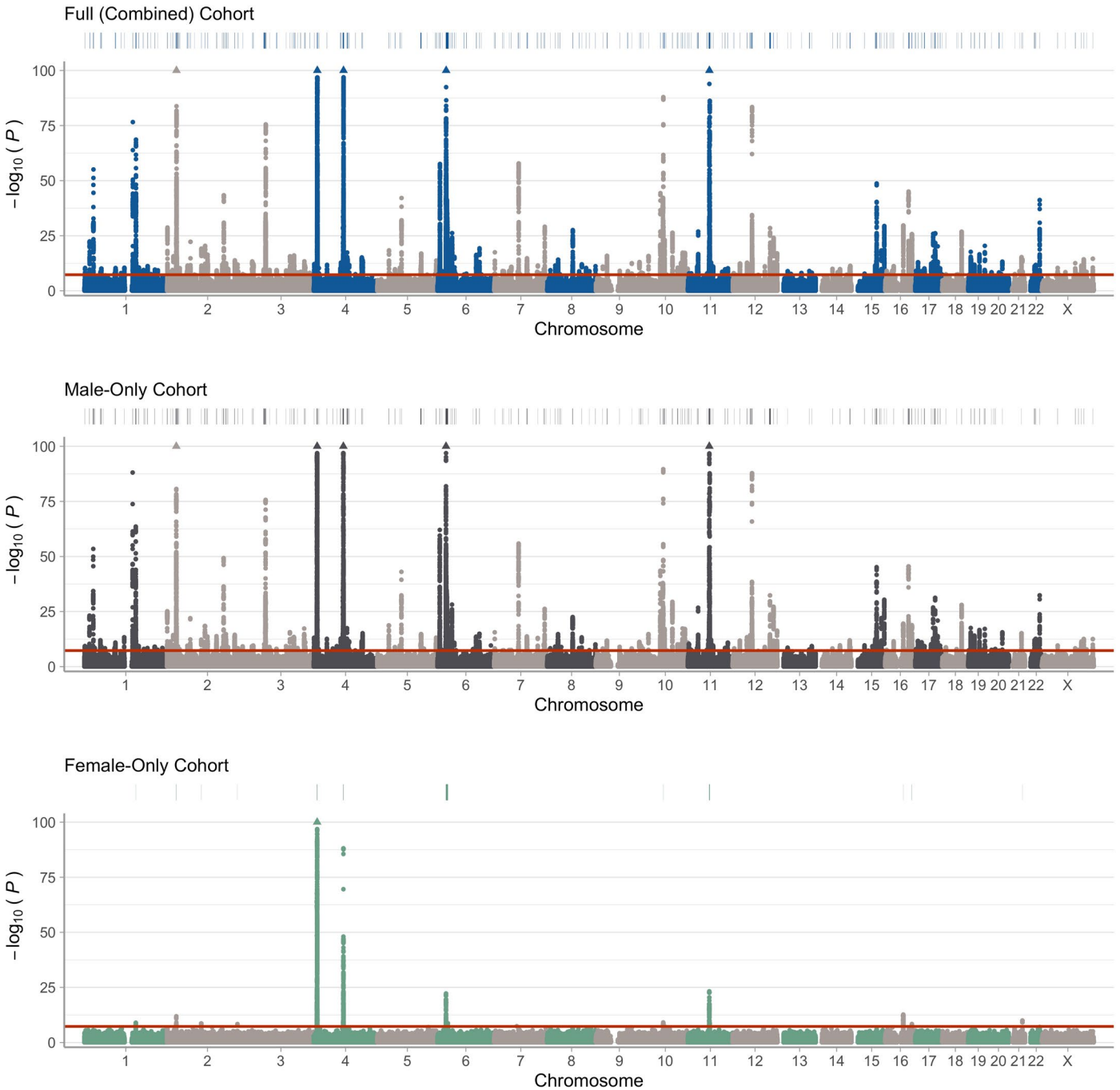
Extended Data Fig. 1 | Manhattan plots of full (combined sexes), male, and female African ancestry analysis. The bars at the top of the Manhattan plots indicate where the genome-wide significant variants ($P \leq 5 \times 10^{-8}$) are located. The red horizontal line indicates the $P = 5 \times 10^{-8}$ significance threshold.

