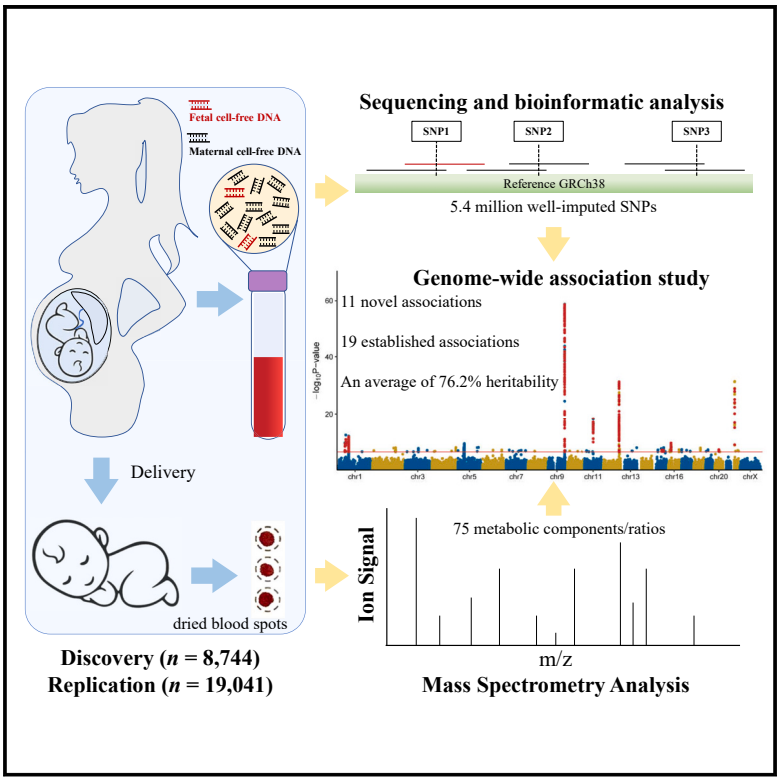
Cell Genomics

Article

A genome-wide association study of neonatal metabolites

Graphical abstract



Highlights

- GWAS of 27,785 low-pass genomes revealed 11 novel associations for neonatal metabolites
- An average of 76.2% heritability for neonatal metabolic components
- Maternal genomes can be utilized in neonatal metabolite GWASs

Authors

Quanze He (贺权泽), Hankui Liu (刘汉奎), Lu Lu (卢露), ..., Lijian Zhao (赵立见), Jianguo Zhang (张建国), Ting Wang (王挺)

Correspondence

jinxin@genomics.cn (X.J.), zhaolijian@genomics.cn (L.Z.), zhangjg@genomics.cn (J.Z.), biowt@njmu.edu.cn (T.W.)

In brief

He et al. imputed 5.4 million SNPs from 27,785 low-pass whole-genome sequencing data obtained through non-invasive prenatal testing in pregnant women. They utilized the SNPs from maternal genomes to identify genetic factors contributing to 75 metabolic components in newborns. The results revealed 19 associations previously established in adults, along with 11 novel associations, and an average heritability of 76.2% for neonatal metabolites.

He et al., 2024, Cell Genomics *4*, 100668
October 9, 2024
https://doi.org/10.1016/j.xgen.2024.100668

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

A genome-wide association study of neonatal metabolites

Quanze He (贺权泽),[1,7,13] Hankui Liu (刘汉奎),[2,5,13] Lu Lu (卢露),[1] Qin Zhang (张芹),[1] Qi Wang (王棋),[1] Benjing Wang (王本敬),[1] Xiaojuan Wu (吴小娟),[1] Liping Guan (管李萍),[2,5] Jun Mao (毛君),[1] Ying Xue (薛莹),[1] Chunhua Zhang (张春花),[1] Xinye Cao (曹昕烨),[3] Yuxing He (贺宇星),[3] Xiangwen Peng (彭向文),[4] Huanhuan Peng (彭欢欢),[5] Kangrong Zhao (赵康容),[1] Hong Li (李红),[1] Xin Jin (金鑫),[6,8,9,12,*] Lijian Zhao (赵立见),[2,5,10,*] Jianguo Zhang (张建国),[2,6,11,*] and Ting Wang (王挺)[1,7,14,*]

[1]The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu Province 215000, China
[2]Hebei Industrial Technology Research Institute of Genomics in Maternal & Child Health, Clin Lab, BGI Genomics, Shijiazhuang 050035, China
[3]Clinical Medicine Department, Xinjiang Medical University, Urumqi, Xinjiang Province 830054, China
[4]Changsha Hospital for Maternal and Child Health Care of Hunan Normal University, Changsha, Hunan Province 431005, China
[5]BGI Genomics, Shenzhen 518083, China
[6]BGI Research, Shenzhen 518083, China
[7]Suzhou Municipal Hospital, Suzhou Jiangsu 215000, China
[8]The Innovation Centre of Ministry of Education for Development and Diseases, School of Medicine, South China University of Technology, Guangzhou 510006, China
[9]Shanxi Medical University-BGI Collaborative Center for Future Medicine, Shanxi Medical University, Taiyuan 030001, China
[10]Medical Technology College, Hebei Medical University, Shijiazhuang 050000, China
[11]School of Public Health, Hebei Medical University, Shijiazhuang 050000, China
[12]Shenzhen Key Laboratory of Transomics Biotechnologies, BGI Research, Shenzhen 518083, China
[13]These authors contributed equally
[14]Lead contact
*Correspondence: jinxin@genomics.cn (X.J.), zhaolijian@genomics.cn (L.Z.), zhangjg@genomics.cn (J.Z.), biowt@njmu.edu.cn (T.W.)
https://doi.org/10.1016/j.xgen.2024.100668

SUMMARY

Genetic factors significantly influence the concentration of metabolites in adults. Nevertheless, the genetic influence on neonatal metabolites remains uncertain. To bridge this gap, we employed genotype imputation techniques on large-scale low-pass genome data obtained from non-invasive prenatal testing. Subsequently, we conducted association studies on a total of 75 metabolic components in neonates. The study identified 19 previously reported associations and 11 novel associations between single-nucleotide polymorphisms and metabolic components. These associations were initially found in the discovery cohort (8,744 participants) and subsequently confirmed in a replication cohort (19,041 participants). The average heritability of metabolic components was estimated to be 76.2%, with a range of 69%–78.8%. These findings offer valuable insights into the genetic architecture of neonatal metabolism.

INTRODUCTION

Metabolites are small molecules that play an important role in biological processes starting from birth.[1,2] Abnormal concentrations of metabolites are potentially harmful in the context of inherited metabolic disorders. In recent years, a multitude of genome-wide association studies (GWASs) have identified hundreds of loci related to metabolites in adults. Several metabolism-related genes underlie inborn errors of metabolism.[3] These findings provide evidence that genetic factors contribute to variations in adult metabolite concentrations.[4–9] In newborns, carriers of 154 pathogenic genes presented abnormal concentrations of metabolites—C3, C3/C2, and C4DC+C5-OH—compared with non-carriers.[10] Despite these findings, the genetic influence on neonatal metabolism remains largely unexplored. Furthermore, metabolism acts as a pivotal link between genetic factors and human traits. For example, *LEP* and *LEPR* loci significantly impact infant growth.[11,12] The leptin-increasing allele of the rs10487505 locus is correlated with a lower infant body mass index (BMI).[12] The transportation of leptin to the brain controls peripheral glucose and lipid metabolism, which, in turn, affects pancreatic function involved in the pathophysiology of obesity and diabetes.[13] Beside anthropometric traits, inherited metabolic illnesses, such as propionic acidemia,[14] mucopolysaccharidoses,[15] and tyrosine hydroxylase deficiency,[16] are caused by pathogenic mutations impacting metabolic pathways.[17] Given the success of newborn screening programs, a combination of exome sequencing and metabolism screening was proposed as an effective diagnostic method to identify inborn errors of metabolism.[18] In the wake of these advances, further research is needed to determine genetic influence on neonatal metabolism.

Currently, routine measurements of DNA in dried blood spots from newborns is not common,[19] posing a significant obstacle for the genetic investigation of neonatal metabolism. Recent improvements in genotype imputation[20,21] have made it possible to obtain reliable genotypes from low-pass whole-genome sequencing (WGS) data of non-invasive prenatal testing (NIPT). Moreover, large-scale low-pass WGS can be used for calculating both GWAS loci with traits and polygenic risk scores (PRSs).[22–24] Previously, we imputed genotypes for 141,431 low-pass genomes from cell-free DNA (cfDNA) from NIPT. This analysis yielded noteworthy findings regarding the genetic associations of height, BMI, and twin pregnancy.[25] By employing a GWAS methodology, which evaluates the direct and/or indirect association between the genetic locus and phenotype,[26] it becomes feasible to link WGS data from the mother with their newborn's phenotypes determined from metabolic screening.

Here, we performed genotype imputation on 27,785 maternal low-pass genomes derived from cfDNA and conducted a two-stage GWAS on 75 metabolic components in neonates. These components included 43 individual metabolites and 32 ratios. We identified 19 previously reported associations and 11 newly discovered associations for 13 neonatal metabolic components. Additionally, we proposed a robust approach for conducting GWASs on neonatal metabolism by using maternal genotypes as a link.

RESULTS

Participant characteristics, sequencing features, variant calling, and genotype imputation

Our study analyzed 27,785 genomes of maternal NIPT collected during a time frame spanning 2015 to 2020. The majority of pregnant women originated from Jiangsu Province, accounting for 60% of the total. The maternal ancestry distribution across different regions of China is as follows: 14% from North China, 79% from Central China, and 7% from South China. A qualification analysis of 43 metabolites, including 11 amino acids, 31 carnitines, and succinylacetone, was conducted on neonatal dried blood spots obtained the third day after delivery. Additionally, 32 ratios were derived based on the measured metabolites (Table S1; Figure S1). By taking into account the systematic bias of sequencing platforms,[25] we designed a two-stage GWAS and integrated the statistical summaries for a meta-analysis. Specifically, the discovery cohort comprised 8,744 genomes sequenced by Ion Torrent, while the replication cohort comprised 19,041 genomes sequenced by Illumina CN500. In the discovery cohort, the average maternal age, maternal height, and neonatal weight were 30.85 ± 4.84 years, 160.87 ± 4.83 cm, and 3,367.51 ± 591.92 g, respectively. Correspondingly, in the replication cohort, these averages were 30.88 ± 4.80 years, 160.80 ± 4.87 cm, and 3,363.36 ± 453.54 g (Table S2).

In the discovery stage, pregnant women underwent low-pass WGS, generating appropriate 5 million single-end reads (Figure S2A) with an average read length of 150 base pairs (bp) (Figure S2B). Reads were aligned to the hg38 reference by bwa software, resulting in an average sequencing depth of 0.28× (Figure S2C) and genome coverage of 25% (Figure S2D). During replication, an average of 5 million single-end reads (each with a read length of 35 bp) resulted in an average sequencing depth of 0.06× and genome coverage of 5% (Figures S2A–S2D). We identified 58 million single-nucleotide variants (SNVs) and estimated the mutant allele frequencies (MuAFs) via BaseVar.[25] For genotype imputation, 21 million single-nucleotide polymorphisms (SNPs) were selected. SNPs with low quality or abnormal depth (Figure S2E) were excluded from this selection. These SNPs had a minor allele frequency (MiAF) > 1% and a transition-to-transversion ratio (Ti/Tv) of 2.12. The imputation process was performed by STITCH[20] utilizing the Han Chinese reference panel of the 1000 Genomes Project. A set of 5.4 million well-imputed SNPs were reserved based on specific criteria: MiAF >1%, p value of Hardy-Weinberg balance test $>1 \times 10^{-5}$, imputation accuracy (I_A) > 0.4, and genotype proportion > 0.9. The majority of common SNPs (91.7%–99.9%) with MiAFs > 5% were found in the dbSNP, ChinaMAP, and/or gnomAD (East Asian) database. Additionally, a percentage (26.1%–36.9%) of low-frequency SNPs with MiAFs < 5% were discovered to be novel (Figure 1A). To estimate the accuracy of imputation, a comparison was made between the MuAFs calculated by STITCH and three other sources of MuAFs, namely the ChinaMAP database, gnomAD database, and BaseVar method. A high density of SNPs was seen along the diagonal line (y = x), accompanied by statistically significant correlations ($r^2 \geq 0.99$, $\beta = 1$) of MuAFs (Figure 1B). These findings indicated that the MuAFs estimated by STITCH can be considered trustworthy. In the process of replication, 19,041 CRAM files were employed to impute the individual genotypes of the 21 million SNPs that were identified during discovery. By applying the same filter criteria, we successfully retained 1.5 million well-imputed SNPs in the replication cohort. Furthermore, we observed a robust MuAF consistency ($r^2 = 0.99$, $\beta = 0.99$) between the initial discovery and subsequent replication stages (Figure 1B).

Utilizing the well-imputed SNPs, we conducted a principal-component analysis (PCA) to discern distinct genetic components among participants from north, central, and south China (Figure 1C). This analysis revealed that the genetics of the participants align with their respective geographical origins (Figure 1D). This finding was consistent with our previous study,[25] which employed genotype data identified by BaseVar. Additionally, we noted a correlation between the second principal component and the proportion of fetal cfDNA, which ranged from 1% to 30% with a median of 10%. This proportion was influenced by the maternal BMI (Figure S3A). This finding indicated that fetal cfDNA has an impact on individual genotypes. Establishing the correlation between the proportion of fetal cfDNA and neonatal metabolism is crucial for identifying the necessary covariates in GWASs. Linear regression analysis did not identify a significant association (Figure S3B), indicating that there is no observable bias for neonatal metabolism GWASs, despite the fact that individual genotypes were affected by the fetal cfDNA. On the other hand, maternal age/height/BMI and neonatal weight/sex were found to strongly influence neonatal metabolism. These factors were utilized as covariates in later GWASs.

