ARTICLE | Article in Press

A genome-wide association study based on plasma high-throughput untargeted metabolomic profiles expands the number of genetically influenced metabolites in pigs

Samuele Bovo · Matteo Bolner · Giuseppina Schiavo · Francesca Bertolini · Maurizio Gallo · Luca Fontanesi

Received: 21 June 2026 / Accepted: 21 August 2026
© The Author(s) 2026

Abstract

Metabolomics is a powerful phenomics approach for investigating the molecular mechanisms linking genotype to phenotype and identifying biologically informative traits for precision livestock breeding. However, the genetic architecture of the plasma metabolome remains incompletely characterised in pigs. Here, we investigated the genetic determinism of plasma metabolites in 268 Italian Duroc pigs using untargeted high-throughput metabolomics and genome-wide association analyses. Plasma metabolomic analysis identified 580 quality-controlled metabolites, which were integrated with genotypes from a panel of approximately 60,000 single nucleotide polymorphisms. Heritability estimates indicated moderate genetic contribution on the plasma metabolome, with ~46% of metabolites showing heritability greater than 0.20. Overall, 89 metabolite quantitative trait loci (mQTL) regions were identified, involving 156 metabolites and metabolite ratios. Most associations were related to lipid and amino acid metabolism, and several loci harboured biologically plausible candidate genes involved in transport, enzymatic activity, and metabolic regulation. Several mQTL exhibited pleiotropic effects, some metabolites showed evidence of polygenic control, and conditional and ratio-based analyses uncovered additional associations not detected by single-metabolite analyses alone. These findings improve the biological interpretation of metabolic pathways underlying economically relevant traits and support the integration of metabolomics and genomics for precision breeding, feeding, and management strategies in pigs.

Keywords Genetically influenced metabolite · Metabolomics · mGWAS · Plasma · *Sus scrofa*

Introduction

Livestock production systems are currently facing major challenges related to resource use efficiency, environmental sustainability, climate adaptation, and increasing societal expectations concerning animal welfare and production ethics[1]. Over recent decades, pig breeding programmes have achieved substantial genetic progress for

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

https://doi.org/10.1038/s41598-026-68614-3

Published online: 07 September 2026

conventional production traits, such as growth rate, feed efficiency, and carcass composition. However, modern breeding objectives are progressively expanding toward more complex phenotypes, including robustness, resilience, adaptability, welfare, and environmental efficiency. These traits are inherently multifactorial, often difficult to define biologically, and challenging to measure accurately at large scale, thereby limiting the rate of genetic improvement and the biological interpretation of selection responses.

To address these limitations, animal science research is increasingly moving towards the adoption of phenomics approaches aimed at the large-scale and multidimensional characterisation of phenotypic variation[2–4]. Alongside advances in digital phenotyping technologies, increasing attention has been devoted to molecular phenomics approaches, including metabolomics, transcriptomics, and proteomics. These molecular layers represent intermediate phenotypes that connect genomic variability to phenotypic variability, therefore offering biologically informative descriptors of functional processes. Among them, metabolomics occupies a particularly relevant position because the metabolome reflects both the genetic background and the physiological state of the animal, integrating information derived from metabolism, nutrition, environmental interactions, and health status[5,6]. Metabolite profiles can therefore be considered functional "proxy phenotypes" that capture ongoing biological processes more directly than conventional production traits. Consequently, elucidating the genetic architecture of the metabolome may facilitate the identification of biomarkers and molecular pathways useful for developing precision breeding, feeding, and management strategies.

In pigs, metabolomics has increasingly been applied to characterise biological variation across biofluids, tissues, breeds, physiological conditions, and production systems, as well as to investigate the metabolic basis of nutrition, growth efficiency, carcass and meat quality, reproductive performance, physiological adaptation, and stress responses (e.g.[7–17]. These studies have highlighted the ability of metabolite profiles to capture biologically meaningful variation associated with both genetic background and environmental influences (e.g.[18–23].

Parallel to these efforts, metabolite genome-wide association studies (mGWAS) have increasingly explored the genetic regulation of metabolite abundance in different pig populations through the identification of genetically influenced metabolites and their associated genomic regions. Although the currently available literature remains heterogeneous with respect to breeds, developmental stages, analytical platforms, metabolite coverage, and statistical methodologies[20,21,24–26], these studies consistently indicate that a substantial fraction of the metabolome is under genetic influence. Moreover, they suggest that metabolite traits often display relatively large genetic effects, enabling the detection of biologically meaningful metabolite-genome associations even in populations of moderate size compared with conventional genome-wide association studies (GWAS) for complex production traits.

Within this context, the Italian pig production system represents a particularly relevant biological and economical model. The production of Protected Designation of Origin (PDO) cured meat products, relies on specific carcass composition and meat quality traits closely linked to metabolism[27,28]. Italian heavy pig production is economically important and is mainly based on Italian Large White and Italian Landrace breeds as maternal lines, and Italian Duroc as the terminal sire line[27]. In particular, Italian Duroc pigs have undergone long-term selection for growth and carcass performance while maintaining characteristics essential for high-quality cured meat production, including intramuscular fat deposition and specific meat quality attributes[29–31]. Consequently, understanding the molecular and genetic regulation of metabolic traits in this breed may support the identification of biomarkers associated with fat deposition, robustness, and adaptation. This information could have potential applications in precision breeding and feeding strategies specific to this breed.

A recent study provided an initial characterisation of genetically influenced metabolites in Italian Duroc pigs through the analysis of the plasma metabolome using the targeted Biocrates metabolomics platform[20]. In the present investigation, we substantially expanded this characterisation by adopting an untargeted metabolomics approach based on the Metabolon platform. Compared with the previous targeted analysis of approximately 180 metabolites, this platform enabled the detection of nearly 600 metabolites distributed across a much broader range of metabolic classes and biological pathways, thereby considerably increasing metabolomic resolution and

pathway coverage. By analysing plasma samples from more than 250 Italian Duroc pigs, we estimated the heritability of metabolite traits and performed metabolome-wide association analyses to investigate the genetic determinants of metabolite abundance. Through this expanded metabolomic coverage, our objective was to achieve a more comprehensive characterisation of genetically influenced metabolites underlying biologically relevant traits in Italian Duroc pigs.

Methods

Animals, sample collection and processing

The study was carried out in compliance with the ARRIVE guidelines. Animals used in the study were kept in compliance with the Italian and European legislation for pig production (DLgs 122/2011 and Council Directive 2008/120/EC). Ethical review and approval were waived for this study because animals were not specifically raised or treated for the study, and blood samples were collected post-mortem during regular economic slaughter provided in a professional slaughterhouse, with procedures following guidelines established by Italian and European Union regulations for animal care and slaughter (Council Regulation (EC) No 1099/2009).

A total of 268 Italian Duroc (IDU) pigs (191 intact gilts and 77 castrated males) were included in the study. These animals were also part of our previous mGWAS investigation[20]. Briefly, animals originated from the sib-testing program were managed by the National Pig Breeder Association (ANAS). Animals were housed under identical environmental conditions at the same facility and provided feed and water semi-ad libitum until they reached a live weight of approximately 155 ± 5 kg. At the end of the performance test, pigs were transported to a commercial abattoir and slaughtered under standard industrial procedures. Prior to slaughter, animals underwent a fasting period of approximately 12 h. Sampling was conducted over 23 slaughter days, all performed in the morning (around 8:00 a.m.). Blood samples were collected immediately after jugulation by drawing approximately 9 mL of blood from the draining carotid artery into EDTA-containing tubes (Vacutest Kima, Padua, Italy). For each animal, the blood sample was divided into two aliquots: one used for DNA extraction and the other for plasma metabolomic profiling. Plasma was obtained by gently inverting the tubes 8–10 times, followed by centrifugation at 2420 × g for 10 min at 4 °C. Two plasma aliquots were then prepared and stored at −80 °C until further analysis.

Genomic data generation and filtering

Genomic DNA was extracted from whole blood using the Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA), according to the manufacturer's instructions. Genotyping was performed using the Illumina PorcineSNP60 v2 BeadChip, which includes 61,565 single nucleotide polymorphism (SNP) markers. Quality control procedures were conducted using PLINK v1.09[32]. SNPs were retained based on the following criteria: call rate > 0.90, minor allele frequency (MAF) > 0.05, and Hardy-Weinberg equilibrium $P > 0.0001$. Markers located on non-autosomal chromosomes, as well as those with ambiguous mapping or lacking genomic coordinates in the Sscrofa11.1 reference genome assembly, were excluded. After quality control filtering, 33,217 autosomal SNPs remained available for downstream analyses.