https://doi.org/10.1038/s41588-024-01921-5



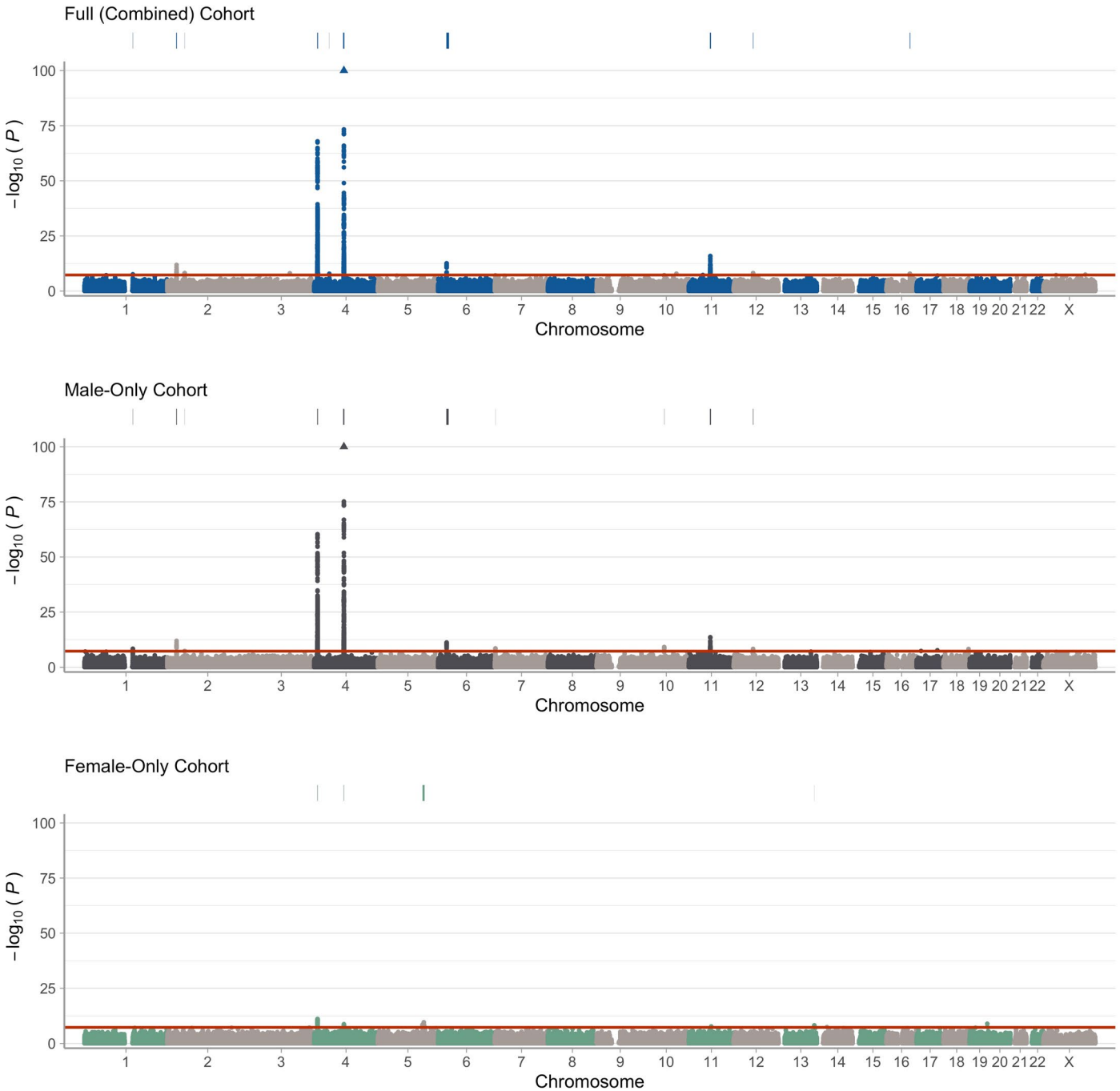
Extended Data Fig. 2 | Manhattan plots of full (combined sexes), male, and female East Asian ancestry analysis. Manhattan plots show variants with $-\log_{10}P \leq 100$ (loci that contain variants with $-\log_{10}P > 100$ are indicated with a triangle mark). The bars at the top of the Manhattan plots indicate where the genome-wide significant variants ($P \leq 5 \times 10^{-8}$) are located. The red horizontal line indicates the $P = 5 \times 10^{-8}$ significance threshold.

