medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Untargeted metabolomic study of autism in newborn screening samples: a pilot study.

Julie Courraud[1,2], Madeleine Ernst[1,2], Susan Svane Laursen[1], David M. Hougaard[1,2], Arieh S. Cohen[1]

[1] Section for Clinical Mass Spectrometry, Department of Congenital Disorders, Danish Center for Neonatal Screening, Statens Serum Institut Copenhagen, Copenhagen, Denmark

[2] iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark

Correspondence: Julie Courraud (julc@ssi.dk), Section for Clinical Mass Spectrometry, Department of Congenital Disorders, Danish Center for Neonatal Screening, Statens Serum Institut, Artillerivej 5, 2300 Copenhagen S, Denmark

ORCID 0000-0002-6797-7129

Abstract:

Background

The etiopathology of autism spectrum disorder (ASD) is unclear. Main risk factors include both genetic and non-genetic factors, especially prenatal and perinatal events. The Danish Neonatal Screening Biobank in connection with registry data provides unique opportunities to study early signs of disease. Therefore, we aimed to study the metabolomic profiles of dried blood spot (DBS) of newborns later diagnosed with ASD.

Methods

From the iPsych cohort, we randomly selected 37 subjects born in 2005 and diagnosed with ASD in 2012 (cases) together with 37 matched controls and submitted their biobanked DBS to an LC-MS/MS-based untargeted metabolomics protocol. Raw data were preprocessed using MZmine 2.41.2 and

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

metabolites were subsequently putatively annotated using mzCloud, GNPS feature-based molecular networking and other metabolome mining tools (MolNetEnhancer). Statistical analyses and data visualization included principal coordinates analyses, PERMANOVAs, t-tests, and fold-change analyses.

Results

4360 mass spectral features were detected, of which 150 could be putatively annotated at a high confidence level. Chemical structure information at a broad level could be retrieved for a total of 1009 metabolites, covering 31 chemical classes including bile acids, various lipids, nucleotides, amino acids, acylcarnitines and steroids. Although the untargeted analysis revealed no clear distinction between cases and controls, 18 compounds repeatedly reported in the ASD literature could be detected in our study and three mass spectral features were found differentially abundant in cases and controls before FDR correction. In addition, our results pinpointed important other factors influencing chemical profiles of newborn DBS samples such as gestational age, age at sampling and month of birth.

Limitations

Inherent to pilot studies, our sample size was insufficient to reveal metabolic markers of ASD. Nevertheless, we were able to establish an efficient metabolomic data acquisition and analysis pipeline and flag main confounders to be considered in future studies.

Conclusions

In this first untargeted DBS metabolomic study, newborns later diagnosed with ASD did not show a significantly different metabolic profile when compared to controls. Nevertheless, our method covered many metabolites associated with ASD in previous studies, suggesting that biochemical markers of ASD are present at birth and may be monitored during newborn screening.

Keywords: autism, dried blood spots, untargeted metabolomics, newborn screening, biomarkers

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Background

The etiopathology of ASD is still unclear and today ASD is diagnosed based on behavioral signs and assessment of communication skills [1,2]. How the condition should be classified is debated [3,4], as well as which tests offer the most reliable conclusions [1]. In Europe, detection services based on behavioral signs are usually accessed on average at 18 months of age, and diagnosis occurs on average at 36 months of age [5]. In this setting, early intervention is a challenge and has been reported to start in Europe at 42 months of age on average [5]. Whether behavioral impairments are reflected in the blood as biochemical abnormalities is still unsure, but the quest for biomarkers is legitimate, as they would represent a useful tool to help in the diagnosis and treatment of ASD and in understanding its underlying molecular mechanisms [6].

The main risk factors for ASD include genetic [7,8] and non-genetic factors, especially exposure during fetal life [1,9–11]. Prenatal stress could influence fetal brain development and interact with genetic predispositions thereby enhancing the risk of future psychiatric disorders [12,13]. Among prenatal outcomes, maternal infection accompanied by fever during the second trimester of pregnancy has been found to increase the risk of ASD twofold approximately [14]. Among perinatal outcomes, preterm birth (<37 weeks) and low birthweight (small for gestational age) have been associated with an increased risk of ASD as well as high frequency ventilation and intracranial hemorrhage [15]. Low Apgar scores, a score used to summarizing vital signs and assess health in newborns [16], also have recently been associated with an increased risk of developing ASD [17].

Gastrointestinal tract disorders are often reported in ASD children, along with certain foods or diets impacting the severity of symptoms [18–21]. There is a growing evidence of strong interactions

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

between gut and brain through microbiota [22,23], and these observations support the notion that ASD is associated with metabolic malfunction such as decrease in sulphation capacity [18], and potentially connected to gut microbial populations and functions [20]. It has also been shown that many small molecules differing between normally-developing and ASD individuals likely result from microbial metabolism [20,24,25]. Recently, plasma and stool metabolites have been associated with poor communication scoring at age 3, and with good prediction of autism by age 8 [26]. Sharon and collaborators (Sharon et al. 2019) have shown that microbiome and metabolome profiles of mice harboring human microbiota predict that specific bacterial taxa and their metabolites modulate ASD behaviors. They found that taurine and 5-amiovaleric acid (5AV) had significantly lower levels in ASD mice and could show that when feeding BTBR mice either taurine or 5AV, autism-like symptoms such as repetitive behavior and decreased social interaction could be decreased. In an intervention study, treatment with *Lactobacillus reuteri* has been shown to have beneficial effects on ASD-related social disturbances in mice [27]. In humans, intestinal microbiota transplantation has shown very promising results, both against gastrointestinal tract symptoms and ASD symptoms, granting the therapy a 'fast-track' status by the FDA [28]. Among the plasma metabolites showing average to good classification capacity between the treated children and the controls, sarcosine, tyramine O-sulfate and inosine 5'-monophosphate were selected as most discriminant [29]. Many of these studies postulate that microbiota-derived molecules are transported across the blood-brain-barrier, acting as neuroactive metabolites [22]. An impaired intestinal permeability or 'leaky gut' could also play a role in the effect of microbiota activity on psychiatric disorders [30,31]. If gut microbial metabolites of potential impact are indeed detectable in blood, this opens the door to blood-based investigations to further study and understand the metabolomic differences between ASD and non-ASD individuals in the context of gut-brain interactions.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Several studies have reported an altered metabolome associated with ASD during childhood, either in blood [32–35,25,36–44], urine [18,38,45–56] or other matrices [25,57]. However, although some biochemical biomarkers or set of biomarkers seem promising [6], none has yet been proven robust enough for clinical practice. Furthermore, it remains unclear at what point in life biochemical abnormalities of ASD are detectable.

To study the early role of genetic, prenatal and perinatal variables on disease development, samples need to be collected shortly after birth. However, it is not practically and ethically straightforward to draw blood from newborns prospectively. In many countries, the newborn screening programs are conducted on dried blood spots (DBS) collected a few days after birth. In Denmark, such DBS are stored in the Danish National Biobank and are available for research purposes for the last 30 years, thereby covering approximately half of the country's population [58]. This allows researchers to alleviate the biases inherent to recruitment in prospective clinical studies and instead retrospectively retrieve the samples that are connected to the relevant metadata stored in centralized health registries.

Taking advantage of this unique resource, we here aimed at studying the strengths and limitations of DBS samples in studying early biochemical abnormalities related to ASD development using an untargeted metabolomics protocol. We compared the metabolomic profiles of newborns that have been diagnosed with ASD at age 7 (cases) to newborns that have not (controls) and investigated potential main confounders.

Methods

Materials and methods

Materials

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methanol (MeOH), acetonitrile (ACN), isopropanol (IPA), water (H_2O) and formic acid (FA) were of Optima™ LCMS-grade and were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Stable-isotope-labeled internal standards (IS) from the NeoBase Non-derivatized MSMS kit (PerkinElmer, Waltham, MA, USA) were used. The exact list of compounds is provided in Additional file 1.

Subjects and samples

Children (n=74) from the iPsych cohort [59] born in 2005 were randomly selected, of which 37 with a diagnosis of autism spectrum disorder (ICD10 F84.0, F84.1, F84.5, F84.8 and/or F84.9) [2] at the date of registry data extraction (2012), and 37 paired controls. Cases and controls were matched based on gender and date of birth. Other metadata such as gestational age (GA), birthweight, age at sampling, month of birth, mother's age at birth, and date of diagnosis, were also collected (when available). Sample size was chosen for several reasons: 1) the unknown variation of metabolites in DBS made power calculations impossible, hence why we needed a pilot study, 2) batch effect is a common technical challenge in metabolomics, and analyzing all samples on one single 96-well plate was expected to reduce technical variability, 3) DBS are highly precious samples.

DBS are full blood from newborns, aged between 48 and 72 hours, blotted onto Ahlstrom #226 filter paper and left to dry for at least 3 hours at room temperature before being sent by mail at ambient temperature to the Department of Congenital Disorders at the Statens Serum Institut in Copenhagen. Subsequent to being used in the newborn screening program the samples are biobanked in the Danish National Biobank (www.nationalbiobank.dk) at -20ºC until they are retrieved for further research analysis.

Sample extraction

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A punch of 3.2-mm diameter was collected from each DBS using a Panthera-Puncher™ 9 blood spot punching system (PerkinElmer) directly into a MicroPlate, non-coated 96-well clear polystyrene plate (PerkinElmer). 100 µL of IS in extraction buffer were added to each well. The IS were labelled amino acids (AA IS) and acylcarnitines (AC IS) diluted in 80% methanol (i.e. dilution factor of 1:330, concentrations in Additional file 1). The plate was heat-sealed and shaken for 45 min at 750 rpm at 25°C in a PHMP-4 incubator. Then it was centrifuged for 30 min at 4000 rpm at 4 °C.

All the transferring steps were performed on a Microlab STAR line automated liquid handling workstation using Venus software (Hamilton, Bonaduz, Switzerland).

The supernatant (75 µL) was transferred to a hard-shell 96-well polypropylene PCR plate (Bio-Rad) and dried down with nitrogen 60 L/min at 25 °C for 1 hour on an EVX-192 (Apricot Designs Evaporex). The residue was reconstituted in 75 µL 2.5% methanol, shaken for 15 min at 750 rpm at 25°C in a PHMP-4 incubator, and centrifuged 10 min at 4000 rpm at 4 °C. 65µL were transferred to a hard-shell 96-well polypropylene PCR plate (Bio-Rad), heat-sealed, and centrifuged again for 5 min at 3000 rpm at 4 °C. The method from sample preparation to MS acquisition is also available as a table according to the guidelines for standardization of LCMS method reporting [60] with adaptation to metabolomics (Additional file 1).

Quality assurance

LC-MS/MS instrument performance was controlled by analyzing 4 pooled extracts, 8 solvent blanks, and 3 paper blanks at regular intervals. Pooled extracts were made of 5 µL of reconstituted extract from each of the samples (cases and controls only, total of 370 µL divided in four wells spread across the plate) and were used to assess the consistency of extraction and data acquisition. Solvent blanks were used to check for carry over and instrument noise, while paper blanks were used to monitor matrix signals from the paper. Internal standards were used to control the quality of the extraction, elution, and

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

signal acquisition. Paired cases and controls were injected after one another but in a random order (first case, then control, or vice-versa). Pairs were randomized over the plate.

