

# Colocalization analysis reveals shared genetic loci contributing to gout and metabolite levels

Riku Takei<sup>1</sup>, Nicholas A. Sumpter<sup>2</sup>, Megan P. Leask<sup>3</sup>, and Tony R. Merriman<sup>1,4</sup>

<sup>1</sup>Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, United States

<sup>2</sup>Department of Internal Medicine and Radboud Institute of Molecular Life Science, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>3</sup>Department of Physiology, University of Otago, Dunedin, New Zealand

<sup>4</sup>Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

## Abstract

Genetic association studies in gout have identified genetic variants in or near genes involved in the biosynthesis and transport of urate and in immunological pathways. However, the causal role of the remaining genetic variants, genes, and pathways in gout is not clear. Here, we present the results from a genetic colocalization analysis of gout-associated signals with metabolite quantitative trait loci (metQTL), shedding light on the metabolites that are likely directly affected by genetic variants associated with gout. We identified 141 candidate metabolites with evidence of colocalization with at least one gout-associated genetic signal, of which 29 showed evidence of a causal relationship with gout by Mendelian randomization. Among the 29 metabolites were lysophosphatidylcholines, which may affect the inflammatory response by binding to the TLR-2/4 receptors, providing plausible candidate metabolites for future studies that link metabolites with inflammatory processes in gout.

## Introduction

Gout is a common inflammatory arthritis caused by an innate immune response to monosodium urate (MSU) crystals that primarily deposit in joints<sup>1</sup>. Though hyperuricemia (characterized by serum urate concentration  $\geq 7\text{mg/dL}$ <sup>1</sup>) is a prerequisite of gout, an inflammatory response by the NLRP3-inflammasome is required to trigger gouty inflammation<sup>1,2</sup>. Genome-wide association studies (GWAS) in gout<sup>3–8</sup> have revealed many genetic loci, some of which have been implicated in biosynthesis and transport of urate.

Metabolomics is a field that involves the measurement of a large number of metabolites within a group of participants to identify metabolite levels that are associated with a phenotype of interest<sup>9,10</sup>. When these cohorts also contain genetic information it allows for the association of genetic variants with metabolite levels<sup>11–13</sup> and provide valuable insights into the biological mechanisms by narrowing the search space for causal metabolites that may be causally related. Recently, there have been several GWAS of metabolomic

profile that associated genetic variants with metabolite levels: Chen et al.<sup>11</sup> studied 1,091 blood metabolites and 309 metabolite ratios in 8,299 individuals from the Canadian Longitudinal Study on Aging (CLSA) cohort, Yin et al.<sup>13</sup> studied 1,391 metabolites in the plasma of 6,136 male participants from the METSIM study, and Schlosser et al.<sup>12</sup> studied 1,296 plasma and 1,399 urine metabolites from 5,023 participants from the German Chronic Kidney Disease (GCKD) study. These studies provide a wealth of resources to study the effect of genetic variations that influence metabolite concentrations in both plasma and urine, which in turn can be used for insights into the biological mechanism of a phenotype of interest.

In the context of gout, however, only a few metabolomics studies have been done; of note are three studies by Renaudin et al.<sup>14</sup>, Lyu et al.<sup>15</sup>, and Joshi and McCormick et al.<sup>16</sup>. Renaudin et al.<sup>14</sup> explored the metabolic profile of bone marrow-derived macrophages (BMDMs) and THP-1 cells stimulated with either MSU crystal or calcium pyrophosphate (CPP) crystal and found that the glycolytic pathway, but not oxidative phosphorylation (OXPHOS), has a pivotal role in the IL-1 $\beta$ -mediated inflammatory process. Lyu et al.<sup>15</sup> recruited subjects in various stages of gouty arthritis including no gout, hyperuricemia, acute gout, intermittently symptomatic gout, and chronic tophaceous gout, to identify biomarkers associated with progression of gouty inflammation. They identified four metabolites (kynurenic acid, N1-Methyl-2-pyridone-5-carboxamide, DL-2-Aminoadipic acid, and 5-hydroxyindole acetic acid) as potential biomarkers for progression of gout. More recently, Joshi et al.<sup>16</sup> utilized metabolomics data from the UK Biobank (UKB) cohort to investigate the metabolomic changes that predict incident gout, using Mendelian randomization to identify glycoprotein acetyls (GlycA) as a causal metabolite of incident gout. Lower levels of glutamine were associated with increased risk of gout in both Renaudin et al.<sup>14</sup> and Joshi et al.<sup>16</sup>, providing consistent evidence of the role of glutamine in the biological process of gout, possibly through glutaminolysis and trained immunity from the accumulation of fumarate<sup>17</sup>. In fact, the most recent genetic study of gout<sup>4</sup> identified that the proline allele of the p.Leu581Pro missense variant of the glutaminase 2 enzyme, which associated with reduced levels of glutamine, also associates with increased risk of gout, supporting the findings from these metabolomics studies that show the important role of glutamine in gout.

Even though these studies provide valuable insights into the biological processes surrounding gout, only a handful of metabolites have been identified and causally associated with gout and progression of gout. Since metabolite levels are direct consequences of enzymatic activities, by combining the largest genetic association data of gout<sup>4</sup> and the recent metabolomics studies<sup>11–13</sup> it may be possible to identify metabolites that are also associated with gout-associated genetic variants and consequently identify causal molecular mechanisms of gout. Here we conducted colocalization analyses of genetic loci from the most recent GWAS of gout<sup>4</sup> with the plasma and urine metabolite quantitative trait loci (metQTL) data from CLSA<sup>11</sup>, METSIM<sup>13</sup>, and GCKD<sup>12</sup> studies in order to identify metabolites that may be affected by genetic loci that also affect gout. Our analysis included pathway enrichment analysis to highlight pathways that are likely involved with the gout-associated metabolites and also include two-sample Mendelian randomization of the metabolites with gout to assess possible causal relationships of the metabolite levels with gout.