https://doi.org/10.1038/s41588-024-01921-5



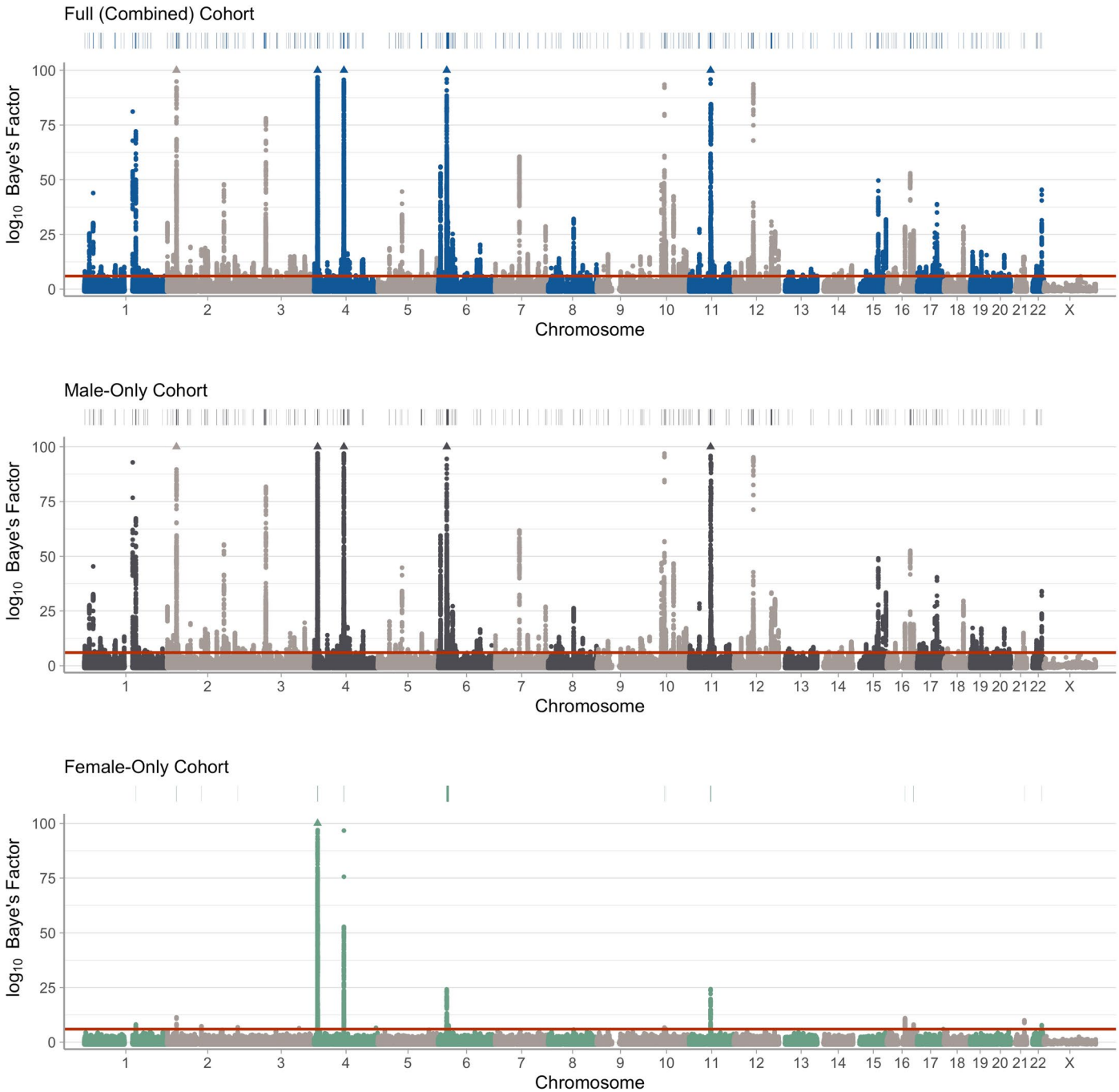
Extended Data Fig. 3 | Manhattan plots of full (combined sexes), male, and female European ancestry analysis. Manhattan plots show variants with $-\log_{10}P \leq 100$ (loci that contain variants with $-\log_{10}P > 100$ are indicated with a triangle mark). The bars at the top of the Manhattan plots indicate where the genome-wide significant variants ($P \leq 5 \times 10^{-8}$) are located. The red horizontal line indicates the $P = 5 \times 10^{-8}$ significance threshold.

