Title

Using Multitissue Multiomics Systems Biology to Understand Tissue-specific Networks of Autism Spectrum Disorders

Author

Gill, Cameron Elias

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Using Multitissue Multiomics Systems Biology to Understand Tissue-specific Networks of Autism Spectrum Disorders

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Physiological Sciences

by

Cameron Elias Gill

2024

ABSTRACT OF THE THESIS

Using Multitissue Multiomics Systems Biology to Understand

Tissue-specific Networks of Autism Spectrum Disorders

by

Cameron Elias Gill

Master of Science in Physiological Sciences

University of California, Los Angeles, 2024

Professor Xia Yang, Chair

The genetic heterogeneity of autism spectrum disorder (ASD) has been a long-standing obstacle in our understanding of the pathogenic mechanisms of the disease, as the genetic risk of ASD is made up of numerous common variants and rare *de novo* or inherited variants. Previous studies have focused primarily on identifying rare variants and their impact on brain cortical cell types, and these mutations have been found to primarily affect neurodevelopment by perturbing neuronal functions. By contrast, common variants have been found to contribute substantially to ASD heritability, but remain understudied. This suggests a need to consider both rare and common variants of ASD to understand the genetic mechanisms of the disease. Furthermore, previous studies have implicated the subcortical areas of the brain and other organ and tissue systems such as the digestive and immune systems in ASD, but tissue-specific mechanisms

remain poorly explored. To address these knowledge gaps, this thesis aims to identify gene networks and pathways informed by ASD common variants in both brain and peripheral tissues across the body and further examine whether these networks also capture genes informed by rare variants. We achieve this by integrating tissue level RNA sequencing data, genome wide association study (GWAS) summary statistics, and tissue-specific transcriptional regulatory networks using the multiomics integration method *Mergeomics*. Furthermore, we infer tissue-specific key regulatory genes governing the pathways and networks of ASD common variants by leveraging tissue-specific Bayesian gene regulatory networks. Lastly, we investigate whether the gene networks informed by ASD common variants converge with those of known ASD rare variants. Our multitissue multiomics systems studies incorporating both common and rare variants reveal the key tissues, biological pathways, and gene network regulators of ASD and identify key similarities and differences between ASD common and rare variants in tissue and network specificity.

The thesis of Cameron Elias Gill is approved.

Weizhe Hong

Roy Wollman

Xia Yang, Committee Chair

University of California, Los Angeles

2024

To my family, for their continual support through my education.

Table of Contents

Acknowledgements

I would like to thank my committee members, Professors Weizhe Hong and Roy Wollman for their helpful feedback during my project. I would like to thank my principal investigator Professor Xia Yang for her endless support and guidance throughout my program. I would like to thank Dr. Yanning Zuo, Dr. Jessica Ding, and Dr. Montgomery Blencowe for their mentorship during my analyses. I would like to thank my family, my partner, my friends, and my lab mates for their uplifting encouragement during graduate school.

Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that manifests through social and communication deficits and various behavioral abnormalities [1, 2]. The prevalence of ASD has continued to grow in recent years, with current epidemiological evaluations estimating that more than 50 million people have been diagnosed with a form of autism, which equates to 1 in 132 individuals in a given population [3]. A diagnosis is based on various criteria established by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 [4]. Social and communication deficits can be categorized as challenges in social-emotional reciprocity, nonverbal communicative behaviors, and developing and understanding relationships. Behavioral abnormalities include restricted, repetitive motor movements, adherence to routines and ritualized patterns, hyperfocused interests, and hyper- or hyporeactivity to sensory stimulation. Severity is ranked on three levels: Level 3, requiring very substantial support; Level 2, requiring substantial support; and Level 1, requiring support. To be diagnosed with ASD, a child must display each of these deficits in a persistent fashion during early developmental stages, and demonstrate at least two of the behavioral abnormalities [4]. These symptoms must also cause clinically significant impacts on everyday functioning in social and occupational contexts.

The wide range of social and communication deficits and behavioral abnormalities that make up the symptoms of ASD, as well as the various levels of severity, indicate that there is a diversity of ASD pathogenesis. This diversity of symptoms is likely rooted in substantial genetic heterogeneity, as numerous rare mutations and common genetic variants with varying effect sizes have been identified in ASD [5]. Despite having a greater risk of causing significant autism syndromes, rare variants comprise only around 1% of patients with autism [6]. Common

variants, by contrast, are more prevalent in a population and collectively contribute to >50% of ASD heritability, but have small effect sizes when compared to rare mutations [5, 7]. Thus, common variants likely have an important role in the pathogenesis of ASD given their prevalence in the population.

Both common and rare variants of ASD likely affect molecular and cellular pathways and functions in key tissues and organ systems related to ASD pathogenesis. Previous studies have identified a number of brain regions that are associated with ASD. For instance, the frontal and temporal cortical regions have shown abnormal gene expression patterns in autistic patients compared to typically developing children [8]. Two regions that are relevant for complex cognitive processes, the anterior cingulate cortex and the amygdala, have exhibited decreased neuronal activity and abnormal growth, respectively, in autistic patients [9, 10]. There has also been evidence of non-symmetric development in the lateral ventricles and hippocampus when comparing autistic and typical children [11]. Other implicated brain regions include the prefrontal, parietal, and visual cortices, cerebellum, caudate nucleus, and various gyri and sulci [12, 13].

In addition to the previously identified brain regions, a number of peripheral organ systems and tissues have also been implicated in ASD association. For example, the immune system has been previously studied in the context of how immune dysregulation causes outcomes such as altered neurodevelopment and behavior [14]. The microbiota-gut-brain axis has also been explored due to the interactions between commensal bacteria, immune cells, enteric nerves, and neurotransmitters, as well as the observation that ASD patients frequently present with gastrointestinal complications [15].

Despite these discoveries, the tissues, brain regions, causal genes, and biological pathways relevant to ASD are not fully understood, and there are currently no medications available to effectively treat ASD [16]. Thus, it is important to examine available omics data to elucidate the most relevant disease-associated mechanisms in a tissue-specific fashion to understand which tissues, pathways, and networks are affected by common versus rare variants. We hypothesize that ASD follows an omnigenic model of pathogenesis, which suggests that hub genes with large effect sizes interact with peripheral genes with smaller effect sizes through highly interconnected networks [17]. It is plausible that rare variants are enriched among the hub genes and essential tissues whereas common variants are enriched among the peripheral genes and a broader range of tissues. Therefore, elucidating how common and rare variants of ASD converge and diverge in tissue-specific gene networks will identify key tissues and gene drivers within gene regulatory networks, which will provide further insights for future mechanistic and therapeutic studies.

Methods

Analysis overview

We utilize a multiomics integration tool, *Mergeomics*, for our analysis of ASD [18, 19]. Briefly, we integrated full summary statistics of an ASD genome-wide association study (GWAS) with tissue-specific expression and splicing quantitative traits (eQTLs/sQTLs) and tissue-specific gene coexpression networks to allow for the ranked identification of pathways and gene subnetworks most associated with ASD based on common variants examined in GWAS. The pipeline then performs a key driver analysis to determine network hub genes, termed “key drivers”, whose neighboring networks are enriched for disease-associated genes within

interconnected gene regulatory networks. The outputs of *Mergeomics* include a ranking of biological pathways and subnetworks informed by ASD GWAS common variants as well as a visualization of key drivers within disease subnetworks. The robustness of *Mergeomics* has been substantiated by experimental validations of its computational predictions, and it has been successfully applied to the analysis of other complex diseases [20-23]. A key advantage of *Mergeomics* is that it utilizes the full disease association strength spectrum and contains a unique test statistic that summarizes disease association enrichment at multiple quantile thresholds to derive stable statistics that are less dependent on any given GWAS significance cutoff and account for discrepancies in sample size and power. **Figure 1** depicts our overall pipeline, datasets utilized, and the three steps of our *Mergeomics* analysis that will be discussed in further detail: marker dependent filtering (MDF), marker set enrichment analysis (MSEA), and key driver analysis (KDA).

Multiomics datasets and gene networks

ASD common variant GWAS Summary Statistics

For our analysis, we utilized the most recent ASD GWAS to retrieve the full summary statistics of ASD association p-values for all analyzed single-nucleotide polymorphisms (SNPs) [24] This study included 18,381 individuals with ASD and 27,969 controls from a primarily Central European (CEU) population. Ricopili [25], a computational pipeline developed by the Psychiatric Genomics Consortium, was used for quality control and principal component analysis. Following this, PLINK [26] was utilized for the primary association analysis and METAL [27] for a meta-analysis. The GWAS summary statistics contained ASD association p-values for more than 9 million SNPs.

ASD Rare Variants

ASD rare variants were compiled from the Simons Foundation Autism Research Initiative (SFARI). The SFARI database is a research consortium database that compiles high confidence rare gene variants of ASD [28]. Genes are stratified based on four levels: Level Syndromic (high confidence in both ASD and a specific syndrome beyond the characteristics of ASD), Level 1 (high confidence in their implication in ASD), Level 2 (strong candidate for ASD association), and Level 3 (moderate evidence based on previous research).

Tissue-specific eQTLs and sQTLs

Tissue-specific eQTL and sQTL from 49 tissues were retrieved from the Genotype Tissue Expression (GTEx) project database [29] for mapping SNPs from the ASD GWAS to genes and further removing SNPs in high linkage disequilibrium (LD) of r2>0.5 based the CEU LD information during MDF. **Table 1** shows the complete list of tissue-specific eQTL/sQTL data used as input for the *Mergeomics* analysis.

Weighted Gene Coexpression Network Analysis (WGCNA) to Define Data-driven Functional Gene Sets

To group genes with functional relevance in individual tissues in a data-driven manner, we used the transcriptome data from the GTEx database to construct tissue-specific WGCNA gene coexpression modules [30]. Typically, these modules contain genes that are coexpressed and functionally related. This provided a means of placing our ASD-associated GWAS genes from MDF into categories that have biological relevance in individual tissue contexts. These

modules are functionally annotated through pathway enrichment analysis using KEGG [31], Reactome [32], and BioCarta [33] databases.

Bayesian gene regulatory networks

To elucidate directional gene regulatory relations, Bayesian networks were constructed from tissue-specific GTEx databases using the RIMANET package [34]. As Bayesian networks from individual datasets are typically sparse, networks from similar tissues were subsequently merged to derive composite networks for brain, digestive, cardiovascular, endocrine, immune, adipose, and reproductive tissues to reduce sparsity and ensure each network contains at least 10k genes. We further merged the immune and adipose tissue networks given the known interactions between adipose and immune cells [35]. The merged networks and their corresponding tissues and sizes are shown in **Table 2**.

Marker Dependent Filtering (MDF)

We first performed SNP-to-gene mapping using SNPs from the ASD GWAS together with tissue-specific eQTLs and sQTLs from the GTEx project database. We also used distanced-based mapping, which maps SNPs to genes at a maximum of ±20 kilobases, as an alternative mapping method. We corrected for linkage disequilibrium (LD) to filter known dependencies between SNPs based on a LD cutoff threshold of r2 > 0.5 from the CEU population, as the ASD GWAS population is mainly CEU. We used -*log10* transformed p-values from the ASD GWAS to represent SNP association strengths for ASD. The output of MDF contained tissue-specific mapping of SNPs to genes based on eQTL, sQTL, and distance-based mapping as well as their ASD association strengths in the form of -log10 p-values. No GWAS cutoffs were applied at this

stage to capture the full spectrum of disease association signals from strong, moderate, to subtle or no association.

Marker Set Enrichment Analysis (MSEA)

In MSEA, tissue-specific ASD GWAS SNP-enriched gene sets were identified. The GWAS-mapped genes from MDF and tissue-specific gene sets derived from WGCNA modules were used as input for MSEA. To determine coexpression modules that are enriched in ASD GWAS, a chi-like statistic was used as the enrichment test statistic in MSEA:

$$\chi = \sum_{i=1}^{n} \frac{O_i - E_i}{\sqrt{E_i} + \kappa}$$

Briefly, in this formula a chi-statistic value is calculated, where *"n"* indicates the number of quantile points in the dataset, which are thresholds used to divide the GWAS SNPs into significant vs non-significant groups. Quantiles, which are rank-based, were used instead of specific p-value cutoffs to enable the normalization of different GWAS datasets that have different sample sizes and statistical power that influence the specific p-value ranges. For this study, we used 10 quantiles ranging from 0.5 to an upper limit that is adjusted based on the median module length to ensure that the distribution of these lengths are appropriately taken into consideration for each module. The expression inside the summation includes *O* and *E*, which are the number of observed and expected positive association signals above each quantile point, respectively. The difference between these two values is divided by the sum of the square root of the expected count of positive signals and a stability parameter, κ, which was set to 1 to account for datasets with extremely low counts.