Reliable genetic associations of maternal height/BMI

We performed a GWAS to investigate the association between maternal height and BMI. Three SNPs linked to maternal height

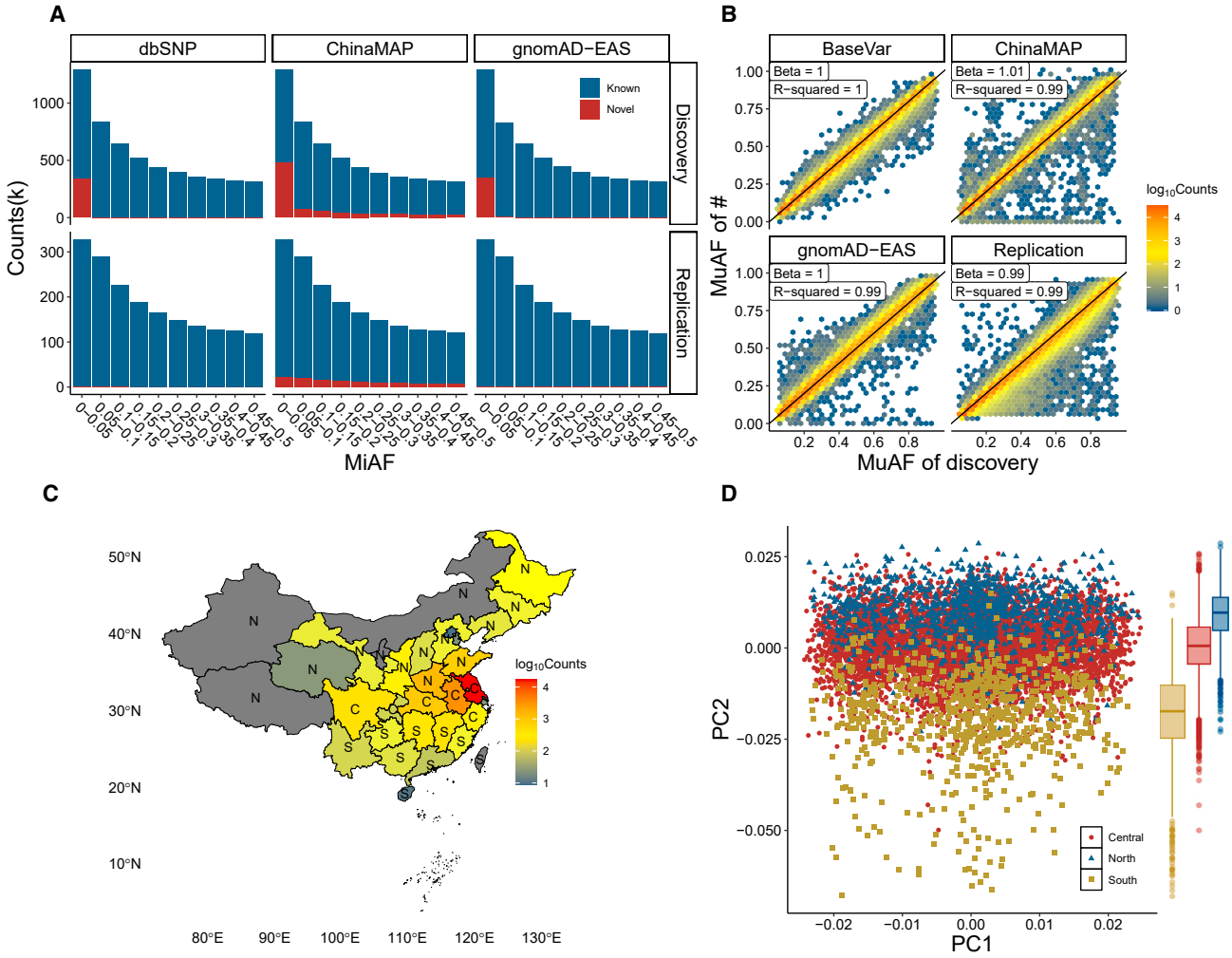


Figure 1. Allele frequency spectrum and genetic structure

(A) Comparison of known and novel SNPs in dbSNP, ChinaMAP, and gnomAD databases.

(B) The accuracy of allele frequencies of well-imputed SNPs.

(C and D) The geographic distribution of participants (C) and their genetic structure (D).

or BMI were found in the discovery phase, which was confirmed in the replication phase (Table S3). Two associations of SNPs, rs7571816 and rs545608, have been reported in the GWAS Catalog, of which[27] *DIS3L2* (rs7571816) is associated with body height[28,29] and *CRYZL2P-SEC16B* (rs545608) is associated with BMI,[30] type 2 diabetes,[31] and body fat.[32] The rs7206410 variant located in the *FTO* gene has been linked to obesity.[33] In our previous study,[25] we also observed an association between rs7206410 and BMI. Collectively, these associations demonstrated the reliability of our study's design and findings.

Genome-wide association of neonatal metabolic components

In order to reveal metabolism-related loci, we established the associations between maternal genetic variants and neonatal metabolic components. The genomic control factors (λ) for 72 metabolic components were <1.1, with the exception of C6DC, C10:1, and C12:1 (Figure S4). This result suggests that the linear regression model, which included neonatal weight/sex, maternal age/height/BMI, and the top three principal components as covariates, effectively adjusted for genetic inflation. Considering a potential additive effect from the maternal effect and neonatal effect (positive correlation for an overestimation and negative correlation for an underestimation), we used the two-stage GWAS to indicate that there is a large/robust effect (Figure S5). Via a threshold of $p < 5 \times 10^{-7}$, we identified 2,628 significant associations between SNPs and metabolic components. Among these associations, 1,863 were seen in the replication cohort. Furthermore, 1,162 associations were successfully replicated by a threshold of $p < 2.68 \times 10^{-5}$ (Bonferroni adjustment for the 1,863 associations). In the meta-analysis, a total of 1,104 associations, consisting of 414 SNPs, were indicated as statistically significant based on a threshold of $p < 6.67 \times 10^{-10}$. This threshold was determined by taking the genome-wide significant

threshold of 5×10^{-8} into account for 75 components. To determine the lead SNP, we selected the SNP exhibiting the highest level of statistical significance, as shown by the *p*-value, inside/near a gene. Subsequently, we kept 30 associations pertaining to 13 neonatal metabolic components (Table 1; Figure 2).

Here, we have linked maternal SNPs and neonatal metabolism. In order to establish the genetic links for metabolism in the neonatal population, we employed a trend test[34] to test the associations between the neonatal allele frequency and the order of different phenotypic groups. Based on the rankings of the quantitative metabolic components, the neonatal participants were categorized into five distinct groups (Figure 3A). In accordance with Hardy-Weinberg balance, the allele frequency in the offspring population should be equivalent to the allele frequency in the parental population. The SNPs associated with metabolism did not fail the Hardy-Weinberg balance test, indicating that they conform to the expected genetic equilibrium. Via a high-depth WGS dataset of 602 trios,[35] we confirmed a significant correlation between the offspring and parental population for each lead SNP (Figure S6A). Furthermore, an analysis of the allele frequencies in the gnomAD database revealed that these frequencies are equal between males and females (Table S4). Based on these facts, we devised an approach (STAR Methods) to estimate the neonatal allele frequency from the maternal allele frequency and gnomAD database, which provides allele frequencies in both males and females. The accuracy of these estimations was subsequently validated by the WGS data of 602 trios (Figure S6B). Through the utilization of a trend test, we demonstrated a positive correlation between the increase of neonatal allele frequencies and the increase of metabolic concentration across the five ranked groups (Figure 3B). These findings offer statistical evidence for the validation of the 30 associations between the unobserved neonatal genotypes and metabolic components.

Among the 30 associations, 19 associations (including 16 genes: *ACADM*, *ASB17*, *CPT2*, *ACADS*, *MAGOH*, *MLEC*, *MSH4*, *CABP1*, *UNC119B*, *SLC44A5*, *RABGGTB*, *IER5L*, *CRAT*, *RPS27L*, *PTPA*, *PRODH*) were documented in the GWAS Catalog (Figure 4A; Table S4). Out of the 11 novel associations, nine associations involving genes (*BCL7A*, *MSH4*, *CYP4A11*, *PPP6R3*, *ACSM2A*, *HPD*, *PSMD9*, and *CFAP251*) were associated with diverse metabolic components. On the other hand, two associations—rs6142615 (*TM9SF4*) and rs75854125 (*LRP8*)—have not been reported in the GWAS Catalog with the involvement of *TM9SF4* in metabolism identified here. The gene *LRP8* is a neighbor to a known metabolism-related locus (*CPT2-MAGOH*). The presence of known SNPs and genes associated with metabolism demonstrated that the maternal NIPT WGS data may have the potential to investigate the genetic associations of neonatal metabolites.

Maternal genetic effect on neonatal metabolic components

The strongest association (β = 0.32 SD ± 0.036 SE) across all metabolic components is the intronic variant rs695948 at the *UNC119B* gene associated with the C4 concentration (Figure 4B; Table S4), in agreement with a previous study.[36] SNP rs1146593 located at the *ACADM* gene is associated with two metabolites, C6 ($\beta = -0.12$ SD ± 0.018 SE) and C8 ($\beta = -0.11$ SD ± 0.018 SE), as well as the ratio of (C3DC+C4-OH)/C10 ($\beta = -0.12$ SD ± 0.018 SE). Additionally, the SNP rs12832102 located at the *ACADS* gene is associated with C4 ($\beta = -0.24$ SD ± 0.028 SE). These associations were known for *ACADM* and *ACADS* in healthy individuals and those with chronic kidney disease.[8,37–39] The *ACADM* and *ACADS* genes encoded medium-chain and short-chain acyl-coenzyme A (CoA) dehydrogenase enzymes, respectively. Deficiency in these enzymes, characterized by abnormal levels of carnitine (Table S5), may confer an increased risk for chronic kidney disease.[40] Furthermore, the SNPs—rs12082305 (*MAGOH*), rs2229291(*CPT2*), and rs75854125 (*LRP8*)—are associated with C6DC with effect sizes of 0.15 SD ± 0.013 SE, 0.14 SD ± 0.021 SE, and 0.14 SD ± 0.021 SE, respectively. While the associations of *CPT2* and *MAGOH* were aligned to C6DC in previous studies,[8,9] the association of *LRP8* and C6DC is a novel finding.

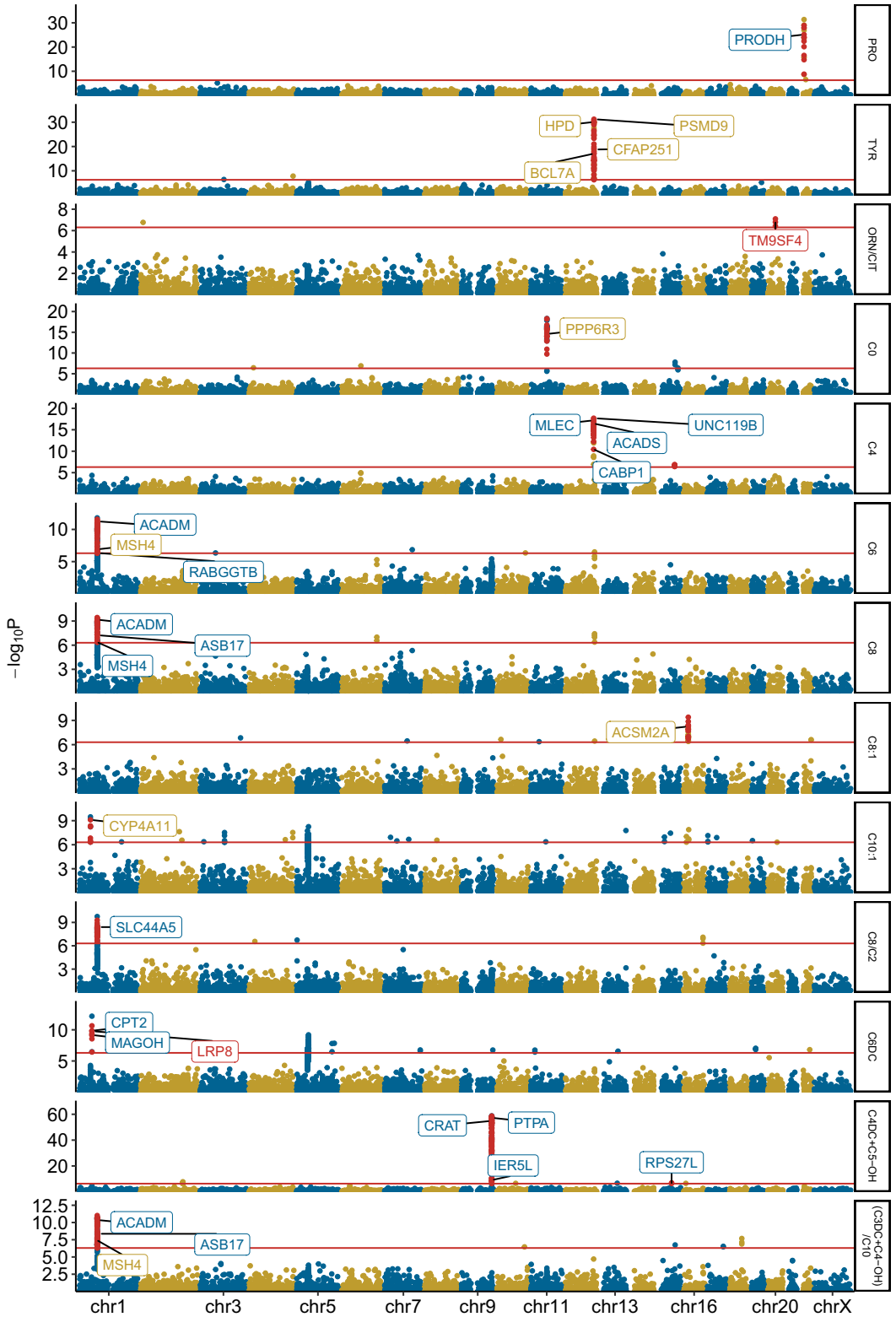
In the context of the 11 novel associations, our findings demonstrated an additive effect of the SNPs in their influence on metabolite concentrations (Figure 4C). The SNP rs695948 of *PPP6R3*, known as the bone density locus,[41] presents a strong association at the C0 concentration ($\beta = -0.29$ SD ± 0.037 SE). Additionally, a novel metabolism-related loci, rs6142615 of *TM9SF4*, has an impact on the ratio of citrulline/ornithine ($\beta = -0.10$ SD ± 0.019 SE). *TM9SF4* was reported to be involved in osteoporosis,[42] with *Tm9sf4*$^{-/-}$ mice exhibiting increase bone mass and a reduction in lipid accumulation within trabecular bones.[42] Treatments involving carnitine, citrulline, and ornithine have been suggested to impact the process of fracture healing.[43–45] The rs2596144 T allele of the *HPD* gene is shown to elevate tyrosine levels in neonates (β = 0.29 SD ± 0.024 SE). The *HPD* gene was documented as a causative factor of type III tyrosinemia (OMIM: 276710)[46] that is characterized by abnormally high levels of tyrosine in the bloodstream (Table S5). Furthermore, a significant interaction of two SNPs was observed in relation to the influence on tyrosine concentration (Figure 4D). The statistical evidence of the interaction between rs1720035 (*PSMD9*) and rs6486782 (*CFAP251*) was observed in both the discovery ($p = 5.32 \times 10^{-7}$) and the replication ($p = 5.62 \times 10^{-7}$) cohorts. This suggests that the impact of rs1720035 is influenced by rs6486782 and that the combined effects of these two SNPs are cumulative. These findings enhanced our comprehension regarding the influence of genetic factors on metabolic concentration and their associations with disorders.

