HHS Public Access

Author manuscript

Brain Behav Immun. Author manuscript; available in PMC 2024 July 01.

Published in final edited form as:

Brain Behav Immun. 2023 July ; 111: 21–29. doi:10.1016/j.bbi.2023.03.026.

Phenotypically driven subgroups of ASD display distinct metabolomic profiles

Nicole Prince[a], **Su H. Chu**[a], **Yulu Chen**[a], **Kevin M. Mendez**[a], **Ellen Hanson**[b], **LeeAnne Green-Snyder**[c], **Elizabeth Brooks**[c], **Susan Korrick**[a,d], **Jessica A. Lasky-Su**[a], **Rachel S. Kelly**[a,*]

[a]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[b]Divisions of Neurology and Developmental Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

[c]Simons Foundation, New York, NY, USA

[d]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Autism Spectrum Disorder (ASD) is a heterogeneous condition that includes a broad range of characteristics and associated comorbidities; however, the biology underlying the variability in phenotypes is not well understood. As ASD impacts approximately 1 in 100 children globally, there is an urgent need to better understand the biological mechanisms that contribute to features of ASD. In this study, we leveraged rich phenotypic and diagnostic information related to ASD in 2001 individuals aged 4 to 17 years from the Simons Simplex Collection to derive phenotypically driven subgroups and investigate their respective metabolomes. We performed hierarchical clustering on 40 phenotypes spanning four ASD clinical domains, resulting in three subgroups with distinct phenotype patterns. Using global plasma metabolomic profiling generated by ultrahigh-performance liquid chromatography mass spectrometry, we characterized the metabolome of individuals in each subgroup to interrogate underlying biology related to the subgroups. Subgroup 1 included children with the least maladaptive behavioral traits (N = 862); global decreases in lipid metabolites and concomitant increases in amino acid and nucleotide pathways were observed for children in this subgroup. Subgroup 2 included children with the highest degree of challenges across all phenotype domains (N = 631), and their metabolome profiles demonstrated aberrant metabolism of membrane lipids and increases in lipid oxidation products. Subgroup 3 included children with maladaptive behaviors and co-occurring conditions that showed the highest IQ scores (N = 508); these individuals had increases in sphingolipid metabolites and fatty acid byproducts. Overall, these findings indicated distinct metabolic patterns

*Corresponding author at: Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA., hprke@channing.harvard.edu (R.S. Kelly).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbi.2023.03.026.

within ASD subgroups, which may reflect the biological mechanisms giving rise to specific patterns of ASD characteristics. Our results may have important clinical applications relevant to personalized medicine approaches towards managing ASD symptoms.

Keywords

Autism spectrum disorder; Simons Simplex Collection; Metabolomics; Lipid metabolism; Amino acid metabolism; Autism severity

1. Introduction

Autism Spectrum Disorder (ASD) is a heterogeneous disorder that encompasses a broad range of phenotypic characteristics related to cognition, language ability, and behavior. It is often accompanied by co-occurring conditions, including sleep disorders, seizures, anxiety, and depression, among others (Lord et al., 2020). ASD affects approximately 1 in 100 children globally (Zeidan et al., 2022), with characteristics that vary across the spectrum of cases (Braden et al., 2022) and contribute to heterogeneity of the disorder. While genetic and environmental contributors have been identified (Braden et al., 2022; Fischbach and Lord, 2010; Hassan and Mokhtar, 2019), additional research is needed to uncover the biological drivers that give rise to varied ASD symptomology. Identifying phenotypically homogenous subgroups of ASD cases may improve the understanding of the molecular mechanisms underlying clinical manifestation and presentation, ultimately aiding the development of novel management strategies targeted to specific subgroups (Persico and Sacco, 2014). Investigating biological patterns within and between subgroups has proven beneficial in neurological research and has enlightened patterns of disease pathogenesis and features (Crouse et al., 2018) to enhance clinical approaches in ASD (Rubin and Panzano, 2002). Subtyping approaches have also been successfully applied to many other heterogeneous conditions that arise from diverse underlying biology (Oron and Elliott, 2017), such as asthma (Kelly et al., 2022), diabetes (Jun et al., 2020; Zaghlool et al., 2022), and cancer (Cao et al., 2021; Gumpenberger et al., 2021) improving personalized medicine approaches for these diseases. These methods may similarly benefit ASD research and facilitate improved understanding of the processes that give rise to the diverse and unique sets of symptoms.

A number of environmental risk factors have been identified in ASD development and progression, including infections during pregnancy, maternal zinc deficiency, and gestational diabetes, among others (Georgiades et al., 2013; Grabrucker, 2012), and these environmental components are thought to interact with underlying genetics to impact ASD risk and presentation (Fischbach and Lord, 2010; Sanders et al., 2015), etiology (Kalsner et al., 2018; Skaar et al., 2005), and symptomology (Clements et al., 2020; Nayar et al., 2021). These genetic-environmental interactions ultimately drive downstream components of the central biological dogma, and the metabolome represents the downstream products of genetic, epigenetic, proteomic, and environmental influences contributing to altered physiological states (Rattray et al., 2018; Walker et al., 2019). This property makes the metabolome an attractive target to understand disruptions to underlying biology that are tied to disease

mechanisms, such as phenotypic patterns expressed by individuals with ASD. Previous metabolomic investigations of children with ASD have identified biomarkers for clinical evaluation (Orozco et al., 2019) and development (Smith et al., 2020) and noted important pathways that distinguish individuals with ASD from typically developing peers (Liang et al., 2020; Smith et al., 2019). However, these studies have not yet attempted to evaluate differences in the metabolome across phenotypically driven subgroups within ASD. Characterizing metabolomic differences across subsets of individuals with ASD who share similar symptomology could reveal important insights into pathway alterations related to specific symptom patterns and severity.

In the current study, we sought to derive phenotypically-driven subgroups of individuals with ASD in 2001 well-characterized children from the Simons Simplex Collection (SSC) (Fischbach and Lord, 2010) using unsupervised hierarchical clustering. We then investigated metabolic differences between subgroups to gain biological insight into pathways associated with ASD traits. Our hypothesis was that phenotypically distinct subgroups would demonstrate differences in the metabolome corresponding to altered underlying biology, thereby identifying pathways that may have clinical relevance for ASD management.

2. Materials and methods

2.1. Simons Simplex Collection (SSC)

The Simons Foundation Autism Research Initiative (SFARI) SSC study recruited families with only one child with ASD across twelve clinical and university sites between 2007 and 2011 (Fischbach and Lord, 2010). Clinical psychologists at each site comprehensively evaluated families to assess diagnostic measures and core features of autism, including intellectual ability, adaptive behaviors, emotional and behavioral problems, motor function, and language. An extensive medical history was obtained for children, including prenatal and perinatal history, developmental milestones, immunizations, medications, dietary supplements, and common behavioral treatments; investigators also collected information on co-occurring conditions and commonly associated “comorbidities” including gastrointestinal complaints, sleep irregularities, and seizures. Children were aged 4 to 17 years, and exclusion criteria included: nonverbal mental age below 18 months, severe neurological deficits, birth trauma, prenatal complications, or genetic evidence of Fragile X or Down syndromes. A detailed description of recruitment and exclusion/inclusion criteria can be found at http://sfari.org. A total of 2001 ASD cases from SSC were included in our study, based on availability of complete phenotype information and suitability of a blood sample for metabolomic profiling.

2.2. Clinical and phenotypic features

A diverse set of phenotypes across multiple symptom domains of ASD were chosen based on previous phenotype analyses in the SSC (Hirota et al., 2020; Matta et al., 2018; Narita et al., 2020; Sullivan et al., 2019) and consultation with clinical psychologists affiliated with the Simons Foundation. The full list of included phenotypes used is shown in Table 1. Phenotypes spanned four clinical domains: 1) core ASD traits measures from diagnostic instruments (Autism Diagnostic Interview-Revised [ADI-R], Autism Diagnostic Observation

Schedule [ADOS], Social Responsiveness Scale [SRS], Vineland Adaptive Behavior Scale-II [VABS-II], Aberrant Behavior Checklist [ABC], and Repetitive Behavior Scale [RBS]); 2) cognitive and adaptive functioning scores (verbal and nonverbal intelligence quotient [IQ], VABS-II), 3) language and communication scores (VABS-II, ADI-R), and 4) maladaptive behavioral and co-occurring conditions (ABC, RBS, Child Behavior Checklist [CBCL], and seizures). The categorization of phenotypes into these domains did not affect the clustering process and was used solely for interpretation purposes.