## Methods

### *Summary data and meta-analysis*

Association summary statistics for 1,400 plasma and plasma metabolite ratios from the CLSA cohort<sup>11</sup>, data from 1,391 plasma metabolites from the METSIM study<sup>13</sup>, and data from 1,296 plasma and 1,399 urine metabolites from the GCKD study<sup>12</sup> were downloaded from the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and METSIM Metabolomics PheWeb (<https://pheweb.org/metsim-metab/>). Full European gout and urate GWAS summary statistics from Major et al.<sup>4</sup> were used in all of the analyses. For the purpose of Mendelian randomisation, fixed-effect meta-analysis was carried out using METAL<sup>18</sup> on the 141 metabolites that colocalized with gout genetic signals and that were also present in all three metabolite data sets.

### *Colocalization analysis*

The regions for colocalization was restricted to the 276 independent gout loci (lead variant  $\pm 500\text{kb}$  window) and only the variants (based on rsID) present in both the gout GWAS and metQTL data were kept. Colocalization of gout summary statistics with each of the metQTL association summary statistics was carried out using the ‘coloc’<sup>19</sup> R<sup>20</sup> package for the 276 loci. Loci that contained less than 100 variants within the  $\pm 500\text{kb}$  window of colocalization region were excluded from the analysis. A locus was considered to be colocalized if the posterior probability of colocalization (hypothesis 4) was 0.8.

### *Mapping of metabolite names*

Not all of the metabolite names used in the three metQTL data sets were consistent, so the metabolite names were mapped to those in the Human Metabolome Database<sup>21</sup> (HMDB version 5.0; downloaded from <https://hmdb.ca/downloads>). The original metabolite name in each of the data set was matched with the metabolite name (or its synonym) as shown in HMDB. If a metabolite name did not match exactly with the HMDB metabolite name, then the name was matched by approximate string matching using the ‘fedmatch’<sup>22</sup> package in R. In cases where a metabolite name matched with multiple candidate names, the names were checked manually in order to determine the metabolite name that was most likely correct based on chemical structure.

### *Pathway enrichment analysis*

Genes were identified from the gout loci (lead variant  $\pm 500\text{kb}$  window) that colocalized with any of the 141 metabolites common to all three metQTL data sets. All genes that had their gene region overlapped with the 1Mb region around the lead variant were considered for the pathway analysis. The list of genes was used for pathway enrichment analysis using DAVID<sup>23,24</sup>.

### *Mendelian randomization*

141 metabolites that colocalized with gout and were present in all three metabolite data sets were consid-

ered for Mendelian randomization (MR) analysis to determine their causal relationship with gout using the ‘MendelianRandomization’<sup>25</sup> package in R. Lead variants from each of the 141 meta-analyzed metabolite summary statistics were identified using the --clump option in PLINKv1.90<sup>26</sup>, using the European ancestry samples from the 1000 Genomes Project<sup>27</sup> as the reference for linkage disequilibrium (LD). To ensure that the lead variants identified were LD-independent, pairwise LD was calculated for all of the lead variants using PLINK and 1000 Genomes Project European ancestry data, and variants with  $r^2 \geq 0.01$  were removed. Any metabolite that had two or less genetically-associated lead variants was not considered for the MR analysis due to technical limitations of the MR methods used. Inverse variance-weighted (IVW) and weighted median (WM) methods were used to test for causality of the metabolite with gout (and vice versa), and the MR-Egger method was used to test for pleiotropy by considering the MR-Egger intercept.

## Results

Colocalization analysis was first carried out for 1,391 plasma metabolites from the METSIM cohort with 276 independent genetic loci from the European gout GWAS from Major et al.<sup>4</sup> (Supplementary table S1). In total, 633 metabolites (45.5% of tested metabolites) colocalized with at least one gout locus implicating 135 (48.9%) of the 276 tested loci, highlighting the extent to which gout genetic loci associate with the human metabolome (Table 1). On average, a metabolite colocalized with 1.56 gout loci in plasma, and there were no more than six colocalized loci for any one plasma metabolite. Conversely, the average number of metabolites that colocalized with a particular gout locus was 7.12, with the largest number of metabolites ( $n = 178$ ) colocalizing at the *GCKR* locus (chr2:26.91MB-28.71MB). These results suggest that metabolites are genetically controlled by a small number of gout-associated loci while a gout genetic locus is able to influence a variety of metabolites.

We repeated the colocalization analysis in metabolite data sets from two independent cohorts: CLSA and GCKD (Table 1 and Supplementary tables S2-S3). Similar to what was observed with the METSIM cohort, 137 (49.6%) gout loci colocalized with 719 (51.4%) metabolites from the CLSA cohort (Table 1; 7.96 metabolites per gout locus). In the GCKD cohort, the number of loci colocalized with plasma and urine metabolites were comparable to the number observed in METSIM (136 (49.3%) and 138 (50.0%) loci), but with less metabolites compared to the METSIM and CLSA cohorts (482 (37.2%) plasma and 430 (30.7%) urine metabolites; 4.74 and 4.17 metabolites per locus, respectively). Furthermore, the average number of gout loci that colocalized with a particular metabolite was comparable to that observed in the METSIM cohort (1.55 in the CLSA cohort, 1.36 and 1.34 loci with plasma and urine metabolites in the GCKD cohort, respectively). Compared to the METSIM data set, the average number of metabolites that colocalized with a gout locus was slightly larger in the CLSA cohort (7.96 metabolites) and lower in the GCKD cohort (4.74 for plasma and 4.17 for urine metabolites), though this may be due to the lower total number of metabolites that colocalized with the gout loci in the GCKD data set. However, the overall trend where there are more colocalized metabolites per locus compared to the number of gout loci that colocalize with a metabolite was

consistent across cohorts.

