



## OPEN Meta-analysis of urinary metabolite GWAS studies identifies novel genome-wide significant loci

Jihan K. Zaki<sup>1,2</sup>, Jakub Tomasik<sup>2</sup>, Jade A. McCune<sup>1</sup>, Oren A. Scherman<sup>1</sup>✉ & Sabine Bahn<sup>2</sup>✉

Genome-wide association studies (GWAS) have substantially enhanced the understanding of genetic influences on phenotypic outcomes; however, realizing their full potential requires an aggregate analysis of numerous studies. Here we represent the first comprehensive meta-analysis of urinary metabolite GWAS studies, aiming to consolidate existing data on metabolite-SNP associations, evaluate consistency across studies, and unravel novel genetic links. Following an extensive literature review and data collection through the EMBL-EBI GWAS Catalog, PubMed, and metabolomix.com, we employed a sample size-based meta-analytic approach to evaluate the significance of previously reported GWAS associations. Our analysis identified 48 independent lead SNPs correlated with the levels of 14 unique urinary metabolites: alanine, 3-aminoisobutyrate, betaine, creatine, creatinine, formate, glycine, glycolate, histidine, 2-hydroxybutyrate, lysine, threonine, trimethylamine, and tyrosine. Notably, the results revealed a novel locus for tyrosine (rs4594899, *SLC12A7*,  $P = 6.6 \times 10^{-9}$ ,  $N = 2623$ ), and three newly associated independent SNPs within known loci: one for glycine (rs1755615, *GLDC*,  $P = 2.4 \times 10^{-10}$ ,  $N = 5319$ ), and two for 3-aminoisobutyrate (rs79053399, *RAI14*,  $P = 6.9 \times 10^{-10}$ ,  $N = 4656$ ; rs36071744, *TTC23L*,  $P = 2.97 \times 10^{-10}$ ,  $N = 4872$ ). These findings underscore the potential of urinary metabolite GWAS meta-analyses in revealing novel genetic factors that may aid in the understanding of disease processes and highlight the necessity for larger and more comprehensive future studies.

The identification of peripheral biomarkers holds significant promise for uncovering the mechanisms of complex diseases, improving diagnostic accuracy, and enabling prognosis and personalized treatment. One way to identify biomarkers is through genome-wide association studies (GWASs), which have transformed our knowledge of how genetic differences, typically in the form of single nucleotide polymorphisms (SNP), affect phenotypic outcomes<sup>1</sup>. While most initial GWAS studies focused on identifying genetic variants associated with specific binary traits or disorders<sup>1</sup>, there has been a subsequent increase in focus towards more quantitative outcomes, such as biological analyte concentrations. The first metabolite GWAS study, conducted in 2008<sup>2</sup>, demonstrated the potential to establish robust causal links between genetic risk factors and target metabolites, and emphasized the importance of identifying relevant molecular intermediaries for a more complete understanding of the mechanistic pathways of disease. This argument has become stronger over the past decade, as the assessment of causal inference of metabolite levels on various diseases has become more accessible due to advancements in Mendelian randomization (MR) methodology and analyte measurement methods.

While numerous metabolite GWAS reports have been published since the first study, these have focused primarily on serum and plasma metabolites as opposed to other sample types, such as urine. This trend is exemplified by the fact that a PubMed search for serum and plasma GWAS studies currently returns nearly twenty times more results than for urine-based studies. Nonetheless, since the first urinary metabolite GWAS study was published in 2011<sup>3</sup>, research in this area has been steadily expanding, even though the overall sample sizes have remained relatively similar. In the face of limited large-scale analyses, a potential way to bolster the power of urinary metabolite GWAS studies is through the implementation of meta-analytic methods<sup>4</sup>. The combination of results from multiple studies could help uncover novel associations and strengthen confidence in previously reported findings.

In the present study, we aimed to conduct a comprehensive meta-analysis of existing urinary metabolite GWAS studies to combine all known metabolite-SNP associations into a single resource, assess the consistency of associations between studies, and identify novel genetic associations to urinary metabolite levels. A

<sup>1</sup>Department of Chemistry, University of Cambridge, Lensfield Rd, Cambridge, Cambridgeshire CB2 1EW, UK.

<sup>2</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK. ✉email: oas23@cam.ac.uk; sb209@cam.ac.uk

comprehensive database search was conducted to identify published and proprietary urinary metabolite GWAS data, followed by a sample size-based meta-analysis to identify significant genetic associations.

Results

The current study consisted of a comprehensive literature review of urinary metabolite GWAS studies carried out to date, followed by extensive data collection and subsequent meta-analysis. Of the identified 191 unique studies of interest, 26 satisfied the inclusion criteria concerning the required urinary metabolite GWAS data and its quality. Selected manuscripts are listed in Supplementary Table S1. Following data enquiries where GWAS summary statistics had not been made publicly available, five studies were included in the final, sample size-based meta-analysis. A summary of the data collection process alongside the list of final included studies is shown in Fig. 1.