Some phenotype scores were modified prior to clustering according to previously utilized methods. ADOS social affect (SA) and restricted and repetitive behaviors (RRB) scores were converted to Calibrated Severity Scores (CSS) to adjust for effects of age and sex using methods described and validated by Hus et al. (2014). Four additional phenotypes based on the SRS questionnaire were altered to account for impacts of age and sex using methods similar to those applied by Hus *et al.* and Frazier *et al.* (Frazier et al., 2014). These were: 1) SRS Parent Total T-scores were adjusted by regressing out age to account for potential confounding; SRS Parent raw scores for 2) Communication, 3) Awareness, and 4) Mannerisms were adjusted by regressing out age and sex to account for potential confounding. Scores for all 40 phenotypes were standardized to a mean of 0 and a standard deviation of 1 prior to clustering.

2.3. Global metabolomic profiling of plasma samples

The SSC collected non-fasting blood samples from study participants during baseline clinical site visits. Blood draws were collected using EDTA tubes, and plasma was separated and stored at −80 °C until analysis. Plasma samples included in this study were selected from the SSC repository, and global metabolomic profiling was performed on 2001 samples by Metabolon, Inc. (NC, USA) through ultrahigh-performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) using their global profiling platform. Details of UPLC-MS/MS methods are available in the Supplementary Methods. Following data collection, missing metabolite values were imputed by replacement with half of the lowest observed value in all samples, for each metabolite individually; metabolites missing data in more than 75% of samples and unnamed features were omitted from analysis, resulting in 989 metabolites included in our analysis. Metabolite values were then log-10 transformed and *Pareto*-scaled.

2.4. Unsupervised hierarchical clustering to generate subgroups

We performed an exploratory, hypothesis-generating analysis using unsupervised hierarchical clustering to identify subgroups of children with similar phenotype patterns. Unsupervised hierarchical clustering was conducted using the 40 ASD phenotypes included in this study (Table 1) and Ward's method. The number of clusters was determined using the Elbow method by plotting total within sum of squares (Thorndike, 1953) as well as visual inspection of the clustering dendrogram. Cluster analysis was performed in R version 4.1.3 (Team, 2020) using the NbClust (Charrad et al., 2014) version 3.0.1 and Cluster (Maechler et al., 2022) version 2.1.4 packages. Generated clusters are hereafter referred to as subgroups. Following subgroup assignments, pairwise comparisons between all pairs of groups were performed for each of the 40 phenotypes using Tukey's Honest

Significant Difference (HSD) test to investigate differences in phenotype scores and patterns of phenotype expressions between subgroups.

2.5. Logistic regression models to assess metabolite associations with subgroups

Associations between the 989 metabolites included in analysis and subgroup assignment were evaluated using logistic regression to identify metabolites associated with membership in each subgroup and to characterize metabolomic profiles of individuals within each of the three phenotype driven subgroups. Metabolites were treated as independent variables, and the dependent variable was a binary yes/no variable assignment into a particular subgroup vs. all other subgroups. Adjustment covariates for potential confounding were based on *a priori* considerations, including age, sex, body mass index (BMI) z score, maternal education level, and household income. Multiple testing correction to control false discovery rate (FDR) was performed using the Benjamini and Hochberg method of P-value correction (Benjamini and Hochberg, 1995). Logistic regression analysis was evaluated using the stats package in R version 4.1.3. For metabolites that were significantly associated with subgroup membership at an FDR < 0.05 threshold, we visually inspected the linearity between the metabolite and membership in each group to assess validity.

2.6. Biological pathway enrichment analysis

Enrichment analysis of biological pathways was performed for each subgroup based on Metabolon sub-pathway assignments and using results of logistic regression models, specifically the odds ratio estimates and P-values for each metabolite association. We utilized the Chemical Similarity Enrichment Analysis (ChemRICH) (Barupal and Fiehn, 2017) package in R to identify enriched pathways and directions of pathway alterations for each ASD subgroup. ChemRICH uses a one-sample Kolmogorov-Smirnov test to calculate a P-value for each biological pathway, including all individual metabolites categorized in that sub-pathway. ChemRICH was advantageous for this purpose given its self-contained p-values and non-reliance on database curation (Barupal and Fiehn, 2017). In order for a pathway to be enriched through ChemRICH analysis, a minimum of three metabolites must be detected in that pathway after quality control procedures were applied to UPLC-MS/MS metabolomics data. Pathways with only one or two metabolites present cannot be identified as enriched. The number of altered metabolites is given by the ChemRICH output, as well as the increase ratio, which indicates the number of metabolites with a positive direction of effect in logistic regression analyses over the total number of altered metabolites detected for a given pathway.

3. Results

3.1. Demographic characteristics of SSC cohort

ASD individuals included in this study ranged in age from 4 to 17 years (mean [SD]: 8.94 [3.34]) and were predominantly male (87.0%) and White (79.6%), which is reflective of the overall makeup of the SSC cohort (Fischbach and Lord, 2010). The average BMI z score was 0.63 with a standard deviation of 1.36. Of the 2001 individuals included in this study, 30.1% of families reported “low” maternal education (high school education or below), 44.2% reported “medium” maternal education (associate or baccalaureate degree),

and 25.7% reported “high” maternal education (graduate degree or above). Household income categories were also assessed, and 16.2% reported household annual income of $50,000 or less, categorized as “low”; 41.4% reported household income between $51,000-$100,000, categorized as “medium”, and 42.7% reported household income above $100,000, categorized as “high.”

3.2. ASD subgroup assignments were associated with differences across phenotype domains

Our analysis identified three subgroups of children with ASD in the SSC. Pairwise comparisons of each of the 40 ASD phenotypes revealed significantly different phenotype patterns across each of the three subgroups. Phenotypes that did not significantly differ between subgroups at FDR < 0.05 are marked with a superscript and footnote in Table 2. A summary of phenotype differences is provided in Fig. 1, and phenotype comparisons are described in detail in the following paragraphs.

Subgroup 1 (N = 862 individuals) showed the lowest degree of maladaptive behaviors and co-occurring conditions. However, they demonstrated greater challenges in the majority of cognitive and adaptive functioning scores and language and communication scores compared to Subgroup 3, but had less challenges in these areas than Subgroup 2. This pattern was retained for 7 core ASD trait measures derived from ADOS and ADI-R scores but not in 13 core ASD traits derived from SRS, ABC, and RBS scores.

Subgroup 2 (N = 631 individuals) included children with the highest degree of challenges in any subgroup. These children reflected the most impaired performance in scores across all four phenotype domains compared to other subgroups, including 37 of the total 40 total phenotypes included, at these findings were significant at FDR < 0.05 (Table 2). They demonstrated higher impairment on 22 of 23 core ASD trait measures, all cognitive and adaptive functioning scores, all language and communication scores, and the majority of maladaptive behaviors and co-occurring conditions.

Subgroup 3 (N = 508 individuals) included children with the second highest impairment in maladaptive behavioral and co-occurring conditions following Subgroup 2, but these individuals reported highest IQ of any subgroup. They also showed the second greatest degree of impairment in 13 core ASD traits measured by SRS, ABC, and RBS questionnaires.

Differences in demographic variables between the three generated subgroups were also evaluated through pairwise comparisons for each pair of subgroups. No significant differences were observed between any pairs of subgroups in sex, age, race, ethnicity, or BMI z score, indicating these were not significantly associated with subgroup differences. However, maternal education category and reported household annual income category differed significantly in Subgroup 2 compared to Subgroups 1 and 3; Subgroups 1 and 3 were not significantly different at FDR < 0.05. A larger proportion of individuals in Subgroup 2 reported “low” and “medium” categories for education and income.

