bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Plasma and Fecal Metabolite Profiles in Autism Spectrum Disorder

Brittany D. Needham[1], Mark D. Adame[1], Gloria Serena[2], Destanie R. Rose[3,4], Gregory M. Preston[5], Mary C. Conrad[5], A. Stewart Campbell[5], David H. Donabedian[5], Alessio Fasano[2], Paul Ashwood[3,4], and Sarkis K. Mazmanian[1]

[1]Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125, USA
[2]Division of Pediatric Gastroenterology and Nutrition, Mucosal Immunology and Biology Research Center, Massachusetts General Hospital for Children, Boston, MA, 02114, USA
[3]Department of Medical Microbiology and Immunology, University of California Davis, Davis, CA, 95616, USA
[4]The M.I.N.D. Institute, University of California, Davis, Sacramento, CA, 95817, USA
[5]Axial Biotherapeutics, Waltham, MA, 02451, USA
Correspondence: bneedham@caltech.edu and sarkis@caltech.edu

Running title: Plasma and Fecal Metabolite Profiles in ASD

Keywords: Autism Spectrum Disorder, ASD, metabolomics, plasma metabolites, fecal metabolites, steroid hormones, mitochondrial dysfunction, phenolic metabolites

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition with hallmark behavioral manifestations including impaired social communication and restricted repetitive behavior. In addition, many affected individuals display metabolic imbalances, immune dysregulation, gastrointestinal (GI) dysfunction, and altered gut microbiome compositions. We sought to better understand non-behavioral features of ASD by determining molecular signatures in peripheral tissues. Herein, we present the untargeted metabolome of 231 plasma and 97 fecal samples from a large cohort of children with ASD and typically developing (TD) controls. Differences in lipid, amino acid, and xenobiotic metabolism discriminate ASD and TD samples. We reveal correlations between specific metabolite profiles and clinical behavior scores, and identify metabolites particularly associated with GI dysfunction in ASD. These findings support a connection between GI physiology, metabolism, and complex behavioral traits, and may advance discovery and development of molecular biomarkers for ASD.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Many diseases are associated with informative metabolic signatures, or biomarkers, that enable diagnoses, predict disease course, and guide treatment strategies. In contrast, autism spectrum disorder (ASD) is diagnosed based on observational evaluation of behavioral symptoms, including reduced social interaction and repetitive/stereotyped behaviors (*1*). The average age of ASD diagnosis is between 3-4 years old (*2*), at which time children can receive behavioral therapy, the gold standard treatment. Because earlier diagnosis improves efficacy of behavioral therapies (*3*, *4*), molecular biomarkers represent an attractive approach for identifying 'at-risk' populations and may aid development of personalized therapies. This prospect is increasingly important given the rising rate of ASD diagnoses, which currently stands at up to 1 in 59 children in the United States (*2*), with no FDA approved drugs for core behavioral symptoms.

Metabolic abnormalities have been reported in ASD (*5*), though most have measured only a small subset of metabolites and many outcomes have not reproduced between cohorts. Mitochondrial disease, which heavily influences systemic metabolism, is estimated to be higher in ASD compared to controls (5% vs ~0.01%) (*6*), and genes crucial for mitochondrial function are risk factors for ASD in humans and rodent models (*7*). The metabolic abnormalities associated with mitochondrial dysfunction in ASD affect cellular energy, oxidative stress, and neurotransmission in the gut and the brain (*7–19*). Other metabolic profiles in ASD implicate aromatic and phenolic metabolites, including derivatives of nicotinic, amino acid, and hippurate metabolism (*20–30*). Various amino acids are detected at differential levels across studies and across sample types, but any consistent patterns are difficult to discern (*15*, *24*, *26*, *29*, *31–34*).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Some of the discrepancies between these studies are likely due to differences in sample number, tissue analyzed, and methodology. Other sources of variability include differences in environmental factors, such as diet and gut bacteria, which differ between ASD and TD populations (*35–37*). Diet is a major source of circulating metabolites, impacting the metabolome directly or indirectly through chemical transformation by the trillions of gut microbes, the microbiome, which has been proposed to modulate complex behaviors in animal models and humans (*38–40*). Such proposed environmental modulators of ASD may integrate with genetic risks to impact behavioral endpoints through the actions of small molecules produced in peripheral tissues outside the brain.

Herein, we present a comprehensive comparison of an extensive panel of identified metabolites in human plasma and feces from a large cohort of matched ASD and TD children. We identified differential levels of metabolites ranging from hormones, amino acids, xenobiotics, and lipids, many of which correlate with clinical behavior and GI scores. To our knowledge, this is the first study to concurrently evaluate paired intestinal and systemic metabolomes in a high-powered analysis with a large number of identified metabolites, allowing direct associations between metabolites previously highlighted in ASD samples and discovering new metabolites of interest. These findings support the emerging concept of evaluating non-behavioral features in the diagnosis of ASD and its GI comorbidities.

METHODS AND MATERIALS

Participants: Samples for this study, aged 3-12 years old, were collected through the UC Davis MIND institute (*41*, *42*). ASD diagnosis was confirmed at the MIND Institute by trained staff

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using the Autism Diagnostic Observation Schedule (ADOS), and the Autism Diagnostic Interview-Revised (ADI-R). Subjects were diagnosed prior to 2013 based on DSM IV. Typically developing (TD) participants were screened using the Social Communication Questionnaire (SCQ). Participants in the TD group scored within the typical range, i.e. below 15, on the SCQ and above 70 on the Mullen Scales of Early Learning (MSEL) and Vineland Adaptive Behavior Score (VABS). Ninety-seven of the participants, who also provided the stool samples, completed an additional evaluation to determine gastrointestinal (GI) symptoms. GI status was determined using the GI symptom survey (GISS), based upon Rome III Diagnostic Questionnaire for the Pediatric Functional GI Disorders (*43*) and described in detail elsewhere (*42*).

See Supplemental materials for additional methods.

RESULTS

Plasma and Fecal Metabolomes Differ between ASD and TD

Plasma samples from 130 ASD and 101 TD children were analyzed along with fecal samples collected from a subset of these same ASD (n=57) and TD individuals (n=40) (Figures S1A-S1G). Samples and metrics of behavioral and GI scores were obtained from the UC Davis MIND institute (*41*, *42*). The ASD group was stratified into subsets of children with GI symptoms (ASD+GI) or without GI symptoms (ASD-GI), to explore potential effects of comorbid intestinal dysfunction in ASD (40 out of 130 ASD samples were +GI). This stratification was based on symptoms associated with ASD including diarrhea, constipation, and irritable bowel syndrome-like symptoms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Samples were analyzed by the global metabolite panel (plasma and fecal samples) and the complex lipid panel (plasma samples) by Metabolon, Inc. (Durham, NC), which identified a total panel of 1,611 plasma and 814 fecal metabolites (Tables S1-S2). Overall, we discovered that 310 metabolites were differentially abundant between the ASD and TD groups in plasma and 112 metabolites were differentially abundant in feces (Figures 1A-1B, S1C, and Tables S1-S2). Using a quantitative assay for a targeted panel of metabolites, we observed a high correlation between relative abundance and precise concentration (Figure S2A). Overall, these data expand on previous evidence that the metabolomic profile of ASD and TD populations display differences not only in the gut compartment, but also in circulation, which may affect the levels of metabolites throughout the body, including the brain (*44*, *45*).

To appreciate the biological relevance of different metabolomes between ASD and TD, individual metabolites were integrated into biochemical pathways for pathway enrichment analysis, revealing the degree of change within each. Here, we identified large scale changes, mostly in lipid, xenobiotic, and nucleotide pathways associated with diverse physiological processes (Figures 1C-1D, S2B). We then used Random Forest machine learning analysis to determine if metabolite profiles can unbiasedly predict whether the sample came from an ASD and TD donor. Overall, the modest predictive accuracy of this machine learning approach was 68% for plasma and 67% for feces. To test whether focusing on the most discriminating metabolites would improve the prediction, we repeated the Random Forest analysis using the top 30 metabolites and found that the predictive accuracy for plasma improved slightly to 69% when using all ASD samples, and to 73% when using only ASD-GI samples. For fecal samples, the predictive accuracy improved to 75% using all ASD and to 73% using only ASD-GI. The top 30

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

metabolites, calculated by measuring the mean decrease in accuracy of the machine learning algorithm, are useful to describe the strongest drivers of overall metabolic differences between ASD and TD populations. These most discriminatory metabolites were primarily from the lipid, amino acid, xenobiotic, and cofactor/vitamin super pathways (Figures 1E-1F). Several of these metabolites have been previously linked to ASD, such as steroids, bile acids, acyl-carnitines, and nicotinamide metabolites (*46–49*). Further, multiple molecules known to be produced or manipulated by the gut microbiota also featured prominently, including 4-ethylphenyl sulfate (4EPS), which is elevated in an ASD mouse model, and indolelactate, a microbe-derived tryptophan metabolite (Figure 1E) (*50*, *51*). The two most discriminatory molecules in plasma (Figures 1G-1H) and feces (Figures 1I-1J) are depicted. Metabolites correlating the strongest with these discriminatory metabolites are closely related on a structural and metabolic level (Figures 1K-1L).

Global Metabolite Levels Correlate with Clinical Behavioral Scores

Using clinical metadata for ASD individuals, we correlated the levels of individual metabolites to the verbal, social, and nonverbal scores of standard diagnostic measures: the Autism Diagnostic Interview, Revised (ADIR), a parent questionnaire, and the cumulative Autism Diagnostic Observation Schedule severity score (ADOS-SS), conducted by trained health professionals (*1*) (Tables S1-S2). Next, grouped correlations between the behavior metrics and entire metabolite pathways were calculated (Figure 1M). We found that verbal and social scores primarily correlate with lipid metabolism pathways and that nonverbal scores have the fewest correlations. The ADOS-SS correlated with diverse metabolite pathways, including amino acids and food/plant component pathways, which may be partly due to the diverse array of symptoms

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

integrated into the ADOS-SS score. Additionally, when stratified by severity of behavioral scores, we observe modest clustering according to plasma metabolites correlated to the disorder between the top and bottom quartiles of behavior severity groups for verbal, social, and ADOSS-SS metrics, but not for nonverbal, and less so with fecal samples (Figures 1N, S1H-S1I). These correlations support the involvement of lipid, amino acid, and xenobiotic metabolism in the etiology of ASD, as previously described (*8*, *48*, *52*), and reveal new candidates for ASD biomarkers that correlate with symptom severity.

Transfer of ASD Fecal Microbiota into Mice is Accompanied by Metabolic Signatures

Since microbial metabolites ranked highly in the Random Forest machine learning analysis, we tested if any of the observed metabolite differences in humans could be transferred to mice via fecal microbial transplant. We selected 4 male donor samples from each of the ASD and TD groups and colonized 2-3 male germ-free mice per donor for three weeks before collecting plasma and fecal samples for metabolite profiling and bacterial DNA sequencing (Figure S2C-S2D), respectively. Global metabolomic analysis revealed that colonized mice modestly cluster by donor and group when statistically significant metabolites are considered in PCA analysis (Figure S2E). We selected the human donors based on 4EPS levels (Figure S2F), due to its involvement in an ASD mouse model (*50*) and dysregulation of similar phenolic compounds in human ASD (*22*, *24*, *28*, *29*, *53*). 4EPS is not produced by the host and is strictly a bacterial metabolite (*50*, *54*). Surprisingly, we observed 4EPS levels in a bimodal distribution in mouse samples (Figure S2G-S2H). In spite of the surprising results with 4EPS, many of the metabolites with the highest fold change and lowest *p*-value are indeed other phenolic molecules such as metabolites of hippurate, tyrosine, and diet-derived phenols (Figures S2I, Table S5). While

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

preliminary, these studies reveal microbiome-mediated effects on xenobiotic pathways and phenolic metabolites dysregulated in ASD.

Lipid and Xenobiotic Metabolites Correlate with GI Symptoms

Considering the prevalence of intestinal issues in ASD, GI-dependent metabolite analysis between individuals may provide useful insights. Accordingly, we found a total of 87 and 24 metabolites in plasma and feces, respectively, that were differentially abundant within the ASD+GI compared to ASD-GI individuals (Figures 2A-2B). At the pathway level, we found differences in free fatty acid and xenobiotic metabolites in the food component and plant pathways between ASD+GI vs ASD-GI plasma (Figure 2C), with no broad GI-dependent pathway alterations in fecal samples (Table S2). In plasma, free fatty acids of multiple chain lengths were lower in ASD+GI compared to ASD-GI samples, including monounsaturated, saturated, and polyunsaturated fatty acids (PUFAs) (Figure 2D). PUFAs are anti-inflammatory and lower levels of these fatty acids may contribute to GI dysfunction directly (*12*).

Several metabolites from the food component and plant pathway also discriminate ASD+GI from ASD-GI, such as lower circulating piperine metabolites in ASD+GI plasma samples (Figures 2E-2G). A lower level of piperine metabolites was also associated with higher ADOS-SS (Figure 2H) and worse nonverbal scores (Table S1). Observed alterations to piperine metabolite levels are supported by the fact that oral administration of piperine has been successfully used as a treatment in preclinical ASD models, presumably due to its antioxidant quality (*55*). While these correlations are interesting, the implications of correlating an altered metabolome and GI symptoms in ASD remain to be determined.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Steroid Hormone Levels are Elevated in ASD

Multiple human ASD studies have examined the levels of specific steroid metabolites within androgenic, pregnenolone and progesterone metabolism, with some finding aberrant levels, positive correlations with ASD severity, and behavioral improvement following treatments that lower levels of certain hormones (*47*, *56–64*). On the other hand, in a recent clinical trial of ASD children given an antioxidant treatment, levels of pregnenolones and androgens increased and correlated with improved behavior (*65*). In our dataset, we found robust increases of almost every detected metabolite within the pregnenolone, androgen, progesterone, and corticosteroid pathways in the plasma of ASD children (Figures 1C, 3A, Table S1). Similarly, some androgenic steroid pathway metabolites were elevated in ASD fecal samples (Figures 3A, Table S2). This is a strong indicator that the physiological pathways associated with the downstream metabolism of cholesterol are significantly altered between ASD and TD populations (Figure 3B). There does not appear to be a global change in steroid metabolism, as most primary bile acid and sterol metabolites were unaffected (Tables S1-S2). We observed some elevation of these hormone levels independent of sex, which is notable considering the male bias in ASD, reflected in our primarily male sample set (7-15% female) (*66*) (Figure S3A-S3B). Because a cluster of our samples are from older individuals in the ASD group, and to account for age-dependent increases in androgens, we stratified by age and still observed heightened androgenic and pregnenolone metabolite levels in ASD subpopulations (Figure S3C). Taken together, these data indicate that steroidal hormone metabolism may be altered in the ASD population relative to TD samples and that these differences are not driven solely by sex or age differences in our cohort.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Lipid Metabolite Levels Differ in ASD

Lipids are crucial for energy storage, cellular membrane integrity and cell signaling. They play a variety of roles in the central nervous system, and their dysregulation has been linked to ASD (*9*, *11*). Phospholipids have been measured at lower levels, while long chain fatty acids are reportedly elevated but PUFAs have been measured at both higher and lower levels, depending on the cohort (*9*, *12*, *13*). Here, we performed an untargeted, quantitative metabolite analysis on complex lipids. The concentrations of 999 lipids were quantified, and 13.7% of these were significantly different in the ASD samples, with 4.6% increased and 9.1% decreased compared to TD samples. Many of these differentially abundant lipids included phospholipids, cholesterol esters and glycerolipids. In general, shorter (14-18 chain length) saturated fatty acids were less abundant in ASD throughout the lipid classes (Figures 3C, S4A-S4B).

PUFA lipid levels were also different, with elevations in diacylglycerols and free fatty acids, and an enrichment of the 18:2 (linolenic) chain length in most lipid classes (Figure 3C). Fecal samples generally trended in the opposite direction from plasma in PUFA lipids (Figure S4C, Table S2). Intriguingly, multiple lipids with linolenic and linoleic (18:3) chains were correlated with social behavior (Tables S1 and S2). PUFA lipids containing linolenic and linoleic acids are precursors to the important PUFAs (arachidonic 20:4 and docosahexaenoic 22:6 acids) for brain development, function, and structural integrity (*12*). Future studies are needed to determine if these specific changes in lipid levels contribute to ASD symptoms.

ASD Correlates with Cellular Energy and Oxidative Stress Metabolites

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Many lipids are markers of mitochondrial and oxidative stress, offering a snapshot into cellular metabolic states. These markers include acyl-carnitines, which have been highlighted in various ASD studies and are established indicators of mitochondrial dysfunction (*11*, *12*, *14–18*, *24*, *25*, *33*, *46*, *58*, *67–71*). Acyl-carnitines are formed to allow transport of lipids across the mitochondrial membranes for beta-oxidation, and abnormal levels of these conjugated lipids accumulate to higher levels with a decrease in beta-oxidation. Interestingly, high levels of short acyl-carnitines are found in rodent models where ASD-like behaviors are induced with the short chain fatty acids, valproic and propionic acids (*72*).

