BMC Medicine

RESEARCH ARTICLE

Open Access

Unraveling the metabolomic architecture of autism in a large Danish population-based cohort

Filip Ottosson[1,2*†], Francesco Russo[1,2†], Anna Abrahamsson[1], Nadia MacSween[1], Julie Courraud[1,2,3,4], Kristin Skogstrand[1,2], Olle Melander[5], Ulrika Ericson[5], Marju Orho-Melander[5], Arieh S. Cohen[1,2,6], Jakob Grove[2,7,8,9], Preben Bo Mortensen[2,10,11], David M. Hougaard[1,2] and Madeleine Ernst[1,2*]

Abstract

Background The prevalence of autism in Denmark has been increasing, reaching 1.65% among 10-year-old children, and similar trends are seen elsewhere. Although there are several factors associated with autism, including genetic, environmental, and prenatal factors, the molecular etiology of autism is largely unknown. Here, we use untargeted metabolomics to characterize the neonatal metabolome from dried blood spots collected shortly after birth.

Methods We analyze the metabolomic profiles of a subset of a large Danish population-based cohort (iPSYCH2015) consisting of over 1400 newborns, who later are diagnosed with autism and matching controls and in two Swedish population-based cohorts comprising over 7000 adult participants. Mass spectrometry analysis was performed by a timsTOF Pro operated in QTOF mode, using data-dependent acquisition. By applying an untargeted metabolomics approach, we could reproducibly measure over 800 metabolite features.

Results We detected underlying molecular perturbations across several metabolite classes that precede autism. In particular, the cyclic dipeptide cyclo-leucine-proline (FDR-adjusted $p = 0.003$) and the carnitine-related 5-aminovaleric acid betaine (5-AVAB) (FDR-adjusted $p = 0.03$), were associated with an increased probability for autism, independently of known prenatal and genetic risk factors. Analysis of genetic and dietary data in adults revealed that 5-AVAB was associated with increased habitual dietary intake of dairy (FDR-adjusted $p < 0.05$) and with variants near SLC22A4 and *SLC22A5* ($p < 5.0e-8$), coding for a transmembrane carnitine transporter protein involved in controlling intracellular carnitine levels.

Conclusions Cyclo-leucine-proline and 5-AVAB are associated with future diagnosis of autism in Danish neonates, both representing novel early biomarkers for autism. 5-AVAB is potentially modifiable and may influence carnitine homeostasis.

Keywords Autism, Metabolomics, Neonatal, Dried blood spot, Metabolome, 5-Aminovaleric acid betaine, Cyclo-leucine-proline

†Filip Ottosson and Francesco Russo contributed equally to this work.

*Correspondence:
Filip Ottosson
faot@ssi.dk
Madeleine Ernst
maet@ssi.dk
Full list of author information is available at the end of the article

Background

Autism refers to a range of neurodevelopmental conditions characterized by impaired social ability, impaired communication, and repetitive behavior. The prevalence of autism in Denmark has been increasing, reaching 1.65% among 10-year-old children in 2017 [1]. Similar trends are seen elsewhere [2, 3]. Although there are several factors associated with autism, including genetic [4] and prenatal factors [5], the molecular etiology of autism is largely unknown. The gut microbiome has been increasingly implicated in autism, as patients with autism commonly experience gastrointestinal issues [6] and have an altered gut microbiota [7, 8]. Dysbiosis of the maternal gut microbiome could even contribute to altered fetal neurodevelopment related to autism [9]. The circulating metabolome reflects genetic [10], gut microbiome [11], and dietary intake variations [12, 13], potentially highlighting mediators between these exposures and autism risk. Detecting autism-related metabolite alterations in early life may facilitate an increased understanding of the molecular mechanisms behind autism, improve diagnosis, and contribute to preventive strategies.

Metabolomics has been used to characterize specific perturbations of the metabolome in children with autism compared to neurotypical controls, including amino acids [14], acylcarnitines [15], and various aromatic and phenolic compounds [15]. Moreover, a selection of metabolites in the folate-dependent one-carbon metabolism have been used to classify autism from typically developed counterparts, indicating the potential for metabolite-based methods to be used in autism diagnosis [16]. However, these studies reflect cross-sectional differences between autism cases and controls, which hampers the attempts to identify early pathophysiological mechanisms of autism, due to the large risk of reverse causation. Metabolites that prospectively associate with autism development have a higher chance of highlighting early protective or detrimental mechanisms of autism that can be targeted in novel preventive strategies.

To unravel underlying disease mechanisms and find tools for early diagnosis of autism, we analyzed the metabolomic profiles of a subset of a large Danish population-based cohort (iPSYCH2015). The subset is a case–control study consisting of over 1400 newborns, including neonates who later were diagnosed with autism, and matching controls. By using untargeted metabolomics of neonatal dried blood spot samples (DBS), we aim at identifying metabolite features associated with autism, in order to better understand the disease etiology. Metabolites that were associated with autism were further investigated in independent cohorts comprising over 7000 adult individuals, in order to describe the genetic and dietary determinants.

Methods

Cohorts

iPSYCH cohorts

The neonatal cohort consisted of individuals who were included in either the population-based cohort or as a case in iPSYCH2015 [17], which is a case-cohort study of psychiatric disorders, nested within the Danish population born between 1981 and 2008. The neonates included in the current case–control study were born between 2003 and 2008. All individuals had available information about gestational age, age at sampling, and neonatal DBS available. Additionally, all samples passed quality control for genotyping. Cases of autism were all male and defined as the first diagnosis of childhood autism (ICD-10: F84.0) in the Danish Psychiatric Central Research Register in 2015 or earlier, at age 1 or older. Controls were selected among males in the population-based cohort in iPSYCH and with none of the following ICD-10 diagnosis codes in the Danish Psychiatric Central Research Register in 2015 or earlier, at age 1 year or older: F84.0, F84.1, F84.2, F84.3, F84.5, F84.8, F84.9, F90.0. Matching of controls was performed based on birth date +/− 14 days, gestational age +/− 4 days, and age at sampling +/− 1 day (neonatal age at time of DBS collection). For neonates born before 2006, autism cases were excluded if diagnosed with mental retardation (F70-F79) before the 10-year birthday in the Danish Psychiatric Central Research Register as of April 2017. Family history of psychiatric disorders was defined as any parent having registry entries (ICD-8:290–315, ICD10:F00-F99) before the end of 2016. Seasons are defined as starting from the 1st day of the month: 1st Dec.–28th Feb. is winter, 1st March–31st May is spring, 1st June–31st August is summer, 1st Sept.–30th Nov. is autumn.

Malmö cohorts

The Malmö Diet and Cancer Study (MDC) is a population-based prospective cohort consisting of 28,449 individuals. The cardiovascular cohort of MDC was designed to study the epidemiology of carotid artery disease. Enrolling participants between 1991 and 1996 [18]. Citrate plasma was available for 3833 among the 5405 participants with fasted blood samples. Dietary intake data was available for 3714 individuals and genome-wide genotyping for 3409 individuals.

Malmö Offspring Study (MOS) is an ongoing population-based cohort study where adult (>18 years old) children and grandchildren from the MDC study are recruited [19]. Participants were invited through letter and visited the research clinic where overnight fasting EDTA plasma samples were collected and anthropometric measurements performed. Plasma samples were

available for 3430 participants. Dietary intake data was available for 1539 individuals.

Metabolomics profiling

DBS samples (3.2-mm-diameter punches) were randomly distributed over nineteen 96-well plates (batches). A batch of DBS consisting of adult blood was created before sample preparation and stored at − 20 °C, referred to as external control (EC) samples. Both EC samples and plate-specific pools of neonatal samples were analyzed for quality control purposes. Sample preparation was performed by extraction in 80% methanol by being incubated for 45 min and subsequent centrifugation. The sample extract (supernatant) was evaporated under nitrogen before being reconstituted in 95% solvent A (99.8% water and 0.2% formic acid) and 5% solvent B (49.9% methanol, 49.9% acetonitrile and 0.2% formic acid). Mass spectrometry analysis was performed using a timsTOF Pro mass spectrometer coupled to a UHPLC Elute LC system, Bruker Daltonics (Billerica, MA, US). The analytical separation was performed on an Acquity HSS T3 (100 Å, 2.1 mm × 100 mm, 1.8 µm) column (Waters, Milford, MA, US). The analysis started with 99% solvent A for 1.5 min, thereafter a linear gradient to 95% solvent B during 8.5 min followed by an isocratic condition at 95% mobile phase B for 2.5 min before going back to 99% mobile phase A and equilibration for 2.4 min. Metabolomics preprocessing was done using the Ion Identity Network workflow in MZmine [20, 21] (version 3.3.0). Before statistical analysis, metabolite features present in less than 25% of the samples were removed and features present in fewer than 75% were treated as binary variables (present or absent). This resulted in a final dataset with a total of 865 metabolite features measured, among which 452 features were continuous and 413 were binary variables. Missing values for metabolite features with continuous measurements were further subjected to imputation using missForest [22] and subsequent batch correction in WaveICA [23]. Annotation of metabolite features was performed using mass spectral molecular networking through the GNPS Platform, unsupervised substructure discovery using MS2LDA, in silico annotation through Network Annotation Propagation, Sirius + CSI:FingerID, MolNetEnhancer and deep neural networks in CANOPUS. Metabolite annotation levels were defined according to the Metabolite Standards Initiative [24]. Metabolites were annotated at level 2 (plausible 2D structure) and level 3 (plausible metabolite class). Detailed descriptions of sample preparation, mass spectrometry analysis, preprocessing, annotation, and quality control procedures can be found in the Additional file 1: supplementary information [20–23, 25–35].