3.3. Biological pathway enrichment reveals differing biology between subgroups

Logistic regression evaluated metabolites significantly associated with membership in a particular subgroup, using a one-versus-all approach. Results of logistic regression analyses of each subgroup are shown in Supplementary File S1 and were used as input for ChemRICH biological pathway analysis to look for metabolic differences between subgroups; enrichment results are summarized in Table 3 and Fig. 1.

Children in Subgroup 1 showed the least maladaptive behavioral challenges and lowest co-occurring conditions, but these individuals demonstrated the second highest challenges in 7 core ASD traits, cognitive & adaptive functioning, and language & communication domains. Eighteen pathways were significantly enriched for this subgroup. Of these eighteen, thirteen were lipid metabolism pathways, ten of which showed decreased metabolite levels (Table 3). Of the remaining 5 enriched pathways, two were amino acid, two were nucleotide, and one was a peptide pathway (Table 3); all but one nucleotide pathway showed increased levels of metabolites. There were nine individual metabolite associations that met FDR-corrected P-value < 0.05 for membership in Subgroup 1, including three metabolites in the Hexosylceramide (HCER) lipid pathway (ORs = 0.034 to 0.12, P = 1.28×10^{-5} to 0.04), one Lysophospholipid metabolite (OR = 0.10, P = 0.02), one Sphingolipid synthesis metabolite (OR = 1.80, P = 0.04) and one Monoacylglycerol metabolite (OR = 0.31, P = 0.04). Associations with individual metabolites are available in Supplementary File S1. Six pathways were significantly enriched across both Subgroups 1 and 2, but five were lipid pathways with opposite directions of effect: Monoacylglycerol, Diacylglycerol, Phosphatidylinositol (PI), Phosphatidylethanolamine (PE), and Lysophospholipid metabolism Supplementary File S2. Only one nucleotide pathway, Adenine containing purine metabolism, showed similar directions of effect between Subgroups 1 and 2. The Tryptophan pathway was not enriched in any subgroup, but increased levels of serotonin (OR = 1.84, P = 0.01; Supplementary File S1) and N-acetylkynurenine (OR = 2.70, P = 0.04) were significantly associated with membership in Subgroup 1. Increased levels of nicotinamide (OR = 2.94, P = 0.04) were associated with Subgroup 1 membership, but this metabolite did not belong to any enriched pathways.

In Subgroup 2 children with the highest degree of impairment, ten biological pathways were enriched, as shown in Table 3. Of these ten pathways, six fell into lipid categories, and the remaining four were categorized into peptide, amino acid, nucleotide, and cofactor classes. All enriched pathways in this subgroup showed increases in metabolite levels, except the cofactor pathway Hemoglobin and Porphyrin Metabolism, which showed decreased levels of metabolites. Only two metabolites demonstrated significant associations with Subgroup 2 membership following FDR correction (Supplementary File S1), and both belonged to the Diacylglycerol pathway: oleoyl-arachidonoyl-glycerol (18:1/20:4) (OR = 4.33, P = 0.03) and linoleoyl-arachidonoyl-glycerol (18:2/20:4) (OR = 4.07, P = 0.04).

Subgroup 3 children were least severe in core ASD traits related to ADOS and ADI-R and cognitive & adaptive functioning domains, but second highest needs in core ASD traits related to SRS, RBS, and ABC, as well as maladaptive behavioral and co-occurring conditions. These children demonstrated high IQ (Table 2), as described in previous

sections. Children of this subgroup showed enrichment of seven pathways; of these seven, six were lipid pathways with four demonstrating increases in metabolite levels, and the remaining pathway was a peptide pathway showing decreased metabolite levels (Table 3). Despite similar patterns in maladaptive behaviors and co-occurring conditions between Subgroups 2 and 3, there were no overlaps in enriched biological pathways between these two groups. Five pathways overlapped between Subgroups 1 and 3, but all five demonstrated opposite directions of enrichment (Table 3). Four of these were lipid pathways: Sphingolipid synthesis, Sphingosines, Sphingomyelins, and Fatty acid-Dicarboxylates; the remaining pathway was metabolism of Dipeptides. No individual metabolites demonstrated significant associations with Subgroup 3 membership at an FDR-corrected P-value threshold of 0.05.

4. Discussion

ASD is a heterogeneous condition characterized by a broad range of phenotype presentations (Lord et al., 2020), but the underlying biological factors associated with differences in phenotype presentations are not well understood. In this analysis, we generated three subgroups of ASD children through unsupervised hierarchical clustering of a diverse range of clinical ASD measures to capture subgroups of individuals. We then interrogated the biological differences between these subgroups by leveraging metabolomic data and demonstrated unique alterations in global metabolomic profiles correlated to each subgroup. Previous studies have illustrated the utility of clustering approaches to form subgroups in ASD (Matta et al., 2018; Narita et al., 2020), observing correlations between the genome and phenotypes, while others have investigated patterns of ASD traits to identify groups that varied in degree of symptoms (Sullivan et al., 2019). To our knowledge, this is the first study to apply hierarchical clustering within only ASD-affected individuals to understand metabolomic alterations in underlying biological pathways related to diverse ASD symptomology and traits. Unlike these previous studies, we incorporated metabolomic data to interrogate altered biological pathways that could ultimately be leveraged to enhance symptom treatment and provide a better understanding of the pathways that may be related to these traits.

Individuals in Subgroup 1 showed the lowest degree of impairment related to behavioral and co-occurring condition domains of any subgroup in this study. As behavioral issues have previously been associated with comorbidities in the SSC cohort (Hirota et al., 2020; Mazurek et al., 2013), it was unsurprising that Subgroup 1 showed consistently lower indices of severity for phenotypes related to both maladaptive behaviors and co-occurring conditions. Metabolomic investigation demonstrated predominant decreases in lipid metabolites belonging to lipid metabolism pathways, unlike Subgroups 2 and 3 that showed only increases in metabolites of lipid metabolism pathways. Several lipid pathways showed opposite directions of effect to other subgroups, such as monoacylglycerol (MAG), diacylglycerol (DAG), lysophospholipid (LPL), and PE pathways compared to Subgroup 2, or sphingomyelin, sphingolipid synthesis, sphingosines, and dicarboxylated fatty acid pathways compared to Subgroup 3. Several metabolites belonging to these pathways showed strong associations with Subgroup 1 membership and could be potential driving metabolites between subgroups or biomarkers for future consideration, including 1-linoleoyl-GPE (18:2) in the LPL pathway, sphinganine of the sphingolipid synthesis pathway, and 1-

oleoylglycerol (18:1) of the MAG pathway. Decreased levels of metabolites of 10 lipid pathways observed for Subgroup 1 may suggest reduced neuroinflammation that has previously been attributed to ASD traits (Chauhan and Chauhan, 2006; Tamiji and Crawford, 2010). This contrast could partially explain the lower degree of maladaptive behaviors and co-occurring conditions compared to the other subgroups. Additionally, increased serotonin levels were observed in children of this subgroup, another likely contributor to these differences. Depleted serotonin has been implicated in behavioral issues of ASD (Muller et al., 2016) and has historically been associated with comorbidities such as depression (Moncrieff et al., 2022), which were lowest in this subgroup. Increased levels of serotonin and N-acetylkynurenine may be reflecting a lesser impact of comorbidities and behavioral issues observed for this subgroup. Concomitant increases in levels of nicotinamide may be related to increases in these two metabolites of the tryptophan pathway, as increased nicotinamide is correlated with increased plasma serotonin (Tian et al., 2013). However, Subgroup 1 individuals displayed some impairment in other ASD domains, including the second highest in severity of cognitive & adaptive functioning scores and language & communication scores. They also showed the second highest severity in core ASD traits derived from ADOS and ADI-R scores, following Subgroup 2. Purine metabolism of adenine nucleotides showed consistently higher metabolite levels across Subgroups 1 and 2 that showed higher impairment of these two domains. Disruptions to purine metabolism have previously been associated with ASD (Geryk et al., 2020; Naviaux, 2018), and our results indicated that aberrant metabolism and a subsequent buildup of purine metabolism products may be associated with these deficits. Finally, increases in metabolites of other pathways previously linked to autism, such as increased glycine, serine, and threonine metabolism (Orozco et al., 2019) and urea cycle metabolism (Page and Coleman, 2000) were observed for Subgroup 1 individuals and may be related to the unique pattern across phenotype domains that characterizes these children.

