Genome-wide association analyses identify 18 new loci associated with serum urate concentrations

Elevated serum urate concentrations can cause gout, a prevalent and painful inflammatory arthritis. By combining data from >140,000 individuals of European ancestry within the Global Urate Genetics Consortium (GUGC), we identified and replicated 28 genome-wide significant loci in association with serum urate concentrations (18 new regions in or near TRIM46, INHBB, SFMBT1, TMEM171, VEGFA, BAZ1B, PRKAG2, STC1, HNF4G, A1CF, ATXN2, UBE2Q2, IGF1R, Nfat5, MAF, HLF, ACVR1B-ACVR1L and B3GNT4). Associations for many of the loci were of similar magnitude in individuals of non-European ancestry.

We further characterized these loci for associations with gout, transcript expression and the fractional excretion of urate. Network analyses implicate the inhibin-activin signaling pathways and glucose metabolism in systemic urate control. New candidate genes for serum urate concentration highlight the importance of metabolic control of urate production and excretion, which may have implications for the treatment and prevention of gout.

Uric acid is a final breakdown product of purine oxidation in humans and is present in the blood as urate. Elevated concentrations of serum urate—hyperuricemia—can cause gout\(^1\). Gout is the most prevalent inflammatory arthritis in developed countries, with an estimated 8.3 million US adults in 2007–2008 having had at least one of the extremely painful attacks\(^2\). Prevalence is increasing, owing in part to population aging, dietary and lifestyle factors, and rising levels of obesity and insulin resistance\(^3–5\). Chronic gout inflicts a considerable social and economic burden resulting from the associated pain and disability\(^6,7\) as well as reduced work-related activity and productivity\(^8\).

Blood urate concentrations are determined by a balance between uric acid production, primarily in the liver, and its disposal via the kidney and gut. High circulating levels are mainly due to the net reabsorption of 90% of filtered urate in the renal proximal tubules\(^9\). Understanding the control of uric acid homeostasis is critical to improving the management and treatment of patients with hyperuricemia and gout.

The heritability of serum urate concentrations is estimated at 40–70% (refs. 10–12), which justifies the search for its genetic determinants. Previous genome-wide association studies (GWAS) have so far identified 11 genomic loci associated with urate concentrations and gout\(^13–18\). Together, SNPs at these loci explained about 5–6% of variance in serum urate concentrations\(^17,18\), suggesting that additional loci remain to be identified. We therefore aimed to identify and validate variants associated with serum urate concentrations in >140,000 individuals of European ancestry and with gout in approximately 70,000 individuals in the GUGC. These variants may provide a basis for the identification of new potential therapeutic targets for gout\(^19\).

RESULTS

Characteristics of the study samples

An overview of the characteristics of the GUGC study samples is provided in Supplementary Table 1. Our study comprises 110,347 individuals from 48 studies contributing to the discovery GWAS meta-analysis of serum urate concentrations and participants of 14 studies contributing to the meta-analysis of gout (2,115 cases and 67,259 controls). Across studies, mean serum urate concentrations ranged from 3.9 to 6.1 mg/dl (median of 5.2 mg/dl), and the proportion with gout ranged from 0.9% to 6.4% (median of 3.3%). Detailed information about the GUGC studies is provided in Supplementary Table 2. Information about study-specific genotyping, imputation, analysis tools and inflation factors for the individual serum urate concentration GWAS (range of 0.98–1.05) are shown in Supplementary Table 3.

Loci associated with serum urate concentrations

A quantile-quantile plot for the 2,450,547 investigated autosomal SNPs in the discovery GWAS meta-analysis showed many more SNPs with low observed P values than expected, even after excluding SNPs in known urate concentration–associated regions (Supplementary Fig. 1). All 2,201 SNPs associated with serum urate concentrations at P < 5 x 10\(^{-8}\) in the discovery stage are listed in Supplementary Table 4. Overall, 37 different genomic loci were identified that contained SNPs associated with serum urate concentrations at P < 1 x 10\(^{-6}\); 26 of these were associated at genome-wide significance (P < 5 x 10\(^{-8}\), 10 known and 16 new loci; Fig. 1). Regional association plots provide a detailed overview of the associated regions (Supplementary Fig. 2). To assess the presence of independently associated SNPs within the regions associated at genome-wide significance and suggestive levels of significance (5 x 10\(^{-8}\) < P < 1 x 10\(^{-6}\)), conditional analyses were carried out (Online Methods). This analysis identified only the SLC22A11-NRXN2 region as containing more than one independent signal (Supplementary Table 5). The association at the previously reported LRRG16A locus\(^17\) was not independent from that at the SLC17A1 locus in our study. For each of the 37 independent loci, the index SNP with the lowest P value was carried forward for replication.