https://doi.org/10.1038/s41588-024-01921-5



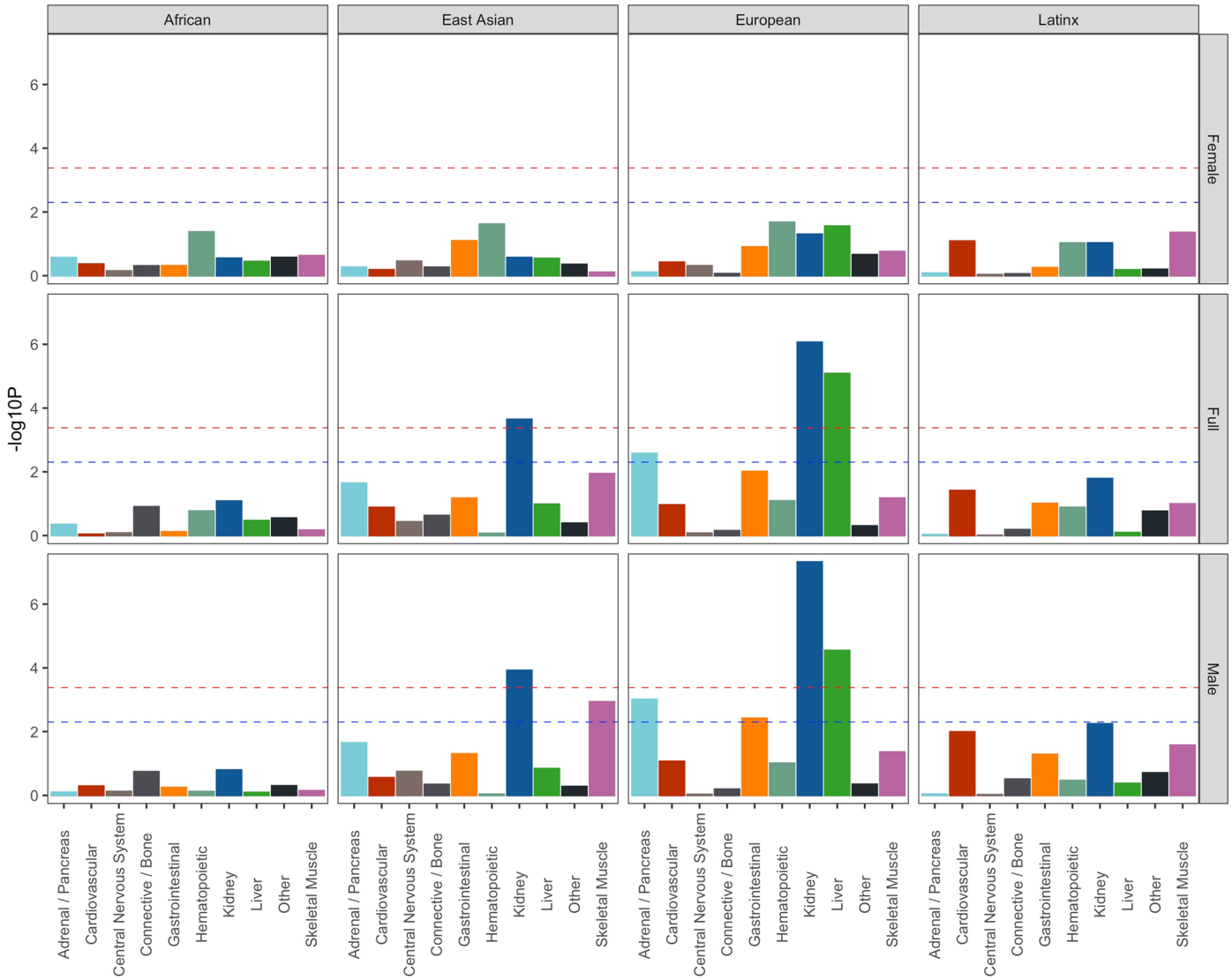
Extended Data Fig. 4 | Manhattan plots of full (combined sexes), male, and female Latinx ancestry analysis. Manhattan plots show variants with $-\log_{10}P \leq 100$ (loci that contain variants with $-\log_{10}P > 100$ are indicated with a triangle mark). The bars at the top of the Manhattan plots indicate where the genome-wide significant variants ($P \leq 5 \times 10^{-8}$) are located.