Description of the included studies

A total of five studies were included in the meta-analysis. The studies are summarized in Supplementary Table S2. The first study, by Nicholson et al.<sup>5</sup>, aimed to identify genetic associations for three urinary analytes in two cohorts, MoTWIN and MoLOBB, comprising 211 participants. Urinary metabolites were measured using <sup>1</sup>H NMR and genotyping was performed using the Illumina 317 K BeadChip SNP array. In the second study, Raffler et al.<sup>6</sup> measured 55 urinary markers using targeted NMR measurements in 3861 participants from the Study of Health in Pomerania (SHIP) and 1691 participants from the Kooperative Gesundheitsforschung in der Region Augsburg (KORA) study. Both analyses used Affymetrix Human SNP Array 6.0 gene chips for genotyping. Sinnott-Armstrong et al.<sup>7</sup> conducted a GWAS study of 4 urine analytes measured using clinical tests in 363 228 UK Biobank participants. Genotyping was conducted using the UK Biobank Axiom Array. Calvo-Serra et al.<sup>8</sup> performed a GWAS study aiming to identify metabolite quantitative trait loci for 44 urinary metabolites measured using proton NMR. The study included 996 children recruited as a part of the Human Early Life Exposome (HELIX) project who were genotyped using the Infinium Global Screening Array (GSA) MD version 1 (Illumina). Finally, Schlosser et al.<sup>9</sup> evaluated genetic associations of 1172 metabolites in 1627 participants from the German Chronic Kidney Disease (GCKD) study and the Study of Health in Pomerania Trend (SHIP-Trend). Metabolites were measured using mass spectrometry, and genotyping was conducted using Illumina Omni2.5Exome BeadChip arrays in the GCKD cohort, and Illumina HumanOmni2.5-Quad in the SHIP-Trend cohort.

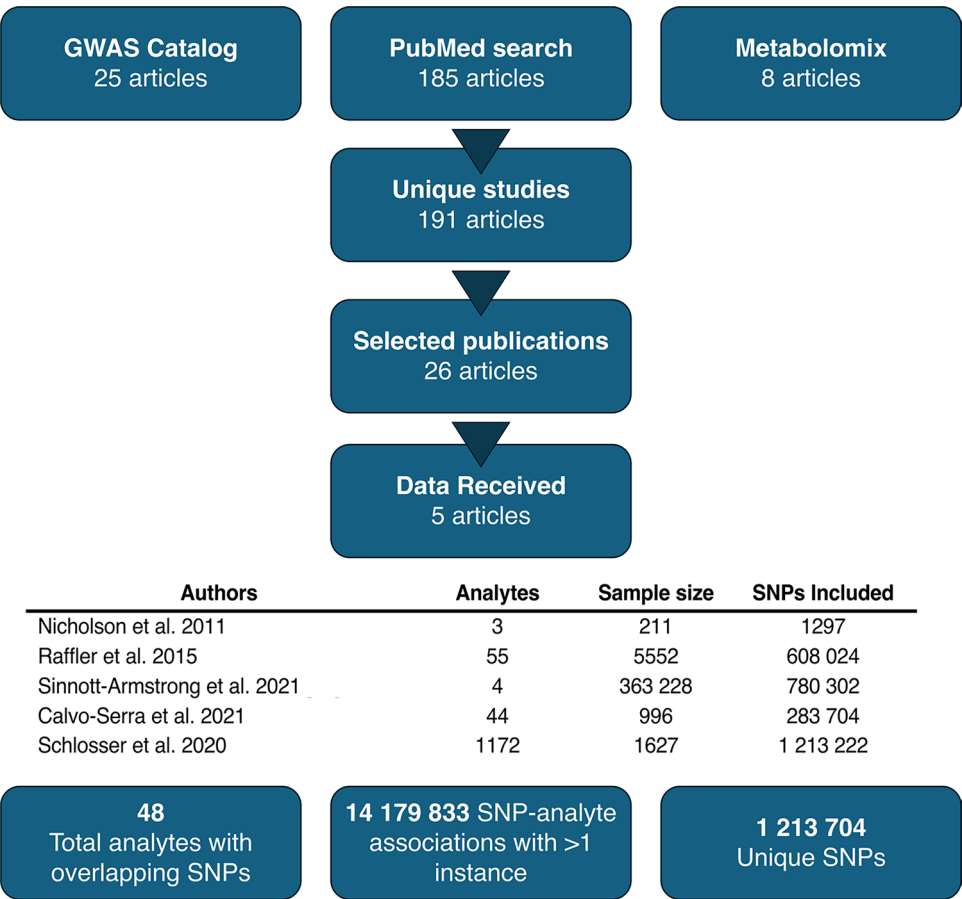


Fig. 1. A summary of the data collection process for the urinary metabolite GWAS meta-analysis.