Subgroup 2 represented children with the highest degree of impairment across all four phenotype domains, showing worse scores for 37 of the 40 total phenotypes explored. In Subgroup 2, we observed increased levels of metabolites belonging to membrane lipid classes such as phosphatidylinositol (PI) and phosphatidylethanolamine (PE) metabolites. Lipid metabolism errors have been linked to inflammation, impaired immunity, and oxidative stress associated with ASD symptoms, as they induce neuroinflammation (Chauhan and Chauhan, 2006), so this could reflect increased neuroinflammation associated with worse phenotype scores. Individual metabolite associations with DAG metabolites further support this, demonstrating increases in second messenger DAG metabolites that function as components of the lipids that make up cellular membranes (Eichmann and Lass, 2015). These increases in PI and PE lipid breakdown showed concomitant increases in pathways to counteract associated inflammation and stress, as corticosteroid, gamma-glutamyl amino acid, and S-adenosylmethionine (SAM) metabolism pathways that were also enriched uniquely in this subgroup. These three pathways involve mechanisms to counteract oxidative species and prevent cellular damage (de Kloet et al., 2008; Ohja et al., 2018; Villalobos et al., 2000; Whitfield, 2001), suggesting subsequent disruption to oxidative balance in response to lipid breakdown in more severe ASD. Our results supported

these findings, implying that lipid oxidative pathways could hold important clinical value to counteract severe ASD symptoms.

Subgroup 3 included children with the highest IQ scores, but these individuals still reported maladaptive behavioral traits and co-occurring conditions, including core ASD trait measures related to behaviors derived from SRS, ABC, and RBS evaluation. The pattern of children with with ASD reporting high IQ that express maladaptive behaviors has been observed in previous subtyping analyses (Reardon et al., 2022), so this may represent a distinct group that would benefit from personalized treatment regimens to address their specific needs. Individuals in this subgroup showed increases in six lipid pathways related to sphingolipid and fatty acid metabolism, and decreased metabolite levels of one dipeptide pathway. Fatty acid acylation of the sphingoid base structure produces ceramides (Levy and Futerman, 2010), so increases in products of fatty acid and sphingolipid metabolism pathways observed with Subgroup 3 membership are likely related by this common conversion mechanism. Further, elevated sphingolipid metabolism has been observed in other neurological conditions (Ben-David and Futerman, 2010), and sphingosine-1-phosphate has even been proposed as a serum biomarker for ASD (Wang et al., 2016), emphasizing the importance of this pathway in ASD. Our findings suggested sphingolipid metabolism pathways could be particularly relevant for children in Subgroup 3. While sphingosines and sphingolipid synthesis pathways were significantly enriched in both Subgroups 1 and 3, they demonstrated opposite directions of effect, further supporting the unique metabolomic profiles of these phenotypically distinct groups. Notably, there were no overlaps in enriched pathways between Subgroups 2 and 3, suggesting that despite some commonalities in phenotype expression related to maladaptive behaviors and co-occurring conditions, diverse underlying biology may give rise to differing degree of challenges.

While there were distinct phenotype patterns observed in our exploratory subgroup analysis, the range of severity in SSC was limited by recruitment criteria. The SSC excluded individuals with nonverbal mental age below 18 months, those with presence of severe neurological deficits, or those with genetic evidence of Fragile X or Down syndromes (Fischbach and Lord, 2010). Future studies may likely find additional subgroups corresponding to ASD populations with other symptoms and/or a broader range of severity. Additionally, the SSC sample was predominantly male and White, so these results may suffer from current bias in ASD evaluation (Durkin et al., 2010) and fail to capture symptoms that characterize more demographically diverse populations. Further, socioeconomic indices of income and education status were significantly lower in children of Subgroup 2 that reflected the highest degree of impairment, including metabolites indicative of increased oxidative stress. We were not able to explicitly evaluate the impacts of socioeconomic factors and stress in this study, but it is an important future consideration for ASD severity and treatment (Kelly et al., 2019). Associations between metabolites and subgroups were based on logistic regression, which may have led to some false negative findings if some metabolites did not meet the assumption for linearity. Further, the enrichment method used required at least two metabolites, which limited pathway-level summary. Lastly, blood draws were non-fasting, which may face confounding by diet. The SSC did not collect comprehensive diet information at the time of blood collection, so the effects of diet on the observed associations with metabolites could not be explored.

Despite some limitations, our analysis revealed novel metabolomic pathways that aid in the understanding of the specific expression of ASD traits.

5. Conclusion

In this study, we interrogated subgroups of individuals with ASD to facilitate improved biological understanding through the metabolome. While previous research has utilized the metabolome to successfully identify metabolomic differences between children with ASD and controls, our analysis is the first, to our knowledge, to interrogate the metabolome between subgroups of only individuals with ASD. Our findings revealed increased levels of membrane lipids and increased oxidative stress may be relevant in children with higher ASD impairment, and elevated lipid metabolism at large may contribute to a higher degree of maladaptive behavioral issues and co-occurring conditions. Higher levels of adenine nucleotides may be relevant to impaired cognition, as this was enriched consistently in subgroups with lower cognitive scores. These findings provide novel information about the underlying biology that gives rise to diverse ASD traits, and these pathways may represent important targets for personalized medicine approaches to alleviate the burden of these symptoms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by a grant from SFARI (674423, awarded to RSK and JALS). We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base as well as access to biospecimens from SSC to perform metabolomic profiling.

Role of the funding source

This project was funded through the Simons Foundation (grant ID 674423, Simons Foundation, United States) awarded to RSK and JALS. Effort from RSK was additionally supported by K01HL146980 from NIH/NHLBI. Effort from NP was supported by NIHT32HL007427 from NIH/NHLBI. The external funders played no role in the design or conduct of the study, collection, management, analysis, or interpretation of the data, preparation, review, approval of, or decision to submit the manuscript.

Abbreviations:

ABC	Aberrant Behavior Checklist
ADHD	Attention-deficit/hyperactivity disorder
ADI-R	Autism Diagnostic Interview, Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism Spectrum Disorder

BMI	body mass index
CSS	calibrated severity score
CBCL	Child Behavior Checklist
DAG	diacylglycero
EDTA	ethylenediaminetetraacetic acid
FDR	false discovery rate
HCER	hexosylceramide
IQ	intelligence quotient
LPL	lysophospholipid
MAG	monoacylglycerol
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PI	phosphatidylinositol
RBS	Repetitive Behavior Scale
RRB	Restricted and Repetitive Behaviors
SFARI	Simons Foundation Autism Research Initiative
SSC	Simons Simplex Collection
SA	Social Affect
SRS	Social Responsiveness Scale
SD	standard deviation
UPLC-MS/MS	Ultrahigh-performance liquid chromatography-tandem mass spectrometry
VABS-II	Vineland Adaptive Behavioral Scale-II

References

Barupal DK, Fiehn O, 2017. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. Sci. Rep. 7, 14567. [PubMed: 29109515]

Ben-David O, Futerman AH, 2010. The role of the ceramide acyl chain length in neurodegeneration: involvement of ceramide synthases. NeuroMol. Med. 12, 341–350.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Stat. Soc.: Ser. B (Methodol.) 57, 289–300.