https://doi.org/10.1038/s41588-024-01921-5



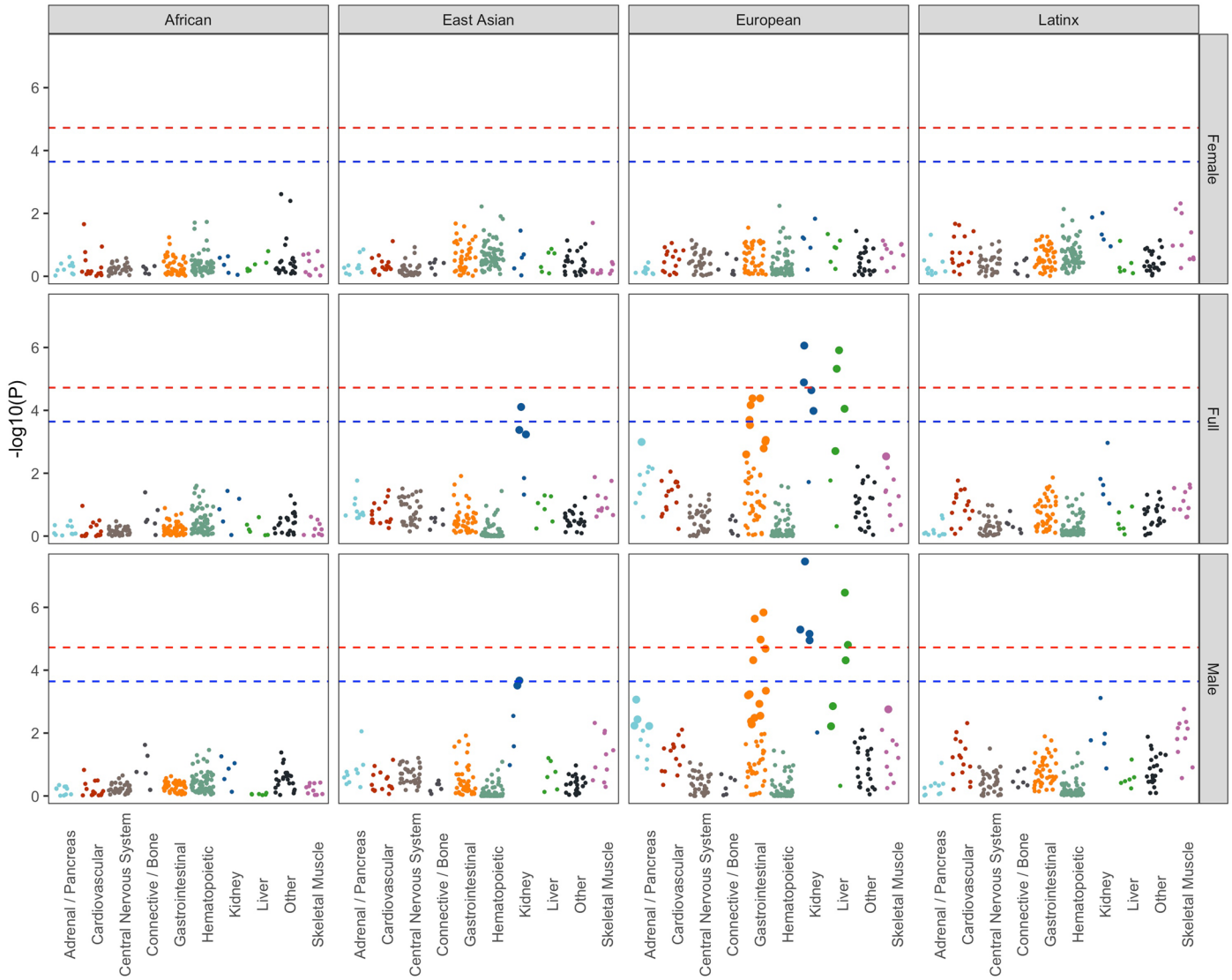
Extended Data Fig. 5 | Manhattan plots of full (combined sexes), male, and female trans-ancestry meta-analysis. Manhattan plots show variants with $\log_{10}BF \leq 100$ (loci that contain variants with $\log_{10}BF > 100$ are indicated with a triangle mark). The bars at the top of the Manhattan plots indicate where the genome-wide significant variants ($\log_{10}BF \geq 6$) are located. The red horizontal line indicates the $\log_{10}BF = 6$ significance threshold.