Metabolomic data generation and filtering

Untargeted metabolomic profiling was performed on plasma samples from all animals using the HD4 platform (Metabolon, Inc., Morrisville, NC, USA), as previously described by Bovo et al.[19]. Metabolite annotation, including compound names, database identifiers, and pathway classifications (super- and sub-pathways), was provided by Metabolon. In total, 722 metabolites were detected, comprising 654 annotated and 68 unnamed compounds

distributed across nine metabolic super-pathways: lipids, amino acids, nucleotides, peptides, carbohydrates, cofactors and vitamins, energy, partially characterized molecules, and xenobiotics. Xenobiotic compounds ($n=60$) were excluded from further analyses, as they primarily reflect exogenous environmental exposures rather than endogenous metabolic variation. Data processing followed the workflow described in our previous study[19]. Briefly, outlier values were identified and treated as missing data. Metabolites and samples with a high proportion of missing values were subsequently excluded, and the remaining missing observations were imputed using the Multivariate Imputation by Chained Equations (MICE) approach with the Predictive Mean Matching method[33,34]. No samples were removed during quality control, whereas 82 metabolites were excluded. Metabolite abundances were then normalized using a Box-Cox transformation and adjusted for relevant covariates, including sex, carcass weight, and sampling date, through linear regression models, as detailed in Bovo et al.[20]. The resulting residuals were used as phenotypic traits in downstream mGWAS analyses. The final dataset consisted of 268 animals and 580 metabolite traits (residuals).

To further investigate metabolic relationships, all possible pairwise metabolite ratios were calculated in a hypothesis-free manner, resulting in 167,910 ratio traits. These ratio traits were processed using the same normalisation and covariate-adjustment procedures applied to the individual metabolites. All statistical analyses were performed in R v4.2.3[35]; packages *mice* and *MASS*) and Python v3.11.7.

Genome-wide association analyses and genetic parameter estimation

Following our previous study[20], genome-wide association analyses were performed for each metabolite and metabolite ratio (residuals) using a univariate linear mixed model implemented in GEMMA v0.94.1[36]. A centered genomic relationship matrix (GRM) was first computed to account for population structure and relatedness. The GRM was also explored in R v4.2.3[35] via Principal Component Analysis (PCA). The model was defined as: $\mathbf{y}=\boldsymbol{W\alpha}+\boldsymbol{x}\beta+\boldsymbol{g}+\boldsymbol{e}$, where y ($n\times1$) is a vector containing the metabolite feature (residuals) for the n animals, $\mathbf{W}$ ($n\times k$) is a covariate matrix with $k=1$ (a column of 1s) and $\boldsymbol{\alpha}$is the k-dimensional vector of covariates effects, $\mathbf{x}$ ($n\times1$) is the vector containing genotypes for the i^{th} SNP (coded as 0, 1, 2, according to the number of copies of the minor allele), β is the additive fixed effect of the i^{th} SNP on the metabolite level, $\mathbf{g}\sim N(\mathbf{0},\sigma^2_e\,\mathbf{K})$ is a multivariate Gaussian polygenic effect, with covariance matrix proportional to the relatedness matrix $\mathbf{K}$ (n × n) and $\mathbf{e}\sim N(\mathbf{0},\sigma^2_e\,\mathbf{I})$ is a multivariate Gaussian vector of uncorrelated residuals. Statistical significance of SNP effects was assessed using the Wald test as implemented in GEMMA v0.94.1[36]. The genomic inflation factors were calculated with R v.4.2.3[35] and evaluated. For both single-metabolite and ratio-based analyses, statistical significance was first defined using a Bonferroni correction ($\alpha=0.05$), corresponding to $P=2.16\times10^{-6}$. For metabolite ratios, significance was further refined using the *p*-gain statistic[36], estimating the critical level as B / ($\alpha\times2$), where B represents the number of tested ratios. Conditional single-metabolite analyses were subsequently performed by including the lead SNP genotype as an additional fixed effect, allowing confirmation of primary signals and detection of secondary ones.

Genomic heritability (h^2_G) was estimated directly from the mixed model implemented in GEMMA v0.94.1[36]. In parallel, pedigree-based narrow-sense heritability (h^2_A) was estimated in R v4.2.3[35]; package *gap* v1.5-3) using four generations of pedigree information.

Metabolomics data and heritability estimates obtained in our previous study based on the Biocrates platform[20] were compared with those generated in the present study using the Metabolon platform through Pearson correlation (r) analyses. Comparisons were restricted to the subset of overlapping metabolites ($n=61$) and samples shared between the two platforms.

Gene annotation and metabolite QTL regions definition

To identify independent association signals, significant SNPs were pruned by retaining only the top-associated marker within each locus. Loci were defined as SNPs located on different autosomes or separated by more than

10 Mb on the same chromosome. Distinct mQTL regions were subsequently declared when association signals were separated by at least 1 Mb without intervening significant markers.

The proportion of phenotypic variance explained (PVE) by each SNP was estimated following the approach described by Shim et al.[38]. Lead SNPs were annotated by identifying protein-coding genes located within ± 500 kb using the Sscrofa11.1 NCBI GFF genome annotation. Functional interpretation of genes in the genomic windows and associated metabolites was based on literature curation and integration of information from multiple public databases, including UniProt[39], GeneCards[40], the GWAS Catalog[41], the Kyoto Encyclopedia of Genes and Genomes (KEGG;[42], and the Human Metabolome Database (HMDB;[43], all accessed on 10 February 2026. From the GWAS Catalog[41], KEGG[42] and HMDB[43] databases, a list of known gene-metabolite pairs was compiled. For each associated metabolite, all genes located within the corresponding genomic window were screened against this list, and matching genes were prioritized as candidate genes. Although the identified associations remain positional, these matches provided independent biological support for candidate gene prioritization. Finally, all prioritized genes, as well as genomic regions without database-supported matches, were manually reviewed using GeneCards[40], the available literature, and the NCBI's Genome Data Viewer to refine the final list of biologically plausible candidate genes.

Results

Descriptive statistics of the pig plasma metabolome

The untargeted plasma metabolome comprised 722 detected metabolites (Fig. 1a), including 594 named endogenous compounds (82%), 60 named xenobiotics (8%), and 68 unnamed compounds of unknown origin (9%). Endogenous metabolites were distributed across eight metabolic super-pathways, with lipids (49%) and amino acids (30%) representing the largest fractions, followed by nucleotides (7%), peptides (4%), carbohydrates (4%), cofactors and vitamins (4%), energy metabolism (1%), and partially characterized molecules (1%). Overall, 102 metabolic sub-pathways were represented (Supplementary Table S1). Xenobiotic compounds were excluded from downstream mGWAS analyses because they primarily reflect exogenous environmental exposures. Quality control procedures resulted in the removal of 82 metabolites due to a high proportion of missing values, yielding a final dataset of 580 metabolites for subsequent analyses. Notably, the relative distribution of metabolites across super-pathways remained essentially unchanged after filtering (Fig. 1a).

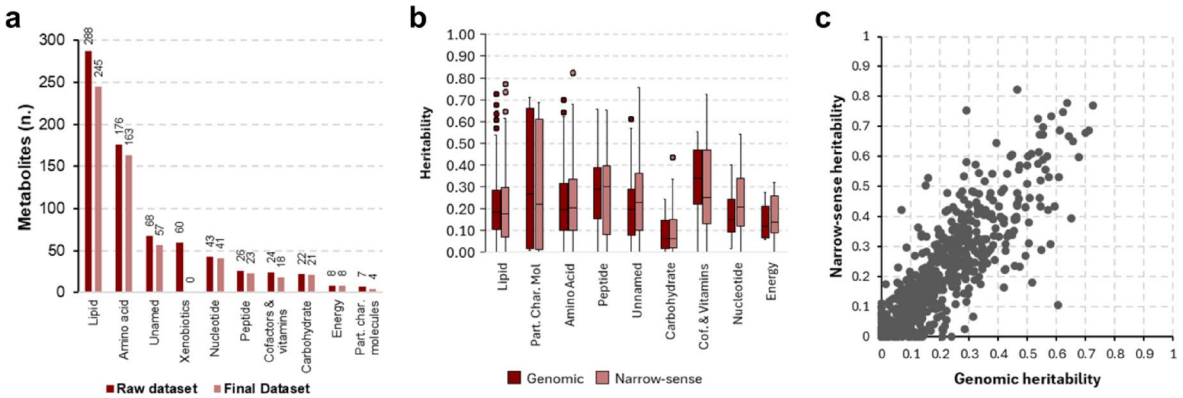

Fig. 1 Metabolites and heritability estimates. (**a**) Distribution of metabolites across super-pathways. (**b**) Distribution of genomic and narrow-sense heritability estimates across super-pathways. (**c**) Correlation between genomic and narrow-sense heritability estimates; each point represents a metabolite.

