Multi-omic analysis along the gut-brain axis points to a functional architecture of autism

James T. Morton[1,2], Dong-Min Jin[3], Robert H. Mills[4], Yan Shao[5], Gibraan Rahman[6, 7], Kirsten Berding[8], Brittany D. Needham[9], María Fernanda Zurita[10], Maude David[11], Olga V. Averina[12], Alexey S. Kovtun[12, 13], Antonio Noto[14], Michele Mussap[15], Mingbang Wang[16], Daniel N. Frank[17], Ellen Li[18], Wenhao Zhou[16], Vassilios Fanos[19], Valery N. Danilenko[12], Dennis P. Wall[20], Paúl Cárdenas[21], Manuel E. Baldeón[22], Ramnik J. Xavier[23, 24, 25], Sarkis K. Mazmanian[9], Rob Knight[7, 26, 27], Jack A. Gilbert[7, 28], Sharon M. Donovan[8], Trevor D. Lawley[5], Bob Carpenter[1], Richard Bonneau[1, 3, 29], and Gaspar Taroncher-Oldenburg[2]

[1]Center for Computational Biology, Flatiron Institute, Simons Foundation, New York, NY, USA
[2]The Simons Foundation Autism Research Initiative, Simons Foundation, New York, NY, USA
[3]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, USA
[4]Precidiag Inc, Watertown, MA, USA
[5]Host-Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, UK
[6]Bioinformatics and Systems Biology Program, University of California San Diego, San Diego, CA, USA
[7]Department of Pediatrics, School of Medicine, University of California San Diego, San Diego, CA, USA
[8]Division of Nutritional Sciences, University of Illinois, Urbana, IL, USA
[9]Division of Biology & Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[10]Microbiology Institute and Health Science College, Universidad San Francisco de Quito, Quito, Ecuador
[11]Departments of Microbiology & Pharmaceutical Sciences, Oregon State University, Corvallis, OR, USA
[12]Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia
[13]Skolkovo Institute of Science and Technology, Skolkovo, Russia
[14]Department of Biomedical Sciences, School of Medicine, University of Cagliari, Cagliari, Italy
[15]Laboratory Medicine, Department of Surgical Sciences, School of Medicine, University of Cagliari, Italy
[16]Children's Hospital of Fudan University, National Center for Children's Health, Shanghai, China
[17]Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[18]Department of Medicine, Division of Gastroenterology and Hepatology, Stony Brook University, Stony Brook, NY, USA
[19]Neonatal Intensive Care Unit and Neonatal Pathology, Department of Surgical Sciences, School of Medicine, University of Cagliari, Italy
[20]Pediatrics (Systems Medicine), Biomedical Data Science, and Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[21]Institute of Microbiology, COCIBA, Universidad San Francisco de Quito, Quito, Ecuador
[22]Facultad de Ciencias Médicas, de la Salud y la Vida, Universidad Internacional del Ecuador, Quito, Ecuador
[23]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[24]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA
[25]Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, MA, USA
[26]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, California, USA
[27]Department of Bioengineering, University of California San Diego, La Jolla, California, USA
[28]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA
[8]Division of Nutritional Sciences, University of Illinois, Urbana, IL, USA
[29]Prescient Design, a Genentech Accelerator, New York, NY, USA

Author Information : These authors contributed equally : James T. Morton, Gaspar Taroncher-Oldenburg
Corresponding authors : Correspondence to James T. Morton or Gaspar Taroncher-Oldenburg

Acknowledgements

We would like to thank Allan Packer, Paul Wang, Natalia Volfovsky, Kelsey Martin and John Spiro for their critical review of the manuscript. Y.S. and T.D.L. are supported by the Wellcome Trust (WT098051);

Contributions

J.T.M. and G.T.-O. conceived and designed the study, developed the software, analyzed the data, interpreted the results and wrote the manuscript; R.B. contributed to study design, data analysis, result interpretation and manuscript editing; R.H.M. contributed to study design, data analysis and manuscript editing; R.J.X. and S.K.M. contributed to study design and manuscript editing; G.R. and B.C. contributed to software development and manuscript editing; D.-M.J. and Y.S. contributed to data analysis and manuscript editing; K.B., B.D.N., M.F.Z., M.D., O.V.A., A.S.K., A.N., M.M., M.W., D.N.F., E.L., W.Z., V.F., V.N.D., D.P.W., M.E.B., R.K., J.G., S.M.D. and T.D.L. provided access to data and contributed to manuscript editing.

Conflict of Interest

R.H.M. is Scientific Director at Precidiag Inc.; T.D.L. is co-founder and Chief Scientific Officer of Microbiotica; S.K.M. is a co-founder and has equity in Axial Therapeutics; R.B is currently Executive Director of Prescient Design, a Genentech Accelerator; G.T.-O. is a Consultant-in-Residence at the Simons Foundation.

Abstract

Autism is a highly heritable neurodevelopmental disorder characterized by heterogeneous cognitive, behavioral and communication impairments. Disruption of the gut-brain axis (GBA) has been implicated in autism, with dozens of cross-sectional microbiome and other omic studies revealing autism-specific profiles along the GBA albeit with little agreement in composition or magnitude. To explore the functional architecture of autism, we developed an age and sex-matched Bayesian differential ranking algorithm that identified autism-specific profiles across 10 cross-sectional microbiome datasets and 15 other omic datasets, including dietary patterns, metabolomics, cytokine profiles, and human brain expression profiles. The analysis uncovered a highly significant, functional architecture along the GBA that encapsulated the overall heterogeneity of autism phenotypes. This architecture was determined by autism-specific amino acid, carbohydrate and lipid metabolism profiles predominantly encoded by microbial species in the genera *Prevotella*, *Enterococcus*, *Bifidobacterium*, and *Desulfovibrio*, and was mirrored in brain-associated gene expression profiles and restrictive dietary patterns in individuals with autism. Pro-inflammatory cytokine profiling and virome association analysis further supported the existence of an autism-specific architecture associated with particular microbial genera. Re-analysis of a longitudinal intervention study in autism recapitulated the cross-sectional profiles, and showed a strong association between temporal changes in microbiome composition and autism symptoms. Further elucidation of the functional architecture of autism, including of the role the microbiome plays in it, will require deep, multi-omic longitudinal intervention studies on well-defined stratified cohorts to support causal and mechanistic inference.

Introduction

Autism spectrum disorder (ASD) encompasses a broad range of neurodevelopmental conditions defined by heterogeneous cognitive, behavioral and communication impairments that manifest early in childhood [1]. To date, over a hundred genes have been identified as putatively associated with ASD, with some genotypes now having a standardized clinical diagnosis [2]. However, most of the genetic variants are still associated with heterogeneous phenotypes, making it difficult to identify molecular mechanisms that might be responsible for particular impairments [3]. Some studies have also looked at the presence of abnormalities in different brain regions in children with ASD [4, 5]. However, whether such neuroanatomical features could mechanistically determine autism, and whether environmental factors could induce analogous ASD-like symptoms, remains unresolved [1].

In addition to risk factors, one comorbidity that has been linked to ASD with high confidence is the occurrence of gastrointestinal (GI) symptoms, such as constipation, diarrhea, or abdominal bloating, but causal insights remain elusive [6, 7, 8]. Mechanistically, much research has been focused on the interplay between the GI system and processes controlled by the neuroendocrine, neuroimmune, and autonomous nervous systems, all of which converge around the GI tract and together modulate the gut-brain axis (GBA) [9, 10, 11].

The GBA facilitates bidirectional communication between the gut and the brain, contributing to brain homeostasis and helping regulate cognitive and emotional functions [9, 12]. Over the past decade, research on the factors modulating the GBA has revealed the central role played by the gut microbiome—the trillions of microbes that colonize the gut—in regulating neuroimmune networks, modifying neural networks, and directly communicating with the brain [13]. Dysregulation of the gut microbiome and the ensuing disruption of the GBA are thought to contribute to the pathogenesis of neurodevelopmental disorders including autism, but the underlying mechanisms and the extent to which the microbiome explains these dynamics is still unknown [14, 15, 16, 17].

Several dozen autism gut metagenomics studies have revealed many, albeit inconsistent, variations in microbial diversity in individuals with ASD compared with neurotypical individuals [18, 19, 17]. Similarly, metagenome-based functional reconstructions and metabolic analyses have also shown strong, albeit inconclusive differences between ASD and neurotypical individuals [20, 21, 22]. Comparative analyses at other omic levels have further shown little agreement across studies [23] raising the question of whether the results obtained so far reflect intrinsic biological differences among cohorts, insufficient statistical power, or experimental biases that preclude meaningful comparisons [24].

A wide range of factors could explain the disagreement across studies, including confounding variation due to batch effects, the application of inappropriate statistical methodologies, and the vast phenotypic and genotypic heterogeneity of ASD. Batch effects can be caused by many factors including misspecified experimental designs, technical variability, geographical location, and demographic composition, and several algorithms have been proposed to correct for them, but a lack of standardized statistical methods further complicates interpretation [25, 26, 27, 28]. Microbiome datasets, like other omic datasets, are compositional, and failure to account for the compositional nature of sequencing counts can lead to high false positive and false negative rates when identifying differentially abundant microbes [29, 30, 31]. Microbiome analysis in ASD is further confounded by the phenotypic and genotypic heterogeneity of the disorder, which is known to be critical for stratifying ASD subtypes and constructing reliable diagnostics but is typically not measured or controlled for [32, 33, 1].

Understanding functional architecture—the network of interactions among different omic levels that determines individual phenotypes—of complex neurodevelopmental disorders such as autism, requires an accurate and comprehensive characterization of the different omic levels contributing to it [34]. Traditionally focused on the human genomic, metabolic, and cellular components of phenotype determination, mounting evidence of the role the GBA plays in phenotype determination through bidirectional modulatory mechanisms raises the need for considering the metagenomic and metabolic contributions of the microbiome as potential key components of the functional architecture of autism [35, 36].

To identify autism-specific omic profiles while reducing cohort-specific confounding factors, we have devised a Bayesian differential ranking algorithm to estimate a distribution of microbial differentials, or relative log-fold changes, [31] across multiple potential ASD subtypes implicit in 25 omic datasets (Table 1). A key feature of this approach was to match individual study participants by sex and age within each study to adjust for confounders in childhood development and cohort-specific batch effects. The preponderance of autism among males is well documented and several potentially sex-dependent mechanisms to explain this phenonmenon have been proposed [37]. Furthermore, the development of the microbiome during childhood is a hallmark of microbiome dynamics in the human gut [38, 39, 40, 41, 42]. Our analysis provides insights into the complexity of the interplay among multiple omic levels in ASD, highlights the inherent limitations of cross-sectional studies for understanding the functional architecture of autism, and provides a framework for further studies aimed at better defining the causal relationship between the microbiome and other omic levels and ASD.

Results

The structure of our analysis consisted of a multi-cohort and multi-omic meta-analysis framework that allowed us to combine independent and dependent omic data sets in one integrated analysis [43, 44]. To minimize issues of compositionality and sequencing depth [45], we modeled overdispersion using a negative binomial distribution for modeling sequencing count data [46] (Box 1 "TACKLING METAGENOMIC UNKNOWNS"). Our differential ranking approach incorporated a case-control matching component consisting of individually pairing ASD children with age- and sex-matched neurotypical control children within each study cohort, allowing us to adjust for confounding variation and batch effects (Supplemental methods). Finally, we cross-referenced the 16S–based microbial differential ranking analysis from eight age-sex matched cohorts against 15 other omic datasets to contextualize the potential functional roles these microbes could play in autism (Figure 1).

Age- and sex-matching increases informational content of cross-sectional ASD datasets

We compared the age- and sex-matched differential ranking analysis to the standard group-averaged differential ranking analysis across eight out of the ten 16S studies [47, 48, 49, 21, 50, 51, 52, 53]. Age- and sex-matched differential analysis outperformed standard group averaging with respect to R^2, and its overall performance strictly improved as more studies were added (Figure S1). This performance boost reflected a reduction in model uncertainty with larger cohorts that was indicative of overlapping differentially abundant taxa across studies and of reduced confounding variation.

Global differential ranking analysis reveals a distribution of significant ASD-microbiome associations

A global, age- and sex-matched differential ranking analysis of the eight 16S datasets selected for this study revealed a clear partitioning of microbial differences with respect to ASD and cohort membership (Figure 2a, Figure S2). The distribution of the overall case-control differences showed a strong ASD-specific signal driven by 142 microbes more commonly found in ASD children and 32 microbes more commonly found in their control counterparts (Table S1). The variability observed is most likely due to confounding factors such as cohort demographics and geographic location, with the eight cohorts originating from Asia, Europe, South America, and North America. Analogous global differential ranking trends could be observed for the virome, shotgun metagenomics sequencing (SMS), and RNA-seq datasets (Figure S3). To determine whether these highly significant microbiome signals (pvalue$<$0.0025) could be used to distinguish ASD subjects from their age- and sex-matched control counterparts, we trained random forest classifiers on train/validation/test splits on data derived from 16S—targeted sequencing of the microbial 16S ribosomal RNA gene—and SMS—whole genome sequencing of microbial communities. Despite the strong microbiome effect size, we faced difficulties fitting generalizable classifiers. Our best classifiers had an average cross-validation accuracy of about 75% (Figure 2b), falling within the range of 52%–90% classification accuracy observed in previous studies [50, 49, 54]. We suspect that the vast heterogeneity across cohorts hampered classification performance. While cohort size did not impact predictiability (Figure 2c), some cohorts with skewed sex ratios or age ranges did exhibit lower classification performance (Figure. 2d-e). In the Zurita et al. cohort, sex-specific factors could confound classification (four girls and 56 boys) [48], and in the Kang et al. cohort, age-associated microbiome development factors could hamper classification accuracy (all children were 10 years or older). In addition, all subjects in the Kang et al. cohort had known GI symptoms [55], further compromising classifier performance because none of the other studies controlled for this variable. As a result, and analogous to the phenotypic and genotypic heterogeneity observed in ASD, the microbiome composition of ASD children also exhibits high heterogeneity, precluding the identification of a homogeneous universal ASD microbial profile and the construction of generalizable classifiers.

Children with ASD exhibit significant individual differences at several omic levels

Differential ranking analysis of three core omic levels—microbiome (16S and SMS) and human transcriptome (RNAseq)—revealed strong and highly significant differences between ASD subjects and their neurotypical

counterparts (p-value < 0.0025) (Figure 2f; Table S2; Table S3). Two additional omic levels—the metabolome and the virome— didn't show significance signals (Figure S4, Table S4). Amongst the models that yielded a statistically significant signal, the 16S and SMS datasets had a larger effect size than the RNAseq datasets (Figure 2f). While each omic dataset by itself showed strong associations with ASD, a side-by-side comparison of the 16S and SMS datasets—two datasets that should show high equivalence—revealed a significant lack of overlap between them, highlighting the outstanding challenge of batch effects in microbiome studies (Figure S5). The reasons for this discrepancy could be many, but most likely center around sample size—our study looked at eight 16S datasets versus only three SMS datasets containing 754 individuals and 166 individuals, respectively. Another major challenge when estimating species profiles with metagenomics reference libraries is assigning a species identification to a read—there are many reads that do not uniquely map to individual species—and as a result, these multi-mapped reads can give rise to numerous false positive taxa [56, 57, 58, 59]. Based on this, we decided to focus primarily on the 16S datasets to define a global differential ranking profile.