We found differential levels of various acyl-carnitines in ASD, creating a pattern of more abundant short chain acyl-carnitines and less abundant long chain acyl-carnitines in the ASD-GI samples compared to TD samples (Figure 4A). Acyl-carnitines are positively correlated with more severe social defects, an effect driven by structures with shorter moieties (C2-C14) (Figure 1M, Table S1). In fecal samples, acetylcarnitine (C2) and free carnitine were elevated in ASD (Figures 4B-4C) and were highly discriminatory (Figure 1F). Other mitochondrial markers in both plasma and feces were also differentially abundant in ASD and are summarized in Figure 4D along with markers of phospholipid metabolism, which occurs largely in the mitochondria and was significantly altered in fecal samples (Figures 4D, 1D). Additionally, the 5 plasma metabolites most positively correlated with ADOS-SS are all involved in these cellular energy pathways (Figure S4D, Table S1). These observed differences in energy markers and lipids could have a neurodevelopmental effect during periods when the high lipid and energy requirement in the brain is crucial (*73–75*), and alterations to levels of tricarboxylic acid (TCA) cycle intermediates have been observed in human ASD prefrontal cortex samples (*49*). Such defects in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cellular metabolism support the theory that mitochondrial dysfunction may not only be comorbid with ASD but also a potential contributing factor, as suggested by numerous previous reports (*6*, *13*, *67*, *69*, *72*).

Amino acid signatures of oxidative stress and mitochondrial function are also present in our dataset. Dysregulated amino acid degradation, homeostasis, and import into the brain have been implicated as a cause of neuronal stress in ASD, and supporting metabolomic data has shown perturbations of various amino acid pathways, such as glutamate, methionine, glutathione, and gamma-glutamyl metabolites (*10*, *17*, *24*, *26*, *29*, *33*, *49*, *52*, *76*), which exhibit differences as well (Figure 4E). We also demonstrate correlations between pathways of oxidative stress (cysteine, methionine, SAM and glutathione pathways) with the ADOS-SS (Figure 4F). Some of these molecules were found in higher levels in the ASD feces and at lower levels in the ASD plasma, such as hypotaurine (Figures 4E, 4G-4H), which might indicate altered fecal production, excretion or differential uptake into the plasma potentially through varied intestinal permeability. Similar to hypotaurine, levels of its precursor, taurine, were significantly increased in ASD fecal samples, although not altered in plasma (Figure S4E and S4F). Taurine plays many roles throughout the host, and has previously been measured at altered levels in ASD, although with little consensus (*17*, *18*, *24–26*, *33*, *65*, *71*, *77*). Hypotaurine and taurine deficiency has been shown to lead to defects in cell differentiation in the brain (*77*) and their dysregulation could alter neuronal signaling (*78*).

Oxidative stress-related glutathione pathway precursors, gamma-glutamyl amino acids, are relevant to ASD through their influence on levels of neurotransmitters such as gamma-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

aminobutyric acid (GABA) (*78*, *79*), which is widely thought to play a role in ASD. In a recent ASD study, an experimental treatment that led to increased gamma-glutamyl AAs and other redox pathway metabolites correlated with improved behavior metrics in children (*65*). Almost every gamma-glutamyl amino acid is negatively correlated with ADOS-SS (Figure 4I). We also observe perturbations in the urea cycle, which processes the amino group of amino acids for excretion in urine (Figure S4G-S4H). Abnormalities in this pathway can be indicative of altered amino acid degradation observed in ASD and can lead to neurotoxic accumulation of nitrogen-containing compounds in the blood (*80*). Together, these results corroborate and extend a growing body of research into altered mitochondrial metabolism and oxidative stress in ASD.

Differential Phenolic Xenobiotic Metabolite Levels in ASD

Phenolic metabolites, a diverse structural class comprised of thousands of molecules containing a phenol moiety, come from dietary ingredients or from biotransformation of aromatic amino acids by the gut microbiota. In their free phenolic forms they can be readily absorbed through intestinal tissues; however, microbial modification of these molecules can significantly alter their absorption, bioavailability, and bioactivity (*81*), leading to various benefits or harm to the host (*81*, *82*). In fact, altered levels of phenolic molecules have been highlighted in many ASD metabolomic studies (*18*, *20–26*, *28*, *31*, *32*, *53*, *65*, *83*, *84*). However, a consensus of enrichment or depletion of these metabolites across ASD vs. TD groups has yet to be reached, and most studies have only measured a small subset of this structural class of metabolites.

We observed altered levels of phenolic metabolites belonging to several interrelated pathways, including tyrosine, benzoate, and food component and plant metabolites (Figure 5A). Some

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

molecules, such as homovanillate and tyramine, are involved in neurotransmitter metabolism (*18*). Others include hippurate derivatives, which we measured at lower levels in the ASD plasma samples (Figure 5A). Phytoestrogens such as daidzein, genistein, and equol derivatives are elevated in our ASD plasma samples (Figure 5A). These metabolites have purported health benefits but are also known disruptors of endocrine signaling (*85*). Additionally, we detected altered levels of molecules with structural similarity to para-cresol sulfate, a toxic molecule elevated in the urine of young ASD children (*20*, *22*). These include, among others, 4-ethylphenyl sulfate (4EPS), 2-ethylphenylsulfate, cresol derivatives, 4-allylphenyl sulfate, and 4-methylbenzenesulfonate, the latter of which was elevated a remarkable 60-fold in a small subset of ASD samples (Figure 5B-5D, Table S1). Some changes in the levels of these phenolic molecules are also observed in animal models of ASD (*50*, *86*). Previously, 4EPS was observed at a 46-fold elevated level in the maternal immune activation mouse model, and daily administration of synthetic 4EPS to wild type mice was sufficient to induce an anxiety-like phenotype (*50*). Here, in ASD plasma samples, 4EPS levels were increased 6.9-fold (Figure 5D). Other phenolic metabolite levels correlate with 4EPS in ASD samples, including 4-acetylphenyl sulfate, a derivative of 4EPS, and others. (Figure 5E, S5A-S5B). These observed alterations in phenolic molecules, combined with mounting evidence from previous studies, suggest that phenolic structural metabolites may play a role in ASD.

DISCUSSION

Changes in the metabolome have been linked to a number of neurodevelopmental and neurodegenerative disorders (*87*, *88*). The current study includes a comprehensive profiling of the metabolites from 231 plasma and 97 matched fecal samples from ASD and TD individuals

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

who have been extensively behavior tested. We observed that specific individual metabolites and metabolic pathways in lipid, amino acid, and xenobiotic metabolism were altered between ASD and TD groups, and that many of these same molecules and pathways correlated to severity of ASD behavioral scores. The core findings are summarized in Figure 6. Various steroid hormone metabolite levels were elevated in ASD samples. PUFA levels, and short chain acyl-carnitines were generally elevated in ASD plasma, while saturated fatty acid levels and long-chain acyl-carnitines were decreased, contributing to a general picture of dysregulated cellular energy and oxidative state, with potential connections to mitochondrial dysfunction in ASD. Some of the lipid and xenobiotic metabolites also showed interesting changes within the ASD group when stratified by the presence of GI symptoms. Finally, many phenolic metabolites, derived largely from host and bacterial metabolism of amino acids, plant polyphenols, and other food components were detected at differential levels in plasma and feces between the comparison groups.

ASD is diagnosed by behavioral tests, with extensive heterogeneity of symptoms, severity, and etiology between individuals and little consensus on molecular mechanisms. This enigmatic spectrum has a strong but complex genetic basis, with hundreds of reported risk genes (*89*). The contributions of risk alleles to behavior is a vast and active area of research. In addition to genetics, understanding of altered metabolite levels in blood, feces, and brains of ASD individuals may provide a glimpse into physiologic aspects of the disorder and hold the potential to advance diagnosis and/or stratification of sub-populations of ASD.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly, multiple mouse models support the notion that gut metabolites are associated with brain development and function (*48*, *90*, *91*). Some metabolites are closely linked with neurological disorders, either to positive or negative outcomes (*92*, *93*). Several examples have been reported of gut metabolites entering circulation and directly affecting the brain, as well as cases where metabolites stimulate pathways in the gut, immune, or autonomic nervous system and exert changes to the brain and to behavior (*94–97*). Comprehensive metabolic profiling in humans and animal models provides insight into the molecular status of disease and how genetic factors and environmental risks interact. Deeper analysis of our dataset along with additional studies, with future empirical studies to validate the relevance of our observations, could illuminate aspects of ASD pathophysiology. The intriguing correlations between ASD behaviors, altered levels of fecal and plasma metabolites, and GI symptoms contribute to the concept that ASD may be viewed as a whole-body condition, and argue for increased investigation into peripheral aspects of disease that may lead to advances in diagnosis and improved stratification of ASD populations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACKNOWLEDGMENTS

This work was supported by grants from Autism Speaks (Grant #7567 to P.A., A.F. and S.K.M.); the Johnson Foundation (to P.A.); the Brain Foundation (to P.A.); Genome, Environment, Microbiome, and Metabolome in Autism (GEMMA) (#825033 to AF); Axial Biotherapeutics (to S.K.M.); and the National Institutes of Health (HD090214 to P.A.) and (MH100556 to S.K.M.). We would like to thank the MIND Institute study staff, and the dedication and commitment of the families who took part in these studies is gratefully acknowledged.

Declaration of Interests

A.S.C., D.H.D. and S.K.M. have financial interest in Axial Biotherapeutics. A.F. has financial interest in Alba Therapeutics. G.M.P. and M.C.C. are employed by Axial Biotherapeutics. B.D.N., M.D.A., G.S., D.R.R., and P.A. report no financial conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. E. Doernberg, E. Hollander, Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. *CNS Spectr.* **21**, 295–299 (2016).

2. J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. Robinson Rosenberg, T. White, M. S. Durkin, P. Imm, L. Nikolaou, M. Yeargin-Allsopp, L.-C. Lee, R. Harrington, M. Lopez, R. T. Fitzgerald, A. Hewitt, S. Pettygrove, J. N. Constantino, A. Vehorn, J. Shenouda, J. Hall-Lande, K. Van Naarden Braun, N. F. Dowling, Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill. Summ.* **67**, 1–23 (2018).

3. G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, J. Varley, Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*. **125**, e17-23 (2010).

4. J. Virués-Ortega, Applied behavior analytic intervention for autism in early childhood: meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clin. Psychol. Rev.* **30**, 387–399 (2010).

5. A. Azhari, F. Azizan, G. Esposito, A systematic review of gut-immune-brain mechanisms in Autism Spectrum Disorder. *Dev. Psychobiol.* (2018), doi:10.1002/dev.21803.

6. D. A. Rossignol, R. E. Frye, Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol. Psychiatry*. **17**, 290–314 (2012).

7. N. Cheng, J. M. Rho, S. A. Masino, Metabolic Dysfunction Underlying Autism Spectrum Disorder and Potential Treatment Approaches. *Front. Mol. Neurosci.* **10** (2017), doi:10.3389/fnmol.2017.00034.

8. A. El-Ansary, L. Al-Ayadhi, Lipid mediators in plasma of autism spectrum disorders. *Lipids Health Dis.* **11**, 160 (2012).

9. A. K. El-Ansary, A. G. B. Bacha, L. Y. Al-Ayahdi, Impaired plasma phospholipids and relative amounts of essential polyunsaturated fatty acids in autistic patients from Saudi Arabia. *Lipids Health Dis.* **10**, 63 (2011).

10. S. J. James, P. Cutler, S. Melnyk, S. Jernigan, L. Janak, D. W. Gaylor, J. A. Neubrander, Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *Am. J. Clin. Nutr.* **80**, 1611–1617 (2004).

11. Q.-Q. Lv, C. You, X.-B. Zou, H.-Z. Deng, Acyl-carnitine, C5DC, and C26 as potential biomarkers for diagnosis of autism spectrum disorder in children. *Psychiatry Res.* **267**, 277–280 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. G. A. Mostafa, L. Y. AL-Ayadhi, Reduced levels of plasma polyunsaturated fatty acids and serum carnitine in autistic children: relation to gastrointestinal manifestations. *Behav. Brain Funct. BBF*. **11** (2015), doi:10.1186/s12993-014-0048-2.

13. E. Pastural, S. Ritchie, Y. Lu, W. Jin, A. Kavianpour, K. Khine Su-Myat, D. Heath, P. L. Wood, M. Fisk, D. B. Goodenowe, Novel plasma phospholipid biomarkers of autism: mitochondrial dysfunction as a putative causative mechanism. *Prostaglandins Leukot. Essent. Fatty Acids*. **81**, 253–264 (2009).

14. H. Wang, S. Liang, M. Wang, J. Gao, C. Sun, J. Wang, W. Xia, S. Wu, S. J. Sumner, F. Zhang, C. Sun, L. Wu, Potential serum biomarkers from a metabolomics study of autism. *J. Psychiatry Neurosci.* **41**, 27–37 (2016).

15. P. R. West, D. G. Amaral, P. Bais, A. M. Smith, L. A. Egnash, M. E. Ross, J. A. Palmer, B. R. Fontaine, K. R. Conard, B. A. Corbett, G. G. Cezar, E. L. R. Donley, R. E. Burrier, Metabolomics as a Tool for Discovery of Biomarkers of Autism Spectrum Disorder in the Blood Plasma of Children. *PLoS ONE*. **9**, e112445 (2014).

16. R. Cozzolino, L. De Magistris, P. Saggese, M. Stocchero, A. Martignetti, M. Di Stasio, A. Malorni, R. Marotta, F. Boscaino, L. Malorni, Use of solid-phase microextraction coupled to gas chromatography–mass spectrometry for determination of urinary volatile organic compounds in autistic children compared with healthy controls. *Anal. Bioanal. Chem.* **406**, 4649–4662 (2014).

17. X. Ming, T. P. Stein, V. Barnes, N. Rhodes, L. Guo, Metabolic Perturbance in Autism Spectrum Disorders: A Metabolomics Study. *J. Proteome Res.* **11**, 5856–5862 (2012).

18. A. Noto, V. Fanos, L. Barberini, D. Grapov, C. Fattuoni, M. Zaffanello, A. Casanova, G. Fenu, A. De Giacomo, M. De Angelis, C. Moretti, P. Papoff, R. Ditonno, R. Francavilla, The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. *J. Matern.-Fetal Neonatal Med. Off. J. Eur. Assoc. Perinat. Med. Fed. Asia Ocean. Perinat. Soc. Int. Soc. Perinat. Obstet.* **27 Suppl 2**, 46–52 (2014).

19. W. Shaw, E. Kassen, E. Chaves, Increased urinary excretion of analogs of Krebs cycle metabolites and arabinose in two brothers with autistic features. *Clin. Chem.* **41**, 1094–1104 (1995).

20. L. Altieri, C. Neri, R. Sacco, P. Curatolo, A. Benvenuto, F. Muratori, E. Santocchi, C. Bravaccio, C. Lenti, M. Saccani, R. Rigardetto, M. Gandione, A. Urbani, A. M. Persico, Urinary p-cresol is elevated in small children with severe autism spectrum disorder. *Biomark. Biochem. Indic. Expo. Response Susceptibility Chem.* **16**, 252–260 (2011).

21. P. Emond, S. Mavel, N. Aïdoud, L. Nadal-Desbarats, F. Montigny, F. Bonnet-Brilhault, C. Barthélémy, M. Merten, P. Sarda, F. Laumonnier, P. Vourc'h, H. Blasco, C. R. Andres, GC-MS-based urine metabolic profiling of autism spectrum disorders. *Anal. Bioanal. Chem.* **405**, 5291–5300 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. S. Gabriele, R. Sacco, S. Cerullo, C. Neri, A. Urbani, G. Tripi, J. Malvy, C. Barthelemy, F. Bonnet-Brihault, A. M. Persico, Urinary p-cresol is elevated in young French children with autism spectrum disorder: a replication study. *Biomark. Biochem. Indic. Expo. Response Susceptibility Chem.* **19**, 463–470 (2014).

23. A. W. Lis, I. Mclaughlin, R. K. Mpclaughlin, E. W. Lis, E. G. Stubbs, Profiles of ultraviolet-absorbing components of urine from autistic children, as obtained by high-resolution ion-exchange chromatography. *Clin. Chem.* **22**, 1528–1532 (1976).

24. M. Lussu, A. Noto, A. Masili, A. C. Rinaldi, A. Dessì, M. D. Angelis, A. D. Giacomo, V. Fanos, L. Atzori, R. Francavilla, The urinary 1H-NMR metabolomics profile of an italian autistic children population and their unaffected siblings. *Autism Res.* **10**, 1058–1066 (2017).

25. L. Nadal-Desbarats, N. Aïdoud, P. Emond, H. Blasco, I. Filipiak, P. Sarda, F. Bonnet-Brilhault, S. Mavel, C. R. Andres, Combined ^{1}H-NMR and ^{1}H– ^{13}C HSQC-NMR to improve urinary screening in autism spectrum disorders. *The Analyst*. **139**, 3460–3468 (2014).