Heritability estimates for 580 plasma metabolites

The genetic contribution to metabolite variability was assessed by estimating both narrow-sense heritability (h^2_A) and genomic heritability (h^2_G) (Fig. 1b and Supplementary Table S2). Mean estimates were 0.22 ± 0.17 for h^2_A and 0.21 ± 0.15 for h^2_G, with the two measures showing a strong positive correlation ($r = 0.83$; Fig. 1c and Supplementary Table S2). Overall, approximately 74% of the metabolites investigated exhibited $h^2_G > 0.10$, 46% showed values exceeding 0.20, and around 2% displayed high heritability estimates (> 0.60). The highest genomic heritability estimate was observed for a hexosylceramide metabolite ($h^2_G = 0.72 \pm 0.11$).

When heritability estimates were evaluated at the metabolic super-pathway level, metabolites belonging to the carbohydrate super-pathway showed the lowest average values ($h^2_A = 0.10 \pm 0.12$, $h^2_G = 0.11 \pm 0.14$), whereas cofactors and vitamins exhibited the highest average heritability estimates ($h^2_A = 0.31 \pm 0.13$, $h^2_G = 0.32 \pm 0.17$) (Fig. 1b and Supplementary Table S2). To assess the consistency of these estimates, we compared them with results from our previous study based on the Biocrates platform, which included a partially overlapping set of samples and metabolites ($n = 61$). Corresponding heritability estimates showed an average correlation of $r = 0.70$ (Supplementary Table S3).

Overview of the significant metabolite QTL regions

Three complementary mGWAS strategies were applied to characterise the genetic architecture of the plasma metabolome. First, single-metabolite GWAS analyses were performed to identify the primary association signals. Second, conditional analyses were conducted to validate these signals and detect additional independent mQTL. Finally, metabolite ratio-based GWAS analyses were used to further expand the spectrum of detectable genetic associations.

Genomic control inflation factor (λ) values of single metabolite GWAS ranged from 0.86 to 1.11, with a mean of 1.02 (s.d. = 0.03) (Supplementary Table S4), indicating that population stratification (Additional file 2: Figure S2) was adequately controlled.

Significant associations were identified across all *Sus scrofa* chromosomes (SSC), with the exception of SSC16. The highest number of mQTL regions was detected on SSC1 ($n = 10$), followed by SSC7 and SSC8 ($n = 9$ each) (Fig. 2).

Overall, 89 genome-wide significant mQTL regions were identified (Supplementary Table S5), involving 109 individual metabolites (~20% of the metabolomic profile) and 67 metabolite ratios. When metabolites contributing

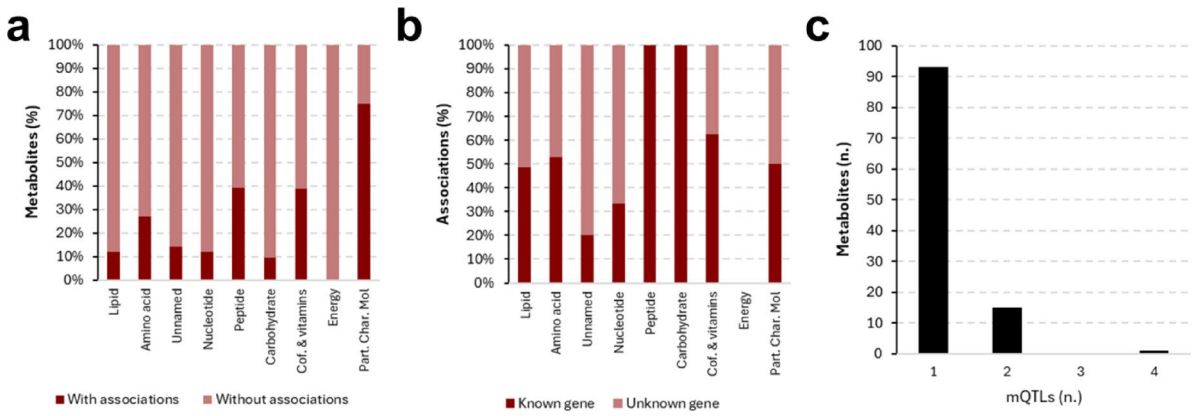

Fig. 2 Metabolite associations and mQTL regions complexity. (**a**) Percentage of metabolites showing significant association signals across metabolite super-pathways. (**b**) Percentage of metabolite-SNP associations linked to known candidate genes or unidentified genes within each super-pathway. (**c**) Distribution of metabolites based on the number of independent mQTL signals.

to significant ratios were also taken into account, the total number of associated metabolites increased to 156 (~27% of the metabolomic profile). Associated metabolites were distributed across all metabolic super-pathways, with the exception of the energy metabolism pathway.

Single-metabolite GWAS signals accounted for 80 of the 89 identified mQTL regions and revealed associations with 109 metabolites. Manhattan plots summarising the mGWAS results are presented in Fig. 3. Conditional analyses identified seven additional SNP-metabolite associations, leading to the discovery of five novel mQTL regions (mQTL #8, #29, #38, #42, and #69; Supplementary Table S5) associated with five metabolites. In addition, two previously identified mQTL regions (mQTL #14 and #64; Supplementary Table S5) were found to be associated with additional metabolites.

Most associated metabolites belonged to the amino acid super-pathway ($n=49$), followed by lipids ($n=37$). Consistently, among the mQTL regions harbouring biologically plausible candidate genes, amino acid-related loci were the most represented ($n=15$), followed by lipid-related regions ($n=10$), while the remaining loci were distributed across the other metabolic super-pathways. However, when considering the proportion of associated metabolites within each super-pathway, partially characterised molecules exhibited the highest fraction of metabolites with significant association signals (Fig. 2a). In contrast, metabolites belonging to the peptide and carbohydrate super-pathways more frequently mapped to mQTL regions containing functionally relevant candidate genes (Fig. 2b).

Most metabolites were associated with a single major mQTL, whereas only a limited number appeared to be under a more polygenic genetic architecture (Fig. 3c). Overall, 23 mQTL regions were associated with two or more metabolites, indicating widespread pleiotropic effects. The strongest pleiotropic signal was observed for mQTL #77 on SSC14 (52.0 Mb), which was associated with eight metabolites (Table 1).

The proportion of variance explained by the lead SNPs averaged 14.2 ± 7.2%, ranging from 7.7% for hypoxanthine to 40.7% for N6,N6-dimethyllysine. The most significant association detected across all analyses was also observed for N6-methyllysine ($P=3.33\times10^{-32}$).

Ratio-based analyses identified nine additional mQTL regions beyond those detected through single-metabolite and conditional GWAS analyses (Supplementary Table S8). Significant metabolite ratios involved combinations of metabolites from multiple super-pathways, although lipid–lipid ratios accounted for the majority of significant associations (77% of all significant pairs). The highest *p*-gain was observed for the ratio tricosanoyl sphingomyelin (d18:1/23:0) / glycosyl ceramide (d18:2/24:1, d18:1/24:2) (*p*-gain $=4.79\times10^{9}$; $P=2.17\times10^{-24}$). In contrast, the most significant association overall by significance level was detected for the ratio 11-palmitoyl-GPC (16:0) / glycosyl ceramide (d18:2/24:1, d18:1/24:2) ($P=9.24\times10^{-33}$; *p*-gain $=2.24\times10^{7}$). Most ratio-derived mQTL regions overlapped with loci identified in the single-metabolite GWAS analyses. Nevertheless, eight regions (#16, #70, #71, #75, #84, #88, #121, and #122; Supplementary Table S5) were uniquely detected through ratio analyses, increasing the total number of identified mQTL regions to 89.

Among all detected mQTL, 37 regions contained genes with biological functions consistent with the associated metabolites (Table 1). Most metabolites were associated with a single major mQTL signal, whereas eight metabolites showed evidence of multiple independent associations. Detailed descriptions of the mQTL regions identified by the different mGWAS approaches, together with their functional relevance within each metabolic super-pathway, are provided in the following sections.