# GWAS data processing

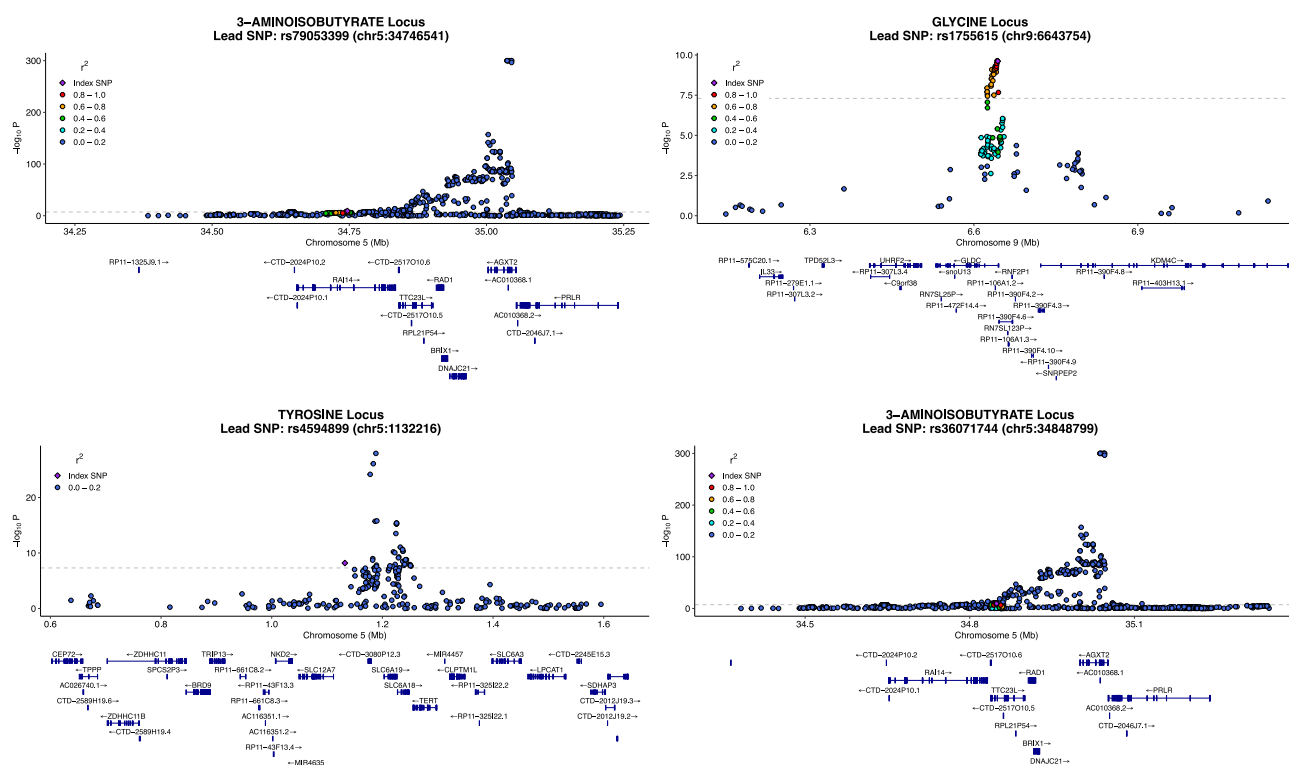
The extracted SNP-analyte association data included 1297 analyte associations to 1297 unique SNPs from the study by Nicholson et al.; 1 799 971 analyte associations to 608 024 unique SNPs from the study by Raffler et al.; 3 113 996 analyte associations with 780 302 unique SNPs from the study by Sinnott-Armstrong et al.; and 12 482 976 analyte associations with 283 704 unique SNPs from the study by Calvo-Serra et al.. Finally, due to the large number of associations in the study by Schlosser et al., only SNP-analyte associations matching the other studies were extracted, which amounted to 1 213 222 unique SNPs and 7 034 023 SNP-analyte associations. Following the removal of associations evaluated only by single studies, 14 179 833 SNP-analyte associations remained in the dataset, corresponding to 1 213 704 unique SNPs and 48 urinary metabolites.