Sibling-matching and unrelated sex- and age-matching show significant discrepancies

To determine whether sex- and age-matched differentials could be universally predictive, we compared the 16S differentials obtained from the age- and sex-matched cohorts with two sibling-matched cohorts [60] [61]. Interestingly, we observed a significant negative correlation between the differentials extracted from the age- and sex-matched cohort and the two sibling-matched cohorts, suggesting that ASD-specific microbes in the sibling-matched studies are enriched in the control group in the age- and sex-matched studies and vice versa (Fig. 2g). Permanova applied to the age-sex matched cohort revealed a strong age-confounder across cohorts (pvalue < 0.001), with little confounding variation due to sex (Table S5). In contrast, Permanova applied to these sibling-matched cohorts revealed that household is a major confounder in both cohorts (pvalue $<$ 0.001), but ruled out age as a confounder and indicated that sex was a confounder only in the David et al. cohort (pvalue < 0.001). The observed discrepancies point to different sets of confounders possibly affecting the analysis: in the case of the age- and sex-matched studies, family confounders aren't typically accounted for, while sibling-matched studies don't typically adjust for age confounders. In addition, and while cohorts such as the one studied in Maude et al. specifically control for the possibility, siblings often exhibit a higher risk of developing ASD compared to the general population [62].

Host cytokine concentrations are correlated with microbial abundances

Immune dysregulation, ranging from circulating 'anti-brain' antibodies and perturbed cytokine profiles to simply having a family history of immune disorders, has been repeatedly associated with ASD [63, 64]. Recently, for example, Zurita et al. showed that concentrations of the inflammatory cytokine transforming growth factor beta (TGF-β) are significantly elevated in ASD children. We reanalyzed this dataset, after age- and sex-matching, and observed that microbial differentials associated with TGF-β and IL-6 concentrations were positively correlated with the global microbial log-fold changes between ASD and control pairs (IL-6 : r=-0.435, p=0; TGF-β : r=0.291, p=0) (Table S6). To validate the integrity of these microbial profiles with respect to the cytokine changes, we calculated the log-ratios of these microbial abundances and showed them to, in turn, be highly correlated with TGF-β and IL-6 concentrations (IL-6 : r=0.50, p=0.0007; TGF-β : r=0.45, p=0.002) (Figure 3 a-d).

Four clusters of microbial genera—*Prevotella*, *Enterococcus*, *Bifidobacteria*, and *Desulfovibrio*—were predominantly associated with the cytokine differentials. Partial mechanistic insights on some of these cytokine-microbe associations have been previously published. Both *B. longum* and *E. faecalis* have shown anti-inflammatory activities: *B. longum* downregulates IL-6 in fetal human enterocytes in vitro [65] and *E. faecalis* has been observed to upregulate TGF-β in human intestinal cells [66]. *P. copri* associations with different cytokines have also been observed in multiple disease contexts [67]. Similarly, *Bifidobacteria* and *Prevotella* both co-occurred with phages enriched in ASD or in neurotypical children (Figure S6, Table S7), but while microbes have previously been reported to mediate viral infections [68, 69], the mechanistic underpinnings of these interactions with the host's immunity remain poorly understood [70, 71, 72].

The microbiome metabolic capacity is reflective of the human brain-associated metabolic capacity in ASD

To determine potential crosstalk between the human brain and the microbiome physiology, we compared the metabolic capacities encoded by the microbial metagenome—combining the individual metabolic capacities of thousands of different microbes—and the differentially expressed human genome in the brain, two omic levels representing entirely different biological contexts. We observed that over 100 human metabolic pathways differentially expressed in the brain tissues of ASD individuals had analogous microbial pathways differentially abundant in the microbiome of children with ASD, suggesting a potential coordination of metabolic pathways across omic levels in ASD (Fig. 3e). Pathways related to amino acid metabolism, carbohydrate metabolism and lipid metabolism were disproportionately represented among the overlapping genes (Table S8).

The microbiome metabolic capacity reflects restrictive diet patterns in children with ASD

Autistic traits in early childhood have been shown to correlate with poor diet quality later in life, however, little is known about how diet quality is directly linked to autistic traits [73]. Here, we re-analyzed the paired microbiome and dietary survey data from Berding et al.. A microbiome-diet co-occurrence analysis revealed startlingly similar amino acid, carbohydrate and lipid metabolism association patterns to those observed in the microbiome-brain metabolic capacity analysis (Table S9). Interestingly, both microbes enriched in ASD and in control subjects co-occurred primarily with amino acid dietary compounds (Figure 3f, Figure S7). Autistic children were less likely to consume foods high in glutamic acid, serine, choline, phenylalanine, leucine, tyrosine, valine and histidine, all compounds involved in neurotransmitter biosynthesis. Even though the metabolomic analysis did not yield statistically significant signals after FDR correction, the metabolites that showed the strongest signal included glutamate and phenylalanine, consistent with the microbiome-diet analysis [74, 75, 76]. Disruptions in the biosynthesis of these neurotransmitter molecules have been implicated in a wide variety of psychiatric disorders, and a recent blood metabolomics study has shown the potential of using branched chain amino acids to define autism subtypes [33]. Due to the incompatibility between the molecular features across datasets, it was not possible to combine any of the metabolomics datasets to boost the statistical power, which remains a major limitation of metabolomics technologies at present (see Methods).

Differential microbial rankings show disease-specific correlations

One major challenge in determining microbiome-disease associations is identifying correlations specific to a particular condition and not generally present across diseases [77, 78]. To determine how specific to ASD our global differential microbiome profile was, we cross-referenced it against differential ranking results obtained from an Inflammatory Bowel Disorder (IBD) dataset [79] and a Type 1 Diabetes (T1D) dataset [80]. IBD shares some comorbidities with ASD [81, 82], while no direct correlation between ASD and T1D has been reported to date. The analysis revealed a notable overlap between microbes enriched in ASD and IBD, and this overlap was stronger than both the overlap between IBD and T1D and between ASD and T1D (Figure S8). Whether this ASD-IBD overlap is suggestive of a common microbial profile or is confounded by the restrictive dietary nature of these two clinical conditions is currently unclear. Higher resolution and properly designed clinical studies will have to be performed to get to a mechanistic understanding of the potential microbiome connection between these two conditions.

ASD microbiome profiles weaken after fecal matter transplant consistent with reported behavioral improvement

While the preceding cross-sectional analyses showed significant associations among several omic levels (virome, microbiome, immunome) or diet and ASD, insights into causality are still limited. By contrast, longitudinal intervention studies provide an opportunity to obtain stronger insights into causality. To test this, we re-analyzed data from a two-year, open label fecal matter transplant (FMT) study with 18 children with ASD [83]. In this study, the children were subjected to a two-week antibiotic treatment and a bowel cleanse followed by two days of high dose FMT treatment and eight weeks of daily maintenance FMT doses. Based on one of the most common evaluation scales for ASD, the Childhood Autism Rating Scale (CARS), significant improvements were achieved after the ten week course of treatment. Two months later the initial gains were largely maintained, and a two-year follow-up showed signs of further improvement in most of the patients. The results are consistent with a potential role of the microbiome in improving autism symptoms, but how the underlying changes in microbiome composition related to those seen in other studies remained unknown.

Here, we re-analyzed the original raw data in the context of the ASD profiles revealed by our cross-sectional differential ranking analysis (Table S10). All microbes associated with ASD in the 18 children prior to the FMT treatment had been identified as ASD-associated microbes in our age- and sex-matched cross-sectional analysis, recapitulating 74% of the cross-sectional profile. Immediately following FMT treatment, the abundances of the ASD-associated microbes decreased in all 8 children (Figure 4). The two-year follow-up analysis revealed that all the ASD-associated microbes, mostly *Enterococcaceae*, continued to be depleted. Consistent with the findings of Kang et al., we also observed *Desulfovibrio sp.*, and *P. copri* increase over the two year period, while *Bifidobacteria sp.* could be found both among depleted and enriched species and other *Prevotella sp.* were depleted, pointing to a potentially wide functional diversity within these genera not noted in the original study.

Discussion

The functional architecture of ASD, and in particular the potential role the microbiome plays in modulating the GBA in the context of autism, remains poorly understood due to disagreements among existing microbiome and other omic studies. Our Bayesian model highlighted a distribution of highly significant microbial differentials obtained from individual age- and sex-pairings between children with ASD and neurotypicals, and parallel analyses at the immunome, human transcriptome, and dietome levels revealed strong associations among omic levels. The virome and the direct metabolome signals, while present, were markedly weaker than the other omic signals. The inferred ASD-specific metabolic profiles from the microbiome and the human transcriptome, on the other hand, showed a high and significant degree of overlap in microbial and human pathways expressed in the gut and in the brain, respectively. The metabolic connection implied by this overlap, which included differentially enriched carbohydrate and amino acid metabolic pathways in

ASD, is a remarkable observation given the fundamental difference between the gut and brain physiologies, which would a priori suggest a reduced overlap in metabolic capacities. The microbiome-diet co-occurrence analysis also highlighted a reduced intake of amino acids and carbohydrates linked to specific microbiome profiles in ASD children. These metabolic and dietary imbalances, particularly regarding glutamate levels, were further apparent, albeit weakly, in the serum, fecal and urine metabolomes we analyzed. This multi-scale overlap we observed along the GBA points to the existence of a functional architecture of ASD driven by the metabolic potential at the genomic and metagenomic levels.

While the differential distributions we determined were highly significant, the global analysis did not provide reliable ASD classifiers or uncover universal microbial 'smoking guns' linked to autism. However, several microorganisms consistently detected across omic levels pointed to potentially interesting functional connections. For example, our analysis suggested that *B. longum* exhibited a down-regulation of IL-6, which has been observed across a number of in vitro and cohort studies[84, 85]. The diet co-occurrence analysis also showed a strong association between *P. copri* and carbohydrate depletion in ASD. The population dynamics of *P. copri* have been reported to be driven primarily by carbohydrates in the diet [67]. Multiple other microbes, including several *Bifidobacteria*, *Enterococcus* and *Desulfovibrio* species, stood out in the immune and viral analyses. In the FMT study, the relative proportions of several *Prevotella*, *Bifidobacteria*, *Desulfovibrio* and *Enterococcus* species also showed strong associations with ASD symptoms, further suggesting a causal role for these microorganisms in shaping autism symptoms.

Despite our inability to determine actual metabolomic profiles at this point (see Methods), our metabolite analysis based on microbiome- and brain-derived metabolite inferences as well as the diet-derived metabolite data reveals a picture of a unifying and distinct ASD functional architecture. With the brain, the immunome and diet as major effectors, the multi-factorial complexity of ASD is reduced to a multi-scale set of interactions centered around human and bacterial metabolism that in turn determines phenotypic, genomic and metagenomic attributes via multiple feedback loops (Figure 5). The association of specific genotypes with ASD has been clearly established [2]; the pivotal role of the immune system in mediating the communication between the gut microbiome and the human brain as well as other peripheral systems is also firmly established [86]; further, the central role of the microbiome in mediating diet-derived nutrient mobilization has been extensively documented [87]; and several hard-wired feedback loops among these effectors such as the hypothalamus-mediated regulation of appetite and diet, have also been described [10].

A major limitation of our meta-analysis is the lack of consistent behavioral, genotypic or electronic healthcare record data that would have allowed subtyping of the ASD subjects in light of environmental confounders. Furthermore, we cannot definitely recommend age- and sex-matching over sibling matching based on our analysis. Age remains a major confounding factor in early childhood microbiome development and controlling for this is key for understanding microbial fluctuations [88]. On the other hand, sibling-matching may help control environmental factors, but mostly rules out the ability to age-match subjects, thus potentially introducing the age confounder [89]. And while our approach revealed strong associations among the microbiome, other omic levels, and ASD, the vast heterogeneity in behavioral patterns and in genotypes is a major obstacle in constructing diagnostics and treatments for ASD symptoms [90, 32].

Our analysis has further exposed the limitations of cross-sectional cohort studies and the need for longitudinal intervention studies to further our understanding of the functional architecture of ASD. Building realistic causal models of autism needs to take into account the multi-factorial complexity underlying different ASD subtypes, which will require a concerted effort to simultaneously analyze several omic levels and at clinically relevant time scales. For instance, understanding the engraftment dynamics of FMT and its functional implications on the recipients' gut microbiomes requires frequent initial sampling of the microbiome, immunome and metabolome, but tracing any behavioral changes over time requires less frequent sampling over periods of up to several years in combination with reliable behavioral, medical and dietary surveys [91, 92]. Collecting and integrating such multi-scale omic datasets presents unique logistical and analytical challenges.

Managing data acquisition and access will require coordinating multiple sites and potentially centralizing some aspects of sample processing. Recent initiatives such as The Environmental Determinants of Diabetes in the Young (TEDDY) study, an international long-term, multi-center initiative to link specific environmental triggers to particular Type 1 diabetes–associated genotypes, provide a blueprint for similar approaches in ASD [93]. A key component of such an initiative would be the establishment of standardized sampling and processing protocols that would minimize technical confounders, one of the top confounders at most omic levels. For instance, our analysis showed major batch effects when comparing 16S and SMS datasets across cohorts (r=-0.023, p=0.48, Figure S5b) as well as within a cohort (r=0.17, p=1e-5, Figure S5b). And while there are extensive efforts underway to calibrate microbiome datasets [94], other omic levels such as the metabolome [26] present even more fundamental technical issues that make it imperative to develop concerted strategies to be able to include them in an integrated analysis.

In addition to the considerable variations in statistical properties across datasets, interactions among omic levels are mostly underdetermined, making the construction of informative models a major challenge. Determining the necessary biologically relevant and unbiased assumptions is a non-trivial process and can inadvertently lead to model mis-specifications resulting in misleading conclusions. As pointed out recently for genetic, environmental and microbiome models in ASD [95, 96], addressing these issues will be critical to inferring causal mechanisms from population-scale studies. In addition, and given the vast heterogeneity of ASD, designing cohort studies that minimize confounding factor effects will be key to furthering our understanding of autism. For example, while our analysis could not identify ASD subtypes implicated in GI symptoms, we have determined stronger associations between gut microbes, host immunity, brain expression and dietary patterns than previously reported, highlighting the potential for boosting the statistical power and biological insight with comprehensive omic analyses. We conclude that multi-omic longitudinal intervention studies on appropriately stratified cohorts, in combination with comprehensive patient metadata, provide an optimal approach to advance our understanding of the etiology of autism to the next level.

References

[1] Catherine Lord, Traolach S Brugha, Tony Charman, James Cusack, Guillaume Dumas, Thomas Frazier, Emily J H Jones, Rebecca M Jones, Andrew Pickles, Matthew W State, Julie Lounds Taylor, and Jeremy Veenstra-VanderWeele. Autism spectrum disorder. *Nat Rev Dis Primers*, 6(1):5, January 2020.

[2] F Kyle Satterstrom, Jack A Kosmicki, Jiebiao Wang, Michael S Breen, Silvia De Rubeis, Joon-Yong An, Minshi Peng, Ryan Collins, Jakob Grove, Lambertus Klei, Christine Stevens, Jennifer Reichert, Maureen S Mulhern, Mykyta Artomov, Sherif Gerges, Brooke Sheppard, Xinyi Xu, Aparna Bhaduri, Utku Norman, Harrison Brand, Grace Schwartz, Rachel Nguyen, Elizabeth E Guerrero, Caroline Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, Catalina Betancur, Edwin H Cook, Louise Gallagher, Michael Gill, James S Sutcliffe, Audrey Thurm, Michael E Zwick, Anders D Børglum, Matthew W State, A Ercument Cicek, Michael E Talkowski, David J Cutler, Bernie Devlin, Stephan J Sanders, Kathryn Roeder, Mark J Daly, and Joseph D Buxbaum. Large-Scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 180(3):568–584.e23, February 2020.

[3] Lilia M Iakoucheva, Alysson R Muotri, and Jonathan Sebat. Getting to the cores of autism. *Cell*, 178(6):1287–1298, September 2019.

[4] Cynthia Mills Schumann, Julia Hamstra, Beth L Goodlin-Jones, Linda J Lotspeich, Hower Kwon, Michael H Buonocore, Cathy R Lammers, Allan L Reiss, and David G Amaral. The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. *Journal of neuroscience*, 24(28):6392–6401, 2004.