Candidate genes for metabolite QTL regions

About 40% of the identified mQTL regions harboured genes previously implicated in metabolic processes related to the associated metabolites or metabolite ratios (Table 1), according to annotations from multiple gene, protein and pathway databases. In contrast, the remaining regions (Supplementary Tables S6–S8) lacked direct functional evidence or biologically plausible candidate genes linked to the corresponding metabolites, suggesting the presence of potentially novel loci influencing metabolite abundances. Functionally annotated mQTL regions (Table 1) included genes encoding transporters, metabolic enzymes, and proteins involved in structural, binding,

https://doi.org/10.1038/s41598-026-68614-3

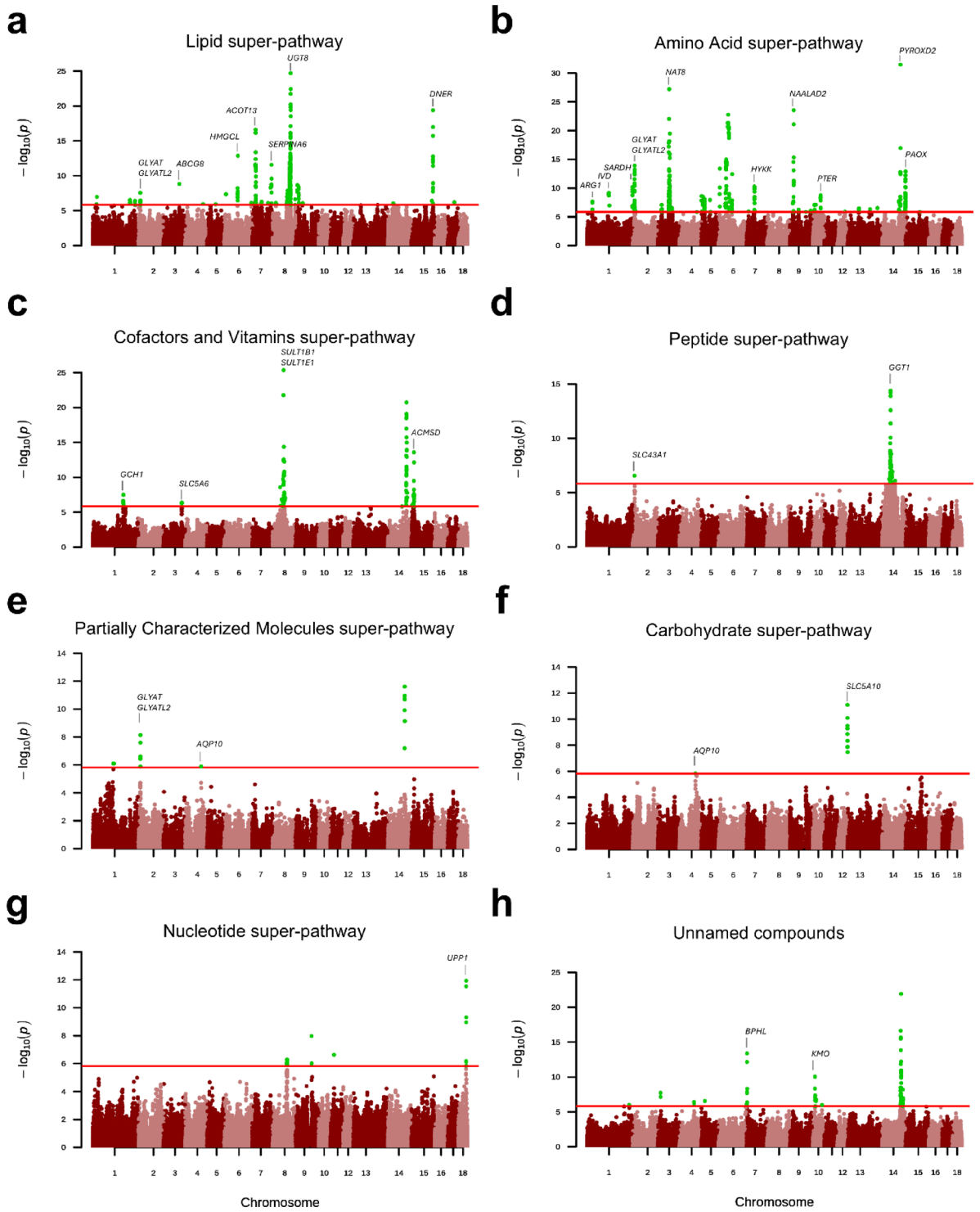

Fig. 3 Manhattan plots of metabolite GWAS results grouped by metabolic super-pathway. (**a**) Lipids; (**b**) Amino acids; (**c**) Cofactors and vitamins; (**d**) Peptides; (**e**) Partially characterized molecules; (**f**) Carbohydrates; (**g**) Nucleotides; (**h**) Unnamed compounds. Each dot represents a SNP. The red horizontal line indicates the genome-wide significance threshold. Significantly associated SNPs are highlighted in green. Candidate genes at selected mQTL regions are reported above the major association peaks.

Table 1 Genome-wide significant metabolite QTL (mQTL) regions with functionally supported candidate genes. Detailed results are reported in Supplementary Tables S5–S7.

mQTL region[a]	SSC: Mb[b]	Candidate gene(s)[c]	Function[d]	*N*. of metabolites (S+*R*)[e]	Metabolite(s) or ratio(s)[f]
4[B–D, LW, M]	1:32	*ARG1*	Hydrolase	1+0	arginine
6 [B–LW,]	1:131.2	*IVD*	Oxidoreductase	1+0	isovalerylcarnitine (C5)
7	1:184	*GCH1*	Hydrolase	1+0	pterin
10 [B–LW]	1:273.2	*SARDH*	Oxidoreductase	1+1	sarcosine
11a	2:11.7–11.9	*SLC43A1*	Transporter	1+0	phenylacetylalanine
11c	2:12.2–12.8	*GLYAT/GLYATL2*	Transferases	4+0	beta-hydroxyisovalerylglycine; isobutyrylglycine; glycine conjugate of C6H10O2;propionylglycine
13	3:26.3	*ACSM3-5/ACSM2B*	Ligase	1+0	phenylacetate
15	3:67.4–73.9	*NAT8*	Transferase	7+1	N-acetylarginine; N-acetyltryptophan; N-acetylasparagine; N-acetylkynurenine; N-acetylphenylalanine; N-acetylleucine; N-acetyltyrosine
16	3:96.9	*ABCG8*	Translocase	1+0	campesterol
17	3:111.9	*SLC5A6*	Transporter	1+0	pantothenate
19	4:95.2–95.5	*AQP10*	Transporter	2+0	arabonate/xylonate; pentose acid
25	5:11.1	*TXN2/FOXRED2*	Oxidoreductases	1+0	6-bromotryptophan
35	6:83.4	*HMGCL*	Lyase	2+0	3-hydroxy-3-methylglutarate; 3-methylglutaconate
37	7:2.3	*BPHL*	Hydrolase	1+0	X-17,676
39	7:19.6	*ACOT13*	Hydrolase	3+0	N-acetyl-2-aminooctanoate; 2-aminooctanoate; 2-hydroxyoctanoate
40a[$;B–D, LW, M]	7:46.1–46.8	*ELOVL5*	Transferase	0+1	1-stearoyl-2-oleoyl-GPS (18:0/18:1)/1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)
40b	7:46.8–47.5	*HYKK*	Transferase	1+1	5-hydroxylysine
41 [B–LW, M]	7:55-55.7	*ABHD2/PLIN1*	Hydrolase / coating protein	2+0	1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6); 1,2-dilinoleoyl-GPC (18:2/18:2)
42 [B–D]	7:87.1	*SV2B/SLCO3A1*	Transporters	1+0	1-palmitoyl-2-oleoyl-GPC (16:0/18:1)*
43 [B–LW, M]	7:94.8	*TMEM229B/RAD51B*	Transmembrane proteins	1+0	1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)
44b[$]	7:109.8	*GALC*	Hydrolase	0+1	sphingomyelin (d18:1/22:2, d18:2/22:1, d16:1/24:2)/lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)
45	7:115	*SERPINA6*	Binding protein	1+0	cortisol
50a	8:66.7–67.5	*SULT1B1/SULT1E1*	Transferases	1+0	gamma-tocopherol/beta-tocopherol
53[P]	8:108.2-109.3	*UGT8*	Transferase	5+48	3-sulfo-palmitoyl-galactosylceramide (d18:1/16:0); glycosyl ceramide (d18:2/24:1, d18:1/24:2); glycosyl ceramide (d18:1/20:0, d16:1/22:0); glycosyl-N-stearoyl-sphingosine (d18:1/18:0); glycosyl-N-palmitoyl-sphingosine (d18:1/16:0)
58b	9:21.2–22.8	*NAALAD2*	Hydrolase	1+1	N-acetyl-aspartyl-glutamate (NAAG)
63	10:8.7	*LYPLAL1*	Hydrolase	1+0	ethylmalonate
64 [B–D, LW, M]	10:12.3–12.4	*KMO*	Oxidoreductase	4+0	X-15,503; N-acetylkynurenine*; kynurenine; kynurenate
66	10:45.5–47	*PTER*	Hydrolase	1+0	N-acetyltaurine
71	12:60.4	*SLC5A10*	Transporter	1+0	1,5-anhydroglucitol (1,5-AG)
76	14:32.6	*ACAD10*	Oxidoreductase	1+0	2-hydroxyoctanoate