Relative quantification of 5-aminovaleric acid betaine in Malmö cohorts

Untargeted metabolomics profiles of plasma samples from MOS and MDC were acquired on a UPLC-QTOF (1290 LC, 6550 MS; Agilent Technologies, Santa Clara, CA) as previously described [36]. Relative abundance of 5-aminovaleric acid betaine was retrieved by integrating the peak areas for the *m/z* putatively annotated as 5-AVAB (*m/z* = 160.13) using Agilent Profinder B.06.00 (Agilent Technologies). Normalization was performed using *m/z* 160.13 measurements in repeatedly injected identical pooled quality control (QC) samples. A low-order nonlinear locally estimated smoothing (LOESS) function was fitted to the *m/z* 160.13 areas in the QC samples as a function of the injection order. The α-parameter (proportion of QC samples used for the correction curve) was set to 2/3. The resulting correction curve was used to normalize the analytical samples as described previously [37]. MS/MS data was acquired at 20 eV and an isolation width of 1.3 m*/z*. Annotation of *m/z* 160.13 was performed using a MASST [38] search of the fragmentation spectra of *m/z* 160.13, where matching spectra where defined as cosine > 0.7, minimum matched peaks = 4, parent mass tolerance < 0.05 Da and ion tolerance < 0.05 Da.

Genome-wide association study and genetic risk score

Genome-wide genotyping of participants from the MDC was conducted using the Illumina GSA Bead Chip (Illumina Inc, San Diego, CA). Genome-wide association studies (GWAS) for DBS and plasma levels of metabolites were performed using a linear additive genetic model adjusting for age, sex (adult cohort only), genetic principal components, and autism status (neonatal cohort only), using PLINK v1.9 [39]. Manhattan plots and locus zoom plots were generated in R, using packages *qqman* and *locuszoomr* packages, respectively. The polygenic score (PRS) is a cross-validated internally trained score that was generated by splitting the iPSYCH sample in 50 random, non-overlapping samples of roughly equal size. For each of the 50 subsamples, a GWAS was run on the complement and the sumstats of that was used to train a PRS in the subsample in question. We employed the SBayesR [40] algorithm as implemented in LDAK [41] following the best practices of LDAK. The score was subsequently standardized by subset strata. Principal components analysis was performed in PLINK v1.9 [39].

Dietary intake assessment

Dietary intakes in the MDC were assessed with a method combining a 7-day menu book, a food frequency questionnaire, and a 45-min interview [42].

In the MOS dietary intake for participants was assessed with a web-based 4-day food record, Riksmaten2010, developed by the Swedish National Food Agency [43]. In both cohorts, six dietary intake groups were energy-adjusted, by dividing intakes with non-alcoholic energy intake.

Statistical analysis

Associations between metabolite features and autism were analyzed using logistic regression models. Nominally significant associations were defined as $P<0.05$, and significant associations as false discovery rate (FDR)-adjusted $P<0.05$. The primary model was adjusted for gestational age, age at sampling, season of birth, and birth year. Significant features (FDR-adjusted $p<0.05$) were analyzed as quartiles (5-AVAB) or the secondary model was additionally adjusted for family history of psychiatric disorders, a genetic risk score for autism, and six genetic principal components. The analyses were also stratified for age of diagnosis (age of diagnosis < 6). Enrichment for metabolite classes was analyzed using Fisher's exact test. Associations between metabolite features and family history of psychiatric disorders were analyzed using logistic regression models. Corresponding analyses for metabolite features and genetic risk for autism were conducted using linear regression. Analyses for family history of psychiatric disorders were adjusted for gestational age, age at sampling, season of birth, and birth year, while analyses on genetic risk score for autism further were adjusted for the first six genetic principal components. Inter-correlation analyses between metabolite features were conducted using Partial Pearson's correlation tests, adjusted for all metabolite features in each respective dataset. Correlations between metabolites and habitual dietary intakes were conducted using Partial Spearman's correlation tests, adjusted for age, sex, and body-mass index. Analyses using partial correlations were performed in the *ppcor* [44] package. Phenome-wide analysis of 5-aminovaleric acid betaine was performed using least absolute shrinkage selection operator (LASSO) regression. A full description of the LASSO method can be found in Additional File 1:supplementary material. All statistical analysis was performed in R 4.2.1 and Jupyter Notebooks are accessible at: https://github.com/ssi-dk/CD-MRG-metabolomics_autism.

Results

Study cohort

We utilized a cohort of 1478 male Danish neonates born between 2003 and 2008, comprising 739 newborns who were diagnosed with autism before the age of 10, and 739 typically developed individuals. Since the neonatal DBS metabolome is influenced by the age at sampling (3 to 8 days), season of birth [45], gestational age [46] (gestational week 34 to 42), and biobank storage time (15–18 years) [47], these factors were used to match each case of incident autism with a control neonate. Cases and controls were very similar in terms of gestational age, age at sampling, season of birth, and age of mother (Table 1). Family history of mental disorders were significantly more common in incident autism cases compared to controls ($p<0.001$). The mean age of autism diagnosis was 6.0 years. A total of 865 metabolites (mass spectral features with unique MS/MS fragmentation patterns) were measured and present in at least 25% of the samples in the discovery cohort. Putative annotation on the metabolite class level was conducted by combining mass spectral molecular networking (GNPS) [26], unsupervised substructure discovery (MS2LDA) [27], in silico annotation through Network Annotation Propagation

Table 1 Characteristics of the neonatal study cohort, presented as mean values for continuous variables (standard deviation within parentheses) and percentage for categorical variables (numbers within parentheses). *P*-values are calculated using *t*-test for continuous variables and chi-square test for categorical variables

	Cases (*N* = 739)	Controls (*N* = 739)	*P*
Age at sampling (years)	5.8 (± 1.2)	5.8 (± 1.1)	0.67
Gestational age (weeks)	40.0 (± 1.3)	40.0 (± 1.2)	0.85
Born in winter	22.7% (168)	22.7% (168)	0.99
Born in spring	24.4% (180)	25.0% (185)	
Born in summer	25.7% (190)	25.4% (188)	
Born in autumn	27.2% (201)	26.8% (198)	
Age of mother (years)	30.8 (± 5.1)	30.8 (± 4.7)	0.96
Age of autism onset (years)	6.0 (± 2.0)	-	-
Family history of mental disorder	30.0% (222)	18.7% (138)	< 0.001

[30], Sirius + CSI:FingerID [48], MolNetEnhancer [32], and deep neural networks in CANOPUS [33]. This resulted in level 2 annotations for 111 (12.8%) metabolite features and level 3 annotations for 229 (26.5%) metabolite features [24].

Metabolites from a wide range of biochemical classes are associated with autism

In total, 99 metabolites had nominally significant associations ($p < 0.05$) with development of autism, as evaluated using logistic regression models adjusted for gestational age, age at sampling, season of birth, and birth year (Fig. 1A) (Additional File 2: Table S1). These included 18 metabolites that could be assigned a structural annotation (level 2 annotation), additionally 15 metabolites with a putative metabolite class, and 66 unknown features. The annotated features included six acylcarnitine-related metabolites (5-aminovaleric acid betaine, C2:0-carnitine, C8:0-carnitine, C10:0-carnitine, C10:3-carnitine, and C16:0-carnitine), cyclic dipeptides (cyclo-leucine-proline and cylo-proline-valine), tryptophan metabolites (indole-3-acetic acid and indole-carboxaldehyde), methionine cycle metabolites (methionine and betaine), alanine, uric acid, creatine, amino adipic acid, glycerophosphocholine and pantothenic acid.