Liquid chromatography

The samples were injected using an autosampler with stack cooler (Open Autosampler UltiMate OAS-3300TXRS (Thermo Fisher Scientific)) and eluted through a Waters Acquity UPLC BEH C18 column (130 Å, 2.1 mm x 50 mm, 1.7-µm particles) preceded by a Waters Acquity UPLC BEH C18 VanGuard pre-column, 130 Å, 2.1 mm x 5 mm, 1.7-µm particles) using a Transcend II, LX-2 with UltiMate pumps (Thermo Fisher Scientific). The pressure limits were set at 0.0 – 1034.0 bar.

The mobile phase consisted of solvent A (97.31% H_2O, 1.25% ACN, 1.25% MeOH and 0.2% FA), and B (2.49% H_2O, 48.66% ACN, 48.66% MeOH and 0.2% FA).

The Wash1 solvent was mobile phase A and the Wash2 solvent mix was 25:25:25:25 v/v MeOH:IPA:H_20:ACN + 0.2% FA. The gradient (0.25 mL/min) started with 100% A : 0% B. After 0.5 min, we applied a gradient ramp to 0% A: 100% B over 8.5 minutes followed by a 0.5-min flow ramp up to 0.9 mL/min and 5 minutes of 100% B. At 15 min, the column was equilibrated for 5.5 min with 100% A. At 17.5 min, the flow was changed back to 0.25 mL/min over 0.5 min. The total run time was 20.5 minutes, including 10 min sample run time and 10.5 min column wash and equilibration. The column temperature was maintained at 60.0°C using a hot pocket column heater and the samples in the autosampler were kept at 4˚C throughout the analysis. The data was acquired in profile mode from 0.20 min and over 9.80 min.

Mass spectrometry

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) was operated with a heated electrospray ionization source (HESI) in positive mode. The instruments were controlled using TraceFinder 4.1 Clinical Research and Aria MX (Thermo Fisher Scientific). Mass range in MS full scan mode was set to 70 to 1050 *m/z* with a resolution of 35,000. Automatic gain control was set to 1.10^6, and maximum injection time at 100 ms. For data dependent-MS2 the resolution was set to 17,500. Automatic gain control was set to 1.10^5, and maximum injection time at 50 ms. Loop count was 5, isolation window 1.5 *m/z* and the stepped NCE 17.5, 35 and 52.5 eV. The spectrum data type was set to Profile. In data dependent settings the Apex trigger was set to 2 to 7 s with 15s dynamic exclusion and charge exclusion on 3-8 and >8. Diisooctylphtalate (391.28429 *m/z*) was selected as lock mass. Other settings included the sheath gas pressure (N_2, 32 psi), the auxiliary gas flow and temperature (N_2, 8 arb. units, 350°C), the S-lens radio frequency level (50.0%), the ion source temperature (350°C), and the spray voltage (3.8 kV between 0-9.8 min and 1.0 kV between 9.8-10 min).

LC-MS data preprocessing

After conversion to .mzML (centroid) using MSConvertGUI v3.0 (ProteoWizard Software Foundation, Palo Alto, CA, USA) [61], raw files were pre-processed using MZmine v2.41.2 [62,63]. All setting details are provided in the batch .xml file (Additional file 2). Briefly, data were cropped based on retention time (RT) 0.27-9.80 min. Masses were detected with a noise threshold of 10,000 for MS1 and of 0 for MS2. The chromatogram was built using the ADAP module [64], with minimum 7 scans per peak, a group intensity threshold of 10,000, a minimum highest intensity of 150,000, and a *m/z* tolerance of 0.001 m/z or 5 ppm. Deconvolution was performed using the Wavelets (ADAP) module, with *m/z* center calculation using median, and ranges for MS2 scan pairing of 0.01 Da and 0.4 min. The isotopes were grouped with a *m/z* tolerance of 0.001 *m/z* or 5 ppm and RT tolerance of 0.1 min. Peaks

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

were aligned with a *m/z* tolerance of 0.001 *m/z* or 5 ppm and RT tolerance of 0.1 min, with 75% weight given to *m/z* and 25% to RT. Finally, peaks were filtered with a minimum of fifteen peaks in a row, and the same RT and peak duration ranges as previously applied. The feature quantification table (.csv) and aggregated MS2 masses list (.mgf) were exported (no merging of MS/MS and filter rows: ALL) for further analysis.

Feature-based molecular networking using GNPS and compound annotation

A molecular network was created with the feature-based molecular networking workflow (https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/) [65] on the GNPS website (http://gnps.ucsd.edu) [66] by uploading the aggregated MS2 mass list. The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor *m/z*. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 4 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. The .graphml network file was then visualized using Cytoscape v3.7.2 [67] where individual sample data and metadata were locally plotted (per sample and metadata sample group relative intensities). To enhance annotation of potential compounds of interest using the mzCloud spectral library (Thermo Fisher Scientific), .raw files were also preprocessed using Compound Discoverer 2.1 (CD2.1) SP1 software

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(Thermo Fisher Scientific). Details regarding the settings are provided in Additional file 3. GNPS and Compound Discoverer (annotation reported when above mzCloud 80% confidence in identity or similarity search) offer annotations with a level 2 confidence according to the Metabolomics Standards Initiative (i.e. putative annotation) [68,69]. To summarize and further enhance chemical structural information within the molecular network, substructure information (https: //ccms-ucsd.github.io/GNPSDocumentation/ms2lda/) [70], information from *in silico* structure annotations from Network Annotation Propagation [71] and Dereplicator [72] were incorporated using the GNPS MolNetEnhancer workflow (https://ccms-ucsd.github.io/GNPSDocumentation/molnetenhancer/) [73] with chemical class annotations retrieved from the ClassyFire chemical ontology [74]. When no chemical structural information could be retrieved through the above searches, the MS/MS spectra were additionally searched via MASST [75] and SIRIUS+CSI:FingerID [76–78]. MASST allows to query a single MS/MS spectrum across all public GNPS datasets giving an idea of the type of samples or matrices where the same MS/MS spectrum has been detected [75]. SIRIUS+CSI:FingerID uses deep learning algorithms to predict the molecular and structural formula of a molecule from MS/MS spectra [76–79].

Contamination filtering and further data curation

Using a Kendrick Mass Filter, we explored the compositionality of our data to assess the potential presence of undesired chemical background [80]. Out of the 4,360 features obtained through MZmine preprocessing, more than 1,100 possessed repeat units typical of polyethylene glycol (PEG) and polypropylene glycol (PPG). Filtering of PEG followed by filtering of PPG was performed using a Kendrick Mass Filter [80] with the following parameters: number of observed signals = 5, Kendrick mass defect = 0.01, and fraction base = 1 (see Additional file 4AB).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Of the 3,253 remaining features, we further excluded those with a maximum intensity in paper blanks / maximum intensity in samples ratio ≥ 0.2, as well as features with 20% or more gaps (i.e. missing value) in cases and/or in controls (1,975 features filtered).

Data visualization and outlier handling

We performed principal coordinates analyses (PCoA) to visualize the data in an interactive 3D-space using EMPeror within QIIME 2 [81,82] using the Bray-Curtis, Canberra, Robust Aitchison [83] and Jaccard distance metrics. This tool allowed us to easily map metadata on each sample in the 3D-space and quickly detect broad patterns in the data, such as assessing the consistency of repeated pool injections (i.e. repeated injections of the same pooled samples should cluster in PCoA space). When performing these calculations on our "raw" unfiltered feature table (4,360 features), all tested distance metrics showed seven samples as clear outliers, of which two controls and five cases. After contamination filtering and data curation (1,281 features), six outliers remained since one outlier (control) was due to PEG contamination. Among the investigated potential explanations for these outliers, no pattern was found when looking at: position on the plate layout, potential RT shift impairing the alignment, and metadata. However, targeted analysis of labeled internal standards and unlabeled endogenous homologs showed that significant (but unexplained) errors occurred during LCMS acquisition, with many undetectable compounds (TraceFinder 4.1 Clinical Research, Thermo Fischer Scientific) (Additional file 5). A heatmap representation of the data (1,281 features) using MetaboAnalyst 4.0 [84] confirmed the six outliers with very low intensities (Additional file 5). Therefore, we decided to exclude these outliers from further statistical analyses and reran the PCoA calculations on the remaining 68 samples.

Statistics

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Using the calculated distance matrices from the PCoAs (1,281 features, 68 samples), we performed Permutational Multivariate Analyses of Variance (PERMANOVAs) [85] to assess how much of the variance in the data is explained by a certain variable in the metadata. We investigated the following variables: autism (yes/no), autism subtype, gender, birthweight, gestational age, age at sampling, month of birth, and injection order. The Adonis R^2 value indicates to what extent the variance is explained by the tested variable. Significance threshold was set at 0.05. Calculations were performed using the *vegan* and *ggplot2* packages in R software v3.6.1 [86].

Finally, the curated dataset (1,281 features, 68 samples, unpaired samples) was processed using MetaboAnalyst 4.0 [84]. We excluded features with more than 50% missing values and replaced the remaining missing values by a small value (half the minimum positive value in the original data). We further filtered non-informative near-constant features based on interquantile range and applied a glog transformation and Pareto scaling. We used a fold-change threshold of 2 (case/control or control/case) and performed t-test and Wilcoxon rank-sum test with FDR correction for multiple comparisons. We could not reliably use the Partial Least Squares Discriminant Analysis (negative Q2 in cross validation). All Jupyter notebooks used for statistical analysis are publicly available on github (https://github.com/madeleineernst/Autism_SupplementaryMaterial).

Results

Subjects

Subjects' characteristics are presented in Table 1 (details in Additional file 6). Cases and controls were similar in terms of GA, birthweight, age at sampling and age of their mother at birth. The most prevalent autism subtype was childhood autism. Most cases had only one diagnosis, but six had both unspecified pervasive development disorder and autism (either childhood autism or atypical autism).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

None had more than two diagnoses. Median age at first diagnosis was 5.6 years (range 1.1-7.8). Most subjects were born at term (GA ≥ 38 weeks). Only three cases and two controls were born preterm.

[Table 1]

Molecular Network analysis

From all features for which a MS2 spectrum has been acquired (2217 features over 4360) a feature-based molecular network was computed via GNPS. Annotation could be retrieved for 150 features (3.4%) of which 103 by matching to GNPS libraries (annotation level 2), and 47 by matching to our in-house library using Trace Finder (annotation level 1, Additional file 7). Using the MolNetEnhancer workflow [73], putative chemical structural information at the chemical class level, corresponding to a level 3 annotation, could be retrieved for an additional 859 features. Hence, nearly 46% (1009) of the mass spectral features could be putatively annotated at a level 1 to 3 (Additional file 7). Annotation covered 31 chemical classes including 53 subclasses and 116 direct parents, such as medium-chain fatty acids, phosphatidylcholines, nucleotides, amino acids, bile acids, steroids, acyl carnitines and catecholamines.