Knowledge annotation of metabolism-related genes

We utilized KEGG enrichment analysis to explore the primary functional attributes of genes. Genes reported in our study are involved in the metabolic pathway of amino acids and lipids (Figure 5A). Additionally, the PhenoScanner[47] tool annotated 26 distinct phenotypes and/or diseases for six novel loci (Figure 5B). The *TM9SF4* gene has been implicated in the regulation of myeloid leukocytes, specifically in relation to monocyte count, and the proportion of monocytes among the total white blood cells.[48] In our data, this gene has an impact on the ratio of citrulline/ornithine. Citrulline could regulate the monocyte percentage of white cells and modulate the regulatory T cell function differently in infantile rats.[49–51] Ornithine could augment the process of autophagy and exert regulatory control on the infection

Table 1. GWAS summary for significant associations

Metabolic components	SNP	Mutant allele frequency		Gene	*p* value				GWAS Catalog
		Our study	ChinaMAP		Discovery	Replication	Meta-analysis	Trend test	
(C3DC+C4-OH)/C10	rs1146593	0.73	0.743	*ACADM*	4.31×10^{-11}	2.00×10^{-12}	2.40×10^{-21}	2.90×10^{-10}	known
(C3DC+C4-OH)/C10	rs79322031	0.282	0.264	*ASB17*	4.33×10^{-9}	2.30×10^{-9}	2.20×10^{-16}	2.40×10^{-9}	known
(C3DC+C4-OH)/C10	rs1144337	0.698	0.725	*MSH4*	3.86×10^{-8}	5.30×10^{-11}	1.90×10^{-17}	1.40×10^{-8}	novel
Adipoylcarnitine (C6DC)	rs2229291	0.21	0.241	*CPT2*	1.41×10^{-10}	3.50×10^{-6}	3.10×10^{-13}	3.20×10^{-11}	known
Adipoylcarnitine (C6DC)	rs75854125	0.204	0.242	*LRP8*	1.62×10^{-10}	8.00×10^{-6}	9.30×10^{-13}	7.80×10^{-10}	novel
Adipoylcarnitine (C6DC)	rs12082305	0.179	0.242	*MAGOH*	6.73×10^{-10}	5.90×10^{-6}	2.00×10^{-12}	1.40×10^{-9}	known
Butyrylcarnitine (C4)	rs12832102	0.09	0.140	*ACADS*	3.48×10^{-17}	7.30×10^{-20}	2.60×10^{-34}	4.60×10^{-14}	known
Butyrylcarnitine (C4)	rs35525135	0.055	0.066	*MLEC*	6.35×10^{-18}	9.40×10^{-11}	2.20×10^{-24}	3.10×10^{-14}	known
Butyrylcarnitine (C4)	rs553328	0.939	0.889	*CABP1*	3.69×10^{-11}	2.60×10^{-12}	3.60×10^{-21}	1.30×10^{-10}	known
Butyrylcarnitine (C4)	rs695948	0.072	0.122	*UNC119B*	2.01×10^{-18}	3.40×10^{-18}	3.40×10^{-32}	2.70×10^{-14}	known
C8/C2	rs814880	0.261	0.256	*SLC44A5*	3.96×10^{-9}	2.10×10^{-20}	5.20×10^{-28}	1.70×10^{-9}	known
Decenoylcarnitine (C10:1)	rs11211403	0.715	0.713	*CYP4A11*	7.78×10^{-10}	1.80×10^{-12}	2.20×10^{-20}	1.80×10^{-6}	novel
Free carnitine (C0)	rs12363572	0.048	0.056	*PPP6R3*	2.63×10^{-15}	1.70×10^{-9}	4.90×10^{-21}	1.30×10^{-12}	novel
Hexanoylcarnitine (C6)	rs1146593	0.73	0.743	*ACADM*	5.49×10^{-12}	5.00×10^{-17}	4.10×10^{-27}	6.70×10^{-11}	known
Hexanoylcarnitine (C6)	rs12703	0.712	0.726	*RABGGTB*	4.53×10^{-7}	4.80×10^{-14}	1.30×10^{-19}	5.00×10^{-7}	known
Hexanoylcarnitine (C6)	rs1251274	0.698	0.725	*MSH4*	1.25×10^{-7}	1.60×10^{-14}	1.10×10^{-20}	1.10×10^{-7}	novel
3-Hydroxy-isovalerylcarnitine1 and methylmalonyl (C4DC+C5-OH)	rs1075650	0.093	0.163	*IER5L*	9.24×10^{-10}	2.40×10^{-25}	1.50×10^{-33}	1.30×10^{-12}	known
3-Hydroxy-isovalerylcarnitine1 and methylmalonyl (C4DC+C5-OH)	rs10988200	0.598	0.594	*CRAT*	1.27×10^{-55}	7.60×10^{-82}	5.90×10^{-134}	3.20×10^{-57}	known
3-Hydroxy-isovalerylcarnitine1 and methylmalonyl (C4DC+C5-OH)	rs11853249	0.502	0.492	*RPS27L*	3.73×10^{-7}	2.60×10^{-5}	4.60×10^{-10}	4.20×10^{-6}	known
3-Hydroxy-isovalerylcarnitine1 and methylmalonyl (C4DC+C5-OH)	rs4713	0.583	0.582	*PTPA*	5.15×10^{-58}	8.30×10^{-84}	2.20×10^{-138}	4.70×10^{-61}	known
Octanoylcarnitine (C8)	rs1146593	0.73	0.743	*ACADM*	6.74×10^{-10}	1.20×10^{-16}	6.30×10^{-25}	2.10×10^{-9}	known
Octanoylcarnitine (C8)	rs77979447	0.194	0.263	*ASB17*	5.71×10^{-8}	9.10×10^{-13}	4.80×10^{-19}	1.30×10^{-6}	known
Octanoylcarnitine (C8)	rs12401729	0.269	0.272	*MSH4*	4.30×10^{-7}	2.40×10^{-14}	5.70×10^{-20}	1.30×10^{-6}	known
Octenoylcarnitine (C8:1)	rs1859000	0.127	0.203	*ACSM2A*	5.59×10^{-9}	1.80×10^{-5}	2.70×10^{-10}	7.00×10^{-6}	novel
ORN/CIT	rs6142615	0.794	0.773	*TM9SF4*	3.40×10^{-7}	2.20×10^{-5}	1.40×10^{-11}	1.20×10^{-7}	novel
Proline (PRO)	rs1808319	0.883	0.818	*PRODH*	7.06×10^{-26}	8.30×10^{-6}	1.10×10^{-21}	2.40×10^{-20}	known
Tyrosine (TYR)	rs2596144	0.881	0.874	*HPD*	8.25×10^{-31}	3.30×10^{-38}	6.60×10^{-66}	1.40×10^{-18}	novel
Tyrosine (TYR)	rs1795967	0.88	0.876	*PSMD9*	5.82×10^{-32}	2.70×10^{-36}	8.90×10^{-65}	9.40×10^{-20}	novel
Tyrosine (TYR)	rs1169093	0.896	0.851	*BCL7A*	8.44×10^{-18}	1.70×10^{-29}	2.50×10^{-45}	4.10×10^{-12}	novel
Tyrosine (TYR)	rs1154513	0.884	0.854	*CFAP251*	1.70×10^{-19}	5.30×10^{-31}	1.80×10^{-48}	1.20×10^{-13}	novel



(legend on next page)

caused by *Mycobacterium tuberculosis* infection.[52] To investigate the specific organs/tissues/cell types where these genes are active, we conducted an expression quantitative trait locus analysis. The metabolism-related genes are active in hepatocytes/liver (Figure 5C). Furthermore, cell type specificity analysis in single-cell transcriptomes showed that the expression of these genes was enriched in hepatocytes (particularly the hepatocyte subtype 3) of fetal/adult liver and the proximal tubules of fetal kidney (Figure 5D). Hepatocyte subtype 3 is characterized by a high level of gene expression associated with lipid, cholesterol, and biosynthesis synthesis.[53] The kidney is a highly metabolically active organ, where renal tubules utilize fatty acid oxidation to produce adenosine triphosphate (ATP).[54] Nevertheless, there was no observed enrichment in any of the cell types of fetal heart or gut.

Heritability of neonatal metabolic components

Genome-wide complex trait analysis (GCTA) estimates heritability by calculating the kinship of individuals based on their genotype matrix rather than directly utilizing the genotype matrix itself. This approach enables the calculation of heritability in neonates using the genetic information from their mothers. To verify our hypothesis, we conducted an observation on the WGS data of 602 trios. The result revealed a noteworthy correlation between the pairwise kinship of the maternal population and that of the neonatal population (Figure S6C). We employed SNPs with a significant threshold of $p < 0.01$ and calculated an average heritability of 76.2% (with a range from 69% to 78.8%) for neonatal metabolites. Furthermore, the heritability of maternal height and maternal BMI was estimated at 79% and 80%, respectively (Figure 6A). The observed high heritability indicated that genetic factors contribute a major component to the variation of metabolite concentration in neonates.

The significant heritability of neonatal metabolism suggests the potential to predict the individual concentration based on genetic data. Taken together with the additive effect of SNPs, we calculate the PRS for each participant using the GWAS summary of SNPs with a threshold of $p < 0.01$. The results showed a correlation between the r^2 of the PRS model and heritability (Figure 6B). The prediction accuracy was indicated by mean-square error (MSE), which was calculated based on the comparison between the predicted values and observed values. To facilitate the comparison of accuracy across various metabolic components, the MSE was divided by the variance (VAR) of the observed values. These results exhibited an accurate prediction of 29 metabolic components (MSE/VAR $\leq$ 0.2), together with that of maternal height and BMI (Figure 6C). Furthermore, we observed a strong correlation ($r^2 > 0.8$) between observation and prediction for 15 metabolic components. However, it is important to mention that the linear model's intercept requires adjustment (Figure S7).

DISCUSSION

Our study carried out a large-scale GWAS on neonatal metabolism. Through this analysis, we were able to identify 19 previously reported associations documented in the GWAS Catalog, as well as 11 novel associations. Moreover, our study demonstrated a significant heritability of neonatal metabolism, which appears to be related to maternal height and BMI, which also exhibit heritability. These findings indicate that metabolic concentrations can be accurately predicted using genetic data, as evidenced by the high heritability and successful application of PRSs. These results deepen our comprehension of the influence of genetic factors on neonatal metabolism.

Post-GWAS analysis revealed connections between metabolism-related genes and their activation in fetal hepatocytes, as well as in kidney's proximal tubules, indicating the involvement of these two critical cell types in the metabolic processes of newborns. According to the current understanding, the liver and kidney are the primary organs implicated in human metabolism and exert a significant impact on cardiovascular risk in the context of metabolic syndrome.[55] Hepatocytes are the primary cellular constituents of the liver, which plays a crucial role in the metabolic processes of carbohydrates, lipids, and proteins.[56] The kidney plays a significant role in the metabolic processes of carbohydrates, with approximately 40% of systemic glucose production occurring in the proximal tubule during periods of fasting and under stress conditions. In contrast, while the intestinal tract contributes to the metabolic responses[57] and supplies energy to the heart through oxidative metabolism,[58] there is currently no statistical evidence to support an involvement of fetal gut or heart in neonatal metabolism. These findings establish a connection among genetic, cellular, and biological processes, shedding light on the underlying mechanism of neonatal metabolism.