After adjusting for multiple testing, only two of the 99 metabolites remained significant, cyclo-Leucine-Proline (cLP) (FDR-adjusted $p = 3.3e-3$) (Additional File 1: Fig. S1) and 5-aminovaleric acid betaine (5-AVAB) (FDR-adjusted $p = 0.031$) (Additional File 1: Fig. S2). Being in the top quartile 5-AVAB was associated with a 69% increased probability of future autism compared to being in the lowest quartile (Fig. 1B). Since measurable levels of cLP were detected in less than 75% of the neonatal cohort (33%), it was treated as a binary variable in the regression analyses. Having measurable levels of cLP was associated

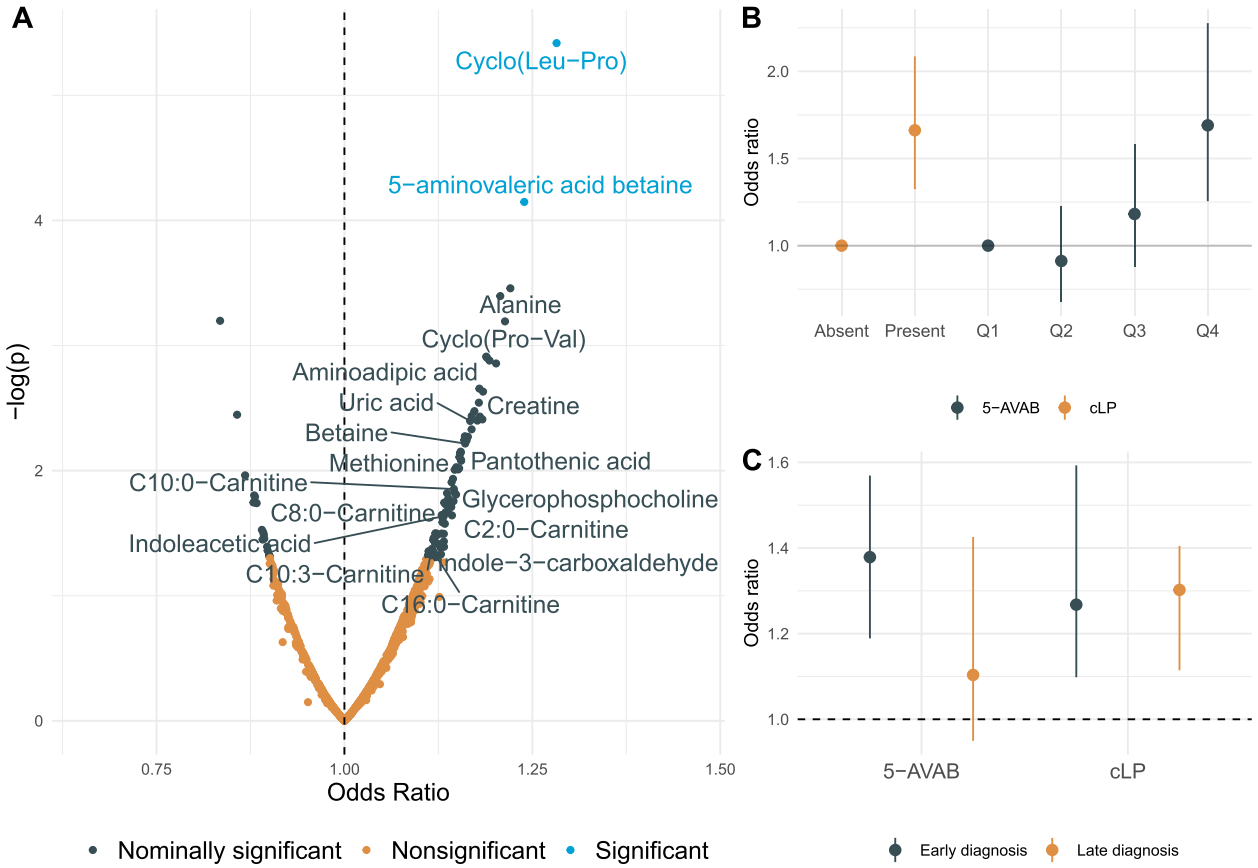

Fig. 1 Associations between metabolites in neonatal DBS and future probability of autism spectrum disorder. Odds ratios are calculated using logistic regression models, adjusted for age at sampling, gestational age, birth year, and season of birth. **A** Associations between all measured metabolite features and autism, with significance threshold defined as false discovery rate adjusted $p < 0.05$. Odds ratios are expressed as increased odds of autism per standard deviation increment of metabolite level. Significant features with an annotation confidence of level 2 or higher are named in the figure. **B** Associations between 5-aminovaleric acid betaine (5-AVAB) (quartiles) and cyclo-leucine-proline (cLP) (binary) and autism. Odds ratios are expressed as compared to the reference category ("Absent" for cLP and quartile 1 [Q1] for 5-AVAB). **C** Associations with autism stratified for age of diagnosis (< 6 years of age) for 5-AVAB and cLP. Odds ratios are expressed per standard deviation increment of metabolite level

with a 70% increased odds of developing autism in the future (Fig. 1B).

Given that autism is partly dependent on genetic factors, we investigated whether the neonatal metabolome was associated with a genetic risk score for autism and with a family history of psychiatric disorders. No metabolites were significantly associated (FDR-adjusted $p<0.05$) with the genetic risk score of autism, but 24 nominally significant associations were found (Additional File 1: Fig. S3). The strength of the association between the genetic risk score (AUC = 0.58) and autism was similar as what was found for cLP (AUC = 0.55) and 5-AVAB (AUC = 0.55). There were eight metabolites associated with family history of psychiatric disorders (FDR-adjusted $p<0.05$), including 5-AVAB, C8:1-carnitine, stachydrine, and trimethyllysine, while a total of 127 metabolites had nominally significant associations ($p<0.05$) (Additional File 1: Fig. S4). However, adjustments for family history of mental disorders and genetic risk score for autism only had a minor impact on the association between metabolites and future autism (Additional File 2: Table S2).

Speculating that age of diagnosis is related to the severity of autism traits and a higher likelihood of metabolic markers of autism being present early in life, we performed analyses stratified on age of diagnosis. Overall, the association between the neonatal metabolome and autism was stronger for individuals diagnosed before ($N=391$), compared to after ($N=348$) the age of six. There were 40 nominally significant associations for early-diagnosed individuals and 13 nominally significant metabolites for individuals diagnosed late. Also, two metabolites, 5-AVAB and an unknown feature (*m/z* 247.13, 3.91 min) remained associated with early diagnosis of autism after adjustments for multiple testing (Additional File 2: Table S3). The association between 5-AVAB and autism was significantly stronger (p for interaction = 0.032) in autism cases diagnosed at the age of 6 years or earlier. For cLP the associations were similar for cases diagnosed before ($p=1.2e-3$) and after ($p=8.8e-4$) the age of six (Fig. 1C).

Cyclo(leucine-proline) and autism

cLP and the structurally similar cyclo(pro-val) belong to the metabolite class diketopiperazines (cyclic dipeptides), both associated with increased probability of autism in the current study. Both metabolites were detected in less than 75% of the neonates and were thus treated as binary variables. For cLP, cyclo(pro-val) was the most strongly correlated metabolite (rho = 0.69), while the caffeine metabolite paraxanthine was the most negatively correlated (rho = −0.10) (Fig. 2A). We did not find any significant correlation between cLP and caffeine in the neonatal DBS (rho = 0.00 $p=0.97$). Moreover, including both cLP and cyclo(pro-val) as covariates in the regression model on autism clearly attenuated the associations for both metabolites (Additional File 2: Table S4), indicating that

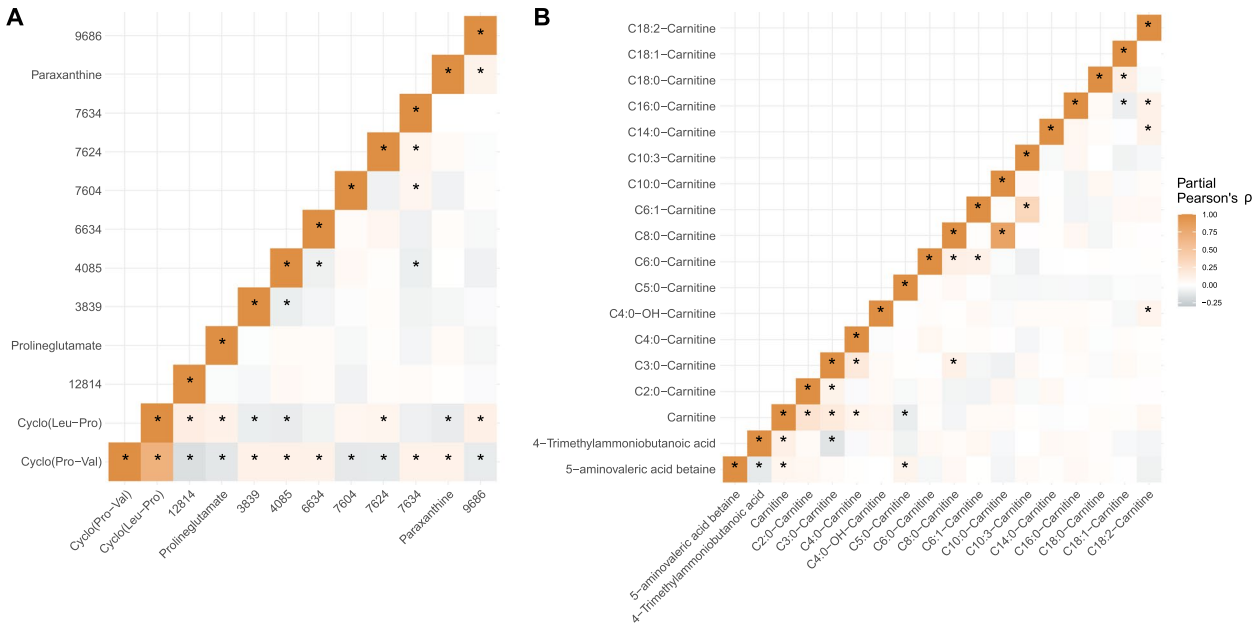

Fig. 2 Correlations between the two metabolites significantly associated with autism and other detected metabolites. Correlation coefficients are partial Pearson's correlation coefficients adjusted for all metabolite features ($N=865$). Significant correlations are indicated with an * at $p<0.05$. **A** Correlations between acylcarnitine-related metabolites. **B** Correlations between cyclo(leucine-proline) and other detected metabolites

the two diketopiperazines may have a shared pathophysiological relation to autism.