Molecular families (independent clusters of nodes) from the 15 predominant putatively annotated chemical classes are presented in Figure 1 (see details in Additional file 7). Plotting the average intensities in the three groups (cases, controls, paper blanks) as well as fold change values (or p-values) on the network nodes allowed for a quick overview of the molecular families with potential biological relevance (See the example of bile acids in Figure 2). This analysis showed the potential of DBS in covering various chemical classes and the power of feature-based molecular network analyses and related metabolome mining tools in expanding the interpretability of complex untargeted metabolomics data.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Statistical analyses

In PCoA space, repeated pool injections clustered satisfactorily showing that the LC-MS/MS data acquisition was of acceptable quality (Figure 3). When looking at the two groups (cases/controls), no clear separation was observed, even after removal of outliers (Figure 3A) and on all tested distance metrics.

The PERMANOVAs (Figure 4, Additional file 8) revealed that the variance in the data was not significantly explained by the grouping (cases/controls) (Bray-Curtis Adonis R^2 = 0.014, P-value = 0.777), even when distinguishing subtypes of ASD, although subtypes had a higher Adonis R^2 (Bray-Curtis Adonis R^2 = 0.070, P-value = 0.707, see Table 1 for details on subtypes of ASD). Similarly, the gender and birthweight did not significantly explain the variance in the data. However, variation in the data explained by gestational age (Bray-Curtis Adonis R^2 = 0.034, P-value = 0.028), age at sampling (Bray-Curtis Adonis R^2 = 0.053, P-value = 0.002) and especially month of birth (Bray-Curtis Adonis R^2 = 0.262, P-value = 0.001, see Figure 3B) was significant, and to an increasing extent respectively (greater Adonis R^2). On a technical level, we also observed a significant effect of injection order on the data variation (Bray-Curtis Adonis R^2 = 0.047, P-value = 0.002), pointing to a limitation in our LC-MS/MS method.

Results of univariate analyses and fold change analysis were carefully scrutinized feature by feature. Considering our small sample size and potential pitfalls inherent to untargeted metabolomics related to contaminants or integration errors, we thought essential to inspect each result to eliminate false positives and spurious findings. Our inspection consisted of a five-step logic starting with peak integration and shape quality (MZmine). We then plotted all individual intensity values to assess whether the case/control difference was driven by four or fewer samples. If not, we reported the extent

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of missing values in each group, checked the consistency of replicated pool injections, and finally checked whether the feature was present in the feature-based molecular network, annotated as a contaminant or in a node cluster with such annotation (Additional file 9). A large proportion of the inspected features were excluded based on these criteria, showing the importance of such a verification in order not to pursue spurious findings in future studies.

Among the 24 features with a fold change (case/control) value <0.5 or >2.0, only one passed manual inspection (Table 2, the full table is in Additional file 9). Eluting quite late (RT = 6.64 min, ID8605), this relatively hydrophobic compound had a detected *m/z* of 1014.4892 and was not connected to any other node in the network analysis (see its mass spectrum in Additional file 10). It could not be annotated, but the algorithm of SIRIUS+CSI:FingerID pointed at a raw formula of $C_{36}H_{63}N_{21}O_{14}$ ([M+H]+, only 7.12% scoring). This compound was more than twice as intense in controls as in cases (FC 0.42, average intensity in cases 2.73E+05 and controls 7.51E+05), and would need further investigation, especially as it was not detected in many samples (Additional file 9). A MASST search was performed, however the feature with *m/z* 1014.4892 was not found in any of the public datasets on GNPS.

[Table 2]

No feature was significantly differentially abundant in cases and controls according to the univariate analyses with FDR correction for multiple comparisons (p values in Table 2).

Features that were differentially abundant before FDR correction are presented in Table 2. As a high proportion of features were deemed irrelevant after inspection, we are presenting only the two relevant features that passed our quality-control criteria. The full list and inspection details can be found in Additional file 9. Methacholine was found to be significantly more abundant in cases when compared

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

to controls (average intensity in cases 4.41E+07 and controls 3.94E+07) both when using a t-test (p = 0.0021) and a Wilcoxon rank-sum test (p = 0.0031). The corresponding node (ID159) in the network analysis was connected to another node with a mass difference of -0.036 *m/z* (225 ppm) which could not be annotated. None of the applied metabolome mining tools was able to retrieve chemical structural information for the second compound significantly more abundant in cases than in controls (ID5593, *m/z* 1014.4892, average intensity in cases 5.71E+05 and controls 4.35E+05). SIRIUS+CSI:Finger ID predicted a molecular formula of $C_{11}H_{22}N_2O_3$ (M+H+, 99.96% scoring). Its RT of 2.78 min could indicate a medium polarity with a logP between -1.0 and 0.5 when compared to tryptophan (RT 2.56 min, HMDB experimental logP -1.06) and hippuric acid (RT 3.04 min, HMDB experimental logP 0.31).

Among the 273 compounds reported in two recent reviews [6,87], 22 were cited at least three times, of which 18 could be linked to features in our study after manual verification (Table 3, Additional file 11).

[Table 3]

Discussion

To assess the early signs of ASD development shortly after birth, we compared DBS samples from newborns that have later on been diagnosed with ASD to newborns that have not. Our pilot study showed the capacity of untargeted metabolomics as an analytical tool applied to biobanked DBS samples to cover several metabolites relevant to ASD, thus suggesting that biochemical markers of ASD are present at birth and could be targeted during neonatal screening. In addition, our method pinpointed other factors which have a strong influence on the metabolic profile of newborn DBS, such

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

as gestational age, age at sampling and month of birth, and which are important to consider when designing metabolomic studies in neonatal, biobanked DBS.

One study from 2013 was performed on newborn DBS samples from 16 autistic children and assessed 90 biomarkers (not only small molecules) using immunoassays [88] of which three sets of five were associated with ASD. Another study was performed on DBS but in ASD children (n=83, age 2-10 years) and was targeting 45 metabolites [37], of which 9 were significantly higher in ASD children. However, the potential of DBS in *untargeted* metabolomics studies has not yet been fully studied, and never in the context of ASD (see recent reviews [6,87]).

Among the 22 compounds that had been repeatedly (≥3 times) reported in the literature to be involved in ASD, 18 could be putatively annotated in our study, showing that our analytical pipeline covers many relevant metabolites, including some specific to gut microbiota activity. Despite thorough curation and inspection of the acquired data, no feature was significantly differentially abundant in cases and controls after FDR correction. The small sample size as well as strong confounders could play a role in the low power of this pilot study. Nonetheless, the PERMANOVA results showed that subtypes of autism seemed to explain more variance in the data than just the case/control division (higher Adonis R^2, Figure 4). This could mean that subtypes have indeed distinct biochemical profiles and should maybe be considered separately or in a stratified manner. Although, the lack of significance could suggest that ASD genetic, prenatal and/or perinatal variables do not (sufficiently) translate into biochemical abnormalities in newborns or that these are not detectable in biobanked DBS using our current experimental setup, we were able to pinpoint 18 out of 22 metabolic markers associated with ASD in previous studies. This finding suggests that rather the sample size and confounders are responsible for the absent signal observed here, and future studies controlling for the here identified confounders and greater sample size could provide more answers.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Among the hits and interesting findings of our pilot, we could show that a methacholine structural analogue could be a relevant marker for ASD, as it was found at a higher although not significant abundance in newborns that have been diagnosed with ASD at age 7. Methacholine is a choline ester drug acting as non-selective muscarinic receptor agonist. It is mainly known as methacholine chloride for its use in assessing bronchial hyper-reactivity in asthmatic patients. Although muscarinic receptors were not associated with autism in children [89], lower estimates of ASD risk among children exposed during fetal life to muscarinic receptor 2 agonists were reported [90]. Higher abundance of methacholine in DBS of ASD cases, as seen in our study, would therefore not be easily explained and demand further investigation. However, detecting a drug metabolite such as methacholine in newborn samples is unexpected, thus it is more likely that this feature is an endogenous choline ester with similar fragmentation behavior to methacholine.

Two other unknown features would benefit from being monitored in future studies. One relatively hydrophobic compound (ID8605, *m/z* 1014.4892) showed an important fold change (much lower in cases) but was not detected in many samples maybe due to low intensities. The second compound (moderately polar, ID5593, *m/z* 1014.4892, $C_{11}H_{22}N_2O_3$) was significantly higher in cases before FDR correction and detected in more than 65% of samples. Including both in an inclusion list for optimized MS2 acquisition could provide higher quality spectra for better annotation. Overall, we observed poor fragmentation of some analytes (4 fragmentation peaks or less) and optimization of fragmentation parameters to a few targeted compounds, could enable better chemical structural annotation, although fragmentation parameters such as other technical and sample preparation parameters, can never be optimized for all compounds equally.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Although our study was not designed specifically to assess the tested confounders, we have shown that gestational age, age at sampling and month of birth are strong drivers of metabolomic profiles in newborns. This demonstrates the importance of considering these confounders when designing a future study in newborn DBS.

Prematurity has been involved in numerous adverse health outcomes [91] and metabolic maturity has been shown to be reflected in the blood and other matrices of infants after birth [92,93]. Although, in the present study, only three cases and two controls were premature (<38 weeks of gestational age), we saw a significant effect of gestational age on the metabolomic profile of newborns thus showing that gestational age is an important factor be controlled for in newborn DBS studies.

Similarly, we found that age at sampling has a significant impact on the newborn blood metabolome. From 3 to 10 days of age, only one week has passed, and yet fundamental metabolic changes occur in the newborn who is not anymore connected to his/her mother's blood, but exposed to various types of nutrition and environmental conditions (healthcare, hospital vs home, etc.). The endogenous anabolism/catabolism balance is in itself a strong variable to consider at that age. From 2009 onwards, the Danish newborn screening program has indeed chosen to standardize the age at DBS sampling to 48 to 72 hours to optimize the window where potential inborn errors of metabolism would be detected best and as early as possible since quick intervention is essential in such cases [94]. The iPsych cohort was based on diagnoses of psychiatric disorders recorded in Danish health registry in 2012 [59]. Such diagnoses are often given after several years of age, which is why the cohort did not cover subjects born after 2005, year at which the age at sampling was not so narrowly standardized. The large sample size of the iPsych cohort (almost 15,000 cases of ASD) should however allow for stricter selection criteria in our future studies.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Another major change that occurs in newborns at birth and in the following days is the gut maturation and its further colonization by microbes [95]. This topic has been under expanding attention in the last decade, and the development and involvement of gut microbiota in neurodevelopment is being scrutinized extensively [22]. The exact dynamics of the microbiota development in the placenta and during the first days of life is still uncertain [95–97], as well as to what extent its activity can be reflected in the blood. A recent study has shown that gut microbial alpha-diversity can be predicted from the human blood metabolome [98] suggesting that microbial metabolites explain a significant amount of the variance in the human blood metabolome. Thus, although sampled at an early stage in life, it is plausible that microbial metabolites mediating health may be found in dried blood spots from newborns [93]. Studying both fecal and blood samples will be essential to answer questions related to the impact of gut microbes on the gut-brain axis, especially in the context of psychiatric disorders where the brain is the main organ concerned but indeed located quite far from the gut. Microbial metabolites would necessarily need to travel in the blood (or lymph) and through the blood-brain barrier to interact with the brain. In our study, some detected metabolites could partly derive from gut microbiota activity such as DL-Indole-3-lactic acid (ID3461, [99,100]), taurine (ID428, level 3, [25]), various bile acids (Additional file 7, [57]), or inosine 5’-monophosphate (ID1133, level 3, [29]).