Replication was attempted in up to 32,813 new study participants. Successful replication was defined as q value < 0.05 in the independent replication sample and genome-wide significance in the
combined sample ($P < 5 \times 10^{-8}$; Online Methods). Summary results for the 26 replicated urate concentration–associated loci are shown (Table 1): there were 16 newly identified regions in or near TRIM46, INHBB, SFMBT1, TMEM171, VEGFA, BAZ1B, PRKAG2, STC1, HNF4G, A1CF, ATXN2, UBE2Q2, IGF1R, NFAT5, MAF and HLF in addition to 10 known urate concentration–associated regions in or near PDZK1, GCKR, SLC2A9, ABCG2, RREB1, SLC17A1, SLC16A9, SLC22A11, NRXN2 and INHBC. The proportion of variance in serum urate concentrations explained by these replicated loci in our data was 7.0%, with 5.2% of this estimate explained by the ten previously known loci and 3.4% explained by SLC2A9 and ABCG2 alone. The moderate proportion of trait variance explained is in line with findings from GWAS of other phenotypes.$^{20}$ The proportion of the age- and sex-adjusted variance in serum urate concentrations explained by all common SNPs rather than just the genome-wide significant index SNPs ranged from 0.27 (standard error (s.e.) of 0.04, ARIC Study) to 0.41 (s.e. of 0.07, SHIP Study) (Online Methods).

Additionally, regions in or near ORC4L, OVOL1 and BCA3 were associated at genome-wide significance but had replication $q$ values of $\geq 0.05$, whereas the QRICH2 locus had a replication $q$ value of $< 0.05$ but did not reach genome-wide significance in the combined sample. Effect estimates, allele frequencies and heterogeneity measures are presented separately for the discovery, replication and combined analyses for all 37 loci tested for replication (Supplementary Table 6). Although the functions of genes within the associated genomic regions were frequently connected to urate transport for the known loci, newly associated regions contained a series of transcription and growth factors encoding genes potentially connected to the metabolic control of serum urate production and excretion (Supplementary Table 7). Ten of the replicated loci also showed significant association with other complex traits, as listed in the National Human Genome Research Institute (NHGRI) GWAS Catalog,$^{21}$ and VEGFA and ATXN2 showed significant associations with ratios of serum metabolites containing a $\gamma$-glutamyl amino acid from a published metabolite-SNP association resource.$^{22}$

**Secondary analyses**

Three secondary analyses were conducted to identify additional urate concentration–associated SNPs. First, as rare monogenic syndromes featuring gout can be caused by X-chromosome mutations in PRPS1 (MIM 300661) and HPRT1 (MIM 300322), we queried 54,926 X-chromosomal SNPs for association with serum urate concentrations in a meta-analysis of 72,026 participants from 25 of the GUGC studies (Supplementary Note). No genome-wide significant associations were observed, either in the combined or the sex-specific analyses (Supplementary Fig. 3); all X-chromosomal SNPs associated with serum urate concentrations at $P < 1 \times 10^{-4}$ are listed (Supplementary Table 8).

Second, because of the higher prevalence of gout in men and the known sex-related differences in the effects of urate concentration–associated variants in SLC2A9 and ABCG2 (refs. 14, 16), we conducted meta-analyses of GWAS separately for 49,825 men and 60,522 women. In addition to SNPs at loci detected in the combined analysis, SNPs were associated with serum urate concentrations at $P < 1 \times 10^{-6}$ in one region in men and in five regions in women. None of these SNPs was significantly associated with serum urate concentrations in the replication samples (Supplementary Table 6), although the index SNP at HNF1A in women can be considered borderline significant with an association at $P = 8.1 \times 10^{-8}$ in the combined discovery and replication data set. Besides SLC2A9 and ABCG2, no additional regions contained SNPs that differed significantly ($P < 5 \times 10^{-8}$) in their association effect sizes between men and women. Sex-related differences for genome-wide significant and suggestive SNPs of the overall and sex-stratified analyses are shown (Supplementary Table 9).

Third, we conducted a search for associated SNPs in genes that are family members of known urate transporter genes and, to the best of our knowledge, had not yet been connected to urate transport in humans. After correcting for multiple testing (Supplementary Note), SNPs in the SLC22A7 region (SLC22A11- SLC22A12 family member) showed significant association with serum urate concentrations in the discovery ($P = 1.9 \times 10^{-3}$) and replication ($q$ value $< 0.05$) samples but did not reach the stringent genome-wide significance level in the combined samples (Table 2). It was recently found that the organic anion transporter 2 (OAT2), encoded by SLC22A7, transports urate in HEK293 cells.$^{23}$ Our findings show that variation in SLC22A7 has a measurable effect on serum urate concentrations in humans, supporting the idea that it is a genuine urate concentration–associated GWAS locus of small effect size. Regional association plots for loci

---

**Figure 1** Multiple genomic loci contain SNPs associated with serum urate concentrations. Truncated Manhattan plot showing $-\log_{10}(P$ values) for all SNPs of the urate discovery GWAS ordered by chromosomal position. The gene closest to the SNP with the lowest $P$ value at each locus (index SNP) is listed. Loci in gray met one but not both replication criteria. Blue triangles represent loci containing SNPs with $P$ values below $1 \times 10^{-25}$.
implicated through sex-specific analyses or the candidate gene approach are also shown (Supplementary Fig. 2).