[5] Cynthia Mills Schumann and David G Amaral. Stereological analysis of amygdala neuron number in autism. *Journal of Neuroscience*, 26(29):7674–7679, 2006.

[6] Radu Lefter, Alin Ciobica, Daniel Timofte, Carol Stanciu, and Anca Trifan. A descriptive review on the prevalence of gastrointestinal disturbances and their multiple associations in autism spectrum disorder. *Medicina*, 56(1):11, December 2019.

[7] Amirhossein Modabbernia, Eva Velthorst, and Abraham Reichenberg. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Mol. Autism*, 8:13, March 2017.

[8] Elaine Y Hsiao. Gastrointestinal issues in autism spectrum disorder. *Harv. Rev. Psychiatry*, 22(2):104, 2014.

[9] Emeran A Mayer. Gut feelings: the emerging biology of gut–brain communication. *Nat. Rev. Neurosci.*, 12(8):453–466, July 2011.

[10] Livia H Morais, Henry L Schreiber, 4th, and Sarkis K Mazmanian. The gut microbiota-brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.*, 19(4):241–255, April 2021.

[11] John F Cryan, Kenneth J O'Riordan, Kiran Sandhu, Veronica Peterson, and Timothy G Dinan. The gut microbiome in neurological disorders. *Lancet Neurol.*, 19(2):179–194, February 2020.

[12] Amar Sarkar, Siobhán Harty, Katerina V-A Johnson, Andrew H Moeller, Rachel N Carmody, Soili M Lehto, Susan E Erdman, Robin I M Dunbar, and Philip W J Burnet. The role of the microbiome in the neurobiology of social behaviour. *Biol. Rev. Camb. Philos. Soc.*, 95(5):1131–1166, October 2020.

[13] Amanda Jacobson, Daping Yang, Madeleine Vella, and Isaac M Chiu. The intestinal neuro-immune axis: crosstalk between neurons, immune cells, and microbes. *Mucosal Immunol.*, 14(3):555–565, February 2021.

[14] Sue Grenham, Gerard Clarke, John F Cryan, and Timothy G Dinan. Brain–Gut–Microbe communication in health and disease. *Front. Physiol.*, 0, 2011.

[15] The Microbiome-Gut-Brain axis in health and disease. *Gastroenterol. Clin. North Am.*, 46(1):77–89, March 2017.

[16] Yoko M Ambrosini, Dana Borcherding, Anumantha Kanthasamy, Hyun Jung Kim, Auriel A Willette, Albert Jergens, Karin Allenspach, and Jonathan P Mochel. The Gut-Brain axis in neurodegenerative diseases and relevance of the canine model: A review. *Front. Aging Neurosci.*, 0, 2019.

[17] Mingyu Xu, Xuefeng Xu, Jijun Li, and Fei Li. Association between gut microbiota and autism spectrum disorder: A systematic review and Meta-Analysis. *Front. Psychiatry*, 10:473, July 2019.

[18] Pedro Andreo-Martínez, María Rubio-Aparicio, Julio Sánchez-Meca, Alejandro Veas, and Agustín Ernesto Martínez-González. A meta-analysis of gut microbiota in children with autism. *J. Autism Dev. Disord.*, May 2021.

[19] Navya Bezawada, Tze Hui Phang, Georgina L Hold, and Richard Hansen. Autism spectrum disorder and the gut microbiota in children: A systematic review. *Ann. Nutr. Metab.*, 76(1):16–29, January 2020.

[20] Narueporn Likhitweerawong, Chanisa Thonusin, Nonglak Boonchooduang, Orawan Louthrenoo, Intawat Nookaew, Nipon Chattipakorn, and Siriporn C Chattipakorn. Profiles of urine and blood metabolomics in autism spectrum disorders. *Metab. Brain Dis.*, August 2021.

[21] Jiang Zhu, Xueying Hua, Ting Yang, Min Guo, Qiu Li, Lu Xiao, Ling Li, Jie Chen, and Tingyu Li. Alterations in gut vitamin and amino acid metabolism are associated with symptoms and neurodevelopment in children with autism spectrum disorder. *J. Autism Dev. Disord.*, July 2021.

[22] Xin-Jie Xu, Xiao-E Cai, Fan-Chao Meng, Tian-Jia Song, Xiao-Xi Wang, Yi-Zhen Wei, Fu-Jun Zhai, Bo Long, Jun Wang, Xin You, and Rong Zhang. Comparison of the metabolic profiles in the plasma and urine samples between autistic and typically developing boys: A preliminary study. *Front. Psychiatry*, 12:657105, June 2021.

[23] Atiqah Azhari, Farouq Azizan, and Gianluca Esposito. A systematic review of gut-immune-brain mechanisms in autism spectrum disorder. *Dev. Psychobiol.*, 61(5):752–771, July 2019.

[24] Patrick D Schloss. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *MBio*, 9(3), June 2018.

[25] Yiwen Wang and Kim-Anh LêCao. Managing batch effects in microbiome data. *Brief. Bioinform.*, 21(6):1954–1970, December 2020.

[26] Wei Han and Liang Li. Evaluating and minimizing batch effects in metabolomics. *Mass Spectrom. Rev.*, November 2020.

[27] Hoa Thi Nhu Tran, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, and Jinmiao Chen. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.*, 21(1):12, January 2020.

[28] Jelena Čuklina, Patrick G A Pedrioli, and Ruedi Aebersold. Review of batch effects prevention, diagnostics, and correction approaches. *Methods Mol. Biol.*, 2051:373–387, 2020.

[29] Gregory B Gloor, Jia Rong Wu, Vera Pawlowsky-Glahn, and Juan José Egozcue. It's all relative: analyzing microbiome data as compositions. *Ann. Epidemiol.*, 26(5):322–329, May 2016.

[30] Doris Vandeputte, Gunter Kathagen, Kevin D'hoe, Sara Vieira-Silva, Mireia Valles-Colomer, João Sabino, Jun Wang, Raul Y Tito, Lindsey De Commer, Youssef Darzi, Séverine Vermeire, Gwen Falony, and Jeroen Raes. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, 551(7681):507–511, November 2017.

[31] James T Morton, Clarisse Marotz, Alex Washburne, Justin Silverman, Livia S Zaramela, Anna Edlund, Karsten Zengler, and Rob Knight. Establishing microbial composition measurement standards with reference frames. *Nat. Commun.*, 10(1):2719, June 2019.

[32] Yuan Luo, Alal Eran, Nathan Palmer, Paul Avillach, Ami Levy-Moonshine, Peter Szolovits, and Isaac S Kohane. A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nat. Med.*, 26(9):1375–1379, September 2020.

[33] Alan M Smith, Joseph J King, Paul R West, Michael A Ludwig, Elizabeth L R Donley, Robert E Burrier, and David G Amaral. Amino acid dysregulation metabotypes: Potential biomarkers for diagnosis and individualized treatment for subtypes of autism spectrum disorder. *Biol. Psychiatry*, 85(4):345–354, February 2019.

[34] Patrick F Sullivan and Daniel H Geschwind. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*, 177(1):162–183, 2019.

[35] Ran Blekhman, Julia K Goodrich, Katherine Huang, Qi Sun, Robert Bukowski, Jordana T Bell, Timothy D Spector, Alon Keinan, Ruth E Ley, Dirk Gevers, et al. Host genetic variation impacts microbiome composition across human body sites. *Genome biology*, 16(1):1–12, 2015.

[36] Braden T Tierney, Yingxuan Tan, Aleksandar D Kostic, and Chirag J Patel. Gene-level metagenomic architectures across diseases yield high-resolution microbiome diagnostic indicators. *Nature communications*, 12(1):1–12, 2021.

[37] Sarah L Ferri, Ted Abel, and Edward S Brodkin. Sex differences in autism spectrum disorder: a review. *Curr. Psychiatry Rep.*, 20(2):9, March 2018.

[38] Maria Gloria Dominguez-Bello, Filipa Godoy-Vitorino, Rob Knight, and Martin J Blaser. Role of the microbiome in human development. *Gut*, 68(6):1108–1114, June 2019.

[39] Marie-Claire Arrieta, Leah T Stiemsma, Nelly Amenyogbe, Eric M Brown, and Brett Finlay. The intestinal microbiome in early life: health and disease. *Front. Immunol.*, 5:427, September 2014.

[40] Gwen Falony, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, Marc Jan Bonder, Mireia Valles-Colomer, Doris Vandeputte, Raul Y Tito, Samuel Chaffron, Leen Rymenans, Chloë Verspecht, Lise De Sutter, Gipsi Lima-Mendez, Kevin D'hoe, Karl Jonckheere, Daniel Homola, Roberto Garcia, Ettje F Tigchelaar, Linda Eeckhaudt, Jingyuan Fu, Liesbet Henckaerts, Alexandra Zhernakova, Cisca Wijmenga, and Jeroen Raes. Population-level analysis of gut microbiome variation. *Science*, 352(6285):560–564, April 2016.

[41] Tanya Yatsunenko, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N Baldassano, Andrey P Anokhin, Andrew C Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J Gregory Caporaso, Catherine A Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I Gordon. Human gut microbiome viewed across age and geography. *Nature*, 486(7402):222–227, May 2012.

[42] Yan Shao, Samuel C Forster, Evdokia Tsaliki, Kevin Vervier, Angela Strang, Nandi Simpson, Nitin Kumar, Mark D Stares, Alison Rodger, Peter Brocklehurst, Nigel Field, and Trevor D Lawley. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*, 574(7776):117–121, October 2019.

[43] Adib Shafi, Tin Nguyen, Azam Peyvandipour, Hung Nguyen, and Sorin Draghici. A Multi-Cohort and Multi-Omics Meta-Analysis framework to identify Network-Based gene signatures. *Front. Genet.*, 10:159, March 2019.

[44] Alexander Kaever, Manuel Landesfeind, Kirstin Feussner, Burkhard Morgenstern, Ivo Feussner, and Peter Meinicke. Meta-analysis of pathway enrichment: combining independent and dependent omics data sets. *PLoS One*, 9(2):e89297, February 2014.

[45] Gregory B Gloor, Jean M Macklaim, Vera Pawlowsky-Glahn, and Juan J Egozcue. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.*, 8:2224, November 2017.

[46] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12):550, 2014.

[47] Kirsten Berding and Sharon M Donovan. Dietary patterns impact temporal dynamics of fecal microbiota composition in children with autism spectrum disorder. *Front Nutr*, 6:193, 2019.

[48] María Fernanda Zurita, Paúl A Cárdenas, María Elena Sandoval, María Caridad Peña, Marco Fornasini, Nancy Flores, Marcia H Monaco, Kirsten Berding, Sharon M Donovan, Thomas Kuntz, Jack A Gilbert, and Manuel E Baldeón. Analysis of gut microbiome, nutrition and immune status in autism spectrum disorder: a case-control study in ecuador. *Gut Microbes*, 11(3):453–464, May 2020.

[49] Zhou Dan, Xuhua Mao, Qisha Liu, Mengchen Guo, Yaoyao Zhuang, Zhi Liu, Kun Chen, Junyu Chen, Rui Xu, Junming Tang, Lianhong Qin, Bing Gu, Kangjian Liu, Chuan Su, Faming Zhang, Yankai Xia, Zhibin Hu, and Xingyin Liu. Altered gut microbial profile is associated with abnormal metabolism activity of autism spectrum disorder. *Gut Microbes*, 11(5):1246–1267, September 2020.

[50] Jennifer Fouquier, Nancy Moreno Huizar, Jody Donnelly, Cody Glickman, Dae-Wook Kang, Juan Maldonado, Rachel A Jones, Kimberly Johnson, James B Adams, Rosa Krajmalnik-Brown, and Catherine Lozupone. The gut microbiome in autism: Study-Site effects and longitudinal analysis of behavior change. *mSystems*, 6(2), April 2021.

[51] Rong Zou, Fenfen Xu, Yuezhu Wang, Mengmeng Duan, Min Guo, Qiang Zhang, Hongyang Zhao, and Huajun Zheng. Changes in the gut microbiota of children with autism spectrum disorder. *Autism Res.*, 13(9):1614–1625, September 2020.

[52] Dae-Wook Kang, James B Adams, Ann C Gregory, Thomas Borody, Lauren Chittick, Alessio Fasano, Alexander Khoruts, Elizabeth Geis, Juan Maldonado, Sharon McDonough-Means, Elena L Pollard, Simon Roux, Michael J Sadowsky, Karen Schwarzberg Lipson, Matthew B Sullivan, J Gregory Caporaso, and Rosa Krajmalnik-Brown. Microbiota transfer therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome*, 5(1):10, January 2017.

[53] Prjna687773. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA687773.

[54] Chloe X Yap, Anjali K Henders, Gail A Alvares, David L A Wood, Lutz Krause, Gene W Tyson, Restuadi Restuadi, Leanne Wallace, Tiana McLaren, Narelle K Hansell, Dominique Cleary, Rachel Grove, Claire Hafekost, Alexis Harun, Helen Holdsworth, Rachel Jellett, Feroza Khan, Lauren P Lawson, Jodie Leslie, Mira Levis Frenk, Anne Masi, Nisha E Mathew, Melanie Muniandy, Michaela Nothard, Jessica L Miller, Lorelle Nunn, Gerald Holtmann, Lachlan T Strike, Greig I de Zubicaray, Paul M Thompson, Katie L McMahon, Margaret J Wright, Peter M Visscher, Paul A Dawson, Cheryl Dissanayake, Valsamma Eapen, Helen S Heussler, Allan F McRae, Andrew J O Whitehouse, Naomi R Wray, and Jacob Gratten. Autism-related dietary preferences mediate autism-gut microbiome associations. *Cell*, November 2021.

[55] Dae-Wook Kang, Zehra Esra Ilhan, Nancy G Isern, David W Hoyt, Daniel P Howsmon, Michael Shaffer, Catherine A Lozupone, Juergen Hahn, James B Adams, and Rosa Krajmalnik-Brown. Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. *Anaerobe*, 49:121–131, February 2018.

[56] A Gonzalez, Y Vázquez-Baeza, J B Pettengill, A Ottesen, D McDonald, and R Knight. Avoiding pandemic fears in the subway and conquering the platypus. *mSystems*, 1(3), May 2016.

[57] Simon H Ye, Katherine J Siddle, Daniel J Park, and Pardis C Sabeti. Benchmarking metagenomics tools for taxonomic classification. *Cell*, 178(4):779–794, August 2019.

[58] Alexa B R McIntyre, Rachid Ounit, Ebrahim Afshinnekoo, Robert J Prill, Elizabeth Hénaff, Noah Alexander, Samuel S Minot, David Danko, Jonathan Foox, Sofia Ahsanuddin, Scott Tighe, Nur A Hasan, Poorani Subramanian, Kelly Moffat, Shawn Levy, Stefano Lonardi, Nick Greenfield, Rita R Colwell, Gail L Rosen, and Christopher E Mason. Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biol.*, 18(1):182, September 2017.

[59] Derrick E Wood and Steven L Salzberg. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.*, 15(3):R46, March 2014.

[60] Joshua S Son, Ling J Zheng, Leahana M Rowehl, Xinyu Tian, Yuanhao Zhang, Wei Zhu, Leighann Litcher-Kelly, Kenneth D Gadow, Grace Gathungu, Charles E Robertson, Diana Ir, Daniel N Frank, and Ellen Li. Comparison of fecal microbiota in children with autism spectrum disorders and neurotypical siblings in the simons simplex collection. *PLoS One*, 10(10):e0137725, October 2015.