# Meta-analysis of urinary metabolite GWAS studies

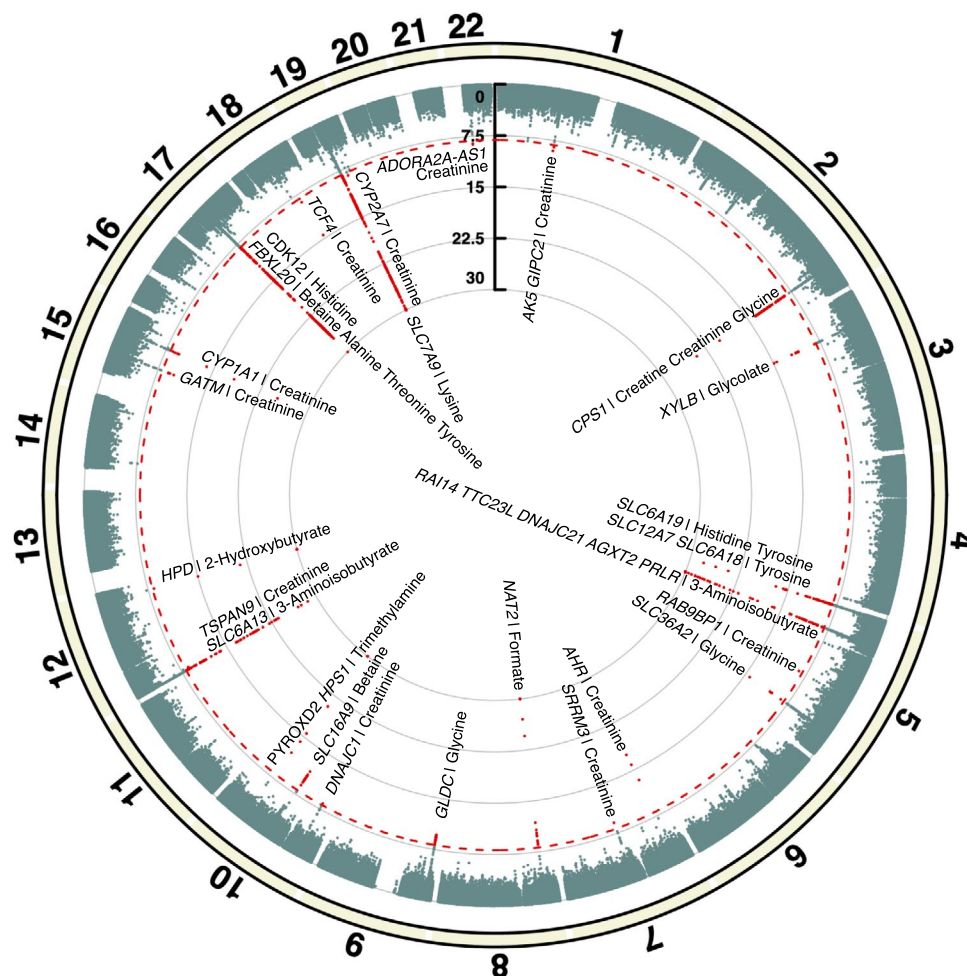
The sample size-based meta-analysis revealed 1226 SNPs with significant ( $P < 7.1 \times 10^{-9}$ ) associations, of which 48 lead SNPs were identified following clumping for 14 urinary analytes. The respective analytes included alanine, 3-aminoisobutyrate, betaine, creatine, creatinine, formate, glycine, glycolate, histidine, 2-hydroxybutyrate, lysine, threonine, trimethylamine, and tyrosine. Details of the significant associations following clumping are shown in Supplementary Table S3. The findings included a novel significant locus for tyrosine (rs4594899, *SLC12A7*,  $P = 6.6 \times 10^{-9}$ ,  $N = 2623$ ), a significant association to glycine (rs1755615, *GLDC*,  $P = 2.4 \times 10^{-10}$ ,  $N = 5319$ ), as well as two significant associations for 3-aminoisobutyrate (rs79053399, *RAI14*,  $P = 6.9 \times 10^{-10}$ ,  $N = 4656$ ; rs36071744, *TTC23L*,  $P = 2.97 \times 10^{-10}$ ,  $N = 4872$ ). An LD assessment between the significant associations for glycine and 3-aminoisobutyrate with previously reported lead SNPs<sup>10</sup> revealed moderate LD ( $r^2 = 0.27$ ) for the *GLDC* SNPs, and low LD for both *RAI14* ( $r^2 = 0.17$ ) and *TTC23L* ( $r^2 < 0.06$ ) SNPs, indicating independent signals within the same gene. The most significant SNP-analyte associations identified were those of rs37369 to 3-aminoisobutyrate ( $P \leq 10^{-308}$ ,  $N = 6145$ ) and rs2147896 to trimethylamine ( $P = 5.0 \times 10^{-165}$ ,  $N = 1207$ ), followed by the associations of rs163908 to 3-aminoisobutyrate ( $P = 2.6 \times 10^{-112}$ ,  $N = 4975$ ) and rs1154510 to 2-hydroxybutyrate ( $P = 1.8 \times 10^{-102}$ ,  $N = 4806$ ). Regional association plots with functional annotation for each novel analyte-SNP association are presented in Fig. 2, and the overall LD assessment of the regions are presented in Supplementary Figure S1.

# Post-hoc analyses

The newly identified loci did not show any pleiotropic effects, as none of them were significantly associated with other phenotypic traits in the PhenoScanner or the GWAS Catalog databases. Furthermore, a relatively large association cluster was identified on chromosome 17, as shown in Fig. 3. The cluster contained significant SNP-metabolite associations to alanine, betaine, histidine, threonine, and tyrosine. A subsequent pathway



**Fig. 2.** Regional association plots for novel SNP associations. The plot shows GWAS association with the target metabolite ( $-\log_{10}$  of  $P$  value, y-axis) for SNPs within a 1 Mb window centered around the lead SNP (purple, diamond). The lead SNP represents the index variant for the novel signal identified following clumping. Points are colored by linkage disequilibrium ( $r^2$ , 1000 Genomes EUR) with the lead SNP. Gene annotations below the plot are based on Ensembl v75 (hg19/GRCh37).



**Fig. 3.** Circular Manhattan plot of all associations identified in the urinary metabolite GWAS meta-analysis. The most significant p-value is presented by each SNP across all 14 analyzed metabolites. Chromosomes are indicated above the circle, numbered through 1–22. Red points on the plot, crossing the red line, indicate significant associations ( $P < 7.1 \times 10^{-9}$ ). Associations exceeding the threshold of  $P < 5 \times 10^{-30}$  have been capped at the threshold for clarity of the plot.

analysis revealed aminoacyl-tRNA biosynthesis as a significant pathway involving four of these metabolites ( $P_{\text{chr17}} = 3.9 \times 10^{-6}$ ). The results of the pathway analysis are presented in Supplementary Figure S2.

## Discussion

The present study aimed to identify novel urinary metabolite loci through a comprehensive meta-analysis of urinary metabolite GWAS studies. To our knowledge, this meta-analysis represents the first such study of urinary metabolite GWASs conducted to date. Three different resources, including the EMBL-EBI GWAS Catalog, PubMed, and metabolomix.com, were searched to identify a set of relevant urinary metabolite GWAS studies, followed by a collection of extensive GWAS summary statistics data. A subsequent sample size-based meta-analysis revealed 48 significant SNP-analyte associations. The results uncovered a novel urinary metabolite locus, *SLC12A7* for tyrosine, and identified three novel independent signals within known loci, including one for glycine (*GLDC*), and two for 3-aminoisobutyrate (*RAI14*, *TTC23L*). Additionally, a region on chromosome 17 was identified to be associated with the levels of multiple urine analytes. Upon further examination, a shared pathway for aminoacyl-tRNA biosynthesis was identified for most of the clustered associations. Most of the newly discovered SNP-analyte associations are consistent with existing literature.