Table 1 (continued)

mQTL region[a]	SSC: Mb[b]	Candidate gene(s)[c]	Function[d]	*N*. of metabolites (S+*R*)[e]	Metabolite(s) or ratio(s)[f]
77	14:52.0	*GGT1*	Transferase	8+0	gamma-glutamylhistidine; gamma-glutamyltyrosine; gamma-glutamyl-alpha-lysine; gamma-glutamylleucine; gamma-glutamylglutamate; gamma-glutamylphenylalanine; gamma-glutamylthreonine; gamma-glutamylglycine
79	14:110.5	*PYROXD2*	Oxidoreductase	2+0	N6-methyllysine; N6,N6-dimethyllysine
82	14:140.6-141.2	*PAOX*	Oxidoreductase	2+1	(N(1)+N(8))-acetylspermidine; N-acetyl-isoputreanine
83	15:14.4–15.8	*ACMSD*	Lyase	2+0	quinolinate; 1-methylnicotinamide
85 B–D, LW, M	15:130.8-131.1	*DNER*	Transmembrane protein	1+1	linoleoylcarnitine (C18:2)
89	18:48.9	*UPP1*	Transferase	2+0	5-methyluridine (ribothymidine); uridine

[a]Identifier of the mQTL region. Regions are sorted by chromosome and position. As some mQTL harbour more than one gene known to be functionally related to the specific metabolite, mQTL have also been split as a, b, c, etc. Some mQTL regions have been previously identified by other studies: B=Bovo et al.[20], in Italian Duroc breed (D), Italian Large White breed (ILW) or in a meta-analysis including these two breeds (M); P=Ponsuksili et al.[21]. mQTL discovered via analysis of ratios are marked with the dollar symbol ($).

[b]*Sus scrofa* chromosome (SSC) and the position in Mb of the most significant markers on Sscrofa11.1 genome version.

[c]The candidate genes potentially explaining a functional effect on metabolites are defined as indicated in Methods. The main source of information is provided in Supplementary Tables S6–S8.

[d]Functions are derived from the UniProt database.

[e]Number of associated single metabolites (S) and metabolite ratios (R).

[f]Associated metabolites. Signals are confirmed via conditional mGWAS. With the star symbol (*) are indicated metabolites identified via conditional mGWAS only (novel associations). Metabolite ratios are reported only when the mQTL has been discovered through metabolite ratio analyses. The full list of associated ratios is reported in Supplementary Table S7.

or regulatory roles (Supplementary Table S9). Overall, these findings highlight the heterogeneous genetic architecture underlying metabolite variability, encompassing both well-established metabolic genes and previously uncharacterised genomic regions. Representative examples of the main associations are described below and organised according to metabolic super-pathway.

Lipid super-pathway

Lipid-related signals were the most prominent, with fatty acid-related sub-pathways contributing the largest number of associated metabolites. Several loci previously implicated in lipid metabolism were confirmed, including acyl-CoA thioesterase 13 (*ACOT13*), associated with octanoate derivatives, and the glycine-N-acyltransferase / glycine-N-acyltransferase-like 2 (*GLYAT*/*GLYATL2*) gene region, associated with propionylglycine. The latter association is particularly noteworthy because this metabolite also intersects with branched-chain amino acid (BCAA) metabolism. In addition, variation at the delta/notch-like epidermal growth factor-related receptor (*DNER*) gene region was associated with linoleoylcarnitine (C18:2).

A substantial proportion of association signals originated from the phosphatidylcholine sub-pathway, in which several loci, including the abhydrolase domain-containing protein 2 (*ABHD2*), perilipin 1 (*PLIN1*), and transmembrane protein 229B / RAD51 paralog B (*TMEM229B*/*RAD51B*) gene regions, were implicated. Moreover, a cluster of five glycosylated or galactosylated metabolites, primarily ceramides and sphingosines, were associated with the UDP glycosyltransferase 8 (*UGT8*) gene region. This finding is consistent with the established enzymatic role of UGT8 in catalysing galactosylceramide biosynthesis, thereby reinforcing the biological coherence of this association signal. Additional evidence linked lipid metabolism with ketogenesis, as 3-hydroxy-3-methylglutaryl-CoA lyase (*HMGCL*) was associated with 3-hydroxy-3-methylglutarate. Transport-related effects were also observed. The association between ATP binding cassette subfamily G member 8 (*ABCG8*) and campesterol

is consistent with the known role of the encoded transporter in sterol trafficking, whereas the association between serpin family A member 6 (*SERPINA6*) and cortisol reflects its function as a glucocorticoid-binding protein. For the remaining lipid-related metabolites, the corresponding mQTL regions did not reveal obvious candidate genes based on current annotations, suggesting either incomplete functional knowledge or the involvement of indirect regulatory mechanisms.

Amino acid super-pathway

Among amino acid-related processes, the BCAA sub-pathway was the most prominently represented, with several mQTL regions overlapping those identified for lipid metabolism. The *GLYAT*/*GLYATL2* region again associated with additional glycine derivatives, while *HMGCL* contributed to further association signals, reinforcing the interconnected nature of lipid and metabolic networks. Other loci identified within this sub-pathway corresponded to genes encoding enzymes directly involved in amino acid catabolism, including isovaleryl-CoA dehydrogenase (*IVD*) for isovalerylcarnitine (C5), N-acetyltransferase 8 (*NAT8*) for N-acetylleucine, and lysophospholipase-like 1 (*LYPLAL1*) for ethylmalonate.

Tryptophan metabolism represented the second most enriched amino acid sub-pathway. Associations within this pathway involved genes with well-established biochemical functions, such as kynurenine 3-monooxygenase (*KMO*), associated with kynurenine and kynurenate, and the thioredoxin 2 / FAD dependent oxidoreductase domain containing 2 (*TXN2*/*FOXRED2*) gene region, associated with 6-bromotryptophan. The urea cycle and arginine-proline metabolism constituted the third most represented category, with associations involving arginase 1 (*ARG1*) for arginine, and *NAT8* for its acetylated form. More broadly, *NAT8* showed consistent associations across several other sub-pathways, including tyrosine, phenylalanine, and asparagine metabolisms, suggesting a broader regulatory role in amino acid acetylation and modification. Within lysine metabolism, hydroxylysine kinase (*HYKK*) and pyridine nucleotide-disulfide oxidoreductase domain 2 (*PYROXD2*) were associated with multiple lysine derivatives, whereas phosphotriesterase related (*PTER*), also known as N-acetyltaurine hydrolase, showed a clear association with N-acetyltaurine. Similarly, polyamine oxidase (*PAOX*) was associated with spermidine catabolites, in agreement with its established role in polyamine degradation. Additional biologically coherent signals included associations between N-acetylated alpha-linked acidic dipeptidase 2 (*NAALAD2*) and N-acetyl-aspartyl-glutamate (NAAG), as well as between sarcosine dehydrogenase (*SARDH*) and sarcosine. For a subset of metabolites, however, no clear functional candidate genes could be identified, highlighting either gaps in current pathway annotations or the presence of more complex regulatory architectures.

Cofactors and vitamins super-pathway

Seven metabolites within this super-pathway were associated with six mQTL regions. Among these, GTP cyclohydrolase 1 (*GCH1*) showed a clear association with pterin, consistent with its established role in tetrahydrobiopterin biosynthesis. Likewise, the association between solute carrier family 5 member 6 (*SLC5A6*), a sodium-dependent vitamin transporter, and pantothenate (vitamin B5) reflects the known function of encoded transporter in vitamin uptake. Signals involving aminocarboxymuconate semialdehyde decarboxylase (*ACMSD*) linked quinolinate and 1-methylnicotinamide to the kynurenine pathway and nicotinamide adenine dinucleotide (NAD^+) biosynthesis. This observation is biologically consistent with the role of *ACMSD* in regulating the conversion of intermediates that can be directed toward quinolinate production. In addition, the sulfotransferase family 1B member 1 / sulfotransferase family 1E member 1 (*SULT1B1*/*SULT1E1*) cluster was associated with tocopherol isoforms, in agreement with the reported sulfation activity of SULT enzymes toward these compounds[44]. In contrast, a strong association signal on SSC14:112.4 Mb for pantoate ($P = 1.82 \times 10^{-21}$) lacked an obvious candidate gene, leaving its biologically interpretation unresolved.