The estimation of heritability is influenced by a different methodology, in that GCTA[59] uses individual genotype data, while LDAK[60] uses statistical summary data. In this study, we estimated the heritability of neonatal metabolism and maternal height from the maternal kinship via the GCTA method. In light of the computational difficulty when analyzing millions of SNPs, we utilized a subset of SNPs with $p < 0.01$ for kinship calculation. While it is possible that heritability may be overestimated when using a major fraction of significant SNPs rather than all SNPs, our findings indicated a heritability of 79% for maternal height. This aligns with previous estimations of 84%–94% in female monozygotic twins and 49%–56% in dizygotic twins.[61] At this particular scale, the average heritability for neonatal metabolites was estimated to be approximately 76.2%, which is higher than the heritability estimates derived from a sample of 381 newborn twins.[62] The following top 10 components identified in Alul et al.'s study[62] show the

Figure 2. Manhattan plots of 13 metabolic components

SNPs that met all of the thresholds in discovery ($<5 \times 10^{-7}$), replication ($<2.68 \times 10^{-5}$), and meta-analysis ($<6.67 \times 10^{-10}$) are indicated by red points. Genes with blue color indicate the 16 genes of 19 known associations documented in the GWAS Catalog. Genes with yellow color indicate the eight genes of nine novel associations with other metabolic components. Genes with red color indicate the genes of two novel associations. The y axis refers to the p value of GWASs in discovery. p values were calculated by linear regression. The red horizontal line refers to the significance threshold (5×10^{-7}). Metabolite abbreviations are defined in Table S1.

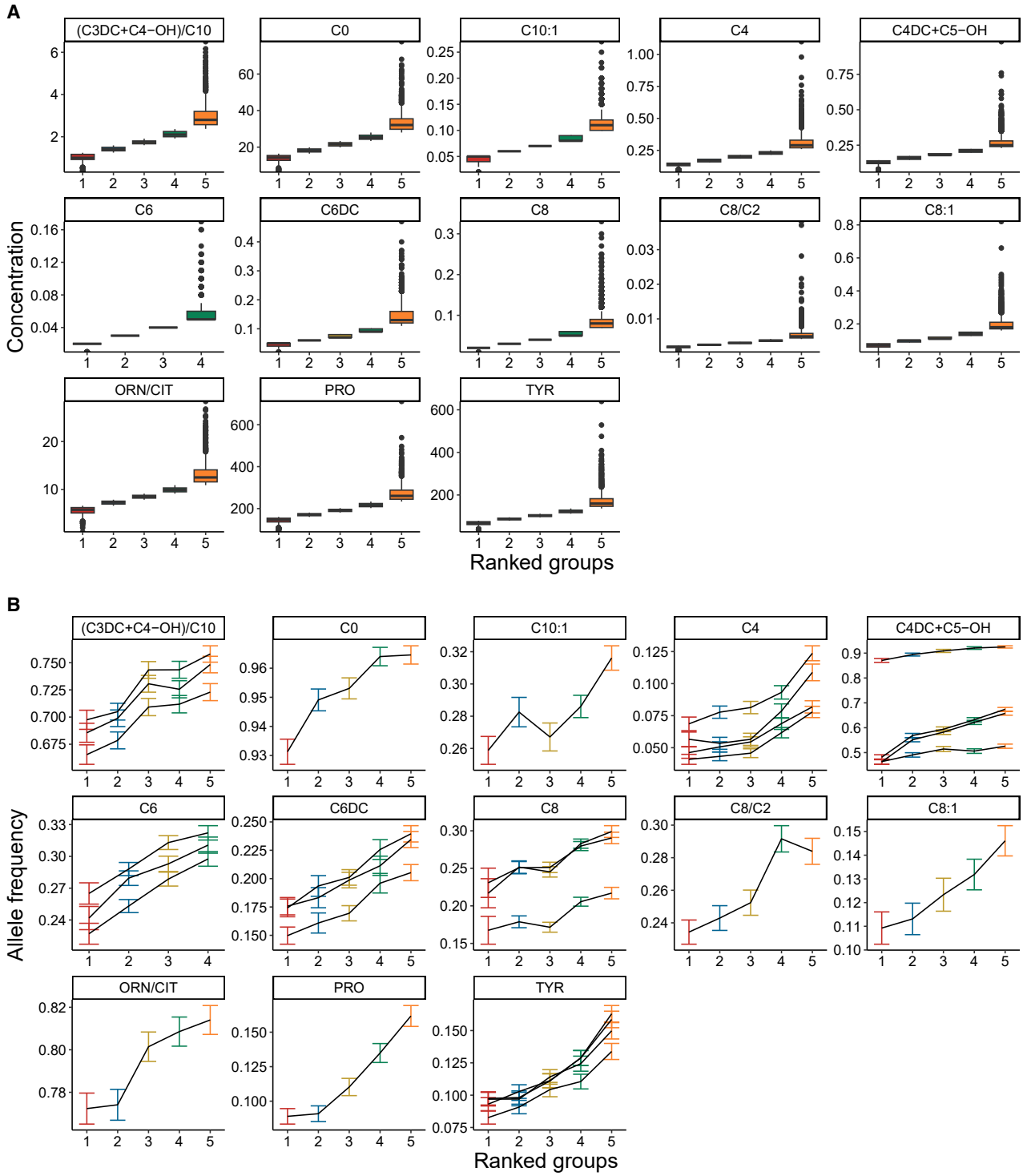
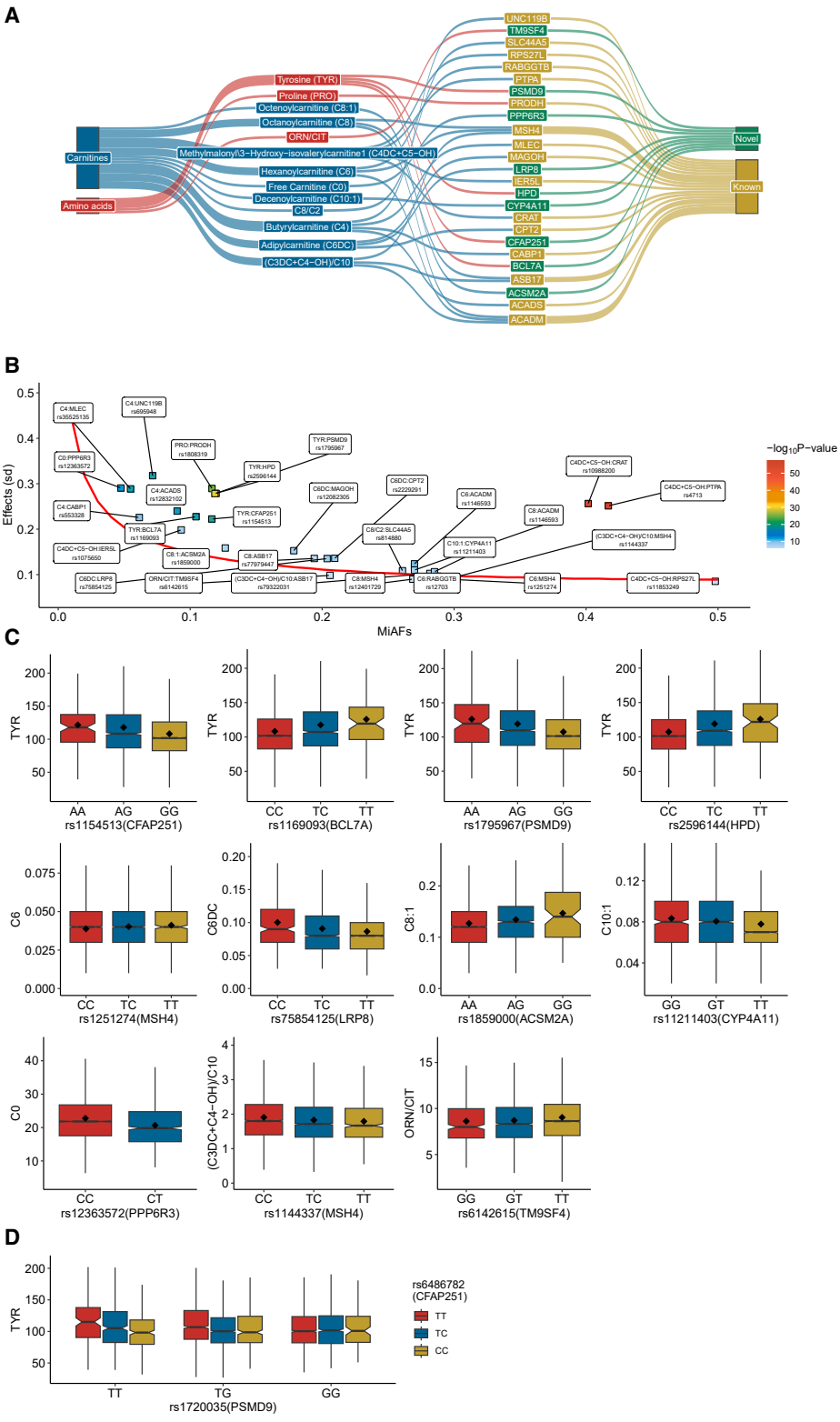


Figure 3. The correlation of allele frequency of SNPs and metabolic concentration in neonates

(A) The distribution of each metabolic concentration of five ranked groups.

(B) Allele frequencies of lead SNPs increased following the increase mean of metabolic concentration among the five ranked groups. Error bar refers to the standard error of allele frequency. Metabolite abbreviations are defined in Table S1.



(legend on next page)

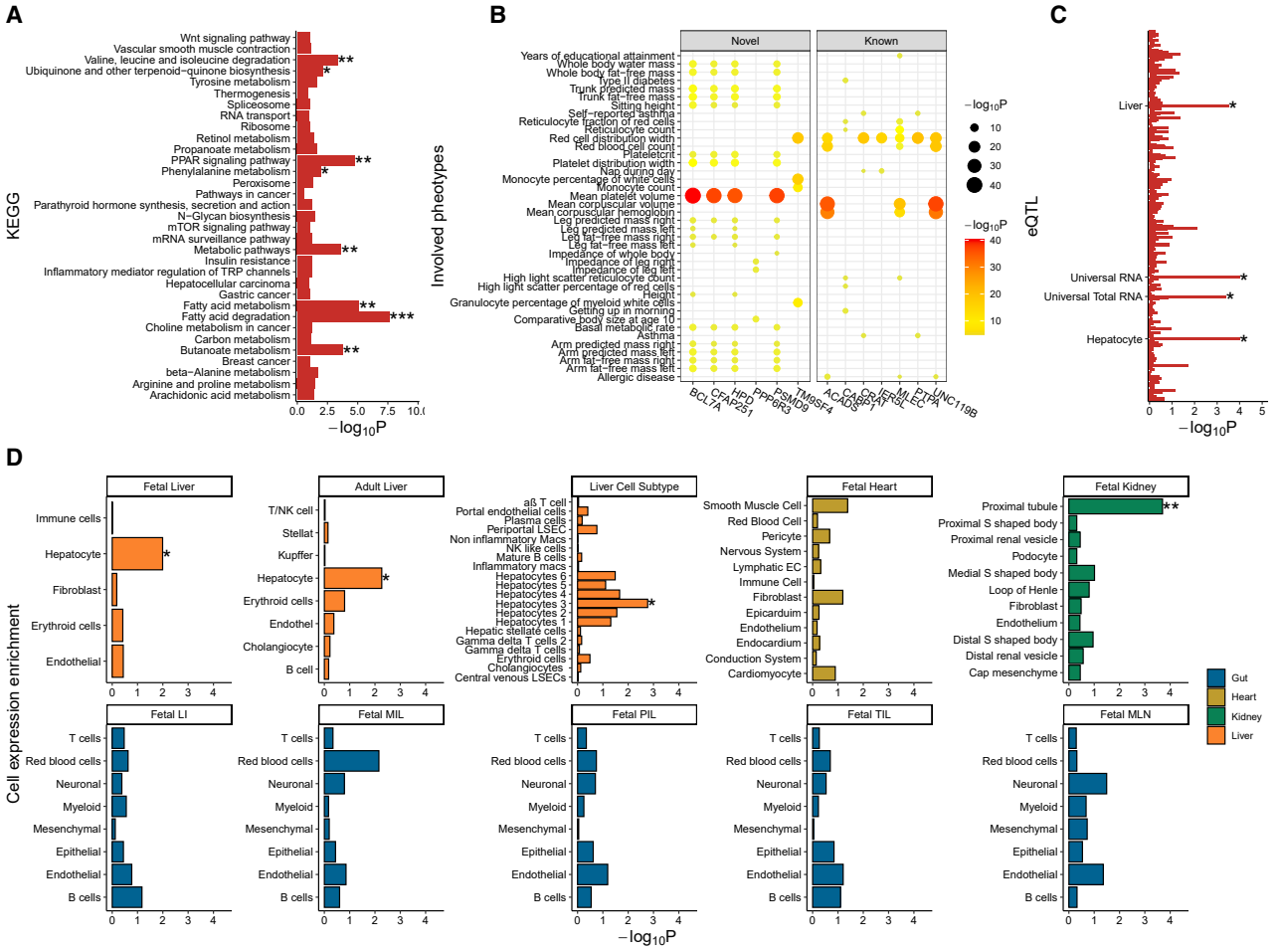


Figure 5. Knowledge annotation of metabolism-related genes

(A) KEGG enrichment analysis annotates eight functional characteristics of the metabolism-related genes. *p* values were calculated by a one-tailed Fisher's exact test.