Lastly, we found that month of birth explains a significant variation in metabolomic profiles of newborns (Figure 3b, Figure 4). Whether there is a yearly cyclic pattern or whether our findings are specific to 2005 remains to be determined. Explanations could include aspects related to pregnancy conditions varying along the year such as diet, weather conditions and sun exposure (e.g. impact on vitamin D levels, type and extent of physical and social activities, mood and stress [101]), exposure to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

"seasonal" infectious diseases (e.g. influenza), exposure to varying air quality (e.g. pollution or pollens [102]), as well as storing conditions which might fluctuate over the year (e.g. sample transport at higher temperatures during summer).

Gender and birthweight were not found to explain a significant part of the variance in the metabolomic profiles of newborns in our study, despite the obvious connection between gestational age and birthweight. The gender misbalance which reflects the gender disparity in ASD (a quarter were girls), non-dedicated design, and small sample size could explain this finding. Some studies have indeed reported that the profile of newborn girls and boys differed in, for instance, blood amino acids and acylcarnitines [103], as well as urine profiles [104]. Despite our finding, we believe that gender and birthweight should be adjusted for and taken into consideration when designing metabolomics studies in newborns. Several of the tested confounders are inter-connected with, for instance, reports of more males being born preterm [105] and females being born lighter [106], both associations being explained by mechanisms that are likely to be reflected in the metabolome such as inflammatory response and insulin resistance, respectively.

On the technical side, the strong effect of injection order is again a reminder that randomizing sample pairs on the plates is fundamental in untargeted metabolomics. Although technical improvements can be considered, it is extremely difficult if not impossible to avoid some level of signal drift and carryover in untargeted metabolomics studies[107].

Limitations and strengths

To avoid batch effect and minimize the use of highly valuable and rare samples, we analyzed only 37 pairs of cases and controls in this pilot study. Despite the small sample size that did not confer enough

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

statistical power for pinpointing strong marker metabolites of ASD, we could detect numerous metabolites associated with ASD in previous studies and identify a number of confounders to be considered in future untargeted metabolomics study using newborn DBS. Other confounders not evaluated in our study will need to be assessed in future studies, including metabolic changes in DBS associated with time and storage conditions. Hematocrit variation could not be measured in our study as we had access to only one punch of paper and did not have the possibility to measure a surrogate marker such as potassium [108]. Furthermore, metabolites detected in this study are inherently reflective of sampling protocols, including extraction protocols and MS acquisition parameters and should be interpreted within these limitations.

Conclusions

This is the first study assessing metabolomic profiles of a psychiatric disorder, ASD in biobanked, newborn DBS samples. The development of biobanks and reuse of systematically collected DBS samples for research purposes in connection with registry data represent many new opportunities to study the physiopathology and early signs of diseases, with extraordinary impacts in prevention, diagnosis and treatment strategies. We showed that untargeted metabolomics on DBS samples offers a wide and relevant coverage of metabolites for the study of ASD and that the new processing tools used in our method largely expand the interpretability of such complex data.

Declarations

- Ethics approval and consent to participate

The Danish ethical committee approved the project (1-10-72-287-12). All blood samples can be stored without explicit informed consent according to Danish law, and be reused for additional analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

provided that projects are approved by the Research Ethics Committee and the Danish Data Protection Agency. Although research projects require informed consent, the Research Ethics Committee can waive this requirement if samples are anonymously processed, and if the projects do not imply any health-related risk or burden to the subjects [109].

- Consent for publication

Not applicable

- Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the risk of compromising individual privacy but are available from the corresponding author on reasonable request and provided that an appropriate collaboration agreement can be agreed upon.

- Competing interests

The authors declare that they have no competing interests.

- Funding

This study was funded by The Lundbeck Foundation, Denmark.

This research has been conducted using the Danish National Biobank resource supported by the Novo Nordisk Foundation grant number is 2010-11-12 and 2009-07-28.

Disclaimer

This funder had no involvement in any aspect of the study.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Grant number

R248-2017-2003 - Period III: 1 March 2018 - 28 February 2021

R155-2014-1724: Period II: 1 March 2015 – 28 February 2018

R102-A9118: Period I: 1 March 2012 – 28 February 2015

- Authors' contributions

JC and AC designed the study.

SSL collected the samples and ran the metabolomics analyses.

ME defined and coded major steps of data processing, chemical structural annotation and statistical analysis and edited the manuscript.

JC processed the data, performed statistical analyses, and drafted the manuscript.

JC, ME, SSL and AC were involved in the manuscript edition and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version of this manuscript.

JC and AC had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

- Acknowledgements

We thank Anders Björkbom for his major contribution to the LCMS method development and Marie Bækvad-Hansen for her help in curating data.

- Authors' information (optional)

Not applicable

References

1. Randall M, Egberts KJ, Samtani A, Scholten RJ, Hooft L, Livingstone N, et al. Diagnostic tests for autism spectrum disorder (ASD) in preschool children. Cochrane Database Syst Rev. 2018;

2. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva: World Health Organization; 1993.

3. Adam D. Mental health: On the spectrum. Nat News. 2013;496:416.

4. Frances A. ICD, DSM and The Tower of Babel. Aust N Z J Psychiatry. 2014;48:371–3.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. Bejarano-Martín Á, Canal-Bedia R, Magán-Maganto M, Fernández-Álvarez C, Cilleros-Martín MV, Sánchez-Gómez MC, et al. Early Detection, Diagnosis and Intervention Services for Young Children with Autism Spectrum Disorder in the European Union (ASDEU): Family and Professional Perspectives. J Autism Dev Disord. 2019;

6. Shen L, Liu X, Zhang H, Lin J, Feng C, Iqbal J. Biomarkers in autism spectrum disorders: Current progress. Clin Chim Acta Int J Clin Chem. 2019;502:41–54.

7. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. JAMA Psychiatry. 2019;76:1035–43.

8. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.

9. Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, et al. The epidemiology of autism spectrum disorders. Annu Rev Public Health. 2007;28:235–58.

10. Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Andrews SV, et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. Genome Med. 2018;10.

11. Braam W, Ehrhart F, Maas APHM, Smits MG, Curfs L. Low maternal melatonin level increases autism spectrum disorder risk in children. Res Dev Disabil. 2018;82:79–89.

12. Abbott PW, Gumusoglu SB, Bittle J, Beversdorf DQ, Stevens HE. Prenatal stress and genetic risk: How prenatal stress interacts with genetics to alter risk for psychiatric illness. Psychoneuroendocrinology. 2018;90:9–21.

13. Fine R, Zhang J, Stevens HE. Prenatal stress and inhibitory neuron systems: implications for neuropsychiatric disorders. Mol Psychiatry. 2014;19:641–51.

14. Croen LA, Qian Y, Ashwood P, Zerbo O, Schendel D, Pinto-Martin J, et al. Infection and Fever in Pregnancy and Autism Spectrum Disorders: Findings from the Study to Explore Early Development. Autism Res. 2019;12:1551–61.

15. Kuzniewicz MW, Wi S, Qian Y, Walsh EM, Armstrong MA, Croen LA. Prevalence and Neonatal Factors Associated with Autism Spectrum Disorders in Preterm Infants. J Pediatr. 2014;164:20–5.

16. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg. 1953;32:260–7.

17. Modabbernia A, Sandin S, Gross R, Leonard H, Gissler M, Parner ET, et al. Apgar score and risk of autism. Eur J Epidemiol. 2019;34:105–14.

18. Alberti A, Pirrone P, Elia M, Waring RH, Romano C. Sulphation deficit in “low-functioning” autistic children: a pilot study. Biol Psychiatry. 1999;46:420–4.

19. Chaidez V, Hansen RL, Hertz-Picciotto I. Gastrointestinal problems in children with autism, developmental delays or typical development. J Autism Dev Disord. 2014;44:1117–27.

20. Krajmalnik-Brown R, Lozupone C, Kang D-W, Adams JB. Gut bacteria in children with autism spectrum disorders: challenges and promise of studying how a complex community influences a complex disease. Microb Ecol Health Dis. 2015;26.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

21. White JF. Intestinal pathophysiology in autism. Exp Biol Med Maywood NJ. 2003;228:639–49.

22. Cerdó T, Diéguez E, Campoy C. Early nutrition and gut microbiome: interrelationship between bacterial metabolism, immune system, brain structure, and neurodevelopment. Am J Physiol-Endocrinol Metab. 2019;317:E617–30.

23. Wang S, Harvey L, Martin R, van der Beek EM, Knol J, Cryan JF, et al. Targeting the gut microbiota to influence brain development and function in early life. Neurosci Biobehav Rev. 2018;95:191–201.

24. De Angelis M, Francavilla R, Piccolo M, De Giacomo A, Gobbetti M. Autism spectrum disorders and intestinal microbiota. Gut Microbes. 2015;6:207–13.

25. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, et al. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. Cell. 2019;177:1600-1618.e17.

26. Kelly RS, Boulin A, Laranjo N, Lee-Sarwar K, Chu SH, Yadama AP, et al. Metabolomics and Communication Skills Development in Children; Evidence from the Ages and Stages Questionnaire. Metabolites. 2019;9:42.

27. Sgritta M, Dooling SW, Buffington SA, Momin EN, Francis MB, Britton RA, et al. Mechanisms Underlying Microbial-Mediated Changes in Social Behavior in Mouse Models of Autism Spectrum Disorder. Neuron. 2019;101:246-259.e6.

28. Adams JB, Borody TJ, Kang D-W, Khoruts A, Krajmalnik-Brown R, Sadowsky MJ. Microbiota transplant therapy and autism: lessons for the clinic. Expert Rev Gastroenterol Hepatol. 2019;13:1033–7.

29. Adams JB, Vargason T, Kang D-W, Krajmalnik-Brown R, Hahn J. Multivariate Analysis of Plasma Metabolites in Children with Autism Spectrum Disorder and Gastrointestinal Symptoms Before and After Microbiota Transfer Therapy. Processes. 2019;7:806.

30. Borre YE, O'Keeffe GW, Clarke G, Stanton C, Dinan TG, Cryan JF. Microbiota and neurodevelopmental windows: implications for brain disorders. Trends Mol Med. 2014;20:509–18.

31. Magistris L de, Familiari V, Pascotto A, Sapone A, Frolli A, Iardino P, et al. Alterations of the Intestinal Barrier in Patients With Autism Spectrum Disorders and in Their First-degree Relatives. J Pediatr Gastroenterol Nutr. 2010;51:418–24.

32. Aldred S, Moore KM, Fitzgerald M, Waring RH. Plasma Amino Acid Levels in Children with Autism and Their Families. J Autism Dev Disord. 2003;33:93–7.

33. Evans C, Dunstan RH, Rothkirch T, Roberts TK, Reichelt KL, Cosford R, et al. Altered amino acid excretion in children with autism. Nutr Neurosci. 2008;11:9–17.

34. De Angelis M, Piccolo M, Vannini L, Siragusa S, De Giacomo A, Serrazzanetti DI, et al. Fecal Microbiota and Metabolome of Children with Autism and Pervasive Developmental Disorder Not Otherwise Specified. Plos One. 2013;8:e76993.

35. Kang D-W, Ilhan ZE, Isern NG, Hoyt DW, Howsmon DP, Shaffer M, et al. Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. Anaerobe. 2018;49:121–31.