Other less represented super-pathways

Within the peptide super-pathway, two loci influenced nine metabolites, primarily related to glutathione metabolism. Several gamma-glutamyl amino acids were associated with the gamma-glutamyltransferase 1 (*GGT1*) gene region, consistent with the central role of the encoded enzyme in glutathione turnover, whereas the *GLYAT*/*GLYATL2* locus was associated with phenylacetylalanine. Partially characterized molecules displayed four distinct associations. Among these, *GLYAT*/*GLYATL2* was associated with a glycine conjugate, while a locus containing aquaporin 10 (*AQP10*) was associated with pentose acid, although no direct functional relationship has yet been established. Within the carbohydrates super-pathway, two metabolites were associated with loci encoding a transporter and an ion channel. Notably, the *AQP10* gene region was associated with arabonate/xylonate, and the solute carrier family 5 member 10 (*SLC5A10*) gene region was associated with 1,5-anhydroglucitol. Both associations are consistent with previous GWAS findings and suggest a role for these genes in carbohydrate transport and homeostasis. In the nucleotide super-pathway, uridine phosphorylase 1 (*UPP1*) was associated with both 5-methyluridine (ribothymidine) and uridine, in agreement with its established role in pyrimidine nucleoside metabolism.

Unnamed molecules

Ten unnamed metabolites showed significant association signals. For two of these compounds, previous evidence reported in the GWAS catalog supported the observed associations[45,46]. Specifically, X-15,503 was associated with kynurenine 3-monooxygenase (*KMO*), suggesting its involvement in the kynurenine pathway, whereas X-17,676 was associated with biphenyl hydrolase-like (*BPHL*). Two particularly strong signals were identified within a ~ 2.3 Mb region on SSC14 (113.9–116.6 Mb), involving X-15,728 ($P = 1.23 \times 10^{-22}$) and X-11,787 ($P = 2.34 \times 10^{-17}$). Despite the strength of these associations, the biochemical identity and functional relevance of these metabolites remains unresolved.

Comparison with the previous biocrates-based metabolite genome-wide association study

The Metabolon and Biocrates platforms shared 61 metabolites included in our previous Biocrates-based mGWAS investigation[20]. For several of these metabolites, overlapping association signals were identified, thereby confirming previously reported mQTL and demonstrating strong cross-platform consistency. Replicated loci included genes with biochemical functions coherent with the associated metabolites. These comprised amino acid-related loci, such as *ARG1* and *KMO*, as well as lipid-related loci, including *PLIN1*, *DNER*, *ELOVL5*, and *SLCO3A1*, among others. In addition, the Metabolon platform enabled the reliable quantification of metabolites that were poorly measured and excluded during quality control in the previous study. For instance, sarcosine was accurately quantified in the present dataset, leading to the identification of a biologically consistent association with the *SARDH* gene.

Discussion

In livestock genomics, increasing attention is being directed toward understanding the biological mechanisms underlying complex phenotypes[47]. Although genomic studies have identified numerous loci associated with economically important traits, the molecular processes linking genotype to phenotype remain only partially understood. Complex traits arise from the interaction of multiple intermediate molecular layers, among which the metabolome represents one of the closest functional readouts of genome activity and physiological regulation. In this context, we investigated the genetic architecture of the plasma metabolome in pigs, a research field that is still

in its early stages, despite recent advances in metabolomic technologies and the increasing number of metabolites detectable in this livestock species[20,21,24–26].

The present study analysed approximately 600 plasma metabolites quantified by mass spectrometry in more than 250 finishing Italian Duroc pigs. Compared with previous porcine mGWAS investigations[20,21, 24–26], this dataset represents a substantial resource in terms of both sample size and metabolite coverage, comparable with studies based on different analytical platforms and pig populations. Nevertheless, previous studies differ markedly in breed composition (purebred versus crossbred populations), developmental stage (from fetal to finishing animals), metabolomic platform, and analytical workflow. Although such heterogeneity complicates direct comparisons across datasets, it also provides the opportunity to identify conserved mechanisms underlying the genetic regulation of metabolism across populations and experimental conditions, while simultaneously generating population-specific information relevant for precision breeding applications. Within this framework, our study provides a comprehensive characterization of metabolite-associated loci in a purebred commercial population, namely the Italian Duroc breed.

A prerequisite for the implementation of metabolomics-informed selection programs is the existence of heritable variation in the traits investigated. In the present study, heritability estimates were moderate overall, averaging 0.22 ± 0.17, with approximately half of the metabolite profile showing $h^2 > 0.20$, indication that the plasma metabolome is characterized by many genetically influenced metabolites. These values are highly consistent with previous reports in pigs, where average metabolite heritability ranged from 0.16 to 0.20[20,21,26]. Furthermore, a subset of the metabolites analysed here had also been investigated in our previous Biocrates-based mGWAS study. For metabolites shared between the two platforms, heritability estimates showed strong agreement ($r = 0.71$), indicating that the genetic component of metabolite variability is robust across metabolomic technologies and supporting the overall stability of the metabolomic signal.

Consistent with these heritability estimates, our results revealed widespread genetic regulation of metabolite abundance through mQTL. By integrating complementary association strategies, including single-metabolite GWAS, conditional analyses, and metabolite ratio-based analyses, we identified 89 mQTL regions associated with 156 metabolites (individual metabolites and metabolites contributing to significant ratios), corresponding to approximately 27% of the investigated metabolome. Most association signals originated from lipid and amino acid pathways, in agreement with their predominance within the plasma metabolome and with previous findings in both pigs and humans. The proportion of variance explained by several lead SNPs was relatively high compared with values typically reported for complex production traits, further supporting the concept that genetically influenced metabolites represent biologically proximal phenotypes positioned closer to gene action. In addition, ratio-based analyses identified loci that were not detected through single-metabolite GWAS alone, indicating that metabolite ratios may capture pathway-level regulation and improve the detection of biologically meaningful biochemical relationships.

The mGWAS analyses provided several relevant insights, including the replication of previously identified loci and the biological interpretation of novel associations. Multiple loci previously identified in our Biocrates-based study[20], including *ARG1*, *KMO*, *DNER*, and *SLCO3A1*, were replicated here, supporting the robustness of the detected genetic signals across metabolomic platforms and the use of the thresholds applied to declare significant results. Conversely, some previously suggested associations were not confirmed. For example, the association between histidine levels and the histidine ammonia-lyase (*HAL*) gene on SSC5 was not detected in the present dataset. However, the signal observed previously was relatively weak, and no suggestive association emerged in the current analyses.

Several novel mQTL were also identified. Among these, *SARDH* associated with sarcosine and *IVD* associated with isovalerylcarnitine (C5) had previously been reported only in Large White populations but were here also detected in Duroc pigs, despite the smaller sample size compared with our previous Biocrates-based Duroc study[20]. This improvement may partly reflect enhanced analytical consistency, as all samples in the present study were processed within a single analytical batch. More broadly, these findings suggest that a substantial fraction

of metabolite regulation is shared across breeds, whereas other loci may contribute to breed-specific metabolic profiles.

The present study also provided additional insights into the kynurenine pathway, a major branch of tryptophan metabolism. The previously identified *KMO* locus associated with kynurenine[20] was replicated and additionally linked to kynurenate, N-acetylkynurenine, and the unnamed compound X-15,503, which had previously been associated with *KMO* in human studies[45,46]. The kynurenine pathway culminates in the production of quinolinate, an intermediate of the de novo nicotinamide adenine dinucleotide (NAD^+) biosynthesis pathway from tryptophan. Quinolinate derives from alpha-amino-beta-carboxymuconate-epsilon-semialdehyde (ACMS), and here the *ACMSD* gene was associated with both quinolinate and 1-methylnicotinamide, extending the evidence for genetic regulation along this pathway. Together, these findings indicate coordinated genetic regulation of multiple branches of tryptophan catabolism and NAD^+ biosynthesis. The kynurenine pathway plays a central role in cellular energy metabolism through NAD^+ synthesis and becomes particularly relevant during immune activation, when energetic demands increase[48]. Moreover, tryptophan is a limiting amino acid in pig nutrition and is directly involved in growth and development[49]. Consequently, understanding the genetic regulation of this pathway may support precision feeding strategies aimed at optimising amino acid utilization while simultaneously informing breeding programs targeting robustness, growth efficiency, and immune resilience.