(B) PhenoScanner identifies 26 phenotypes related to six novel loci.

(C and D) The metabolism-related genes harbor significant enrichments of expression quantitative trait loci in four tissue/organs (C) and are specifically expressed in the hepatocytes of fetal/adult liver and proximal tubules of fetal kidney (D). *p* values were calculated from the distribution of null hypothesis in simulation. No significant association was observed in any cell types of fetal heart and five regions of fetal gut, including the large intestine (LI), middle ileum (MIL), proximal ileum (PIL), terminal ileum (TIL), and mesenteric lymph node (MLN).

differences compared with our study: C4DC (83% vs. 77%), C4 (66% vs. 77%), C5 (61% vs. 76%), C2 (50% vs. 74.6%), C0 (45% vs. 76.2%), C3 (44% vs. 77%), C3/C2 (64% vs. 77.2%), C4/C3 (61% vs. 76.1%), PHE/TYR (51% vs. 77.5%), and TYR (47% vs. 75%). These differences may be attributed to the higher proportion of dizygotic twins, which accounted for 71% of Alul et al.'s study cohort. In the adult population, the average heritability of metabolites is approximately 50%.[63] As the phenotypic VAR was divided into genetic VAR and environmental VAR, the VAR in metabolism was largely influenced by genetic factors in newborns, and the influence of environmental factors increased in adults.

In addition to the aforementioned results, we showed that several neonatal metabolism-related loci ($p < 5 \times 10^{-8}$ in meta-analysis) have significant impacts on maternal metabolites and long-term health, such as pubertal growth and type 2

Figure 4. Summaries for associations and SNP effects

(A) The connections of metabolic components and related genes.

(B) SNPs effects (SD) for metabolic components and corresponding minor allele frequencies. The red line indicates the 80% power for detecting these significant associations.

(C) Metabolic concentrations are differed by the genotypes of SNPs from the 11 novel significant associations.

(D) A SNP-SNP interaction between rs1720035 and rs6486782. Metabolite abbreviations are defined in Table S1.

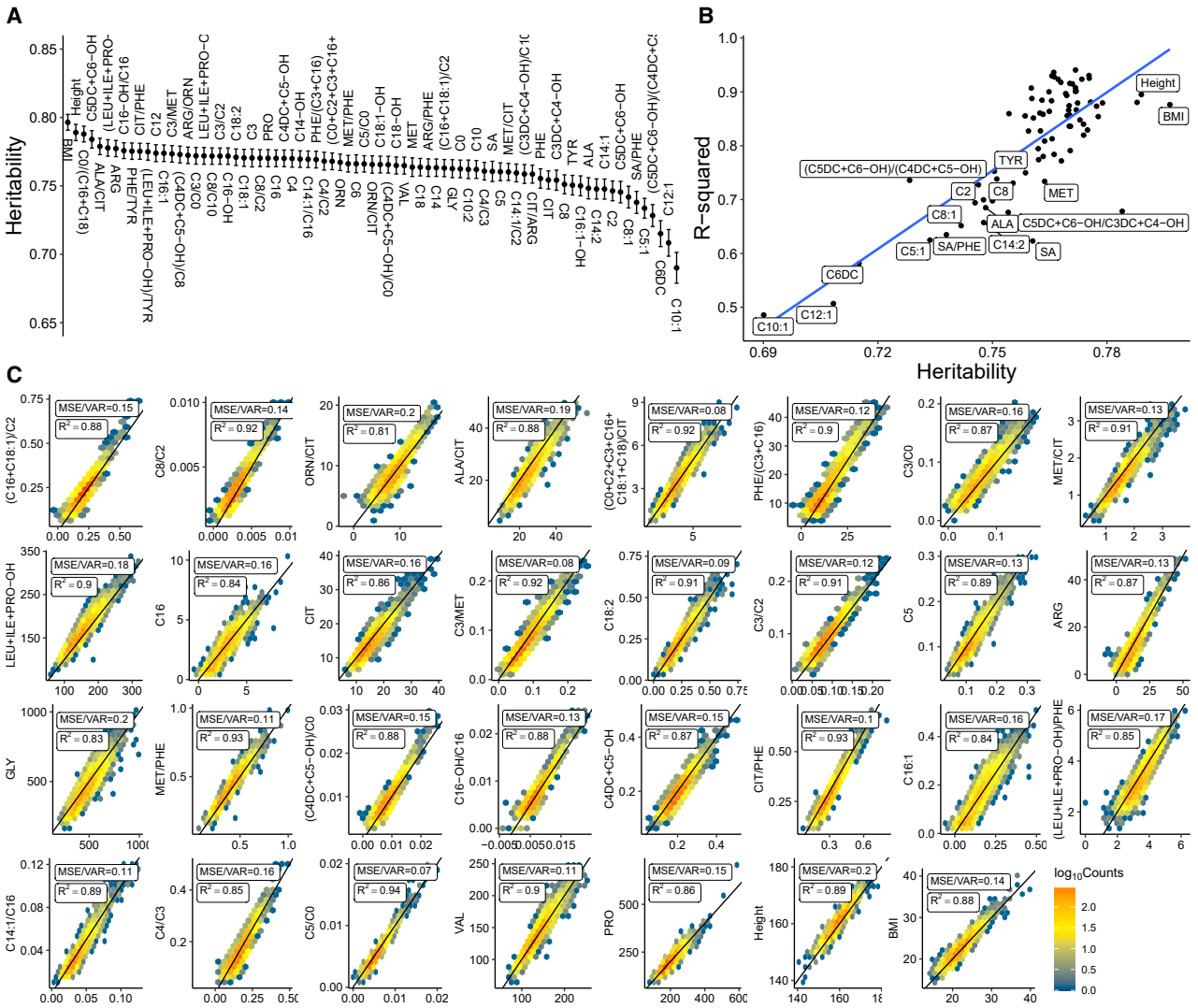


Figure 6. Metabolic heritability and PRS prediction

(A) The black dot and error bars show the estimated heritability and standard error, respectively.

(B) A correlation was observed between the heritability and r^2 of the PRS model. The accuracy of prediction was calculated by MSE/VAR.

(C) A total of 29 neonatal metabolic components and maternal height/BMI can be accurately predicted by the PRS model using a threshold of MSE/VAR $\leq$ 0.2. The x and y axes refer to the prediction and observation of the concentration of metabolites, respectively. Metabolite abbreviations are defined in Table S1.

diabetes. Research that independently investigates the genetic influence on maternal metabolites[64] and pregnancy phenotypes is underway.[65] In conjunction with these two studies,[64,65] our research showed a co-association of the *HMCN2* gene locus for citrulline in both maternal and neonatal subjects and a co-association of the *MARCH8-ZFAND4* locus for maternal mean corpuscular volume (and mean corpuscular hemoglobin) and neonatal C2 concentration. Upon comparison with public GWAS summaries, our non-significant association ($p_{discovery}$ = 8.94 × 10^{-8}, $p_{replication}$ = 0.0023, $p_{meta\text{-}analysis}$ = 3.53 × 10^{-7}) of arginine/ornithine for rs2608953 (*MED23* intron) may be detectable in larger cohorts. This SNP is a known locus associated with type 2 diabetes in a cross-ancestry population.[66] The disease-risk-increasing allele T (odd ratio = 1.042) is associated with higher ratios (β = 0.086 SD) in our data. Cochran's *Q* statistic (p = 0.15) indicated no significant heterogeneity in multi-ancestry for this locus.[66] Cao et al. showed that arginine was positively associated with the risk of type 2 diabetes, while the correlation of risk for ornithine was negatively associated.[67] Karjalainen et al. showed that *MED23* is associated with glutamine,[68] a metabolite that improves glycemic control and the levels of incretins in diabetes mellitus.[69] The locus of rs369065 (*LIN28B* intron) with a suggestive association ($p_{discovery}$ = 8.14 × 10^{-6}, $p_{replication}$ = 0.0001, $p_{meta\text{-}analysis}$ = 5.80 × 10^{-9}) for C4DC+C5-OH has a significant impact on pubertal growth in Europeans. The concentration-increasing allele T (β = 0.066 SD) is associated with higher height growth (β = 0.078 SD) from age 14 to adult.[70] Although the compared population is European, common variants were

shown to have similar effects across ancestries for height[71] and metabolic traits.[68,72] We did not observe a co-association between neonatal metabolism and childhood obesity[73] or childhood BMI at 11 timings from birth to age 7.[11]

The primary inquiry of the study is to utilize genetic data derived from a collection of maternal and fetal cfDNA for GWASs. Fetal cfDNA contributes a proportion ranging from 1% to 30% of the total cfDNA, hence influencing the genotype within maternal blood. The influence of genotype bias in a NIPT-WGS GWAS necessitates thorough discussion. If there is no correlation between the fraction of cfDNA and phenotype, then the imputed genotypes are not biased for GWASs. However, if there is indeed a correlation, then it is recommended to include the proportion of cfDNA and principal components as covariates in the association model. Our investigation did not find any statistical evidence to support an association between the proportion of cfDNA and neonatal metabolism. The fraction of fetal cfDNA at the gestational duration of NIPT does not affect neonatal metabolism after delivery, suggesting that the proportion of cfDNA does not introduce bias to GWASs. However, a significant association was seen between the fraction of cfDNA and maternal BMI, in agreement with a previous report,[74] indicating that the inclusion of the cfDNA proportion as a covariate is necessary in association tests of maternal BMI. It is also noteworthy that our study focused on GWASs of neonatal phenotypes. We showed that the allele frequency of SNPs is in Hardy-Weinberg balance and congruent between neonates and their mothers. Furthermore, we have devised a trend test based on allele frequency to validate the association established by maternal genotypes. Our findings of 19 known associations, nine associations for diverse metabolic components, and one association with a gene located near known loci support the feasibility of this approach in neonatal metabolism GWASs. The widespread global application of NIPT technology, together with subsequent newborn testing, has rendered the utilization of the maternal genotype as a method to conduct GWASs on neonatal traits highly valuable.

Limitations of the study

The genetic connections in neonates have been demonstrated; nevertheless, the precise genetic effect in neonates and the biological mechanism underlying the association between maternal genotype and neonatal metabolism remains elusive. While the initial observation of the link was made in relation to the maternal genotype, it is important to note that we did not propose a direct influence of the maternal genotype on neonatal metabolism. In this study, we have put forth a hypothetical mechanism by which the effective allele is transmitted from the mother and/or potentially the unobserved father to the offspring, thereby exerting an influence on the concentration of the related metabolite. Our study demonstrated a positive correlation between the allele frequency in neonatal groups and the concentration of neonatal metabolites. The obtained outcome provided evidence in favor of our supposition. Brand et al. proposed an approach for extracting the fetal genome from cfDNA found in the mother's blood.[75] By isolating fetal cfDNA from maternal cfDNA and including the genome of the father, we can improve our comprehension of the genetic architecture of heritable phenotypes. Environmental covariates (e.g., delivery way,[76] gestational duration[77]) are also important for estimating an adjusted genetic effect and an interacting effect. Further studies are needed to determine the effects from the combination of genetics and the environment.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Ting Wang (biowt@njmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Public data employed in our study are listed in the key resources table. The GWAS summary was released at The National Genomics Data Center (website: https://ngdc.cncb.ac.cn/; Project ID: PRJCA018290, OMIX ID: OMIX004498). The necessary codes of bioinformatic and statistical analyses were released at GitHub (https://github.com/liuhankui/NIPT) and Zenodo (https://doi.org/10.5281/zenodo.13692304).

ACKNOWLEDGMENTS

We thank the participants for their involvement in this study. We thank Zhangwei Xie, You Zhou, and Ziwei Wang for their assistance with data collection. This study was supported by grants from Primary Research and Development Plan of Jiangsu Province, China (BE2022736); Jiangsu Maternal and Children Health Care Key Discipline, China (FXK202142); Jiangsu Provincial Medical Key Discipline (Laboratory) Cultivation Unit, China (JSDW202214); Shenzhen Municipal of Government of China, China (JCYJ20180507183615145); Key Research and Development Project of Hebei Province, China (21377720D); S&T Program of Hebei, China (215A9907D); Clinical Medical Technology Demonstration Base for Genetic Research of Fetal Congenital Heart Disease in Hunan Province, China (2021SK4036); Project of Changsha Science and Technology Bureau, China (KH2201045); Natural Science Foundation of Hunan Province, China (2023JJ30063); Changsha Science and Technology Bureau Natural Science Surface Project, China (kq2202030); and National Key Research and Development Program of China, China (2022YFC2703102).

AUTHOR CONTRIBUTIONS

Conceptualization, Q.H., H. Liu, and X.J.; methodology, H. Liu, Q.H., and X.J.; software, H. Liu and Q.H.; formal analysis, Q.H., H. Liu, L.L., Q.Z., X.W., Y.X., C.Z., X.C., Y.H., X.P., and K.Z.; statistical analysis, H. Liu; data curation, Q.W., B.W., J.M., L.G., H.P., and H. Li.; validation, Q.H., H. Liu, J.Z., and X.J.; visualization, H. Liu; interpretation, J.Z., X.J., Q.H., and H. Liu; writing – original draft, H. Liu and Q.H.; writing – review & editing, X.J., J.Z., L.Z., T.W., Q.H., H. Liu, L.L., L.G., J.M., and X.P.; supervision, T.W., J.Z., L.Z., and X.J.; resources, L.Z., J.Z., and T.W.; project administrator, T.W., J.Z., and L.Z.; funding acquisition, T.W., J.Z., L.Z., and X.P. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100668.