26. I. K. S. Yap, M. Angley, K. A. Veselkov, E. Holmes, J. C. Lindon, J. K. Nicholson, Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *J. Proteome Res.* **9**, 2996–3004 (2010).

27. J. B. Adams, L. J. Johansen, L. D. Powell, D. Quig, R. A. Rubin, Gastrointestinal flora and gastrointestinal status in children with autism--comparisons to typical children and correlation with autism severity. *BMC Gastroenterol.* **11**, 22 (2011).

28. M. D. Angelis, M. Piccolo, L. Vannini, S. Siragusa, A. D. Giacomo, D. I. Serrazzanetti, F. Cristofori, M. E. Guerzoni, M. Gobbetti, R. Francavilla, Fecal Microbiota and Metabolome of Children with Autism and Pervasive Developmental Disorder Not Otherwise Specified. *PLOS ONE*. **8**, e76993 (2013).

29. D.-W. Kang, Z. E. Ilhan, N. G. Isern, D. W. Hoyt, D. P. Howsmon, M. Shaffer, C. A. Lozupone, J. Hahn, J. B. Adams, R. Krajmalnik-Brown, Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. *Anaerobe*. **49**, 121–131 (2018).

30. L. Wang, C. T. Christophersen, M. J. Sorich, J. P. Gerber, M. T. Angley, M. A. Conlon, Elevated Fecal Short Chain Fatty Acid and Ammonia Concentrations in Children with Autism Spectrum Disorder. *Dig. Dis. Sci.* **57**, 2096–2102 (2012).

31. T. Bitar, S. Mavel, P. Emond, L. Nadal-Desbarats, A. Lefèvre, H. Mattar, M. Soufia, H. Blasco, P. Vourc'h, W. Hleihel, C. R. Andres, Identification of metabolic pathway disturbances using multimodal metabolomics in autistic disorders in a Middle Eastern population. *J. Pharm. Biomed. Anal.* **152**, 57–65 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

32. F. Gevi, L. Zolla, S. Gabriele, A. M. Persico, Urinary metabolomics of young Italian autistic children supports abnormal tryptophan and purine metabolism. *Mol. Autism*. **7** (2016), doi:10.1186/s13229-016-0109-5.

33. H. Kuwabara, H. Yamasue, S. Koike, H. Inoue, Y. Kawakubo, M. Kuroda, Y. Takano, N. Iwashiro, T. Natsubori, Y. Aoki, Y. Kano, K. Kasai, Altered metabolites in the plasma of autism spectrum disorder: a capillary electrophoresis time-of-flight mass spectroscopy study. *PloS One*. **8**, e73814 (2013).

34. J. S. Orozco, I. Hertz-Picciotto, L. Abbeduto, C. M. Slupsky, Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome. *Transl. Psychiatry*. **9**, 243 (2019).

35. B. J. Ferguson, K. Dovgan, D. Severns, S. Martin, S. Marler, K. Gross Margolis, M. L. Bauman, J. Veenstra-VanderWeele, K. Sohl, D. Q. Beversdorf, Lack of Associations Between Dietary Intake and Gastrointestinal Symptoms in Autism Spectrum Disorder. *Front. Psychiatry*. **10**, 528 (2019).

36. M. R. Sanctuary, J. N. Kain, K. Angkustsiri, J. B. German, Dietary Considerations in Autism Spectrum Disorders: The Potential Role of Protein Digestion and Microbial Putrefaction in the Gut-Brain Axis. *Front. Nutr.* **5** (2018), doi:10.3389/fnut.2018.00040.

37. S. A. Cermak, C. Curtin, L. G. Bandini, Food selectivity and sensory sensitivity in children with autism spectrum disorders. *J. Am. Diet. Assoc.* **110**, 238–246 (2010).

38. B. D. Needham, W. Tang, W.-L. Wu, Searching for the gut microbial contributing factors to social behavior in rodent models of autism spectrum disorder. *Dev. Neurobiol.* **78**, 474–499 (2018).

39. J. Pulikkan, A. Mazumder, T. Grace, Role of the Gut Microbiome in Autism Spectrum Disorders. *Adv. Exp. Med. Biol.* **1118**, 253–269 (2019).

40. H. E. Vuong, J. M. Yano, T. C. Fung, E. Y. Hsiao, The Microbiome and Host Behavior. *Annu. Rev. Neurosci.* **40**, 21–49 (2017).

41. I. Hertz-Picciotto, L. A. Croen, R. Hansen, C. R. Jones, J. van de Water, I. N. Pessah, The CHARGE Study: An Epidemiologic Investigation of Genetic and Environmental Factors Contributing to Autism. *Environ. Health Perspect.* **114**, 1119–1125 (2006).

42. D. R. Rose, H. Yang, G. Serena, C. Sturgeon, B. Ma, M. Careaga, H. K. Hughes, K. Angkustsiri, M. Rose, I. Hertz-Picciotto, J. Van de Water, R. L. Hansen, J. Ravel, A. Fasano, P. Ashwood, Differential immune responses and microbiota profiles in children with autism spectrum disorders and co-morbid gastrointestinal symptoms. *Brain. Behav. Immun.* **70**, 354–368 (2018).

43. D. A. Drossman, The Functional Gastrointestinal Disorders and the Rome III Process. *Gastroenterology*. **130**, 1377–1390 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

44. R. S. Kelly, A. Boulin, N. Laranjo, K. Lee-Sarwar, S. H. Chu, A. P. Yadama, V. Carey, A. A. Litonjua, J. Lasky-Su, S. T. Weiss, Metabolomics and Communication Skills Development in Children; Evidence from the Ages and Stages Questionnaire. *Metabolites*. **9**, 42 (2019).

45. A. Weiler, A. Volkenhoff, H. Hertenstein, S. Schirmeier, Metabolite transport across the mammalian and insect brain diffusion barriers. *Neurobiol. Dis.* **107**, 15–31 (2017).

46. R. E. Frye, S. Melnyk, D. F. MacFabe, Unique acyl-carnitine profiles are potential biomarkers for acquired mitochondrial disease in autism spectrum disorder. *Transl. Psychiatry*. **3**, e220 (2013).

47. C. Gillberg, E. Fernell, E. Kočovská, H. Minnis, T. Bourgeron, L. Thompson, C. S. Allely, The role of cholesterol metabolism and various steroid abnormalities in autism spectrum disorders: A hypothesis paper. *Autism Res. Off. J. Int. Soc. Autism Res.* **10**, 1022–1044 (2017).

48. A. V. Golubeva, S. A. Joyce, G. Moloney, A. Burokas, E. Sherwin, S. Arboleya, I. Flynn, D. Khochanskiy, A. Moya-Pérez, V. Peterson, K. Rea, K. Murphy, O. Makarova, S. Buravkov, N. P. Hyland, C. Stanton, G. Clarke, C. G. M. Gahan, T. G. Dinan, J. F. Cryan, Microbiota-related Changes in Bile Acid & Tryptophan Metabolism are Associated with Gastrointestinal Dysfunction in a Mouse Model of Autism. *EBioMedicine*. **24**, 166–178 (2017).

49. I. Kurochkin, E. Khrameeva, A. Tkachev, V. Stepanova, A. Vanyushkina, E. Stekolshchikova, Q. Li, D. Zubkov, P. Shichkova, T. Halene, L. Willmitzer, P. Giavalisco, S. Akbarian, P. Khaitovich, Metabolome signature of autism in the human prefrontal cortex. *Commun. Biol.* **2**, 234 (2019).

50. E. Y. Hsiao, S. W. McBride, S. Hsien, G. Sharon, E. R. Hyde, T. McCue, J. A. Codelli, J. Chow, S. E. Reisman, J. F. Petrosino, P. H. Patterson, S. K. Mazmanian, Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. *Cell*. **155**, 1451–1463 (2013).

51. H. M. Roager, T. R. Licht, Microbial tryptophan catabolites in health and disease. *Nat. Commun.* **9**, 1–10 (2018).

52. H.-F. Zheng, W.-Q. Wang, X.-M. Li, G. Rauw, G. B. Baker, Body fluid levels of neuroactive amino acids in autism spectrum disorders: a review of the literature. *Amino Acids*. **49**, 57–65 (2017).

53. B. Diémé, S. Mavel, H. Blasco, G. Tripi, F. Bonnet-Brilhault, J. Malvy, C. Bocca, C. R. Andres, L. Nadal-Desbarats, P. Emond, Metabolomics Study of Urine in Autism Spectrum Disorders Using a Multiplatform Analytical Methodology. *J. Proteome Res.* **14**, 5273–5282 (2015).

54. L. Santamaría, I. Reverón, F. L. de Felipe, B. de las Rivas, R. Muñoz, Ethylphenol Formation by Lactobacillus plantarum: Identification of the Enzyme Involved in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reduction of Vinylphenols. *Appl. Environ. Microbiol.* **84** (2018), doi:10.1128/AEM.01064-18.

55. B. Pragnya, J. S. L. Kameshwari, B. Veeresh, Ameliorating effect of piperine on behavioral abnormalities and oxidative markers in sodium valproate induced autism in BALB/C mice. *Behav. Brain Res.* **270**, 86–94 (2014).

56. B. Auyeung, J. Ahluwalia, L. Thomson, K. Taylor, G. Hackett, K. J. O'Donnell, S. Baron-Cohen, Prenatal versus postnatal sex steroid hormone effects on autistic traits in children at 18 to 24 months of age. *Mol. Autism*. **3**, 17 (2012).

57. S. Baron-Cohen, B. Auyeung, B. Nørgaard-Pedersen, D. M. Hougaard, M. W. Abdallah, L. Melgaard, A. S. Cohen, B. Chakrabarti, L. Ruta, M. V. Lombardo, Elevated fetal steroidogenic activity in autism. *Mol. Psychiatry*. **20**, 369–376 (2015).

58. M. H. Hassan, T. Desoky, H. M. Sakhr, R. H. Gabra, A. H. Bakri, Possible Metabolic Alterations among Autistic Male Children: Clinical and Biochemical Approaches. *J. Mol. Neurosci. MN* (2019), doi:10.1007/s12031-018-1225-9.

59. M. P. Mamidala, A. Polinedi, P. T. V. P. Kumar, N. Rajesh, O. R. Vallamkonda, V. Udani, N. Singhal, V. Rajesh, Maternal hormonal interventions as a risk factor for Autism Spectrum Disorder: an epidemiological assessment from India. *J. Biosci.* **38**, 887–892 (2013).

60. P. M. Whitaker-Azmitia, M. Lobel, A. Moyer, Low maternal progesterone may contribute to both obstetrical complications and autism. *Med. Hypotheses*. **82**, 313–318 (2014).

61. M. D. Majewska, M. Hill, E. Urbanowicz, P. Rok-Bujko, P. Bieńkowski, I. Namysłowska, P. Mierzejewski, Marked elevation of adrenal steroids, especially androgens, in saliva of prepubertal autistic children. *Eur. Child Adolesc. Psychiatry*. **23**, 485–498 (2014).

62. J. Willing, C. K. Wagner, Exposure to the Synthetic Progestin, 17α-Hydroxyprogesterone Caproate During Development Impairs Cognitive Flexibility in Adulthood. *Endocrinology*. **157**, 77–82 (2016).

63. A. Quartier, L. Chatrousse, C. Redin, C. Keime, N. Haumesser, A. Maglott-Roth, L. Brino, S. Le Gras, A. Benchoua, J.-L. Mandel, A. Piton, Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biol. Psychiatry*. **84**, 239–252 (2018).

64. D. A. Geier, J. K. Kern, P. G. King, L. K. Sykes, M. R. Geier, An evaluation of the role and treatment of elevated male hormones in autism spectrum disorders. *Acta Neurobiol. Exp. (Warsz.)*. **72**, 1–17 (2012).

65. S. Bent, B. Lawton, T. Warren, F. Widjaja, K. Dang, J. W. Fahey, B. Cornblatt, J. M. Kinchen, K. Delucchi, R. L. Hendren, Identification of urinary metabolites that correlate with clinical improvements in children with autism treated with sulforaphane from broccoli. *Mol. Autism*. **9** (2018), doi:10.1186/s13229-018-0218-4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

66. L. Kanner, Autistic disturbances of affective contact. *Acta Paedopsychiatr.* **35**, 100–136 (1968).

67. K. K. Griffiths, R. J. Levy, Evidence of Mitochondrial Dysfunction in Autism: Biochemical Links, Genetic-Based Associations, and Non-Energy-Related Mechanisms. *Oxid. Med. Cell. Longev.* **2017** (2017), doi:10.1155/2017/4314025.

68. A. Legido, R. Jethva, M. J. Goldenthal, Mitochondrial dysfunction in autism. *Semin. Pediatr. Neurol.* **20**, 163–175 (2013).

69. J. Lombard, Autism: a mitochondrial disorder? *Med. Hypotheses*. **50**, 497–500 (1998).

70. M. Coleman, J. P. Blass, Autism and lactic acidosis. *J. Autism Dev. Disord.* **15**, 1–8 (1985).

71. S. Mavel, L. Nadal-Desbarats, H. Blasco, F. Bonnet-Brilhault, C. Barthélémy, F. Montigny, P. Sarda, F. Laumonnier, P. Vourc′h, C. R. Andres, P. Emond, 1H–13C NMR-based urine metabolic profiling in autism spectrum disorders. *Talanta*. **114**, 95–102 (2013).

72. D. F. MacFabe, Enteric short-chain fatty acids: microbial messengers of metabolism, mitochondria, and mind: implications in autism spectrum disorders. *Microb. Ecol. Health Dis.* **26** (2015), doi:10.3402/mehd.v26.28177.

73. C.-Y. Chang, D.-S. Ke, J.-Y. Chen, Essential fatty acids and human brain. *Acta Neurol. Taiwanica*. **18**, 231–241 (2009).

74. C. W. Kuzawa, H. T. Chugani, L. I. Grossman, L. Lipovich, O. Muzik, P. R. Hof, D. E. Wildman, C. C. Sherwood, W. R. Leonard, N. Lange, Metabolic costs and evolutionary implications of human brain development. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13010–13015 (2014).

75. J. S. O'Brien, E. L. Sampson, Lipid composition of the normal human brain: gray matter, white matter, and myelin. *J. Lipid Res.* **6**, 537–544 (1965).

76. A. M. Smith, J. J. King, P. R. West, M. A. Ludwig, E. L. R. Donley, R. E. Burrier, D. G. Amaral, Amino Acid Dysregulation Metabotypes: Potential Biomarkers for Diagnosis and Individualized Treatment for Subtypes of Autism Spectrum Disorder. *Biol. Psychiatry*. **85**, 345–354 (2019).

77. H. Ripps, W. Shen, Review: Taurine: A "very essential" amino acid. *Mol. Vis.* **18**, 2673–2686 (2012).

78. A. Dickinson, M. Jones, E. Milne, Measuring neural excitation and inhibition in autism: Different approaches, different findings and different interpretations. *Brain Res.* **1648**, 277–289 (2016).

79. C. A. Olson, H. E. Vuong, J. M. Yano, Q. Y. Liang, D. J. Nusbaum, E. Y. Hsiao, The Gut Microbiota Mediates the Anti-Seizure Effects of the Ketogenic Diet. *Cell*. **173**, 1728-1741.e13 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

80. A. Liu, W. Zhou, L. Qu, F. He, H. Wang, Y. Wang, C. Cai, X. Li, W. Zhou, M. Wang, Altered Urinary Amino Acids in Children With Autism Spectrum Disorders. *Front. Cell. Neurosci.* **13** (2019), doi:10.3389/fncel.2019.00007.

81. A. Duda-Chodak, T. Tarko, P. Satora, P. Sroka, Interaction of dietary compounds, especially polyphenols, with the intestinal microbiota: a review. *Eur. J. Nutr.* **54**, 325–341 (2015).

82. L. Valdés, A. Cuervo, N. Salazar, P. Ruas-Madiedo, M. Gueimonde, S. González, The relationship between phenolic compounds from diet and microbiota: impact on human health. *Food Funct.* **6**, 2424–2439 (2015).

83. J. Kałużna-Czaplińska, E. Żurawicz, W. Struck, M. Markuszewski, Identification of organic acids as potential biomarkers in the urine of autistic children using gas chromatography/mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **966**, 70–76 (2014).

84. W. J. Johannsen, S. H. Friedman, E. I. Feldman, A. Negrete, A Re-examination of the Hippuric Acid--Anxiety Relationship. *Psychosom. Med.* **24**, 569 (1962).

85. H. B. Patisaul, W. Jefferson, The pros and cons of phytoestrogens. *Front. Neuroendocrinol.* **31**, 400–419 (2010).

86. J. C. Naviaux, M. A. Schuchbauer, K. Li, L. Wang, V. B. Risbrough, S. B. Powell, R. K. Naviaux, Reversal of autism-like behaviors and metabolism in adult mice with single-dose antipurinergic therapy. *Transl. Psychiatry*. **4**, e400–e400 (2014).