[61] Maude M David, Christine Tataru, Jena Daniels, Jessey Schwartz, Jessica Keating, Jarrad Hampton-Marcell, Neil Gottel, Jack A Gilbert, and Dennis P Wall. Children with autism and their typically developing siblings differ in amplicon sequence variants and predicted functions of Stool-Associated microbes. *mSystems*, 6(2), April 2021.

[62] Ewa Pisula and Karolina Ziegart-Sadowska. Broader autism phenotype in siblings of children with ASD–A review. *Int. J. Mol. Sci.*, 16(6):13217–13258, June 2015.

[63] Lisa A Croen, Yinge Qian, Paul Ashwood, Julie L Daniels, Daniele Fallin, Diana Schendel, Laura A Schieve, Alison B Singer, and Ousseny Zerbo. Family history of immune conditions and autism spectrum and developmental disorders: Findings from the study to explore early development, 2019.

[64] Amory Meltzer and Judy Van de Water. The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology*, 42(1):284–298, January 2017.

[65] Di Meng, Weishu Zhu, Kriston Ganguli, Hai Ning Shi, and W Allan Walker. Anti-inflammatory effects of bifidobacterium longum subsp infantis secretions on fetal human enterocytes are mediated by TLR-4 receptors. *Am. J. Physiol. Gastrointest. Liver Physiol.*, 311(4):G744–G753, October 2016.

[66] Shugui Wang, Lydia Hui Mei Ng, Wai Ling Chow, and Yuan Kun Lee. Infant intestinal enterococcus faecalis down-regulates inflammatory responses in human intestinal cell lines. *World J. Gastroenterol.*, 14(7):1067–1076, February 2008.

[67] Adrian Tett, Kun D Huang, Francesco Asnicar, Hannah Fehlner-Peach, Edoardo Pasolli, Nicolai Karcher, Federica Armanini, Paolo Manghi, Kevin Bonham, Moreno Zolfo, Francesca De Filippis, Cara Magnabosco, Richard Bonneau, John Lusingu, John Amuasi, Karl Reinhard, Thomas Rattei, Fredrik Boulund, Lars Engstrand, Albert Zink, Maria Carmen Collado, Dan R Littman, Daniel Eibach, Danilo Ercolini, Omar Rota-Stabelli, Curtis Huttenhower, Frank Maixner, and Nicola Segata. The prevotella copri complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe*, 26(5):666–679.e7, November 2019.

[68] Zhenda Shi and Andrew T Gewirtz. Together forever: Bacterial-Viral interactions in infection and immunity. *Viruses*, 10(3), March 2018.

[69] Ursula Neu and Bernardo A Mainou. Virus interactions with bacteria: Partners in the infectious dance. *PLoS Pathog.*, 16(2):e1008234, February 2020.

[70] Luis F Camarillo-Guerrero, Alexandre Almeida, Guillermo Rangel-Pineros, Robert D Finn, and Trevor D Lawley. Massive expansion of human gut bacteriophage diversity. *Cell*, 184(4):1098–1109.e9, February 2021.

[71] Mohammadali Khan Mirzaei and Corinne F Maurice. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.*, 15(7):397–408, July 2017.

[72] Andrey N Shkoporov and Colin Hill. Bacteriophages of the human gut: The "known unknown" of the microbiome. *Cell Host Microbe*, 25(2):195–209, February 2019.

[73] Holly A Harris, Yuchan Mou, Gwen C Dieleman, Trudy Voortman, and Pauline W Jansen. Child autistic traits, food selectivity and diet quality: A Population-Based study. *J. Nutr.*, December 2021.

[74] Hitoshi Kuwabara, Hidenori Yamasue, Shinsuke Koike, Hideyuki Inoue, Yuki Kawakubo, Miho Kuroda, Yosuke Takano, Norichika Iwashiro, Tatsunobu Natsubori, Yuta Aoki, Yukiko Kano, and Kiyoto Kasai. Altered metabolites in the plasma of autism spectrum disorder: a capillary electrophoresis time-of-flight mass spectroscopy study. *PLoS One*, 8(9):e73814, September 2013.

[75] Brittany D Needham, Mark D Adame, Gloria Serena, Destanie R Rose, Gregory M Preston, Mary C Conrad, A Stewart Campbell, David H Donabedian, Alessio Fasano, Paul Ashwood, and Sarkis K Mazmanian. Plasma and fecal metabolite profiles in autism spectrum disorder. *Biol. Psychiatry*, 89(5):451–462, March 2021.

[76] Antonio Noto, Vassilios Fanos, Luigi Barberini, Dmitry Grapov, Claudia Fattuoni, Marco Zaffanello, Andrea Casanova, Gianni Fenu, Andrea De Giacomo, Maria De Angelis, Corrado Moretti, Paola Papoff, Raffaella Ditonno, and Ruggiero Francavilla. The urinary metabolomics profile of an italian autistic children population and their unaffected siblings. *J. Matern. Fetal. Neonatal Med.*, 27 Suppl 2:46–52, October 2014.

[77] Claire Duvallet, Sean M Gibbons, Thomas Gurry, Rafael A Irizarry, and Eric J Alm. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.*, 8(1):1784, December 2017.

[78] Braden T Tierney, Yingxuan Tan, Aleksandar D Kostic, and Chirag J Patel. Gene-level metagenomic architectures across diseases yield high-resolution microbiome diagnostic indicators. *Nat. Commun.*, 12(1):2907, May 2021.

[79] Jason Lloyd-Price, Cesar Arze, Ashwin N Ananthakrishnan, Melanie Schirmer, Julian Avila-Pacheco, Tiffany W Poon, Elizabeth Andrews, Nadim J Ajami, Kevin S Bonham, Colin J Brislawn, David Casero, Holly Courtney, Antonio Gonzalez, Thomas G Graeber, A Brantley Hall, Kathleen Lake, Carol J Landers, Himel Mallick, Damian R Plichta, Mahadev Prasad, Gholamali Rahnavard, Jenny Sauk, Dmitry Shungin, Yoshiki Vázquez-Baeza, Richard A White, 3rd, IBDMDB Investigators, Jonathan Braun, Lee A Denson, Janet K Jansson, Rob Knight, Subra Kugathasan, Dermot P B McGovern, Joseph F Petrosino, Thaddeus S Stappenbeck, Harland S Winter, Clary B Clish, Eric A Franzosa, Hera Vlamakis, Ramnik J Xavier, and Curtis Huttenhower. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569(7758):655–662, May 2019.

[80] Aleksandar D Kostic, Dirk Gevers, Heli Siljander, Tommi Vatanen, Tuulia Hyötyläinen, Anu-Maaria Hämäläinen, Aleksandr Peet, Vallo Tillmann, Päivi Pöhö, Ismo Mattila, Harri Lähdesmäki, Eric A Franzosa, Outi Vaarala, Marcus de Goffau, Hermie Harmsen, Jorma Ilonen, Suvi M Virtanen, Clary B Clish, Matej Orešič, Curtis Huttenhower, Mikael Knip, DIABIMMUNE Study Group, and Ramnik J Xavier. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*, 17(2):260–273, February 2015.

[81] Finale Doshi-Velez, Paul Avillach, Nathan Palmer, Athos Bousvaros, Yaorong Ge, Kathe Fox, Greg Steinberg, Claire Spettell, Iver Juster, and Isaac Kohane. Prevalence of inflammatory bowel disease among patients with autism spectrum disorders. *Inflamm. Bowel Dis.*, 21(10):2281–2288, October 2015.

[82] Isaac S Kohane, Andrew McMurry, Griffin Weber, Douglas MacFadden, Leonard Rappaport, Louis Kunkel, Jonathan Bickel, Nich Wattanasin, Sarah Spence, Shawn Murphy, and Susanne Churchill. The co-morbidity burden of children and young adults with autism spectrum disorders. *PLoS One*, 7(4):e33224, April 2012.

[83] Dae-Wook Kang, James B Adams, Devon M Coleman, Elena L Pollard, Juan Maldonado, Sharon McDonough-Means, J Gregory Caporaso, and Rosa Krajmalnik-Brown. Long-term benefit of microbiota transfer therapy on autism symptoms and gut microbiota. *Sci. Rep.*, 9(1):5821, April 2019.

[84] Valery Danilenko, Andrey Devyatkin, Mariya Marsova, Madina Shibilova, Rustem Ilyasov, and Vladimir Shmyrev. Common inflammatory mechanisms in COVID-19 and parkinson's diseases: The role of microbiome, pharmabiotics and postbiotics in their prevention. *J. Inflamm. Res.*, 14:6349–6381, November 2021.

[85] M Čitar, B Hacin, G Tompa, M Štempelj, I Rogelj, J Dolinšek, M Narat, and B Bogovič Matijašić. Human intestinal mucosa-associated lactobacillus and bifidobacterium strains with probiotic properties modulate IL-10, IL-6 and IL-12 gene expression in THP-1 cells. *Benef. Microbes*, 6(3):325–336, 2015.

[86] Kevin R Foster, Jonas Schluter, Katharine Z Coyte, and Seth Rakoff-Nahoum. The evolution of the host microbiome as an ecosystem on a leash. *Nature*, 548(7665):43–51, August 2017.

[87] Marcel Van de Wouw, Harriët Schellekens, Timothy G Dinan, and John F Cryan. Microbiota-gut-brain axis: modulator of host metabolism and appetite. *J. Nutr.*, 147(5):727–745, 2017.

[88] Juan Miguel Rodríguez, Kiera Murphy, Catherine Stanton, R Paul Ross, Olivia I Kober, Nathalie Juge, Ekaterina Avershina, Knut Rudi, Arjan Narbad, Maria C Jenmalm, Julian R Marchesi, and Maria Carmen Collado. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.*, 26:26050, February 2015.

[89] Se Jin Song, Christian Lauber, Elizabeth K Costello, Catherine A Lozupone, Gregory Humphrey, Donna Berg-Lyons, J Gregory Caporaso, Dan Knights, Jose C Clemente, Sara Nakielny, Jeffrey I Gordon, Noah Fierer, and Rob Knight. Cohabiting family members share microbiota with one another and with their dogs. *Elife*, 2:e00458, April 2013.

[90] Stelios Georgiades, Peter Szatmari, and Michael Boyle. Importance of studying heterogeneity in autism. *Neuropsychiatry*, 3(2):123–125, April 2013.

[91] Brooke C Wilson, Tommi Vatanen, Thilini N Jayasinghe, Karen S W Leong, José G B Derraik, Benjamin B Albert, Valentina Chiavaroli, Darren M Svirskis, Kathryn L Beck, Cathryn A Conlon, Yannan Jiang, William Schierding, David J Holland, Wayne S Cutfield, and Justin M O'Sullivan. Strain engraftment competition and functional augmentation in a multi-donor fecal microbiota transplantation trial for obesity. *Microbiome*, 9(1):107, May 2021.

[92] Christopher S Smillie, Jenny Sauk, Dirk Gevers, Jonathan Friedman, Jaeyun Sung, Ilan Youngster, Elizabeth L Hohmann, Christopher Staley, Alexander Khoruts, Michael J Sadowsky, Jessica R Allegretti, Mark B Smith, Ramnik J Xavier, and Eric J Alm. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. *Cell Host Microbe*, 23(2):229–240.e5, February 2018.

[93] TEDDY Study Group. The environmental determinants of diabetes in the young (TEDDY) study: study design. *Pediatr. Diabetes*, 8(5):286–298, October 2007.

[94] Jacob T Nearing, André M Comeau, and Morgan G I Langille. Identifying biases and their potential solutions in human microbiome studies. *Microbiome*, 9(1):113, May 2021.

[95] Dan Bai, Benjamin Hon Kei Yip, Gayle C Windham, Andre Sourander, Richard Francis, Rinat Yoffe, Emma Glasson, Behrang Mahjani, Auli Suominen, Helen Leonard, et al. Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA psychiatry*, 76(10):1035–1043, 2019.

[96] James T Morton, Sharon Donovan, and Gaspar Taroncher-Oldenburg. Decoupling diet from microbiome dynamics results in model mis-specification that implicitly annuls potential associations between the microbiome and disease phenotypes—ruling out any role of the microbiome in autism (yap et al. 2021) likely a premature conclusion. 2022.

[97] Bo Yang, Yong Wang, and Pei-Yuan Qian. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 17:135, March 2016.

[98] L Palkova, A Tomova, G Repiska, K Babinska, B Bokor, I Mikula, G Minarik, D Ostatnikova, and K Soltys. Evaluation of 16S rRNA primer sets for characterisation of microbiota in paediatric patients with autism spectrum disorder. *Sci. Rep.*, 11(1):6781, March 2021.

[99] William Walters, Embriette R Hyde, Donna Berg-Lyons, Gail Ackermann, Greg Humphrey, Alma Parada, Jack A Gilbert, Janet K Jansson, J Gregory Caporaso, Jed A Fuhrman, Amy Apprill, and Rob Knight. Improved bacterial 16S rRNA gene (v4 and v4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, 1(1), January 2016.

[100] Christopher Wilks, Shijie C Zheng, Feng Yong Chen, Rone Charles, Brad Solomon, Jonathan P Ling, Eddie Luidy Imada, David Zhang, Lance Joseph, Jeffrey T Leek, Andrew E Jaffe, Abhinav Nellore, Leonardo Collado-Torres, Kasper D Hansen, and Ben Langmead. recount3: summaries and queries for large-scale RNA-seq expression and splicing. May 2021.

[101] Qiyun Zhu, Shi Huang, Antonio Gonzalez, Imran McGrath, Daniel McDonald, Niina Haiminen, George Armstrong, Yoshiki Vázquez-Baeza, Julian Yu, Justin Kuczynski, Gregory D Sepich-Poore, Austin D Swafford, Promi Das, Justin P Shaffer, Franck Lejzerowicz, Pedro Belda-Ferre, Aki S Havulinna, Guillaume Méric, Teemu Niiranen, Leo Lahti, Veikko Salomaa, Ho-Cheol Kim, Mohit Jain, Michael Inouye, Jack A Gilbert, and Rob Knight. OGUs enable effective, phylogeny-aware analysis of even shallow metagenome community structures. April 2021.

[102] Jethro S Johnson, Daniel J Spakowicz, Bo-Young Hong, Lauren M Petersen, Patrick Demkowicz, Lei Chen, Shana R Leopold, Blake M Hanson, Hanako O Agresta, Mark Gerstein, et al. Evaluation of 16s rrna gene sequencing for species and strain-level microbiome analysis. *Nature communications*, 10(1):1–11, 2019.

[103] Levi Waldron. Data and statistical methods to analyze the human microbiome. *mSystems*, 3(2), March 2018.

[104] Andrew D Fernandes, Jennifer Ns Reid, Jean M Macklaim, Thomas A McMurrough, David R Edgell, and Gregory B Gloor. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*, 2:15, May 2014.

[105] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, January 2010.

[106] Huang Lin and Shyamal Das Peddada. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *npj Biofilms and Microbiomes*, 6(1):1–13, December 2020.

[107] Jyoti Shankar. Insights into study design and statistical analyses in translational microbiome studies. *Ann Transl Med*, 5(12):249, June 2017.

[108] P R Rosenbaum and D B Rubin. The bias due to incomplete matching. *Biometrics*, 41(1):103–116, March 1985.

[109] Stan Development Team. Stan modeling language users guide and reference manual, 2022.

[110] Huang Lin and Shyamal Das Peddada. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.*, 11(1):3514, July 2020.

[111] James T Morton, Alexander A Aksenov, Louis Felix Nothias, James R Foulds, Robert A Quinn, Michelle H Badri, Tami L Swenson, Marc W Van Goethem, Trent R Northen, Yoshiki Vazquez-Baeza, Mingxun Wang, Nicholas A Bokulich, Aaron Watters, Se Jin Song, Richard Bonneau, Pieter C Dorrestein, and Rob Knight. Learning representations of microbe–metabolite interactions. *Nat. Methods*, 16(12):1306–1314, November 2019.