**Gout GWAS**

We conducted a genome-wide gout discovery meta-analysis including 2,538,056 autosomal SNPs. Two previously reported loci had associations that reached genome-wide significance in the overall and sex-stratified analyses: ABCG2 (rs1481012, \(P = 2.0 \times 10^{-12}\)) and SLC2A9 (rs4475146, \(P = 4.1 \times 10^{-26}\) (Supplementary Fig. 4 and Supplementary Table 10). The quantile-quantile plot of the \(P\) values from the combined analysis showed more observed low \(P\) values than expected by chance, even after excluding the previously known urate concentration–associated regions (Supplementary Fig. 5). The index SNPs in SLC2A9 and ABCG2 from the combined analysis of gout were in high linkage disequilibrium (LD; \(r^2 > 0.9\)) with the index SNPs from the serum urate concentration meta-analysis at these loci.

**Serum urate concentration–associated loci and gout**

Because elevated serum urate levels are a key risk factor for the development of gout, we investigated the association of the urate

### Table 1 SNPs associated with serum urate concentrations in individuals of European ancestry

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>bp (Build 36)</th>
<th>Closest gene</th>
<th>GRAIL gene</th>
<th>A1</th>
<th>A2</th>
<th>Freq. A1</th>
<th>Effect</th>
<th>s.e.</th>
<th>(P) value</th>
<th>OR</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1471633</td>
<td>1</td>
<td>144435096</td>
<td>PDZK1</td>
<td>PDZK1</td>
<td>A</td>
<td>C</td>
<td>0.46</td>
<td>0.059</td>
<td>0.005</td>
<td>(1.2 \times 10^{-29})</td>
<td>1.03</td>
<td>(2.8 \times 10^{-1})</td>
</tr>
<tr>
<td>rs2160326</td>
<td>2</td>
<td>27584444</td>
<td>GCKR</td>
<td>GCKR</td>
<td>T</td>
<td>C</td>
<td>0.41</td>
<td>0.074</td>
<td>0.005</td>
<td>(1.2 \times 10^{-44})</td>
<td>1.14</td>
<td>(8.2 \times 10^{-6})</td>
</tr>
<tr>
<td>rs21498742</td>
<td>4</td>
<td>9553150</td>
<td>SLC2A9</td>
<td>SLC2A9</td>
<td>A</td>
<td>G</td>
<td>0.77</td>
<td>0.373</td>
<td>0.006</td>
<td>(0)</td>
<td>1.56</td>
<td>(1.9 \times 10^{-31})</td>
</tr>
<tr>
<td>rs2231142</td>
<td>4</td>
<td>89271347</td>
<td>ABCG2</td>
<td>ABCG2</td>
<td>T</td>
<td>G</td>
<td>0.11</td>
<td>0.217</td>
<td>0.009</td>
<td>(1.0 \times 10^{-13})</td>
<td>1.73</td>
<td>(1.7 \times 10^{-39})</td>
</tr>
<tr>
<td>rs675209</td>
<td>6</td>
<td>7047083</td>
<td>RREB1</td>
<td>RREB1</td>
<td>T</td>
<td>C</td>
<td>0.27</td>
<td>0.061</td>
<td>0.006</td>
<td>(1.3 \times 10^{-23})</td>
<td>1.09</td>
<td>(1.1 \times 10^{-2})</td>
</tr>
<tr>
<td>rs1165151</td>
<td>6</td>
<td>25929595</td>
<td>SLC1A7</td>
<td>SLC1A7</td>
<td>T</td>
<td>G</td>
<td>0.47</td>
<td>-0.091</td>
<td>0.005</td>
<td>(7.0 \times 10^{-70})</td>
<td>0.86</td>
<td>(5.3 \times 10^{-7})</td>
</tr>
<tr>
<td>rs1171614</td>
<td>10</td>
<td>61139544</td>
<td>SLC16A9</td>
<td>SLC16A9</td>
<td>T</td>
<td>G</td>
<td>0.22</td>
<td>-0.079</td>
<td>0.007</td>
<td>(2.3 \times 10^{-28})</td>
<td>0.91</td>
<td>(1.7 \times 10^{-2})</td>
</tr>
<tr>
<td>rs2078267</td>
<td>11</td>
<td>64094060</td>
<td>SLC22A11</td>
<td>SLC22A11</td>
<td>T</td>
<td>C</td>
<td>0.51</td>
<td>-0.073</td>
<td>0.006</td>
<td>(9.4 \times 10^{-38})</td>
<td>0.88</td>
<td>(2.3 \times 10^{-5})</td>
</tr>
<tr>
<td>rs478607</td>
<td>11</td>
<td>64234639</td>
<td>NRNX2</td>
<td>SLC22A12</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>-0.047</td>
<td>0.007</td>
<td>(4.4 \times 10^{-31})</td>
<td>0.97</td>
<td>(4.1 \times 10^{-1})</td>
</tr>
<tr>
<td>rs3741414</td>
<td>12</td>
<td>56130316</td>
<td>INHBC</td>
<td>INHBE</td>
<td>T</td>
<td>C</td>
<td>0.24</td>
<td>-0.072</td>
<td>0.007</td>
<td>(2.2 \times 10^{-25})</td>
<td>0.87</td>
<td>(2.7 \times 10^{-4})</td>
</tr>
</tbody>
</table>