87. M. Grochowska, T. Laskus, M. Radkowski, Gut Microbiota in Neurological Disorders. *Arch. Immunol. Ther. Exp. (Warsz.)*. **67**, 375–383 (2019).

88. G. Sharon, T. R. Sampson, D. H. Geschwind, S. K. Mazmanian, The Central Nervous System and the Gut Microbiome. *Cell*. **167**, 915–932 (2016).

89. L. M. Iakoucheva, A. R. Muotri, J. Sebat, Getting to the Cores of Autism. *Cell*. **178**, 1287–1298 (2019).

90. L. Desbonnet, G. Clarke, A. Traplin, O. O'Sullivan, F. Crispie, R. D. Moloney, P. D. Cotter, T. G. Dinan, J. F. Cryan, Gut microbiota depletion from early adolescence in mice: Implications for brain and behaviour. *Brain. Behav. Immun.* **48**, 165–173 (2015).

91. K. M. Neufeld, N. Kang, J. Bienenstock, J. A. Foster, Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterol. Motil. Off. J. Eur. Gastrointest. Motil. Soc.* **23**, 255–264, e119 (2011).

92. M. Kujawska, J. Jodynis-Liebert, Polyphenols in Parkinson's Disease: A Systematic Review of In Vivo Studies. *Nutrients*. **10**, 642 (2018).

93. T. R. Sampson, J. W. Debelius, T. Thron, S. Janssen, G. G. Shastri, Z. E. Ilhan, C. Challis, C. E. Schretter, S. Rocha, V. Gradinaru, M.-F. Chesselet, A. Keshavarzian, K. M. Shannon,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

R. Krajmalnik-Brown, P. Wittung-Stafshede, R. Knight, S. K. Mazmanian, Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell*. **167**, 1469-1480.e12 (2016).

94. M. Jaglin, M. Rhimi, C. Philippe, N. Pons, A. Bruneau, B. Goustard, V. Daugé, E. Maguin, L. Naudon, S. Rabot, Indole, a Signaling Molecule Produced by the Gut Microbiota, Negatively Impacts Emotional Behaviors in Rats. *Front. Neurosci.* **12** (2018), doi:10.3389/fnins.2018.00216.

95. J. C. O'Connor, M. A. Lawson, C. André, M. Moreau, J. Lestage, N. Castanon, K. W. Kelley, R. Dantzer, Lipopolysaccharide-induced depressive-like behavior is mediated by indoleamine 2,3-dioxygenase activation in mice. *Mol. Psychiatry*. **14**, 511–522 (2009).

96. P. Tian, G. Wang, J. Zhao, H. Zhang, W. Chen, Bifidobacterium with the role of 5-hydroxytryptophan synthesis regulation alleviates the symptom of depression and related microbiota dysbiosis. *J. Nutr. Biochem.* **66**, 43–51 (2019).

97. D. Wang, L. Ho, J. Faith, K. Ono, E. M. Janle, P. J. Lachcik, B. R. Cooper, A. H. Jannasch, B. R. D'Arcy, B. A. Williams, M. G. Ferruzzi, S. Levine, W. Zhao, L. Dubner, G. M. Pasinetti, Role of intestinal microbiota in the generation of polyphenol-derived phenolic acid mediated attenuation of Alzheimer's disease β-amyloid oligomerization. *Mol. Nutr. Food Res.* **59**, 1025–1040 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURE TITLES AND LEGENDS

Figure 1. Plasma and Fecal Metabolomes Differ between ASD and TD and Global Metabolite Levels Correlate with Clinical Behavioral Scores

(A-B) The number of significantly elevated and decreased metabolites (pval<0.05) in ASD samples compared to the TD control group by ANOVA contrasts in plasma and feces, respectively. Samples are stratified by all samples or samples without or with GI symptoms (-GI, +GI). *See also Table S1 and S2.* (C-D) Pathway analysis results of human plasma and fecal comparisons (all samples), indicating which metabolomic pathways are the most altered between groups, with enrichment value plotted and p-value to the right of each bar. Metabolites within each pathway could be observed at either higher or lower levels, as this plot only indicates significant changes. (E-F) Top 30 most distinguishing metabolites between each group in plasma and feces by random forest analysis, with mean decrease accuracy along the x-axis. Metabolites known to be produced by (asterisks) or influenced by (triangles) the gut microbiota are denoted. The super pathway to which each metabolite belongs to is indicated by color of sphere and defined in the legend. *See also Figure S1.* (G-H) Scaled intensity values indicating relative levels of the most distinguishing molecules between ASD and TD (all samples) in plasma. Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 2.87E-09(G) and 1.57E-06(H). Data are represented as mean ± SEM. (I-J) Scaled intensity values indicating relative levels of the most distinguishing molecules between ASD and TD (all samples) in feces. Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 1.37E-05(I) and 0.002(J). Data are represented as mean ± SEM. (K) Top correlated plasma metabolites that covary with margaroylcarnitine and indolelactate. All p-values <0.0001. (L) Top correlated fecal metabolites that covary with nicotinamide and 9-HOTrE. All p-values <0.0001. (M) Spearman correlations between behavior scores of ASD children in the ADIR diagnostic test (Verbal, Social, and Nonverbal metrics) and the ADOS-SS and metabolite pathways in ASD samples. Directionality of correlation is indicated in the legend at bottom. Colors of pathways are defined at the top left of the chart. A split box means that both positive and negative correlations occur with metabolites within that pathway. *See also Tables S1 and S2.* (N) PCA plot comparing the metabolic profile least and most severe ~quartiles within verbal, social and ADOS-SS scores. PCA input included all metabolites significantly associated with the behavior. Clustering is denoted by ellipses of the 95% confidence interval. LPC, lysophosphatidylcholine; CE, cholesterol ester; FFA, free fatty acid; androst., androstane; hydroxypreg, hydroxypregnenalone; PFOS, perfluorooctanesulfonic acid; hydroxy-CMPF, hydroxy-3-carboxy-4-methyl-5-propyl-2- furanpropionate; DHEA-s, dehydroepiandrosterone sulfate; 9-HOTrE, 9S-hydroxy-10E,12Z,15Z-octadecatrienoic acid; AMP, adenosine monophosphate; HExCer, Hexosylceramide; PC, phosphatidylcholine; DAG, diacylglycerol; TAG, triacylglycerol; MFA, monounsaturated fatty acid; PE plas, phosphatidylethanolamine plasmalogens; Endocann, endocannabinoid; Met, methionine; Cys, cysteine; SAM, s-adenosyl methionine; Ala, alanine; Asp, aspartate; ɣ-glutAA, gamma-glutamyl amino acids; Second. Bile, secondary bile; MAG, monoacylglycerol; Gly, glycine; Ser, serine; thr, threonine. *See also Figures S1 and S2.*

Figure 2. Lipid and Xenobiotic Metabolites Correlate with GI Symptoms

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(A-B) Numbers of significantly (p value<0.05) elevated and decreased metabolites within the ASD samples, comparing ASD+GI to ASD-GI in plasma and feces, respectively, by ANOVA contrasts. (C) All pathways significantly altered in the comparison between ASD+GI and ASD-GI human plasma samples, with enrichment value plotted and p-value to the right of each bar. Metabolites within each pathway could be observed at either higher or lower levels, as this plot only indicates significant changes. (D) Complex lipid panel results for the ASD+GI vs ASD-GI plasma samples. Along the top, length of lipid chain of the free fatty acid is displayed (chain length:number of unsaturations) and are grouped into saturated (SAT), monounsaturated (MUFA), and polyunsaturated (PUFA) fatty acids. Color legend is displayed below. (E-G) Scaled intensity values indicating relative levels of free and conjugated piperine in ASD+GI and ASD-GI symptoms. Data are represented as mean ± SEM. Asterisks indicate significance in unpaired t test with Welch's correction. P-values: 0.03(E), 0.01(F), and 0.003(G). Free piperine does not reach statistical significance in ANOVA comparison performed on total dataset (p value=0.16). Individual p values (all<0.05) of each sulfated and glucuronidated form in ANOVA comparison can be found in *Table S1*. (H) Spearman correlation of piperine and derivatives with ADOS-SS. *See also Figure S1.* FFA, free fatty acid; monounsat., monounsaturated.

Figure 3. Steroid Hormone Levels are Elevated in ASD and other Lipid Metabolite Levels Differ in ASD

(A) Significant alterations to levels of all metabolites detected in the pregnenolone, progestin, and androgen steroid pathways in plasma (P) and feces (F), with colors indicating significance and fold change according to legend. (B) Schematic summarizing the metabolism of these pathways. (C) Complex lipid panel results for all the ASD plasma samples compared to TD controls with acyl chain length of lipids across the top, described by chain length, degree of unsaturation and categorized by saturated (SAT), monounsaturated (MUFA), and polyunsaturated (PUFA) fatty acids. Lipid classes are listed along the left. Direction of change and significance are indicated by the legend. Significance determined by ANOVA contrasts. *See also Figure S3 and Table S3.*

Figure 4. ASD Correlates with Cellular Energy and Oxidative Stress Metabolites

(A) Log2 fold change of acyl-carnitines in the plasma of ASD-GI samples compared to controls. Significance indicated by color according to legend below, determined by ANOVA contrasts. (B-C) Scaled intensity values indicating relative levels of acetylcarnitine(C2) and carnitine, respectively, in ASD fecal samples compared to TD controls (all samples). Data are represented as mean ± SEM. Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 6.44E-05(B) and 2.55E-03(C). (D) Schematic of mitochondrial markers and other metabolites closely associated with cellular energy in plasma (within center box) and feces (boxed to left). *=significant only in ASD-GI. Color of text indicates direction and significance of change according to legend above. (E) Differences in levels of metabolites from the cysteine, methionine, and glutathione pathways, comparing ASD vs TD. Refer to legends below and in D. (F) Correlation of ADOS-SS with metabolites from the cysteine, methionine, and glutathione pathways. Significant and trending metabolites corresponding to the linear regression in the graph are listed along with Pearson coefficients and p-values. Insignificant correlation values are not listed here. Refer to color legend in D. (G-H) Scaled intensity values indicating relative levels of hypotaurine in feces (G) and plasma (H) (all

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

samples). Data are represented as mean ± SEM. Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 1.61E-05(G) and 1.35E-3(H). (I) Correlations of gamma-glutamyl amino acids with ADOS-SS, color legend in D, with Pearson coefficients and p-values to the right.
PC, phosphatidylcholine; PE, phosphatidylethanolamine; GPG, glycerophosphoglycerol; GPC, glycerophosphocholine; GPE, glycerophosphoethanolamine; GPS, glycerophosphoserine; GPI, glycerophosphoinositol; TAG, triacylglycerol; DAG, diacylglycerol; FFA, free fatty acid; SAT, saturated fatty acid; PUFA, polyunsaturated fatty acid. Data are represented as mean ± SEM. *See also Figure S4 and Tables S1 and S2.*

Figure 5. Differential Phenolic Xenobiotic Metabolite Levels in ASD

(A) Phenolic metabolites, belonging to the benzoate, tyrosine, and food component/plant pathways that are significantly different between ASD vs TD groups in plasma (P), and feces (F). Directionality and significance defined in legend. (B) Scaled intensity values indicating relative levels of 4-allylphenol sulfate in plasma (all samples). Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p value: 2.50-03. (C) Scaled intensity values indicating relative levels of 2-ethylphenyl sulfate in plasma (all samples). (D) Scaled intensity values indicating relative levels of 4-ethylphenyl sulfate (4EPS) in plasma (all samples). Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p value: 1.16E-03. Data in (B-D) are represented as mean ± SEM. (E) Nearest neighbor correlation between plasma 4EPS and 4-acetylphenol sulfate, $\log_2$ scale. Pearson correlation is indicated. p-values: 0.06(TD) and <1E-04(ASD). *See also Figure S5 and Tables S1 and S2.*

Figure 6. Summary chart of core findings

Categorized into metabolites of lipid, mitochondrial function marker, and xenobiotic and phenolic pathways as well as potential biomarkers, the metabolites of most interest (left) and observations made from the data (right) are summarized.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Figure S1. Related to Figures 1 and 2. Group characteristics. (A) Numbers of plasma samples: sex, GI symptoms, age range, and total n. (B) Numbers of fecal samples: sex, GI symptoms, age range, and total n. (C) PCA plot of plasma (left) and feces (right) with all metabolites as input. (D-F) Data are represented as mean ± SEM. (D) Ages of donors of all plasma samples (p=0.007), all plasma samples 11 years and younger (p=0.051), and according to GI status. (p=0.26) (E) Ages of donors of all fecal samples (p=0.15) and samples according to GI status (p=0.075). (F) ADOS-SS graphed according to GI status within the ASD sample group. (p=0.018) (G) Correlations between age (in months) ASD plasma sample individuals and behavior metrics are shown, with linear regression line, spearman correlation value, and p-value. (H) PCA of plasma metabolites significantly correlated with nonverbal behavior, grouped by severity. Ellipses indicate 95% confidence interval. (I) PCA of fecal metabolites significantly correlated with behavior scores, grouped by severity. Ellipses indicate 95% confidence interval.

Figure S2. Related to Figures 1 and 2. Relative vs absolute values for select metabolites, pathway enrichment analysis of ASD-GI and ASD+GI comparisons to TD controls, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transfer of human fecal microbiota into mice. (A) Correlation plots comparing the relative abundance measured in the untargeted analysis and the follow-up quantitated values for a subset of 139 plasma samples. Correlations for 4-ethylphenyl sulfate, p-cresol sulfate, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF), 3-hydroxyhippurate, 3-indoxylsulfate, N-acetylserine, and p-cresol glucuronide are shown. (B) Significantly affected pathways in ASD+GI vs TD and ASD-GI vs TD comparisons with enrichment value plotted and p-value to the right of each bar. At the top plasma sample results are displayed, with fecal sample results below. (C) 16s profiles for ASD and TD colonized mice. When identified, genus and species are indicated. When family was the most specific identifier, noted by f in legend. (D) Volcano plot for comparison of 16s sequencing in recipient ASD compared to TD mice, genus level. Arrow indicating significant result of *Bacteroides* genus, W value=39. (E) PCA of ASD and TD colonized mice, with all significantly different metabolites from ANOVA contrasts used as input (top) and all metabolites as input (bottom). Ellipses indicate 95% confidence intervals (F) Scaled intensity values indicating relative levels of 4EPS levels in plasma from donors used to colonize mice. p-value 0.23 (G) Scaled intensity values indicating relative levels of 4EPS levels in mice colonized with TD or ASD donors, colored according to donor. Log10 scaled intensity of biochemical peaks is plotted along the y-axis. (H) LeFSe analysis of fecal bacteria at the genus level from mice exhibiting high or low 4EPS levels (regardless of donor diagnosis) after colonization with human samples. (I) Volcano plot of -log10 p-value and log2 fold change of all mouse metabolites analyzed by mixed effects model. All metabolites above a log2 fold change of +/-0.5 and a -log10 p-value of 0.3 are colored. Triangles denote phenolic metabolites. Those phenolic metabolites also significantly altered in the human samples are labeled by name. *See also Table S5.*

Figure S3. Related to Figure 3. Steroid hormones by sex and age. (A) Levels of androgens, pregnenolones and progestins in all female plasma samples. Scaled intensity values for each female plasma sample. p-values of Welch's corrected t test are shown about each graph. (B) Levels of androgens, pregnenolones and progestins in all male plasma samples. Scaled intensity values for each male plasma sample shown. p-values of Welch's corrected t test are shown about each graph. (C) Pregnenolone, progestin, and androgen steroid metabolites clustered by tighter age groupings, denoted in years or months. Dark red indicates significant difference (pval<0.05), light red indicates a trend (0.05>pval>0.1), with fold change in text.

Figure S4. Related to Figures 3 and 4. Complex lipid panel, cellular energy, and amino acid results. (A) Sphingolipids with significantly altered levels in plasma of ASD compared to TD (all samples). (B) Biochemical importance plot of Random Forest analysis using complex lipid panel results for all plasma samples. (C) Scaled intensity values indicating relative levels of two representative PUFA (18:2)-containing fecal lipids that differ in directionality from the plasma phenotype shown in main text Figure 3B. Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 4.5E-04(left) and 5.9E-04(right). Data are represented as mean ± SEM. (D) The top 5 most positively correlated plasma metabolites with ADOS-SS. The metabolite, spearman correlation value, and p-value are listed. (E) Scaled intensity values indicating relative levels of taurine in feces (all samples). Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p value: 0.012. Data are represented as mean ± SEM. (F) Scaled intensity values indicating relative levels of taurine in plasma (all samples). p-value from ANOVA contrasts performed on total metabolomics

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dataset. Scaled intensity of taurine levels in plasma (all samples). (G) Abbreviated pathway schematic of significantly different metabolites in the urea cycle in plasma. Asterisk indicates significance only in the ASD-GI/TD comparison. (H) Scaled intensity values indicating relative levels of urea cycle metabolites that are altered in feces: urea, norvaline, and argininate (all samples). Asterisks indicate significance in ANOVA contrasts performed on total metabolomics dataset. p values: 1.7E-04(urea), 0.02(norvaline), and 0.014(argininate). Data are represented as mean ± SEM.