	Total number of metabolites	Number of unique metabolites colocalized with gout loci	Number of gout loci colocalized with metabolites	Average number of metabolites that colocalized with gout locus <sup>1</sup>	Average number of loci that colocalized with a metabolite
METSIM	1,391	633	135	7.12	1.56
CLSA	1,400	719	137	7.96	1.55
GCKD (plasma)	1,296	482	136	4.74	1.36
GCKD (urine)	1,399	430	138	4.17	1.34

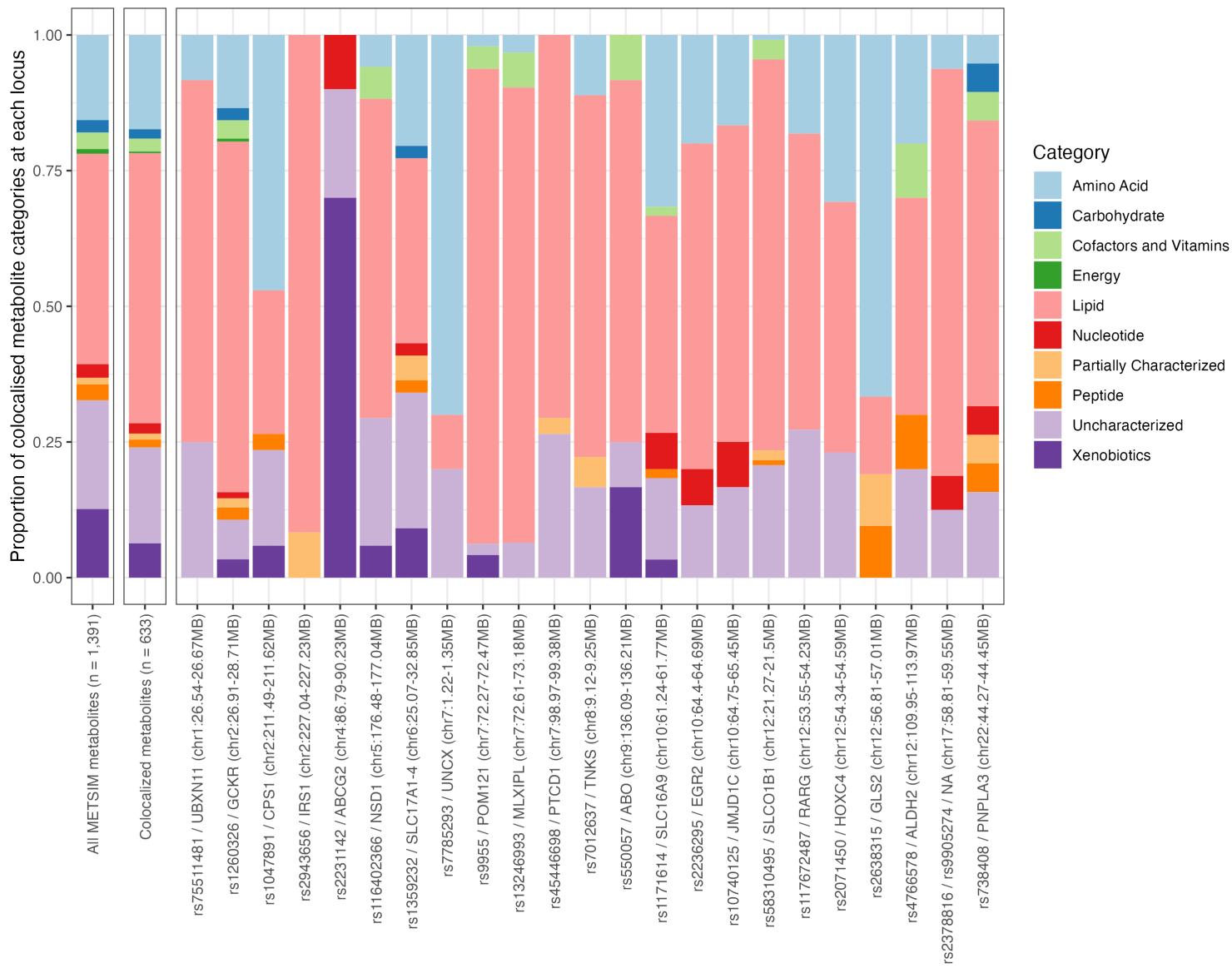
**Table 1:** Summary of the number of metabolites and gout genetic loci colocalized with one another

Next, we investigated the types of metabolites that colocalized at each of the gout loci in the METSIM cohort using the ten category descriptions (amino acids, carbohydrate, cofactors and vitamins, energy, lipid, nucleotide, partially characterized, peptide, uncharacterized, and xenobiotics; Supplementary table S5). Among the 633 metabolites that colocalized with a gout locus, lipid metabolites were the most common (proportion of metabolites that were lipids = 0.498), followed by uncharacterized metabolites (0.177), and amino acids (0.174). These proportions were similar to the overall proportions observed in all METSIM metabolites ( $n = 1,391$  metabolites; Figure 1 and Supplementary table S5). The proportions from the colocalized metabolites were similar in loci that had  $\geq 10$  colocalized metabolites (23 loci in total), except at the *ABCG2* locus (rs2231142, chr4:86.79MB-90.23MB; Figure 1). At the *ABCG2* locus, 70% (7/10) of the metabolites that colocalized were xenobiotics, likely reflecting the role of ABCG2 as a universal transporter for a variety of drugs, metabolites, and molecules<sup>28</sup>, including urate (one of the colocalized metabolites at this locus and causal of gout). This result was not able to be compared with the CLSA and GCKD cohorts, as the category information was not available in these metabolite datasets.

To identify plasma metabolites that were colocalized in the three data set (METSIM, CLSA, and GCKD), the metabolite names from each data sets were first mapped to standard metabolite names from the Human Metabolome Database (HMDB; Methods). In total, 359 ( $359/633 = 56.7\%$ ), 326 ( $326/719 = 45.3\%$ ), and 274 ( $274/482 = 56.8\%$ ) plasma metabolites from METSIM, CLSA, and GCKD cohorts matched with an HMDB metabolite name, respectively, of which 141 ( $141/529 = 26.7\%$ ) metabolites were common in all three data sets (Figure 2 and Supplementary table S6). Of the 141 common plasma metabolites 89 (63.1%) did not colocalize in urine (out of 218 urine metabolites that had a HMDB metabolite name; Figure 2 and Supplementary tables S7 and S8).