[112] James T Morton, Jon Sanders, Robert A Quinn, Daniel McDonald, Antonio Gonzalez, Yoshiki Vázquez-Baeza, Jose A Navas-Molina, Se Jin Song, Jessica L Metcalf, Embriette R Hyde, et al. Balance trees reveal microbial niche differentiation. *MSystems*, 2(1):e00162–16, 2017.

[113] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.

[114] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[115] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

[116] Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *arXiv preprint arXiv:2006.10256*, 2020.

[117] S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *In revision, J. Open Res. Software*, 2017.

[118] Ravin Kumar, Colin Carroll, Ari Hartikainen, and Osvaldo Antonio Martín. Arviz a unified library for exploratory analysis of bayesian models in python. 2019.

[119] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[120] Daniel McDonald, Jose C Clemente, Justin Kuczynski, Jai Ram Rideout, Jesse Stombaugh, Doug Wendel, Andreas Wilke, Susan Huse, John Hufnagle, Folker Meyer, et al. The biological observation matrix (biom) format or: how i learned to stop worrying and love the ome-ome. *Gigascience*, 1(1):2047–217X, 2012.

[121] The scikit-bio development team. scikit-bio: A bioinformatics library for data scientists, students, and developers, 2020.

[122] Thomas P Quinn, Ionas Erb, Greg Gloor, Cedric Notredame, Mark F Richardson, and Tamsyn M Crowley. A field guide for the compositional analysis of any-omics data.

[123] Michelle Pistner Nixon, Jeffrey Letourneau, Lawrence David, Sayan Mukherjee, and Justin D Silverman. A statistical analysis of compositional surveys. January 2022.

[124] Rasko Leinonen, Hideaki Sugawara, Martin Shumway, and International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.*, 39(suppl_1):D19–D21, 2010.

[125] Kalen Cantrell, Marcus W. Fedarko, Gibraan Rahman, Daniel McDonald, Yimeng Yang, Thant Zaw, Antonio Gonzalez, Stefan Janssen, Mehrbod Estaki, Niina Haiminen, Kristen L. Beck, Qiyun Zhu, Erfan Sayyari, James T. Morton, George Armstrong, Anupriya Tripathi, Julia M. Gauglitz, Clarisse Marotz, Nathaniel L. Matteson, Cameron Martino, Jon G. Sanders, Anna Paola Carrieri, Se Jin Song, Austin D. Swafford, Pieter C. Dorrestein, Kristian G. Andersen, Laxmi Parida, Ho-Cheol Kim, Yoshiki Vázquez-Baeza, and Rob Knight. Empress enables tree-guided, interactive, and exploratory analyses of multi-omic data sets. *mSystems*, 6(2), 2021.

Methods

Search strategy and inclusion criteria

We performed a systematic search for published and/or publicly deposited or not yet published and/or publicly available human microbiome, metabolome, immunome, transcriptome AND autism/ASD datasets in several NCBI databases (PubMed, SRA, and BioProject), UCSD's MassIVE resource, the PsychENCODE consortium, the American Gut Project, and from individual research groups worldwide. About half of 70+ studies we identified were already deposited on public data repositories or were made directly available to us by the research groups.

Most studies consisted of heterogenous—no genotype or phenotype stratification—ASD and neurotypical age- and sex-matched cohorts and had one or two datasets (microbiome [16S, shotgun metagenomic sequencing (SMS)], metabolome [urine/serum/fecal], immunome [cytokines], transcriptome [RNAseq], dietary survey, behavioral survey) associated with them, with only a few studies having three or more omic datasets associated with them (Table 1). We adopted a multi-cohort and multi-omics meta-analysis framework that allowed us to combine independent and dependent omic data sets in one overall analysis[43, 44]. In total, we analyzed 597 ASD-control pairs. To reduce the batch effects and noise associated with primer choice in the 16S datasets, a major confounder in microbiome analyses, we restricted the 16S datasets to include only those targeting the variable region V4 of the bacterial ribosomal RNA, a region exhibiting higher heterogeneity and lower evolution rates than other variable regions[97, 98, 99]. Our analysis included 16S datasets obtained targeting the V4 region exclusively, the V3-V4 region, or the V4-V5 region.

The final metabolomic meta-analysis we present here consists of the combined analysis of only four independently preprocessed, normalized, and analyzed metabolomic datasets. Despite several more ASD-related datasets being available, the disparity in mass spectrometric technologies used to generate them, which results in the detection of different subsets of metabolites, precluded their side-by-side comparison (Table 1). For example, targeted mass spectrometry enables the precise determination of concentrations for

a finite number of metabolites, whereas untargeted mass spectrometry detects up to two or three orders of magnitude more metabolites but is compositional in nature and thus does not yield absolute abundances. Furthermore, batch effects due to sample-processing such as differences in reagents, sample storage and mass-spectrometry instruments can introduce unwanted variation in both the abundances and the detected molecular features [26]. One additional obstacle we encountered was the proprietary nature of many of the metabolomic datasets that made it impossible to access the raw data and run standardized workflows.

Of the 40 transcriptomic datasets that were available in recount3 [100], the vast majority were obtained from studies with model animals, and only four of them had been obtained from postmortem processing of brain samples from autistic and neurotypical individuals. These four datasets collected different brain tissue types, including from the amygdala, the prefrontal cortex, the anterior cingulate and the dorsolateral prefrontal cortex.

Data processing

16S amplicon and shotgun metagenomics samples were downloaded from the SRA. The 16S amplicon samples were processed using Deblur and subsequently mapped to bacterial whole genomes captured in the Web of Life using Woltka[101]. This is done in order to make the amplicon data comparable to shotgun metagenomics data. Bacterial abundances were extracted from shotgun metagenomics samples using Woltka and Bowtie2. Viral abundances were extracted from shotgun metagenomics samples using GPD and BWA. RNA expression data were obtained directly from recount3 [100]; the four metabolomics datasets were provided by the authors.

The approach of mapping both 16S amplicon sequences and SMS samples to a common set of microbial reference genome provided a consistent taxonomic annotatation between the different 16S amplicon types and SMS datasets. However, there are notable limitations in taxonomic resolution in all of these datasets. For instance, multiple Bifidobacterium species that are associated with non-human microbiomes were found including *B. asteroides* (honeybee), *B. callitrichos* (marmoset), *B. choerinum* (pig) and *B. sanguine* (tamarin). We observed that these 16S amplicons were multi-mapped across many different Bifidobacterium genomes, which highlights the lack of species level resolution highlighted in previous studies [102]. Similarly, taxonomic profiles obtained from shotgun sequencing are known to have elevated false positive rates taxonomic identifications due to high genome similarity between microbial reference genomes [59].

To enable age- and sexmatching, a bipartite matching between ASD and neurotypical subjects was performed using age and sex covariates. Subjects that could not be matched were excluded from the meta-analysis. Amongst the 16S and SMS datasets, there were multiple longitudinal datasets. To integrate these datasets into the cross-sectional analysis, we only picked the first time point for each subject.

Differential ranking analysis

One of the most common approaches to evaluating microbiome and other omic studies consists of determining differences in the abundances of microbial taxa, human metabolites or other omic features between cases and controls [103]. Such differential abundance analysis is typically performed by computing the log fold changes between the case and control groups[104, 46, 105]. However, confounders such as sex-, age-, and geography-related batch effects, compositionality, high-dimensionality, over-dispersion, and sparsity, prevented a reliable estimation of differential abundances and thus compromised the side-by-side comparison of these differential abundances across studies in the manner of a traditional meta-analysis [106, 107]. Here, we set out to overcome these inherent limitations of traditional meta-analyses by developing a generalizable approach for controlling for select confounders that would help reveal a comprehensive picture of ASD-specific omic signals.

To minimize confounder effects, we developed a Bayesian differential ranking algorithm that used bipartite matching to optimize the age- and sex-based pairing of ASD and control subjects within each dataset. This approach helped both control for potential age and sex confounders and minimize batch effects such as sample collection method, sample processing protocol, and geographical provenance [108]. These Bayesian models were fitted via MCMC using Stan [109]. Conceptually, this allowed us to compute log-fold change differences of microbes between age- and sex-matched subjects, but because we did not have absolute abundance information we could only estimate this log-fold change up to a constant [31] (Supplemental methods). To determine if there was a significant difference between the age- and sex-matched pairs, we constructed an effect size metric utilizing our model's uncertainty estimation (see Supplemental methods for more details). To show that this model is relevant for biological data, we built a simulation benchmark using the 16S count data. Specifically, we fitted the Bayesian model on the 16S cross-sectional cohort, and simulated microbial counts based on those estimated parameters. We showed that the ground truth log-fold changes across all of the microbes are within the 95% credible intervals estimated by our algorithm. When we evaluated our Bayesian model fit on the 16S, SMS and RNAseq datasets, our models fits achieved Rhat values below 1.1 and ESS values above 300, indicating that the draws from the posterior distribution are reliable [109].

To identify microbes that were ASD-specific or neurotypical specific, we fitted a Gaussian mixture model on top of the estimated log-fold changes, binning the taxa into three different groups, those taxa hypothesized to be more abundant in neurotypical controls, those more abundant in ASD children and those that are equally prevalent in both groups, or are "neutral".

This strategy was inspired by the work done with ANCOM-BC [110]. The major difference in our approach compared to ANCOM-BC is that our approach assumes a negative binomial distribution for modeling counts and allows for Bayesian model uncertainty quantification. The reference frame in the cross-sectional analysis refers to the average abundance of the microbes that are categorized as neutral in Figure 1a. These same microbes were used to construct a reference frame in the FMT analysis to standardize all of the time

points. The FMT analysis used the same matching strategy, but instead of matching on age and sex, the matchings were performed on the subjects to compare different time points.

The heatmap shown in Figure 1 displayed the log-fold changes for each case-control pair. To do this, a robust CLR transform was performed and all zeros were imputed to the mean abundance for visualization purposes. The case-control log-fold changes were computed for each pair as highlighted in Figure S7.

Other methods

We fitted Random Forests models on nine 16S datasets and on three SMS datasets. We randomly split the samples into 90/10 training and test splits, performed a 10-fold cross-validation on the training datasets to obtain optimal model parameters, and computed predictions on the held-out test dataset. PERMANOVA with Bray-Curtis distances was used to determine if confounding variation due to household, age and sex were statistically significant in the sibling cohorts.

We used MMvec [111] to perform the diet-microbe co-occurrence analysis. Here, microbes were used to predict dietary intake. This analysis enabled the estimation of conditional probabilities, namely the probability of observing a dietary compound given the microbe was already observed. To estimate these conditional probabilities, MMvec performs a matrix factorization, identifying the factors that explain the most information in these interactions. We compared the MMvec microbial factors against the cross-sectional log-fold changes. We then compared the MMvec dietary factors against t-statistics that measure the differences in dietary compounds between ASD and neurotypical children.

To identify candidate viral-microbe interactions, we ran MMvec on each of the SMS datasets. We then pulled out the top co-occurring viral taxa for each microbe that had a conditional log-probability greater than 1, amounting to 78580 microbe-viral interactions. Then we filtered out the microbe-viral interactions that were not present in the GPD [70], leaving 31276 microbial-viral interactions.

We used Songbird [31] to perform the cytokine-microbe analysis via a multinomial regression that used the cytokines to predict microbial abundances. We reported biased microbial log-fold changes with respect to cytokine concentration differences. Pearson correlation was used to determine the agreement between the 16S cross-sectional microbial differentials and the microbe-cytokine differentials. To directly link these microbial abundances to the cytokine concentrations, we computed log-ratios, or balances, of microbes for each sample. For example, for IL-6 the numerator consisted of the top 30 microbes that are estimated to increase the most in abundance when IL-6 concentration increased, and the denominator consisted of the bottom 30 microbes which are estimated to be the most decreased when IL-6 concentration increases. Once these partitions are defined, the balances for each sample are computed by taking the log-ratio of the average abundance of the numerator group and the denominator group. See Morton et al 2017 for more details behind balances [112]. Pearson correlation between these balances and the cytokine concentrations are then computed to measure the agreement between the microbial abundances and the cytokine concentrations.

To identify key microbial genes, we performed a comparative genomic analysis in which we binned the microbial genomes into those associated with ASD and those associated with control subjects. Using a binomial test, we were able to determine if a particular gene was more commonly observed in ASD-associated microbes than by random chance. Significant microbial genes and RNA transcripts were subsequently mapped to KEGG pathways. To directly compare the two contrasting omics levels and gauge metabolic similarity, we retrieved all the molecules involved in both the microbial and human pathways and calculated their intersection. Since the metabolomics datasets were not directly comparable, we performed Wilcoxon tests on age- and sex-matched metabolomics samples within each cohort separately. While our analysis revealed multiple metabolites that were below the 0.05 p-value threshold, none of these metabolites passed the FDR corrected threshold.

Software Availability

Software implementation of our Bayesian age-sex matched differential ranking algorithm can be found at https://github.com/flatironinstitute/q2-matchmaker

We want to acknowledge Matplotlib [113], Seaborn [114], Scipy [115], Numpy [116], Xarray [117], Arviz [118], Scikit-learn [119], biom-format [120] and Scikit-bio [121] for providing the software foundation that this work was built upon.