Figure S5. Related to Figure 5. Nearest neighbor analysis of 4EPS. (A) Nearest neighbor analysis of 4EPS plasma levels resulting from correlations with all fecal metabolite levels. Red star denotes phenolic compounds. (B) Nearest neighbor analysis of 4EPS plasma levels against all plasma metabolite levels (minus 4-acetylphenol sulfate from Main Figure 5C). Red star denotes phenolic compounds.

Table S1. Related to Figures 1-6. Fold change and statistics of human plasma metabolites and correlations to behavior scores. ANOVA contrasts, p-value, q-value, % filled values, and pearson correlation r is provided.

Table S2. Related to Figure 1-6. Fold change and statistics of human fecal metabolites and correlations to behavior scores. ANOVA contrasts, p-value, q-value, % filled values, and pearson correlation r is provided.

Table S3. Related to Figure 6. Fold change and p value (Welch's t test) of human donor plasma metabolites.

Table S4. Related to Figure 6. Fold change and p value (Welch's t test) of human donor fecal metabolites.

Table S5. Related to Figure 6. Fold change and statistics of mouse plasma metabolites.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

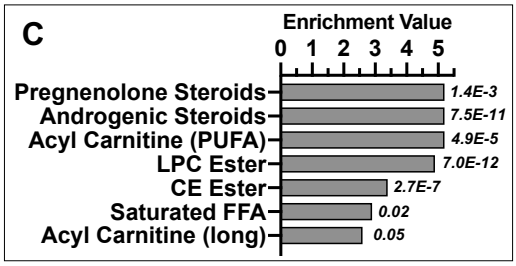
Figure 1

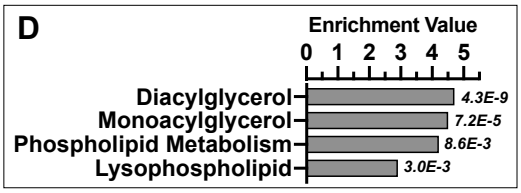
A Plasma

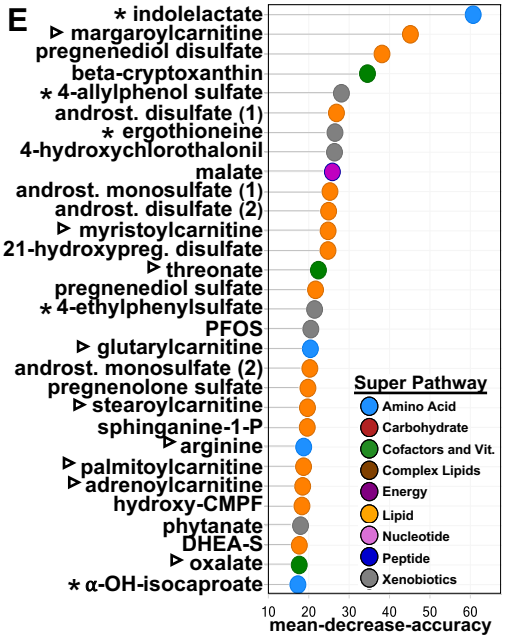
Compared to TD	↑	↓	Total
All ASD	62	248	310
ASD-GI	72	182	254
ASD+GI	54	206	260

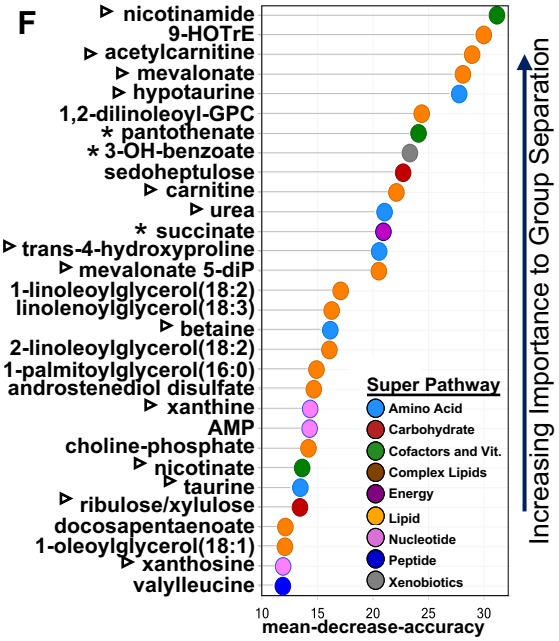
B Feces

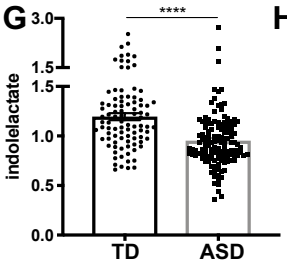
Compared to TD	↑	↓	Total
All ASD	25	87	112
ASD-GI	28	53	81
ASD+GI	14	66	80

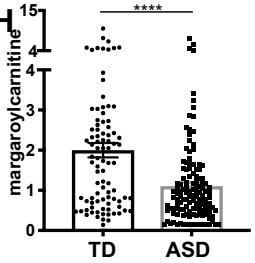


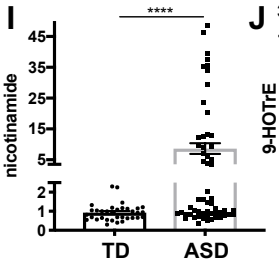


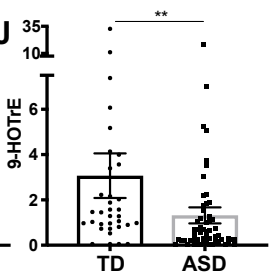


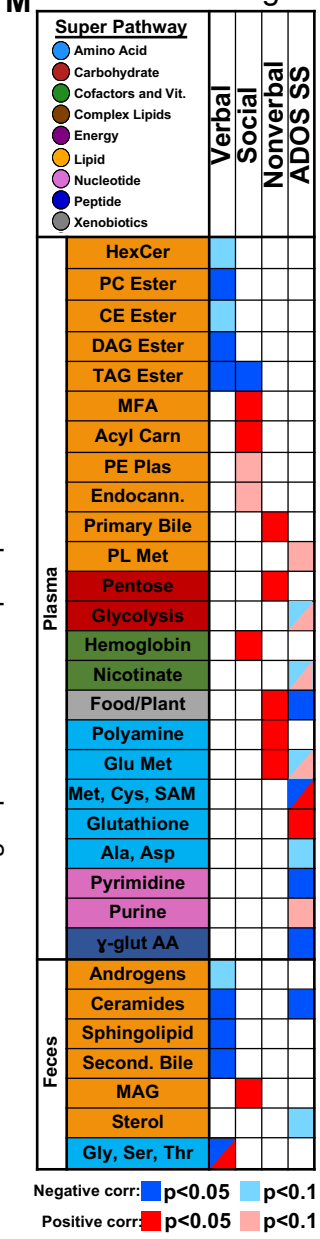












K

	Correlated Metabolite	r	pval
Margaroylcarnitine	myristoylcarnitine (C14)	0.88	****
	stearoylcarnitine (C18)	0.88	****
	palmitoylcarnitine (C16)	0.87	****
	palmitoleoylcarnitine (C16:1)	0.81	****
	oleoylcarnitine (C18:1)	0.79	****
	cysteinylglycine	0.78	****
	LPC(17:0)	0.77	****
	dihomo-linoleoylcarnitine (C20:2)	0.77	****
Indolelactate	phenyllactate (PLA)	0.51	****
	3-(4-hydroxyphenyl)lactate	0.51	****
	alpha-hydroxyisovalerate	0.48	****
	2-hydroxy-3-methylvalerate	0.46	****
	N-acetyltryptophan	0.41	****
	kynurenate	0.40	****
	imidazole lactate	0.38	****
	LPC(14:0)	0.38	****

L

	Correlated Metabolite	r	pval
Nicotinamide	hypotaurine	0.74	****
	choline phosphate	0.67	****
	glycerophosphorylcholine (GPC)	0.63	****
	ergothioneine	0.57	****
	acetylcarnitine (C2)	0.56	****
	glycerol 3-phosphate	0.56	****
	carnitine	0.53	****
	adenosine 5'-monophosphate (AMP)	0.53	****
9-HOTrE	1-linolenoylglycerol (18:3)	0.68	****
	diacylglycerol (16:1/18:2, 16:0/18:3)	0.67	****
	13-HODE + 9-HODE	0.66	****
	palmitoyl-linolenoyl-glycerol (16:0/18:3)	0.63	****
	glycerol	0.57	****
	oleoyl-linoleoyl-glycerol (18:1/18:2)	0.57	****
	palmitoyl-linoleoyl-glycerol (16:0/18:2)	0.56	****
	enterodiol	0.54	****

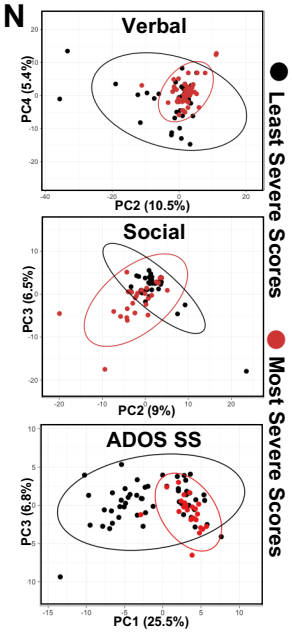


Figure 2

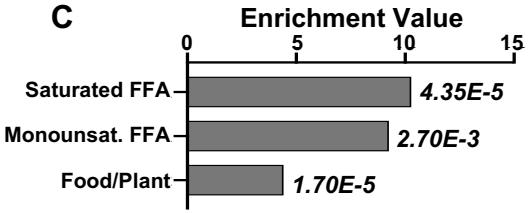
A

Plasma	↑	↓	Total
ASD+GI / ASD-GI	30	57	87

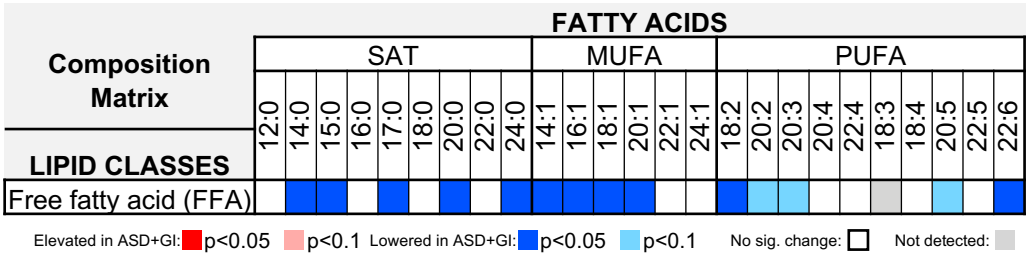
B

Feces	↑	↓	Total
ASD+GI / ASD-GI	3	21	24

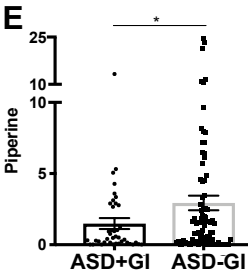
C



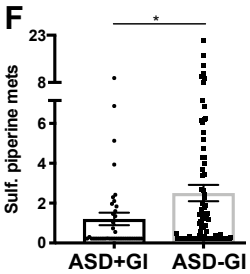
D



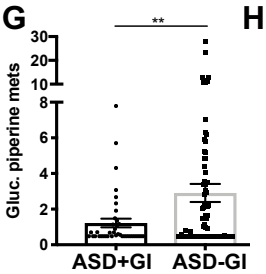
E



F



G



H

Mets	Corr	P-value
Pip	-0.12	<0.01
Pip-sulf	-0.15	<0.01
Pip-gluc	-0.14	<0.01

Figure 3

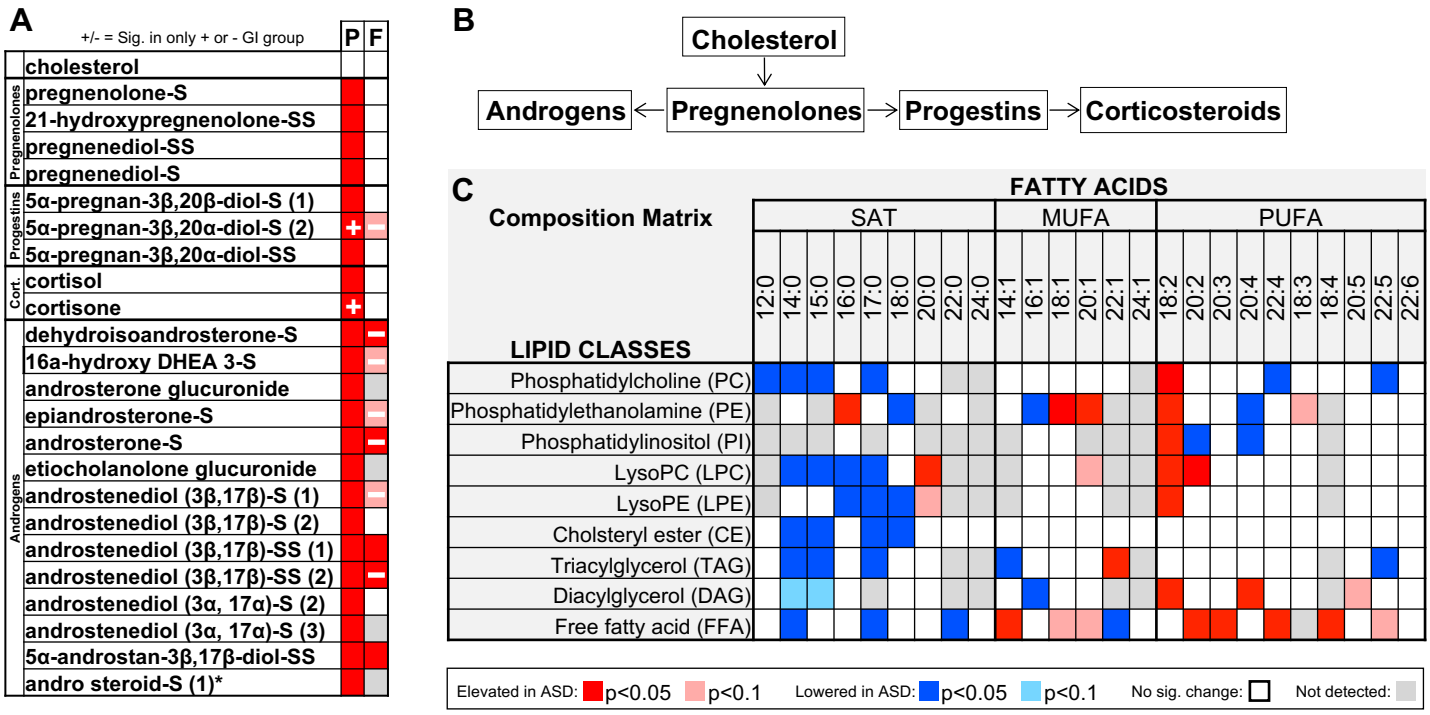
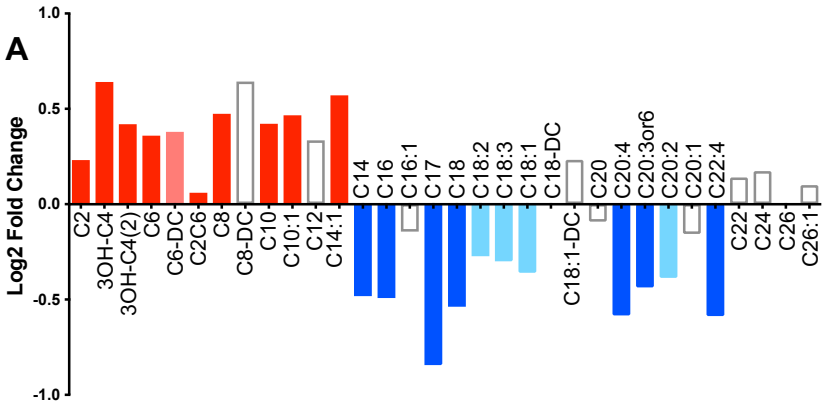
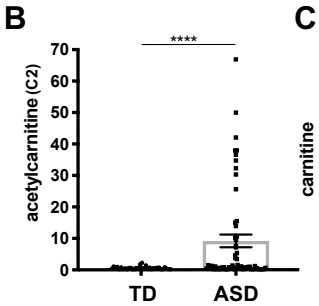
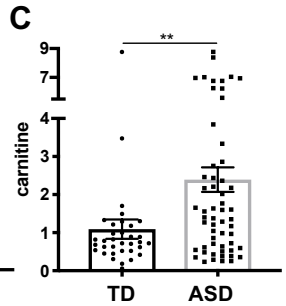
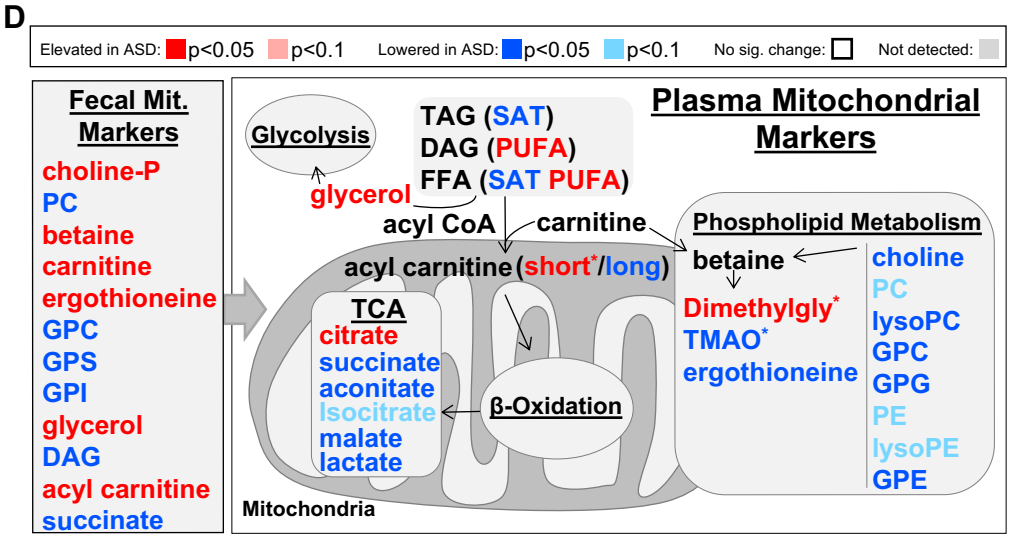