https://doi.org/10.1038/s41588-024-01921-5



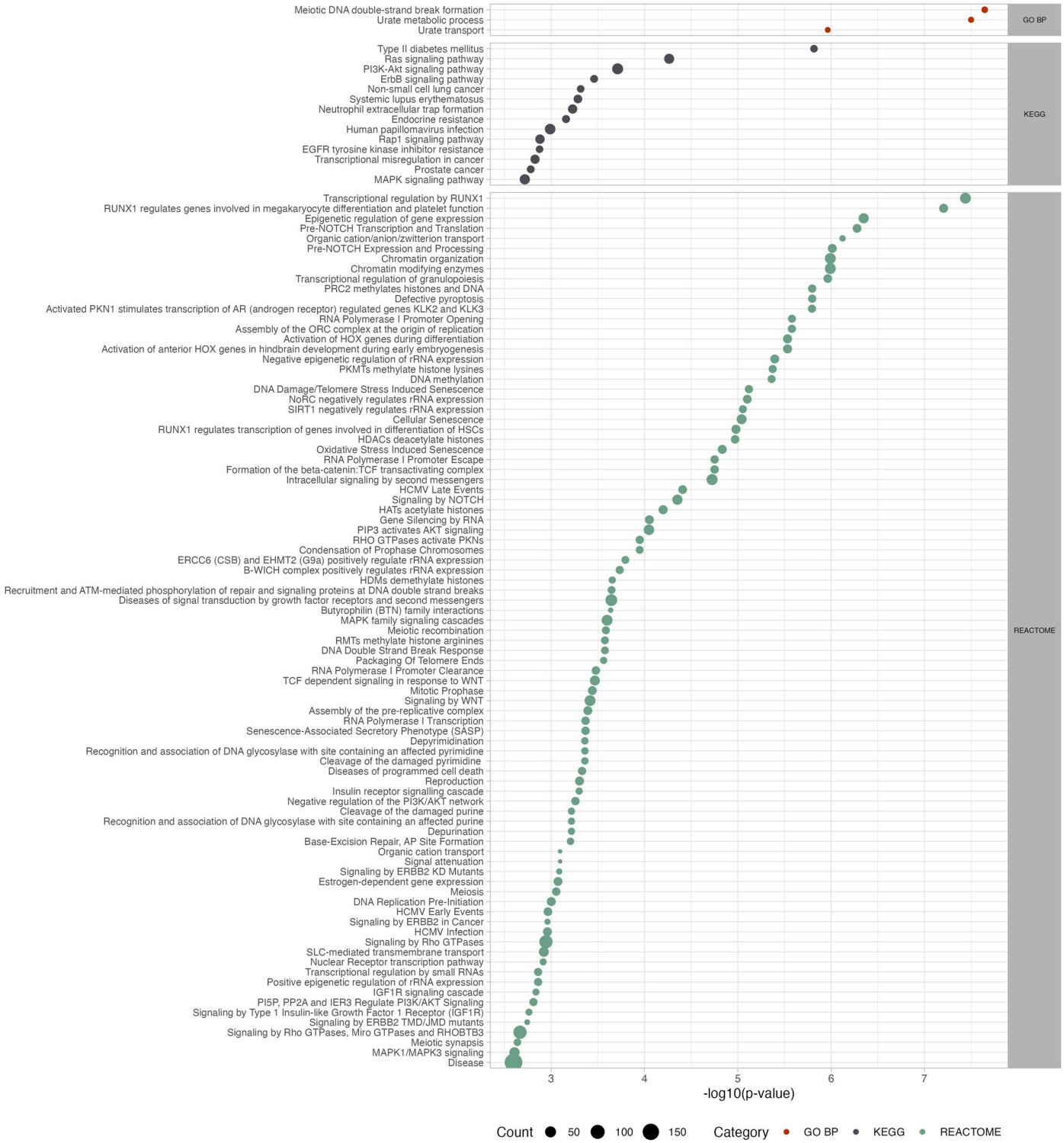
Extended Data Fig. 6 | Bar plot of cell-type group enrichment of African, East Asian, European, and Latino gout GWAS for full, male-, and female-specific analyses. Red dashed line: conservative p-value threshold of 4.2×10^{-4}; Blue dashed line: p-value threshold of 5.0×10^{-3}.

https://doi.org/10.1038/s41588-024-01921-5



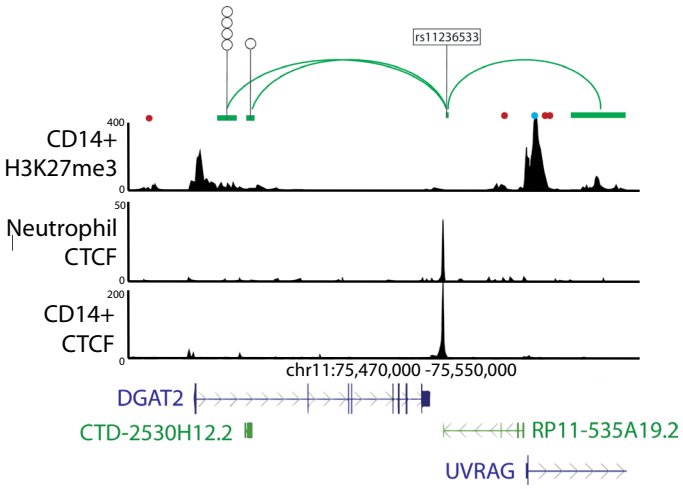
Extended Data Fig. 7 | Bar plot of cell-type specific enrichment of African, East Asian, European, and Latino gout GWAS for full, male-, female-specific analyses. Red dashed line: conservative p-value threshold of 1.9×10^{-5}; Blue dashed line: p-value threshold of 2.2×10^{-4}. Large dots: cell types with FDR-adjusted p-value ≤ 0.05.