Tables and Figures

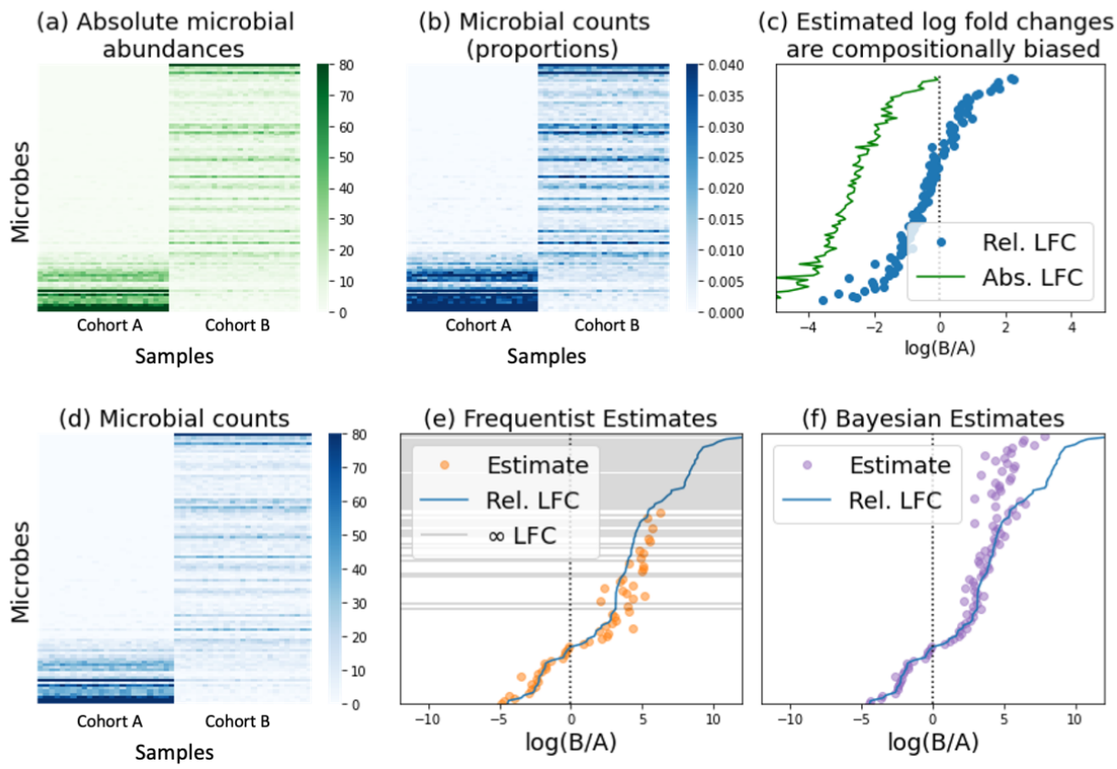

Figure BOX 1: **TACKLING METAGENOMIC UNKNOWNS**
Metagenomic sequence data present unique quantification challenges due to a lack of total microbial load measurements, which precludes the determination of absolute microbe abundances, and to limitations brought about by sampling and sequencing depth limitations, which result in an incomplete representation of the metagenome. We devised a Bayesian differential ranking algorithm to address both these challenges, the compositional challenge and the zero-inflation challenge.
The compositional challenge: Most sequencing count datasets lack absolute abundance information in the form of cells, colony forming units, or transcripts per volume. This limitation preempts the reliable estimation of log fold changes and is a defining characteristic of compositional data that can lead to excessive false positives or false negatives depending on the magnitude of the change in absolute abundances [31, 45]. As illustrated in panels a) through c), microbial counts (a) are typically converted into proportional abundances (b) that are then used to compute log-fold ratios. Fold change calculations adopt the general formula $\frac{B}{A} = \frac{N_B p_B}{N_A p_A} = \frac{p_B}{p_A} \times \frac{N_B}{N_A}$, where A and B represent the two samples being compared, p_A and p_B represent the microbial proportions in A and B, and N_A and N_B represent the total number of microbes in A and B, also known as the ground truth. A key limitation of sequencing count data is their lack of proportionality to the corresponding absolute abundances in the original samples due to sequencing depth constraints [122]. Our inability to observe N_A and N_B introduces a bias that ultimately prevents us from performing FDR correction to identify differentially abundant microbes [123]. This bias depends on the change in microbial population size, with large population shifts leading to increased false positive and false negative rates, and an overall skewed representation of the ground truth (c).
The zero-inflation challenge: Sampling errors and shallow sequencing lead to disproportionately high numbers of zero counts, especially for microbes present in low abundances (d). Multinomial, Poisson and Negative Binomial distributions have been used to explicitly handle zero counts [46]. However, estimating log-fold differentials remains problematic when microbes are not observed in any of the samples in one group since log 0 is $-\infty$ and thus the true log-fold change of a zero-count microbe can not be determined (e). Bayesian inference avoids this problem by introducing a prior that prevents nonsensical log-fold change estimates (f). Specifically, this introduces a rounded-zero assumption whereby all microbes have a non-zero chance of being observed. Panel h highlights what these log-fold changes would look like using a Dirichlet prior, where every microbe has the same probability of being observed before collecting data.

Organism	Body type	Data type	Num Studies	Num Subject pairs	References
Human	Postmortem brain tissue	RNAseq	4	49	(Velmeshev et al. 2019; Wright et al. 2017; Herrero et al. 2020) SRP072713
Human	Serum	Immune markers	1	22	(Zurita et al. 2020)
Human	Serum	Metabolome	2	50	(Needham et al. 2021; Kuwabara et al. 2013)
Human	Urine	Metabolome	1	26	(Noto et al. 2014)
Human	NA	Dietary survey	1	26	(Berding and Donovan 2019)
Human	NA	Behavioral survey	1	28	(Kang et al. 2019)
Microbial	Fecal	Metabolome	2	43	(Needham et al. 2021; Kang et al. 2018)
Microbial	Fecal	16S amplicon	10	346	(Berding and Donovan 2019; Zurita et al. 2020; Dan et al. 2020; J. Zhu et al. 2021; Fouquier et al. 2021; Zou et al. 2020; Kang et al. 2019, 2017; Son et al. 2015; David et al. 2021) SRP299486
Microbial	Fecal	Shotgun metagenomics	3	83	(Averina et al. 2020; M. Wang et al. 2019; Dan et al. 2020)

Table 1: ASD omic datasets included in this study All sequencing datasets were retrieved from the SRA. [124].

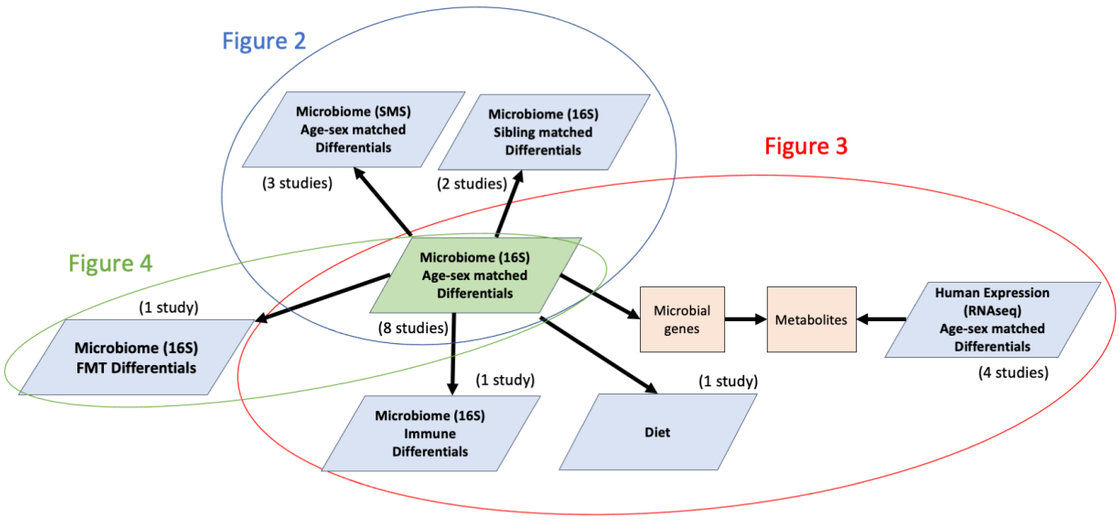

Figure 1: The structure of our meta-analysis across multiple omics levels. For Figure 2, microbial differential abundance on 16S data from age-sex matched cohorts was cross-referenced against sibling matched cohorts and SMS data from other age-sex matched cohorts. For Figure 3, these same 16S microbial differential abundances were cross-referenced against cytokine profiles, dietary surveys and pathways from RNAseq. For Figure 4, the differentially abundant microbes from the age-sex matched analysis was cross-referenced against the Kang et al FMT trial.

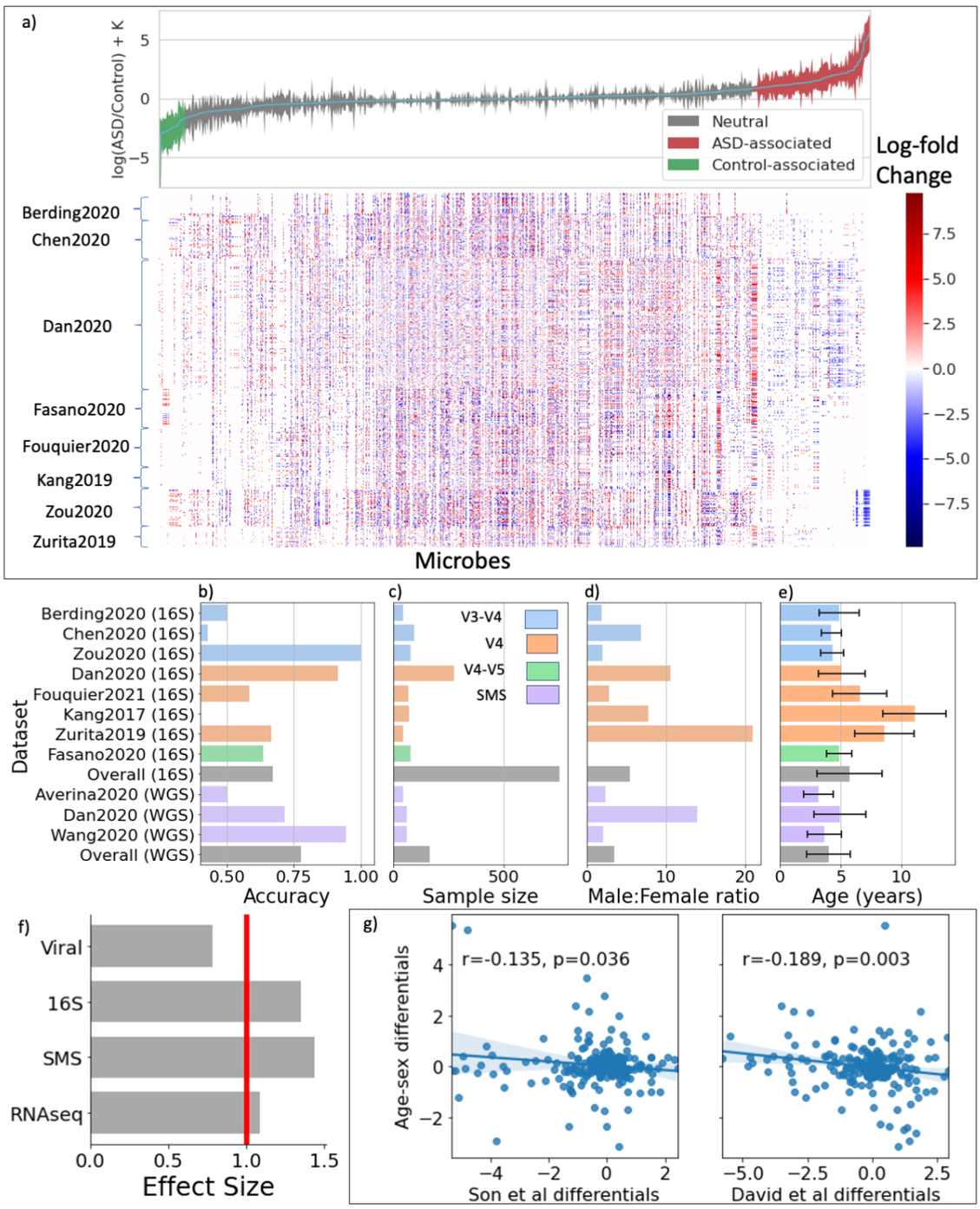

Figure 2: Differential ranking analysis across omics levels. (a) Global microbial 16S log-fold changes between age- and sex-matched ASD and control individuals. Error bars represent the 95% credible intervals. Heatmap showing all CLR transformed microbial differentials for each age- and sex- matched ASD-control pair across all cohorts. (b-e) Held-out random forests ASD classification accuracy, sample size, male:female ratio and age distributions across all 16S and shotgun metagenomics datasets analyzed in this study. V3-V4, V4, and V4-V5 refer to the variable region of the bacterial ribosomal RNA analyzed; SMS refers to shotgun metagenomic sequencing. (f) Effect sizes of different omics levels: viral, 16S, SMS, and RNAseq. (g) Comparisons of sibling-matched differentials from two different studies to the global age- and sex-matched differentials.

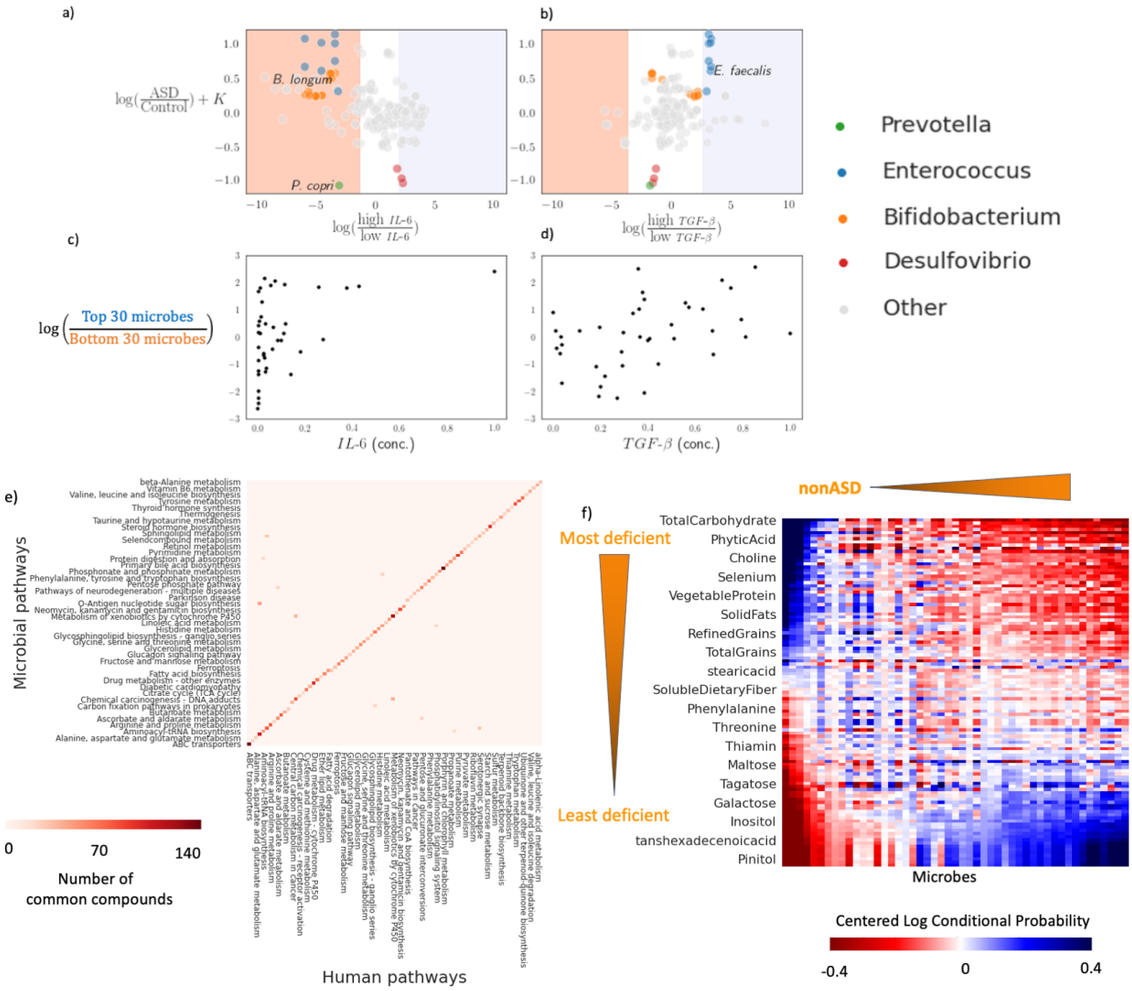

Figure 3: Characterizing the associations between differentially abundant microbes in ASD and cytokines, gene expression in the brain, and dietary patterns. (a-b) Comparison of microbial differentials obtained from age- and sex-matching and cytokine analysis. LFC denotes log-fold change of microbial abundances with respect to a specific cytokine. (c-d) Microbial log-ratios constructed from 30 top and bottom most differentially abundant microbes corresponding to each cytokine. (e) Heatmap showing the overlap of molecules between ASD-enriched pathways in the microbiome and in the brain. (f) Co-occurrence analysis between diet and microbes in ASD.