A
+/- = Sig. in only + or - GI group
P
F
cholesterol
Pregnenolones
pregnenolone-S
21-hydroxypregnenolone-SS
pregnenediol-SS
pregnenediol-S
Progestins
5α-pregnan-3β,20β-diol-S (1)
5α-pregnan-3β,20α-diol-S (2)
5α-pregnan-3β,20α-diol-SS
Cort.
cortisol
cortisone
Androgens
dehydroisoandrosterone-S
16a-hydroxy DHEA 3-S
androsterone glucuronide
epiandrosterone-S
androsterone-S
etiocholanolone glucuronide
androstenediol (3β,17β)-S (1)
androstenediol (3β,17β)-S (2)
androstenediol (3β,17β)-SS (1)
androstenediol (3β,17β)-SS (2)
androstenediol (3α, 17α)-S (2)
androstenediol (3α, 17α)-S (3)
5α-androstan-3β,17β-diol-SS
andro steroid-S (1)*
B
Cholesterol
Androgens
Pregnenolones
Progestins
Corticosteroids
C
Composition Matrix
FATTY ACIDS
SAT
MUFA
PUFA
12:0
14:0
15:0
16:0
17:0
18:0
20:0
22:0
24:0
14:1
16:1
18:1
20:1
22:1
24:1
18:2
20:2
20:3
20:4
22:4
18:3
18:4
20:5
22:5
22:6
LIPID CLASSES
Phosphatidylcholine (PC)
Phosphatidylethanolamine (PE)
Phosphatidylinositol (PI)
LysoPC (LPC)
LysoPE (LPE)
Cholsteryl ester (CE)
Triacylglycerol (TAG)
Diacylglycerol (DAG)
Free fatty acid (FFA)
Elevated in ASD: p<0.05 p<0.1
Lowered in ASD: p<0.05 p<0.1
No sig. change:
Not detected:

Figure 4







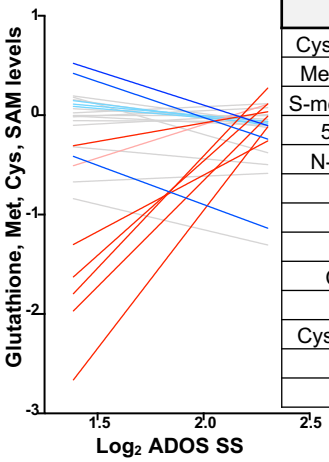


E

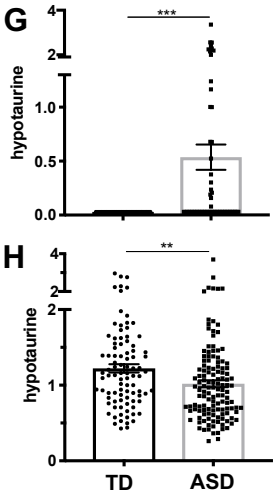
	P	F
N-acetylmethionine	+	
N-formylmethionine	+	
S-methylmethionine	−	
SAH		
cysteine		
cystathionine	+	
S-methylcysteine	−	
S-methylcys sulfoxide		
hypotaurine		
cysteine-S-sulfate	+	
taurine		
lanthionine		
N-acetyltaurine		
cystine	+	−
5-oxoproline		
2-aminobutyrate		
cysteinylglycine		

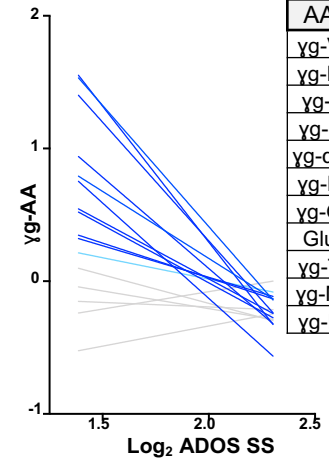
+/- = Sig. in only + or - GI group

F



Metabolite	Corr	pval
Cysteine sulfinic acid	-0.25	.005
Methionine sulfoxide	-0.23	.009
S-methylcys sulfoxide	-0.20	.027
5-methylthioribose	-0.17	.051
N-formylmethionine	-0.15	.090
2-aminobutyrate	-0.15	.097
cysteinylglycine	0.17	.052
hypotaurine	0.20	.028
Cysteine s-sulfate	0.21	.020
Cysteine-GSSG	0.29	.001
Cysteinylgly disulfide	0.30	.001
cystine	0.31	.001
Cys-gly, oxidized	0.32	3E-4





AA	Corr	pval
γg-V	-0.36	3E-5
γg-H	-0.32	2E-4
γg-I	-0.31	3E-4
γg-L	-0.31	4E-4
γg-αL	-0.30	7E-4
γg-E	-0.29	.001
γg-G	-0.26	.003
Glu	-0.25	.004
γg-T	-0.20	.025
γg-M	-0.20	.026
γg-F	-0.15	.083

Figure 5

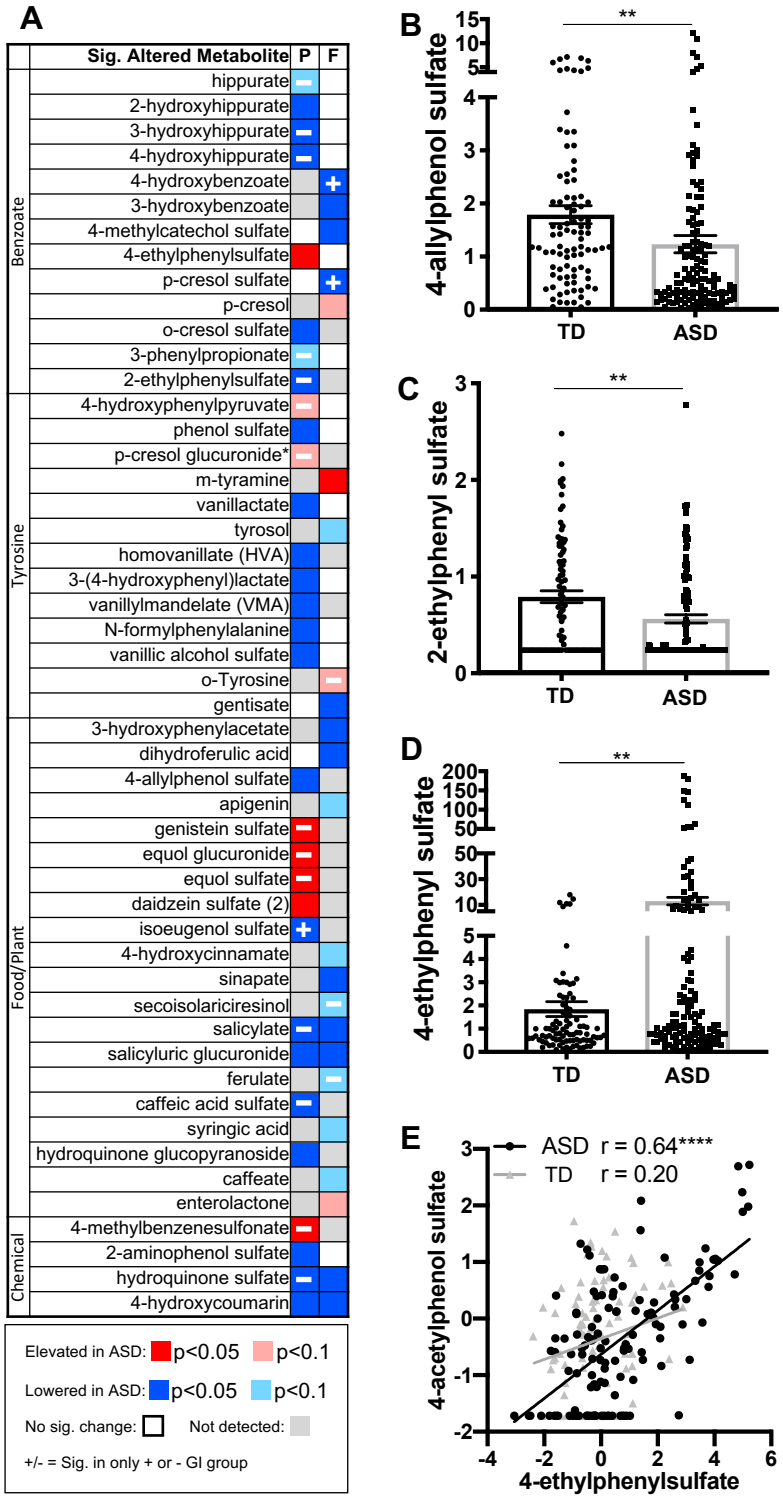

A
Sig. Altered Metabolite
P
F
Benzoate
hippurate
2-hydroxyhippurate
3-hydroxyhippurate
4-hydroxyhippurate
4-hydroxybenzoate
3-hydroxybenzoate
4-methylcatechol sulfate
4-ethylphenylsulfate
p-cresol sulfate
p-cresol
o-cresol sulfate
3-phenylpropionate
2-ethylphenylsulfate
Tyrosine
4-hydroxyphenylpyruvate
phenol sulfate
p-cresol glucuronide*
m-tyramine
vanillactate
tyrosol
homovanillate (HVA)
3-(4-hydroxyphenyl)lactate
vanillylmandelate (VMA)
N-formylphenylalanine
vanillic alcohol sulfate
o-Tyrosine
gentisate
Food/Plant
3-hydroxyphenylacetate
dihydroferulic acid
4-allylphenol sulfate
apigenin
genistein sulfate
equol glucuronide
equol sulfate
daidzein sulfate (2)
isoeugenol sulfate
4-hydroxycinnamate
sinapate
secoisolariciresinol
salicylate
salicyluric glucuronide
ferulate
caffeic acid sulfate
syringic acid
hydroquinone glucopyranoside
caffeate
enterolactone
Chemical
4-methylbenzenesulfonate
2-aminophenol sulfate
hydroquinone sulfate
4-hydroxycoumarin
Elevated in ASD: p<0.05 p<0.1
Lowered in ASD: p<0.05 p<0.1
No sig. change:
Not detected:
+/- = Sig. in only + or - GI group
B
4-allylphenol sulfate
**
TD
ASD
C
2-ethylphenyl sulfate
**
TD
ASD
D
4-ethylphenyl sulfate
**
TD
ASD
E
ASD r = 0.64****
TD r = 0.20
4-acetylphenol sulfate
4-ethylphenylsulfate

Figure 6

	Metabolite(s)	Core Findings
Lipids	PC, DAG, and TAG acyl esters	Pathways significantly enriched; Negative correlations with verbal score severity
	Steroid hormones	Pathways significantly enriched; hormone levels elevated
	PUFAs	Elevated levels in DAG, FFA, and many with chain length 18:2
	Saturated FAs	Pathways significantly enriched; broad decrease in levels
	Free FAs	Decreased levels in ASD+GI
Markers of Mit. Function	Acyl-carnitines	Pathways significantly enriched; shorter chain moieties elevated, longer chain moieties decreased; positive correlation with social score severity; acetyl- and free carnitine elevated in feces
	Phospholipid metabolism	Pathway significantly enriched in feces
	TCA cycle	Increase in citrate; decrease in succinate, aconitate, malate, lactate levels
	Pathways rel. to oxidative stress (Cys, Met., SAM; glutathione; gg-AA metabolites)	Altered levels in ASD; Significant correlations with nonverbal scores and ADOS SS
Xenobiotic and Phenolic Metabolites	Metabolites of Food and Plant Components	Positive correlations to nonverbal scores and negative correlations to ADOS SS; Pathway significantly enriched in ASD+GI
	Phenolic structure mets.: 4-ethylphenyl sulfate (4EPS), genistein-S, equol mets., daidzein sulfate, 4-methylbenzene sulfonate	Elevated in ASD plasma samples
	Phenolic structure mets.: hippurate and benzoic acid mets., some tyrosine mets., salicylic mets.	Decreased in ASD plasma samples
	Phenolic structure mets.: equol-S, 3-phenylpropionate, hippurate, 4-hydroxyhippurate, caffeic acid-S, phenol-S, daidzein-S, genistein-S	Display shifted levels in mouse recipients of human ASD fecal samples compared to TD fecal samples
Potential biomarkers	Acyl-carnitines and lipids	Rank in the top 30 most discriminatory metabolites, assessed by Random Forest analysis
	Metabolites known to be influenced by the gut microbiota	Rank in the top 30 most discriminatory metabolites, assessed by Random Forest analysis
	Indolelactate, margaroylcarnitine	Most discriminatory plasma metabolites
	Nicotinamide, 9-HOTrE	Most discriminatory fecal metabolites

Supplementary figure 1.

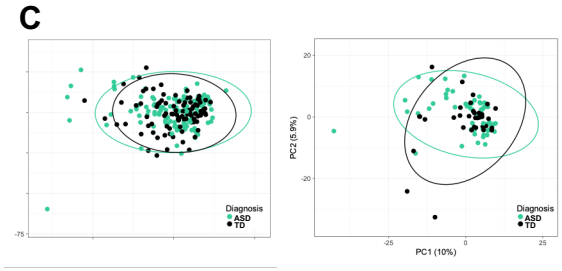
A

Plasma	Sex		GI Symptoms			Age range (mos)	Mean Age (mos)	Total N
Group	Male	Female	+GI	-GI	ND			
TD	89	12	0	90	2	39-151	80	101
ASD	110	20	40	86	4	37-153	88	130

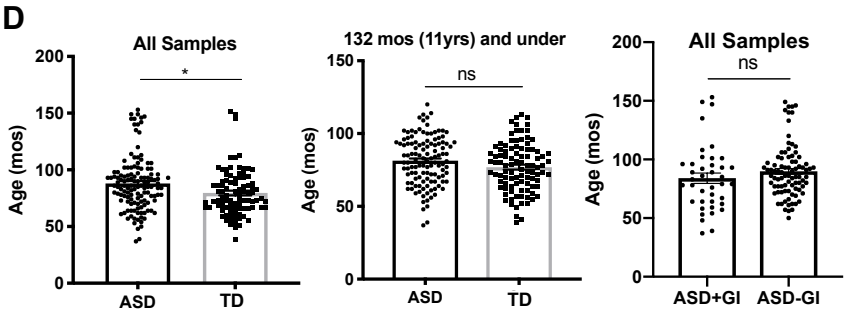
B

Feces	Sex			GI Symptoms			Age range (mos)	Mean Age (mos)	Total N
Group	Male	Female	ND	+GI	-GI	ND			
TD	36	3	1	0	33	1	39-151	86	40
ASD	48	7	2	22	31	4	37-149	91	57