5-Aminovaleric acid betaine and autism

5-AVAB is a key regulator of intracellular carnitine availability, since it influences both cellular uptake [49] and biosynthesis of carnitine [50]. As aforementioned, we also found that elevated levels of several acylcarnitines were associated ($p < 0.05$) with the later development of autism. Metabolite class enrichment analysis showed that acylcarnitines were significantly enriched among nominally significant associations with future autism ($p = 0.015$) and with family history of mental disorders ($p = 0.013$) (Additional File 2: Table S5). Partial correlation analysis revealed that 5-AVAB was correlated with several metabolites in the acylcarnitine class (Fig. 2B). Notably, 5-AVAB showed positive correlations with C5:0-carnitine and free carnitine but negative correlation with the carnitine precursor 4-trimethylamoniumbutanoic acid. Adjusting the model between 5-AVAB and autism for the autism-associated acylcarnitines (C2:0-carnitine, C8:0-carnitine, C10:0-carnitine, C10:3-carnitine, and C16:0-carnitine), the associations were attenuated but remained significant (OR = 1.22, $P = 1.5e-3$).

If a causal link between 5-AVAB and autism probability can be established, it is important to examine how neonatal 5-AVAB levels can be modified. We have previously shown that 5-AVAB is vertically transferred from mother to infant, since the maternal levels, both during gestational week 24 and 1 week postpartum, are strongly correlated with the neonatal levels [51]. This suggests that the neonatal levels of 5-AVAB to a large extent are a reflection of the prenatal exposure of 5-AVAB. Understanding the key determinants of 5-AVAB in adults, should highlight potential strategies to modulate the maternal levels and thus the resulting prenatal exposure.

Dietary and genetic determinants of 5-aminovaleric acid betaine

To investigate determinants of 5-AVAB levels in adults, we hypothesized that their levels could be influenced by dietary or genetic factors. We used plasma metabolomics, genome-wide genotyping, and dietary intake data from two Swedish population-based cohorts, MDC ($N = 3833$) and MOS ($N = 3430$) (Additional File 2: Table S6). A feature with correct MS1 match (*m/z* error < 5 ppm) to 5-AVAB (*m/z* 160.13) was putatively annotated in both cohorts by matching acquired *m/z* 160.13 fragmentation spectra to online databases using MASST [38]. One spectrum achieved a match in both MDC (cosine = 0.83, mass difference < 0.01 and 5 shared peaks) and MOS (cosine = 0.81, mass difference < 0.01 and 4 shared peaks) with the uploaded 5-AVAB spectrum (Additional File 1: Fig. S5), and is thus referred to as 5-AVAB from here onwards. In both MDC and MOS, 5-AVAB was correlated similarly with acylcarnitine-related metabolites as in the cohort of neonates. This includes overall positive correlations with short-chain acylcarnitines and free carnitine along with negative correlations with 4-trimethylamoniumbutanoic acid for MOS only (Additional File 1: Fig. S6–S7).

By investigating six dietary intake groups in MDC ($N = 3714$), 5-AVAB correlated significantly (FDR-adjusted $p < 0.05$) with increased intake of dairy (rho = 0.18), decreased intake of meat (rho = −0.09), and decreased intake fruit and vegetables (rho = −0.07). Among specific dairy-related intakes, increased intake of milk (rho = 0.22) and yogurt (rho = 0.09) were significantly correlated (Fig. 3A). Similar correlations were seen for 5-AVAB in MOS ($N = 1539$), where dairy intake (rho = 0.15) was correlated with 5-AVAB (Fig. 3B). Intake of milk (rho = 0.12), but not other specific dairy intakes, was associated with 5-AVAB in both cohorts. In the stratified analysis, both total dairy and milk intake were significantly associated in men and women separately. In MDC, stronger correlations were observed for dairy in men compared to women, although no differences were observed in MOS (Additional File 2: Table S7).

By conducting a genome-wide association study (GWAS) in MDC ($N = 3409$), we found several genome-wide significant ($p < 5.e-8$) variants near the *SLC22A5* and *SLC22A4* genes (Fig. 4A), with rs272889 being the most strongly associated SNP ($p = 3.6E-20$) (Fig. 4B) (Additional File 2: Table S8). Stronger associations were seen in men (Additional File 1: Fig. S8) compared to women (Additional File 1: Fig. S9) for variants near both SLC22A4 and SLC22A5. Both dietary intake of dairy and the top SNP was among the 10 most significantly correlated variables in a phenome-wide scan for 5-AVAB determinants performed using 140 clinical, dietary, genetic, and metabolomic variables in MDC (Additional File 1: Fig. S10). In the neonatal cohort, there was no genome-wide significant association between SNPs and either 5-AVAB (Additional File 1: Fig. S11) (Additional File 2: Table S10) or cLP (Additional File 1: Fig. S12) (Additional File 2: Table S11).

Discussion

The main finding of our study is that levels of 5-AVAB and cLP are higher in newborns who later develop autism compared to neurotypical controls. We found additional putative associations between metabolites previously linked to autism in cross-sectional studies, indicating that these alterations may be present already at birth. To our knowledge, this is the first large-scale study that investigated the untargeted neonatal metabolome in

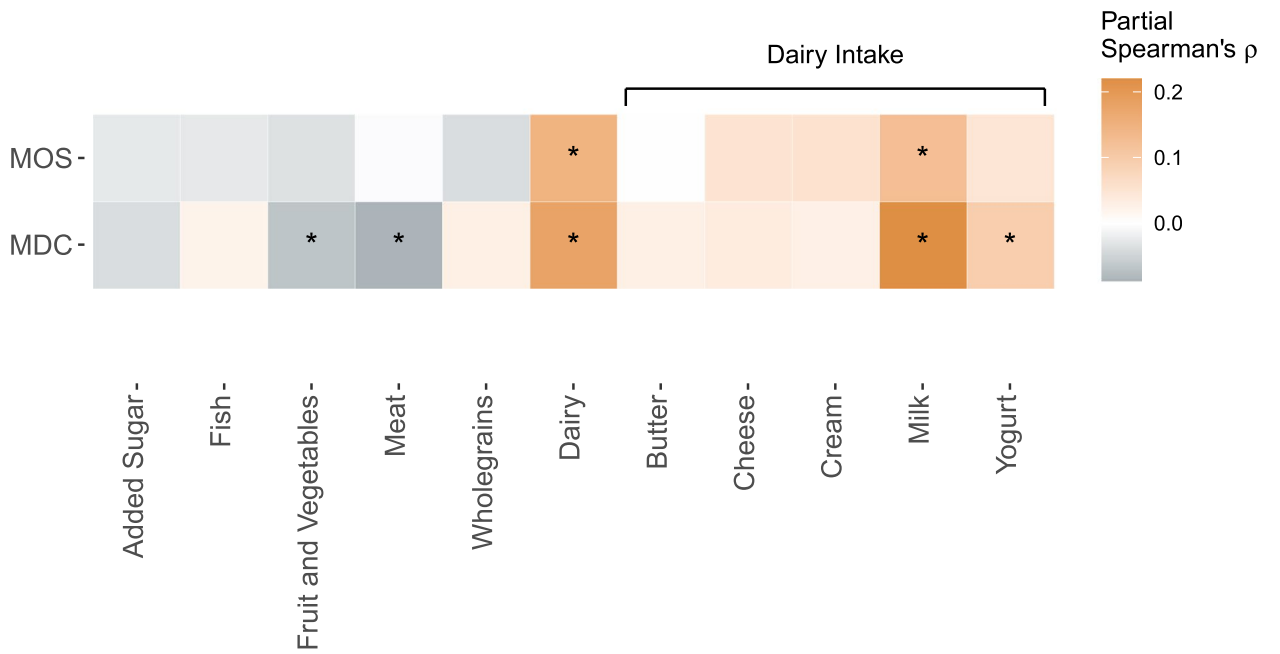

Fig. 3 Associations between plasma levels of 5-AVAB and dietary intakes in the Malmö Diet and Cancer Study ($N = 3714$) and the Malmö Offspring Study ($N = 1539$). The heatmap shows partial Spearman's correlations (adjusted for age, sex, and body mass index) between eleven dietary intake groups and 5-AVAB. Significance was defined as FDR-adjusted $p < 0.05$

relation to autism, paving the way for the identification of early metabolite biomarker candidates. We suggest that neonatal 5-AVAB levels largely reflect the maternal levels, which may be modified by dietary interventions.