### Table 2 Urate concentration–associated SNPs from secondary analyses

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>bp (Build 36)</th>
<th>Gene</th>
<th>A1</th>
<th>A2</th>
<th>Freq. A1</th>
<th>Effect</th>
<th>s.e.</th>
<th>(P) value</th>
<th>OR</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1449178</td>
<td>6</td>
<td>43380166</td>
<td>SLC22A7</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>-0.034</td>
<td>0.007</td>
<td>(1.2 \times 10^{-6})</td>
<td>0.95</td>
<td>(2.2 \times 10^{-1})</td>
</tr>
<tr>
<td>rs675318</td>
<td>12</td>
<td>52265330</td>
<td>SLC22A7</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>-0.034</td>
<td>0.007</td>
<td>(1.2 \times 10^{-6})</td>
<td>0.95</td>
<td>(2.2 \times 10^{-1})</td>
</tr>
<tr>
<td>rs675319</td>
<td>12</td>
<td>52265330</td>
<td>SLC22A7</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>-0.034</td>
<td>0.007</td>
<td>(1.2 \times 10^{-6})</td>
<td>0.95</td>
<td>(2.2 \times 10^{-1})</td>
</tr>
</tbody>
</table>

\(P\) values are corrected for inflation using genomic control. Median imputation quality was calculated across all cohorts for urate concentration–associated SNPs in the overall sample and ranged from 0.83 to 1 (median of 0.98). Allele frequencies are presented for the discovery sample. Gout estimates are provided for the combined discovery and validation samples. Sample sizes available for the individual SNPs are listed in Supplementary Table 6. Chr., chromosome; A1, allele 1, effect allele; freq., frequency; s.e., standard error.

These SNPs met one but not both criteria required for replication. \(a^2P\) value \(< 1 \times 10^{-700}\).
A weighted genetic urate score was constructed on the basis of the genetic urate concentration–associated SNPs with gout in the data set described. We observed a positive linear correlation between the genetic effect on serum urate concentrations and the log odds of gout for the replicated loci (Pearson's correlation = 0.93; Fig. 2). In addition, information was available on the associations with incident gout over a period of up to 22 years in two independent studies that were not part of either the gout or serum urate concentration discovery analyses. The 1,036 cases of gout in these studies met the American College of Rheumatology Criteria\(^2^4,2^5\). In all gout samples combined (3,151 cases and 68,350 controls), 17 out of 26 of the replicated urate concentration–associated SNPs showed nominal association with gout \((P < 0.05); \) Table 1. SNP associations with prevalent and incident gout were of generally comparable direction and magnitude overall, as well as among men and women separately (Supplementary Table 11).

**Genetic urate risk score is associated with gout**

A weighted genetic urate score was constructed on the basis of the number of risk alleles across loci from the main analysis as described previously\(^2^6\) and scaled to a risk allele count range as outlined (Supplementary Note). The association between urate scores and gout was evaluated in studies with a large number of cases (prevalent gout: ARIC, SHIP and KORA F4; incident gout: NHS and HPFS). Risk scores ranged from 10 to 45 (mean of 31 ± 5). Risk scores were significantly associated with increased odds of prevalent gout (odds ratio \(OR = 1.11\) per risk score unit increase, 95% confidence interval \((CI) = 1.09–1.14\); \(P = 2.5 \times 10^{-29}; n = 693\) cases) and incident gout over a period of up to 22 years \((OR = 1.10, 95\% CI = 1.08–1.13; P = 3.7 \times 10^{-21}; n = 1,036\) cases). Gout prevalence increased from <1% to 18% across risk score categories in the population-based studies, and the increased prevalence of gout with higher genetic urate score categories could be replicated in the independent studies recording incident cases of gout (Supplementary Fig. 6).

**Characterization of serum urate concentration–associated loci**

To gain insight into how the identified variants might contribute to altered serum urate levels, we evaluated their association with the fractional excretion of uric acid (FEUA, \(n = 6,799\); Supplementary Table 12). FEUA is the proportion of urate filtered by the glomeruli that is eventually excreted in the urine. Physiological values are typically well below 10%. SNPs at 10 replicated loci showed nominal association with FEUA (Table 3); SNPs at the SLC2A9, GCKR and IGF1R loci passed a multiple testing–corrected threshold for 26 replicated SNPs \((P < 1.9 \times 10^{-5})\). In all ten instances, the allele associated with higher serum urate concentrations was associated with lower FEUA.