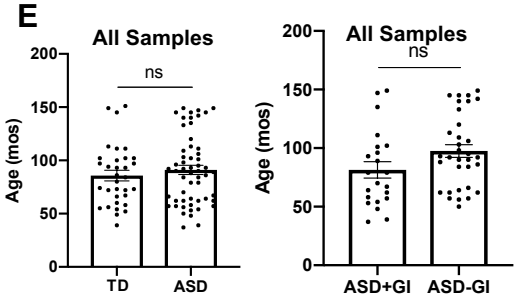
C



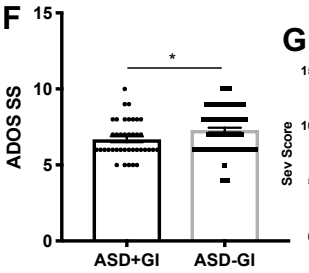
D



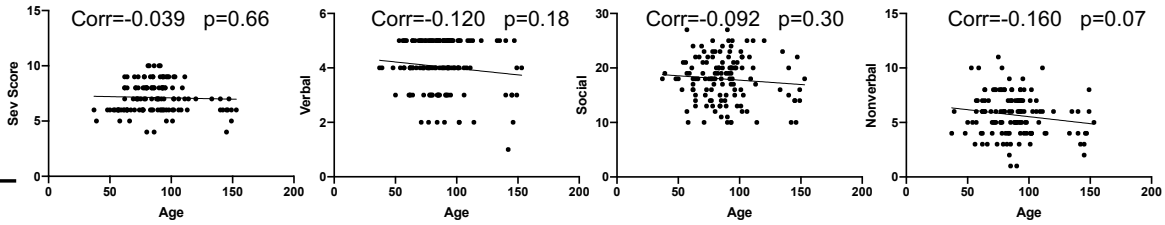
E



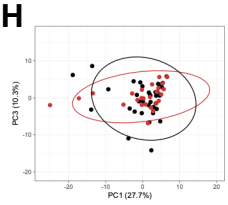
F



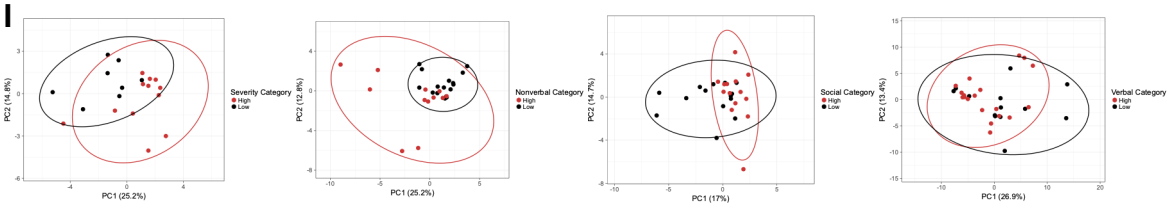
G

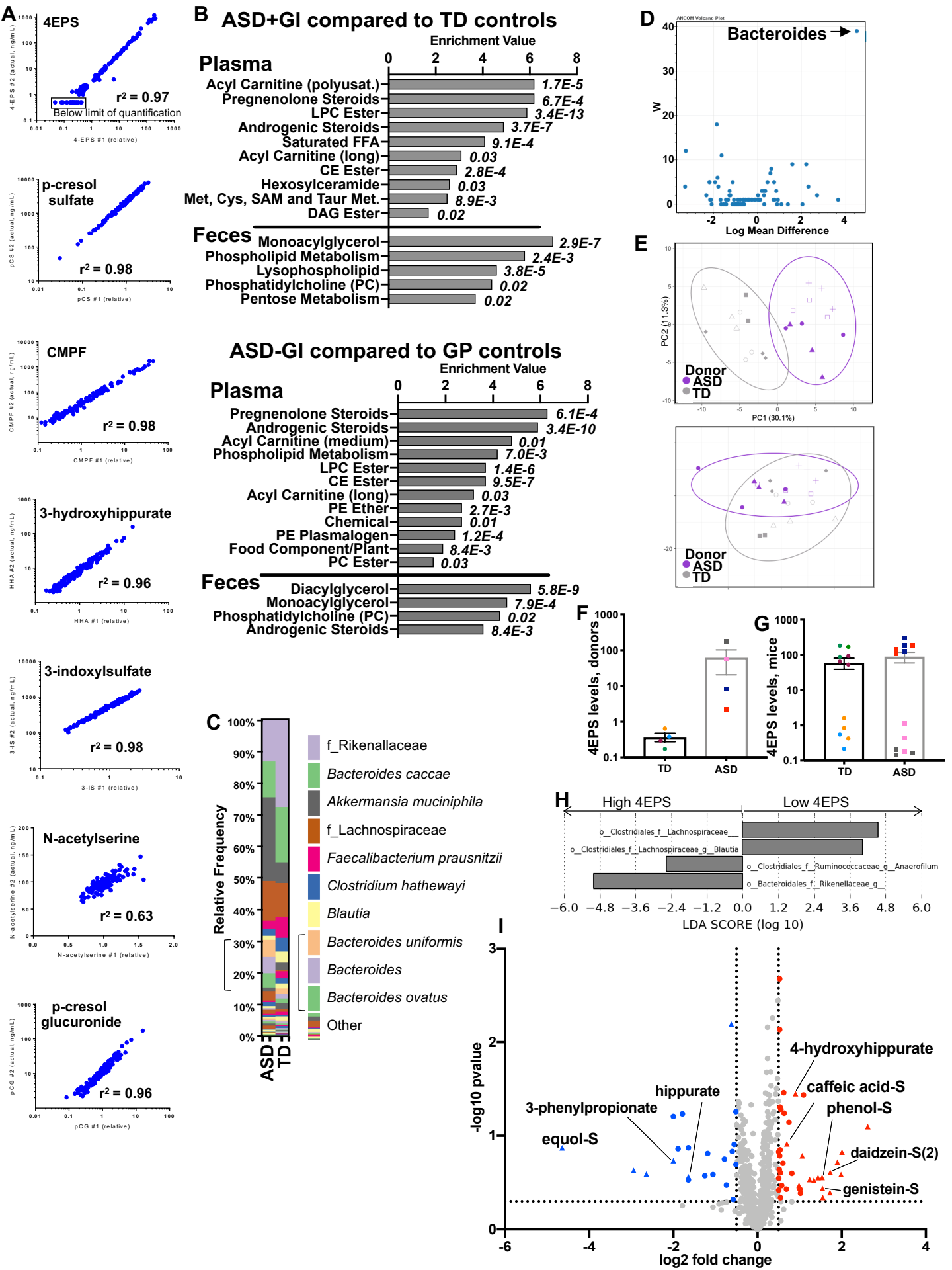


H

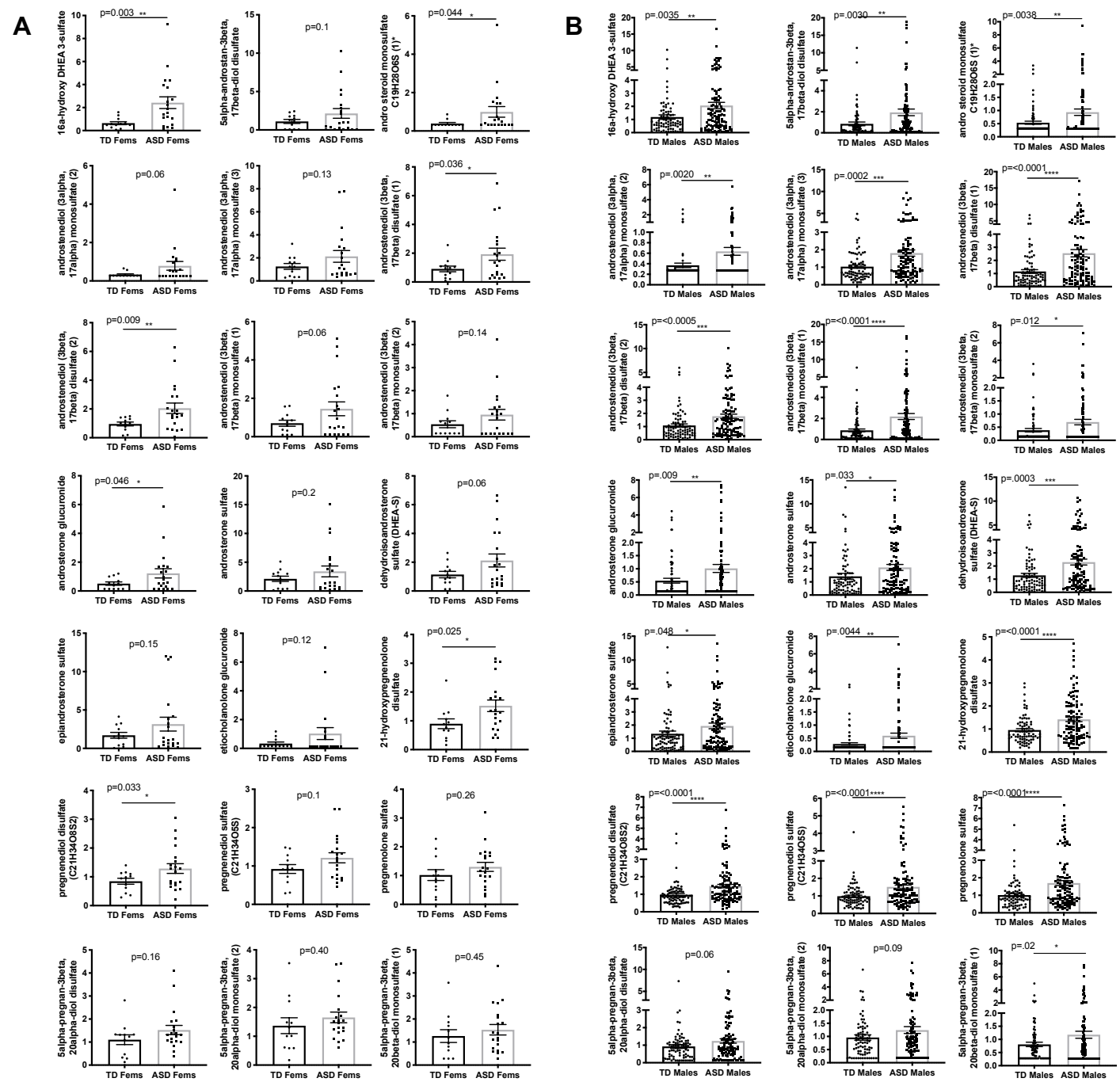


I




A
4EPS
r² = 0.97
Below limit of quantification
4-EPS #2 (actual, ng/mL)
4-EPS #1 (relative)
p-cresol sulfate
r² = 0.98
pCS #2 (actual, ng/mL)
pCS #1 (relative)
CMPF
r² = 0.98
CMPF #2 (actual, ng/mL)
CMPF #1 (relative)
3-hydroxyhippurate
r² = 0.96
HHA #2 (actual, ng/mL)
HHA #1 (relative)
3-indoxylsulfate
r² = 0.98
3-IS #2 (actual, ng/mL)
3-IS #1 (relative)
N-acetylserine
r² = 0.63
N-acetylserine #2 (actual, ng/mL)
N-acetylserine #1 (relative)
p-cresol glucuronide
r² = 0.96
pCG #2 (actual, ng/mL)
pCG #1 (relative)
B
ASD+GI compared to TD controls
Enrichment Value
Plasma
Acyl Carnitine (polyusat.) 1.7E-5
Pregnenolone Steroids 6.7E-4
LPC Ester 3.4E-13
Androgenic Steroids 3.7E-7
Saturated FFA 9.1E-4
Acyl Carnitine (long) 0.03
CE Ester 2.8E-4
Hexosylceramide 0.03
Met, Cys, SAM and Taur Met. 8.9E-3
DAG Ester 0.02
Feces
Monoacylglycerol 2.9E-7
Phospholipid Metabolism 2.4E-3
Lysophospholipid 3.8E-5
Phosphatidylcholine (PC) 0.02
Pentose Metabolism 0.02
ASD-GI compared to GP controls
Enrichment Value
Plasma
Pregnenolone Steroids 6.1E-4
Androgenic Steroids 3.4E-10
Acyl Carnitine (medium) 0.01
Phospholipid Metabolism 7.0E-3
LPC Ester 1.4E-6
CE Ester 9.5E-7
Acyl Carnitine (long) 0.03
PE Ether 2.7E-3
Chemical 0.01
PE Plasmalogen 1.2E-4
Food Component/Plant 8.4E-3
PC Ester 0.03
Feces
Diacylglycerol 5.8E-9
Monoacylglycerol 7.9E-4
Phosphatidylcholine (PC) 0.02
Androgenic Steroids 8.4E-3
C
Relative Frequency
ASD
TD
f_Rikenallaceae
Bacteroides caccae
Akkermansia muciniphila
f_Lachnospiraceae
Faecalibacterium prausnitzii
Clostridium hathewayi
Blautia
Bacteroides uniformis
Bacteroides
Bacteroides ovatus
Other
D
ANCOM Volcano Plot
Bacteroides
W
Log Mean Difference
E
PC2 (11.3%)
PC1 (30.1%)
Donor
ASD
TD
F
4EPS levels, donors
TD
ASD
G
4EPS levels, mice
TD
ASD
H
High 4EPS
Low 4EPS
o__Clostridiales_f__Lachnospiraceae__
o__Clostridiales_f__Lachnospiraceae_g__Blautia
o__Clostridiales_f__Ruminococcaceae_g__Anaerofilum
o__Bacteroidales_f__Rikenellaceae_g__
LDA SCORE (log 10)
I
-log10 pvalue
log2 fold change
4-hydroxyhippurate
caffeic acid-S
phenol-S
hippurate
3-phenylpropionate
equol-S
daidzein-S(2)
genistein-S

Supplementary figure 3.

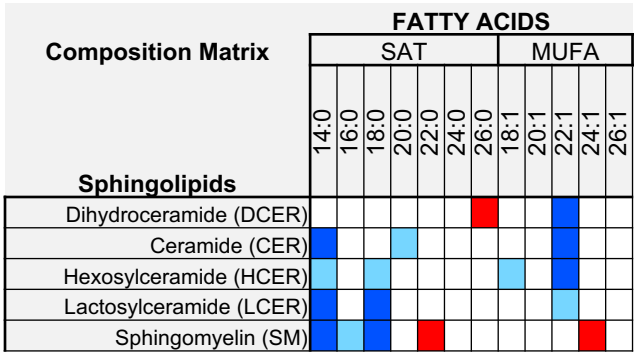


C

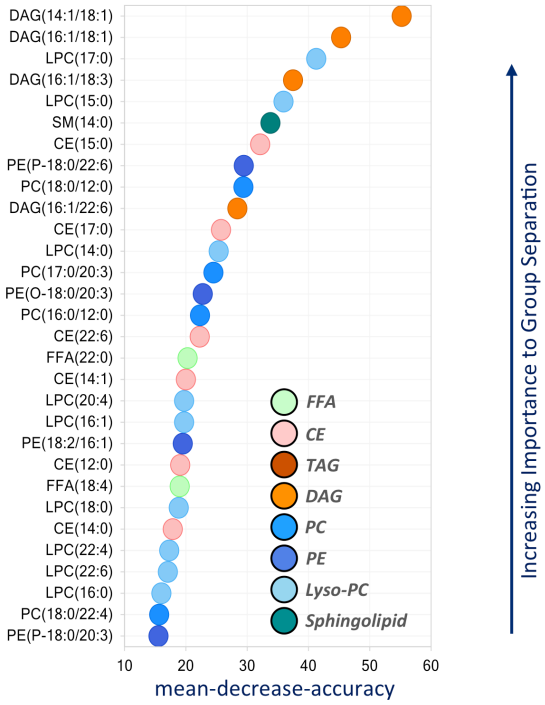
			Pregnenolones				Progestins			Androgens														
			21-hydroxypregnenolone disulfate	pregnenediol sulfate	pregnenediol	pregnenolone sulfate	5alpha-pregnan-3beta,20alpha-disulfate	5alpha-pregnan-3beta,20alpha-monosulfate	5alpha-pregnan-3beta,20beta-diol	16a-hydroxy DHEA 3-sulfate	5alpha-androstan-3beta,17beta-diol disulfate	andro steroid monosulfate	androstenediol (3alpha, 17alpha)	androstenediol (3alpha, 17alpha) monosulfate (2)	androstenediol (3beta,17beta) monosulfate (1)	androstenediol (3beta,17beta) monosulfate (2)	androstenediol (3beta,17beta) monosulfate (3)	androstenediol (3beta,17beta) disulfate (1)	androstenediol (3beta,17beta) disulfate (2)	androsterone glucuronide	androsterone sulfate	dehydroisoandrosterone sulfate	epiandrosterone sulfate	etiocholanolone glucuronide
3-8	36-107	110/95	1.4	1.3	1.3	1.4	1.2	1.1	1.2	1.7	1.9	1.7	1.5	1.4	1.9	1.5	2	1.5	1.7	1.3	1.5	1.3	2.1
9-12	108-153	19/6	1.8	2.3	1.9	2.6	2.6	2	1.9	2.1	1.5	1.9	1.7	2.2	2.1	1.5	2.1	1.7	1.3	1.2	1.7	1.2	1.5
3-4	36-59	12/15	1.2	1.6	1.3	1.4	2.2	1.6	1.5	1.8	7	1	1	2.3	3.8	1.9	3.7	1.3	1.6	3.8	3.1	4	1.8
5-6	60-83	18/26	1.4	1.4	1.4	1.5	1.3	1.2	1.7	1.7	1.2	1.5	1.3	0.8	1.2	1.1	1.2	1.1	1.5	0.8	1	0.8	1.1
7-8	84-107	54/28	1.5	1.4	1.4	1.4	1	1	1.2	1.8	1.6	1.8	1.9	1.4	1.6	1.5	1.7	1.5	2.1	1.2	1.4	1.2	2.1
9-11	108-143	13/3	2.2	3	2.2	2.4	1.9	1.3	1.9	3.4	4.1	4	4.5	3.3	4.7	2.8	3.5	11	2	2.4	3	2.3	1.8
12	145-153	7/3	1.8	2.2	2	3.2	3.8	3.4	2.2	2	1.3	1.8	1.5	2.1	1.7	1.3	2.1	1.2	1.1	0.9	1.6	1	1.4

Supplementary figure 4.