To determine the biological pathways represented by the colocalized metabolites, pathway enrichment analysis was carried out using the genes present in the loci that colocalized with 141 plasma metabolites that were common among all three data sets. Three sets of protein coding genes (1,489, 1,245, and 908 genes) were identified from the loci that colocalized with all 141 plasma, 89 plasma-specific, or 52 metabolites present in both plasma and urine, respectively (Methods). Fifteen, eight, and nine pathways were

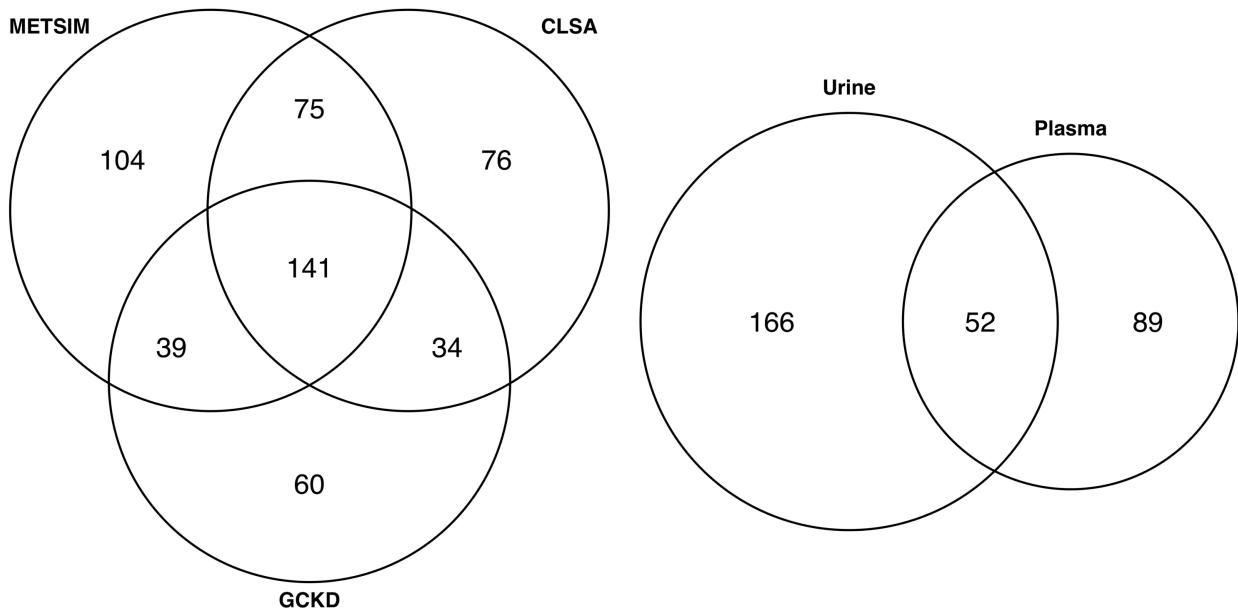
<sup>1</sup>Note that the average number of metabolites per gout locus includes metabolites colocalized to multiple loci.



**Figure 1:** Proportion of broad categories for all 1,391 METSIM metabolites (left panel), 633 colocalized metabolites (middle panel), and 23 gout loci with at least ten colocalized metabolites (right panel).

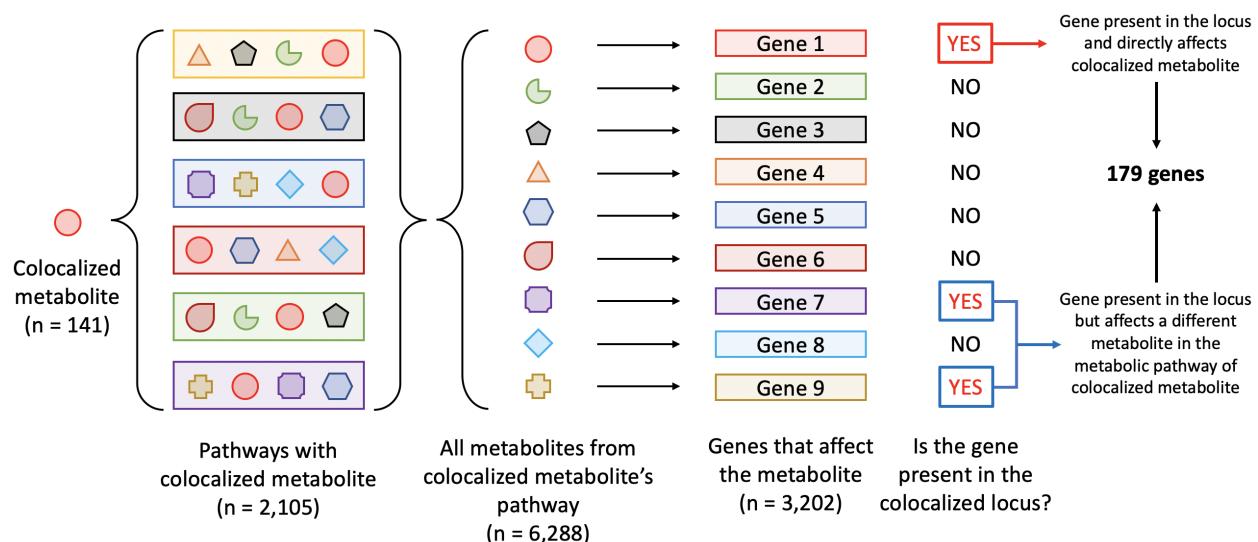
significantly enriched (FDR-adjusted  $P \leq 0.05$ ) with the three gene sets (Supplementary tables S9-S11), respectively. Notable pathways identified from plasma-specific genes included glycolysis/gluconeogenesis, pyruvate metabolism, ethanol oxidation, and retinol metabolic process, indicating the genes present in the loci that colocaled specifically with the plasma metabolites may be derived from metabolic processes for energy generation. In contrast, pathways that were enriched by the genes present in the loci that colocaled with both plasma and urine mainly included transportation, rather than metabolism, of molecules and compounds, such as urate metabolism and transport, transmembrane and organic anion transport, and bile transport and secretion. Many of the plasma-specific and plasma and urine pathways were present in the full list of genes (set 1), though there were some unique pathways that did not appear in the other two sets including the PI3K-AKT signaling pathway, and the insulin resistance and signaling pathway.