The glycine lead SNP, rs1755615 is an intronic variant in the *GLDC* gene region which encodes glycine dehydrogenase, also known as the P protein. This protein constitutes a central enzyme in the glycine cleavage system (GCS), where its primary role is to catalyze the decarboxylation of glycine<sup>11</sup>. Glycine is a major amino acid, primarily synthesized in the kidneys and the liver, and has significant roles in metabolic regulation, neurological function, and response to oxidative stress<sup>12</sup>. An excess of glycine in the body can lead to glycine encephalopathy, also known as nonketotic hyperglycinemia, a condition primarily caused by dysfunction in the GCS<sup>12</sup>. Additionally, imbalances of glycine and glycine-related compounds have been further linked to various neurobiological disorders such as bipolar disorder<sup>13–15</sup>. Our findings indicate that the genetic influences on the



metabolic glycine degradation system extend to detectable changes in urinary glycine levels. This result could pave the way for future studies exploring the role of peripheral glycine in disease development and brain health.

The second of the identified metabolites with novel associations, 3-aminoisobutyrate, also known as beta-aminoisobutyric acid or BAIBA, is a product of thymine and valine metabolism. Considered an important player in carbohydrate and lipid metabolism, it is strongly upregulated by exercise, and therefore also associated with a lower risk of cardiometabolic diseases<sup>16–19</sup>. rs79053399, is an intronic variant of the retinoic acid induced 14 (*RAI14*) gene which encodes the Rai14 protein, also known as ankycorbin. Rai14 is known to regulate mechanical properties of cells by contributing to membrane protein binding and shaping<sup>20</sup>, and plays a role in actin organization and function. As such, it is involved in processes like cell motility, adhesion, polarity, and cytokinesis<sup>21</sup>, and also acts as a key linking scaffold during micropinocytosis and cell migration<sup>20</sup>. While the mechanistic link between BAIBA and Rai14 is not clear, SNPs within the *RAI14* gene have been previously linked to other metabolite alterations, such as those in creatinine<sup>22</sup> and uric acid levels<sup>23</sup>. This suggests that Rai14 could potentially regulate peripheral levels of multiple analytes in a non-specific manner. However, further research is required to elucidate this possible link.

In turn, the second independent SNP associated with BAIBA, rs36071744, is an intron variant in the *TTC23L* gene encoding the Tetratricopeptide Repeat Protein 23-Like protein (TTC23L). Tetratricopeptide repeat and Tetratricopeptide repeat-like proteins are known to play many roles such as mediating protein–protein interaction and acting as structural protein scaffolding<sup>24</sup>. Past GWAS studies have associated SNPs in the *TTC23L* gene with numerous blood measurement related traits such as red blood cell count<sup>25</sup>, platelet count<sup>26</sup>, erythrocyte count<sup>26</sup>, and reticulocyte count<sup>26</sup>. The exact underlying mechanism connecting urinary BAIBA and Tetratricopeptide repeat-like proteins is unclear, however the combined signals from plasma and urine measurements suggest that the association is robust. For both rs79053399 and rs36071744 it is however notable that they are in close proximity with SNPs within the *AGXT2* gene, which are more significant than both novel associations by an order of magnitude. While the LD reference panel utilized infers that there is little correlation between the two SNPs and the *AGXT2* SNPs, this may not necessarily be the case in the underlying individual level data within the included studies. However, multiple previous associations to urinary BAIBA at the same locus support a true association between *TTC23L* and BAIBA<sup>10</sup>. These associations should be carefully validated in further studies to confirm our observations.

The newly identified locus, rs4594899, is a variant of the *SLC12A7* gene which encodes the Solute carrier family 12 member 7 (SLC12A7) protein, which is a membrane transport protein. SLC12A is known to regulate cell volume through acting as a potassium-chloride transporter, and its upregulation has been associated with numerous cancer types and other diseases such as adrenocortical carcinoma, gynecological and breast cancers, and congenital hydrocephalus<sup>28–31</sup>. Tyrosine is linked with cancers primarily through interactions with tyrosine kinase, an enzyme that phosphorylates and therefore activates tyrosine residues<sup>32</sup>. In turn, tyrosine kinase inhibitors are novel oncological treatments which have been trialed against numerous cancers, however with limited success, especially with regards to adrenocortical carcinoma<sup>33,34</sup>. Interestingly, the rs4594899 SNP resides in close proximity to other, more significant SNPs within the region near other Solute carrier family genes, which could imply that the identified effect may be an artifact. However, broader LD analysis (Supplementary Figure S1) revealed it resides within a tight independent LD block whose constituent SNPs were largely unassessed for tyrosine association. Crucially, rs4594899 shows low LD with other nearby tyrosine-significant SNPs, supporting an independent association signal rather than an artifact driven by linkage to these known loci. Therefore, the identified association should still be considered biologically plausible, and further research could shed light on the metabolomic association between the Solute carrier family proteins and urinary tyrosine.