**Associations in individuals of non-European ancestry**

To evaluate whether the observed SNP associations are generalizable to individuals of non-European ancestry, we investigated the replicated and genome-wide significant urate concentration–associated SNPs in 8,340 individuals of Indian ancestry, 5,820 African-Americans and 15,286 Japanese (Supplementary Table 13). Although allele frequencies at the index SNPs varied considerably across the groups (Fig. 3, right), the effects on serum urate concentrations were of comparable magnitude and identical direction for the majority of SNPs (Fig. 3, left), further supporting the idea that these index SNPs represent true urate concentration–associated loci.

**Association with transcript expression in cis**

The association of implicated SNPs with transcript expression in \(cis\) can potentially provide insights into the gene underlying the association signal. We therefore queried existing expression quantitative trait locus (eQTL) databases containing data from various tissues (Supplementary Note). For ten of the replicated loci, there were

---

**Table 3**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Closest gene</th>
<th>A1</th>
<th>A2</th>
<th>Serum urate (mg/dl)</th>
<th>FEUA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect</td>
<td>(P) value</td>
</tr>
<tr>
<td>rs1260326</td>
<td>GCKR</td>
<td>T</td>
<td>C</td>
<td>0.074</td>
<td>1.2 (\times 10^{-44})</td>
</tr>
<tr>
<td>rs12498742</td>
<td>SLC2A9</td>
<td>A</td>
<td>G</td>
<td>0.373</td>
<td>0*</td>
</tr>
<tr>
<td>rs2231142</td>
<td>ABCG2</td>
<td>T</td>
<td>G</td>
<td>0.217</td>
<td>1.0 (\times 10^{-134})</td>
</tr>
<tr>
<td>rs675209</td>
<td>RREB1</td>
<td>T</td>
<td>C</td>
<td>0.061</td>
<td>1.3 (\times 10^{-23})</td>
</tr>
<tr>
<td>rs2078267</td>
<td>SLC2A11</td>
<td>T</td>
<td>C</td>
<td>-0.073</td>
<td>9.4 (\times 10^{-38})</td>
</tr>
<tr>
<td>rs478607</td>
<td>NFXN2</td>
<td>A</td>
<td>G</td>
<td>-0.047</td>
<td>4.4 (\times 10^{-11})</td>
</tr>
<tr>
<td>rs1394125</td>
<td>UBE2Q2</td>
<td>A</td>
<td>G</td>
<td>0.043</td>
<td>2.5 (\times 10^{-13})</td>
</tr>
<tr>
<td>rs6598541</td>
<td>IGF1R</td>
<td>A</td>
<td>G</td>
<td>0.043</td>
<td>4.8 (\times 10^{-15})</td>
</tr>
<tr>
<td>rs7193778</td>
<td>NFAT5</td>
<td>T</td>
<td>C</td>
<td>-0.046</td>
<td>8.2 (\times 10^{-10})</td>
</tr>
<tr>
<td>rs7224610</td>
<td>HLF</td>
<td>A</td>
<td>C</td>
<td>-0.042</td>
<td>5.4 (\times 10^{-17})</td>
</tr>
</tbody>
</table>

Serum urate concentration estimates are provided for the combined discovery and replication sample. Sample size for the SNPs for FEUA ranged from 6,798 to 6,799. A1, allele 1, effect allele; s.e., standard error.

*\(p < 1 \times 10^{-7}\)
Supplementary Table 6

Japanese

Indian ancestry

was consistent with the epidemiologically observed associations, as was the case for the pleiotropic ATXN2 locus and blood pressure.

In aggregate, a weighted risk score composed of the 26 replicated urate concentration–increasing alleles was significantly associated (P < 4.5 × 10⁻³) only with plasma CRP concentrations after correcting for the number of traits investigated (Supplementary Note). This association was driven by the pleiotropic GCKR locus rather than being a general effect of urate concentration–associated variants on the inflammatory marker CRP; the association between the score and CRP concentrations was abolished in a sensitivity analysis when rs1260326 in GCKR was excluded from the genetic urate score. The absence of significant associations between the genetic urate score and the investigated traits, including blood pressure, might suggest a lack of causal relationships between serum urate concentrations and these traits, given that the present study had adequate statistical power to detect significant associations, as shown previously in a smaller subset. However, the methodology used has limitations, and alternative explanations cannot be ruled out.

Network analysis identifies additional loci

To identify further urate concentration–associated genomic regions and in order to place these into a biologically interpretable context, we incorporated previous knowledge on molecular interactions in which the gene products of implicated genes operate. First, GRAIL was used to infer the most likely gene underlying the association at each of the 37 loci from the discovery analyses of the sex-combined network analyses that passed a multiple testing–corrected threshold (P < 6.8 × 10⁻⁵, Bonferroni correction for SNPs in all implicated genes). Replication was attempted for the resulting 17 SNPs in up to 23,078 participants (Supplementary Table 6). Replication criteria, defined as in the discovery analysis, were met for two regions: B3GNT4 and ACVR1B-ACVR1L. Moreover, ARTNT had a suggestive replication q value of <0.05 but did not reach genome-wide significance. Regional association plots are shown (Supplementary Fig. 7).