Given that not all of the genes will be directly involved with the biological pathways that affect the metabolite level, we focused on identifying genes directly involved with generation of the 141 colocaled



**Figure 2:** Venn diagram showing the overlap of the colocalized plasma metabolites in METSIM, CLSA, and GCKD cohorts (left), and the overlap of 141 common metabolites with colocalised GCKD urine metabolites (right).

plasma metabolites (Figure 3). We first identified 2,105 pathways that included at least one of the 141 plasma metabolites based on the pathway information in HMDB and extracted 6,288 metabolites from these pathways to ensure that any upstream or downstream metabolites, and therefore the genes that may affect the level of the 141 plasma metabolites, were included. Based on the information available in the Human Metabolome Database (HMDB), 3,202 genes encoding enzymes and proteins that directly affected the 6,288 metabolites were extracted and matched back to the loci that colocalized with the 141 plasma metabolites, of which 179 genes were located within the colocalized loci. This list of 179 genes is the most relevant set of genes that may affect the 141 plasma metabolite levels directly or indirectly, as the protein products of these genes affect either the colocalized metabolite itself, or affect the metabolite that is up- or downstream to the metabolite of interest in a pathway (Supplementary table S12). Of the 179 genes, however, only 17 at 9 gout loci encoded proteins that directly metabolizes or generates the colocalized metabolite; for ex-



**Figure 3:** Schematic of the identification of metabolites directly affected by genes present in the gout loci that colocalized with metabolite levels.

ample, the *GLS2* (encoding glutaminase 2) was identified at the rs58310495 locus (chr12:21.27MB-21.5MB region) which colocalized with L-glutamine (Supplementary table S13). Pathway enrichment analysis of the 179 genes identified 71 enriched pathways (FDR-adjusted  $P \leq 0.05$ ), of which 63 were not identified in the previous analyses, though some were closely related, such as fatty acid metabolism, metabolism of carbohydrates, and transport of organic anions (Supplementary table S14). Among the 63 significant pathways were estrogen metabolic process, androgen biosynthetic process, steroid biosynthesis, mitochondrial fatty acid beta-oxidation, a variety of amino acid metabolic processes (including, but not limited to, glutamine, arginine, proline, valine, leucine, and isoleucine), metabolism of water-soluble vitamins and cofactors.

Lastly, two-sample Mendelian randomization (MR) analysis was carried out to evaluate causality of the plasma metabolites with gout. In order to focus the search for metabolites that are most likely causal for gout, we focused on the 141 plasma metabolites that had metQTL colocalized with gout genetic association signals that were common amongst the three plasma metabolite data sets. The summary statistics of all 141 plasma metabolites of the three data sets were meta-analyzed and genetically-independent lead variants genetically associated with the level of each metabolite were identified (Methods). Of the 141 plasma metabolites, 126 metabolites had at least three independent lead variants for the MR analysis (Methods). 29 metabolites (including urate) showed evidence of causality with gout (Bonferroni-corrected  $P \leq 3.97 \times 10^{-4}$ ) in either the inverse variance-weighted (IVW) or weighted median (WM) methods, of which four metabolites (3,5-dichloro-2,6-dihydroxybenzoic acid, mannose, myo-inositol, and perfluorooctanoate) showed evidence of pleiotropy (MR-Egger intercept  $P$  of 0,  $2.31 \times 10^{-14}$ ,  $3.05 \times 10^{-11}$ , and 0, respectively; Supplementary table S15). Urate was not pleiotropic with gout (MR-Egger intercept  $P = 0.32$ ). Among the 29 metabolites were lysophosphatidylcholines (LPCs; such as 1-oleoyl-GPC, 1-palmitoleyl-GPC, and 1-palmitoyl-2-palmitoleoyl-GPC), acylcarnitines (such as (R)-3-hydroxybutyrylcarnitine and 3-hydroxyhexanoylcarnitine), amino acid derivatives (such as gamma-glutamylglutamine, N-acetylalanine, and N-acetylglycine), and sugar molecules (such as mannose and myo-inositol).

The effect estimates of two of the three acylcarnitines ((R)-3-hydroxybutyrylcarnitine and 3-hydroxyhexanoylcarnitine) were in a positive direction, indicating that the increased level of these metabolites increase the risk of gout. In contrast, all three lysophosphatidylcholines (1-oleoyl-GPC, 1-palmitoleyl-GPC, and 1-palmitoyl-2-palmitoleoyl-GPC) showed negative effect estimates, suggesting a protective role of LPCs in gout. To evaluate reverse causality, that is gout having a causal effect on the metabolite level, we reversed the MR analysis to test the causality of gout on the metabolite levels and found urate to be the only metabolite that had evidence of causal relationship with gout ( $P_{IVW} = 3.38 \times 10^{-4}$  and  $P_{WM} = 7.77 \times 10^{-3}$ ; Supplementary table S16). However, gout also had evidence of pleiotropy with urate (MR-Egger intercept  $P = 5.67 \times 10^{-4}$ ), reflecting the fact that the gout variants affect both urate levels and gout, making it difficult to determine the direction of causality. Since urate is a known causal risk factor of gout, we also tested the causality of these 28 metabolites (29 metabolites sans urate) with Major et al.<sup>4</sup> urate GWAS and found 24 of 28 metabolites with evidence of causality (Bonferroni-corrected  $P \leq 1.79 \times 10^{-3}$ ; Supplementary

table S17), with gamma-glutamylthreonine, glycocholenate-sulfate, N-acetylalanine, and octadecanedioate showing no evidence of causality with urate. There were six metabolites that showed evidence of pleiotropy (1-oleoyl-GPC, eicosanedioate, homocitrulline, mannose, myo-inositol, and perfluorooctanoate; Supplementary table S17). Testing for reverse causality, there was no evidence of urate having causal effect on the levels of the 28 metabolites (Supplementary table S18), thus 18 metabolites (24 urate-associating metabolites – 6 pleiotropic metabolites) may be causal of gout via their effect on urate.