Received: March 9, 2023
Revised: December 16, 2023
Accepted: September 11, 2024
Published: October 9, 2024

REFERENCES

1. Wu, G. (2009). Amino acids: metabolism, functions, and nutrition. Amino Acids *37*, 1–17. https://doi.org/10.1007/s00726-009-0269-0.
2. Ferreira, G.C., and McKenna, M.C. (2017). L-Carnitine and Acetyl-L-carnitine Roles and Neuroprotection in Developing Brain. Neurochem. Res. *42*, 1661–1675. https://doi.org/10.1007/s11064-017-2288-7.
3. Cheng, Y., Schlosser, P., Hertel, J., Sekula, P., Oefner, P.J., Spiekerkoetter, U., Mielke, J., Freitag, D.F., Schmidts, M., et al.; GCKD Investigators (2021). Rare genetic variants affecting urine metabolite levels link population variation to inborn errors of metabolism. Nat. Commun. *12*, 964. https://doi.org/10.1038/s41467-020-20877-8.
4. Lotta, L.A., Pietzner, M., Stewart, I.D., Wittemans, L.B.L., Li, C., Bonelli, R., Raffler, J., Biggs, E.K., Oliver-Williams, C., Auyeung, V.P.W., et al. (2021). A cross-platform approach identifies genetic regulators of human metabolism and health. Nat. Genet. *53*, 54–64. https://doi.org/10.1038/s41588-020-00751-5.
5. Rhee, E.P., Yang, Q., Yu, B., Liu, X., Cheng, S., Deik, A., Pierce, K.A., Bullock, K., Ho, J.E., Levy, D., et al. (2016). An exome array study of the plasma metabolome. Nat. Commun. *7*, 12360. https://doi.org/10.1038/ncomms12360.
6. Tabassum, R., Rämö, J.T., Ripatti, P., Koskela, J.T., Kurki, M., Karjalainen, J., Palta, P., Hassan, S., Nunez-Fontarnau, J., Kiiskinen, T.T.J., et al. (2019). Genetic architecture of human plasma lipidome and its link to cardiovascular disease. Nat. Commun. *10*, 4329. https://doi.org/10.1038/s41467-019-11954-8.
7. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523. https://doi.org/10.1038/s41588-018-0222-9.
8. Feofanova, E.V., Chen, H., Dai, Y., Jia, P., Grove, M.L., Morrison, A.C., Qi, Q., Daviglus, M., Cai, J., North, K.E., et al. (2020). A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos. Am. J. Hum. Genet. *107*, 849–863. https://doi.org/10.1016/j.ajhg.2020.09.003.
9. Yin, X., Chan, L.S., Bose, D., Jackson, A.U., VandeHaar, P., Locke, A.E., Fuchsberger, C., Stringham, H.M., Welch, R., Yu, K., et al. (2022). Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. Nat. Commun. *13*, 1644. https://doi.org/10.1038/s41467-022-29143-5.
10. Guo, F., Zhou, L., Zhang, F., Yu, B., Yang, Y., and Liu, Z. (2024). Abnormal biochemical indicators of neonatal inherited metabolic disease in carriers. Orphanet J. Rare Dis. *19*, 145. https://doi.org/10.1186/s13023-024-03138-5.
11. Helgeland, Ø., Vaudel, M., Sole-Navais, P., Flatley, C., Juodakis, J., Bacelis, J., Koløen, I.L., Knudsen, G.P., Johansson, B.B., Magnus, P., et al. (2022). Characterization of the genetic architecture of infant and early childhood body mass index. Nat. Metab. *4*, 344–358. https://doi.org/10.1038/s42255-022-00549-1.
12. Helgeland, Ø., Vaudel, M., Juliusson, P.B., Lingaas Holmen, O., Juodakis, J., Bacelis, J., Jacobsson, B., Lindekleiv, H., Hveem, K., Lie, R.T., et al. (2019). Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. Nat. Commun. *10*, 4448.
13. Duquenne, M., Folgueira, C., Bourouh, C., Millet, M., Silva, A., Clasadonte, J., Imbernon, M., Fernandois, D., Martinez-Corral, I., Kusumakshi, S., et al. (2021). Leptin brain entry via a tanycytic LepR–EGFR shuttle controls lipid metabolism and pancreas function. Nat. Metab. *3*, 1071–1090. https://doi.org/10.1038/s42255-021-00432-5.
14. Rousson, R., and Guibaud, P. (1984). Long term outcome of organic acidurias: survey of 105 French cases (1967–1983). J. Inherit. Metab. Dis. *7*, 10–12.
15. Muenzer, J. (2004). The mucopolysaccharidoses: a heterogeneous group of disorders with variable pediatric presentations. J. Pediatr. *144*, S27–S34.
16. Bräutigam, C., Wevers, R.A., Jansen, R.J.T., Smeitink, J.A.M., Andel, J.F. de R., Gabreëls, F.J.M., and Hoffmann, G.F. (1998). Biochemical hallmarks of tyrosine hydroxylase deficiency. Clin. Chem. *44*, 1897–1904.
17. Fukao, T., and Nakamura, K. (2019). Advances in inborn errors of metabolism. J. Hum. Genet. *64*, 65. https://doi.org/10.1038/s10038-018-0535-7.
18. Adhikari, A.N., Gallagher, R.C., Wang, Y., Currier, R.J., Amatuni, G., Bassaganyas, L., Chen, F., Kundu, K., Kvale, M., Mooney, S.D., et al. (2020). The role of exome sequencing in newborn screening for inborn errors of metabolism. Nat. Med. *26*, 1392–1397. https://doi.org/10.1038/s41591-020-0966-5.
19. Ding, Y., Owen, M., Le, J., Batalov, S., Chau, K., Kwon, Y.H., Van Der Kraan, L., Bezares-Orin, Z., Zhu, Z., Veeraraghavan, N., et al. (2023). Scalable, high quality, whole genome sequencing from archived, newborn, dried blood spots. NPJ Genom. Med. *8*, 5. https://doi.org/10.1038/s41525-023-00349-w.
20. Davies, R.W., Flint, J., Myers, S., and Mott, R. (2016). Rapid genotype imputation from sequence without reference panels. Nat. Genet. *48*, 965–969.
21. Davies, R.W., Kucka, M., Su, D., Shi, S., Flanagan, M., Cunniff, C.M., Chan, Y.F., and Myers, S. (2021). Rapid genotype imputation from sequence with reference panels. Nat. Genet. *53*, 1104–1111.
22. Chat, V., Ferguson, R., Morales, L., and Kirchhoff, T. (2021). Ultra Low-Coverage Whole-Genome Sequencing as an Alternative to Genotyping Arrays in Genome-Wide Association Studies. Front. Genet. *12*, 790445. https://doi.org/10.3389/fgene.2021.790445.
23. Li, J.H., Mazur, C.A., Berisa, T., and Pickrell, J.K. (2021). Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. Genome Res. *31*, 529–537. https://doi.org/10.1101/gr.266486.120.
24. Wasik, K., Berisa, T., Pickrell, J.K., Li, J.H., Fraser, D.J., King, K., and Cox, C. (2021). Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. BMC Genom. *22*, 197. https://doi.org/10.1186/s12864-021-07508-2.
25. Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S.S., Fang, L., Li, Z., Lin, L., Liu, R., et al. (2018). Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. Cell *175*, 347–359.e14.
26. Balding, D.J. (2006). A tutorial on statistical methods for population association studies. Nat. Rev. Genet. *7*, 781–791.

27. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012.

28. Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Hosono, N., Ohmiya, H., Daigo, Y., Yamamoto, K., Kubo, M., Nakamura, Y., and Kamatani, N. (2010). A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. Hum. Mol. Genet. *19*, 2303–2312. https://doi.org/10.1093/hmg/ddq091.

29. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518. https://doi.org/10.1038/s41586-019-1310-4.

30. Hotta, K., Nakamura, M., Nakamura, T., Matsuo, T., Nakata, Y., Kamohara, S., Miyatake, N., Kotani, K., Komatsu, R., Itoh, N., et al. (2009). Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. J. Hum. Genet. *54*, 727–731.

31. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat. Genet. *52*, 680–691. https://doi.org/10.1038/s41588-020-0637-y.

32. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. Nat. Commun. *10*, 339. https://doi.org/10.1038/s41467-018-08000-4.

33. Lan, N., Lu, Y., Zhang, Y., Pu, S., Xi, H., Nie, X., Liu, J., and Yuan, W. (2020). FTO - A Common Genetic Basis for Obesity and Cancer. Front. Genet. *11*, 559138. https://doi.org/10.3389/fgene.2020.559138.

34. Zeng, Z., Liu, H., Xu, H., Lu, H., Yu, Y., Xu, X., Yu, M., Zhang, T., Tian, X., Xi, H., et al. (2021). Genome-wide association study identifies new loci associated with risk of HBV infection and disease progression. BMC Med. Genom. *14*, 84. https://doi.org/10.1186/s12920-021-00907-0.

35. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004.

36. Krumsiek, J., Suhre, K., Evans, A.M., Mitchell, M.W., Mohney, R.P., Milburn, M.V., Wägele, B., Römisch-Margl, W., Illig, T., and Adamski, J. (2012). Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet. *8*, e1003005. https://doi.org/10.1371/journal.pgen.1003005.

37. Surendran, P., Stewart, I.D., Au Yeung, V.P.W., Pietzner, M., Raffler, J., Wörheide, M.A., Li, C., Smith, R.F., Wittemans, L.B.L., Bomba, L., et al. (2022). Rare and common genetic determinants of metabolic individuality and their effects on human health. Nat. Med. *28*, 2321–2332.

38. Schlosser, P., Scherer, N., Grundner-Culemann, F., Monteiro-Martins, S., Haug, S., Steinbrenner, I., Uluvar, B., Wuttke, M., Cheng, Y., Ekici, A.B., et al. (2023). Genetic studies of paired metabolomes reveal enzymatic and transport processes at the interface of plasma and urine. Nat. Genet. *55*, 995–1008.

39. Li, Y., Sekula, P., Wuttke, M., Wahrheit, J., Hausknecht, B., Schultheiss, U.T., Gronwald, W., Schlosser, P., Tucci, S., Ekici, A.B., et al. (2018). Genome-wide association studies of metabolites in patients with CKD identify multiple loci and illuminate tubular transport mechanisms. J. Am. Soc. Nephrol. *29*, 1513–1524.

40. Mason, E., Hindmarch, C.C.T., and Dunham-Snary, K.J. (2023). Medium-chain Acyl-COA dehydrogenase deficiency: Pathogenesis, diagnosis, and treatment. Endocrinol. Diabetes Metab. *6*, e385.

41. He, D., Liu, H., Wei, W., Zhao, Y., Cai, Q., Shi, S., Chu, X., Qin, X., Zhang, N., Xu, P., and Zhang, F. (2023). A longitudinal genome-wide association study of bone mineral density mean and variability in the UK Biobank. Osteoporos. Int. *34*, 1907–1916.

42. Yu, L., Xie, M., Zhang, F., Wan, C., and Yao, X. (2021). TM9SF4 is a novel regulator in lineage commitment of bone marrow mesenchymal stem cells to either osteoblasts or adipocytes. Stem Cell Res. Ther. *12*, 573. https://doi.org/10.1186/s13287-021-02636-8.

43. Wijnands, K.A.P., Brink, P.R.G., Weijers, P.H.E., Dejong, C.H.C., and Poeze, M. (2012). Impaired fracture healing associated with amino acid disturbances. Am. J. Clin. Nutr. *95*, 1270–1277. https://doi.org/10.3945/ajcn.110.009209.

44. Meesters, D.M., Hannemann, P.F., van Eijk, H.M., Schriebl, V.T., Brink, P.R., Poeze, M., and Wijnands, K.A. (2020). Enhancement of fracture healing after citrulline supplementation in mice. Eur. Cell. Mater. *39*, 183–192. https://doi.org/10.22203/eCM.v039a12.

45. Aydin, A., Halici, Z., Albayrak, A., Polat, B., Karakus, E., Yildirim, O.S., Bayir, Y., Cadirci, E., Ayan, A.K., and Aksakal, A.M. (2015). Treatment with carnitine enhances bone fracture healing under osteoporotic and/or inflammatory conditions. Basic Clin. Pharmacol. Toxicol. *117*, 173–179.

46. Tomoeda, K., Awata, H., Matsuura, T., Matsuda, I., Ploechl, E., Milovac, T., Boneh, A., Scott, C.R., Danks, D.M., and Endo, F. (2000). Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. Mol. Genet. Metabol. *71*, 506–510.

47. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics *32*, 3207–3209. https://doi.org/10.1093/bioinformatics/btw373.

48. Paolillo, R., Spinello, I., Quaranta, M.T., Pasquini, L., Pelosi, E., Lo Coco, F., Testa, U., and Labbaye, C. (2015). Human TM9SF4 Is a New Gene Down-Regulated by Hypoxia and Involved in Cell Adhesion of Leukemic Cells. PLoS One *10*, e0126968. https://doi.org/10.1371/journal.pone.0126968.

49. Lee, Y.-C., Su, Y.-T., Liu, T.-Y., Tsai, C.-M., Chang, C.-H., and Yu, H.-R. (2018). L-Arginine and L-Citrulline Supplementation Have Different Programming Effect on Regulatory T-Cells Function of Infantile Rats. Front. Immunol. *9*, 2911. https://doi.org/10.3389/fimmu.2018.02911.

50. Geiger, R., Rieckmann, J.C., Wolf, T., Basso, C., Feng, Y., Fuhrer, T., Kogadeeva, M., Picotti, P., Meissner, F., Mann, M., et al. (2016). L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. Cell *167*, 829–842.e13. https://doi.org/10.1016/j.cell.2016.09.031.