Another important aspect emerging from this study concerns the genetic regulation of metabolite abundance itself, providing insight into both polygenic and pleiotropic architectures. Most metabolites were characterized by a single major regulatory locus; however, a subset displayed evidence of multiple independent mQTL. For instance, 2-hydroxyoctanoate was associated with both *ACOT13*, involved in CoA ester hydrolysis, and *ACAD10*, which participates in fatty acid β-oxidation. These results suggest that the regulation of individual metabolites may involve multiple components within the same biochemical pathway. Alongside polygenicity, pleiotropic effects were widespread. Several loci acted as major regulators of structurally related metabolites, including *NAT8* for N-acetylated compounds, *GLYAT/GLYATL2* for glycine derivatives, *UGT8* for glycosylated lipids, and *GGT1* for gamma-glutamyl amino acids. Similarly, *HMGCL* contributed to association signals connecting lipid metabolism and branched-chain amino acid metabolism, further highlighting the strong biochemical interconnection among metabolic pathways. Such pleiotropic architectures are particularly relevant for metabolomics-assisted breeding because selection targeting one metabolite or pathway may indirectly influence multiple correlated metabolic traits. This aspect is especially important in the context of precision breeding strategies aimed at improving efficiency, robustness, or resilience while minimizing unintended metabolic trade-offs.

Beyond the genetic architecture itself, several identified loci were biologically relevant in relation to potential applications of metabolite biomarkers. As discussed for the kynurenine pathway, metabolites involved in tryptophan metabolism are closely connected to immune function, growth, and energy homeostasis, traits directly relevant for sustainable livestock production from both economic and animal welfare perspectives. Consistent with welfare-related mechanisms, cortisol levels were associated with *SERPINA6*, a gene involved in the hypothalamic-pituitary-adrenal (HPA) axis and previously linked to glucocorticoid regulation in pigs. The HPA axis has been extensively investigated in this species because of the central role of glucocorticoids in stress responsiveness, immune modulation, and energy metabolism[50–53]. Here, this association was reported for the first time in the Italian Duroc population, suggesting that metabolomics may help identify biomarkers relevant to genetic differences in stress responsiveness and adaptive capacity.

Additional biologically meaningful associations involved vitamin metabolism, lipid pathways, and neurotransmission-related processes. Pantothenate (vitamin B5) was associated with the multivitamin transporter *SLC5A6*. Considering the essential role of pantothenate in coenzyme A synthesis, this finding further supports the extensive genetic regulation of lipid metabolism observed across multiple pathways. Similarly, *UGT8* regulated several glycolipids involved in membrane integrity and cellular signalling[54]. We also identified an association between NAAG, the most abundant dipeptide neurotransmitter in the brain and an important modulator of glutamatergic neurotransmission, and *NAALAD2*, encoding a dipeptidase involved in NAAG metabolism. Because neuronal

activity and neurotransmission are closely linked to behaviour and stress responsiveness, plasma signatures of neuromodulatory pathways may represent informative biomarkers for future precision management applications.

Finally, several association signals involved unnamed metabolites, including X-15,503, X-15,728, and X-11,787. Although the biochemical identity of these compounds remains unresolved, their robust genetic associations strongly suggest that they represent biologically meaningful metabolic traits. These findings highlight the potential of mGWAS not only to characterize known metabolic pathways, but also to contribute to the functional annotation of currently unidentified metabolites.

We also compared the genomic results obtained here with our previous studies investigating breed-related differences in metabolite abundances[19]. Several metabolites showing breed-related variation overlapped with metabolites harbouring major mQTL, supporting the existence of breed-specific genomic backgrounds shaping the metabolome. Interestingly, some identified mQTL were also located close to, or overlapping with, genomic regions previously associated with production traits in the Italian Duroc population[31]. For example, the *NAALAD2* locus associated with N-acetyl-aspartyl-glutamate on SSC9:22.8 Mb was located approximately 0.2 Mb from a genomic window previously associated with average daily gain (SSC9:23.0–24.0 Mb). This observation may suggest a potential neuro-metabolic contribution to growth-related phenotypes in Italian Duroc pigs. Similarly, the *UPP1* locus on SSC18:48.9 Mb overlapped with a region previously associated with lean cut percentage (SSC18:48.0–49.0 Mb).

Despite these findings, some limitations should be considered. Although the sample size is broadly in line with the average size of currently reported for mGWAS in pigs, it remains moderate compared with conventional GWAS, and may have limited the detection minor mQTLs. Moreover, to preserve statistical power in this study, the significance threshold was corrected for the number of SNPs tested but not for the number of metabolites. This approach increased sensitivity for detecting potentially relevant associations, although some signals should be interpreted cautiously. Because many metabolites are biochemically correlated and therefore do not represent 580 fully independent phenotypes, a Bonferroni correction across all metabolites would likely be overly conservative, particularly given the moderate sample size. Nevertheless, future analyses should estimate the effective number of independent metabolite tests, with independent replication needed to confirm the robustness of the reported associations. In addition, while the reported candidate genes are supported by positional associations, biological annotation and replication in both pigs and humans, the underlying causal variants remain to be identified. Fine-mapping approaches based on whole-genome sequencing, followed by functional validation, will be required to distinguish whether the detected associations are driven by regulatory or protein-coding variants. In addition, the integration of additional molecular layers, such as transcriptomics and proteomics, would be needed to further improve the biological interpretation of these loci, linking variability at the genetic level with changes in gene expression and enzyme abundance, as both can contribute to metabolic fluxes and ultimately influence metabolite concentrations. Finally, although the untargeted metabolomics platform substantially expanded metabolome coverage, a proportion of detected compounds remained unidentified, limiting the biological interpretation of some robust association signals.

Conclusions

Our results further confirm that a substantial fraction of the plasma metabolome in the Italian Duroc population is under measurable genetic influence. The identification of numerous biologically coherent mQTL indicates that many metabolites are heritable molecular traits shaped by specific genomic regions, thereby improving our understanding of the molecular mechanisms linking genotype to phenotype in pigs. These findings are directly relevant for precision livestock breeding and nutrition, as metabolite-associated loci may contribute to the identification of biomarkers related to production efficiency, growth, immune competence, robustness, and animal welfare. Overall, the integration of metabolomic and genomic information, leading to the identification of many genetically influenced metabolites, represents a promising strategy to support metabolite-informed breeding,

nutrition, feeding, and management approaches in pigs, ultimately contributing to the development of more sustainable, efficient, and resilient livestock production systems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-68614-3.

Acknowledgements The authors thank personnel of the National Pig Breeders Association (ANAS), Stefania Dall'Olio, Paolo Zambonelli, Anisa Ribani and Valeria Taurisano (University of Bologna) for the collaboration in this study.

Author contributions SB: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing. MB: Data curation; Methodology; Writing – original draft; Writing – review & editing. GS: Data curation; Funding acquisition; Methodology; Writing – review & editing. FB: Data curation; Funding acquisition; Methodology; Writing – review & editing. MG: Data curation; Funding acquisition; Methodology; Resources; Writing – review & editing. LF: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.

Funding This study was supported by: (1) PRIN2017 PigPhenomics that has received funding from the Italian MUR (for the production of the genomic and metabolomic data) (to LF); (2) the Horizon Europe Re-Livestock project that has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement No. 101059609 (for the elaboration and analysis of the data) (to LF and MG); (3) the Italian PRIN2022 project FEEDTHEPIG, Funded by the European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice PRIN 2022 PNRR (DD N. 1409 del 14/09/2022), proposal code P2022FZMJ9 - CUP J53D23018310001 for the analysis of part of the data (to FB); (4) the Italian PRIN2022 HamCapture, funded by the European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice Prin 2022 - DD N. 104 del 2/2/2022, proposal code 202238NP9N – CUP J53D23009570001 (for the analysis of part of the data) (to GS); (5) the Programma di Sviluppo Rurale Nazionale (PSRN) SUIS-2, co-funded by the European Agricultural Fund for Rural Development of the European Union and by the MASAF (for pig sequencing data) (to LF and MG); (6) the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022), within the Agritech National Research Center (Agritech Spoke 1) (to LF and GS).

Data availability None of the data were deposited in an official repository. Data may be available upon request by contacting the corresponding author.

Declarations

Competing interests The authors declare no competing interests.

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/.

References

1. Papakonstantinou, G. I. et al. Precision livestock farming technology: Applications and challenges of animal welfare and climate change. *Agriculture* **14**, 620 (2024).
2. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: The next challenge. *Nat. Rev. Genet.* **11**, 855–866 (2010).
3. Koltes, J. E. et al. A vision for development and utilization of high-throughput phenotyping and big data analytics in livestock. *Front. Genet.* **10**, 1197 (2019).
4. Pérez-Enciso, M. & Steibel, J. P. Phenomes: The current frontier in animal breeding. *Genet. Sel. Evol.* **53**, 22 (2021).
5. Fiehn, O. Metabolomics—The link between genotypes and phenotypes. *Plant. Mol. Biol.* **48**, 155–171 (2002).