36. Naushad SM, Jain JMN, Prasad CK, Naik U, Akella RRD. Autistic children exhibit distinct plasma amino acid profile. Indian J Biochem Biophys. 2013;50:474–8.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

37. Barone R, Alaimo S, Messina M, Pulvirenti A, Bastin J, Group M-A, et al. A Subset of Patients With Autism Spectrum Disorders Show a Distinctive Metabolic Profile by Dried Blood Spot Analyses. Front Psychiatry. 2018;9.

38. Anwar A, Abruzzo PM, Pasha S, Rajpoot K, Bolotta A, Ghezzo A, et al. Advanced glycation endproducts, dityrosine and arginine transporter dysfunction in autism - a source of biomarkers for clinical diagnosis. Mol Autism. 2018;9:3.

39. Smith AM, King JJ, West PR, Ludwig MA, Donley ELR, Burrier RE, et al. Amino Acid Dysregulation Metabotypes: Potential Biomarkers for Diagnosis and Individualized Treatment for Subtypes of Autism Spectrum Disorder. Biol Psychiatry. 2019;85:345–54.

40. West PR, Amaral DG, Bais P, Smith AM, Egnash LA, Ross ME, et al. Metabolomics as a Tool for Discovery of Biomarkers of Autism Spectrum Disorder in the Blood Plasma of Children. PLoS ONE. 2014;9.

41. Rangel-Huerta OD, Gomez-Fernández A, de la Torre-Aguilar MJ, Gil A, Perez-Navero JL, Flores-Rojas K, et al. Metabolic profiling in children with autism spectrum disorder with and without mental regression: preliminary results from a cross-sectional case-control study. Metabolomics Off J Metabolomic Soc. 2019;15:99.

42. Bitar T, Mavel S, Emond P, Nadal-Desbarats L, Lefèvre A, Mattar H, 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.

43. Delaye J-B, Patin F, Lagrue E, Le Tilly O, Bruno C, Vuillaume M-L, et al. Post hoc analysis of plasma amino acid profiles: towards a specific pattern in autism spectrum disorder and intellectual disability. Ann Clin Biochem. SAGE Publications; 2018;55:543–52.

44. Kuwabara H, Yamasue H, Koike S, Inoue H, Kawakubo Y, Kuroda M, et al. Altered Metabolites in the Plasma of Autism Spectrum Disorder: A Capillary Electrophoresis Time-of-Flight Mass Spectroscopy Study. PLoS ONE. 2013;8.

45. Retey J. The Urocanase Story: A Novel Role of NAD+ as Electrophile. Arch Biochem Biophys. 1994;314:1–16.

46. Shaw W. Increased urinary excretion of a 3-(3-hydroxyphenyl)-3-hydroxypropionic acid (HPHPA), an abnormal phenylalanine metabolite of Clostridia spp. in the gastrointestinal tract, in urine samples from patients with autism and schizophrenia. Nutr Neurosci. 2010;13:135–43.

47. Mavel S, Nadal-Desbarats L, Blasco H, Bonnet-Brilhault F, Barthélémy C, Montigny F, et al. 1H-13C NMR-based urine metabolic profiling in autism spectrum disorders. Talanta. 2013;114:95–102.

48. Yap IKS, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. J Proteome Res. 2010;9:2996–3004.

49. Ming X, Stein TP, Barnes V, Rhodes N, Guo L. Metabolic perturbance in autism spectrum disorders: a metabolomics study. J Proteome Res. 2012;11:5856–62.

50. Liu A, Zhou W, Qu L, He F, Wang H, Wang Y, et al. Altered Urinary Amino Acids in Children With Autism Spectrum Disorders. Front Cell Neurosci. 2019;13.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

51. Kałużna-Czaplińska J. Noninvasive urinary organic acids test to assess biochemical and nutritional individuality in autistic children. Clin Biochem. 2011;44:686–91.

52. Lussu M, Noto A, Masili A, Rinaldi AC, Dessì A, Angelis MD, et al. The urinary 1H-NMR metabolomics profile of an italian autistic children population and their unaffected siblings. Autism Res. 2017;10:1058–66.

53. Chen Q, Qiao Y, Xu X, You X, Tao Y. Urine Organic Acids as Potential Biomarkers for Autism-Spectrum Disorder in Chinese Children. Front Cell Neurosci. 2019;13.

54. Emond P, Mavel S, Aïdoud N, Nadal-Desbarats L, Montigny F, Bonnet-Brilhault F, et al. GC-MS-based urine metabolic profiling of autism spectrum disorders. Anal Bioanal Chem. 2013;405:5291–300.

55. Noto A, Fanos V, Barberini L, Grapov D, Fattuoni C, Zaffanello M, et al. The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. J Matern Fetal Neonatal Med. 2014;27:46–52.

56. Gevi F, Zolla L, Gabriele S, Persico AM. Urinary metabolomics of young Italian autistic children supports abnormal tryptophan and purine metabolism. Mol Autism. 2016;7:47.

57. Wang M, Wan J, Rong H, He F, Wang H, Zhou J, et al. Alterations in Gut Glutamate Metabolism Associated with Changes in Gut Microbiota Composition in Children with Autism Spectrum Disorder. mSystems. American Society for Microbiology Journals; 2019;4.

58. Nørgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening Biobank. J Inherit Metab Dis. 2007;30:530–6.

59. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-Hansen M, et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2018;23:6–14.

60. Vogeser M, Schuster C, Rockwood AL. A proposal to standardize the description of LC–MS-based measurement methods in laboratory medicine. Clin Mass Spectrom. 2019;13:36–8.

61. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30:918–20.

62. Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinforma Oxf Engl. 2006;22:634–6.

63. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010;11:395.

64. Myers OD, Sumner SJ, Li S, Barnes S, Du X. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. Anal Chem. 2017;89:8696–703.

65. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based Molecular Networking in the GNPS Analysis Environment. bioRxiv. 2019;812404.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

66. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34:828–37.

67. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

68. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. J Am Soc Mass Spectrom. 2016;27:1897–905.

69. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3:211–21.

70. Hooft JJJ van der, Wandy J, Barrett MP, Burgess KEV, Rogers S. Topic modeling for untargeted substructure exploration in metabolomics. Proc Natl Acad Sci. 2016;113:13738–43.

71. Silva RR da, Wang M, Nothias L-F, Hooft JJJ van der, Caraballo-Rodríguez AM, Fox E, et al. Propagating annotations of molecular networks using in silico fragmentation. PLOS Comput Biol. 2018;14:e1006089.

72. Mohimani H, Gurevich A, Mikheenko A, Garg N, Nothias L-F, Ninomiya A, et al. Dereplication of peptidic natural products through database search of mass spectra. Nat Chem Biol. 2017;13:30–7.

73. Ernst M, Kang KB, Caraballo-Rodríguez AM, Nothias L-F, Wandy J, Chen C, et al. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. Metabolites. 2019;9:144.

74. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminformatics. 2016;8:61.

75. Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, et al. Mass spectrometry searches using MASST. Nat Biotechnol. 2020;38:23–6.

76. Böcker S, Dührkop K. Fragmentation trees reloaded. J Cheminformatics. 2016;8:5.

77. 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.

78. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods. 2019;16:299–302.

79. Shen H, Dührkop K, Böcker S, Rousu J. Metabolite identification through multiple kernel learning on fragmentation trees. Bioinforma Oxf Engl. 2014;30:i157-164.

80. da Silva RR, Vargas F, Ernst M, Nguyen NH, Bolleddu S, del Rosario KK, et al. Computational Removal of Undesired Mass Spectral Features Possessing Repeat Units via a Kendrick Mass Filter. J Am Soc Mass Spectrom. 2019;30:268–77.

81. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852–7.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

82. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. GigaScience. 2013;2.

83. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. mSystems. 2019;4:e00016-19.

84. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46:W486–94.

85. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46.

86. R Core Team. R: a language and environment for statistical computing [Internet]. R Found. Stat. Comput. Vienna Austria. 2019 [cited 2019 Nov 21]. Available from: https://www.R-project.org/

87. Glinton KE, Elsea SH. Untargeted Metabolomics for Autism Spectrum Disorders: Current Status and Future Directions. Front Psychiatry. 2019;10.

88. Mizejewski GJ, Lindau-Shepard B, Pass KA. Newborn screening for autism: in search of candidate biomarkers. Biomark Med. 2013;7:247–60.

89. Lee M, Martin-Ruiz C, Graham A, Court J, Jaros E, Perry R, et al. Nicotinic receptor abnormalities in the cerebellar cortex in autism. Brain J Neurol. 2002;125:1483–95.

90. Janecka M, Kodesh A, Levine SZ, Lusskin SI, Viktorin A, Rahman R, et al. Association of Autism Spectrum Disorder With Prenatal Exposure to Medication Affecting Neurotransmitter Systems. JAMA Psychiatry. 2018;75:1217–24.

91. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. The Lancet. 2008;371:261–9.

92. Gil AM, Duarte D. Biofluid Metabolomics in Preterm Birth Research. Reprod Sci. 2018;25:967–77.

93. Ernst M, Rogers S, Lausten-Thomsen U, Bjorkbom A, Laursen SS, Courraud J, et al. Gestational-age-dependent development of the neonatal metabolome. medRxiv. Cold Spring Harbor Laboratory Press; 2020;2020.03.27.20045534.

94. Dionisi-Vici C, Deodato F, Roschinger W, Rhead W, Wilcken B. “Classical” organic acidurias, propionic aciduria, methylmalonic aciduria and isovaleric aciduria: Long-term outcome and effects of expanded newborn screening using tandem mass spectrometry. J Inherit Metab Dis. Springer; 2006;29:383–9.

95. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol Mol Biol Rev MMBR. 2017;81.

96. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015;17:690–703.

97. Rosa PSL, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, et al. Patterned progression of bacterial populations in the premature infant gut. Proc Natl Acad Sci. 2014;111:12522–7.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

98. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, et al. Blood metabolome predicts gut microbiome α-diversity in humans. Nat Biotechnol. 2019;37:1217–28.

99. Meng D, Sommella E, Salviati E, Campiglia P, Ganguli K, Djebali K, et al. Indole-3-lactic acid, a metabolite of tryptophan, secreted by Bifidobacterium longum subspecies infantis is anti-inflammatory in the immature intestine. Pediatr Res. Nature Publishing Group; 2020;1–9.

100. Laursen MF, Sakanaka M, Burg N von, Andersen D, Mörbe U, Rivollier A, et al. Breastmilk-promoted bifidobacteria produce aromatic lactic acids in the infant gut. bioRxiv. Cold Spring Harbor Laboratory; 2020;2020.01.22.914994.

101. Keller MC, Fredrickson BL, Ybarra O, Cote S, Johnson K, Mikels J, et al. A warm heart and a clear head - The contingent effects of weather on mood and cognition. Psychol Sci. Blackwell Publishing; 2005;16:724–31.

102. D'Amato G, Holgate ST, Pawankar R, Ledford DK, Cecchi L, Al-Ahmad M, et al. Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the World Allergy Organization. World Allergy Organ J. Elsevier; 2015;8:UNSP 25.

103. Ruoppolo M, Scolamiero E, Caterino M, Mirisola V, Franconi F, Campesi I. Female and male human babies have distinct blood metabolomic patterns. Mol Biosyst. 2015;11:2483–92.