51. Reizine, F., Grégoire, M., Lesouhaitier, M., Coirier, V., Gauthier, J., Delaloy, C., Dessauge, E., Creusat, F., Uhel, F., Gacouin, A., et al. (2022). Beneficial effects of citrulline enteral administration on sepsis-induced T cell mitochondrial dysfunction. Proc. Natl. Acad. Sci. USA *119*, e2115139119. https://doi.org/10.1073/pnas.2115139119.

52. Sivangala Thandi, R., Radhakrishnan, R.K., Tripathi, D., Paidipally, P., Azad, A.K., Schlesinger, L.S., Samten, B., Mulik, S., and Vankayalapati, R. (2020). Ornithine-A urea cycle metabolite enhances autophagy and controls Mycobacterium tuberculosis infection. Nat. Commun. *11*, 3535. https://doi.org/10.1038/s41467-020-17310-5.

53. MacParland, S.A., Liu, J.C., Ma, X.-Z., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., et al. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat. Commun. *9*, 4383.

54. Gewin, L.S. (2021). Sugar or Fat? Renal Tubular Metabolism Reviewed in Health and Disease. Nutrients *13*, 1580. https://doi.org/10.3390/nu13051580.

55. Carbone, F., Montecucco, F., Mach, F., Pontremoli, R., and Viazzi, F. (2013). The liver and the kidney: two critical organs influencing the atherothrombotic risk in metabolic syndrome. Thromb. Haemostasis *110*, 940–958. https://doi.org/10.1160/TH13-06-0499.

56. Gong, J., Tu, W., Liu, J., and Tian, D. (2022). Hepatocytes: A key role in liver inflammation. Front. Immunol. *13*, 1083780. https://doi.org/10.3389/fimmu.2022.1083780.

57. DeWitt, R.C., and Kudsk, K.A. (1999). The gut's role in metabolism, mucosal barrier function, and gut immunology. Infect. Dis. Clin. *13*, 465–481. https://doi.org/10.1016/s0891-5520(05)70086-6.

58. Kartha, C.C. (2021). Energy Metabolism in Cardiomyocyte. In Cardiomyocytes in Health and Disease (Springer), pp. 73–92. https://doi.org/10.1007/978-3-030-85536-9_7.

59. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

60. Zhang, Q., Privé, F., Vilhjálmsson, B., and Speed, D. (2021). Improved genetic prediction of complex traits from individual-level data or summary statistics. Nat. Commun. *12*, 4192. https://doi.org/10.1038/s41467-021-24485-y.

61. Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D.I., Cornes, B.K., Davis, C., Dunkel, L., De Lange, M., Harris, J.R., Hjelmborg, J.V.B., et al. (2003). Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. *6*, 399–408. https://doi.org/10.1375/136905203770326402.

62. Alul, F.Y., Cook, D.E., Shchelochkov, O.A., Fleener, L.G., Berberich, S.L., Murray, J.C., and Ryckman, K.K. (2013). The heritability of metabolic profiles in newborn twins. Heredity *110*, 253–258. https://doi.org/10.1038/hdy.2012.75.

63. Hagenbeek, F.A., Pool, R., van Dongen, J., Draisma, H.H.M., Jan Hottenga, J., Willemsen, G., Abdellaoui, A., Fedko, I.O., den Braber, A., Visser, P.J., et al. (2020). Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. Nat. Commun. *11*, 39. https://doi.org/10.1038/s41467-019-13770-6.

64. Liu, S., Yao, J., Lin, L., Lan, X., Wu, L., He, X., Kong, N., Li, Y., Deng, Y., Xie, J., et al. (2024). Genome-wide association study of maternal plasma metabolites during pregnancy. Cell Genom. *4*, 100657. https://doi.org/10.1016/j.xgen.2024.100657.

65. Xiao, H., Li, L., Yang, M., Zhang, X., Zhou, J., Zeng, J., Zhou, Y., Lan, X., Liu, J., Lin, Y., et al. (2024). Genetic analyses of 104 phenotypes in 20,900 Chinese pregnant women reveal pregnancy-specific discoveries. Cell Genom. *4*, 100633. https://doi.org/10.1016/j.xgen.2024.100633.

66. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima, H., Yu, G.Z., Rüeger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet. *54*, 560–572. https://doi.org/10.1038/s41588-022-01058-3.

67. Cao, Y.-F., Li, J., Zhang, Z., Liu, J., Sun, X.-Y., Feng, X.-F., Luo, H.-H., Yang, W., Li, S.-N., Yang, X., and Fang, Z.Z. (2019). Plasma levels of amino acids related to urea cycle and risk of type 2 diabetes mellitus in Chinese adults. Front. Endocrinol. *10*, 50.

68. Karjalainen, M.K., Karthikeyan, S., Oliver-Williams, C., Sliz, E., Allara, E., Fung, W.T., Surendran, P., Zhang, W., Jousilahti, P., Kristiansson, K., et al. (2024). Genome-wide characterization of circulating metabolic biomarkers. Nature *628*, 130–138. https://doi.org/10.1038/s41586-024-07148-y.

69. Jafari-Vayghan, H., Varshosaz, P., Hajizadeh-Sharafabad, F., Razmi, H.R., Amirpour, M., Tavakoli-Rouzbehani, O.M., Alizadeh, M., and Maleki, V. (2020). A comprehensive insight into the effect of glutamine supplementation on metabolic variables in diabetes mellitus: a systematic review. Nutr. Metab. *17*, 80. https://doi.org/10.1186/s12986-020-00503-6.

70. Cousminer, D.L., Berry, D.J., Timpson, N.J., Ang, W., Thiering, E., Byrne, E.M., Taal, H.R., Huikari, V., Bradfield, J.P., Kerkhof, M., et al. (2013). Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. Hum. Mol. Genet. *22*, 2735–2747.

71. Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., et al. (2022). A saturated map of common genetic variants associated with human height. Nature *610*, 704–712. https://doi.org/10.1038/s41586-022-05275-y.

72. Cho, C., Kim, B., Kim, D.S., Hwang, M.Y., Shim, I., Song, M., Lee, Y.C., Jung, S.-H., Cho, S.K., Park, W.-Y., et al. (2024). Large-scale cross-ancestry genome-wide meta-analysis of serum urate. Nat. Commun. *15*, 3441.

73. Bradfield, J.P., Vogelezang, S., Felix, J.F., Chesi, A., Helgeland, Ø., Horikoshi, M., Karhunen, V., Lowry, E., Cousminer, D.L., Ahluwalia, T.S., et al. (2019). A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. Hum. Mol. Genet. *28*, 3327–3338.

74. Kim, S.K., Hannum, G., Geis, J., Tynan, J., Hogg, G., Zhao, C., Jensen, T.J., Mazloom, A.R., Oeth, P., Ehrich, M., et al. (2015). Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. Prenat. Diagn. *35*, 810–815. https://doi.org/10.1002/pd.4615.

75. Brand, H., Whelan, C.W., Duyzend, M., Lemanski, J., Salani, M., Hao, S.P., Wong, I., Valkanas, E., Cusick, C., Genetti, C., et al. (2023). High-resolution and noninvasive fetal exome screening. N. Engl. J. Med. *389*, 2014–2016.

76. Bouhanick, B., Ehlinger, V., Delpierre, C., Chamontin, B., Lang, T., and Kelly-Irving, M. (2014). Mode of delivery at birth and the metabolic syndrome in midlife: the role of the birth environment in a prospective birth cohort study. BMJ Open *4*, e005031.

77. Mericq, V., Martinez-Aguayo, A., Uauy, R., Iñiguez, G., Van der Steen, M., and Hokken-Koelega, A. (2017). Long-term metabolic risk among children born premature or small for gestational age. Nat. Rev. Endocrinol. *13*, 50–62. https://doi.org/10.1038/nrendo.2016.127.

78. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

79. Cao, Y., Li, L., Xu, M., Feng, Z., Sun, X., Lu, J., Xu, Y., Du, P., Wang, T., Hu, R., et al. (2020). The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals. Cell Res. *30*, 717–731.

80. Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., and Guo, J. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. *49*, W317–W325. https://doi.org/10.1093/nar/gkab447.

81. Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371. https://doi.org/10.1038/s41586-019-1652-y.

82. Stewart, B.J., Ferdinand, J.R., Young, M.D., Mitchell, T.J., Loudon, K.W., Riding, A.M., Richoz, N., Frazer, G.L., Staniforth, J.U.L., Vieira Braga, F.A., et al. (2019). Spatiotemporal immune zonation of the human kidney. Science *365*, 1461–1466. https://doi.org/10.1126/science.aat5031.

83. Elmentaite, R., Kumasaka, N., Roberts, K., Fleming, A., Dann, E., King, H.W., Kleshchevnikov, V., Dabrowska, M., Pritchard, S., Bolt, L., et al. (2021). Cells of the human intestinal tract mapped across space and time. Nature *597*, 250–255. https://doi.org/10.1038/s41586-021-03852-1.

84. Miao, Y., Tian, L., Martin, M., Paige, S.L., Galdos, F.X., Li, J., Klein, A., Zhang, H., Ma, N., Wei, Y., et al. (2020). Intrinsic Endocardial Defects Contribute to Hypoplastic Left Heart Syndrome. Cell Stem Cell *27*, 574–589.e8. https://doi.org/10.1016/j.stem.2020.07.015.

85. Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience *7*, 1–6. https://doi.org/10.1093/gigascience/gix120.

86. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

87. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008.

88. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.

89. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

90. Agrawal, A., Chiu, A.M., Le, M., Halperin, E., and Sankararaman, S. (2020). Scalable probabilistic pca for large-scale genetic variation data. PLoS Genet. *16*, e1008773.

91. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122.

92. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

93. Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience *8*, giz082. https://doi.org/10.1093/gigascience/giz082.

94. Slowikowski, K., Hu, X., and Raychaudhuri, S. (2014). SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. Bioinformatics *30*, 2496–2497.

95. Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F.L., Yang, H.M., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., et al. (2003). The International HapMap Project. Nature *426*, 789–796. https://doi.org/10.1038/nature02168.

96. The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.

97. Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

98. Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3. 0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum. Mutat. *37*, 235–241.

99. Liu, S., Liu, Y., Gu, Y., Lin, X., Zhu, H., Liu, H., Xu, Z., Cheng, S., Lan, X., Li, L., et al. (2024). Utilizing Non-Invasive Prenatal Test Sequencing Data Resource for Human Genetic Investigation. Cell Genom. *4*, 100669. https://doi.org/10.1016/j.xgen.2024.100669.

100. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat. Rev. Genet. *11*, 499–511.

101. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909. https://doi.org/10.1038/ng1847.

102. Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E., and Greenwood, C.M.T.; UK10K Consortium (2014). Estimating genome-wide significance for whole-genome sequencing studies. Genet. Epidemiol. *38*, 281–290.

103. The FANTOM Consortium and the RIKEN PMI and CLST DGT (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

104. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

105. Liu, H.-K., He, S.-J., and Zhang, J.-G. (2021). A bioinformatic study revealed serotonergic neurons are involved in the etiology and therapygenetics of anxiety disorders. Transl. Psychiatry *11*, 297.

106. Liu, H., Guan, L., Deng, M., and Bolund, L. (2023). Integrative genetic and single cell RNA sequencing analysis provides new clues to the amyotrophic lateral sclerosis neurodegeneration. Front. Neurosci. *17*, 1116087. https://doi.org/10.3389/fnins.2023.1116087.

107. Wang, C., Liu, H., Li, X.-Y., Ma, J., Gu, Z., Feng, X., Xie, S., Tang, B.-S., Chen, S., Wang, W., et al. (2024). High-depth whole-genome sequencing identifies structure variants, copy number variants and short tandem repeats associated with Parkinson's disease. NPJ Parkinsons Dis. *10*, 134. https://doi.org/10.1038/s41531-024-00722-1.