The calculated χ value is a sum of the output of the expression at each quantile point from a given test gene set. To create a null distribution, we generated random gene sets matching the gene number of the test gene set and calculate the χ values from the random gene sets. The following null hypothesis is then tested: *Given the set of all distinct markers from a set of N genes, these markers contain an equal proportion of positive association study findings when compared to all the distinct markers from a set of N random genes [18]*. The distribution is estimated by randomly shuffling the genes mapped from disease-associated markers and approximating the parameters to best represent the data (i.e. a parametric model is fit for the distribution). From the null distribution, we are able to calculate a corresponding Z score, which is a measurement of the number of standard deviations an observed value is from the mean of a distribution. The Z score that is determined from the distance between our actual χ value and the mean of the null distribution provides us with an enrichment score for each gene set (i.e., a tissue-specific coexpression module in the current study), which is then used to rank tissue-specific modules for their enrichment in ASD.

Key Driver Analysis

Key gene drivers of ASD and their associated neighbors within gene regulatory networks were identified in KDA. Significant ASD-GWAS enriched coexpression modules generated from MSEA and tissue-specific probabilistic Bayesian gene regulatory networks are used as input for KDA. Next, key drivers are predicted by first identifying hub nodes that are in the top 25% in terms of the number of edge connections. For each hub node and its subnetwork, KDA utilizes a method that is similar to MSEA to assess enrichment in ASD-associated gene sets identified from MSEA. The proportion of nodes in the subnetwork that are among ASD-

associated modules from MSEA is determined, and an enrichment statistic is calculated by creating a null distribution of reshuffled subnetworks for the key driver. This is done to observe the likelihood of obtaining the same proportion of disease-associated genes for a random network of the same size. As output for KDA, tissue-specific key drivers and their first neighbors within gene regulatory networks are identified. These networks are ranked by their enrichment for ASD GWAS-informed gene sets from MSEA and are visualized in Cytoscape [36]. Using EnrichR, a gene set enrichment annotation tool [37-39], we performed pathway enrichment analysis in order to understand the functions of predicted key driver subnetworks. Furthermore, genes in the subnetworks that contain known rare variants were annotated based on information from the SFARI database. We considered the rare variants both as a collective across all four ASD levels and at each individual stratification for the analysis.

Results

Marker Set Enrichment Analysis reveals tissue-specific coexpression modules enriched for ASD GWAS signals

The MSEA analysis identified 47 tissues (10 brain regions, 37 peripheral tissues) from which at least one coexpression module was significantly enriched for ASD associations in GWAS below an FDR cutoff of 5% (**Figure 2**). Across these tissue, there were 196 ASD-enriched WGCNA modules after MSEA. **Table 3** shows the top 10 modules based on the statistical significance of ASD GWAS enrichment, where coexpression modules from 5 brain regions and 5 peripheral tissues with diverse annotations were observed. The anterior cingulate cortex, which is involved in emotional regulation and cognitive control, and the amygdala, another region crucial for emotional response, contained coexpression modules with the highest

ASD enrichment. The top annotations for these two modules highlight complex neuronal and cross-system processes that suggest how impairments caused by ASD pathogenesis contribute to dysregulation. The other top brain tissue coexpression modules (from the frontal cortex (Brodmann Area 9), cerebellum, and cortex), consist of vital neuronal system activity annotations. Together these findings support how impaired brain functioning plays a significant role in ASD pathogenesis. Interestingly, peripheral tissues from the digestive system, reproductive system, and immune system contained highly significant coexpression modules relevant to mRNA splicing, immune pathways, cell cycle, and mTOR signaling.

When comparing the statistical significance between brain and peripheral modules, we did not find significant difference in the average false discovery rate between brain and peripheral modules (**Figure 3**). Across all significant modules, pathway annotation revealed a broad range of consistent pathways including cell cycle, gene regulation (particularly splicing), neuronal signaling, oxidative phosphorylation, immune system, and mammalian target of rapamycin (mTOR) signaling (**Figure 4**).

To further explore the pathways within the enriched coexpression modules of tissues with the most relevance to ASD, we categorized tissues into seven regions: Adipose/Immune, Brain, Cardiovascular, Digestive, Endocrine, Female Reproductive, and Male Reproductive. There were nearly 800 unique pathways across all regions, and we found that 80 were shared among all seven regions (**Figure 5**). We also observed a vast array of shared and unique pathways, which provided us a unique perspective on region-specific pathway annotations within ASD-enriched coexpression modules while also highlighting terms that are abundant in nearly all tissues in the analysis (e.g., immune pathways and cell cycle regulation). For the pathways shared across all regions, and the regions that contained pathways shared amongst many tissues (e.g., Brain,

Adipose/Immune), we highlighted annotations that were either abundant and highly enriched in multiple tissues, or present in one tissue yet remained significant. Of note, brain region-specific pathways displayed associations with crucial neuronal processes (NMDA receptor activation) and neurodegenerative conditions (Alzheimer's disease, Amyotrophic Lateral Sclerosis). Several regions also displayed associations with immune system regulation, cellular signaling, growth and proliferation, and protein interactions. These associations suggest the impacts of dysregulation and abnormalities of these pathways on diseases such as ASD, which are rooted in neurodevelopmental processes and cellular function.

Key Driver Analysis identifies distinctions between brain and peripheral tissue associations in ASD

Using the significant tissue-specific coexpression modules identified from MSEA, we ran KDA to identify tissue-specific key drivers. We also assessed whether the key drivers and their subnetworks capture both common and rare variants of ASD. These key drivers were further intersected with genes containing common and rare variants. As seen in **Figure 6A**, we observed higher numbers of rare and common variants in the key drivers from our brain Bayesian networks compared to all of the peripheral networks, supporting the importance of the brain in ASD, as expected. However, when normalizing the gene counts of rare and common variants against the total number of genes in the peripheral and brain Bayesian networks, the normalized count was higher for periphery tissue networks (**Figure 6B**), due to their much smaller network size compared to more complex brain network (**Table 2**). This finding supports that periphery tissue genes possess sizable contributions to the overall genetic burden of the disease.

We next sought to assess the relative distribution of common and rare variants among key drivers from the brain and peripheral tissue networks. **Figure 7** shows the percentage of key drivers across both brain and peripheral tissues that are also either known rare variants from the SFARI database, or common variants from the GWAS. We observed higher numbers of known rare ASD variants among brain tissue-derived key drivers compared to peripheral tissue-derived key drivers. As the rare variants mostly affect brain development and neuronal functions and have larger effect sizes, our results aligns with the central role of brain tissues in the pathogenesis of ASD. By contrast, there is a similar number of common variants among key drivers between brain and peripheral tissues, suggesting that common variants of ASD are less discriminative between the brain and the peripheral tissues.

To further test whether an observed overlap was statistically significant, we utilized hypergeometric testing to assess if the overlap between the key drivers and common or rare variants could be observed by random chance. We first examined all key drivers of a given tissue, and then explored key drivers within the context of their associated subnetwork. As shown in **Figure 8,** brain tissue key drivers showed much stronger enrichment for both the rare and common variants than peripheral key drivers, further supporting the importance of the brain in ASD. We also observed that there is a significantly higher enrichment for rare variants than for common variants among the brain key drivers. As rare variants have larger effect sizes than common variants, the observation of stronger enrichment of rare variants among brain key drivers also supports the stronger influence of brain networks in ASD.

Prioritization of key drivers based on rare/common variant enrichment in key driver subnetworks

To prioritize key drivers, we assessed the significance of their subnetworks for common and rare variant enrichment. **Figure 9** and **Figure 10** show the top 15 subnetworks in terms of their rare variant overlap enrichment and common variant overlap enrichment, respectively. There is a higher abundance of brain tissue key driver subnetworks (12 out of 15) that were enriched for rare variants, and an increasing representation of digestive tissue key driver subnetworks (5 out of 15) that had enrichment for common variants. We also found that there is a higher proportion of brain tissue key driver subnetworks with are enriched for both rare and common variants (**Figure 11** and **Table 4)**, supporting convergence between rare and common variants in these key driver subnetworks.

The predicted key driver subnetworks of Synaptotagmin 1 and Adducin 2 support crucial roles as regulators of neuronal processes and neurodevelopment.

Two brain tissue key driver subnetworks stood out due to their high overlap enrichment of either common or rare variants. Synaptotagmin 1 (SYT1), shown in **Figure 12**, held the highest rare variant overlap enrichment in its subnetwork across all key drivers. Its prediction in our analysis as a key driver underscores its significance as a known rare variant and also implies this gene's potential impact on other rare (e.g., RELN, NCKAP1) and common variants (e.g., CACNB3, SV2B) within its subnetwork (**Figure 12A**). This gene is a known syndromic ASD rare variant that causes severe neurodevelopmental abnormalities, and it serves as a membrane protein of synaptic vesicles involved in neurotransmitter release during calcium binding and thus

plays a vital role in neuronal processes [40, 41]. The top annotation terms from the SYT1 subnetwork agree with these crucial neuronal processes (**Figure 12B**).

The other key driver subnetwork of interest, Adducin 2 (ADD2), also showed a high enrichment for both rare (e.g., SHANK2, SCN8A) and common variants (e.g., ACTN2, DMN1) (**Figure 13A**). Adducin genes encode cytoskeleton proteins that are critical for osmotic rigidity and cell shape by regulating the formation of the spectrin-actin membrane skeleton [42]. ADD2 is expressed in the brain, and its knock-out results in the loss of activity-dependent connection formation between neurons when knocked out [43]. This suggests ADD2's relevance to learning and development and its contribution to ASD as a key component of the disease pathogenesis. Similar to SYT1, this is also reflected in the top annotations of ADD2's subnetwork, which show important nervous system processes (**Figure 13B**).

Discussion

ASD is a neurodevelopmental disorder that possesses complex heterogeneity in its genetic architecture [16]. Previous research has explored the role of both common and rare variants in the pathogenesis of ASD, and have implicated numerous genes, tissues, and biological pathways [5, 44, 45]. Therefore, a leading challenge of the disease is the complexity in both how it manifests in an individual and the underlying genetic mechanisms that vary significantly. Numerous studies have identified relevant aspects of the central nervous system as well as various peripheral regions with ASD association, and have highlighted the abundant interactivity between organs, tissues, and genes [8-15, 46].Thus, given the nature of ASD and the lack of current pharmacological treatments, continued research aims to fully deconstruct the genetic intricacies of the disease.

In our study, we utilize summary statistics from an ASD GWAS [24], publicly available QTL data from the GTEx database [29], and gene regulatory network information to investigate ASD with use of a computational tool, *Mergeomics* [18, 19]. Taking into account both common and rare variants, we sought to understand the molecular interactions within and between brain and peripheral tissues to uncover tissues, genes, biological pathways, and gene networks that are enriched for common and/or rare variants of ASD. Furthermore, we aimed to predict key drivers within gene regulatory networks that may be of interest for further research and experimental validation.

From our marker set enrichment analysis (MSEA), we found that there are a diverse range of ASD-enriched tissues and gene sets. While brain tissues possessed the highest disease association enrichment, seen in the anterior cingulate cortex and the amygdala, there were also peripheral tissues that showed relevance to ASD genetic signals (**Table 3**). The coexpression module enrichments of the anterior cingulate cortex and amygdala align with previous results that observed structural abnormalities of this cortex in both ASD mouse models and adult ASD individuals [47, 48], and that have identified various dysregulations and alterations in the amygdala of those with ASD [49]. With regards to the enriched modules of the other brain regions from **Table 3** (frontal cortex [Brodmann Area 9], cerebellum, and cortex), previous neuroanatomical analyses have included these tissues as regions of the brain that display abnormal development [50-52]. Thus, our findings provide additional evidence suggesting the importance of cognitive functioning, motor coordination, and complex neuronal system activity in the context of ASD.

In addition to these confirmatory outcomes of various brain tissues, we found that tissues in the digestive, reproductive, endocrine, and immune systems also displayed ASD enrichment

within particular coexpression modules. The digestive system has recently become heavily implicated in ASD pathogenesis as research has continued to explore the interactions within the microbiota-gut-brain axis, and how there are distinctive gastrointestinal complications present in those with ASD [15, 53]. There has also been extensive research on how maternal factors and its influence on fetal development contribute to ASD risk [54-59]. This influence can occur at a hormonal level or through immune pathways, which demonstrate not only the cross-system interactions at play within the pathogenesis of ASD but also highlight its complexity across both brain and peripheral regions of the body.