https://doi.org/10.1038/s41588-024-01921-5



Extended Data Fig. 8 | Functional and pathway enrichment analyses of gout candidate genes. The DAVID database was used to identify GO Biological Function term, KEGG and REACTOME pathways enriched in the gout GWAS dataset. Significance (FDR) of the enrichment is denoted on the x-axis, size of the circle denotes number of genes contributing to the enrichment term.

https://doi.org/10.1038/s41588-024-01921-5



Extended Data Fig. 9 | DGAT2: Example of genome organization at a candidate immune-priming lncRNA locus. ENCODE H3K4me3 signal track from CD14+ monocytes indicates enrichment at the promoters of DGAT2, the lncRNA RP11-535A19.2 and UVRAG. ENCODE CTCF signal track from neutrophils and CD14+ monocytes indicates CTCF binding at rs11236533. Genehancer connections are in green and illustrate physical connections (Hi-C) between rs11236533 which disrupts a CTCF site and additional maximally-associated SNPs at two Genehancer regulatory elements. Red and blue dots indicate the CpG locations that are associated with co-localized meQTL and the color denotes direction of effect of the gout risk allele (red = higher methylation, blue = lower methylation). White lollipops represent maximally-associated SNPs at the locus within predicted *cis* regulatory elements.

nature portfolio

Corresponding author(s): Tony Merriman

Last updated by author(s): Jul 23, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☒	☐	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This is described in the manuscript (page 37)
Data analysis	This is described in the manuscript (pp 50-52). Data collection/data analysis software/tools/algorithms/packages used are PLINKv1.9, GATK, FINEMAP, PAINTOR, LD Score Regression, cov-LD Score Regression, SLALOM, CoDeS3D, GCTA-COJO, GREP. Custom codes are available at https://github.com/MerrimanLab/Gout_GWAS_Code and https://github.com/Geeketics/LocusZooms.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Datasets will be made available at no cost for academic use (https://research.23andme.com/