104. Diaz SO, Pinto J, Barros AS, Morais E, Duarte D, Negrão F, et al. Newborn Urinary Metabolic Signatures of Prematurity and Other Disorders: A Case Control Study. J Proteome Res. 2016;15:311–25.

105. Challis J, Newnham J, Petraglia F, Yeganegi M, Bocking A. Fetal sex and preterm birth. Placenta. 2013;34:95–9.

106. Wilkin TJ, Murphy MJ. The gender insulin hypothesis: why girls are born lighter than boys, and the implications for insulin resistance. Int J Obes. 2006;30:1056–61.

107. Wehrens R, Hageman JosA, van Eeuwijk F, Kooke R, Flood PJ, Wijnker E, et al. Improved batch correction in untargeted MS-based metabolomics. Metabolomics. 2016;12.

108. Petrick L, Edmands W, Schiffman C, Grigoryan H, Perttula K, Yano Y, et al. An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies. Metabolomics Off J Metabolomic Soc. 2017;13.

109. Folketinget. Bekendtgørelse af lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter - (The Danish Parliament. Order of law on scientific treatment of health scientific research projects) [Internet]. retsinformation.dk. 2017 [cited 2019 Oct 28]. Available from: https://www.retsinformation.dk/Forms/r0710.aspx?id=192671

Tables

Table 1. Subjects characteristics

	All subjects n = 74						After outlier removal n = 68					
	cases			controls			cases			controls		
Age at 1st Jan. 2006 (months, median (range))	7.6	(0.8-11.6)		7.6	(0.8-11.6)		7.3	(0.8-11.6)		7.5	(0.8-11.6)	
Gender (girls / boys)	8 / 29			8 / 29			7 / 25			8 / 28		
Classification of cases (ICD10)[1]												
- **F84.0 Childhood autism**	17			-			15			-		
- **F84.1 Atypical autism**	7			-			6			-		
- **F84.5 Asperger syndrome**	4			-			3			-		
- **F84.8 Other pervasive developmental disorders**	4			-			4			-		
- **F84.9 Unspecified pervasive developmental disorders**	11			-			10			-		
Gestational age (weeks, median (range), NA)	40	(33-42)	*5*	39	(30-42)	*2*	40	(33-41)	*4*	39	(30-42)	*2*
Birthweight (g, median (range), NA)	3500	(2210-4880)	*1*	3480	(977-4850)	*2*	3498	(2210-4880)	*0*	3490	(977-4850)	*2*
Age at sampling (days, median (range), NA)	6	(3-9)	*3*	6	(4-10)	*3*	6	(3-9)	*2*	6	(4-10)	*3*
Age of mother at birth (years, median (range))	31.7	(20.8-41.5)		31.7	(18.3-41.2)		32.3	(20.8-41.5)		31.8	(18.3-41.2)	

[1] ICD10 classification [2]

More details are provided in Additional file 6.

Table 2. Differentially abundant features in univariate analyses without FDR correction (p<0.01) and/or with high fold-change meeting inspection criteria

Putative annotation of relevant compounds	Annotation level[1]	*m/z*	RT (min)	ID	p-value t-test		p-value Wilcoxon rank-sum test		FC[2]	Network connections[3]
					without FDR correction	with FDR correction	without FDR correction	with FDR correction		
Methacholine $C_8H_{18}NO_2^+$	2[4]	160.13315	0.45	**159**	**0.0021**	0.9174	**0.0031**	0.9434	1.25	connected to 1853 (-0.036 m/z)
SIRIUS 99.96%: $C_{11}H_{22}N_2O_3$ (M+H+)	4	231.17005	2.78	**5593**	**0.0072**	0.9174	0.0138	0.9434	1.46	Single node
SIRIUS 7.12%: $C_{36}H_{63}N_{21}O_{14}$ (M+H+)	4	1014.48923	6.64	**8605**	0.0414	0.9174	0.0179	0.9434	**0.42**	Single node

[1] Annotation level of confidence according to the Metabolomics Standards Initiative (i.e. putative annotation) [68,69].

[2] FC: Fold-change (case/control)

[3] Network connections in GNPS feature-based molecular network.

[4] Source of annotation mzCloud (89.9% score). See its mass spectrum in Additional file 10.

Inspection criteria: peak integration or shape quality, initial missing values, single values plot, presence and consistence in replicated pool injections, annotation or connection to contaminants. For details see Additional file 7.

Table 3. Compounds reported in the literature three or more times as being associated with ASD

Compound name	Annot-ation level[1]	Raw formula	*m/z* [M+H]+	RT (min)	Feature ID (MZmine 2.41)	Detected by Compound Discoverer 2.1	HMDB ID	Literature reference
Arginine	1	C6H14N4O2	175.11895	0.35	1450	ND	HMDB0000517	[38,44,50]
Aspartic acid	1	C4H7NO4	134.04478	0.41	1073	ND	HMDB0000191	[34,35,40,41,50,57]
Citric acid	4	C6H8O7	193.03428	0.35	1776	yes	HMDB0000094	[40,42,51]
Creatine	2	C4H9N3O2	132.07675	0.40	16	yes	HMDB0000064	[42,47,52]
Creatinine	2	C4H7N3O	114.06619	0.40	281	yes	HMDB0000562	[40,50,52,53]
Decanoylcarnitine	1	C17H33NO4	316.24823	6.00	3633	yes	HMDB0000651	[37,41,57]
Glutamic acid	1	C5H9NO4	148.06043	0.38	136	yes	HMDB0000148	[34,35,38,40–43,52]
Glutamine	2	C5H10N2O3	147.07642	0.40	107	yes	HMDB0000641	[35,38,39]
Glycine	3	C2H5NO2	76.03930	0.38	1177	ND	HMDB0000123	[34,39,43,47,49,52]
Glycolic acid	-	C2H4O3	77.02332	-	ND	ND	HMDB0000115	[53–55]
Hippuric acid	2	C9H9N1O3	180.06552	3.04	5174	ND	HMDB0000714	[48,51,52,54]
Histidine	2	C6H9N3O2	156.07675	0.32	342	yes	HMDB0000177	[34,49,56]
Lactic acid	-	C3H6O3	91.03897	-	ND	ND	HMDB0000190	[35,44,52]
p-cresol	-	C7H8O1	109.06479	-	ND	ND	HMDB0001858	[34,35,53,56]
Phenylalanine	1	C9H11NO2	166.08625	1.70	594+5370+287	yes	HMDB0000159	[34,56,57]
Serine	2	C3H7NO3	106.04987	0.40	437	ND	HMDB0000187	[34,40,42,43,49]
Succinic acid	-	C4H6O4	119.03388	-	ND	ND	HMDB0000254	[40,47,48,54]
Taurine	3	C2H7NO3S	126.02194	0.43	428	ND	HMDB0000251	[25,44,47–49,52]
Threonine	2	C4H9NO3	120.06552	0.40	476	ND	HMDB0000167	[38,42,49,50]
Tryptophan	2	C11H12N2O2	205.09715	2.53	164	yes	HMDB0000929	[38,41,52,55,56]
Tyrosine	1	C9H11NO3	182.08117	0.72	58	yes	HMDB0000158	[35,42,57]
Valine	2	C5H11NO2	118.08625	0.42	ND	yes	HMDB0000883	[34,39,52]

ND: not detected

[1] Annotation level of confidence according to the Metabolomics Standards Initiative (i.e. putative annotation) [68,69].

When the [M+H]+ adduct could not be found (+/- 5 ppm), common adducts were searched including [M+Na]+, [M+K]+, [M+2H]2+, $[M+H-H_2O]+$.

See full list of compounds considered and more details in Additional file 11.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

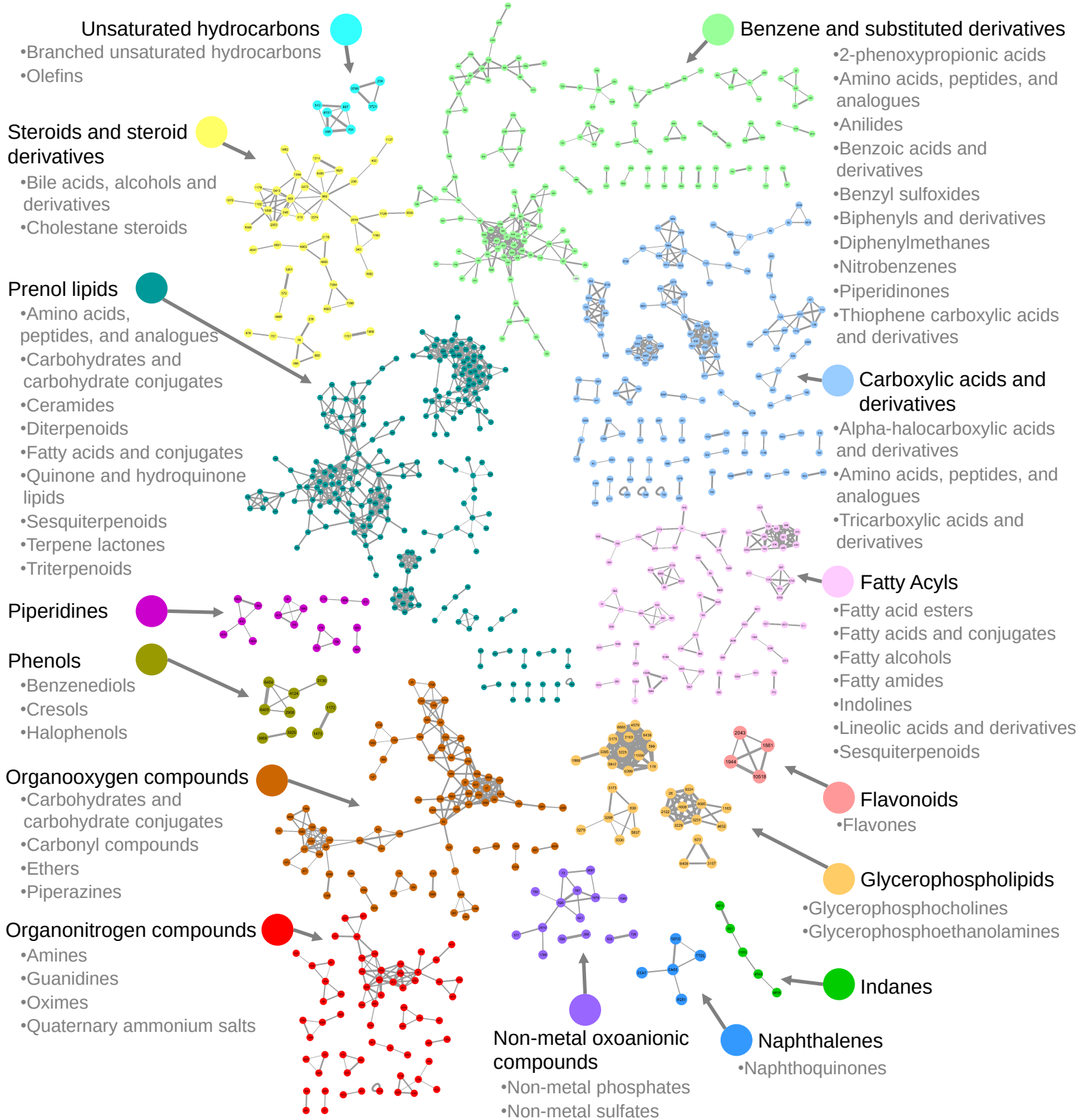
Figures

Figure 1. Feature-based molecular network displaying the 15 predominant putative chemical classes and their subclasses

Nodes represent mass spectral features and are used as a proxy for a metabolite. Connected nodes represent high tandem mass spectral similarity, and thus high chemical structural similarity. The thickness of the grey edges connecting nodes varies according to the cosine score representing to what extent two connected metabolites are chemically similar (based on MS2 spectra, from 0.7: less similar and thin edge to 1.0: identical and thick edge).