Our exploration of pathway annotations from these disease-enriched coexpression modules further emphasize how the complexity of ASD spans throughout the body. We found that immune, cellular signaling, and cell growth and regulation pathways were noticeably abundant in various regions (**Figure 4** and **Figure 5**). In **Figure 5**, we found that Alzheimer's disease and Amyotrophic Lateral Sclerosis (ALS) were amongst brain region pathways that were consistently enriched, which interestingly aligns with how this has been area of research given that these two disorders are rooted in neurological dysfunction, similar to ASD. As there exists a spectrum of neurological disorders ranging from neurodevelopment to neurodegeneration, with ASD, ALS, and Alzheimer's falling within these confines, research has explored potential associations within this spectrum [60-64].

We also noticed how involved the endocrine and immune systems were within various regions. The male and female reproductive, digestive, and adipose systems contained various pathways involved in immune regulation (e.g., hyaluronan uptake and degradation, Interleukin-6, CD40 pathways) and endocrine function (e.g., cysteine and methionine metabolism, fatty acid synthesis), or cellular signaling, processes, and regulation (e.g., MAPK pathways, EDG1

pathways, ENOS activation). As research continues to explore the role of immune dysregulation and metabolic dysfunction in ASD pathogenesis [65-69], we recognize that our observed relationships further emphasize the interconnected nature between pathways and across systems. Dysregulation in one region is likely to impact many others, which thus supports the idea that these relationships collectively contribute to the diversity and complexity of ASD pathogenesis. As a whole, our MSEA findings recapitulate previous studies that highlight the impact of altered neural functioning in various brain regions, and we also emphasize the rising understanding of the roles that digestive, reproductive, endocrine, and immune system have in ASD.

In our key driver analysis, we discovered that the brain outperforms peripheral tissues in terms of abundance of significant rare and common variants. Furthermore, we also identified key drivers and their subnetworks that possessed significant enrichment in either rare or common variants. A greater number of brain tissue key driver subnetworks contained a significant number of both rare and common variants, which suggest that these key drivers are crucial regulators within ASD pathogenesis given the convergence of these two types of variants. The rankings of key driver subnetworks in terms of their rare or common variant overlap enrichment highlight the central role of the brain in ASD, yet also point towards the digestive system as the leading peripheral region for common variant abundance. We visualized the subnetwork of two genes, SYT1 and ADD2, which we identified as key drivers given the number of edge connections they contained.

Overall, we found that our approach to studying ASD aligns with previous research findings while also revealing new avenues for future exploration. Each of our top 10 coexpression modules (**Table 2**) held a degree of relevance in ASD pathogenesis based on prior knowledge, and the most replicated pathways from MSEA (**Figure 4** and **Figure 5**)

demonstrated how the various cellular processes and immune pathways were abundant across brain and peripheral tissues. While the insignificant difference in the collective adjusted p-values between brain and peripheral coexpression modules (**Figure 3**) may also be a factor of significance dilution due to the increased complexity of brain tissue genes, it may reflect the growing body of research that suggests ASD relevance within other regions of the body, namely the digestive, endocrine, and immune systems [14, 15, 53, 56, 59, 70].

From our key driver analysis, we highlight regulatory relationships of two key gene drivers, SYT1 (**Figure 11**) and ADD2 (**Figure 12**). While SYT1 has previously been found to be a known rare variant of ASD [28], and ADD2 a common variant [24], our network analysis highlights unexplored tissue-specific network interactions with both types of variants. These networks provide the opportunity for future experimentation and validation. In addition, the key driver that indicated convergence of both variant types in their subnetworks (e.g., SCN8A, AMPH, ATP9A) also provided a new perspective for genes that possess particularly unique regulatory interactions in ASD pathogenesis.

Overall, using common variant inputs, we discover significant ASD relevance in both brain and peripheral regions as well as notable convergence of known rare and common variants in the brain. Our predicted subnetworks identify a number of key drivers, some of which were previously implicated, and possess diverse potential interactions and regulatory mechanisms on neighboring genes.

Limitations of the study

Despite our findings, there are several limitations that are present as well as ample opportunities for improvement. First, the data used as input is slightly dated. The GWAS was

published in 2019 [24], and the most recent version of the GTEx data used for the eQTL and sQTL data was collected in 2019 as well [29]. Furthermore, the demographics of the GTEx database tissue donors is heavily skewed against minority populations, with 84.6% of donors identifying as White. ASD research is constantly updating and new discoveries continue to be made, so a primary area of improvement is to utilize more recent and culturally diverse data as it becomes available. Resources that are actively updated such as the Simons Foundation for Autism Research Initiative [28], the Australia Autism Biobank [71], and the MSSNG database [72] will undoubtedly provide useful data for our computational tool. Furthermore, there are also studies that have performed transcriptome, metabolome, and epigenome level analyses of ASD and which are likely to expand our understandings as well [73-75].

We also acknowledge that we have only performed a tissue-level analysis with the use of sex-combined data. Cell-level analysis is a fast-growing area of research across numerous fields, and ASD is a disease that demands this level of cellular specificity. Furthermore, there is a gender bias in the disease prevalence [70], which emphasizes that a sex-specific analysis is likely to reveal key insights.

Utilizing updated datasets and carrying out cell-level sex-specific analyses are key areas of future work. Doing such in conjunction with our initial results will undoubtedly shed further light on ASD's genetic complexities. Overall, we hope that our findings will assist in the generation of new hypotheses and experimental validations to help uncover new mechanisms for neurodevelopmental disorders such as ASD.

Tables

Table 1. Summary of eQTLs and sQTLs from the Genotype Tissue Expression Portal (GTEx) database for Mergeomics analysis.
eQTLs and sQTLs from 14 brain regions and 35 peripheral tissues were retrieved from the GTEx database to be used for tissue-specific mapping of SNPs from the ASD GWAS to genes that are potentially regulated by the SNPs in each tissue.

Region	Tissue	eQTL Count	sQTL Count
Peripheral	Adipose: Subcutaneous	1501725	649559
Peripheral	Adipose: Visceral Omentum	1116992	501342
Peripheral	Adrenal Gland	606549	266580
Brain	Amygdala	241511	84934
Brain	Anterior Cingulate Cortex (Brodmann Area 24)	360811	121320
Peripheral	Artery: Aorta	1137522	446515
Peripheral	Artery: Coronary	468104	231587
Peripheral	Artery: Tibial	1528215	615729
Brain	Basal Ganglia: Caudate	592299	183573
Brain	Basal Ganglia: Nucleus Accumbens	580998	199042
Brain	Basal Ganglia: Pumen	477659	137298
Peripheral	Breast Mammary Tissue	915137	478320
Peripheral	Cells: Cultured Fibroblasts	1558412	604866
Peripheral	Cells: EBV-transformed Lymphocytes	303559	215630
Brain	Cerebellar Hemisphere	689639	234153
Brain	Cerebellum	838323	276313
Peripheral	Colon: Sigmoid	861273	367653
Peripheral	Colon: Transverse	946365	392568
Brain	Cortex	675220	208046
Peripheral	Esophagus: Gastroesophageal Junction	891705	375284
Peripheral	Esophagus: Mucosa	1374993	480793
Peripheral	Esophagus: Muscularis	1344979	499704
Brain	Frontal Cortex (Brodmann Area 9)	511541	168139
Peripheral	Heart: Atrial Appendage	962118	360824
Peripheral	Heart: Left Ventricle	850330	286330
Brain	Hippocampus	374601	128103
Brain	Hypothalamus	385593	147721
Peripheral	Kidney Cortex	76350	35421
Peripheral	Liver	407126	158972
Peripheral	Lung	1261501	567609
Peripheral	Minor Salivary Gland	294404	143965
Peripheral	Muscle: Skeletal	1405527	574674
Peripheral	Nerve: Tibial	1720751	659871
Peripheral	Ovary	370124	197505
Peripheral	Pancreas	818024	251970
Brain	Pituitary	713991	326113
Peripheral	Prostate	514066	258156
Peripheral	Skin: Non-Sun-Exposed Suprapubic	1409730	564897
Peripheral	Skin: Sun-Exposed Lower leg	1632271	638497
Peripheral	Small Intestine Terminal Ileum	414021	195522
Brain	Spinal Cord: Cervical C1	293087	111631
Peripheral	Spleen	770698	279622
Peripheral	Stomach	706670	295700
Brain	Substantia Nigra	198962	75770
Peripheral	Testis	1554886	954055
Peripheral	Thyroid	1765762	682833
Peripheral	Uterus	216551	138022
Peripheral	Vagina	224692	139644
Peripheral	Whole Blood	1276546	377472

Table 2: Summary of tissue-specific Bayesian networks constructed using GTEx transcriptome data.
Bayesian networks were constructed for individual tissues and further merged into seven Bayesian networks representing major biological systems to reduce sparsity.

Network Name	Individual Tissues	Node Number	Edge Number
Adipose	Adipose (Subcutaneous), Adipose (Visceral Omentum)	2262	17758
Brain	Amygdala, Anterior Cingulate Cortex, Caudate, Cerebellar Hemisphere, Cortex, Frontal Cortex, Hippocampus, Hypothalamus, Nucleus Accumbens, Putamen, Pituitary, Spinal Cord, Substantia Nigra	37866	223949
Cardiovascular	Artery (Aorta), Artery (Coronary), Artery (Tibial), Heart (Atrial Appendage), Heart (Left Ventricle)	11241	41168
Digestive	Colon (Sigmoid), Colon (Transverse), Esophagus (Gastroesophageal Junction), Esophagus (Mucosa), Esophagus (Muscularis), Liver, Minor Salivary Gland, Pancreas, Small Intestine (Terminal Ileum), Stomach	14259	68401
Endocrine	Adrenal, Ovary, Pituitary, Testis, Thyroid	14720	37690
Female Reproductive	Breast Mammary Tissue, Endocervix, Ectocervix, Fallopian Tube, Ovary, Uterus, Vagina	10020	25117
Immune	EBV Lymphocytes, Spleen, Whole Blood	10430	21815
Immune-Adipose	Adipose (Subcutaneous), Adipose (Visceral Omentum), EBV Lymphocytes, Spleen, Whole Blood	12692	39573

Table 3: Top 10 coexpression modules enriched for ASD GWAS signals from Marker Set Enrichment Analysis.
The top ten gene sets from MSEA were ranked based on false discovery rate (FDR) of ASD GWAS enrichment. Tissue corresponds to the sample name from GTEx data, with the module name indicating which coexpression module within each tissue showed ASD GWAS enrichment. The top annotations indicate the pathway terms that had the highest enrichment p-value based on pathway annotation of the genes in each tissue-specific coexpression module.

	Tissue	Module	FDR	Top Annotations
1	Anterior Cingulate Cortex	blue	1.27E-24	Protein Degredation, Long Term Potentiation, Immune System
2	Amygdala	black	3.07E-24	Spliceosome Activity, Proteasome Activity, Electron Transport
3	Liver	turquoise	5.59E-23	Cell Cycle Regulation, DNA Repair, mTORC1 Activity
4	EBV Lymphocytes	blue	2.01E-19	TNF-α Signaling, Pancreatic Cancer, Neurotrophin Signalling
5	Minor Salivary Gland	lightgreen	6.69E-19	RNA Splicing, TGF-β signaling, Membrane Trafficking
6	Frontal Cortex	gray60	7.77E-19	Immune Response and Regulation, Neuronal, Synaptic Processes
7	Uterus	turquoise	3.33E-16	Transcription, mRNA processing
8	Cerebellum	purple	1.04E-15	Energy Metabolism, Cellular Stress Response, Oxygen Transport
9	Cortex	blue	2.49E-15	Olfactory Signal Transduction
10	Testis	magenta	2.52E-13	Cell Cycle Regulation, DNA Repair

Table 4: Table displaying the convergence of brain and peripheral tissues in terms of their rare and common variant overlap enrichment scores among network key drivers.
Specific overlap enrichment scores for sixteen key driver subnetworks that possessed an overlap enrichment p-value of less than 0.05 for both rare and common variants.