Among the significant associations confirmed by the present study, a notable cluster of five closely situated loci was observed on chromosome 17. These loci were linked with alanine, betaine, histidine, threonine, and tyrosine urine levels. Pathway analysis indicated that, except for betaine, these analytes are involved in the aminoacyl-tRNA biosynthesis pathway. Aminoacyl-tRNA synthesis is a process where amino acids are attached to the respective transfer-RNA molecules by aminoacyl-tRNA synthetases prior to translation<sup>35,36</sup>. We hypothesize that the identified associations may converge on a common intermediary factor in the aminoacyl-tRNA synthesis pathway, possibly through modulating enzyme activity, thereby affecting the levels of multiple analytes. Alternatively, the findings could reflect underlying correlations among the urinary concentrations of tyrosine, threonine, and histidine, as all three analytes exhibit significant SNP associations in close proximity also elsewhere in the genome (chr 5). The present results highlight the need for further research to unravel the mechanisms behind the clustering of these associations.

Finally, our analysis provides valuable insights for enhancing future urinary metabolite GWAS studies. A pivotal observation from this research is the narrow focus of existing urinary metabolite GWAS studies, which predominantly target metabolites already known to be associated with specific diseases. This is exemplified by the fact that only 48 analytes were eligible for the current meta-analysis, as compared to nearly 4,000 compounds discovered in urine to date<sup>37</sup>. While such a focused approach has inherent value in disease-specific contexts, it severely constrains the general utility of these studies for other applications. In contrast, blood-based GWAS studies often rely on hypothesis-free metabolomic profiling which has proven to be broadly applicable, shedding light on a wide array of physiological and pathological states<sup>2,38,39</sup>. In contrast to blood samples, the non-invasive nature of collecting urinary metabolite readouts makes them promising candidates for broader research applications. Therefore, future studies should adopt a more generalist viewpoint, targeting a wide range of metabolites that can serve a multitude of research objectives. Furthermore, the accessibility of complete GWAS summary statistic data from the assessed studies was very limited, and many were not available even after contacting the authors. Therefore, open data access practices are critical for facilitating more rigorous meta-analyses and cross-disciplinary research.

The interpretation of our findings is subject to several limitations. First, in the absence of better alternatives, this meta-analysis utilized a sample size-based method to calculate statistical significance. This approach, while practical, is not ideal as it does not allow for the assessment of combined effect sizes and may lead to results being disproportionately affected by the largest of the included studies. Consequently, fine-mapping was not feasible due to the limited applicability of current fine-mapping tools to such meta-analysis summary statistics, mainly due to challenges in accurately estimating linkage disequilibrium and potential confounding by inter-study heterogeneity. Second, the number of meta-analyzed studies was low, and their ethnic composition was predominantly Caucasian, which may impact the robustness and generalizability of the findings. Furthermore, we could not report exact demographic details of the analyzed cohorts, as the majority of the studies included did not specify their sub-cohort demographics and instead only referred to the larger cohort studies. Third, the variations in participant demographics, such as young age or impaired kidney function, could have influenced urinary metabolite expression, potentially inflating or masking certain analyte-SNP associations. Finally, urinary analyte concentrations are inherently variable, both between and within individuals, due to factors such as diet, hydration, or activity levels<sup>40,41</sup>. This could significantly influence the outcomes of the original GWAS studies included in our analysis, affecting the reliability of the findings.

## Conclusion

In conclusion, the present meta-analysis of five urinary metabolite GWAS studies identified four novel SNP-metabolite associations, including rs4594899 with tyrosine, rs79053399 and rs36071744 with 3-aminoisobutyrate, and rs1755615 with glycine. The associations involving tyrosine and glycine are biologically plausible and corroborate existing literature, while the potential link between 3-aminoisobutyrate and the proteins Rai14 and TTC23L warrants further exploration. Our study highlights the necessity for an increased scope and larger sample size in future urinary metabolite GWAS studies, as well as the importance of open data access practices.

## Methods

### Metabolite data collection

The present meta-analysis included GWAS data obtained from three different sources, including the EMBL-EBI GWAS catalog under the trait urinary metabolite measurement (EFO-0005116); urinary metabolite GWAS studies compiled by Professor Karsten Suhre on metabolomix.com; and the PubMed library using the search term: ((GWAS) OR (genome-wide association) OR (Mendelian randomization)) AND (urine) AND ((metabolite) OR (analyte) OR (measurement) OR (biomarker)). Studies were considered eligible provided they included healthy participants or did not study known specific metabolite-disorder associations, and had no overlapping participants with other studies. In cases where the GWAS summary statistic data were not publicly available, a data request was sent to the corresponding author(s), followed by a reminder. Only those metabolites that appeared in two or more studies, with two or more overlapping SNPs, were included in the meta-analysis. Metabolites whose identities were uncertain, due to the limitations of untargeted identification methods, were excluded to increase the accuracy of compound identification.

### Data processing

Information was extracted from the obtained GWAS summary statistics data regarding the measured analyte name, chromosome, position, the reference SNP cluster identifier (rsID) tag, effect allele, minor allele, beta coefficient, standard error, p-value, sample size, and effect allele frequency. Each evaluated analyte-SNP combination was harmonized to ensure allele consistency, i.e. SNPs with opposite alleles were flipped by changing the direction of the effect sizes as well as allele frequencies, and SNPs with non-matching effect alleles following flipping were removed.