The name of annotated metabolites (levels 1 and 2), details on chemical classes with fewer than 4 metabolites (absent on this figure), chemical classification scores [73], all unknowns, and group intensities for all features (average, standard deviations) are detailed in Additional file 7.

See Figure_1.pdf

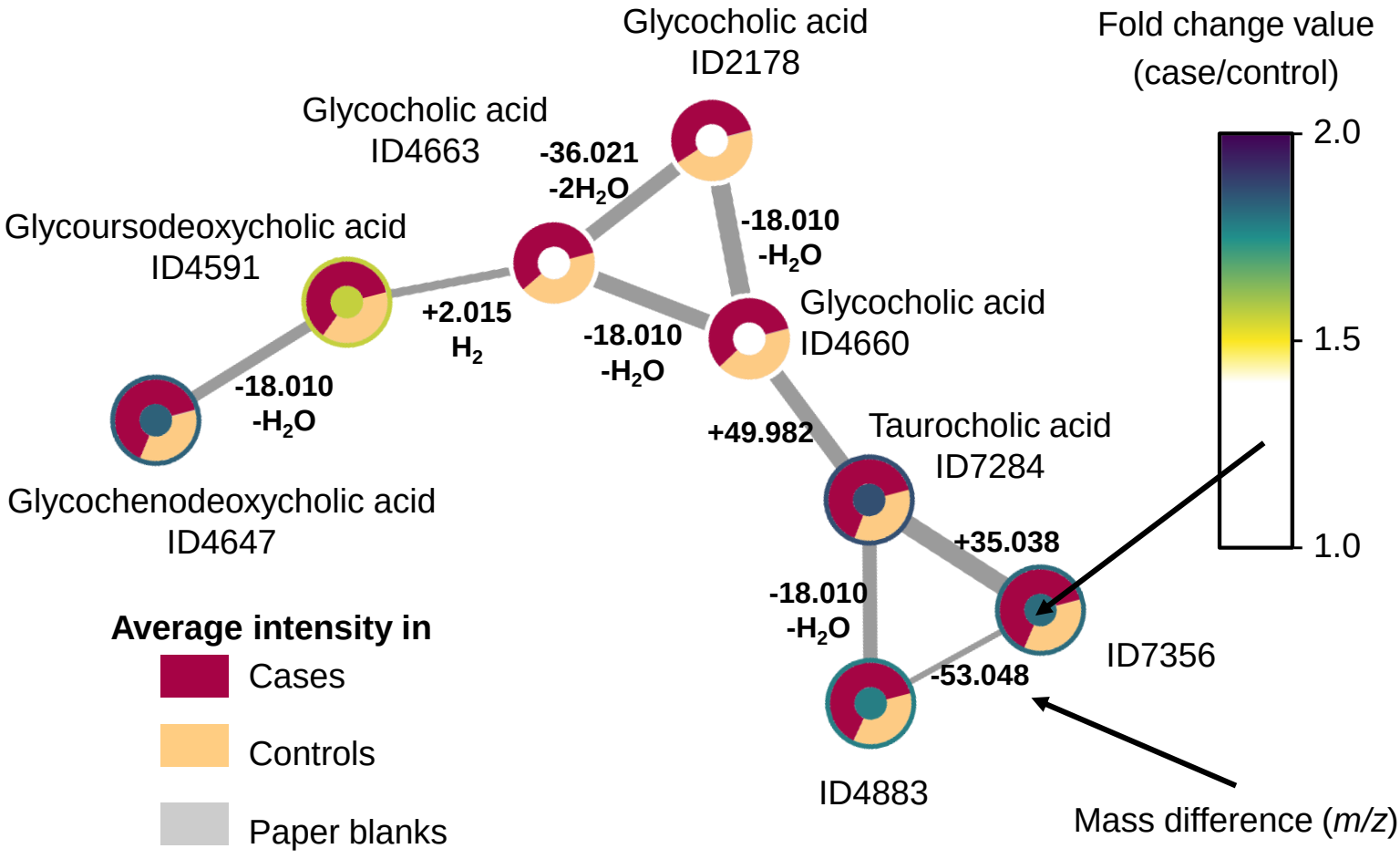

Unsaturated hydrocarbons
•Branched unsaturated hydrocarbons
•Olefins
Steroids and steroid derivatives
•Bile acids, alcohols and derivatives
•Cholestane steroids
Prenol lipids
•Amino acids, peptides, and analogues
•Carbohydrates and carbohydrate conjugates
•Ceramides
•Diterpenoids
•Fatty acids and conjugates
•Quinone and hydroquinone lipids
•Sesquiterpenoids
•Terpene lactones
•Triterpenoids
Piperidines
Phenols
•Benzenediols
•Cresols
•Halophenols
Organooxygen compounds
•Carbohydrates and carbohydrate conjugates
•Carbonyl compounds
•Ethers
•Piperazines
Organonitrogen compounds
•Amines
•Guanidines
•Oximes
•Quaternary ammonium salts
Benzene and substituted derivatives
•2-phenoxypropionic acids
•Amino acids, peptides, and analogues
•Anilides
•Benzoic acids and derivatives
•Benzyl sulfoxides
•Biphenyls and derivatives
•Diphenylmethanes
•Nitrobenzenes
•Piperidinones
•Thiophene carboxylic acids and derivatives
Carboxylic acids and derivatives
•Alpha-halocarboxylic acids and derivatives
•Amino acids, peptides, and analogues
•Tricarboxylic acids and derivatives
Fatty Acyls
•Fatty acid esters
•Fatty acids and conjugates
•Fatty alcohols
•Fatty amides
•Indolines
•Lineolic acids and derivatives
•Sesquiterpenoids
Flavonoids
•Flavones
Glycerophospholipids
•Glycerophosphocholines
•Glycerophosphoethanolamines
Indanes
Naphthalenes
•Naphthoquinones
Non-metal oxoanionic compounds
•Non-metal phosphates
•Non-metal sulfates

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2. Network of molecular features putatively annotated as bile acids with average group intensities, fold change values, mass differences and cosine scores displayed.

Molecular family #75 is composed of eight bile acid structural analogues (see details in Additional File 7). Coloring according to the fold change values makes it easier to spot the families with differential abundance in cases vs. controls. Displaying average intensities for the three groups (cases, controls, paper blanks) allows for a quick control of the noise (paper blanks, here no noise detected) and confirmation of fold change. On edges, while the thickness of the connection represents to what extent two metabolites are chemically similar, the mass difference is essential to support annotation as it translates into how molecules differ from one another (e.g. water loss, conjugation, adducts, etc.).

See Figure_2.pdf


Glycocholic acid
ID2178
Fold change value
(case/control)
2.0
1.5
1.0
Glycocholic acid
ID4663
-36.021
-2H2O
-18.010
-H2O
Glycoursodeoxycholic acid
ID4591
+2.015
H2
-18.010
-H2O
Glycocholic acid
ID4660
-18.010
-H2O
+49.982
Taurocholic acid
ID7284
+35.038
Glycochenodeoxycholic acid
ID4647
-18.010
-H2O
ID7356
-53.048
ID4883
Average intensity in
Cases
Controls
Paper blanks
Mass difference (m/z)

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

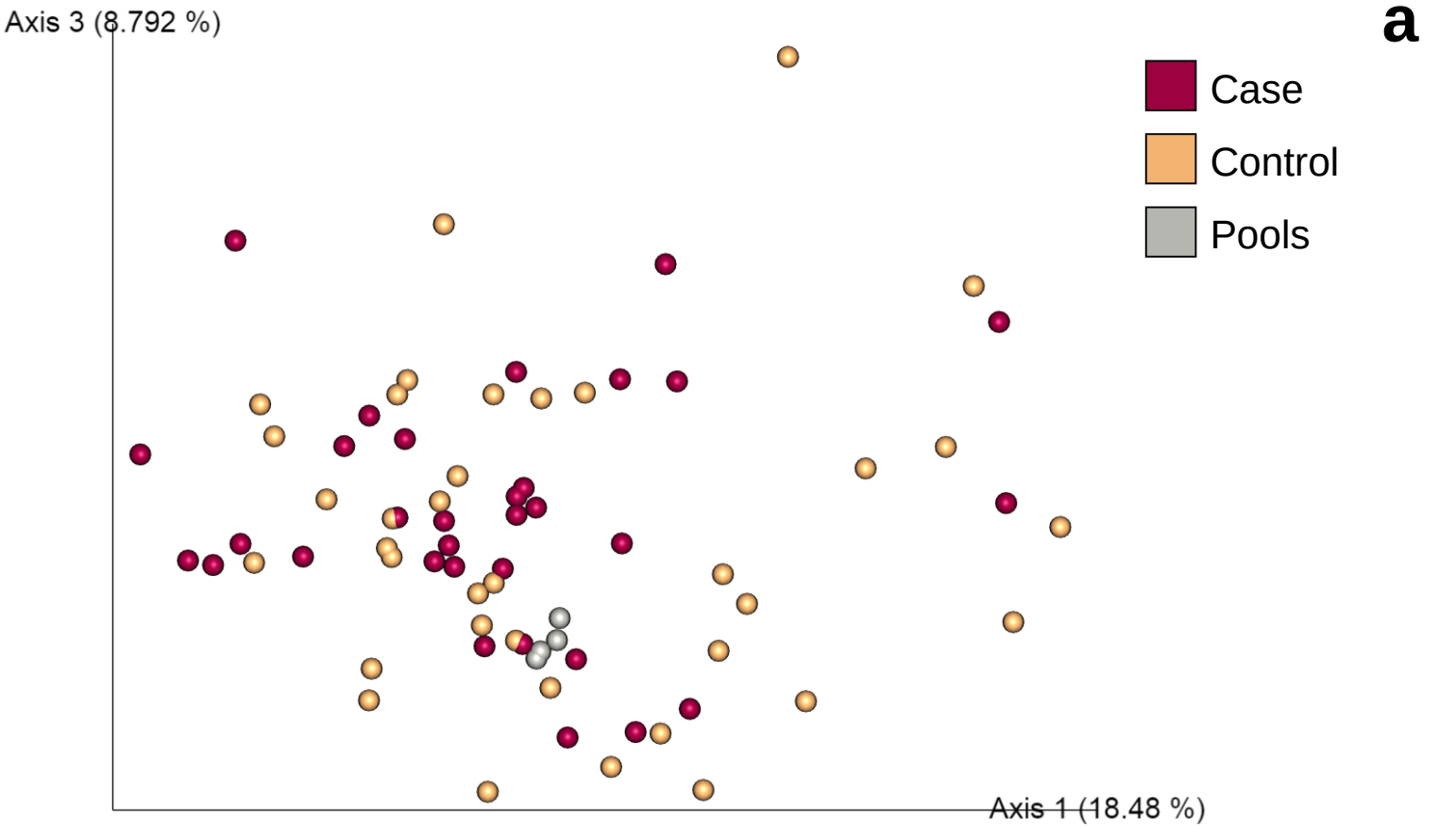
Figure 3. Bray-Curtis principal coordinates analysis of the 68 samples after outlier removal. Each sphere represents one sample. Axes are principal components 1 (x) and 3 (y) explaining 18.48% and 8.79% of the variability in the data, respectively. The four replicated pool injections cluster satisfactorily.