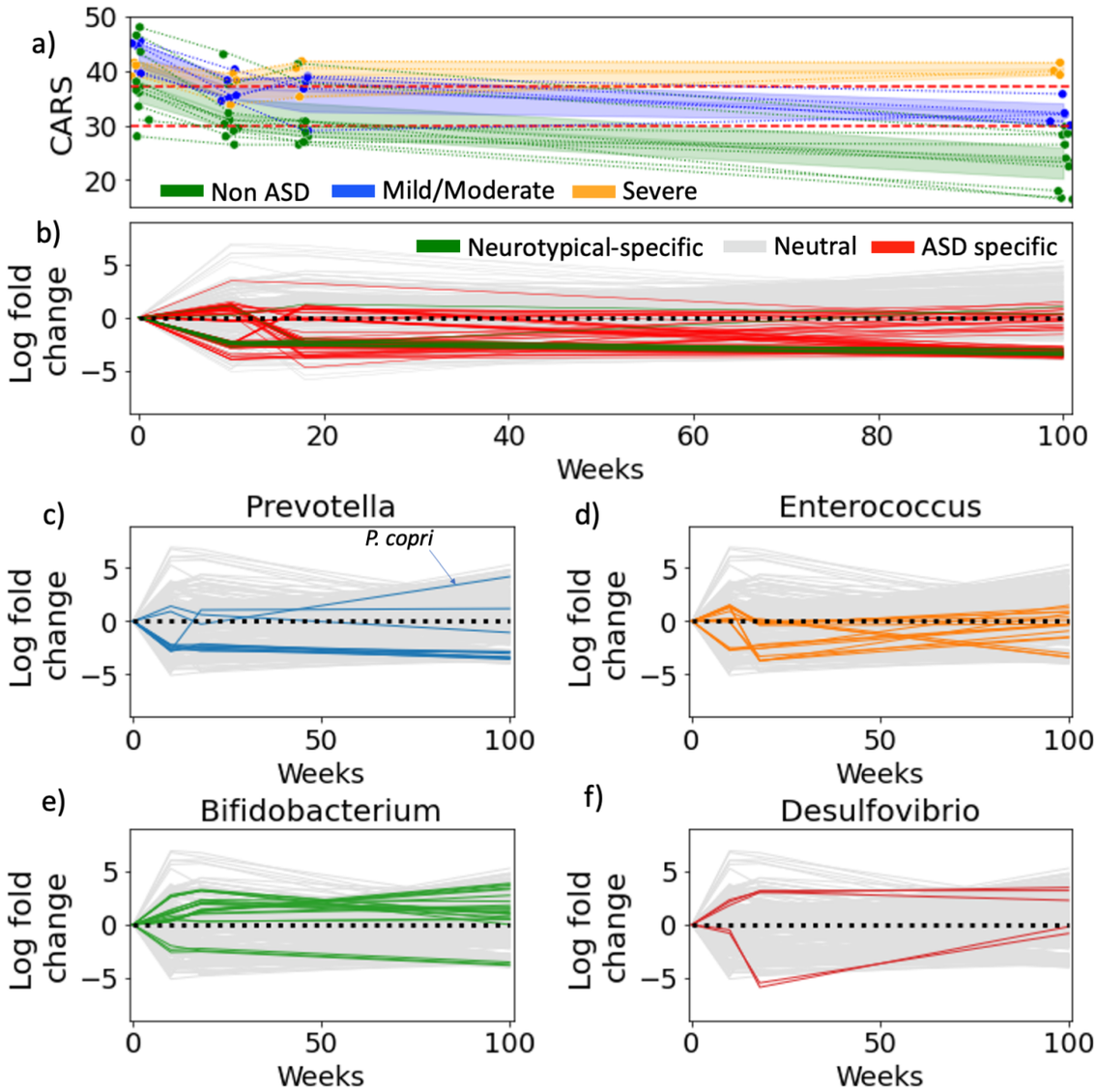

Figure 4: Fecal matter transplants have long-lasting effects on autism gut microbiomes. (a) The improvement of CARS for each ASD child over time. The children are split into 3 groups, non-ASD, mild/moderate and severe based on whether their CARS score fell below 30, between 30-37 or greater than 37 (b) Microbial log-fold changes over time: the time series was generated by calculating log-fold changes between time points for each microbe. ASD-specific microbes highlighted in red were determined in the cross-sectional study. (c-f) Microbial log-fold changes are re-colored with genera highlighted in cytokine comparisons.

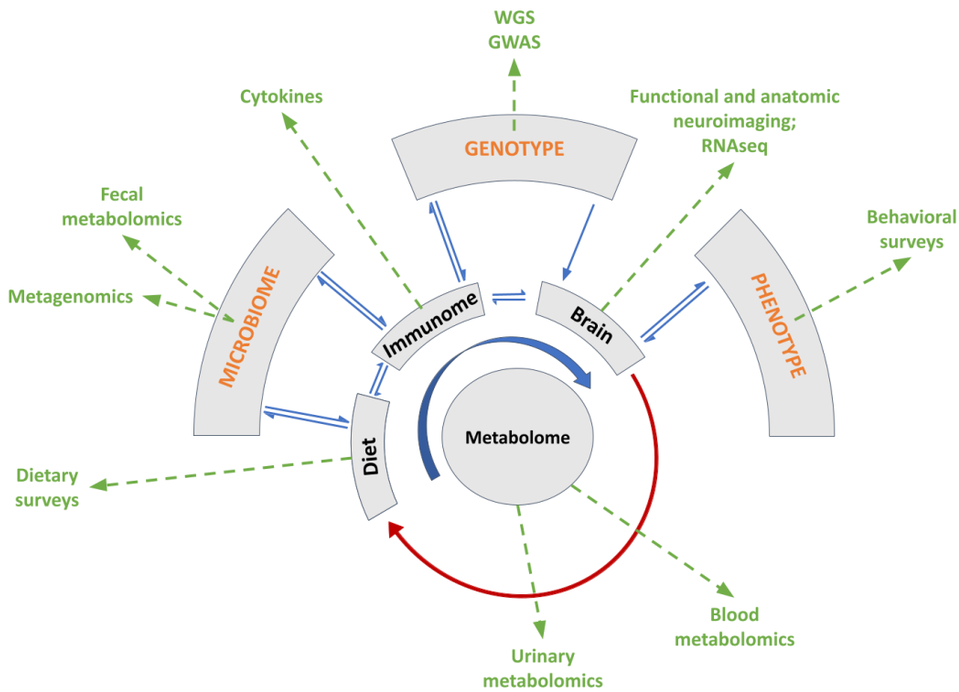

Figure 5: Proposed functional architecture of ASD Hypothesized causal graphs underlying relevant omic levels along the GBA in ASD, and experimental considerations for future studies. Blue arrows denote causal direction, red arrows indicate feedback loops and green arrows indicate measurable data types.

Supplemental Materials

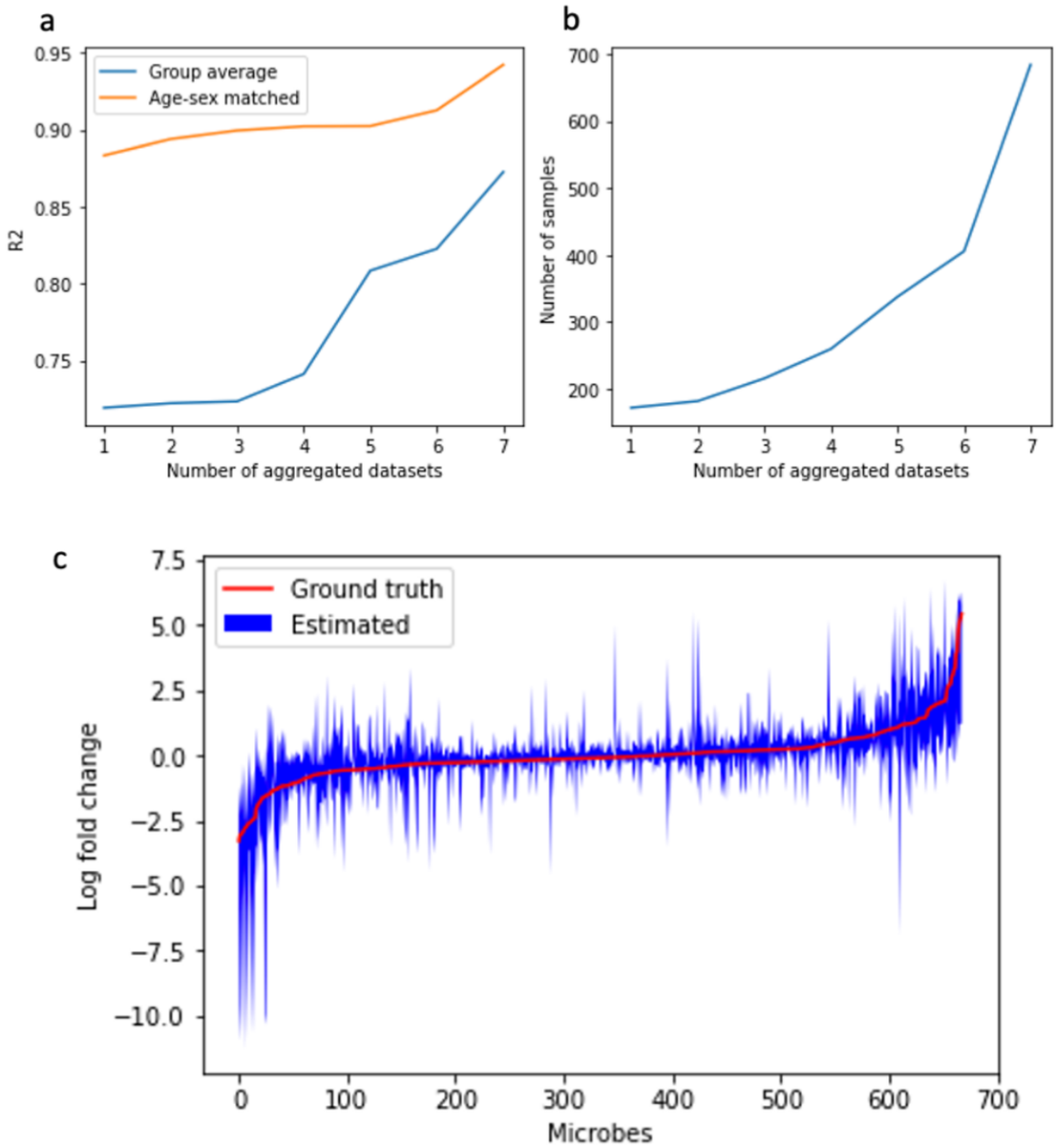

Figure S1: Benchmarks. (a) Comparison of age- and sex-matching approach compared to standard group averaging with respect to dataset size across 7 of the 10 16S studies (excluding Kang et al, David et al and Son et al). (b) Number of samples analyzed. The x-axis represents the number of aggregated datasets, the y-axis on the left panel is the average R^2 metric to measure the model error, and the y-axis on the right panel is the number of samples in the aggregated dataset. (c) Differential abundance estimation derived from a simulated datasets modeled from the cross-sectional cohort from the 8 16S datasets.

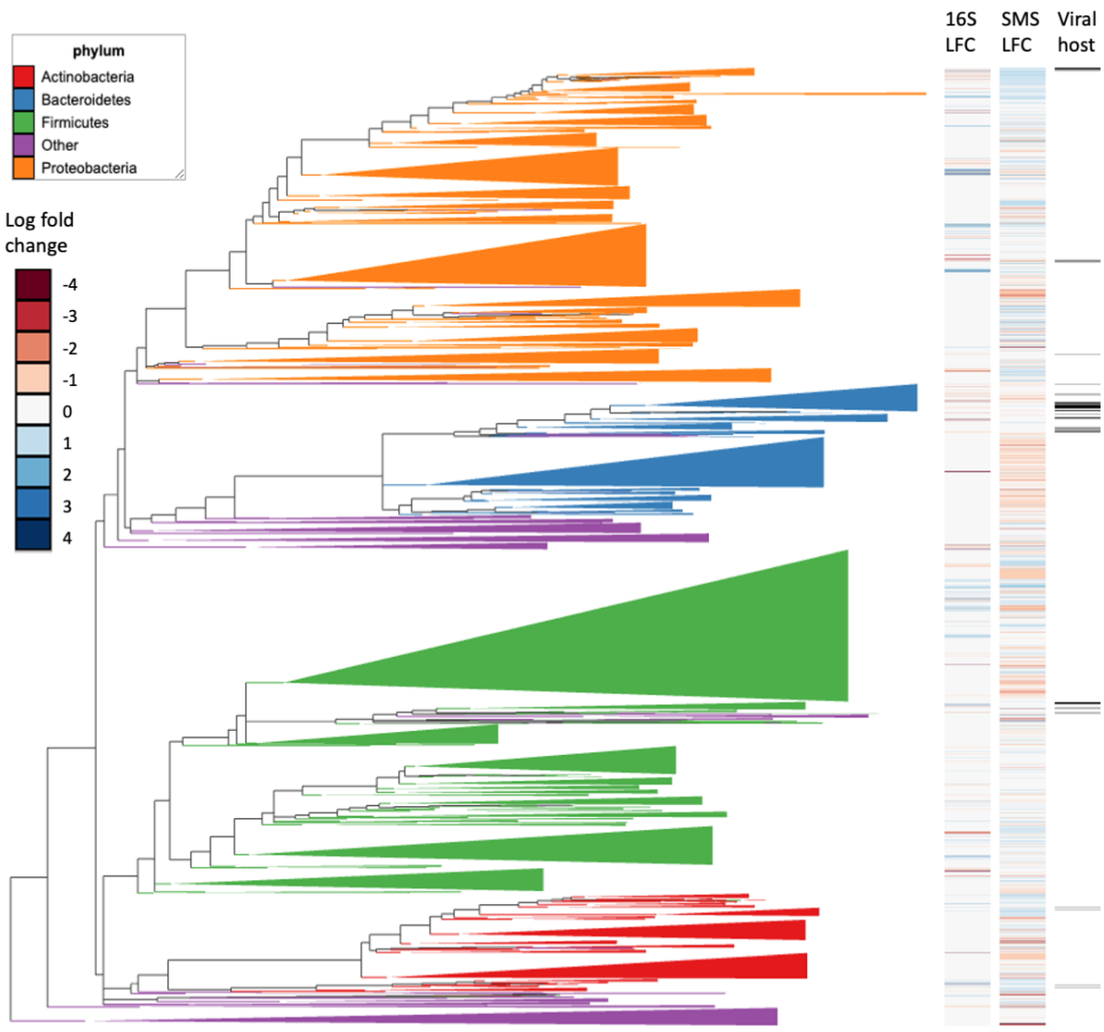

Figure S2: Phylogenetic visualization of microbes with respect to the differentials computed from 16S and SMS differentials. Microbes that were annotated as viral hosts were also highlighted. The phylogenetic visualization was generated using Empress [125].