Braden BB, Pagni BA, Monahan L, Walsh MJM, Dixon MV, Delaney S, Ballard L, Ware JE Jr., 2022. Quality of life in adults with autism spectrum disorder: influence of age, sex, and a controlled,

randomized mindfulness-based stress reduction pilot intervention. Qual. Life Res. 31, 1427–1440. [PubMed: 34655389]

Cao J, Gong J, Li X, Hu Z, Xu Y, Shi H, Li D, Liu G, Jie Y, Hu B, Chong Y, 2021. Unsupervised Hierarchical Clustering Identifies Immune Gene Subtypes in Gastric Cancer. Front. Pharmacol. 12, 692454. [PubMed: 34248641]

Charrad M, Ghazzali N, Boiteau V, Niknafs A, 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J. Stat. Softw. 61.

Chauhan A, Chauhan V, 2006. Oxidative stress in autism. Pathophysiology 13, 171–181. [PubMed: 16766163]

Clements CC, Sparding T, Schultz RT, Yerys BE, Watkins MW, 2020. DAS-II Cognitive Profiles Are Not Diagnostically Meaningful For Autism: A ROC Analysis. Autism Res. 13, 2143–2154. [PubMed: 32696622]

Crouse JJ, Moustafa AA, Bogaty SER, Hickie IB, Hermens DF, 2018. Parcellating cognitive heterogeneity in early psychosis-spectrum illnesses: A cluster analysis. Schizophr. Res. 202, 91–98. [PubMed: 30042029]

de Kloet ER, Karst H, Joels M, 2008. Corticosteroid hormones in the central stress response: quick-and-slow. Front. Neuroendocrinol. 29, 268–272. [PubMed: 18067954]

Durkin MS, Maenner MJ, Meaney FJ, Levy SE, DiGuiseppi C, Nicholas JS, Kirby RS, Pinto-Martin JA, Schieve LA, 2010. Socioeconomic inequality in the prevalence of autism spectrum disorder: evidence from a U.S. cross-sectional study. PLoS One 5, e11551. [PubMed: 20634960]

Eichmann TO, Lass A, 2015. DAG tales: the multiple faces of diacylglycerol–stereochemistry, metabolism, and signaling. Cell. Mol. Life Sci. 72, 3931–3952. [PubMed: 26153463]

Fischbach GD, Lord C, 2010. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron 68, 192–195. [PubMed: 20955926]

Frazier TW, Ratliff KR, Gruber C, Zhang Y, Law PA, Constantino JN, 2014. Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the social responsiveness scale-2. Autism 18, 31–44. [PubMed: 24019124]

Georgiades S, Szatmari P, Boyle M, Hanna S, Duku E, Zwaigenbaum L, Bryson S, Fombonne E, Volden J, Mirenda P, Smith I, Roberts W, Vaillancourt T, Waddell C, Bennett T, Thompson A, Pathways in, A.S.D.S.T., 2013. Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach. J. Child Psychol. Psychiatry 54, 206–215. [PubMed: 22862778]

Geryk J, Krsicka D, Vlckova M, Havlovicova M, Macek M Jr., Kremlikova Pourova R, 2020. The Key Role of Purine Metabolism in the Folate-Dependent Phenotype of Autism Spectrum Disorders: An In Silico Analysis. Metabolites 10. [PubMed: 33375435]

Grabrucker AM, 2012. Environmental factors in autism. Front. Psychiatry 3, 118. [PubMed: 23346059]

Gumpenberger T, Brezina S, Keski-Rahkonen P, Baierl A, Robinot N, Leeb G, Habermann N, Kok DEG, Scalbert A, Ueland PM, Ulrich CM, Gsur A, 2021. Untargeted Metabolomics Reveals Major Differences in the Plasma Metabolome between Colorectal Cancer and Colorectal Adenomas. Metabolites 11. [PubMed: 35050133]

Hassan M, Mokhtar H, 2019. Investigating autism etiology and heterogeneity by decision tree algorithm. Inf. Med. Unlocked 16.

Hirota T, Deserno M, McElroy E, 2020. The Network Structure of Irritability and Aggression in Individuals with Autism Spectrum Disorder. J. Autism Dev. Disord. 50, 1210–1220. [PubMed: 31897854]

Hus V, Gotham K, Lord C, 2014. Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. J. Autism Dev. Disord. 44, 2400–2412. [PubMed: 23143131]

Jun G, Aguilar D, Evans C, Burant CF, Hanis CL, 2020. Metabolomic profiles associated with subtypes of prediabetes among Mexican Americans in Starr County, Texas, USA. Diabetologia 63, 287–295. [PubMed: 31802145]

Kalsner L, Twachtman-Bassett J, Tokarski K, Stanley C, Dumont-Mathieu T, Cotney J, Chamberlain S, 2018. Genetic testing including targeted gene panel in a diverse clinical population of children

with autism spectrum disorder: Findings and implications. Mol. Genet. Genomic Med. 6, 171–185. [PubMed: 29271092]

Kelly RS, Mendez KM, Huang M, Hobbs BD, Clish CB, Gerszten R, Cho MH, Wheelock CE, McGeachie MJ, Chu SH, Celedon JC, Weiss ST, Lasky-Su J, 2022. Metabo-Endotypes of Asthma Reveal Differences in Lung Function: Discovery and Validation in Two TOPMed Cohorts. Am. J. Respir. Crit. Care Med. 205, 288–299. [PubMed: 34767496]

Kelly B, Williams S, Collins S, Mushtaq F, Mon-Williams M, Wright B, Mason D, Wright J, 2019. The association between socioeconomic status and autism diagnosis in the United Kingdom for children aged 5–8 years of age: Findings from the Born in Bradford cohort. Autism 23, 131–140. [PubMed: 29113453]

Levy M, Futerman AH, 2010. Mammalian ceramide synthases. IUBMB Life 62, 347–356. [PubMed: 20222015]

Liang Y, Ke X, Xiao Z, Zhang Y, Chen Y, Li Y, Wang Z, Lin L, Yao P, Lu J, 2020. Untargeted Metabolomic Profiling Using UHPLC-QTOF/MS Reveals Metabolic Alterations Associated with Autism. Biomed Res. Int. 2020, 6105608. [PubMed: 32964039]

Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, Jones EJH, Jones RM, Pickles A, State MW, Taylor JL, Veenstra-VanderWeele J, 2020. Autism spectrum disorder. Nat. Rev. Dis. Primers 6, 5. [PubMed: 31949163]

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, 2022. cluster: Cluster Analysis Basics and Extensions.

Matta J, Zhao J, Ercal G, Obafemi-Ajayi T, 2018. Applications of node-based resilience graph theoretic framework to clustering autism spectrum disorders phenotypes. Appl. Netw. Sci. 3, 38. [PubMed: 30839816]

Mazurek MO, Kanne SM, Wodka EL, 2013. Physical aggression in children and adolescents with autism spectrum disorders. Res. Autism Spectr. Disord. 7, 455–465.

Moncrieff J, Cooper RE, Stockmann T, Amendola S, Hengartner MP, Horowitz MA, 2022. The serotonin theory of depression: a systematic umbrella review of the evidence. Mol. Psychiatry.

Muller CL, Anacker AMJ, Veenstra-VanderWeele J, 2016. The serotonin system in autism spectrum disorder: From biomarker to animal models. Neuroscience 321, 24–41. [PubMed: 26577932]

Narita A, Nagai M, Mizuno S, Ogishima S, Tamiya G, Ueki M, Sakurai R, Makino S, Obara T, Ishikuro M, Yamanaka C, Matsubara H, Kuniyoshi Y, Murakami K, Ueno F, Noda A, Kobayashi T, Kobayashi M, Usuzaki T, Ohseto H, Hozawa A, Kikuya M, Metoki H, Kure S, Kuriyama S, 2020. Clustering by phenotype and genome-wide association study in autism. Transl. Psychiatry 10, 290. [PubMed: 32807774]

Naviaux RK, 2018. Antipurinergic therapy for autism-An in-depth review. Mitochondrion 43, 1–15. [PubMed: 29253638]

Nayar K, Sealock JM, Maltman N, Bush L, Cook EH, Davis LK, Losh M, 2021. Elevated Polygenic Burden for Autism Spectrum Disorder Is Associated With the Broad Autism Phenotype in Mothers of Individuals With Autism Spectrum Disorder. Biol. Psychiatry 89, 476–485. [PubMed: 33229037]

Ohja K, Gozal E, Fahnestock M, Cai L, Cai J, Freedman JH, Switala A, El-Baz A, Barnes GN, 2018. Neuroimmunologic and Neurotrophic Interactions in Autism Spectrum Disorders: Relationship to Neuroinflammation. NeuroMol. Med. 20, 161–173.