KD	RV Score	CV Score	Tissue	Tissue Region
ADD2	2.07E-09	8.12E-07	Substantia Nigra, Pituitary, Frontal Cortex	Brain
AMPH	0.000143	0.004631	Substantia Nigra	Brain
ATP9A	8.83E-06	0.000623	Cerebellum	Brain
CALD1	0.00222	0.024888	Colon Sigmoid	Digestive
CD74	0.033097	0.000532	Colon Sigmoid	Digestive
DYNC1H1	0.01168	0.007854	Cerebellum	Brain
DYNC1I1	0.030137	0.011718	Cerebellum	Brain
MAP2K1	0.000191	0.031176	Caudate	Brain
PCDH7	1.61E-05	0.036614	Colon Sigmoid	Digestive
PLPP3	0.0006	0.036894	Putamen, Frontal Cortex	Brain
RBFOX2	2.42E-05	0.013884	Substantia Nigra	Brain
SCN8A	0.000599	0.00065	Substantia Nigra	Brain
SNAP91	9.16E-08	0.016832	Substantia Nigra	Brain
SYP	0.000796	0.021754	Substantia Nigra, Pituitary	Brain
TMEM130	3.45E-08	0.001244	Pituitary	Brain
XKR4	0.021331	0.006791	Colon Sigmoid	Digestive

Figures

Figure 1: Analysis workflow.
The first step of the overall analysis pipeline for *Mergeomics* is marker dependent filtering (MDF) to correct of linkage disequilibrium (LD) in GWAS SNPs. Summary statistics from the genome-wide association study are taken with tissue-specific eQTL and sQTL data to map genetic markers to corresponding genes. ASD-enriched modules are identified in marker set enrichment analysis (MSEA) by organizing the marker-mapped genes from MDF into coexpression gene sets and assessing their enrichment against a null distribution. By assessing their disease enrichment in a gene regulatory network, important regulatory genes are identified in key driver analysis (KDA). Figure created with BioRender.com.

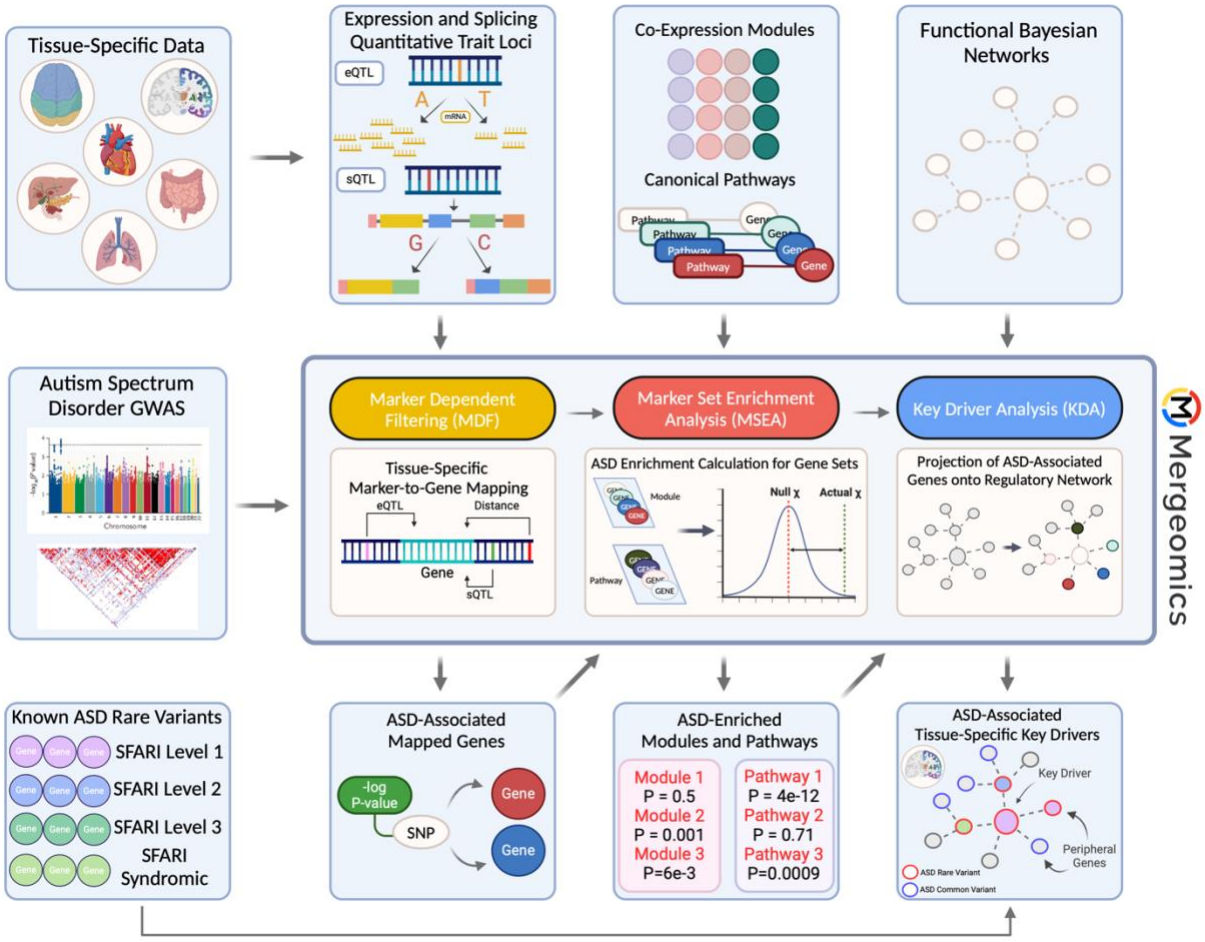


Figure 2: Numbers of ASD GWAS enriched modules by tissue from Marker Set Enrichment Analysis.
Tissues were ranked by the number of WGCNA modules that were below an FDR cutoff of 5% for ASD GWAS association enrichment.

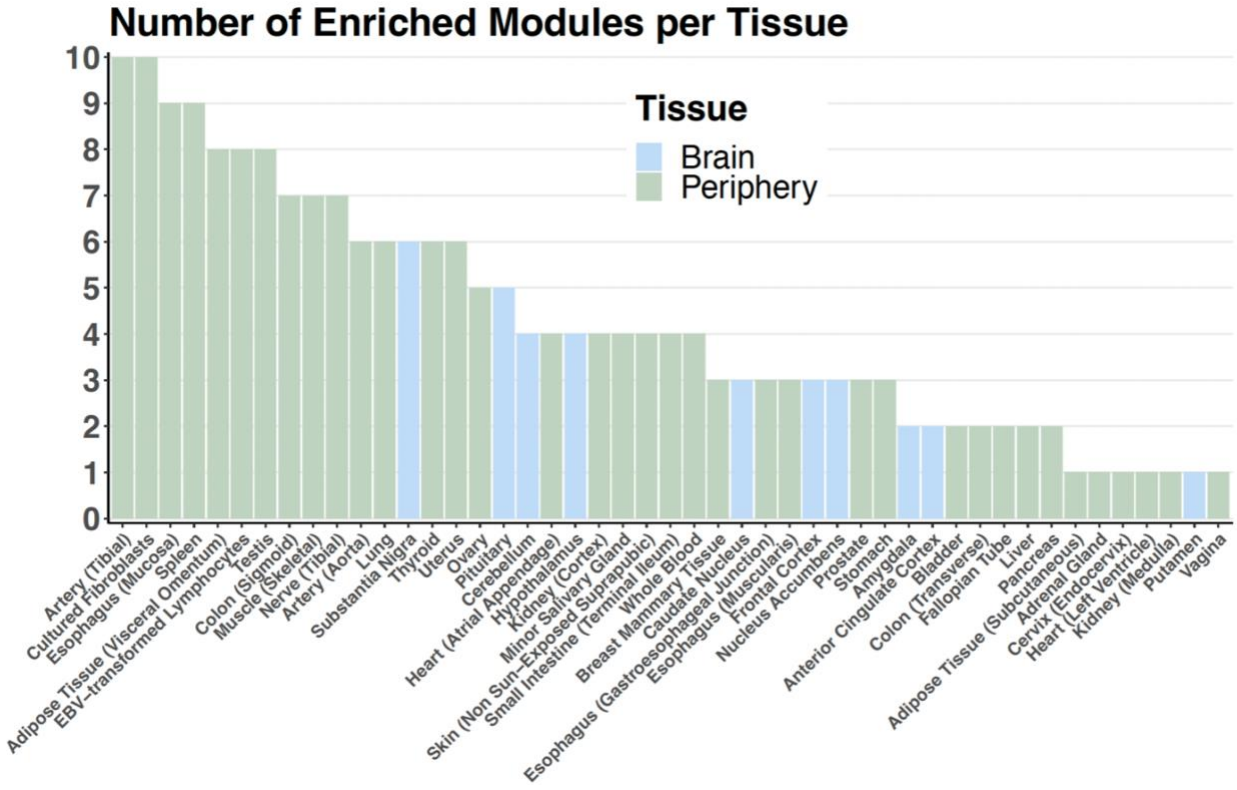


Figure 3: No significant difference was observed in the ASD GWAS enrichment significance between brain and periphery coexpression modules in MSEA.
The difference between the -*log10* false discovery rate of brain and peripheral modules was calculated utilizing a two-sided Wilcoxon test given the non-parametric distribution of the data based on Shapiro-Wilk test (W = 0.64228, p-value < 2.2e-16).

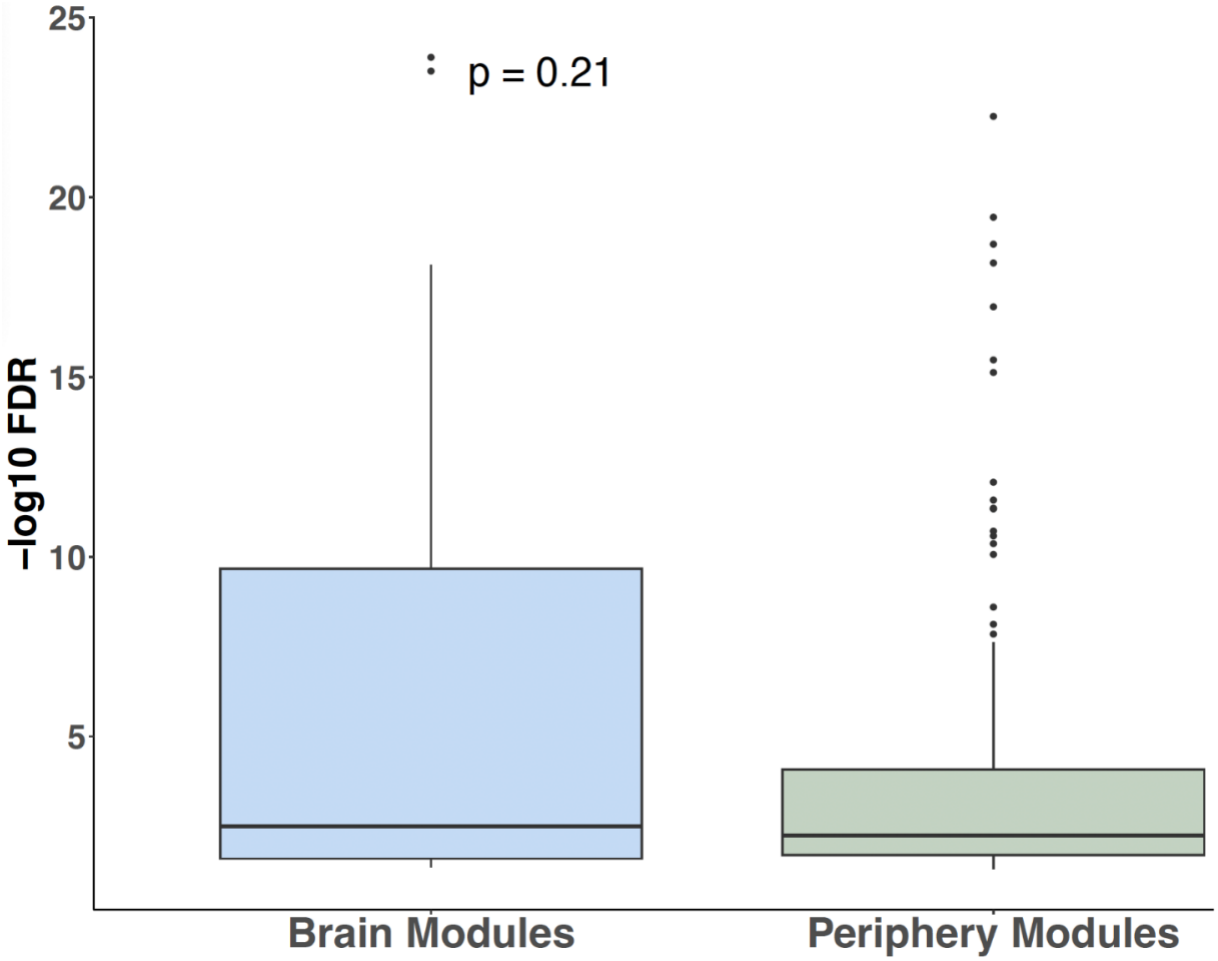


Figure 4: Most consistent functional terms across significant coexpression modules from Marker Set Enrichment Analysis show cellular processes, mRNA splicing, immune system, and mTOR signaling pathways.
All pathway annotation terms for genes within ASD-associated modules were collected and ranked based on the number of modules enriched for each pathway.