## Discussion

There have been multiple genome-wide association studies performed for gout<sup>3–8</sup> and even though the most recent genetic study of gout<sup>4</sup> revealed weak evidence of causal role of the clonal hematopoiesis of indeterminate potential (CHIP) *DNMT3A* pathway with gout, definitive implication of causal pathways of gouty inflammation is not clear (aside from genetic variants linked to innate immune cytokines that have an obvious causal link)<sup>4</sup>. Many genetic studies have linked the genetic signals to other functional information (such as expression quantitative trait loci (eQTL)), but very few have explored causal connections with metabolite levels. There have been several direct metabolomics studies that test for observational association of metabolite levels with gout<sup>14–16</sup>, but these studies have only been able to identify a handful of candidate metabolites, most likely due to several factors including sample size and the onerous number of metabolites being explored (and therefore suffering from limited power owing to multiple testing correction). Furthermore, given that these studies are observational, no conclusions regarding causality are able to be drawn.

In this study, we explored the association of metabolite levels with gout by first considering the genetic loci that were identified from the most recent and largest gout GWAS to date, thus narrowing down the search space for candidate metabolites that may be important in gout, and narrowing to metabolites that have some genetic implication in gout. We observed an average of 7.12 metabolites genetically colocalizing with each gout locus and an average of 1.56 gout loci colocalizing with each metabolite, highlighting the fact that a gout genetic locus likely affects several metabolites within the same metabolic pathway, but each metabolite is directly affected by one, or at most a few, gout genetic loci (Table 1). Since we used multiple metQTL data sets, we were able to overlap the presence of metabolites that consistently colocalized with gout among these data sets and curated 141 candidate metabolites that may have a role in gout (Supplementary table S6). Furthermore, by using and comparing with the urine metabolite data set from the GCKD cohort we were able to identify 89 plasma metabolites that may have their concentration affected by metabolic processes, rather than through excretion or reabsorption via the kidney.

In order to identify the most relevant pathways, we first focused on the pathways that included any of the 141 plasma metabolites. We then identified 179 genes present in the colocalized loci that directly affected the breakdown or production of metabolites within the pathways that involved the 141 plasma metabolites (Supplementary table S12). Surprisingly, among the 179 genes, there were only 17 genes at 9 gout loci that

directly metabolized or generated the colocalized metabolites. This suggests that genetic loci identified from GWAS rarely include genes that encode proteins that directly affect the metabolite levels. The 179 genes that were curated from this approach yielded an additional 63 pathways that were not identified by using the list of all of the genes present in any of the colocalized gout loci (Supplementary table S14). For example, estrogen, androgen, and steroid biosynthesis pathways were three of the pathways identified from the analysis. Gout is a disease manifested mainly in men<sup>1,29</sup> and the fact that we identified several pathways that are related to sex hormones and steroid biosynthesis supports these pathways as an underlying biological mechanism that puts men at a higher risk for gout.

Other pathways of note are the amino acid metabolic processes, in particular the glutamine metabolic process. Glutamine has been identified as associated with gout in published metabolomics studies<sup>14,16</sup> and has been reported to be one of the key metabolites in establishing trained immunity via the tricarboxylic acid (TCA) cycle. Trained immunity is a phenomenon where innate immune cells (such as natural killer (NK) cells and monocytes) gain an increased and non-specific immune response by epigenomic remodelling against subsequent infections after being exposed to the initial pathogen<sup>17,30</sup>, and many of the substrates, for example fumarate, succinate, and citrate, promote (e.g. citrate converted to acetyl-CoA thus providing the acetyl group for acetylation<sup>31</sup>) or inhibit (e.g. fumarate directly inhibits KDM5 family of histone demethylases<sup>32,33</sup>) the epigenetic reprogramming that is essential for training. Training by soluble urate has also been experimentally implicated in enhanced monocyte response to MSU crystal<sup>34</sup>. This makes glutamine a strong candidate metabolite involved in the inflammatory process of gout. In fact, rs2638315 (chr12:56.81MB-57.01MB) colocalized with glutamine and is in complete LD with the missense variant rs2657879 for *GLS2*<sup>4</sup> (encodes glutaminase 2) which directly metabolizes glutamine (Supplementary table S13), providing an immediate candidate gene that can be tested for functional validation. *CAD* (encoding carbamoyl-phosphate synthetase 2) at the rs1260326 (chr2:26.91MB-28.71MB) locus colocalized and directly involves glutamine, and is involved in the first few rate-limiting steps of de novo pyrimidine biosynthesis<sup>35</sup>. However, it is not clear whether the gout risk is mediated through changes in glutamine levels and pyrimidine biosynthesis pathway by *CAD* or by another more well-established gout gene, *GCKR*, in the same locus, or both.

Mendelian randomization analysis of the 141 plasma metabolites was carried out to establish causal relationship of the metabolites with gout and to identify directionality of association, identifying 29 metabolites with causal evidence for gout (Supplementary table S15 and S16). Other than urate, some of the metabolites identified were lysophosphatidylcholines (LPCs), acylcarnitines, amino acid derivatives, and sugar molecules. LPCs are phospholipids that can directly bind to toll-like receptors (TLR) 4 and TLR-2/1<sup>36</sup> and, consistent with our MR results, lower plasma level of LPCs have been observed in patients with active rheumatoid arthritis<sup>37,38</sup> and have also been associated with other diseases (such as diabetes, schizophrenia, polycystic ovary syndrome, Alzheimer disease, pulmonary arterial hypertension, aging, asthma, and liver cirrhosis)<sup>39</sup>. However, proposing any inflammatory property of LPCs in gout is complicated as they can either activate or

inhibit the TLR-mediated signaling pathway depending on the absence or presence of classical TLR ligands, respectively<sup>36,39</sup>. TLR-2 and TLR-4 receptors are central in the inflammatory response to MSU crystal where, upon stimulation by an external signal, the receptors induce expression of the proteins involved in formation of the NLRP3 inflammasome, thus mediating the cleavage of pro-IL-1 $\beta$  to pro-inflammatory cytokine IL-1 $\beta$ <sup>40</sup>. There have been no functional experiments conducted for LPCs in the context of gout and establishing their biological role in gout would be of great interest.