Oron O, Elliott E, 2017. Delineating the Common Biological Pathways Perturbed by ASD's Genetic Etiology: Lessons from Network-Based Studies. Int. J. Mol. Sci. 18. [PubMed: 29267212]

Orozco JS, Hertz-Picciotto I, Abbeduto L, Slupsky CM, 2019. Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome. Transl. Psychiatry 9, 243. [PubMed: 31582732]

Page T, Coleman M, 2000. Purine metabolism abnormalities in a hyperuricosuric subclass of autism. BBA 1500, 291–296. [PubMed: 10699370]

Persico AM, Sacco R, 2014. Endophenotypes in Autism Spectrum Disorders. In: Patel VB, Preedy VR, Martin CR (Eds.), Comprehensive Guide to Autism. Springer, New York, New York, NY, pp. 77–95.

Rattray NJW, Deziel NC, Wallach JD, Khan SA, Vasiliou V, Ioannidis JPA, Johnson CH, 2018. Beyond genomics: understanding exposotypes through metabolomics. Hum. Genomics 12, 4. [PubMed: 29373992]

Reardon AM, Li K, Langley J, Hu XP, 2022. Subtyping Autism Spectrum Disorder Via Joint Modeling of Clinical and Connectomic Profiles. Brain Connect. 12, 193–205. [PubMed: 34102874]

Rubin WV, Panzano PC, 2002. Identifying meaningful subgroups of adults with severe mental illness. Psychiatr. Serv. 53, 452–457. [PubMed: 11919359]

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, Goldberg AP, Jinlu C, Keaney JF 3rd, Klei L, Mandell JD, Moreno-De-Luca D, Poultney CS, Robinson EB, Smith L, Solli-Nowlan T, Su MY, Teran NA, Walker MF, Werling DM, Beaudet AL, Cantor RM, Fombonne E, Geschwind DH, Grice DE, Lord C, Lowe JK, Mane SM, Martin DM, Morrow EM, Talkowski ME, Sutcliffe JS, Walsh CA, Yu TW, Autism Sequencing C, Ledbetter DH, Martin CL, Cook EH, Buxbaum JD, Daly MJ, Devlin B, Roeder K, State MW, 2015. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215–1233. [PubMed: 26402605]

Skaar DA, Shao Y, Haines JL, Stenger JE, Jaworski J, Martin ER, DeLong GR, Moore JH, McCauley JL, Sutcliffe JS, Ashley-Koch AE, Cuccaro ML, Folstein SE, Gilbert JR, Pericak-Vance MA, 2005. Analysis of the RELN gene as a genetic risk factor for autism. Mol. Psychiatry 10, 563–571. [PubMed: 15558079]

Smith AM, King JJ, West PR, Ludwig MA, Donley ELR, Burrier RE, Amaral DG, 2019. Amino Acid Dysregulation Metabotypes: Potential Biomarkers for Diagnosis and Individualized Treatment for Subtypes of Autism Spectrum Disorder. Biol. Psychiatry 85, 345–354. [PubMed: 30446206]

Smith AM, Natowicz MR, Braas D, Ludwig MA, Ney DM, Donley ELR, Burrier RE, Amaral DG, 2020. A Metabolomics Approach to Screening for Autism Risk in the Children's Autism Metabolome Project. Autism Res. 13, 1270–1285. [PubMed: 32558271]

Sullivan MO, Gallagher L, Heron EA, 2019. Gaining Insights into Aggressive Behaviour in Autism Spectrum Disorder Using Latent Profile Analysis. J. Autism Dev. Disord. 49, 4209–4218. [PubMed: 31292900]

Tamiji J, Crawford DA, 2010. The neurobiology of lipid metabolism in autism spectrum disorders. Neurosignals 18, 98–112. [PubMed: 21346377]

Team, R.C., 2020. R: A language and environment for statistical computing. R Foundation or Statistical Computing, Vienna, Austria.

Thorndike RL, 1953. Who belongs in the family? Psychometrika 18, 267–276.

Tian YJ, Li D, Ma Q, Gu XY, Guo M, Lun YZ, Sun WP, Wang XY, Cao Y, Zhou SS, 2013. Excess nicotinamide increases plasma serotonin and histamine levels. Sheng Li Xue Bao 65, 33–38. [PubMed: 23426511]

Villalobos MA, De La Cruz JP, Cuerda MA, Ortiz P, Smith-Agreda JM, Sanchez De La Cuesta F, 2000. Effect of S-adenosyl-L-methionine on rat brain oxidative stress damage in a combined model of permanent focal ischemia and global ischemia-reperfusion. Brain Res. 883, 31–40. [PubMed: 11063985]

Walker DI, Valvi D, Rothman N, Lan Q, Miller GW, Jones DP, 2019. The metabolome: A key measure for exposome research in epidemiology. Curr Epidemiol Rep 6, 93–103. [PubMed: 31828002]

Wang H, Liang S, Wang M, Gao J, Sun C, Wang J, Xia W, Wu S, Sumner SJ, Zhang F, Sun C, Wu L, 2016. Potential serum biomarkers from a metabolomics study of autism. J. Psychiatry Neurosci. 41, 27–37. [PubMed: 26395811]

Whitfield JB, 2001. Gamma glutamyl transferase. Crit. Rev. Clin. Lab. Sci. 38, 263–355. [PubMed: 11563810]

Zaghlool SB, Halama A, Stephan N, Thangam M, Ahlqvist E, Albagha OME, AbouSamra AB, Suhre K, 2022. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. medRxiv, 2022.2001.2013.22269204.

Zeidan J, Fombonne E, Scorah J, Ibrahim A, Durkin MS, Saxena S, Yusuf A, Shih A, Elsabbagh M, 2022. Global prevalence of autism: A systematic review update. Autism Res. 15, 778–790. [PubMed: 35238171]

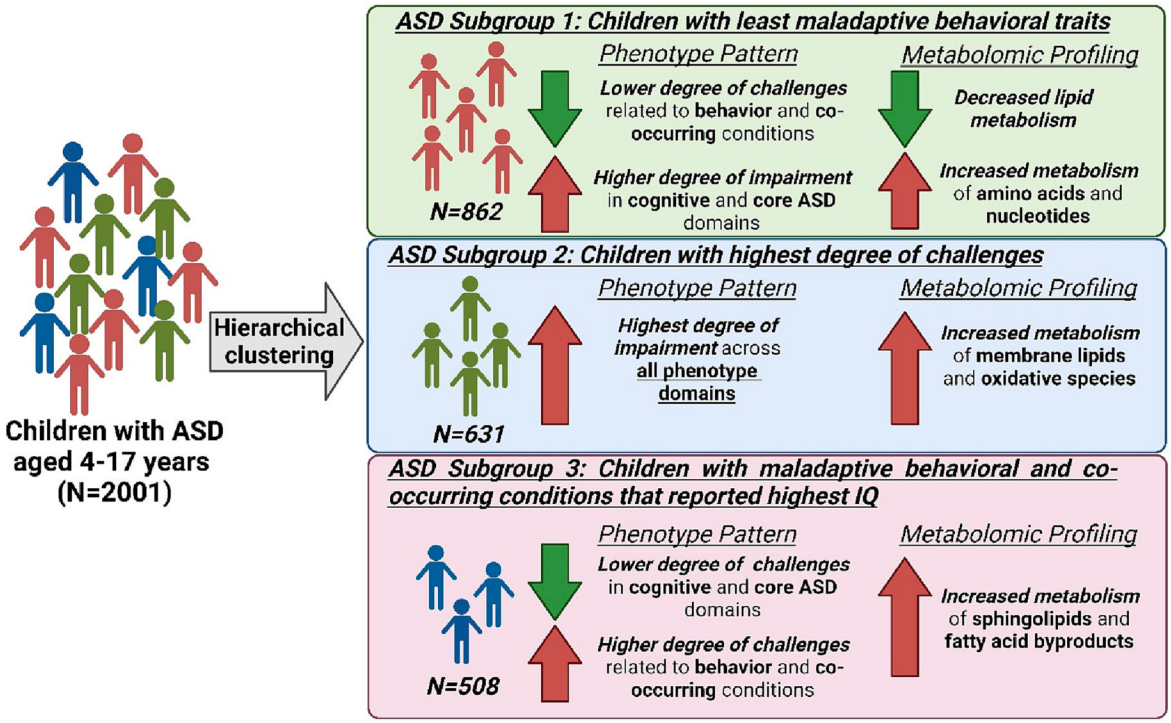

Fig. 1.