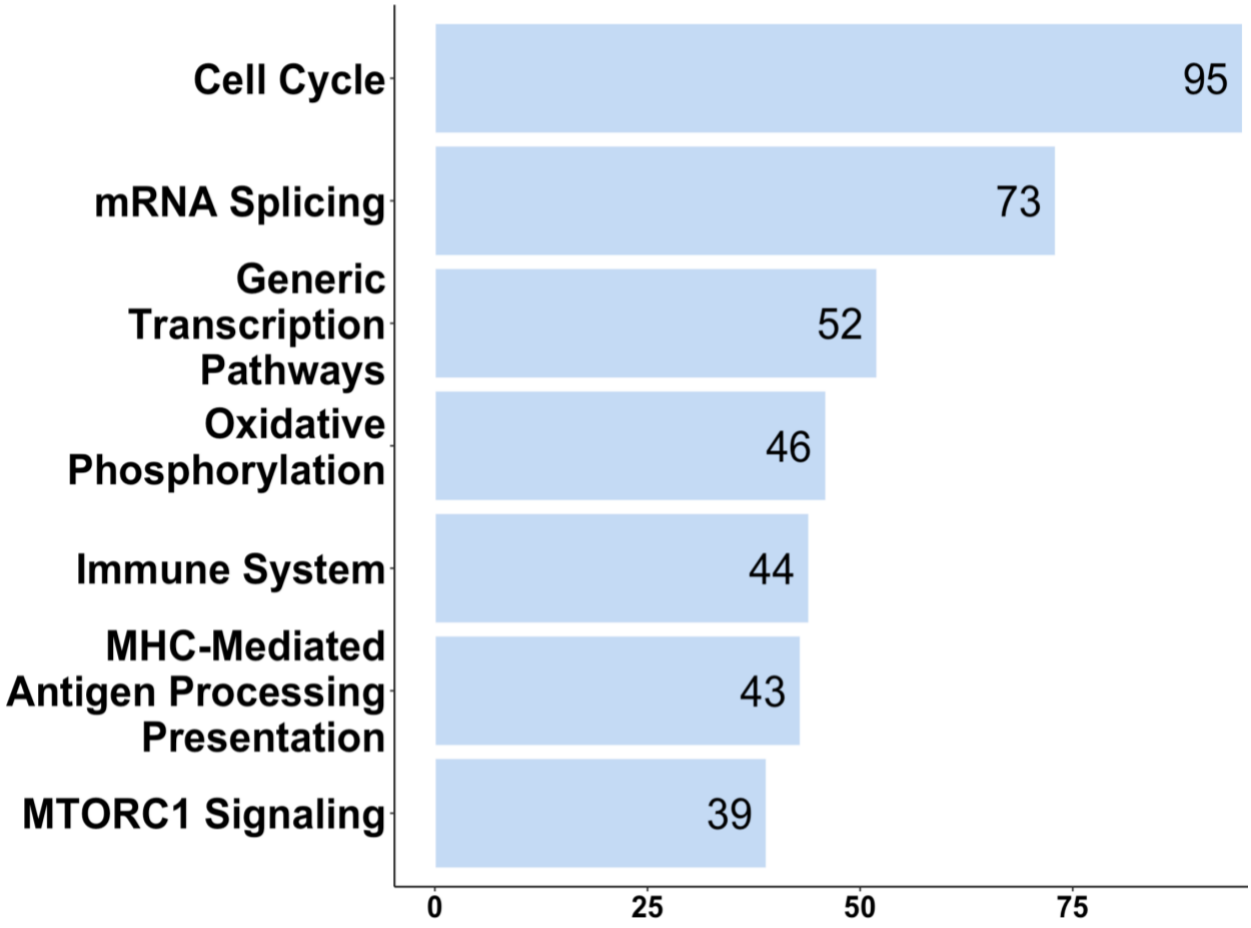


Figure 5: Shared and unique terms across MSEA coexpression module pathway annotations between seven groups.

The seven groups were Adipose/Immune, Brain, Cardiovascular, Digestive, Endocrine, Female Reproductive, and Male Reproductive. The pathways were ranked based on the median values of the -*log10* transformed adjusted p-values if the pathway was derived from multiple coexpression modules, and the top three terms (where applicable) were subsequently depicted on the figure. Adipose/Immune, Brain, and the group of pathways shared across all tissues were extensive enough that we categorized pathways based on whether they were shared across >3 tissues in the groups (and >30 tissues in the case of the shared group), or if they were still moderately enriched but only present in one tissue within this region. Adipose/Immune: Phospholipid Metabolism (Whole Blood), Metabolism of Porphyrins (Spleen), ENOS Activation and Regulation (Cultured Fibroblasts). Brain: Olfactory Signaling and Transduction (Cortex), ALS Pathway (Putamen), Cytokines and Inflammatory Response (Caudate). Shared: Because at a minimum, pathways in this group must be shared by at least one tissue in each of the seven regions, the three "uniquely enriched" terms came from 9 tissues at a minimum.

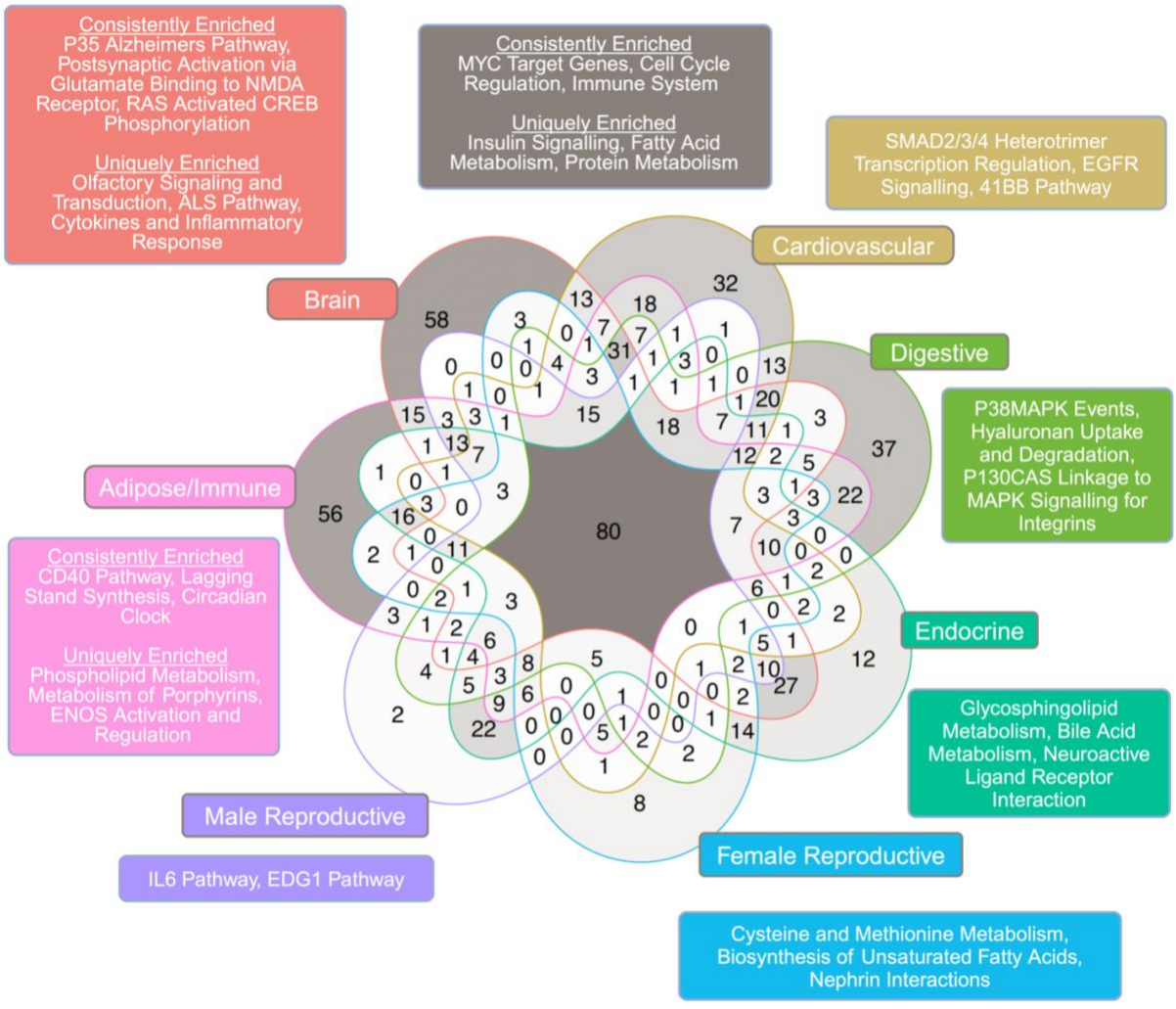


Figure 6: Numbers of common and rare ASD variants among brain and peripheral tissue Bayesian networks.
A) Raw counts of the number of genes in each network that were either known rare variants or common variants from the ASD GWAS were calculated and compared across each merged Bayesian network. B) Normalized counts of the number of genes in each network.

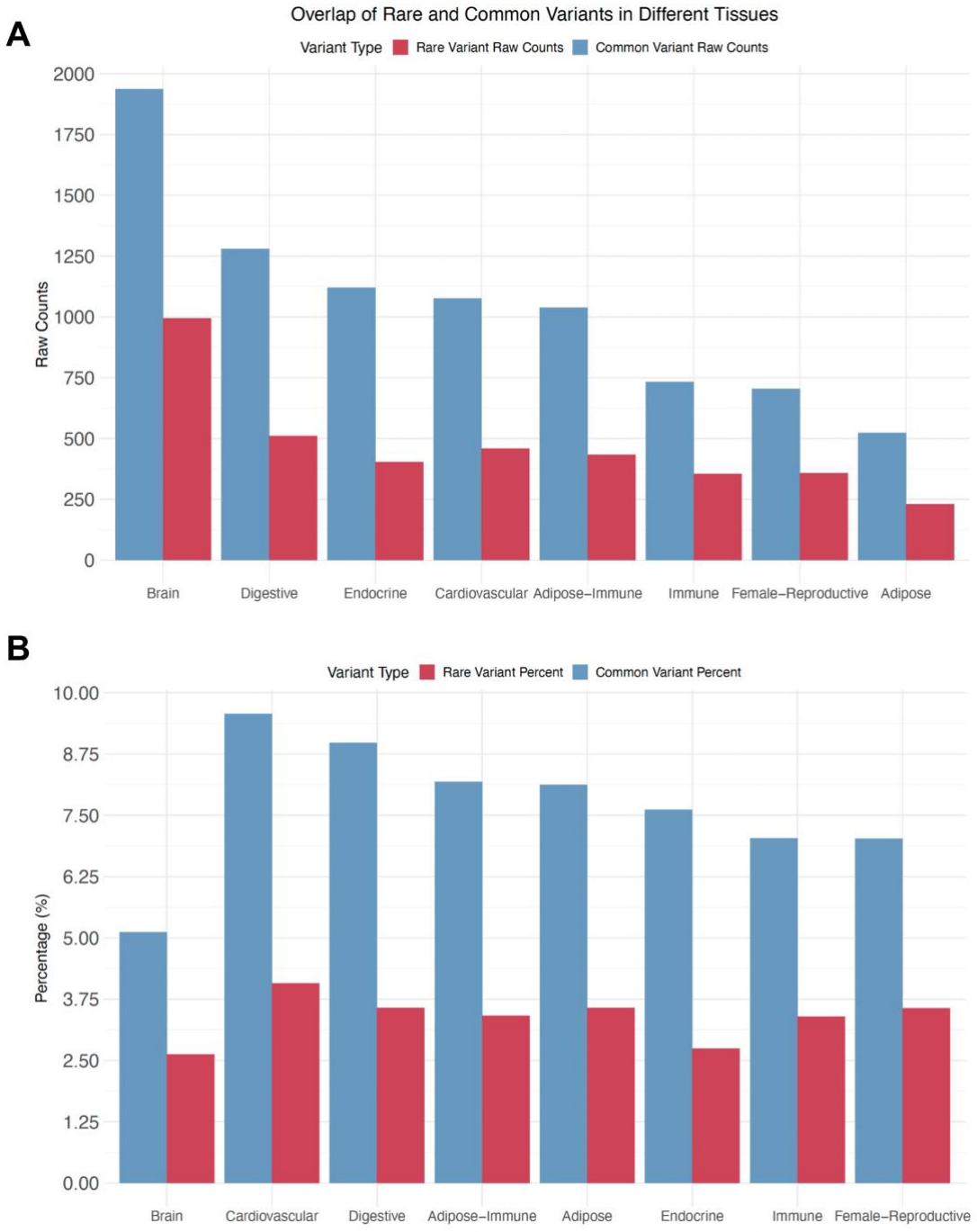


Figure 7 : Average percentage of key drivers overlapping with known ASD rare variants and common variants.
The percentage of key drivers for a given tissue and their overlap with known rare variants from the SFARI database and common variants from the ASD GWAS was calculated for brain and peripheral key drivers. Each dot corresponds to all the key drivers from a given brain or peripheral tissue. Given the non-normal distribution of the data and the percentages of the Shapiro-Wilk test (W = 0.95837, p-value = 0.008316), we performed a Wilcoxon test to determine the significance between groups. As we expected the medians of the brain tissues to be higher than that of the peripheral tissues, this was one-sided.