Acylcarnitines are a group of molecules formed from carnitine and acyl-coenzyme A (acyl-CoA) and their role is to transport fatty acids from the mitochondrion to the cytoplasm for energy production via  $\beta$ -oxidation. In fact, lipid metabolic processes and  $\beta$ -oxidation (a process where acylcarnitine is sequentially cleaved to generate acetyl-CoA, which is then fed into the TCA cycle for energy production) were identified in the pathway enrichment analysis, supporting the possible role of acylcarnitines in the pathogenesis of gout. A previous study by Kolz et al.<sup>41</sup> associated DL-carnitine and propionyl-L-carnitine with serum urate via a variant in the *SLC16A9* locus (carnitine efflux transporter), where the variant associated with lower levels of DL-carnitine and propionyl-L-carnitine, and higher serum urate. The TCA cycle has a pivotal role in the establishment of innate immunity<sup>17,31,32</sup> through epigenomic remodelling, and acylcarnitines, via generation of acetyl-CoA, could be a plausible causal metabolite of gout through trained immunity. Innate immune cells are trained by epigenetic modifications such as mono- and tri-methylation of lysine 4 of histone 3 (H3K4me1 and H3K4me3) and acetylation of lysine 27 at histone 3 (H3K27ac) at cytokine genes such as *TNFA*, *IL6*, and *TLR4* for rapid and non-specific response to subsequent infections<sup>31,42</sup>. In fact, pathway analysis of genes genetically-associated with gout<sup>4</sup> revealed chromatin modifications as one of the significantly enriched pathways and many histone methyltransferase genes were included in the analysis, suggesting an emerging role of histone modifications in gout most likely via trained immunity. Other mechanisms such as glycolysis<sup>32</sup>, lipid biosynthesis<sup>43</sup>, and the pentose phosphate pathway<sup>17,32</sup> have been implicated in trained immunity, many of which are closely linked to the TCA cycle metabolites. Furthermore, MSU crystal has been demonstrated to act as a danger-associated molecular pattern (DAMP) and elicits an innate immune response and soluble urate can induce epigenetic reprogramming of innate immune cells and contribute to trained immunity<sup>44,45</sup>. Both urate and carnitine derivatives have been causally associated with gout in this study through Mendelian randomization, indicating the possibility of carnitines and urate contributing to trained immunity through the TCA cycle and/or altered levels of serum urate.

There are several limitations to this study. First, the inconsistent names of the metabolites in different metabolomics studies have made it difficult to consolidate specific metabolites in all data sets. Even though we used HMDB to optimize consistent naming across the data sets, not all metabolites were able to be mapped to the names in HMDB. In fact, only about half of the metabolite names in the studies were able to be assigned a consistent name based on HMDB, thus almost half of the metabolites were not assessed for a possible role in gout. Other than to encourage future metabolomics studies to use a consistent metabolite naming scheme, this will be a limitation for any study that uses multiple publicly available

metabolomics studies. A second limitation is the underrepresentation of transporters cataloged in HMDB. In our study, we used the protein information present in HMDB to link the genes (and hence gene product) to the metabolite(s) in order to generate a list of genes that directly affected the 141 plasma metabolites for the pathway enrichment analysis. Since many of the genes that have the strongest association signals in gout are transporters, the consequence of lower representation of transporters in HMDB was large. However, even allowing for the lack of transporter genes, we were able to identify many relevant pathways that involved metabolites that we found to have a causal link to gout from the MR analysis. A third limitation is the difficulty of determining whether the causal metabolites are acting solely through urate or through both urate and the inflammatory response to MSU crystal. Mendelian randomization of metabolite levels on urate was carried out to address this issue (Supplementary table S17 and S18); however, it is difficult to determine with certainty whether the causal mechanism of the metabolites on gout is truly through urate, due to the fact that the underlying genetics of urate and gout has considerable overlap and it is very difficult to distinguish the loci that specifically affect urate from those that specifically affect the risk of gout. Regardless of whether urate mediates the effect of the metabolite levels on gout risk, the 29 metabolites warrant further functional investigation to determine their roles in the mechanism of gout. Lastly, due to the nature of metabolomics studies it is very difficult to gather a large cohort, the sample sizes of the metabolomics studies were relatively small compared to traditional GWAS. Since there were less samples, the power to detect association signals was lower and posed a challenge with lead variant selection for an instrumental variable for association with gout. We ameliorated this by meta-analyzing the three data sets in order to increase the likelihood of identifying more association signals for the MR analysis. In the future, we expect to find more metabolites that associate with gout as larger metabolomics studies are performed.

Metabolomics is a rapidly expanding field in genetics and the wealth of information provided by these studies is extremely valuable. By utilizing the data from metabolomics study, we identified 141 plasma metabolites that genetically colocalized with gout and 29 metabolites that have evidence of a causal relationship with gout, of which acylcarnitines and lysophosphatidylcholines represent strong candidates for future functional studies. These metabolites, especially those that showed causal relationship with gout, have potential as biomarkers.

## References

- [1] Dalbeth, N., Gosling, A. L., Gaffo, A. & Abhishek, A. Gout. *The Lancet* **397**, 1843–1855 (2021).
- [2] Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
- [3] Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nature Genetics* **51**, 1459–1474 (2019).