6. Fontanesi, L. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Anim. Front.* **6**, 73–79 (2016).
7. Bovo, S. et al. Deconstructing the pig sex metabolome: Targeted metabolomics in heavy pigs revealed sexual dimorphisms in plasma biomarkers and metabolic pathways. *J. Anim. Sci.* **93**, 5681–5693 (2015).
8. Bovo, S. et al. Metabolomics evidences plasma and serum biomarkers differentiating two heavy pig breeds. *Animal* **10**, 1741–1748 (2016).
9. Bovo, S. et al. Comparative targeted metabolomic profiles of porcine plasma and serum. *Animal* **17**, 101029 (2023).
10. Bovo, S. et al. Description of metabolic differences between castrated males and intact gilts obtained from high-throughput metabolomics of porcine plasma. *J. Anim. Sci.* **103**, skaf178 (2025).
11. Qu, H. & Ajuwon, K. M. Metabolomics of heat stress response in pig adipose tissue reveals alteration of phospholipid and fatty acid composition during heat stress. *J. Anim. Sci.* **96**, 3184–3195 (2018).
12. Luise, D. et al. Targeted metabolomic profiles of piglet plasma reveal physiological changes over the suckling period. *Livest. Sci.* **231**, 103890 (2020).
13. Wang, H., Xia, P., Lu, Z., Su, Y. & Zhu, W. Metabolome-microbiome responses of growing pigs induced by time-restricted feeding. *Front. Vet. Sci.* **8**, 681202 (2021).
14. Peukert, M. et al. Sexual dimorphism of metabolite profiles in pigs depends on the genetic background. *Metabolites* **11**, 261 (2021).
15. Zhang, Y. et al. Metabolomic analysis and identification of sperm freezability-related metabolites in boar seminal plasma. *Animals* **11**, 1939 (2021).
16. Hou, X. et al. Metabolomics and lipidomics profiles related to intramuscular fat content and flavor precursors between Laiwu and Yorkshire pigs. *Food Chem.* **404**, 134699 (2023).
17. Metzler-Zebeli, B. U. et al. Creep feeding and weaning influence the postnatal evolution of the plasma metabolome in neonatal piglets. *Metabolites* **13**, 214 (2023).
18. Bovo, S. et al. Comparative analysis of targeted metabolomic profiles reveals plasma metabolite differences across three Italian heavy pig breeds. *Sci. Rep.* **15**, 39296 (2025).
19. Bovo, S. et al. High-throughput untargeted metabolomics reveals metabolites and metabolic pathways that differentiate two divergent pig breeds. *Animal* **19**, 101393 (2025).
20. Bovo, S. et al. Merging metabolomics and genomics provides a catalog of genetic factors that influence molecular phenotypes in pigs linking relevant metabolic pathways. *Genet. Sel. Evol.* **57**, 11 (2025).
21. Ponsuksili, S. et al. Genetic regulation and variation of fetal plasma metabolome in the context of sex, paternal breeds and variable fetal weight. *Open. Biol.* **15**, 240285 (2025).
22. Bertolini, F. et al. Identification of biomarkers for feed efficiency and growth rate by exploring the plasma metabolome of divergent heavy pigs. *Animal* **20**, 101725 (2026).
23. Bolner, M. et al. Comparison of untargeted metabolomic patterns between Italian Landrace and Italian Large White pigs reveals distinct molecular phenotypes. *Ital. J. Anim. Sci.* **25**, 190–205 (2026).
24. Wang, X. & Kadarmideen, H. N. Metabolite genome-wide association study (mGWAS) and gene-metabolite interaction network analysis reveal potential biomarkers for feed efficiency in pigs. *Metabolites* **10**, 201 (2020).
25. Dervishi, E. et al. GWAS and genetic and phenotypic correlations of plasma metabolites with complete blood count traits in healthy young pigs reveal implications for pig immune response. *Front. Mol. Biosci.* **10**, 1140375 (2023).
26. Dervishi, E. et al. Heritability and genetic correlations of plasma metabolites of pigs with production, resilience and carcass traits under natural polymicrobial disease challenge. *Sci. Rep.* **11**, 20628 (2021).
27. Bosi, P. & Russo, V. The production of the heavy pig for high quality processed products. *Ital. J. Anim. Sci.* **3**, 309–321 (2004).
28. Consorzio del Prosciutto di Parma. *Disciplinare di produzione della Denominazione di Origine Protetta Prosciutto di Parma*. (2025). Available at: https://www.prosciuttodiparma.com/wp-content/uploads/2023/09/Disciplinare_Prosciutto_di_Parma_06_03_2023_ITA.pdf
29. Fontanesi, L. et al. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. *J. Anim. Breed. Genet.* **132**, 239–246 (2015).
30. Schiavo, G. et al. Twenty years of artificial directional selection have shaped the genome of the Italian Large White pig breed. *Anim. Genet.* **47**, 181–191 (2016).
31. Bertolini, F. et al. Genome-wide association studies for seven production traits highlight genomic regions useful to dissect dry-cured ham quality and production traits in Duroc heavy pigs. *Animal* **12**, 1777–1784 (2018).
32. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
33. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: What is it and how does it work? *Int. J. Methods Psychiatr Res.* **20**, 40–49 (2011).
34. Faquih, T. et al. A workflow for missing values imputation of untargeted metabolomics data. *Metabolites* **10**, 486 (2020).
35. R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, 2025).

36. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
37. Petersen, A. K. et al. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinform.* **13**, 120 (2012).
38. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One*. **10**, e0120758 (2015).
39. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).
40. Stelzer, G. et al. The GeneCards Suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinf.* **54**, 1301–13033 (2016).
41. Cerezo, M. et al. The NHGRI-EBI GWAS Catalog: Standards for reusability, sustainability and diversity. *Nucleic Acids Res.* **53**, D998–D1005 (2025).
42. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
43. Wishart, D. S. et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).
44. Hashiguchi, T. et al. Enzymatic sulfation of tocopherols and tocopherol metabolites by human cytosolic sulfotransferases. *Biosci. Biotechnol. Biochem.* **75**, 1951–1956 (2011).
45. Yin, X. et al. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nat. Commun.* **13**, 1644 (2022).
46. Chen, Y. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.* **55**, 44–53 (2023).
47. Lappalainen, T., Li, Y. I., Ramachandran, S. & Gusev, A. Genetic and molecular architecture of complex traits. *Cell* **187**, 1059–1075 (2024).
48. Savitz, J. The kynurenine pathway: A finger in every pie. *Mol. Psychiatry*. **25**, 131–147 (2020).
49. Fan, J. et al. Tryptophan regulates food intake in growing pigs by modulating hypothalamic AMPK-mTOR signalling pathway. *Br. J. Nutr.* **133**, 269–276 (2025).
50. Guyonnet-Dupérat, V. et al. Functional implication of an Arg307Gly substitution in corticosteroid-binding globulin, a candidate gene for a quantitative trait locus associated with cortisol variability and obesity in pig. *Genetics* **173**, 2143–2149 (2006).
51. Murani, E., Reyer, H., Ponsuksili, S., Fritschka, S. & Wimmers, K. A substitution in the ligand binding domain of the porcine glucocorticoid receptor affects activity of the adrenal gland. *PLoS One*. **7**, e45518 (2012).
52. Görres, A., Ponsuksili, S., Wimmers, K. & Muráni, E. Analysis of non-synonymous SNPs of the porcine SERPINA6 gene as potential causal variants for a QTL affecting plasma cortisol levels on SSC7. *Anim. Genet.* **46**, 239–246 (2015).
53. Terenina, E., Iannuccelli, N., Billon, Y., Fève, K. & Larzul, C. Genetic determinism of cortisol levels in pig. *Front. Genet.* **16**, 1461385 (2025).
54. Meech, R. et al. The UDP-glycosyltransferase (UGT) superfamily: New members, new functions, and novel paradigms. *Physiol. Rev.* **99**, 1153–1222 (2019).

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

Authors and Affiliations

Samuele Bovo[1] · Matteo Bolner[1] · Giuseppina Schiavo[1] · Francesca Bertolini[1] · Maurizio Gallo[2] · Luca Fontanesi[1]

✉ Samuele Bovo
samuele.bovo@unibo.it

✉ Luca Fontanesi
luca.fontanesi@unibo.it

[1] Animal and Food Genomics Group, Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

[2] Associazione Nazionale Allevatori Suini, Via Nizza 53, 00198 Roma, Italy