Diketopiperazines, such as cLP, have been proposed to be involved in psychiatric disorders [52], have an observed ability to cross the blood–brain barrier [53], and are present in a variety of food and beverages [54]. Diet is a potential determinant of cLP, given that its level has been associated with increased intake of coffee [55] and alcohol [56] and decreased intake of starchy vegetables [56] in previous studies. However, we were not able to identify cLP in the Malmö adult cohorts, where data on nutrition was available, and can therefore only speculate on potential sources of cLP in our study. It is unlikely that maternal coffee consumption represents the major contribution to neonatal cLP, since caffeine was not correlated with cLP in the neonates. Results from GWAS show that genetic variants in *CYP1A2* and *CYP2C19* have been shown to be associated with cLP levels in adults [57]. Our finding that cLP was correlated with the caffeine metabolite paraxanthine, supports that neonatal levels of cLP could be partly regulated by CYP1A2 activity, since it is the enzyme mainly responsible for the degradation of caffeine. CYP2C19 is an enzyme involved in the metabolism of several drugs used for psychiatric disorders, including benzodiazepines and selective serotonin receptor inhibitors (SSRI) [58] and elevated expression has been associated with lower hippocampal volume in mice [59] and increased enzymatic capacity is associated with depression in humans [60]. Prenatal exposure to SSRIs is associated with an increased probability of autism [61], although meta-analyses indicate that at least some of the probability increase can be attributed to genetic confounding [62]. Furthermore, cLP could potentially originate from the gut microbiome, as it previously has been proposed to function as a quorum-sensing molecule for certain bacteria, while inhibiting the growth of other bacteria and fungi. Thus, gut microbes might produce cLP to induce growth suppression among competing bacterial species [63].

The finding that neonatal levels of the metabolite 5-AVAB is associated with increased risk of future development of autism confirms the results from our previous pilot study in 37 neonates and matching controls [45]. Based on our finding that 5-AVAB is more strongly associated with autism diagnosed before the age of six, we speculate that neonatal levels are connected to the severity of autism traits. Several studies in mice have indicated that 5-AVAB is involved in brain development, both in the fetus and adult mice. For instance, fetuses of germ-free (GF) mice lack several microbially produced metabolites, including 5-AVAB, and simultaneously display a reduced thalamocortical axonogenesis compared to conventionally raised mice. This effect was reversed by injection of 5-AVAB in the GF mice, preserving normal

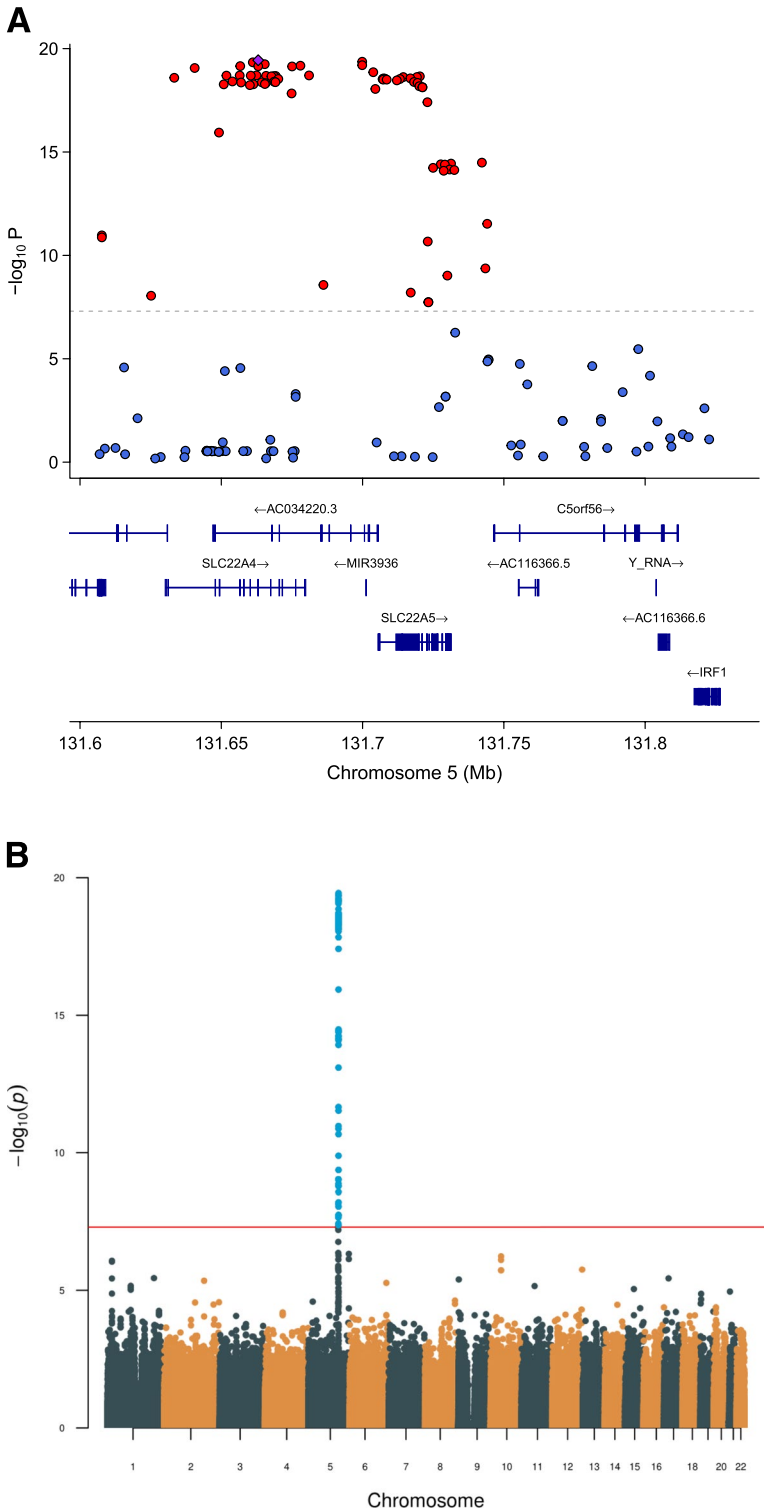

Fig. 4 Associations between single nucleotide polymorphisms (SNPs) and plasma levels of 5-amino valeric acid betaine (5-AVAB) in the Malmö Diet and Cancer Study (MDC) ($N = 3409$). **A** Locus zoom plot, indicating associations between SNPs near *SLC22A5*. **B** Manhattan plot of the genome-wide association study of plasma levels of 5-AVAB. Genome-wide significance threshold is indicated at $p < 5.0e-8$. Genome-wide significant SNPs are colored in blue and are located in the *SLC22A5* and *SLC22A4* genes

thalamocortical axonogenesis [64]. Given that early brain growth is often faster in children who later develop autism [65], one could speculate that high fetal brain levels of 5-AVAB may contribute to stimulating brain overgrowth. On the other hand, levels of 5-AVAB have been shown to be increased in the blood and brain of older adults and have negative effects on learning and memory [66]. Overall, these findings indicate that 5-AVAB may be involved in brain development, although the detailed functions are largely unknown.

Speculating that 5-AVAB is related to brain development, it is crucial to investigate how neonatal 5-AVAB levels can be modulated. We have previously shown that 5-AVAB may be directly transferred from the mother to the neonate, as the maternal plasma levels of 5-AVAB are strongly correlated with the neonatal levels of DBS [51]. Thus, it is possible that the neonatal 5-AVAB levels could be considered a proxy for the prenatal exposure to the maternal 5-AVAB.

The current and previous studies indicate that 5-AVAB in adults originates from endogenous production, dietary intake, and the gut microbiome, which is briefly discussed below. Endogenous production of 5-AVAB from the carnitine precursor trimethyllysine (TML) has been reported [67, 68] and has been proposed to occur via oxidative deamination followed by decarboxylation [69]. The exact production pathway and to what extent the total pool of circulating 5-AVAB originates from endogenous production is however not known. More evidence for an endogenous production of 5-AVAB comes from a recent study investigating the effects of predicted loss-of-function variants in the X-chromosome gene *TMLHE*[70]. *TMLHE* codes for the enzyme trimethyllysine dioxygenase, responsible for catalyzing the conversion of TML to hydroxytrimethyllysine, the first step in the carnitine biosynthesis. Predicted loss-of-function variants in *TMLHE* associate with increased levels of TML and 5-AVAB and with decreased levels of hydroxytrimethyllysine and 4-trimethylammoniobutanoic, both downstream of *TMLHE* in the carnitine biosynthesis pathway. As expected, the effects of the loss-of-function variants are significantly larger in men compared to women [70].

Although endogenous production is likely an important source of circulating 5-AVAB, there is mounting evidence that the gut microbiome also is a major contributor [49, 50]. These include several reports of lack of 5-AVAB in germ-free mice [49, 64], strong associations between microbial species in the human gut and levels of 5-AVAB in the brain [66] and plasma [11], and correlations with gut microbial diversity [11]. Moreover, levels of 5-AVAB in mice fetuses are distinctly diminished when maternal microbiota is lacking [71]. When maternal microbiota is present, 5-AVAB appears to localize to a wide variety of tissues, including the brain [71]. The relative contribution of the endogenous and the gut microbial production of 5-AVAB are not known and interspecies differences may exist. For instance, substantial differences in the biosynthesis of carnitine, where carnitine is produced from TML much more efficiently in mice compared to humans [72], may increase the bioavailability of TML for the production of 5-AVAB in humans. Although the gut microbiota is likely an important source of human circulating 5-AVAB, the exact microbial pathways to produce 5-AVAB are unknown. In recent years, the gut microbiome has increasingly been implicated in autism, as children with autism experience an increased prevalence of gastrointestinal issues [6] and have an altered gut microbiome [7, 8] compared to typically developing children. Maternal gut health has also been linked to the probability of autism in the offspring [9], suggesting that the maternal gut microbiome may influence health outcomes for the offspring. In this context, we propose that 5-AVAB may be a mediator in the gut-brain axis.