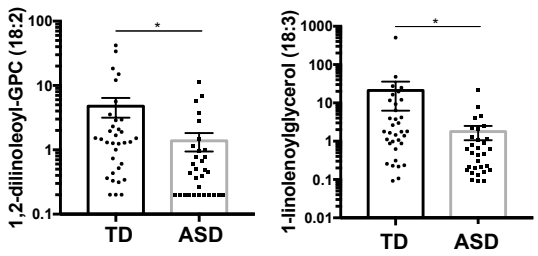
A



B



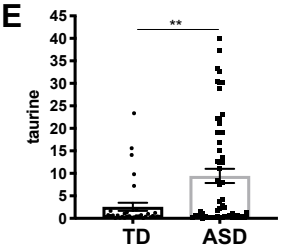
C



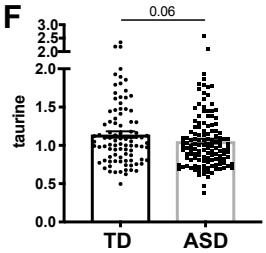
D

Met	Corr.	p
isocitrate	0.32	0.0002
3-phosphoglycerate	0.33	0.0002
phosphoethanolamine	0.33	0.0001
choline-phosphate	0.34	0.0001
glycerol 3-phosphate	0.35	0.0001

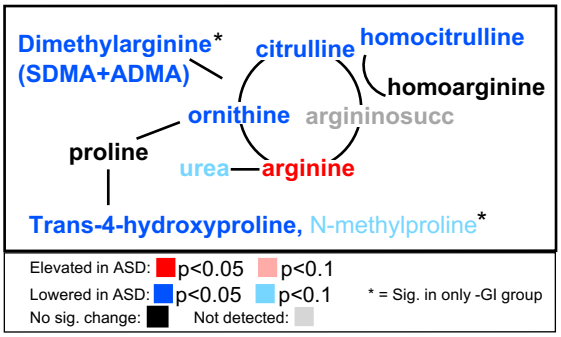
E



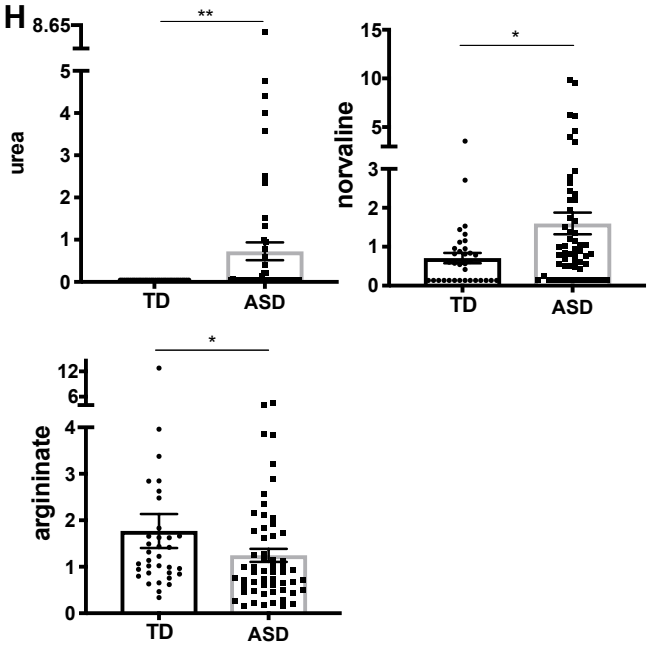
F



G

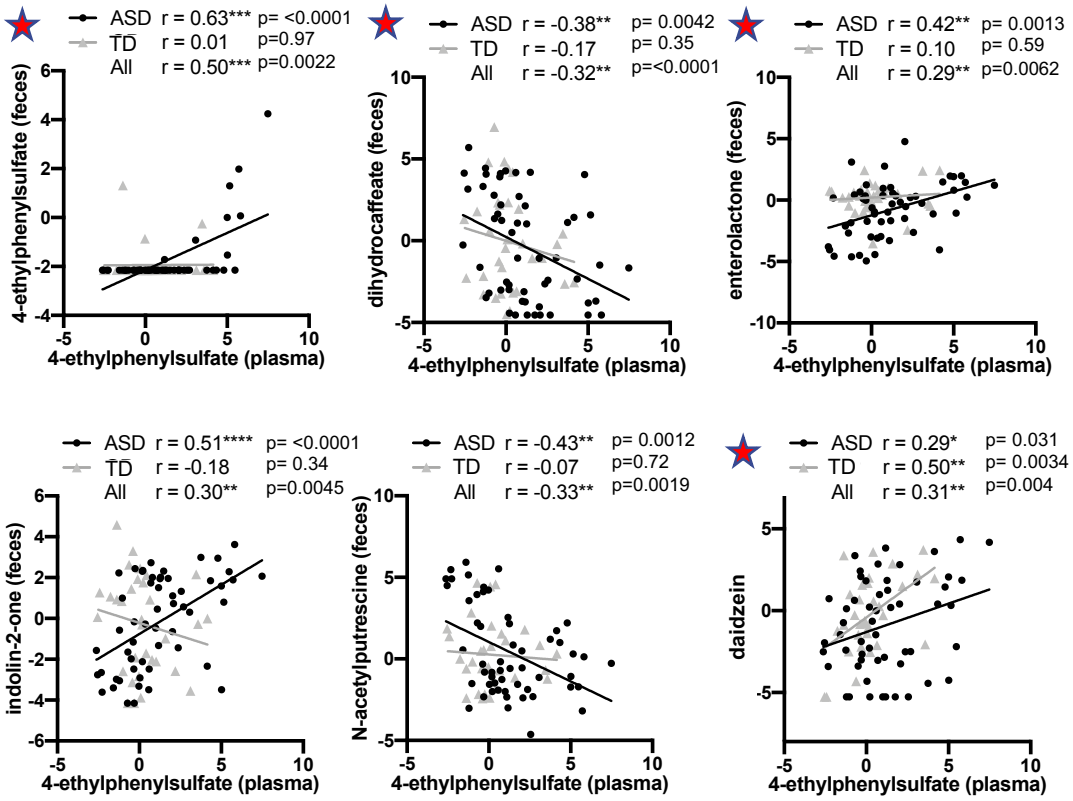


H



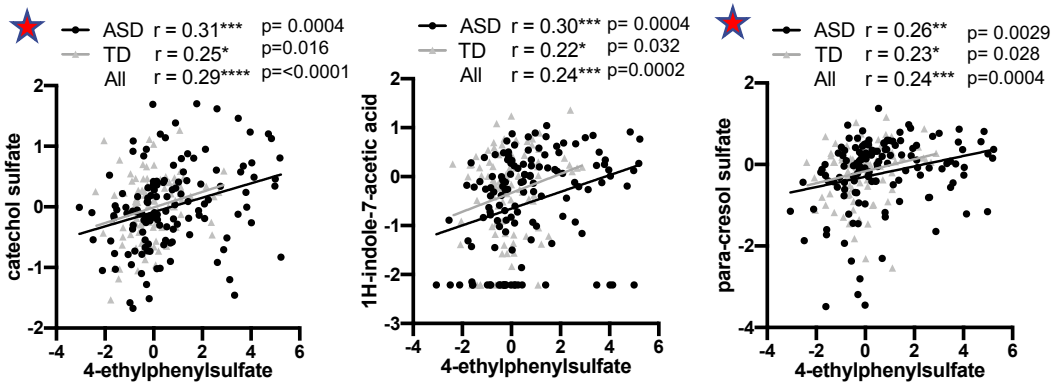
A

Nearest neighbor analysis: 4EPS in plasma correlated to fecal metabolites



B

Nearest neighbor analysis: 4EPS in plasma correlated to plasma metabolites



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPLEMENTAL METHODS

Extended Participant Information

Briefly, the GISS consists of 7 sections, each section asking a series of 1 to 6 questions to determine if the participant met the criteria for diarrhea, constipation or irritable bowel syndrome (IBS)- like symptoms, and if stooling/symptoms have been consistent for the last 6 months. Any individual with antibiotic use within 3 months was excluded, but multivitamins and over the counter supplements were permitted. Based on their responses, participants were placed in one of four groups, ASD+GI, ASD-GI, TD+GI or TD-GI. Due to the low incidence of GI issues in typically developing children enrolled in the study, we only have 9 individuals in the TD+GI groups. Our original statistical analysis resulted in many seemingly intriguing metabolites with differential levels between the ASD+GI and TD+GI samples. However, upon closer inspection, most of these differences were driven by the fact that the TD+GI group (with a small n) was very different from all other samples rather than the ASD+GI being unique. There were occasional differences that were indeed convincing between the ASD+GI and TD+GI, but these differences also arose in the comparison of -GI samples, so they were not specific to the GI phenotype and presenting them in an ASD+GI compared to TD+GI context would have been misleading. Thus, the TD+GI group was removed from the analysis and is not presented here.

This study was approved by institutional review boards for the State of California and the University of California, Davis. Informed consent is obtained from a legal guardian for all study participants prior to data collection in accordance with the UC Davis IRB protocol.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Blood and stool collection: Peripheral blood was collected from participants in acid-citrate dextrose Vacutainers (BD Biosciences; San Jose, Ca). Blood was centrifuged at 2100 rpm for 10 minutes followed by the collection of plasma into cryovials. Plasma was stored at -80 °C until analysis. In addition to blood, stool samples were also obtained, parents were given collection containers with RNA later to collect stool samples at home and asked to store the samples in the freezer and brought back frozen to the clinic within 24 hrs.

Metabolite analysis: All metabolite analysis, identification, quality control were performed by standard procedures at Metabolon Inc. as follows.

Sample Preparation: All samples were maintained at -80ºC until processed. Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

QA/QC: Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS): All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Bioinformatics: The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

LIMS: The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data Extraction and Compound Identification: Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Complex Lipids Platform: Lipids were extracted from samples in methanol:dichloromethane in the presence of internal standards. The extracts were concentrated under nitrogen and reconstituted in 0.25mL of 10mM ammonium acetate dichloromethane:methanol (50:50). The extracts were transferred to inserts and placed in vials for infusion-MS analysis, performed on a Shimazdu LC with nano PEEK tubing and the Sciex SelexIon-5500 QTRAP. The samples were analyzed via both positive and negative mode electrospray. The 5500 QTRAP scan was performed in MRM mode with the total of more than 1,100 MRMs. Individual lipid species were quantified by taking the peak area ratios of target compounds and their assigned internal standards, then multiplying by the concentration of internal standard added to the sample. Lipid class concentrations were calculated from the sum of all molecular species within a class, and fatty acid compositions were determined by calculating the proportion of each class comprised by individual fatty acids.

Peak Quantification, Normalization and Statistical analysis: Peaks were quantified using area-under-the-curve. The present dataset comprises a total of 1611 and 814 compounds of known identity for the human plasma and fecal samples, respectively. The mouse dataset comprises a total of 746 known biochemicals. Following log transformation and imputation of missing values, if any, with the minimum observed value for each compound, ANOVA contrasts using ArrayStudio (Qiagen) were used to identify biochemicals that differed significantly between experimental groups. Analysis by two-way ANOVA identified biochemicals exhibiting significant interaction and main effects for experimental parameters of disease and GI symptoms. For mouse samples, in addition to two-way ANOVA contrasts, in order to account for non-independence (match to donor), we ran a linear mixed effect model where the donor match is a

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

random effect. The analysis was done in R.gui version 3.5.1. Correlation analysis was performed comparing the relationship between observed changes in metabolite levels and measures of severity. An estimate of the false discovery rate (q-value) is calculated to take into account the multiple comparisons that normally occur in metabolomic-based studies. In all graphs of individual metabolites, the significance noted in the graph is based on p-value from the ANOVA and can be found in the corresponding Supplemental Tables along with the q-values.

Principle Components Analysis (PCA): All PCA plots were generated using the ClustVis web tool. In cases where ASD samples were subdivided by severity, samples were separated and grouped by behavioral score as close to quartiles as possible without splitting samples with the same score into separate groups. The groupings were as follows: for human feces scores for lower and higher severity groups, respectively: ADOS scores 4-5 and 8-10; nonverbal scores 1-4 and 7-10, social scores 10-14 and 20-27, verbal scores 1-3 and 5. Groupings for human plasma scores for lower and higher severity groups, respectively: ADOS scores 4-5 and 9-10; nonverbal scores 1-4 and 8-11, social scores 10-14 and 22-27, verbal scores 1-3 and 5.

Random Forest Analysis: Random Forest Analysis was performed as described previously [1]. A random subset of the data with identifying true class information was selected to build the tree ("bootstrap sample" or "training set"), and then the remaining data, the "out-of-bag" (OOB) variables, were passed down the tree to obtain a class prediction for each sample. This process was repeated thousands of times to produce the forest. The final classification of each sample was determined by computing the class prediction frequency for the OOB variables over the whole forest. When the full forest is grown, the class predictions were compared to the true

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

classes, generating the "OOB error rate" as a measure of prediction accuracy. The mean decrease in accuracy was determined by randomly permuting a variable, running the observed values through the trees, and then reassessing the prediction accuracy. Thus, the Random Forest analysis provides an "importance" rank ordering of biochemicals.

The further Random Forest using the top 30 metabolites was performed by first leaving out one sample at a time, fitting the Random Forest, keeping the top 30 metabolites, then using the top 30 to predict the observation that was left out. This was repeated for each sample and then the actual result was compared to the predicted and error was calculated.

Pathway Enrichment Analysis: For each individual pair-wise comparison, pathway enrichment determined the number of statistically significantly different compounds relative to all detected compounds in a sub-pathway, compared to the total number of statistically significantly different compounds relative to all detected compounds in the study. A pathway enrichment value greater than one indicates that the pathway contains more significantly changed compounds relative to the study overall, suggesting that the pathway may be a target of interest for further investigation. Enrichment Value = (# of significant metabolites in pathway(k) / total # of detected metabolites in pathway(m)) / (total # of significant metabolites(n) / total # of detected metabolites(N)) (k/m)/(n/N).

Animal husbandry and sample collection: All mouse housing and experiments were approved by the California Institute of Technology IACUUC. Fecal samples were thawed, 300mls of 1.5% bicarbonate in PBS was added, samples were vortexed, and then allowed to settle. 150uL of fecal sample was then delivered by oral gavage to 5-week-old germ free mice that were then

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

maintained in sterile, microisolator cages for 3 weeks. After three weeks, fecal samples were collected for 16s sequencing. Blood was collected by cardiac puncture following euthanasia by CO_2, plasma was isolated on ice using EDTA-treated collection tubes (Thermo), and samples were stored at -80C until analysis. One replicate GP sample was thawed prematurely and thus removed from further analysis.

Fecal analysis for 16s sequencing: Fecal samples collected and immediately put into empty sterile tubes, flash frozen, and maintained at -80C until processing. Total DNA was isolated with Qiagen DNeasy powersoil extraction kit (Qiagen) following manufacturer instructions. Hypervariable V4 region of the 16s gene was amplified by PCR using 5Xprime master mix (Prime). Barcoded 806 reverse primers and unique forward 515 primer (IDT) were used as previously described. The amplification was confirmed by electrophoresis and the amplified products were purified with Quiaquick PCR purification kit (Qiagen).

Samples were sent to MGH NGS Core facility to be sequenced on the Illumina MiSeq instrument using MiSeq v2 500-cycle sequencing kit, resulting in approximately 25 million paired-end 250 bp reads covering amplicon regions. Data were analyzed QIIME2 software package at the Bioinformatic core at MGH. Low quality score sequencing reads (average $Q < 25$) were truncated to 240bp followed by filtering using *deblur* algorithm with default setting and the high quality reads were aligned to the reference library using *mafft*. The aligned reads were masked to remove highly variable positions, and a phylogenetic tree was generated from the masked alignment using the FastTree method. Alpha and beta diversity metrics and Principal Component Analysis plots based on Jaccard distance were generated using default QIIME2

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.17.098806; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plugins. Taxonomy assignment was performed using *feature-classifier* method and naïve Bayes classifier trained on the Greengenes 13_8 99% operational taxonomic units (OTUs). Linear discriminant analysis Effect Size (LEfSe) was performed as described previously[2] using the Galaxy web application.

References

1. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
2. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biology* **12**, R60 (2011).