Summary of phenotype and metabolic pathway patterns between the three ASD subgroups identified in this study. A total of 2001 individuals with ASD from the SSC cohort were included and split into subgroups via unsupervised hierarchical clustering of 40 diverse ASD characteristics and traits. Assessment of phenotype patterns revealed Subgroup 1 children showed the least maladaptive behaviors of any subgroup, Subgroup 2 children had the highest degree of impairment in all domains, and Subgroup 3 children demonstrated the highest IQ scores but still reported the second highest maladaptive behaviors and co-occurring conditions. These were associated with metabolic pathway disruptions to lipids, amino acids, and nucleotides in each subgroup.

Table 1

Means and standard deviations (SD) of 40 phenotype scores utilized in clustering. Means and SDs are shown prior to standardization for clustering. Adjusted phenotypes and additional information are denoted with superscripts and footnotes.

Domain	Phenotype	Mean (SD)
Core ASD Trait measures	ADOS Total CSS	7.44 (1.71)
	ADOS Social affect CSS[1]	7.24 (1.76)
	ADOS Restricted/repetitive behaviors CSS[1]	7.68 (2.06)
	ADI-R A Social score	19.75 (5.63)
	ADI-R B Verbal score	8.87 (3.42)
	ADI-R B Non-verbal score	16.52 (4.29)
	ADI-R C Restricted/stereotyped behavior score	6.67 (2.56)
	ADI-R D Abnormality score	3.70 (1.10)
	ADI-R q86 Age at first abnormality	3.63 (0.56)
	VABS-II Social standard score	73.00 (11.55)
	SRS Parent Total t-score[2]	79.02 (10.47)
	SRS Parent Communication raw score[3]	32.79 (9.77)
	SRS Parent Awareness raw score[3]	12.29 (3.50)
	SRS Parent Mannerisms raw score[3]	18.49 (6.76)
	ABC Stereotypy	4.60 (2.08)
	ABC Lethargy	9.30 (6.87)
	ABC Inappropriate speech	3.78 (2.94)
	RBS Compulsive behaviors	4.00 (3.90)
	RBS Self-injury	1.93 (2.75)
	RBS Stereotypy	4.20 (3.17)
	RBS Ritualistic behaviors	5.24 (3.98)
	RBS Restricted behaviors	3.65 (2.80)
	RBS Sameness behaviors	7.73 (5.90)
Cognitive & Adaptive Functioning scores	Verbal IQ score	84.00 (26.11)
	Non-Verbal IQ score	89.27 (22.52)
	VABS-II Daily living skills score	78.62 (12.75)
Language & Communication scores	VABS-II Communication standard score	79.92 (12.70)
	ADI-R Language delay	4.35 (2.59)
Maladaptive behaviors and Co-occurring conditions	ABC Irritability	11.34 (8.59)
	ABC Hyperactivity	16.45 (10.33)
	ABC Depressed mood[4]	0.32 (0.63)
	CBCL Internalizing T-score	60.47 (9.54)
	CBCL Externalizing T-score	56.66 (10.74)
	CBCL Attention problems T-score	61.41 (8.16)
	CBCL Affective problems T-score	60.91 (9.20)
	CBCL Aggressive behavior T-score	61.46 (8.67)

Domain	Phenotype	Mean (SD)
	CBCL Anxiety problems T-score	59.49 (9.47)
	RBS Sleeping problems[5]	0.99 (1.02)
	Febrile seizures[6]	0.06 (0.31)
	Non-febrile seizures[7]	0.15 (0.54)

Abbreviations: ABC, Aberrant Behavior Checklist; ADOS, Autism Diagnostic Observation Schedule; ADI-R, Autism Diagnostic Interview-Revised; CBCL, Child Behavior Checklist; CSS, Calibrated Severity Score; IQ, Intelligence quotient; RBS, Repetitive Behavior Scale; SRS, Social Responsiveness Scale; VABS-II Vineland Adaptive Behavior Scale-II.

[1] Calibrated severity scores for ADOS social affect and restricted/repetitive behaviors were derived using methods from Hus et al. (2014).

[2] SRS parent Total t-score residuals were used in clustering after regressing out the effect of age.

[3] SRS parent Communication, Awareness, and Mannerisms raw score residuals were used in clustering after regressing out age and sex.

[4] Depressed mood scores were derived from the ABC questionnaire and categorized on a scale of 0–3 as: never a problem (0), slight problem (1), moderately serious problem (2), or severe problem (3).

[5] Sleep problems scores were derived from the RBS questionnaire and categorized on a scale of 0–3 as: behavior does not occur (0), behavior occurs and is a mild problem (1), behavior occurs and is a moderate problem (2), or behavior occurs and is a severe problem (3).

[6] Febrile seizure scores were coded on a scale of 0–2 as: no evidence for presence of febrile seizures (0), possible presence of febrile seizures (1), or reported febrile seizures (2).

[7] Non-febrile seizures scores were coded on a scale of 0–3 as: no evidence for presence of non-febrile seizures (0), possible experience of non-febrile seizures (1), likely presence of non-febrile seizures (2), or epilepsy reported (3).

Table 2

Trends in phenotype scores between ASD subgroups. Means and standard deviations for phenotype scores are provided for each subgroup. For each phenotype, relative trends in scores are color-coded, with darker blue indicating higher degree of impairment and light blue indicating lower degree of impairment. Any phenotypes that did not significantly differ at an FDR-corrected P-value < 0.05 are noted with superscript and footnote and are indicated with only include two shades of blue, with the non-significantly different groups colored by the same shade.

Domain	Phenotype	Subgroup 1 Children with least maladaptive behavioral traits (N=862)	Subgroup 2 Children with highest degree of challenges (N=631)	Subgroup 3 Children with maladaptive behaviors and co-occurring conditions that reported high IQ (N=508)
Core ASD Trait Measures	ADOS Total CSS	7.46 (1.66)	7.92 (1.69)	6.82 (1.61)
	ADOS Social affect CSS	7.25 (1.71)	7.66 (1.68)	6.70 (1.80)
	ADOS Restricted/repetitlve behaviors CSS	7.67 (2.01)	8.03 (2.01)	7.26 (2.13)
	ADI-R A Social score	19.60 (5.65)	22.62 (4.76)	16.45 (4.62)
	ADI-R B Verbal score	9.09 (3.35)	10.43 (3.01)	6.55 (2.72)
	ADI-R B Non-verbal score	16.64 (4.20)	18.63 (3.81)	13.69 (3.32)
	ADI-R C Restricted/stereotyped behavior score[1]	6.13 (2.52)	7.72 (2.34)	6.28 (2.49)
	ADI-R D Abnormality score[2]	3.85 (1.05)	3.82 (1.04)	3.30 (1.17)
	ADI-R q86 Age at first abnormality	3.67 (0.50)	3.77 (0.45)	3.36 (0.66)
	VABS-II Social standard score[1]	75.82 (11.82)	66.83 (10.05)	75.89 (9.67)
	SRS Parent Total t-score	72.54 (10.72)	87.17 (4.64)	79.91 (7.45)
	SRS Communication score	27.35 (9.02)	40.39 (7.08)	32.57 (7.33)
	SRS Awareness raw score	10.75 (3.26)	14.48 (3.10)	12.21 (2.90)
	SRS Mannerisms raw score	14.58 (6.03)	23.82 (5.20)	18.52 (4.83)
	ABC Stereotypy	3.02 (2.98)	7.39 (4.42)	3.81 (3.44)
	ABC Lethargy	6.10 (4.92)	14.22 (7.32)	8.64 (5.53)
	ABC Inappropriate speech	2.66 (2.41)	5.60 (3.04)	3.43 (2.54)
	RBS Compulsive behaviors	2.48 (2.66)	6.44 (4.63)	3.57 (3.11)
	RBS Self-injury	0.97 (1.57)	3.26 (3.57)	1.89 (2.45)
	RBS Stereotypy	2.99 (2.35)	6.28 (3.49)	3.67 (2.65)
	RBS Ritualistic behaviors	3.46 (2.92)	7.72 (4.33)	5.17 (3.35)