Dietary intake could be an alternative source of 5-AVAB and previous studies have shown that plasma 5-AVAB is correlated with increased intake of wholegrain [73] and milk [74, 75], which we further support by showing associations between habitual intake of total dairy, milk, and yogurt with increased plasma 5-AVAB in two large middle-age Swedish cohorts.

We found that plasma 5-AVAB in the adult was genome-wide associated with variants near the *SLC22A4* and *SLC22A5* genes, coding for a transmembrane transporter for carnitines and 5-AVAB, supporting results from previous GWAS [76, 77]. Our findings also indicated that the associations with variants near SLC22A5 and *SLC22A4* were stronger in men compared to women. Similar associations were not seen in neonates, indicating that other factors are more important in determining the neonatal level of 5-AVAB. It is however noteworthy that the neonatal GWAS was performed in a significantly smaller population, which limits the ability to rule out genetic determinants. Overall, our findings indicate that 5-AVAB levels are influenced by both genetic and dietary factors, in addition to the well-established link with the gut microbiome. If any of the above causal links between prenatal exposure to high 5-AVAB and autism development can be proven, 5-AVAB may represent an early and modifiable biomarker of autism.

The most well-described function of 5-AVAB is its ability to influence lipid metabolism via regulating the cellular uptake of carnitine and acylcarnitines. Cellular uptake of 5-AVAB is dependent on the carnitine transport protein *SLC22A5*, whereby it can compete with free carnitine and acylcarnitines for uptake into tissues [49, 67, 78] and across the blood–brain barrier [79]. Moreover,

5-AVAB has been shown to regulate the endogenous synthesis of carnitine in mice by competitive inhibition of γ-butyrobetaine hydroxylase, the enzyme catalyzing the rate-limiting step in the carnitine biosynthesis pathway, the hydroxylation of 4-trimethylammoniobutanoic acid (γ-butyrobetaine) to carnitine [50]. Thus, 5-AVAB regulates carnitine homeostasis by decreasing carnitine biosynthesis and lowering cellular uptake, which in concert may decrease fatty acid oxidation [49, 50] and lead to mitochondrial dysfunction [50], due to intracellular carnitine deficiency. Our study also indicates an intricate relation between 5-AVAB and acylcarnitines, where levels of 5-AVAB were correlated with higher circulating carnitine and short-chain acylcarnitines but lower levels of 4-trimethylammoniobutanoic acid. Deletions in *TMLHE* are relatively common (1/350 males), cause carriers to be auxotrophic for carnitine, and have also been linked to an increased probability of developing autism in boys [80, 81]. Several previous cross-sectional metabolomics studies show that children with autism have altered circulating levels of acylcarnitines [15, 82, 83] and free carnitine [15]. Carnitine supplementation in children with autism has resulted in improved behavioral scores [84], but larger studies are needed to confirm their efficacy. The directionality of the associations has not been completely consistent and one study indicates that the acylcarnitines in children with autism differ depending on the fatty acid side chain, where short-chain acylcarnitines generally were more abundant in autism compared to controls and vice versa for long-chain acylcarnitines [15]. In our study, this is reflected in correlations between acylcarnitines and 5-AVAB, with a gradient in correlations going from positive for short-chain to negative for long-chain features. However, the distinction between long- and short-chain features was not seen in the association with future autism. This leads us to propose two potential mechanisms linking 5-AVAB to autism. The results from the current studies are consistent with the theory of brain carnitine deficiency as a causal factor for autism and that 5-AVAB may act as a regulator of intracellular uptake of carnitine. An alternative hypothesis, where there is a causal link between 5-AVAB and autism, independent of carnitine, should also be further investigated. This could also explain the links between *TMLHE* deletions in males and autism, since *TMLHE* deficiency also results in increased TML and 5-AVAB levels.

Finally, several of the nominally significant metabolites have previously been linked to autism. Altered products of gut bacterial tryptophan metabolism have previously been reported in children with autism [85]. In our study, this was reflected by increased levels of indole-3-acetic acid and indole-3-carboxaldehyde. Uric acid has been reported to be elevated in the urine of children with autism [86]. The methionine cycle, including methionine and betaine, has been suggested to be altered in autism, but conflicting findings make its involvement in autism etiology unclear [87].

We acknowledge several limitations with our study. Although the described autism-associated alterations in the neonatal metabolome are present shortly after birth and precede autism diagnosis by several years, our study is observational and should be interpreted within that context. It is thus not possible to infer causal relationships between metabolites and the development of autism. Similarly, the associations shown by the GWAS do not prove a causative link between the gene near associated variants and 5-AVAB. However, our findings generate hypotheses for mechanisms linking the neonatal metabolites to autism, which should be further investigated in future studies, preferably using targeted metabolomics. Exclusion of co-occurring mental disorders was performed for neonates born between 2003 and 2005, but not for later births. A broader coverage of the metabolome could have been achieved by using a combination of different extraction solvents, chromatographic separation, and ionization mode. Since the sample material was limited, we aimed at using a single method with as broad coverage of annotated semi-polar metabolites as possible. We were unable to measure cLP in the independent middle-aged cohort, which limits our potential to investigate its genetic and dietary determinants. Moreover, the neonatal cohort lacks information about socioeconomic status and maternal parity, which are potential confounders in the association between the neonatal metabolome and autism. Finally, the neonatal cohort consisted of males only, which hinders us from drawing conclusions about the general population.

Conclusions

Cyclo-leucine-proline and 5-AVAB are associated with future diagnosis of autism in Danish neonates, both representing novel early biomarkers for autism. 5-AVAB is potentially modifiable and may influence carnitine homeostasis.

Abbreviations

5-AVAB	5-Aminovaleric acid betaine
AUC	Area under the curve
cLP	Cyclo-leucinepProline
DBS	Dried blood spot
EC	External control
FDR	False discovery-rate
GF	Germ free
GWAS	Genome-wide association study
ICD-10	International Statistical Classification of Diseases and Related Health Problems—revision 10
LASSO	Least absolute shrinkage selection operator
LC	Liquid chromatography
LOESS	Locally estimated scatterplot smoothing
MDC	The Malmö Diet and Cancer Study

MOS	The Malmö Offspring Study
PRS	Polygenic risk score
QTOF	Quadrupole time-of-flight
SNP	Single nucleotide polymorphism
SSRI	Selective serotonin reuptake inhibitor
TML	Trimethyllysine
UHPLC	Ultra-high performance liquid chromatography

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-024-03516-7.

Additional file 1: Figure S1. Mirror plot showing the cyclo(leucine-proline) fragmentation spectrum acquired in the neonatal cohort and corresponding library match. Mirror plot generated using the Metabolomics Spectrum Resolver. Figure S2. Mirror plot showing the 5-aminovaleric acid betaine fragmentation spectrum acquired in the neonatal cohort and corresponding library match. Mirror plot generated using the Metabolomics Spectrum Resolver. Figure S3. Associations between neonatal metabolites and genetic risk score for autism. Associations were evaluated using linear regression models adjusted for major factors causing variation in the neonatal dried blood spot metabolome, including gestational age, age at sampling, season and year of birth. Nominally significant associations were defined as $P < 0.05$, and significant associations as FDR-adjusted $p < 0.05$. Figure S4. Associations between neonatal metabolite levels and family history of psychiatric disorders. Nominally significant associations were defined as $P < 0.05$, and significant associations as FDR-adjusted $P < 0.05$. Figure S5. Mirror plot showing the m/z 160.13 fragmentation spectrum acquired in the Malmö Offspring Study and corresponding library match to 5-AVAB (cosine = 0.81, 4 matching fragments, mass difference < 0.01). Figure S6. Partial Pearson's correlation coefficients between carnitine-related metabolites in the Malmö Diet and Cancer Study ($N = 3833$). Figure S7. Partial Pearson's correlation coefficients between carnitine-related metabolites in the Malmö Offspring Study ($N = 3430$). Figure S8. Associations between single nucleotide polymorphisms (SNP) located near SLC22A5 and 5-amino valeric acid betaine. Analysis is done in male participants in the Malmö Diet and Cancer Study ($N = 1382$). Figure S9. Associations between single nucleotide polymorphisms (SNP) located near SLC22A5 and 5-amino valeric acid betaine. Analysis is done in female participants in the Malmö Diet and Cancer Study ($N = 2026$). Figure S10. Phenome-wide scan for 5-amino valeric acid betaine (5-AVAB). Analysis performed in the Malmö Diet and Cancer Study ($N = 3077$). (A) Correlation between 5-AVAB and predicted 5-AVAB. 5-AVAB prediction was performed using LASSO regression. R2 indicates the model's explained variance of 5-AVAB. (B) Top 30 variables in prediction of 5-AVAB. LASSO coefficients indicate contribution to the model. Figure S11. Associations between single nucleotide polymorphisms (SNPs) and blood levels of 5-amino valeric acid betaine (5-AVAB) in the Neonatal cohort ($N = 1239$). The Manhattan plot of the genome-wide association study shows genome-wide significance threshold is indicated at $p < 5.0e\text{-}8$. Figure S12. Associations between single nucleotide polymorphisms (SNPs) and blood levels of cyclo-Leucine-Proline (cLP) in the Neonatal cohort ($N = 1239$). The Manhattan plot of the genome-wide association study shows genome-wide significance threshold is indicated at $p < 5.0e\text{-}8$. Figure S13. Relative standard deviation (RSD) for metabolite features ($N = 281$) measured in all external control samples