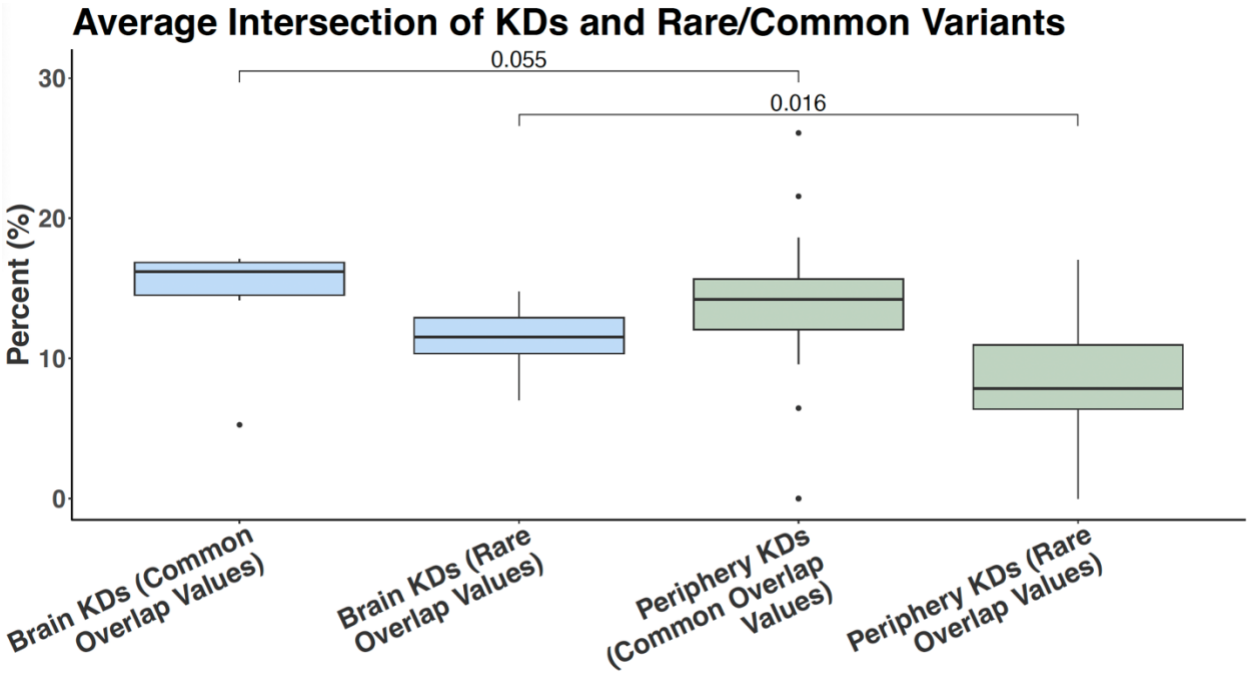


Figure 8: Overlap enrichment score comparisons of key drivers with rare variants and common variants.
Hypergeometric testing was utilized to compare a set of key drivers for a given tissue and the collective set of rare variants, or alternatively the common variants that correspond to that given tissue based on marker dependent filtering. For this comparison, all brain tissues were pooled into one category and all peripheral tissues into the other category. The significance was assessed in the context of a background set of genes, which for rare variants were the respective merged Bayesian networks that corresponded to the tissue, combined with all SFARI rare variants. For common variants, the background set of genes was the corresponding merged Bayesian network combined with the set of GWAS common variants that were mapped from the respective tissue. A two-sided Wilcoxon test was used due to the non-parametric nature of the data based on the Shapiro-Wilk test (W = 0.45273, p-value = 6.194e-16).

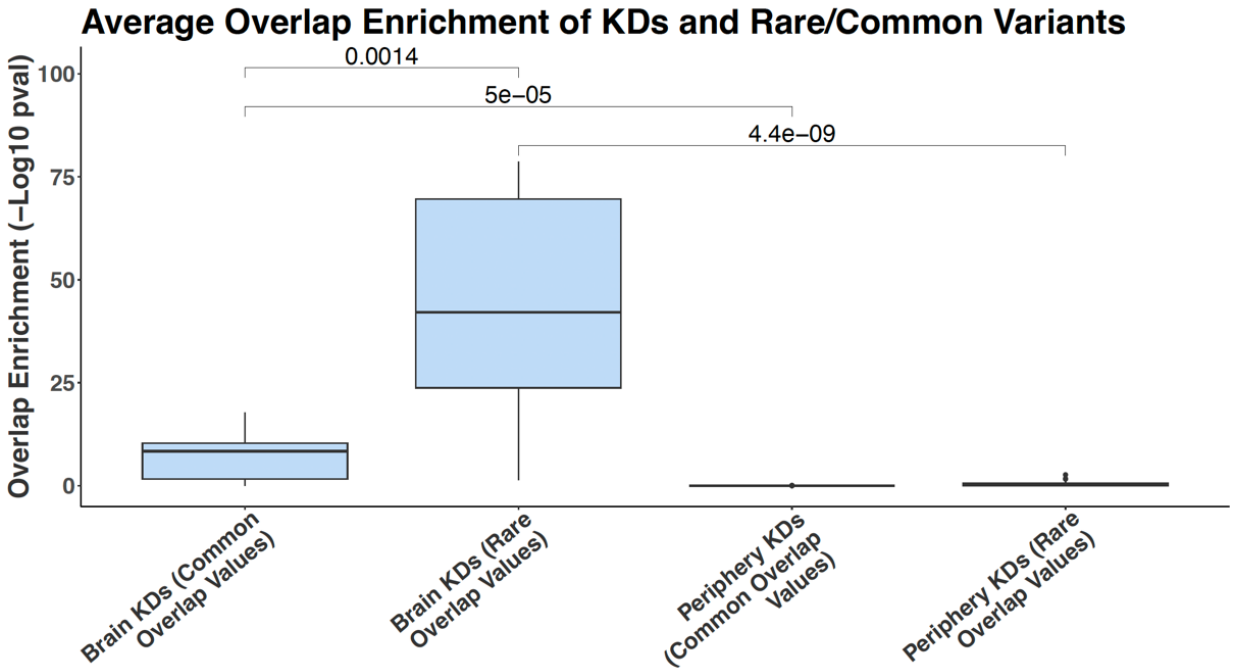


Figure 9: Top 15 key driver subnetworks ranked by rare variant overlap enrichment. Key drivers were ranked by their -*log10* transformed p-value generated from hypergeometric testing of the key driver subnetwork and its overlap with SFARI rare variants. Within each bar is the key driver's tissue(s) with the name of the gene set in parentheses. The numbers to the right of each bar display the fold enrichment for each subnetwork, which indicates the subnetwork's enrichment of key driver-regulated genes for known rare variants. The background set of genes for this analysis were all genes in the GTEx transcriptome data for each tissue. SN: Substantia nigra. PT: Pituitary. FC: Frontal Cortex.

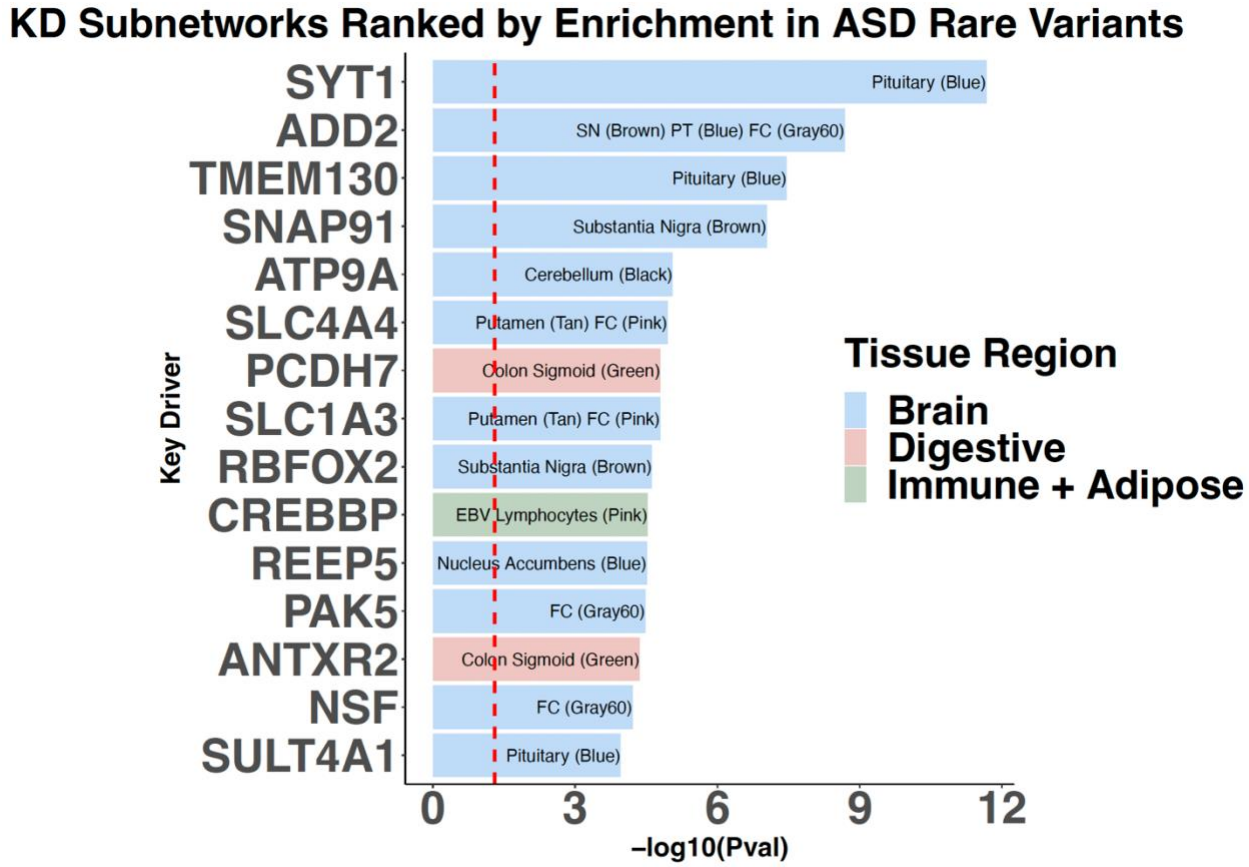


Figure 10: Top 15 key driver subnetworks ranked for common variant overlap enrichment. Key drivers were ranked on their -*log10* transformed p-value generated from hypergeometric testing of the key driver subnetwork and its overlap with ASD GWAS common variants. Within each bar is the key driver's tissue(s) with the name of the gene set in parentheses. The numbers to the right of each bar display the fold change for each subnetwork, which indicates the subnetwork's enrichment or depletion of key driver-regulated genes in relation to the common variants associated with the tissue of the key driver. The background set of genes for this analysis were in the GTEx transcriptome data for each tissue. SN: Substantia nigra. PT: Pituitary. FC: Frontal Cortex.