108. Liu, H., Guan, L., Su, X., Zhao, L., Shu, Q., and Zhang, J. (2024). A broken network of susceptibility genes in the monocytes of Crohn's disease patients. Life Sci. Alliance *7*, e202302394.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
gnomAD database	Karczewski et al.[78]	https://gnomad.broadinstitute.org/news/2020-10-gnomad-v3-1
ChinaMAP database	Cao et al.[79]	http://www.mbiobank.com
KOBAS database	Bu et al.[80]	http://kobas.cbi.pku.edu.cn
GWAS Catalog database	Buniello et al.[80]	https://www.ebi.ac.uk/gwas
602 trios WGS dataset	Byrska-Bishop et al.[35]	https://www.internationalgenome.org/data-portal/data-collection
Liver scRNA dataset	Popescu et al.[81]	E-MTAB-7407
Liver scRNA dataset	MacParland et al.[53]	GEO: GSE115469
Kidney scRNA dataset	Stewart et al.[82]	http://www.kidneycellatlas.org
Gut scRNA dataset	Elmentaite et al.[83]	https://www.gutcellatlas.org
Heart scRNA dataset	Miao et al.[84]	GEO: GSE138979
Pubertal growth summary	Cousminer et al.[70]	https://egg-consortium.org/pubertal-growth.html
T2D summary	Mahajan et al.[66]	https://diagram-consortium.org/downloads.html
Childhood obesity summary	Bradfield et al.[73]	https://egg-consortium.org/childhood-obesity-2019.html
Childhood BMI summary	Helgeland et al.[11]	https://www.fhi.no/en/ch/studies/moba/for-forskere-artikler/gwas-data-from-moba/
Neonatal metabolites summary	Our study	https://ngdc.cncb.ac.cn/omix/release/OMIX004498
Software and algorithms		
SOAPnuke	Chen et al.[85]	https://github.com/BGI-flexlab/SOAPnuke
bwa	Li et al.[86]	https://github.com/lh3/bwa
samtools/bcftools	Danecek et al.[87]	http://samtools.github.io
sambamba	Tarasov et al.[88]	https://github.com/biod/sambamba
GATK	McKenna et al.[89]	https://github.com/broadinstitute/gatk
SeqFF	Kim et al.[74]	https://github.com/HyunbinCho/seqff
propca	Agrawal et al.[90]	https://github.com/sriramlab/ProPCA
BaseVar	Liu et al.[25]	https://github.com/Zilong-Li/BaseVarC
STITCH	Davies et al.[25]	https://github.com/rwdavies/STITCH
vawk	–	https://github.com/cc2qe/vawk
VEP	McLaren et al.[91]	https://github.com/Ensembl/ensembl-vep
METAL	Willer et al.[92]	https://github.com/statgen/METAL
GCTA	Yang et al.[59]	https://yanglab.westlake.edu.cn/software/gcta
PRSice	Choi et al.[93]	https://choishingwan.github.io/PRSice
SNPsea	Slowikowski et al.[94]	http://broadinstitute.org/mpg/snpsea
PhenoScanner	Staley et al.[47]	http://www.phenoscanner.medschl.cam.ac.uk
Necessary code	Our study	https://github.com/liuhankui/NIPT https://doi.org/10.5281/zenodo.13692304

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study subjects

According to the maternal ID of NIPT and neonatal ID of newborn metabolic screening between 2015 and 2020, 27,785 pairs of maternal NIPT low-pass WGS data and neonatal metabolic data were selected. Participants with positive results on chromosomal abnormalities, abnormal ultrasound fetuses, or abnormal neonatal metabolic screening result have been excluded. 8,744 participants were sequenced by the Ion Torrent platform and 19,041 participants were sequenced by the Illumina CN500 platform. Plasma was divided from maternal peripheral blood (5 mL) by centrifuging and then 600 μL (Ion Torrent) or 1.4 mL (Illumina CN500) plasma was used to extract cell-free DNA. The library construction, quality control, pooling, and sequencing were performed according to their respective sequencing experiments instruction of the Ion Torrent platform and Illumina CN500 platform, respectively. Neonatal

metabolites (11 kinds of amino acid, 31 kinds of carnitine, and succinylacetone) were quantified by mass spectrum (Waters TQ Detector) in neonatal dried blood spots.

The written informed consent was approved by Institutional Review Board of Suzhou Municipal Hospital at 2015 and obtained from each participant from 2015 to 2020. The study was approved by the Institutional Review Board of Suzhou Municipal Hospital (K-2021-024-H01) and BGI group (BGI-IRB 23097). The genetic data collecting is approved ([2022] CJ2133) by the Human Genetic Resources Administration of China.

METHOD DETAILS

SNV calling

Raw sequencing reads with a proportion of gap or low-quality base ($Q < 20$) at a threshold of >0.1 were filtered by SOAPnuke.[85] Clean reads were aligned to human genome reference (GRCh38) by bwa.[86] Duplicate reads probably generated from PCR were excluded by sambamba.[88] Reads around known insertions and/or deletions were re-aligned by the indel realignment modules of GATK.[89] Read qualities were rebuilt by base quality recalibration modules of GATK with references of variants from the HapMap project,[95] 1000 Genomes Project,[96] and dbSNP database.[97] Recalibrated reads were stored in cram format by samtools.[87] The fetal cfDNA proportion was estimated by SeqFF[74] from depth coverage calculated from cram. SNVs were detected and the corresponding mutant allele frequencies were estimated by BaseVar[25] from the cram files of the discovery cohort. Based on a suggestion from the author of BaseVar, we set 100 counts into a batch and set a no-overlap sliding window with 2 million bases to detect SNVs. SNVs met any of the three filter conditions were excluded form subsequent analysis: (1) quality <100; (2) minor allele frequency <1%; (3) read depth <500 or >5000. Variant Effect Predictor,[91] gnomAD,[78] ChinaMAP,[79] and dbNSFP[98] databases were used to annotate SNVs for obtaining useful information, including gene symbol, variant consequence, and allele frequency.

Genotype imputation

The individual genotypes of SNPs in the discovery cohort were imputed by STITCH.[20] A reference panel of the Chinese population of the 1000 Genomes Project[96] was used to initialize the ancestral haplotypes in an expectation–maximization algorithm of STITCH. We set 10 haplotypes in 100 generations as described in our previous study[25]/protocol[99] and imputed the SNPs in 5 million bases window with 250 kilobases buffer size. The imputed SNPs met any of the following indicators were excluded from subsequent association study: (1) p value (two-tailed) of Hardy-Weinberg balance test $<1 \times 10^{-5}$; (2) MiAF <1%; (3) the rate of genotyped individuals <90%. (4) $I_A < 0.4$. I_A was defined in IMPUTE2[100] as an information metric for estimating the accuracy of imputation. I_A takes values between 0 and 1, where a value near 1 indicates that an SNP has been imputed with high certainty.

Principal component analysis

Principal component analysis was performed by propca[90] on the imputed genotypes of SNPs with MiAF >5%. To decrease the memory and time of calculation, we pick up the first SNP from every 50 SNPs in sort VCF file and obtained 100,000 SNPs for PCA. The genotype of wild-type, heterogeneous, and homogeneous were re-coded to 0, 1, and 2 correspondingly and stored in a modified EigenStrat[101] format. The genetic structure of participants was displayed by the top two principal components. Linear regression was used to investigate the association between the proportion of fetal cfDNA and the top three principal components.

QUANTIFICATION AND STATISTICAL ANALYSIS

Association study

In the pre-GWAS, as the individual genotypes were imputed from a pool of maternal and fetal cfDNA and the fetal cfDNA was shown to contribute to the second principal component of genotype matrix, it is necessary to confirm any bias of imputed genotype for subsequent GWAS of phenotypes. A linear regression model was used to estimate the association between the proportion of fetal cfDNA and neonatal metabolism. p values (two-tailed) of multi-test were adjusted by Bonferroni method. No statistical evidence was observed for an association between fetal cfDNA and neonatal metabolism.

In the GWAS, a linear regression model was used to estimate the significance of the association between SNPs and each metabolic component by glm function in R-program, respectively. The neonatal weight/sex, maternal age/height/BMI, top three principal components of PCA were used as covariates in the regression model. p values (two-tailed) at a threshold $<5 \times 10^{-7}$ were used for determining genome-wide significance[102] in the discovery cohort. To validate the associations, the same method was used to calculate the significance of SNPs in the replication cohort. The statistical summary (β, p value, and sample size) of association in the discovery and replication cohort were used for meta-analysis by METAL.[92] SNPs with all p values of discovery, replication, and meta-analysis pass through the threshold of $<5 \times 10^{-7}$, $<2.68 \times 10^{-5}$, and $<6.67 \times 10^{-10}$ were considered to make a significant contribution to the variation of metabolite concentration. SNP with the smallest p value in a gene or in one million bases window of the intergenic region was defined as the lead SNP. Besides neonatal metabolic components, we also included two maternal traits—maternal height and BMI—for GWAS. The top three principal components of PCA and maternal age were used as covariates in linear regression model.

The gestational weeks and the proportion of fetal cfDNA were used as additional covariates for maternal BMI. The thresholds for the p value ($P_{discovery} < 5 \times 10^{-7}$, $P_{replication} < 2.68 \times 10^{-5}$, $P_{meta\text{-}analysis} < 5 \times 10^{-10}$) are consistent with that employed in the GWAS of metabolism.

Trend test of allele frequency and ranked phenotypic groups

Because the association is estimated from neonatal metabolism and maternal genotype, which is unobserved in neonates in our study, we designed a trend test to estimate the association between ranked phenotypic groups and corresponding allele frequency. Firstly, we sorted the neonatal metabolite concentrations and classified the neonates into five ranked groups. Subsequently, we estimated the allele frequency for each neonatal group. Finally, we performed a trend test via prop.trend.test R function on the allele frequencies of five ranked groups.

The allele frequency of each neonatal group in the second step was estimated by:

We estimated the allele frequency in the un-observed paternal population (p_{father}) from the allele frequency in our maternal population (p_{mother}), male population (p_{XY}) and female population (p_{XX}) of gnomAD database:

$$p_{father} = \widehat{w}\, p_{mother}$$

$$\widehat{w} = p_{xy}/p_{xx}$$

The proportions of genotypes (AA, Aa, and aa) in the un-observed paternal population were calculated according to H-W balance:

$$P(AA) = p^2, P(Aa) = 2pq, P(aa) = q^2$$

$$q = 1 - p$$

We assumed the mate-pairs of father and mother are random and they translate their allele to the offspring independently, the allele translation probability from one parent was estimated by:

$$\begin{aligned} P(Child = A, Parent) &= \sum_{i}^{n} P(\rightarrow A|G_i)P(G_i) \\ &= P(\rightarrow A|AA)P(AA) + P(\rightarrow A|Aa)P(Aa) + P(\rightarrow A|aa)P(aa) \\ &= P(AA) + 0.5\,P(Aa) = p^2 + pq = p \end{aligned}$$

Where $P(\rightarrow A|AA) = 1$, $P(\rightarrow A|Aa) = 0.5$, $P(\rightarrow A|aa) = 0$

The proportion of neonatal genotypes (*AA*, *aa*, and *Aa*) were calculated by:

$$P(Child = AA) = P(Child = A, Father) \times P(Child = A, Mother) = p_{mother}\, p_{father}$$

$$P(Child = aa) = P(Child = a, Father) \times P(Child = a, Mother) = q_{mother}\, q_{father}$$

$$P(Child = Aa) = 1 - p_{mother}\, p_{father} - q_{mother}\, q_{father} = p_{mother} + p_{father} - 2p_{mother}p_{father}$$

The allele frequency in neonatal population was calculated by:

$$P(A) = \frac{2P(AA)+P(Aa)}{2} = \frac{p_{mother}+p_{father}}{2} = \frac{1}{2}p_{mother}\left(1 + \frac{p_{xy}}{p_{xx}}\right)$$

Variant-variant interaction

SNPs with p value at a threshold of <0.01 in genome-wide association were used to study the interaction of pairwise SNPs. We set an interaction of two SNPs in linear model. p value thresholds ($P_{discovery} < 5 \times 10^{-7}$, $P_{replication} < 5 \times 10^{-7}$) were used to identify the significant interaction of pairwise SNPs.

Enrichment analysis

KEGG pathway enrichment analysis was performed in the KOBAS database.[80] p values were calculated by a one-tailed Fisher's exact test and adjusted by the false discovery rate (FDR) method. SNPsea[94] was used to screen the tissues and cells that are likely to be influenced by metabolism-associated loci. We employed 533 organs/tissues/cell types/cell lines from FANTOM[103] for screening. Empirical p values were calculated by permutation from null SNP sets.[104] Cell type expression enrichment analysis was performed in large-scale single-cell transcriptome datasets of liver,[53,81] heart,[84] kidney,[82] and gut[83] vi method described in our previous studies.[105–108] Briefly, the method first calculated the specificity of a gene in each cell type and then estimated a background distribution for the gene specificity by a bootstrap method at 10,000 times resampling. p values (one-tailed) of genes were

calculated by the cumulative density function of the background distribution. *p* values were adjusted by the FDR method and the adjusted *p* value at a threshold of <0.05 was used to indicate the significant enrichment.

Heritability and polygenic risk score

Kinship matrix of all participants was calculated from individual genotypes and used for heritability estimation in GCTA GREML[59] model:

$$A_{jk} = \frac{1}{N}\sum_{i}^{N} A_{ijk} = \begin{cases} \frac{1}{N}\sum_{i}^{N}\frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}, j \neq k \\ 1 + \frac{1}{N}\sum_{i}^{N}\frac{x_{ij}^2 - (1+2p_i)x_{ij} + 2p_i^2}{2p_i(1 - p_i)}, j = k \end{cases}$$

A refers to the kinship matrix, *j* and *k* refers to the *j* th and *k* th individual, *i* refers to the *i* th SNP of the *N* SNPs used for calculation, *x* refer to the expect dosage of genotype, *p* refers to the allele frequency of SNP. This formula indicate that we can used the kinship of mother to instead the kinship of neonates for neonatal heritability estimation. To demonstrate our hypothesis, we used the accuracy genotype from the high-depth WGS of 602 trios and showed that the pairwise correlation in neonatal kinship were highly correlated with that in maternal kinship. SNPs with *p* value at a threshold of <0.01 were used to estimate the heritability of neonatal metabolism and maternal height/BMI. Genotypes of SNPs were transformed to plink bed format and the kinship matrix of participants was calculated. Heritability was estimated by GREML analysis of GCTA with neonatal weight/gender, maternal age/BMI, and the top three principal components as covariates. To predict the neonatal metabolites concentration and maternal height/BMI from individual genotypes, SNPs with *p* value at a threshold of <0.001, <0.005, <0.01 and corresponding regression coefficient were used to calculate the polygenic risk score by PRSice.[93] A linear regression model with maternal height/BMI as covariates was used to estimate the association between polygenic risk score and phenotype. A training set of 5,000 participants in the discovery cohort was randomly selected to estimate the best-fit regression model. The remaining 3,744 participants in the discovery cohort were used to test the performance of the best-fit regression mode. Mean-square error (MSE) was used to estimate the accuracy of polygenic risk score model. The variance (VAR) was calculated from the observation. MSE/VAR was used to compare the accuracy between different metabolic components and a threshold of MSE/VAR $\leq$0.2 was set to indicate the high-accuracy model.