Additional file 2: Table S1. All metabolites associated with autism ($p < 0.05$) in logistic regression model 1, adjusted for age at sampling, season of birth and gestational age. Proof of annotations from GNPS, SIRIUS, MS2LDA and notes for manual annotations are presented. Table S2. Odds ratios and p-values for logistic regression model 1, model 2 and model 3. All significant ($p < 0.05$) metabolites in model 1 are shown. Table S3. Associations with autism stratified by age of diagnosis. Early indicates autism diagnosis < 6 years old and late > = 6 years old. Results are from logistic regression model 1. Table S4. Autism associations for cyclic dipeptides cyclo(Leu-Pro) and cyclo(Pro-Val) when included as covariates in the same logistic regression model (model 1). Table S5. Metabolite class enrichment analysis. *P*-values are calculated using Fisher's exact test. Table S6. Baseline characteristics description of participants from the Malmö Diet and Cancer Study (MDC) and the Malmö Offspring Study (MOS). Table S7. Correlations between 5-AVAB and dietary intake in the Malmö Diet and Cancer Study (MDC) and the Malmö Offspring Study (MOS) stratifed by sex. Table S8. Top 500 SNPs from genomewide association study of 5-AVAB in the Malmö Diet and Cancer Study. Table S9. Relative standard deviations (%) in repeated injections of quality control (QC) samples for 16 metabolites. QC samples consist of pooled adult blood spotted on filter paper. Table S10. Top 500 SNPs from genomewide association study of 5-AVAB in the neonatal cohort. Table S11. Top 500 SNPs from genomewide association study of cLP in the neonatal cohort.

Acknowledgements

This project was funded by the Lundbeck Foundation through The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), grant number R248-2017-2003—Period III: 1 March 2018 to 28 February 2021; R155-2014-1724: Period II: 1 March 2015 to 28 February 2018; R102-A9118: Period I: 1 March 2012 to 28 February 2015. This research has been conducted using the Danish National Biobank resource supported by the Novo Nordisk Foundation. Some of the computing for this project was performed on the GenomeDK cluster. We would like to thank GenomeDK and Aarhus University for providing computational resources and support that contributed to these research results.

Code availability

Source code and jupyter notebooks used for analysis are publicly available at https://github.com/ssi-dk/CD-MRG-metabolomics_autism.

Authors' contributions

F.O. performed the analysis and visualization. M.E. led the investigation. M.E., F.R., D.H., P.B.M., A.S.C., J.C., and K.S. contributed to neonatal study design. F.O., M.E., and F.R. drafted the manuscript and analyzed data. F.O., M.E., A.A., and N.M. contributed to the mass spectrometry methodology of the neonatal DBS. N.M and A.A acquired the neonatal DBS metabolomics data. O.M., M.O.M., and U.E. contributed to the design of the adult cohort study. F.O. contributed to the mass spectrometry methodology, acquisition, and analyses of adult plasma samples. F.O. and U.E. contributed to dietary assessment analyses. J.G. provided input to genetic analyses. All authors provided interpretation of the results and critical feedback on the manuscript.

Availability of data and materials

The data underlying this study are not publicly available due to the Danish Data Protection Act and European Regulation 2016/679 of the European Parliament and of the Council (GDPR) that prohibit the distribution of personal data. The data are available from the corresponding authors upon reasonable request and under a data transfer and collaboration agreement.

Declarations

Ethics approval and consent to participate

The neonatal metabolomics study was conducted according to the principles of the Declaration of Helsinki and approved by the Danish Ethics Committee (1–10-72–287-12) and the ethics committee of Lund University approved the study protocols for MOS (DNR 2012/594) and MDC (DNR2009/633).

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Section for Clinical Mass Spectrometry, Danish Center for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Copenhagen, Denmark. [3]Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, 15771 Panepistimiopolis, ZografouAthens, Greece. [4]Department of Clinical Therapeutics, School of Medicine, National and Kapodistrian University of Athens, Alexandra Hospital, 11528 Athens, Greece. [5]Department of Clinical Sciences, Lund University, Malmö, Sweden. [6]Testcenter Denmark, Statens

Serum Institut, Copenhagen, Denmark. [7]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [8]Bioinformatics Research Center, Aarhus University, Aarhus, Denmark. [9]Center for Genomics and Personalized Medicine, Aarhus, Denmark. [10]NCRR - National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [11]CIRRAU - Centre for Integrated Registerbased Research at Aarhus University, Aarhus, Denmark.

Received: 23 January 2024 Accepted: 2 July 2024
Published online: 19 July 2024

References

1. Schendel DE, Thorsteinsson E. Cumulative Incidence of Autism Into Adulthood for Birth Cohorts in Denmark, 1980–2012. JAMA. 2018;320:1811–3.
2. Russell G, et al. Time trends in autism diagnosis over 20 years: a UK population-based cohort study. J Child Psychol Psychiatry. 2022;63:674–82.
3. Van Naarden Braun K, et al. Trends in the prevalence of autism spectrum disorder, cerebral palsy, hearing loss, intellectual disability, and vision impairment, metropolitan atlanta, 1991–2010. PLoS ONE. 2015;10:e0124120.
4. Grove J, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
5. Leavey A, Zwaigenbaum L, Heavner K, Burstyn I. Gestational age at birth and risk of autism spectrum disorders in Alberta. Canada J Pediatr. 2013;162:361–8.
6. Hsiao EY. Gastrointestinal issues in autism spectrum disorder. Harv Rev Psychiatry. 2014;22:104–11.
7. Strati F, et al. New evidences on the altered gut microbiota in autism spectrum disorders. Microbiome. 2017;5:24.
8. West KA, et al. Multi-angle meta-analysis of the gut microbiome in Autism Spectrum Disorder: a step toward understanding patient subgroups. Sci Rep. 2022;12:17034.
9. Kim E, et al. Maternal gut bacteria drive intestinal inflammation in offspring with neurodevelopmental disorders by altering the chromatin landscape of CD4+ T cells. Immunity. 2022;55:145-158.e7.
10. Shin S-Y, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46:543–50.
11. Dekkers KF, et al. An online atlas of human plasma metabolite signatures of gut microbiome composition. Nat Commun. 2022;13:1–12.
12. Smith E, et al. A healthy dietary metabolic signature is associated with a lower risk for type 2 diabetes and coronary artery disease. BMC Med. 2022;20:122.
13. Li J, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur Heart J. 2020;41:2645–56.
14. West PR, et al. Metabolomics as a tool for discovery of biomarkers of autism spectrum disorder in the blood plasma of children. PLoS ONE. 2014;9:e112445.
15. Needham BD, et al. Plasma and Fecal Metabolite Profiles in Autism Spectrum Disorder. Biol Psychiatry. 2021;89:451–62.
16. Vargason T, et al. Classification of autism spectrum disorder from blood metabolites: Robustness to the presence of co-occurring conditions. Res Autism Spectr Disord. 2020;77:101644.
17. Bybjerg-Grauholm J, et al. The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. bioRxiv. (2020) https://doi.org/10.1101/2020.11.30.20237768.
18. Rosvall M, Janzon L, Berglund G, Engström G, Hedblad B. Incident coronary events and case fatality in relation to common carotid intima-media thickness. J Intern Med. 2005;257:430–7.
19. Brunkwall L, et al. The Malmö Offspring Study (MOS): design, methods and first results. Eur J Epidemiol. 2021;36:103–16.
20. Schmid R, et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. Nat Commun. 2021;12:3832.
21. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11:395.
22. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8.
23. Deng K, et al. WaveICA: A novel algorithm to remove batch effects for large-scale untargeted metabolomics data based on wavelet analysis. Anal Chim Acta. 2019;1061:60–9.
24. Sumner LW, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3:211–21.
25. Nothias L-F, et al. Feature-based molecular networking in the GNPS analysis environment. Nat Methods. 2020;17:905–8.
26. Wang M, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34:828–37.
27. van der Hooft JJJ, Wandy J, Barrett MP, Burgess KEV, Rogers S. Topic modeling for untargeted substructure exploration in metabolomics. Proc Natl Acad Sci U S A. 2016;113:13738–43.
28. Wandy J, et al. Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. Bioinformatics. 2018;34:317–8.
29. Rogers S, et al. iphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. Faraday Discuss. 2019;218:284–302.
30. da Silva RR, et al. Propagating annotations of molecular networks using in silico fragmentation. PLoS Comput Biol. 2018;14: e1006089.
31. Dührkop K, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods. 2019;16:299–302.
32. Ernst M, et al. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. Metabolites. 2019;9:144.
33. Dührkop K, et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat Biotechnol. 2021;39:462–71.
34. Djoumbou Feunang Y, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform. 2016;8:61.
35. Bittremieux W, et al. Universal MS/MS Visualization and Retrieval with the Metabolomics Spectrum Resolver Web Service. bioRxiv. 2020.05.09.086066 (2020) https://doi.org/10.1101/2020.05.09.086066.
36. Ottosson F, et al. The gut microbiota-related metabolite phenylacetylglutamine associates with increased risk of incident coronary artery disease. J Hypertens. 2020;38:2427–34.
37. Dunn WB, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6:1060–83.
38. Wang M, et al. Mass spectrometry searches using MASST. Nat Biotechnol. 2020;38:23–6.
39. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
40. Lloyd-Jones LR, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun. 2019;10:5086.
41. Zhang Q, Privé F, Vilhjálmsson B, Speed D. Improved genetic prediction of complex traits from individual-level data or summary statistics. Nat Commun. 2021;12:4192.
42. Callmer E, Riboli E, Saracci R, Akesson B, Lindgärde F. Dietary assessment methods evaluated in the Malmö food study. J Intern Med. 1993;233:53–7.
43. Nybacka S, Bertéus Forslund H, Hedelin M. Validity of a web-based dietary questionnaire designed especially to measure the intake of phyto-oestrogens. J Nutr Sci. 2016;5: e36.
44. Kim S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Commun Stat Appl Methods. 2015;22:665–74.
45. Courraud J, Ernst M, Svane Laursen S, Hougaard DM, Cohen AS. Studying Autism Using Untargeted Metabolomics in Newborn Screening Samples. J Mol Neurosci. 2021;71:1378–93.
46. Ernst M, et al. Gestational age-dependent development of the neonatal metabolome. Pediatr Res. 2021;89:1396–404.
47. Ottosson F, et al. Effects of Long-Term Storage on the Biobanked Neonatal Dried Blood Spot Metabolome. J Am Soc Mass Spectrom. 2023. https://doi.org/10.1021/jasms.2c00358.
48. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci U S A. 2015;112:12580–5.
49. Liu KH, et al. Microbial metabolite delta-valerobetaine is a diet-dependent obesogen. Nat Metab. 2021;3:1694–705.