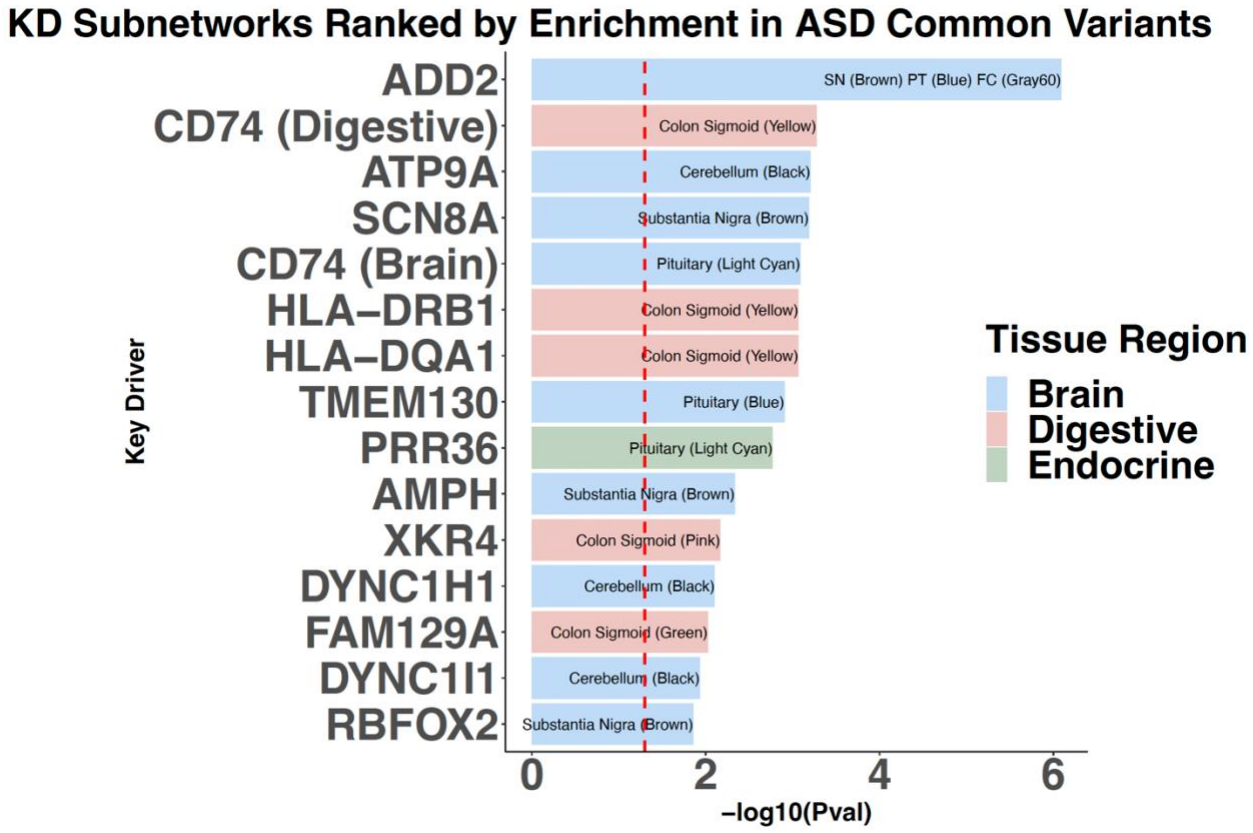


Figure 11: Venn Diagrams displaying the convergence of brain and peripheral tissues in terms of their rare and common variant overlap enrichment scores.
Sixteen key driver subnetworks shared an overlap enrichment p-value of less than 0.05 for both rare and common variants. The two diagrams separate the data by region and indicate that brain key driver subnetworks were more abundant. KD = Key Driver. CV = Common Variant. RV = Rare Variant.

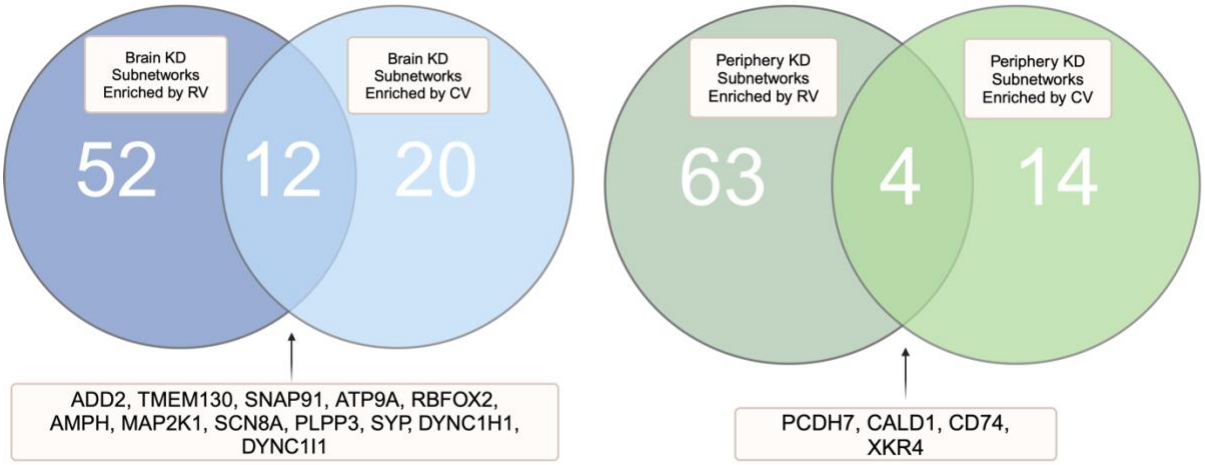


Figure 12: Key Driver Subnetwork for Synaptotagmin 1 (SYT1), a key driver with high rare variant overlap enrichment.
A) SYT1 and its first neighbors visualized using Cytoscape and colored based on their SFARI database ASD confidence stratification or by their ASD GWAS common variant disease association strength. B) EnrichR was used to analyze the key driver subnetwork genes to generate pathway annotation terms. Terms were ranked based on their *-log10* false discovery rate.

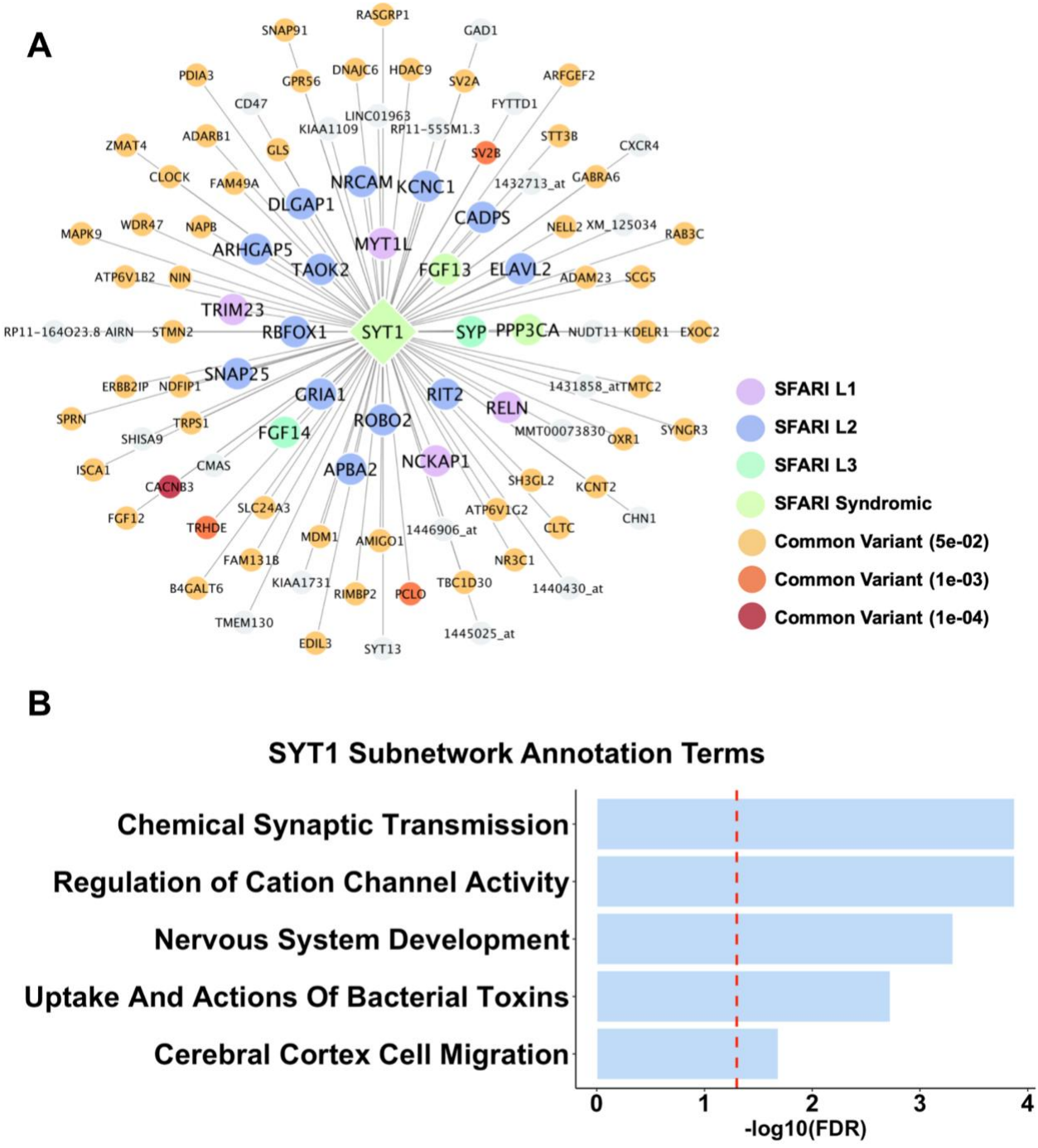
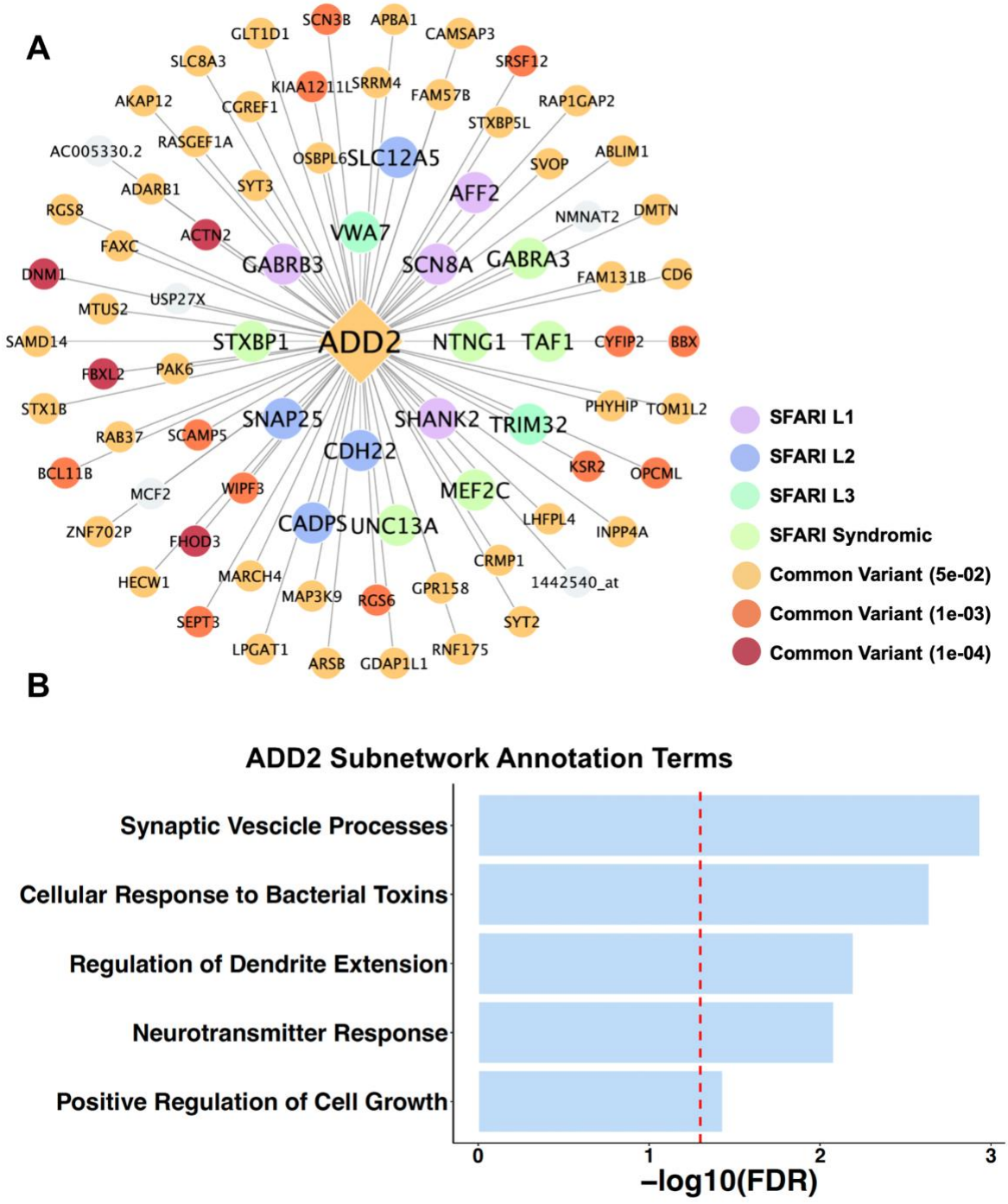


Figure 13: Key Driver Subnetwork for Adducin 2, (ADD2) a key driver with high rare and common variant overlap enrichment.
A) ADD2 and its first neighbors visualized using Cytoscape and colored based on their SFARI database ASD confidence level stratification or by their ASD GWAS common variant disease association strength. B) EnrichR was used to analyze the key driver subnetwork genes to generate pathway annotation terms. Terms were ranked based on their *-log10* false discovery rate.