A specific subnetwork from the analysis around the inhibins-activins pathway was highlighted as part of the functional association network analysis, in which three seed genes (INHBB, INHBC-INHBE and ACVR2A) were connected with three genes implicated by the functional association network analysis (ACVR1B-ACVR1L, ACVR1C and BMPR2), one of which replicated (Supplementary Fig. 8). The gene products operate in one pathway and are inhibitors-activins receptors and their ligands. Detailed information on all replicated genes in the inhibins-activins network is provided in the Supplementary Note. Another pathway linking a large number of GWAS discovery association signals contains genes involved in growth factor signaling through tyrosine kinase receptors implicated in the control of cell growth and differentiation (Supplementary Fig. 9). One of these genes, PKRL, is also highlighted in the Supplementary Note, which provides information about identified associated loci that include candidate genes with a role in glucose metabolism. The functional

![Figure 3 SNP effects on urate concentrations (mg/dl) are similar among individuals of European ancestry, African-Americans, Indians and Japanese, whereas allele frequencies vary. Comparison of SNP effects at the replicated loci and at four additional loci meeting one of the replication criteria. Left, effects on serum urate concentrations per minor allele (defined in individuals of European ancestry) sorted by effect size. Right, comparison of allele frequencies across the four different samples. Symbols are absent if the data could not be provided for reasons related to allele frequency (not polymorphic, low minor allele frequency) or poor imputation quality.](image-url)
network associations that formed the basis for Supplementary Figures 8 and 9 are provided in Supplementary Table 17. Pathway analyses using Ingenuity Pathway Analysis software (Supplementary Note) showed functional network associations with gene expression, cellular organization, carbohydrate metabolism, molecular transport and endocrine system disorders (lowest $P = 1 \times 10^{-28}$; Supplementary Table 18).

**DISCUSSION**

In GUGC, we identified and replicated 28 genome-wide significant urate concentration–associated loci, 18 of which were newly identified, using GWAS ($n = 26$) and pathway ($n = 2$) approaches. The loci from our GWAS meta-analysis were characterized in detail, including analyzing their associations with serum urate concentrations in samples of non-European ancestry and with the fractional excretion of urate. The serum urate concentration–increasing allele at all loci was associated with higher risk of gout. The functional association network analysis supports a new role for inhibins-activins pathways in regulating urate homeostasis.

We confirmed the presence of urate concentration–associated SNPs from previous GWAS for all 11 previously reported loci,$^{13-18}$ 6 of which map to genes encoding proteins related to urate transport across membranes. Conversely, none of the genes at the newly associated loci from this study seem to be obvious candidates for factors involved in urate transport. A primary candidate for having an influence on urate generation is PRPSAPI, which encodes a protein involved in the regulation of purine synthesis.$^{28}$ Many of the remaining index SNPs map to genes coding for transcription or growth factors with broad downstream responses. Some of these may ultimately affect the activity of known urate transporters, such as the transcription factor HNF-1α, encoded by a suggestive urate concentration–associated locus that has response elements upstream of ABCG2, SLC17A3, SLC22A11, SLC16A9 and PDZK1.

Many of the newly identified loci can be connected into functional or biochemical pathways. All pathway approaches evaluated in our analyses highlight terms related to carbohydrate metabolism, such as regulation of glycolysis, glucose, insulin and pyruvate (Supplementary Tables 16 and 18). In addition to the previously identified GCKR locus, the gene products of five of the newly discovered replicated loci (Supplementary Note) directly modulate glucose flux. The association of serum urate concentrations with these glucose metabolism–related loci may be explained by increased flow of glucose-6-phosphate through the pentose phosphate pathway and/or high levels of aerobic glycolysis with lactate production as observed in proliferating cells (Warburg effect)$^{29-31}$. The pentose phosphate pathway generates ribose-5-phosphate, a key precursor of de novo purine synthesis and thereby of uric acid production. The amount of lactate generated can influence urate transport across membranes and can therefore alter urate elimination by the kidney. Detailed gene information and an illustration of how several of these genes are connected in regulatory networks targeting the rate-limiting glycolytic enzymes glucokinase and pyruvate kinase are provided (Supplementary Note).

Moreover, insulin and insulin-like growth factor 1, the receptors of which showed suggestive and significant association with urate concentrations, induce glucokinase gene (GCK) expression in the liver and pancreas, respectively.$^{32}$

In addition to known urate transporters expressed in the renal proximal tubules, three loci linked to glucose metabolism and/or insulin response (GCKR, IGF1R and NPAT5) were nominally associated with FEUA. The urate concentration–increasing alleles at all loci were associated with lower FEUA, which could either be owing to increased urate reabsorption in the kidney by the mechanisms discussed or possibly to the reported effect of insulin on decreasing renal urate clearance and sodium excretion in healthy people$^{33,34}$. Despite the connection of these new candidate genes for urate regulation in networks related to glucose metabolism, there was no significant aggregate effect of urate concentration–associated loci on glyceremic traits or measures of insulin resistance. A possible explanation might be the absence of a sizeable effect of SNPs in heterogeneous pathways on a single related physiological process or the lack of statistical power of current GWAS efforts to detect small effect sizes.