50. Zhao M, et al. TMAVA, a Metabolite of Intestinal Microbes, Is Increased in Plasma From Patients With Liver Steatosis, Inhibits γ-Butyrobetaine Hydroxylase, and Exacerbates Fatty Liver in Mice. Gastroenterology. 2020;158:2266-2281.e27.
51. Olarini A, et al. Vertical Transfer of Metabolites Detectable from Newborn's Dried Blood Spot Samples Using UPLC-MS: A Chemometric Study. Metabolites. 2022;12:94.
52. Semon BA. Dietary cyclic dipeptides, apoptosis and psychiatric disorders: a hypothesis. Med Hypotheses. 2014;82:740–3.
53. Teixidó M, Zurita E, Malakoutikhah M, Tarragó T, Giralt E. Diketopiperazines as a tool for the study of transport across the blood-brain barrier (BBB) and their potential use as BBB-shuttles. J Am Chem Soc. 2007;129:11802–13.
54. Borthwick AD, Da Costa NC. 2,5-diketopiperazines in food and beverages: Taste and bioactivity. Crit Rev Food Sci Nutr. 2017;57:718–42.
55. Guertin KA, et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. Am J Clin Nutr. 2015;101:1000–11.
56. Guertin KA, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr. 2014;100:208–17.
57. Yin X, et al. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. Nat Commun. 2022;13:1644.
58. van der Weide J, Hinrichs JWJ. The influence of cytochrome P450 pharmacogenetics on disposition of common antidepressant and antipsychotic medications. Clin Biochem Rev. 2006;27:17–25.
59. Persson A, et al. reased hippocampal volume and increased anxiety in a transgenic mouse model expressing the human CYP2C19 gene. Mol Psychiatry. 2014;19:733–41.
60. Jukić MM, et al. Elevated CYP2C19 expression is associated with depressive symptoms and hippocampal homeostasis impairment. Mol Psychiatry. 2017;22:1155–63.
61. King BH. Association Between Maternal Use of SSRI Medications and Autism in Their Children. JAMA. 2017;317:1568–9.
62. Morales DR, Slattery J, Evans S, Kurz X. Antidepressant use during pregnancy and risk of autism spectrum disorder and attention deficit hyperactivity disorder: systematic review of observational studies and methodological considerations. BMC Med. 2018;16:6.
63. Kang JD, et al. Bile Acid 7α-Dehydroxylating Gut Bacteria Secrete Antibiotics that Inhibit Clostridium difficile: Role of Secondary Bile Acids. Cell Chem Biol. 2019;26:27-34.e4.
64. Vuong HE, et al. The maternal microbiome modulates fetal neurodevelopment in mice. Nature. 2020;586:281–6.
65. Hazlett HC, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature. 2017;542:348–51.
66. Mossad O, et al. Microbiota-dependent increase in δ-valerobetaine alters neuronal function and is responsible for age-related cognitive decline. Nature Aging. 2021;1:1127–36.
67. Haikonen R, Kärkkäinen O, Koistinen V, Hanhineva K. Diet- and microbiota-related metabolite, 5-aminovaleric acid betaine (5-AVAB), in health and disease. Trends Endocrinol Metab. 2022;33:463–80.
68. Rebouche CJ, Engel AG. Significance of renal gamma-butyrobetaine hydroxylase for carnitine biosynthesis in man. J Biol Chem. 1980;255:8700–5.
69. Hoppel CL, Cox RA, Novak RF. N6-Trimethyl-lysine metabolism. 3-Hydroxy-N6-trimethyl-lysine and carnitine biosynthesis. Biochem J. 1980;188:509–19.
70. Scherer, N. *et al.* Coupling of metabolomics and exome sequencing reveals graded effects of rare damaging heterozygous variants on gene function and resulting traits and diseases. bioRxiv. 2023. https://doi.org/10.1101/2023.10.17.23297094.
71. Pessa-Morikawa T, et al. Maternal microbiota-derived metabolic profile in fetal murine intestine, brain and placenta. BMC Microbiol. 2022;22:46.
72. Vaz FM, Wanders RJA. Carnitine biosynthesis in mammals. Biochem J. 2002;361:417–29.
73. Kärkkäinen O, et al. Diets rich in whole grains increase betainized compounds associated with glucose metabolism. Am J Clin Nutr. 2018;108:971–9.
74. Pallister T, et al. Metabolites of milk intake: a metabolomic approach in UK twins with findings replicated in two European cohorts. Eur J Nutr. 2017;56:2379–91.
75. Pallister T, et al. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. PLoS ONE. 2016;11:e0158568.
76. Al-Khelaifi F, et al. Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. Sci Rep. 2019;9:19889.
77. Feofanova EV, et al. 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. 2020;107:849–63.
78. Kärkkäinen O, et al. Whole grain intake associated molecule 5-aminovaleric acid betaine decreases β-oxidation of fatty acids in mouse cardiomyocytes. Sci Rep. 2018;8:13036.
79. Okura T, Kato S, Deguchi Y. Functional expression of organic cation/carnitine transporter 2 (OCTN2/SLC22A5) in human brain capillary endothelial cell line hCMEC/D3, a human blood-brain barrier model. Drug Metab Pharmacokinet. 2014;29:69–74.
80. Celestino-Soper PBS, et al. A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. Proc Natl Acad Sci U S A. 2012;109:7974–81.
81. Beaudet AL. Brain carnitine deficiency causes nonsyndromic autism with an extreme male bias: A hypothesis. BioEssays. 2017;39:1700012.
82. Barone R, et al. A Subset of Patients With Autism Spectrum Disorders Show a Distinctive Metabolic Profile by Dried Blood Spot Analyses. Front Psychiatry. 2018;9:636.
83. Bitar T, et al. Identification of metabolic pathway disturbances using multimodal metabolomics in autistic disorders in a Middle Eastern population. J Pharm Biomed Anal. 2018;152:57–65.
84. Fahmy SF, El-hamamsy MH, Zaki OK, Badary OA. l-Carnitine supplementation improves the behavioral symptoms in autistic children. Res Autism Spectr Disord. 2013;7:159–66.
85. Khan ZUN, et al. Urinary metabolomics using gas chromatography-mass spectrometry: potential biomarkers for autism spectrum disorder. BMC Neurol. 2022;22:101.
86. Page T, Moseley C. Metabolic treatment of hyperuricosuric autism. Prog Neuropsychopharmacol Biol Psychiatry. 2002;26:397–400.
87. Main PAE, Angley MT, Thomas P, O'Doherty CE, Fenech M. Folate and methionine metabolism in autism: a systematic review. Am J Clin Nutr. 2010;91:1598–620.

Publisher's Note