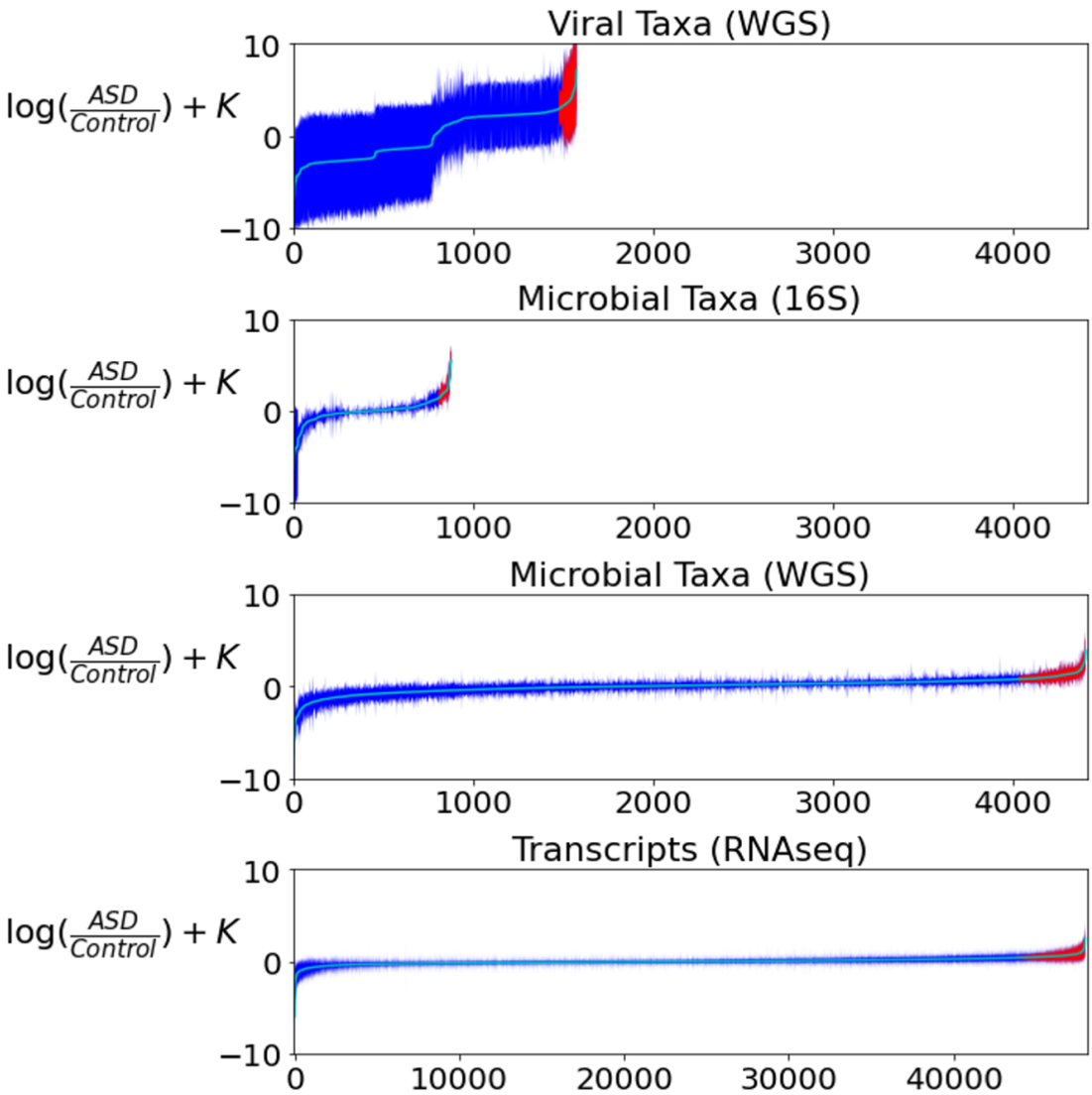

Figure S3: Global differential ranking trends observed for the virome, 16S, SMS, and RNAseq datasets analyzed in this study. The x axis for the virome, 16S and SMS datasets is equivalent to showcase the differences in feature counts; the x axes for the RNAseq dataset is larger by a factor of 10, illustrating the stark difference in number of features of this dataset compared to the other three.

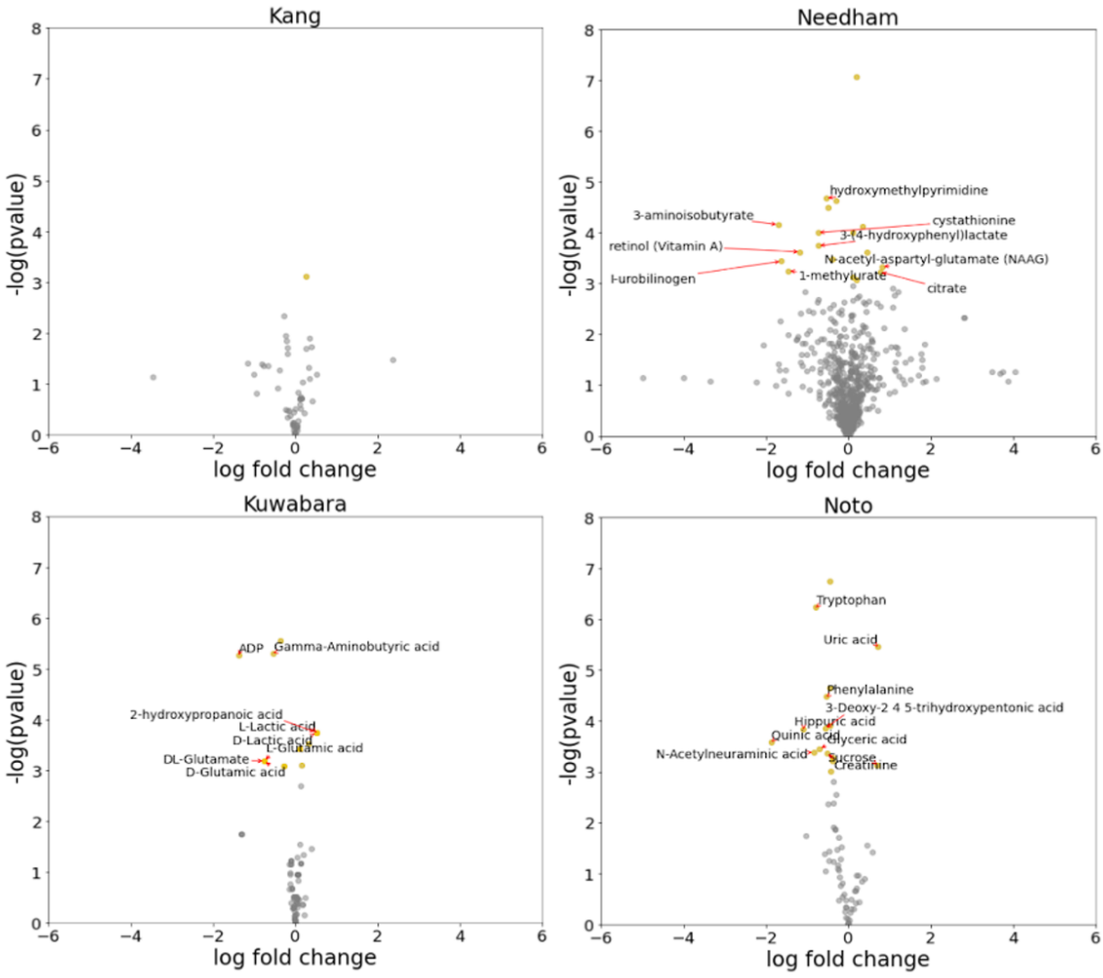

Figure S4: Metabolomics differential ranking analysis across four studies. Paired t-tests were performed to identify differentially abundant metabolites. The metabolites shown in Needham et al consist of both fecal and serum metabolites. None of the metabolites had significant log-fold changes after applying FDR correction.

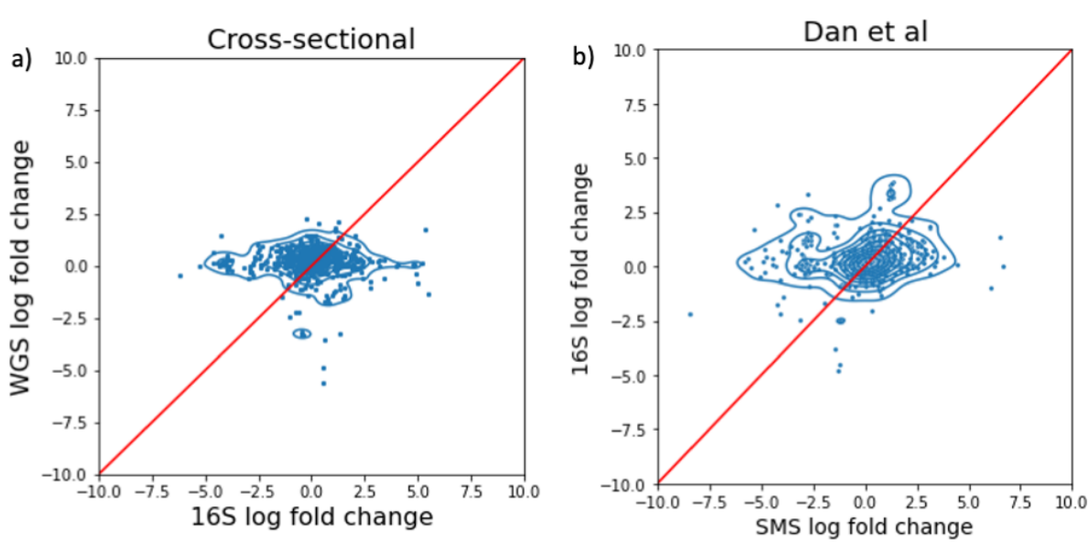

Figure S5: Comparison of log-fold changes computed from 16S and SMS. (a) Comparison of differentials obtained from 16S and SMS on the cross-sectional datasets. (b) Comparison of differentials obtained from 16S and SMS on the same samples from Dan et al.

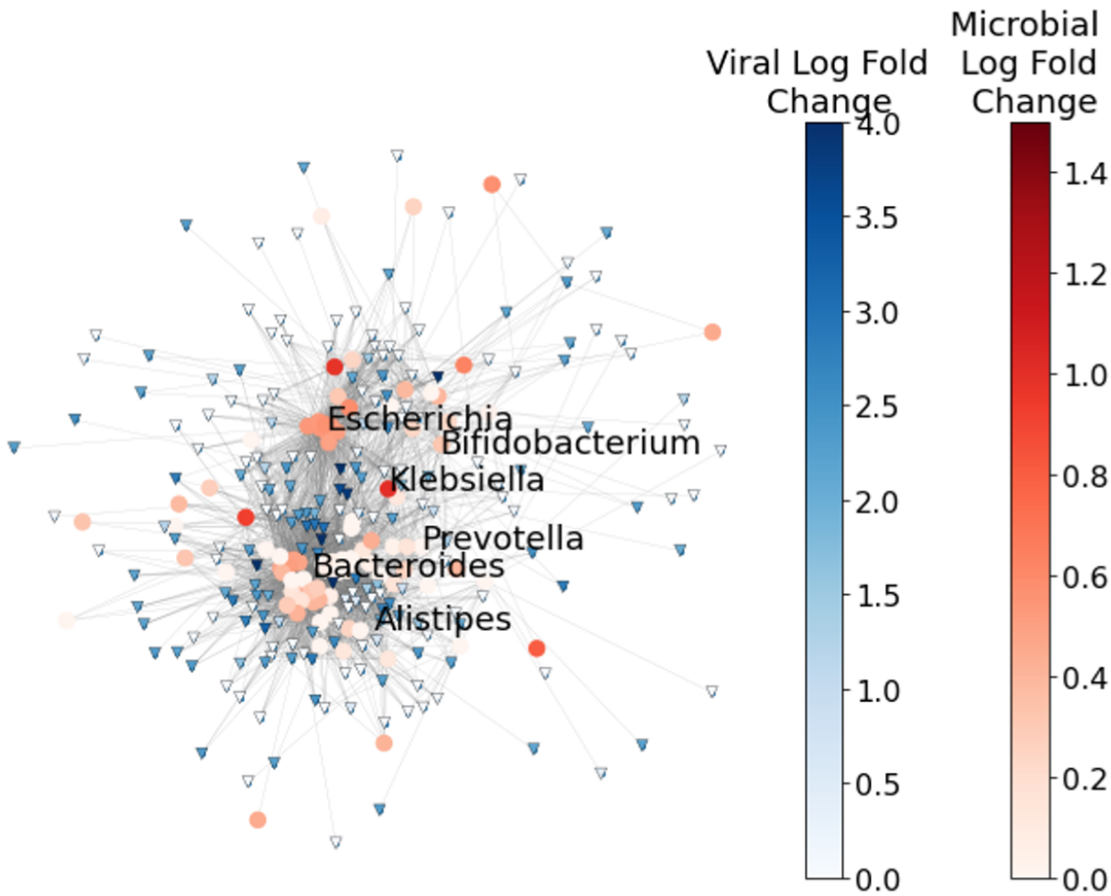

Figure S6: Microbe-viral co-occurrence network estimated using MMvec. Microbes are colored red and viruses are colored blue. Edges are drawn between microbes and viruses if they are highly co-occurring and the interaction was annotated in GPD.

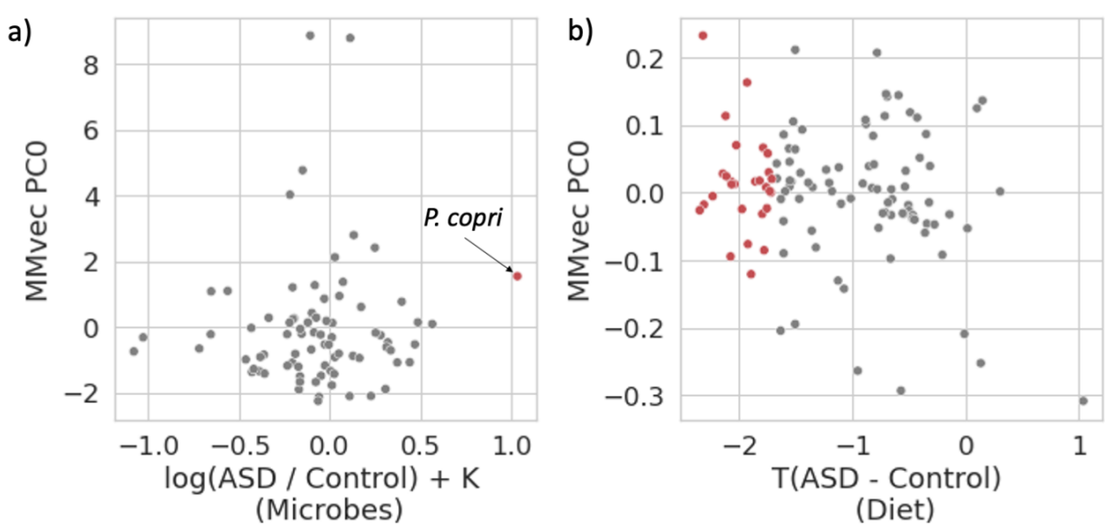

Figure S7: Principal components analysis of microbe-diet interactions. The top principal component explaining the variation in the microbe-diet co-occurrences is compared against the (a) microbial log fold change and (b) dietary differences computed from a t-statistic. Dietary compounds that are significantly significant before FDR correction are highlighted in red.

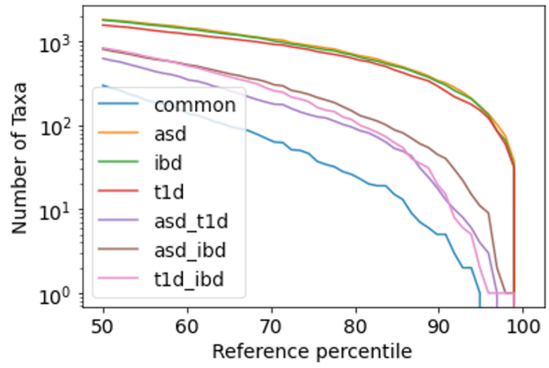

Figure S8: Comorbidity analysis. The number of taxa in common between diseases from differential ranking analysis are shown, with a focus on the intersections between ASD, IBD and T1D.

Table S1: Table of statistics for 16S differentials, including mean log-fold change, standard deviation log fold change, 95% credible intervals and taxonomy for each microbe.

Table S2: Table of statistics for SMS differentials, including mean log-fold change, standard deviation log fold change, 95% credible intervals and taxonomy for each microbe.

Table S3: Table of statistics for RNAseq differentials, including mean log-fold change, standard deviation log fold change, 95% credible intervals and taxonomy for each transcript.

Table S4: Table of statistics for viral differentials, including mean log-fold change, standard deviation log fold change, 95% credible intervals for each virus.

Table S5: Permanova breakdown of sibling matched cohorts looking at the confounding variation due to age, sex and household.

Table S6: Microbial log fold changes due to cytokine differences, including mean log-fold change for each cytokine.

Table S7: Microbe virus co-occurrences, where entries represent the centered log-probability of a microbe and a virus both present for a given sample

Table S8: A list of paired microbe and human pathways in addition to the number of overlapping metabolites

Table S9: Microbe diet co-occurrences, where entries represent the centered log-probability of a microbe and a dietary compound both present for a given subject.

Table S10: Microbial log fold changes between paired time points across all of the subjects in the FMT study.