References

1. Lai, M.C., M.V. Lombardo, and S. Baron-Cohen, *Autism.* Lancet, 2014. **383**(9920): p. 896-910.

2. Lord, C., et al., *Autism spectrum disorder.* Lancet, 2018. **392**(10146): p. 508-520.

3. Baxter, A.J., et al., *The epidemiology and global burden of autism spectrum disorders.* Psychol Med, 2015. **45**(3): p. 601-13.

4. *Diagnostic and statistical manual of mental disorders. 5th ed*. 2013: American Psychiatric Association.

5. Iakoucheva, L.M., A.R. Muotri, and J. Sebat, *Getting to the Cores of Autism.* Cell, 2019. **178**(6): p. 1287-1298.

6. Yasuda, Y., et al., *Genetics of autism spectrum disorders and future direction.* J Hum Genet, 2023. **68**(3): p. 193-197.

7. Gaugler, T., et al., *Most genetic risk for autism resides with common variation.* Nat Genet, 2014. **46**(8): p. 881-5.

8. Parikshak, N.N., et al., *Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism.* Nature, 2016. **540**(7633): p. 423-427.

9. Balsters, J.H., et al., *Disrupted prediction errors index social deficits in autism spectrum disorder.* Brain, 2017. **140**(1): p. 235-246.

10. Shen, M.D., et al., *Subcortical Brain Development in Autism and Fragile X Syndrome: Evidence for Dynamic, Age- and Disorder-Specific Trajectories in Infancy.* Am J Psychiatry, 2022. **179**(8): p. 562-572.

11. Richards, R., et al., *Increased hippocampal shape asymmetry and volumetric ventricular asymmetry in autism spectrum disorder.* Neuroimage Clin, 2020. **26**: p. 102207.

12. Hashem, S., et al., *Genetics of structural and functional brain changes in autism spectrum disorder.* Transl Psychiatry, 2020. **10**(1): p. 229.

13. Weston, C.S.E., *Four Social Brain Regions, Their Dysfunctions, and Sequelae, Extensively Explain Autism Spectrum Disorder Symptomatology.* Brain Sci, 2019. **9**(6).

14. Meltzer, A. and J. Van de Water, *The Role of the Immune System in Autism Spectrum Disorder.* Neuropsychopharmacology, 2017. **42**(1): p. 284-298.

15. Taniya, M.A., et al., *Role of Gut Microbiome in Autism Spectrum Disorder and Its Therapeutic Regulation.* Front Cell Infect Microbiol, 2022. **12**: p. 915701.

16. Masi, A., et al., *An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options.* Neurosci Bull, 2017. **33**(2): p. 183-193.

17. Boyle, E.A., Y.I. Li, and J.K. Pritchard, *An Expanded View of Complex Traits: From Polygenic to Omnigenic.* Cell, 2017. **169**(7): p. 1177-1186.

18. Shu, L., et al., *Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems.* BMC Genomics, 2016. **17**(1): p. 874.

19. Ding, J., et al., *Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics.* Nucleic Acids Res, 2021. **49**(W1): p. W375-W387.

20. Arneson, D., et al., *Systems spatiotemporal dynamics of traumatic brain injury at single-cell resolution reveals humanin as a therapeutic target.* Cell Mol Life Sci, 2022. **79**(9): p. 480.

21. Koplev, S., et al., *A mechanistic framework for cardiometabolic and coronary artery diseases.* Nat Cardiovasc Res, 2022. **1**(1): p. 85-100.

22. Ye, F., et al., *Shared Genetic Regulatory Networks Contribute to Neuropathic and Inflammatory Pain: Multi-Omics Systems Analysis.* Biomolecules, 2022. **12**(10).

23. Zuo, Y., et al., *Chronic adolescent exposure to cannabis in mice leads to sex-biased changes in gene expression networks across brain regions.* Neuropsychopharmacology, 2022. **47**(12): p. 2071-2080.

24. Grove, J., et al., *Identification of common genetic risk variants for autism spectrum disorder.* Nat Genet, 2019. **51**(3): p. 431-444.

25. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.

26. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.

27. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans.* Bioinformatics, 2010. **26**(17): p. 2190-1.

28. Abrahams, B.S., et al., *SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs).* Mol Autism, 2013. **4**(1): p. 36.

29. Keen, J.C. and H.M. Moore, *The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine.* J Pers Med, 2015. **5**(1): p. 22-9.

30. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.

31. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-D361.

32. Gillespie, M., et al., *The reactome pathway knowledgebase 2022.* Nucleic Acids Res, 2022. **50**(D1): p. D687-D692.

33. Rouillard, A.D., et al., *The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins.* Database (Oxford), 2016. **2016**.

34. Zhu, J., et al., *Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks.* Nat Genet, 2008. **40**(7): p. 854-61.

35. Wensveen, F.M., et al., *Interactions between adipose tissue and the immune system in health and malnutrition.* Semin Immunol, 2015. **27**(5): p. 322-33.

36. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

37. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. **14**: p. 128.

38. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.

39. Xie, Z., et al., *Gene Set Knowledge Discovery with Enrichr.* Curr Protoc, 2021. **1**(3): p. e90.

40. Gene, S., *SYT1 - Synaptotagmin 1*. 2023.

41. National Center for Biotechnology Information, U.S.N.L.o.M. *SYT1 Synaptotagmin 1 [Homo Sapiens (Human)] - Gene - NCBI*. 2024; Available from: www.ncbi.nlm.nih.gov/gene/6857.

42. Database, G.T.H.G., *ADD2 Gene - Adducin 2*. 2024.

43. Qi, C., et al., *Variants in ADD1 cause intellectual disability, corpus callosum dysgenesis, and ventriculomegaly in humans.* Genet Med, 2022. **24**(2): p. 319-331.

44. Chang, J., et al., *Genotype to phenotype relationships in autism spectrum disorders.* Nat Neurosci, 2015. **18**(2): p. 191-8.

45. Klei, L., et al., *How rare and common risk variation jointly affect liability for autism spectrum disorder.* Mol Autism, 2021. **12**(1): p. 66.

46. van Rooij, D., et al., *Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group.* Am J Psychiatry, 2018. **175**(4): p. 359-369.

47. Laidi, C., et al., *Decreased Cortical Thickness in the Anterior Cingulate Cortex in Adults with Autism.* J Autism Dev Disord, 2019. **49**(4): p. 1402-1409.

48. Guo, B., et al., *Anterior cingulate cortex dysfunction underlies social deficits in Shank3 mutant mice.* Nat Neurosci, 2019. **22**(8): p. 1223-1234.

49. Wang, S. and X. Li, *A revisit of the amygdala theory of autism: Twenty years after.* Neuropsychologia, 2023. **183**: p. 108519.

50. Fetit, R., et al., *The neuropathology of autism: A systematic review of post-mortem studies of autism and related disorders.* Neurosci Biobehav Rev, 2021. **129**: p. 35-62.

51. Donovan, A.P. and M.A. Basson, *The neuroanatomy of autism - a developmental perspective.* J Anat, 2017. **230**(1): p. 4-15.

52. Carper, R.A. and E. Courchesne, *Localized enlargement of the frontal cortex in early autism.* Biol Psychiatry, 2005. **57**(2): p. 126-33.

53. Goralczyk-Binkowska, A., D. Szmajda-Krygier, and E. Kozlowska, *The Microbiota-Gut-Brain Axis in Psychiatric Disorders.* Int J Mol Sci, 2022. **23**(19).

54. Firestein, M.R., et al., *Elevated prenatal maternal sex hormones, but not placental aromatase, are associated with child neurodevelopment.* Horm Behav, 2022. **140**: p. 105125.

55. Rotem, R.S., et al., *Maternal Thyroid Disorders and Risk of Autism Spectrum Disorder in Progeny.* Epidemiology, 2020. **31**(3): p. 409-417.

56. Rosenfeld, C.S., *The placenta-brain-axis.* J Neurosci Res, 2021. **99**(1): p. 271-283.

57. Gardener, H., D. Spiegelman, and S.L. Buka, *Prenatal risk factors for autism: comprehensive meta-analysis.* Br J Psychiatry, 2009. **195**(1): p. 7-14.

58. Tioleco, N., et al., *Prenatal maternal infection and risk for autism in offspring: A meta-analysis.* Autism Res, 2021. **14**(6): p. 1296-1316.

59. Ravaccia, D. and T. Ghafourian, *Critical Role of the Maternal Immune System in the Pathogenesis of Autism Spectrum Disorder.* Biomedicines, 2020. **8**(12).

60. Younesian, S., et al., *The DNA Methylation in Neurological Diseases.* Cells, 2022. **11**(21).

61. Rhodus, E.K., et al., *Behaviors Characteristic of Autism Spectrum Disorder in a Geriatric Cohort With Mild Cognitive Impairment or Early Dementia.* Alzheimer Dis Assoc Disord, 2020. **34**(1): p. 66-71.

62. Rhodus, E.K., et al., *Comparison of behaviors characteristic of autism spectrum disorder behaviors and behavioral and psychiatric symptoms of dementia.* Aging Ment Health, 2022. **26**(3): p. 586-594.

63. Liu, M., et al., *A common spectrum underlying brain disorders across lifespan revealed by deep learning on brain networks.* iScience, 2023. **26**(11): p. 108244.

64. Khan, S.A., et al., *Alzheimer's Disease and Autistic Spectrum Disorder: Is there any Association?* CNS Neurol Disord Drug Targets, 2016. **15**(4): p. 390-402.

65. Mari-Bauset, S., et al., *Systematic review of prenatal exposure to endocrine disrupting chemicals and autism spectrum disorder in offspring.* Autism, 2022. **26**(1): p. 6-32.

66. Loyacono, N., et al., *Gastrointestinal, nutritional, endocrine, and microbiota conditions in autism spectrum disorder.* Arch Argent Pediatr, 2020. **118**(3): p. e271-e277.

67. Manzi, B., et al., *Autism and metabolic diseases.* J Child Neurol, 2008. **23**(3): p. 307-14.

68. McLellan, J., et al., *Maternal Immune Dysregulation and Autism-Understanding the Role of Cytokines, Chemokines and Autoantibodies.* Front Psychiatry, 2022. **13**: p. 834910.

69. Erbescu, A., et al., *Re-emerging concepts of immune dysregulation in autism spectrum disorders.* Front Psychiatry, 2022. **13**: p. 1006612.

70. Park, H.R., et al., *A Short Review on the Current Understanding of Autism Spectrum Disorders.* Exp Neurobiol, 2016. **25**(1): p. 1-13.

71. Alvares, G.A., et al., *Study protocol for the Australian autism biobank: an international resource to advance autism discovery research.* BMC Pediatr, 2018. **18**(1): p. 284.

72. Trost, B., et al., *Genomic architecture of autism from comprehensive whole-genome sequence annotation.* Cell, 2022. **185**(23): p. 4409-4427 e18.

73. Chung, M.K., et al., *Plasma metabolomics of autism spectrum disorder and influence of shared components in proband families.* Exposome, 2021. **1**(1): p. osab004.

74. Sun, W., et al., *Histone Acetylome-wide Association Study of Autism Spectrum Disorder.* Cell, 2016. **167**(5): p. 1385-1397 e11.

75. Rodriguez-Fontenla, C. and A. Carracedo, *UTMOST, a single and cross-tissue TWAS (Transcriptome Wide Association Study), reveals new ASD (Autism Spectrum Disorder) associated genes.* Transl Psychiatry, 2021. **11**(1): p. 256.