Another major pathway highlighted by our analyses is the inhibins-activins growth factor system (Supplementary Fig. 8 and Supplementary Note). This pathway is thought to mediate a very wide range of processes, which could affect urate levels through a variety of known functions, such as energy balance, insulin release, apoptosis, inflammation and sex hormone regulation. In addition, our approach demonstrates the value of the functional association network analyses in identifying additional loci from conventional GWAS results.

The mainstay of gout therapy is allopurinol, which inhibits the formation of urate and is an inexpensive and effective preventative agent for chronic gout. Although other therapeutic options exist, the overall number of urate concentration–lowering agents approved for clinical use is limited. Data from a clinical trial indicate that only 21% of patients offered the most common dose of allopurinol (300 mg/d) achieve optimal levels of serum urate$^{35}$. In addition, there are rare, serious side-effects, and reduced doses of allopurinol are recommended in individuals with impaired renal function. The identification of additional therapeutic options to lower serum urate levels is therefore an active area of research. Our finding that gene loci implicated in glucose homeostasis influence serum urate concentrations fits with the observation that drugs that decrease insulin resistance, such as thiazolidinediones and metformin, tend to decrease serum urate levels$^{36-40}$. Thus, the identification of new gene loci in the present study may open new avenues for research to improve the treatment and prevention of gout.

The major strengths of our study are the large sample size with $>$110,000 individuals for the discovery step, the independent replication of associations and the possibility to examine identified loci for their relationship to prevalent and American College of Rheumatology Criteria–confirmed incident gout. The comprehensive evaluation of urate concentration–associated variants provided additional insights by detecting associations with fractional urate excretion and transcript expression.

Limitations of our study include the relatively modest sample size available for the replication step, which compromises power and potentially results in an inability to validate true urate concentration–associated loci such as ORCAL, OVOL1 and BCS3. The definition of gout varied across several cohorts, which may have led to some misclassification but is unlikely to result in false positive results. As with other GWAS, our analysis focused mostly on common SNPs, and we could therefore not assess the contributions of rare variants, such as a recently described risk variant for gout$^{41}$.

We identified 28 genome-wide significant loci associated with serum urate concentrations. We found that alleles associated with increased serum urate concentrations were also associated with increased risk of gout. The modulation of urate production and excretion by signaling processes that influence metabolic pathways, such as glycolysis and the pentose phosphate pathway, seem to be central pathways including the genes from the newly identified associated loci. These findings may have implications for further research into urate concentration–lowering drugs to treat and prevent the common inflammatory arthritis gout.

Center, Rotterdam, The Netherlands. 124Institute of Clinical Chemistry, University Hospital of Zurich, Zurich, Switzerland. 125Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. 126Medical Department, University of Leipzig, Leipzig, Germany. 127Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, Leipzig, Germany. 128Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland. 129Institute of Molecular Medicine, University of Helsinki, Helsinki, Finland. 130Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. 131Harvard Medical School, Boston, Massachusetts, USA. 132Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. 133Department of Medicine, University of Turku and Turku University Hospital, Turku, Finland. 134Department of Internal Medicine II, University Hospital Regensburg, Regensburg, Germany. 135Department of Neurology, Mayo Clinic, Jacksonville, Florida, USA. 136Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. 137Imperial College Cerebrovascular Research Unit, Imperial College London, London, UK. 138Laboratory for International Alliance, Center for Genomic Medicine, RIKEN, Yokohama, Japan. 139Nephrology Division, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland. 140Section of Rheumatology, Boston University School of Medicine, Boston, Massachusetts, USA. 141Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, Massachusetts, USA. 142Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. 143University of Texas–Houston Institute of Molecular Medicine, Houston, Texas, USA. 144University of Texas–Houston School of Public Health, Houston, Texas, USA. 145Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. 146Renal Program, Institute of Internal Medicine, Columbus-Gemelli University Hospital, Catholic University, Rome, Italy. 147Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 148Synlab Centre of Laboratory Diagnostics, Heidelberg, Germany. 149Division of Nephrology, Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 150On behalf of PROCARDIS. 151Institute of Molecular Genetics, CNR, Pavia, Italy. 152Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. 153Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. 154Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. 155University Institute of Social and Preventive Medicine, Lausanne, Switzerland. 156These authors contributed equally to this work. 157These authors jointly directed this work. Correspondence should be addressed to A. Köttgen (anna.koettgen@uniklinik-freiburg.de), V.V. (veronique.vitart@igmmed.ac.uk), M. Bochud (murielle.bochud@chuv.ch) or C.G. (christian.gieger@helmholtz-muenchen.de).
ALL meta-analyses were carried out in duplicate by two independent analysts. Meta-analysis was performed on the results of all genome-wide scans using a fixed-effects model applying inverse variance weighting as implemented in METAL. Results were confirmed by comparing the results to those from a 2 score–based meta-analysis. SNPs that were present in <75% of all samples contributing to the respective meta-analysis were excluded, and the remaining SNPs were used as the basis for all subsequent analyses.