- [4] Major, T. J. *et al.* A genome-wide association analysis reveals new pathogenic pathways in gout\*. *Nat Genet* 1–15 (2024).
- [5] Li, Z. *et al.* Replication of Gout/Urate Concentrations GWAS Susceptibility Loci Associated with Gout in a Han Chinese Population. *Scientific Reports* **7**, 4094 (2017).
- [6] 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. *Annals of the Rheumatic Diseases* **79** (2020).
- [7] Kawamura, Y. *et al.* Genome-wide association study revealed novel loci which aggravate asymptomatic hyperuricaemia into gout. *Annals of the Rheumatic Diseases* **78** (2019).
- [8] Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, 100192 (2022).
- [9] Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451–459 (2016).
- [10] Li, R., Liang, N., Tao, Y. & Yin, H. Metabolomics in Hyperuricemia and Gout. *Gout, Urate, and Crystal Deposition Disease* **1**, 49–61 (2023).
- [11] Chen, Y. *et al.* Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat Genet* **55**, 44–53 (2023).
- [12] Schlosser, P. *et al.* Genetic studies of paired metabolomes reveal enzymatic and transport processes at the interface of plasma and urine. *Nat Genet* **55**, 995–1008 (2023).
- [13] Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nat Commun* **13**, 1644 (2022).
- [14] Renaudin, F. *et al.* Gout and pseudo-gout-related crystals promote GLUT1-mediated glycolysis that governs NLRP3 and interleukin-1 activation on macrophages. *Annals of the Rheumatic Diseases* **79** (2020).
- [15] Lyu, S. *et al.* Metabolomics analysis reveals four biomarkers associated with the gouty arthritis progression in patients with sequential stages. *Seminars in Arthritis and Rheumatism* **55**, 152022 (2022).
- [16] Joshi, A. D. *et al.* Prediagnostic Glycoprotein Acetyl Levels and Incident and Recurrent Flare Risk Accounting for Serum Urate Levels: A Population-Based, Prospective Study and Mendelian Randomization Analysis. *Arthritis & Rheumatology* **75**, 1648–1657 (2023).
- [17] Arts, R. J. W. *et al.* Glutaminolysis and Fumarate Accumulation Integrate Immunometabolic and Epigenetic Programs in Trained Immunity. *Cell Metabolism* **24**, 807–819 (2016).
- [18] Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

- [19] Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).
- [20] R Core Team. R: A Language and Environment for Statistical Computing (2023).
- [21] Wishart, D. S. *et al.* HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Research* **50**, D622–D631 (2022).
- [22] Cohen, G. J. *et al.* The U.S. syndicated loan market: Matching data. *Journal of Financial Research* **44**, 695–723 (2021).
- [23] Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
- [24] Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13 (2009).
- [25] Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology* **46**, 1734–1739 (2017).
- [26] Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742–015–0047–8 (2015).
- [27] Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- [28] Austin Doyle, L. & Ross, D. D. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). *Oncogene* **22**, 7340–7358 (2003).
- [29] Chen-Xu, M., Yokose, C., Rai, S. K., Pillinger, M. H. & Choi, H. K. Contemporary Prevalence of Gout and Hyperuricemia in the United States and Decadal Trends: The National Health and Nutrition Examination Survey, 2007–2016. *Arthritis & Rheumatology* **71**, 991–999 (2019).
- [30] Netea, M. G., Quintin, J. & van der Meer, J. W. M. Trained Immunity: A Memory for Innate Host Defense. *Cell Host & Microbe* **9**, 355–361 (2011).
- [31] van der Heijden, C. D. *et al.* Epigenetics and Trained Immunity. *Antioxidants & Redox Signaling* **29**, 1023–1040 (2018).
- [32] Arts, R. J. W. *et al.* Immunometabolic Pathways in BCG-Induced Trained Immunity. *Cell Reports* **17**, 2562–2571 (2016).
- [33] Xiao, M. *et al.* Inhibition of -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26** (2012).
- [34] Straton, A. R., Kischkel, B., Crișan, T. O. & Joosten, L. A. B. Epigenomic Reprogramming in Gout. *Gout, Urate, and Crystal Deposition Disease* **2**, 325–338 (2024).

- [35] Li, G., Li, D., Wang, T. & He, S. Pyrimidine Biosynthetic Enzyme CAD: Its Function, Regulation, and Diagnostic Potential. *International Journal of Molecular Sciences* **22**, 10253 (2021).
- [36] Carneiro, A. B. *et al.* Lysophosphatidylcholine Triggers TLR2- and TLR4-Mediated Signaling Pathways but Counteracts LPS-Induced NO Synthesis in Peritoneal Macrophages by Inhibiting NF-B Translocation and MAPK/ERK Phosphorylation. *PLOS ONE* **8**, e76233 (2013).
- [37] Koh, J. H. *et al.* Lipidome profile predictive of disease evolution and activity in rheumatoid arthritis. *Exp Mol Med* **54**, 143–155 (2022).
- [38] Fuchs, B., Schiller, J., Wagner, U., Häntzschel, H. & Arnold, K. The phosphatidylcholine/lysophosphatidylcholine ratio in human plasma is an indicator of the severity of rheumatoid arthritis: Investigations by <sup>31</sup>P NMR and MALDI-TOF MS. *Clinical Biochemistry* **38**, 925–933 (2005).
- [39] Knuplez, E. & Marsche, G. An Updated Review of Pro- and Anti-Inflammatory Properties of Plasma Lysophosphatidylcholines in the Vascular System. *International Journal of Molecular Sciences* **21**, 4501 (2020).
- [40] Major, T. J., Dalbeth, N., Stahl, E. A. & Merriman, T. R. An update on the genetics of hyperuricaemia and gout. *Nature Reviews Rheumatology* **14**, 341–353 (2018).
- [41] Kolz, M. *et al.* Meta-Analysis of 28,141 Individuals Identifies Common Variants within Five New Loci That Influence Uric Acid Concentrations. *PLOS Genetics* **5**, e1000504 (2009).
- [42] Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
- [43] Bekkering, S. *et al.* Metabolic Induction of Trained Immunity through the Mevalonate Pathway. *Cell* **172**, 135–146.e9 (2018).
- [44] Badii, M. *et al.* Urate-induced epigenetic modifications in myeloid cells. *Arthritis Research & Therapy* **23**, 202 (2021).
- [45] Cabău, G., Crișan, T. O., Klück, V., Popp, R. A. & Joosten, L. A. B. Urate-induced immune programming: Consequences for gouty arthritis and hyperuricemia. *Immunological Reviews* **294**, 92–105 (2020).