Domain	Phenotype	Subgroup 1 Children with least maladaptive behavioral traits (N=862)	Subgroup 2 Children with highest degree of challenges (N=631)	Subgroup 3 Children with maladaptive behaviors and co-occurring conditions that reported high IQ (N=508)
	RBS Restricted behaviors	2.48 (2.17)	5.49 (3.03)	3.33 (2.23)
	RBS Sameness behaviors	4.67 (3.94)	11.90 (6.44)	7.75 (4.68)
Cognitive & Adaptive Functioning Scores	Verbal IQ score	82.48 (25.20)	77.38 (27.90)	94.81 (21.65)
	Non-Verbal IQ score	89.82 (22.09)	82.24 (23.82)	97.06 (18.57)
	VABS-II Dally living skills score[1]	80.37 (12.97)	73.85 (12.24)	81.59 (11.26)
Language & Communication Scores	VABS-II Communication standard score	81.23 (13.08)	75.10 (11.23)	83.68 (11.98)
	ADI-R Language delay	4.55 (2.51)	5.42 (2.41)	2.67 (2.03)
Maladaptive Behaviors & Co-occurring Conditions	ABC Irritability	6.08 (5.66)	17.40 (8.63)	12.74 (7.06)
	ABC Hyperactivity	10.80 (8.13)	23.42 (9.42)	17.41 (9.12)
	ABC Depressed mood	0.11 (0.36)	0.56 (0.83)	0.37 (0.59)
	CBCL Internalizing t-score	54.94 (8.56)	65.53 (8.40)	63.57 (7.33)
	CBCL Externalizing t-score	50.00 (8.95)	62.80 (9.42)	60.34 (8.56)
	CBCL Attention problems t-score	57.27 (6.60)	65.80 (7.85)	62.99 (7.54)
	CBCL Affective problems t-score	56.94 (6.86)	66.20 (8.70)	63.26 (7.51)
	CBCL Aggressive behavior t-score	54.07 (5.86)	64.90 (10.32)	61.96 (8.42)
	CBCL Anxietv problems t-score	56.77 (7.57)	65.13 (9.26)	62.70 (8.68)
	RBS Sleeping problems	0.66 (0.85)	1.45 (1.10)	0.98 (0.97)
	Febrile seizures[3]	0.12 (0.45)	0.02 (0.12)	0.00 (0.00)
	Non-febrile seizures[3]	0.27 (0.75)	0.08 (0.32)	0.03 (0.17)

[1]No significant difference between Subgroups 1 and 3 at FDR < 0.05 if denoted with this superscript.

[2]No significant difference between Subgroups 1 and 2 at FDR < 0.05 if denoted with this superscript.

[3]No significant difference between Subgroups 2 and 3 at FDR < 0.05 if denoted with this superscript.

Table 3

Enriched biological pathways in ASD subgroups. Enrichment P-values and the proportion of increased metabolites are shown for each pathway. Heatmap-style coloring is utilized to show degree of increase, with red indicating 100% of metabolites showed increased levels in a particular subgroup, and green indicating 0% of metabolites showed increased levels in a particular subgroup. Shades of light green, yellow, and orange indicate degree of change between green and red, in agreement with ratio columns.

Pathway Name	Type	Subgroup 1 (N=862) *Children with least maladaptive behavioral traits*		Subgroup 2 (N=631) *Children with highest degree of challenges*		Subgroup 3 *Children with maladaptive behavioral and co-occurring condition that reported, highest IQ*	
		Pathway P-value[1]	Increased metabolites/ total metabolites (ratio)[2]	Pathway P-value[1]	Increased metabolites/total metabolites (ratio)[2]	Pathway P-value[1]	Increased metabolites/total metabolites (ratio)[2]
Monoacylglycerol	Lipid	3.20×10^{-6}	0/6 (0)	5.40×10^{-6}	4/4 (1)		
Diacylglycérol	Lipid	3.20×10^{-5}	0/6 (0)	1.50×10^{-5}	5/5 (1)		
Phosphatidylinositol (pi)	Lipid	2.20×10^{-4}	0/2 (0)	6.00×10^{-5}	2/2 (1)		
Phosphatidylethanol amine (PE)	Lipid	3.90×10^{-3}	0/4 (0)	8.00×10^{-4}	3/3 (1)		
Gamma-glutamyl Amino Acid	Peptide			0.015	2/2 (1)		
Corticosteroids	Lipid			0.038	2/2 (1)		
Lysophospholipid	Lipid	5.80×10^{-3}	0/6 (0)	3.40×10^{-3}	2/3 (0.67)		
Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid			3.50×10^{-3}	1/2 (0.5)		
Purine Metabolism, Adenine containing	Nucleotide	3.50×10^{-3}	2/8 (0.25)	0.044	1/2 (0.5)		
Hemoglobin and Porphyrin Metabolism	Cofactor			0.043	0/3 (1)		
Lysoplasmalogen	Lipid	9.80×10^{-3}	2/2 (1)				
Dipeptide	Peptide	2.80×10^{-8}	9/9 (1)			5.50×10^{-6}	0/3 (0)
Sphingolipid Synthesis	Lipid	0.015	2/2 (1)			1.60×10^{-4}	1/3 (0.33)
Purine Metabolism, (Hypo)Xanthine/ lnos ine containing	Nucleotide	0.016	2/2 (1)				
Glycine, Serine and Threonine Metabolism	Amino Acid	0.017	3/3 (1)				
Sphingosines	Lipid	1.40×10^{-4}	2/3 (0.67)			5.40×10^{-4}	1/3 (0.33)
Urea cycle; Arginine and Proline Metabolism	Amino Acid	0.033	4/6 (0.67)				

Pathway Name	Type	Subgroup 1 (N=862) *Children with least maladaptive behavioral traits*		Subgroup 2 (N=631) *Children with highest degree of challenges*		Subgroup 3 *Children with maladaptive behavioral and co-occurring condition that reported, highest IQ*	
		Pathway P-value[1]	Increased metabolites/ total metabolites (ratio)[2]	Pathway P-value[1]	Increased metabolites/total metabolites (ratio)[2]	Pathway P-value[1]	Increased metabolites/total metabolites (ratio)[2]
Fatty Acid, Dicarboxylate	Lipid	2.00×10^{-6}	0/13 (0)			1.30×10^{-3}	8/8 (1)
Phosphatidylcholine (PC)	Lipid	3.00×10^{-4}	0/6 (0)				
Fatty Acid, Monohydroxy	Lipid	3.00×10^{-4}	0/6 (0)				
Hexosylceramide (HCER)	Lipid	3.40×10^{-4}	0/4 (0)				
Sphingomyelins	Lipid	0.019	0/2 (0)			2.80×10^{-4}	7/7 (1)
Ceramides	Lipid					1.80×10^{-3}	3/3 (1)
Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	Lipid					0.036	2/2 (1)

[1] Pathway P-values were generated by the ChemRICH package, as described in Barupal and Fiehn (2017).

[2] The increased ratio describes the number of increased metabolites over the total number of metabolites significantly associated with that subgroup.