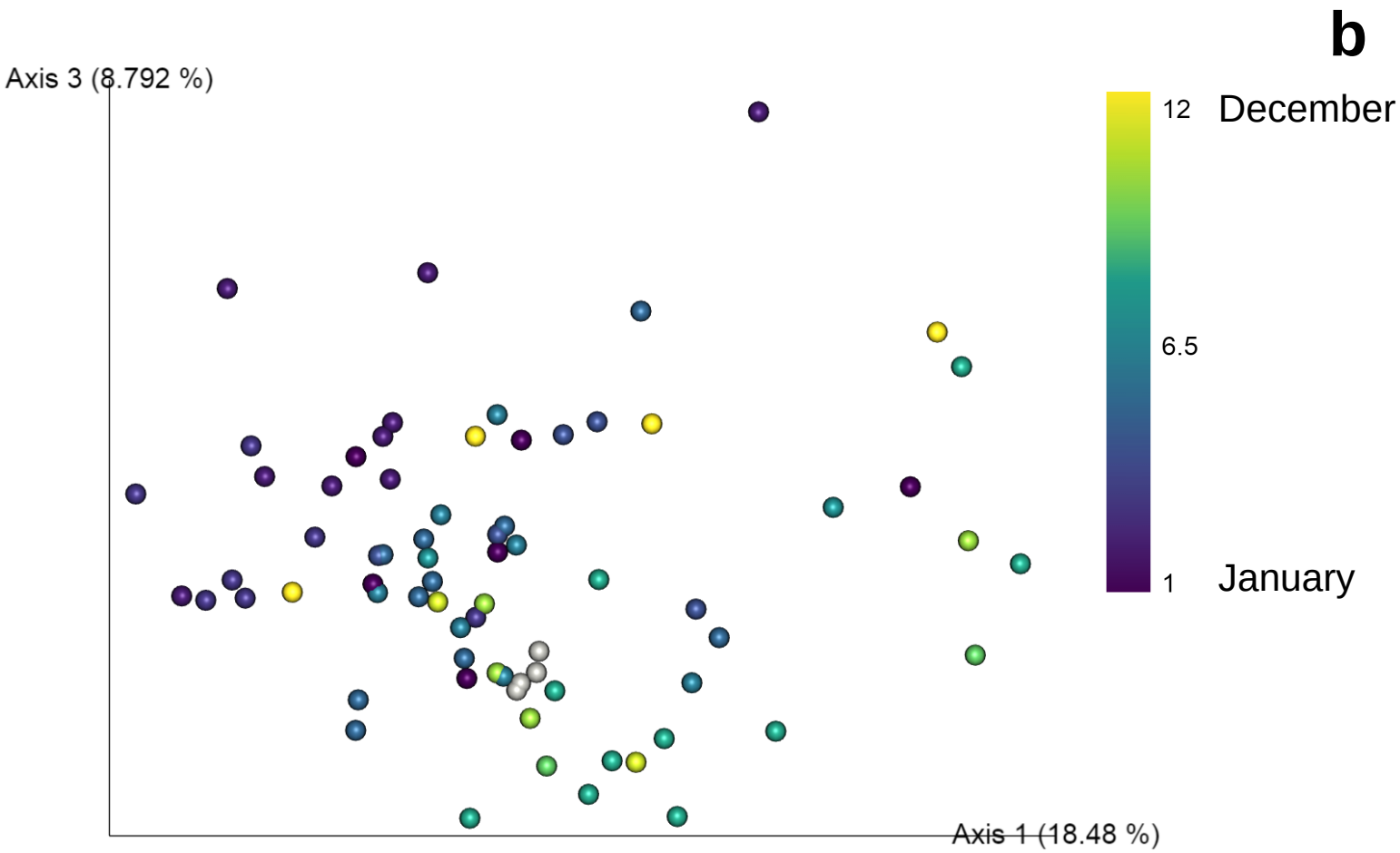
a) Coloring reflects the type of samples, i.e. cases, controls and four replicated pool injections. No clear distinction between cases and controls can be observed (PERMANOVA Adonis $R^2 = 0.014$, P-value = 0.777).

b) Coloring reflects the month of birth for the 68 samples as well as the four replicated pool injections. Samples collected in the winter (dark purple and yellow, extremities of the colors scale) are positioned away from the samples collected in summer (green and turquoise, middle of the color scale). 26.2% of the variation in the data can be explained by month of birth (Adonis $R^2 = 0.262$, P-value = 0.001)

See Additional file 8 for detailed values with and without FDR correction and for other distance matrices.

See Figure_3ab.pdf


a
Axis 3 (8.792 %)
Axis 1 (18.48 %)
Case
Control
Pools


b
Axis 3 (8.792 %)
Axis 1 (18.48 %)
12
December
6.5
1
January

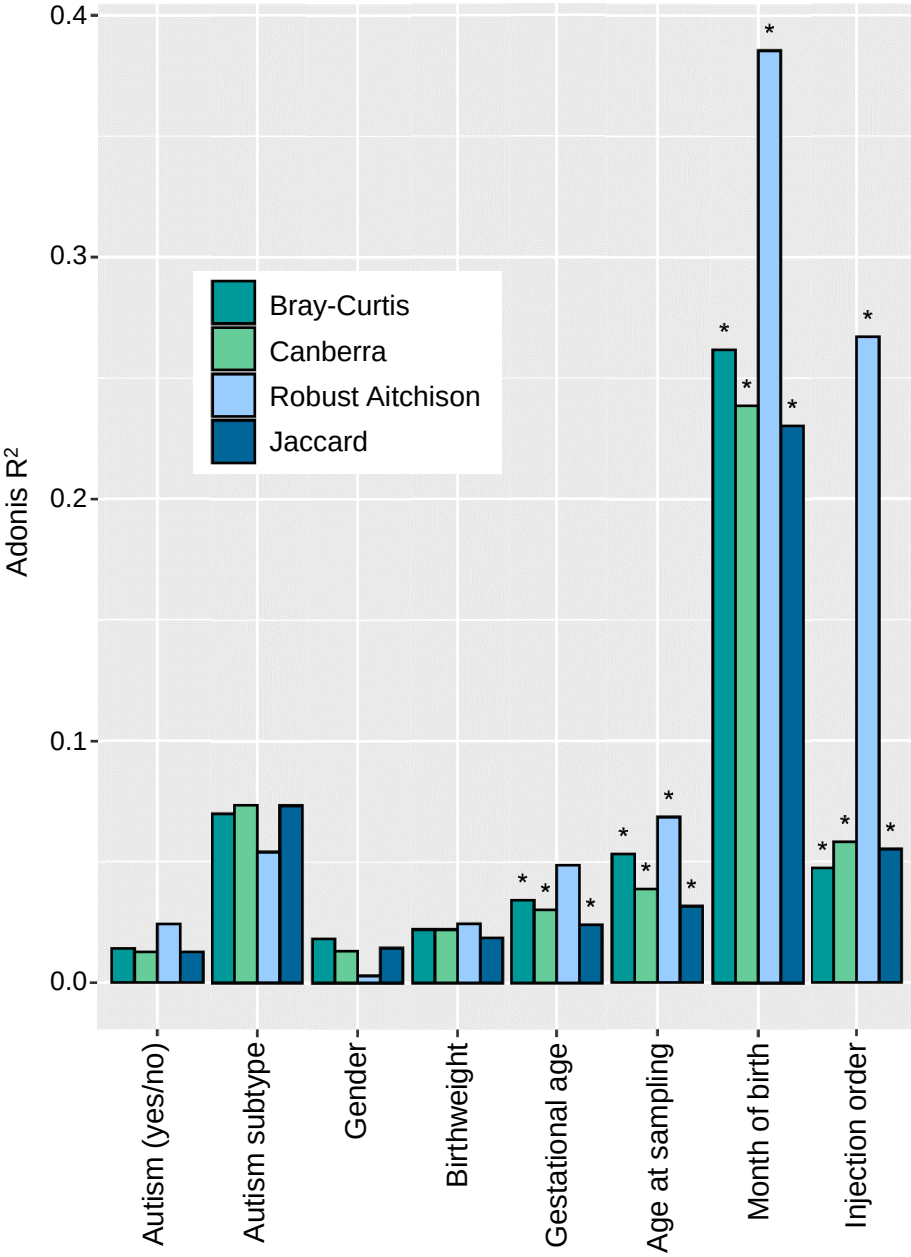
medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 4. PERMANOVAs of the four tested distance matrices showing how much of the variation (Adonis R^2) is explained by a metadata variable.

P-values include FDR correction. A star is present when the corrected p-values was <0.05.

All exact values are available in Additional file 8 and detailed metadata (subject characteristics) are available in Table 1 and Additional file 6.

See Figure_4.pdf


Adonis R²
0.4
0.3
0.2
0.1
0.0
Bray-Curtis
Canberra
Robust Aitchison
Jaccard
Autism (yes/no)
Autism subtype
Gender
Birthweight
Gestational age
Age at sampling
Month of birth
Injection order

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Additional files

Additional file 1:

Standardized reporting of untargeted metabolomics LC-MS/MS method according to [60] containing the list of internal standards from the Neobase Non-derivatized MSMS kit and their concentration in the extraction buffer

See .xlsx

Additional file 2:

MZmine batch .xml file used to preprocess the raw data

See .xml

Additional file 3:

Compound Discoverer 2.1 preprocessing workflow settings

See .pdf

Additional file 4ab:

PEG (4a) and PPG (4b) filtering scripts using the Kendrick Mass filter according to [80]

See the two .ipynb

Can be opened using https://jupyter.org/

Additional file 5:

Targeted analysis of outliers using TraceFinder (IS and unlabeled homologs) and heatmap of untargeted analysis

See .pdf

Additional file 6:

Subjects characteristics in details

See .xlsx

Additional file 7:

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069153; this version posted April 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

All features including annotated compounds. Out of the 4360 features detected, 150 could be annotated by GNPS library matching (annotation level 2) or in-house Trace Finder library (annotation level 1) and an additional 859 by MolNetEnhancer (annotation level 3).

See .xlsx

Additional file 8:

PERMANOVAs Adonis R^2 values and p-values calculated with (74 samples) and without (68 samples) outliers

See .xlsx

Additional file 9:

Full list of compounds with t-test p value <0.01 without FDR correction and/or with Wilcoxon rank-sum test p value <0.01 without FDR correction and/or case/control fold change value <0.5 or >2.

See .xlsx

Additional file 10: Fragmentation profiles of the two unknown features to be monitored in future studies as well as methacholine as shown in Table 2.

See .pdf

Additional file 11: Full list of compounds reported in the literature as involved in ASD and considered in this study.

See .xlsx