### Sample size-based meta-analysis

The present analysis was conducted using R version 4.2.0. The meta-analysis utilized a sample size-weighted approach<sup>42</sup>. This method offers substantial flexibility, allowing for the integration of results even in the absence of effect size estimates or when the beta coefficients and standard errors from individual studies are not measured on the same scale. It was selected as the primary analysis method because most of the metabolomic GWAS data in the literature have incomparable effect sizes and standard errors, attributable to inconsistent measurement methods and inconsistent data transformation practices. The method applies Eqs. 1 and 2 to identify a combined Z-score from combined SNP summary statistics, which is then used to calculate the new, composite P-values. In brief, the method applies a weighted sum on calculated Z-scores from each individual study to identify a combined Z-score, where higher Z-scores correspond to lower P-values, and vice versa.

$$Z = \frac{\sum_i \left( \Phi^{-1} \left( \frac{P_i}{2} \right) \times \text{sign}(\beta_i) \times \sqrt{N_i} \right)}{\sqrt{\sum_i \left( \sqrt{N_i^2} \right)}} \quad (1)$$

$$P = 2\Phi(|-Z|) \quad (2)$$

where:  $\beta_i$ : Effect size estimate of study,  $\Phi^{-1}$ : Quantile function,  $\Phi$ : Distribution function,  $N_i$ : Sample size of study,  $P$ : P-value,  $Z$ : Z-score.

## Post-processing and post-hoc analyses

A Bonferroni-corrected significance threshold of  $P < 7.1 \times 10^{-9}$  was applied to adjust for multiple comparisons. To identify the final set of lead SNPs among the significant SNPs, all closely positioned (within 1000 kb) and correlated ( $r^2 > 0.1$ ) significant associations were combined using clumping, with the European population from the 1000 Genomes Project used as a reference panel<sup>43</sup>. Regional association plots for novel loci at each analyte were generated using the LocusZoomR package<sup>44</sup> around the identified lead SNPs (within 1000 kb), and linkage disequilibrium (LD) was assessed using the same reference panel described above. Functional annotation for all lead SNPs was carried out using the Genome Reference Consortium human reference genome build 37 (GRCh37/hg19), consistent with its use across all studies included in the meta-analysis. Comprehensive visualization of LD blocks in proximity to lead SNPs was achieved by generating LD plots incorporating all genotyped SNPs within the region, irrespective of their testing status against target metabolites. A Manhattan plot of all evaluated associations was generated using the CMplot package in R. Potential confounding effects were assessed by reviewing the respective literature and searching the GWAS Catalog and the PhenoScanner database<sup>45,46</sup>. Additionally, a pathway analysis was conducted for metabolites with significant associations clustered in nearby chromosomal regions using the MetaboAnalyst platform<sup>47</sup>. The hypergeometric test was applied as the enrichment evaluation method, and the analysis was based on the Kyoto Encyclopedia of Genes and Genomes Homo sapiens reference pathway library. All significant associations from the meta-analysis were evaluated against past associations in the literature, and each signal at previously reported loci was assessed for independence using LD.

## Data availability

The GWAS summary statistic data used in the present study can be found from the original studies, which are all cited in the manuscript. Direct links to data repositories can be found in Supplementary Table 2.

Received: 29 November 2024; Accepted: 16 June 2025

Published online: 01 July 2025

## References

- Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Primers* **1**, 1–21 (2021).
- Gieger, C. et al. Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).
- Suhre, K. et al. A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* **43**, 565–569 (2011).
- Evangelou, E. & Ioannidis, J. P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).
- Nicholson, G. et al. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.* **7**, e1002270 (2011).
- Raffler, J. et al. Genome-wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality. *PLoS Genet.* **11**, e1005487 (2015).
- Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
- Calvo-Serra, B. et al. Urinary metabolite quantitative trait loci in children and their interaction with dietary factors. *Hum. Mol. Genet.* **29**, 3830–3844 (2021).
- Schlosser, P. et al. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat. Genet.* **52**, 167–176 (2020).
- Valo, E. et al. Genome-wide characterization of 54 urinary metabolites reveals molecular impact of kidney function. *Nat. Commun.* **2024**(16), 1–16 (2025).
- Kikuchi, G., Motokawa, Y., Yoshida, T. & Hiraga, K. Glycine cleavage system: Reaction mechanism, physiological significance, and hyperglycinemia. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **84**, 246–263 (2008).
- Wang, W. et al. Glycine metabolism in animals and humans: Implications for nutrition and health. *Amino Acids* **45**, 463–477 (2013).
- Hoekstra, R. et al. Bipolar mania and plasma amino acids: Increased levels of glycine. *Eur. Neuropsychopharmacol.* **16**, 71–77 (2006).
- Ghasemi, M. et al. The role of NMDA receptors in the pathophysiology and treatment of mood disorders. *Neurosci. Biobehav. Rev.* **47**, 336–358 (2014).
- Zaki, J. K., Tomasik, J., McCune, J., Scherman, O. A. & Bahn, S. Discovery of urinary metabolite biomarkers of psychiatric disorders using two-sample Mendelian randomization. medRxiv 2023.09.26.23296078 (2023) <https://doi.org/10.1101/2023.09.26.23296078>.
- Yi, X. et al. Signaling metabolite  $\beta$ -aminoisobutyric acid as a metabolic regulator, biomarker, and potential exercise pill. *Front. Endocrinol. (Lausanne)* **14**, 1192458 (2023).
- Roberts, L. D. et al.  $\beta$ -aminoisobutyric acid induces browning of white fat and hepatic  $\beta$ -oxidation and is inversely correlated with cardiometabolic risk factors. *Cell Metab.* **19**, 96–108 (2014).
- Begriche, K., Massart, J. & Fromenty, B. Effects of  $\beta$ -aminoisobutyric acid on leptin production and lipid homeostasis: Mechanisms and possible relevance for the prevention of obesity. *Fundam. Clin. Pharmacol.* **24**, 269–282 (2010).
- Tanianskii, D. A., Jarzebska, N., Birkenfeld, A. L., O'sullivan, J. F. & Rodionov, R. N. Beta-aminoisobutyric acid as a novel regulator of carbohydrate and lipid metabolism. *Nutrients* **11**, 524 (2019).
- Lobos Patorniti, N. et al. Rai14 is a novel interactor of Invariant chain that regulates macropinocytosis. *Front Immunol.* **14**, 1182180 (2023).
- Peng, Y. F. et al. Ankycorbin: A novel actin cytoskeleton-associated protein. *Genes Cells* **5**, 1001–1008 (2000).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- Nakatochi, M. et al. Genome-wide meta-analysis identifies multiple novel loci associated with serum uric acid levels in Japanese individuals. *Commun. Biol.* **2**, 115 (2019).
- Perez-Riba, A. & Itzhaki, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr. Opin. Struct. Biol.* **54**, 43–49 (2019).
- Chen, M. H. et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020).
- Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020).