collaborate/#dataset-access/). Table S40 contains summary statistics from 9,980 independently-associated SNPs. For the full European GWAS, we took all SNPs at P<1x10-4 and used UK Biobank as LD reference to clump at R2<0.01 within windows of 5 Mb from each lead SNP. This resulted in 2,480 SNPs. For each of the African, East Asian and Latinx GWAS, we removed the SLC2A9 and ABCG2 regions (chr4:9.32Mb-11.21Mb and chr4:86.79Mb-90.23Mb, respectively) then took the 2,500 most significant remaining SNPs for each. Summary statistics from meta-analyses without the 23andMe data set are available at GWAS Catalog (https://www.ebi.ac.uk; ID: GCP000871).
1000 Genomes Project Phase 3 data were downloaded from http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/; information on UK Biobank cohort can be viewed at https://www.ukbiobank.ac.uk/; GWAS Catalog (https://www.ebi.ac.uk/gwas/); GTEx data were downloaded from https://gtexportal.org/home/ datasets; variant information from dbSNP was downloaded from https://ftp.ncbi.nih.gov/ snp/latest_release/VCF/GCF_000001405.25.gz; CUPs used for SLALOM were downloaded from https://github.com/cathaloruaidh/genomeBuildConversion; GoDMC methylation QTL data were downloaded from http://mqtldb.godmc.org.uk/downloads; RELI transcription factor ChIP-seq data https://tf.cchmc.org/external/RELI/RELI_public_data.tar.bz2; baseline LD score v1.1, cell-type specific, and cell-type group annotations were downloaded from https://alkesgroup.broadinstitute.org/LDSCORE/; functional annotations for PAINTOR were downloaded from https://ucla.box.com/s/x47apvgv51au1rlmuat8m4zdjhcniv2d; LD score regression summary statistics of UK Biobank traits from round 2 analysis were downloaded from https://nealelab.github.io/UKBB_ldsc/downloads.html; ABC enhancer-gene connection data were downloaded from ftp://ftp.broadinstitute.org/outgoing/lincRNA/ABC/ AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt.gz); FATHMM scores for non-coding variants were obtained from http://fathmm.biocompute.org.uk/; ImmuNexUT eQTL data were downloaded from https://humandbs.biosciencedbc.jp/en/hum0214-v6; OneK1K eQTL data were downloaded from https://onek1k.s3.ap-southeast-2.amazonaws.com/onek1k_eqtl_dataset.zip; CHIP summary statistics were downloaded from https://ftp.ebi.ac.uk/pub/databases/gwas/ summary_statistics/GCST90102001-GCST90103000/; Susztak Kidney Biobank (https://susztaklab.com/); a list of differentially expressed genes in stimulated monocytes can be obtained from Table S2 in the original paper133; GWAS data for white blood cell traits used in gene prioritization analysis (https://www.ebi.ac.uk/gwas/); FANTOM5 TSS data were downloaded from https://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/ TSS_classifier/; HaploReg is available at https://pubs.broadinstitute.org/mammals/haploreg/ haploreg.php; GeneHancer tracks were accessed through USCS; KEGG, REACTOME and GO pathway analyses were conducted in https://david.ncifcrf.gov/; and transcription factor binding site enrichment was carried out in https://maayanlab.cloud/Enrichr/.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Combined and sex-specific analyses have been reported.
Reporting on race, ethnicity, or other socially relevant groupings	Ancestral-specific analyses were carried out (African, East Asian, European, Latinx).
Population characteristics	Sex, age and gout status were reported. No sample size prior calculations were performed. Sample size was determined by the combined size of the various cohorts included in the study. A final sample size of 120,295 people with gout and 2,503,807 without gout was deemed sufficient on the basis that a sample of this size is widely accepted as being sufficiently powered to detect loci of weak effect that confer risk to a complex condition.
Recruitment	Recruitment of participants is described in the publications cited on page 37.
Ethics oversight	For the 23andMe Inc. cohort, participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services (www.eandireview.com). Ethical & Independent Review Services was recently acquired, and its new name as of July 2022 is Salus IRB (https://www.versiticlinicaltrials.org/salusirb). The UK Biobank (UKBB) was undertaken with ethical approval from the North West Multi-Centre Research Ethics Committee of the UK. GlobalGout obtained ethical approval from the following committees: the New Zealand Multiregional Ethics Committee (MEC05/10/130); the Northern Y Region Health Research Ethics Committee (Ngāti Porou Hauora Charitable Trust study; NTY07/07/074); Research and Ethics Committee, Repatriation General Hospital, South Australia (32/08); Research Ethics Committee, University of New South Wales; Ethikkommission, Technische Universität Dresden (EK 8012012); South East Scotland Research Ethics Committee (04/S1102/41); Commission Cantonale (VD) D'éthique de la Recherche sur l'être Humain, Université de Lausanne; Commissie Mensgebonden Onderzoek regio Arnhem—Nijmegen; Partners Health Care System Institutional Review Board. The Institutional Review Board of the Kaiser Foundation Research Institute provided ethical approval for the Kaiser Permanente sample set. The FinnGen study was approved by the Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa. The ethics review board at the Affiliated Hospital of Qingdao University approved the study in China. In Japan ethical approvals were provided by the institutional ethical committees of the National Defense Medical College, Nagoya University and RIKEN. The Korean Association Resource (KARE) was approved by the institutional review board of the Korea National Institute of Health. The FAST study and Generation Scotland was ethically approved by the UK Multi-Centre Research Ethics Committee (Reference number: 2011-001883-23) and the NHS Tayside Committee on Medical Research Ethics (REC Reference Number: 05/S1401/89).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes are disclosed in Table S1.
Data exclusions	Cohort-specific exclusions are described in the cohort-specific citations on page 37. Controls with hyperuricemia, diabetes, cancer and other arthritis-related illnesses were excluded from the China Gout cohort in order to increase power to detect gout-related loci. Women were excluded from the Japan Gout / Japan Mulit-Institutional Collaborative Cohort given the very low prevalence of gout in women in Japan. Patients with a history of myocardial infarction or stroke in the previous 6 months and those with congestive heart failure (New York Heart Association [NYHA] class III or IV) or severe renal impairment were excluded from the FAST trial according to the protocol of this RCT from which people with gout were sourced for this study. There were no pre-determined exclusion criteria for this study.
Replication	The study design did not include replication. The reason is that all available samples were included in the main discovery GWAS in order to maximize power.
Randomization	There was no randomization. Sex and age, as appropriate were includes in logistic regression analyses as co-variables.
Blinding	Blinding was not relevant to this study as the cohorts used were already separately genome-wide scanned for association with gout.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*