Consequently, data for 2,450,547 genotyped or imputed autosomal SNPs were available for the primary analysis of serum urate concentrations, and data for 2,538,056 SNPs were available for the primary gout analysis. After all meta-analyses, a second genomic control correction was applied to primary analyses (genomic inflation factor = 1.12 for urate and 1.03 for gout) as well as secondary sex-stratified analyses (genomic inflation factor for urate = 1.07 (men), 1.08 (women) and gout = 1.04 (men), 1.03 (women)). The top three SNPs from the gout meta-analysis results among women were removed because they were thought to represent false positive results on the basis of low MAF (0.02), very low P values and an F heterogeneity measure of >99%.

Genome-wide significance was defined as a P value of ≤ 5 × 10⁻⁸ after the second correction for genomic control, corresponding to a Bonferroni correction of 1 million independent tests. All loci containing SNPs of suggestive significance (P < 1 × 10⁻⁶) in the discovery-stage meta-analyses were considered for further analysis and assessed for the presence of independent SNPs within each locus.

Assessment of the presence of independent signals at each locus. Two steps were performed to identify independent SNPs in the same region that associated with serum urate concentration in the overall analyses. First, SNPs with P values of < 1 × 10⁻⁵ were aggregated on the basis of the LD structure from the HapMap release 28 Utah residents of Northern and Western European ancestry (CEU) data set using PLINK (settings r² > 0.01, 1-Mb distance). Second, to verify the potential independent associations from the first step, a multiple-regression model was calculated in 32 of the studies, adjusting for all lead SNPs at once, and meta-analysis was performed, followed by a comparison of the effect estimate for each SNP with those from the single-SNP association model using a t test (Supplementary Table 5). Independent SNPs were defined as those with (i) a t-test P value of >0.05 and (ii) a difference of the effect estimates between the single- and the multiple-SNP models of ≤ 20% compared to the single-SNP model. rs1178977 in BAZ1B did not fulfill the criteria for independence (20.7% change in effect estimate). Yet, it was treated as an independent SNP, as the adjacent associated genomic region was located >75 Mb away, suggesting that the change in β estimate might be due to other factors than a nearby SNP in LD. Analyses were performed in R.

Subsequently, 37 independent loci of suggestive significance (P < 1 × 10⁻⁶) were carried forward for replication.

Secondary analyses. To further identify genetic variants associated with serum urate concentrations and gout, several additional analyses were carried out. These included X-chromosome analyses, sex-stratified analyses, a gene-based test and urate transporter candidate analyses, as well as secondary analyses to improve the characterization of associated SNPs. To this end, we investigated associations with FEUA, associations in individuals of non-European descent.
ancestry, associations with transcript expression, associations with serum metabolite concentrations, risk score analyses and associations with other urate-related phenotypes. Secondary analyses are described in detail in the Supplementary Note.

As a result of the secondary analyses, seven additional SNPs were identified for further replication testing: one from the analysis in men only, five from the analysis in women and one from the urate transporter candidate analysis.

A new part of the secondary analyses was the implementation of functional association networks as described in the Supplementary Note. As detailed there, the approach based on functional associations among the urate genes, mostly protein-protein interactions, led to the identification of 17 additional independent SNPs in newly identified genomic regions that were subjected to replication testing. Altogether, we therefore tested 61 SNPs (37 overall, 6 sex specific, 1 urate transporter candidate and 17 network) for replication in additional study samples.

**Replication analyses.** Up to 18,821 participants contributed information to the *in silico* replication analyses of the 61 independent variants tested, and additional *de novo* genotyping was conducted in up to 13,992 individuals. The cohorts are described in Supplementary Table 1. We excluded SNPs with call rate of <0.9 or imputation quality of <0.3 (MACH) or <0.4 (Impute, BEAGLE). The statistical methods used to estimate the associations were identical to the ones used in the discovery cohorts.

Details about *de novo* genotyping, including platforms, array design and quality control, are provided in the Supplementary Note. To assess evidence for replication, test statistics of all *in silico* and *de novo* replication cohorts were combined using fixed-effects and inverse variance weighting using METAL. The false discovery rate was estimated in the independent replication samples as a q value on the basis of the P-value distribution of all SNPs tested for replication using the qvalue package in R. Replication was defined, for each SNP, as q value < 0.05 in the replication samples alone, indicating that <5% false positives would be expected among these variants. To place an additional filter on these results, only SNPs that were associated with genome-wide significance (P < 5 × 10⁻⁸) in the combined analysis were considered to have been successfully replicated; combined estimates were obtained by meta-analysis of the double-corrected test statistics of the discovery GWAS with the replication estimates.

---