27. Rhee, E. P. et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab.* **18**, 130 (2013).
28. Jin, S. C. et al. SLC12A ion transporter mutations in sporadic and familial human congenital hydrocephalus. *Mol. Genet. Genom. Med.* **7**, e892 (2019).
29. Brown, T. C., Murtha, T. D., Rubinstein, J. C., Korah, R. & Carling, T. SLC12A7 alters adrenocortical carcinoma cell adhesion properties to promote an aggressive invasive behavior. *Cell Commun. Signal* **16**, 27 (2018).
30. Brown, T. C. et al. DNA copy amplification and overexpression of SLC12A7 in adrenocortical carcinoma. *Surgery* **159**, 250–258 (2016).
31. Brown, T. C. et al. Insulin-like growth factor and SLC12A7 dysregulation: A novel signaling hallmark of non-functional adrenocortical carcinoma. *J. Am. Coll. Surg.* **229**, 305–315 (2019).
32. Paul, M. K. & Mukhopadhyay, A. K. Tyrosine kinase—role and significance in cancer. *Int. J. Med. Sci.* **1**, 101 (2004).
33. De Filipo, G., Mannelli, M. & Canu, L. Adrenocortical carcinoma: Current treatment options. *Curr. Opin. Oncol.* **33**, 16–22 (2021).
34. Konda, B. & Kirschner, L. S. Novel targeted therapies in adrenocortical carcinoma. *Curr. Opin. Endocrinol. Diabetes Obes.* **23**, 233 (2016).
35. Gomez, M. A. R. & Ibba, M. Aminoacyl-tRNA synthetases. *RNA* **26**, 910 (2020).
36. Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650 (2003).
37. Wishart, D. S. et al. HMDB 5.0: The human metabolome database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).
38. Shin, S. Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**(6), 543–550 (2014).
39. Nag, A. et al. Effects of protein-coding variants on blood metabolite measurements and clinical biomarkers in the UK Biobank. *Am. J. Hum. Genet.* **110**, 487 (2023).
40. Chen, Y. Variations of human urinary proteome. *Adv. Exp. Med. Biol.* **845**, 91–94 (2015).
41. Nagaraj, N. & Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J. Proteome Res.* **10**, 637–645 (2011).
42. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
43. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. Boughton, A. P. et al. LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics* **37**, 3017–3018 (2021).
45. Kamat, M. A. et al. PhenoScanner V2: An expanded tool for searching human genotype–phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
46. Staley, J. R. et al. PhenoScanner: A database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
47. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37**, W652 (2009).

## Author contributions

Conceptualization: SB, OAS, JT, JKZ; Methodology: SB, OAS, JT, JKZ; Data curation: JKZ; Data analysis: JKZ; Resources: SB, OAS; Writing—Original Draft: JKZ; Writing—Review & Editing: All co-authors; Supervision: SB, OAS, JT; Funding acquisition: SB, OAS, JKZ.

## Funding

This work was supported by the Stanley Medical Research Institute (grant number: O7R-1888) by grants to Sabine Bahn, and by the Oskar Huttunen Foundation grant to Jihan K. Zaki.

## Declarations

## Competing interests

Dr Tomasik has a patent pending for dried blood spot biomarkers for bipolar disorder, and received licensing and consulting fees from Psymics Ltd for unrelated work. Dr Bahn reported grants from Stanley Medical Research Institute and Psymics during the conduct of the study; is a founder and shareholder in Psymics; is Director of Psynova Neuro-tech outside the submitted work; and has a patent pending for dried blood spot biomarkers for bipolar disorder. The other authors have no conflicts to declare.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-07518-0>.

**Correspondence** and requests for materials should be addressed to O.A.S. or S.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025