Investigating Individual Differences in Autism Spectrum Disorder Through Genetic and Functional Connectivity Variability

Julianna Pijar

Advisors: Stefano Anzellotti, Ph.D. and Aidas Aglinskas, Ph.D.

Boston College
Department of Psychology and Neuroscience
Undergraduate Senior Thesis
May 2023

TABLE OF CONTENTS

ABSTRACT

Autism Spectrum Disorder (ASD) displays uniquely in every individual, creating disparities in symptom severity, genetics, and functional connectivity. Examining the relationship between genetic and functional connectivity variability could help to better understand individual differences in ASD. From this, improved diagnosis, treatment, and understanding of ASD can be developed. To resolve individual differences in symptom severity and presentation, I generated matrices of subject functional connectivity data and compared this to gene expression maps. Multivariate regression analysis was performed on the data to anticipate ASD symptoms from these correlation matrices and to establish which genes have the largest impact on these predictions. The ANOVAs ran on the data were not significant, but there were several genes implicated in specific aspects of ASD. STX1A, MVP, CDKL5, and RABEP2 were the only genes correlated across more than one subtype of ASD. These results pave the way for future research to investigate the roles of these genes in a larger size of ASD subjects.

I. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social situations, communication, and restrictive repetitive behaviors (RRB). These can display as deficits in nonverbal communication, social reciprocation, and social relationships, as well as fixated narrow interests, rigidity in routines, or hypersensitization to sensory stimuli in the environment (APA Manual, 2019).

Autism is a spectrum due to the heterogeneity between individuals' symptom presentation and severity. This displays uniquely in every individual, creating disparities in symptom severity, genetics, and functional connectivity (Georgiades, Szatmari, and Boyle, 2013). Lombardo et al (2019) is a review paper that tries to better comprehend and model these data within the autism spectrum. The authors discussed the terminology used to describe ASD and explained how simplifying it could further point to discerning the heterogeneity of different types of models within it. There are several factors to weigh in when analyzing the variability within ASD, such as development, which can lead to decomposing this variability in ASD. Shifting to a focus on the evolution and history of diagnostic criteria, they noted how previous research is not as expansive and inclusive as current findings. The historical idea of there being one definitive type of autism to explain all has significantly affected treatment for individuals with ASD. Treatment commonly follows an approach of "one size fits all", which the paper vehemently argues against. Interpreting heterogeneity is imperative for progress in precision medicine. Problems associated with current ASD studies include small sample sizes, sample bias, and a lack of available data. Variability from small sample sizes is often not a correct representation of the ASD population, but typically an exaggeration of the effects of the true population. Solutions they presented to fix this would be through the top down approach, bottom up approach, and performing longitudinal studies. Despite the challenges in acquiring such data,

the researchers remain optimistic that future research can prevent this trend and positively support the ASD community by finding enhanced treatment. Disentangling this heterogeneity is crucial to provide better diagnoses, treatment, and understanding of ASD. From this, individualized treatments can be created. Recognizing the subtypes of ASD and identifying potential genetic or functional connectivity biomarkers could help clarify these discrepancies.

Like ASD, depression is another disorder that shows high levels of heterogeneity and has little research or knowledge widely known about it. Drysdale et al (2017) used Functional Magnetic Resonance Imaging (fMRI) to divide depression into four subtypes, with the goal of better treating and diagnosing patients with depression. Ultimately, they wished to resolve which groups would respond best to Transcranial Magnetic Stimulation (TMS). Employing statistical analyses, they clustered participants with similar levels of atypical functional connectivity in resting state networks. After exclusion criteria were applied, the participants were placed in the fMRI scanner. The researchers employed a seed based approach to estimate functional connectivity to identify where the matrices are. From this, they were able to define depression subtypes with similar clinical-symptom profiles and create connectivity biomarkers. These connectivity biomarkers could predict how patients would respond to TMS and which regions to effect to have a maximal response. They further tested this system on patients with generalized anxiety disorder and schizophrenia, finding that it also placed them under one of the connectivity subtypes. Using biomarkers to cluster subjects in both of these studies helped to generate more individualized treatment, yet more research still needs to be conducted before treatment can be subtyped. The discovery of identifiable biomarkers in depression proves that it is possible to find them for other disorders that have the obstacle of individual differences. I hope to uncover

particular genes in my research similar to how these biomarkers helped to identify individual variation in subjects with depression.

Genetics play a role in the etiology of ASD, but the specific influences each gene has are yet to be determined (Chaste and Leboyer, 2012). Exploring this link, Silva et al (2022) sought to discover risk mechanisms across the genome implicated in the pathophysiology of neurodevelopmental disorders. To investigate this, they used Genetic Copy Number Variants (CNVs) present in a variety of neurodevelopmental disorders. Deletions and duplications at specific genes can pose a risk and serve as biomarkers for distinct disorders. The researchers set out to uncover how these established genes lead to a particular disorder and what factors determine how someone acquires it. Neuroimaging methods in CNV cohorts were applied to examine surface area, total brain volume, and cortical thickness. The genetic regions they examined were 15q11.2, 16p11.2 distal, 16p11.2 proximal, 1q21.1 distal, 7q11.23 William syndrome critical region, and 22q11.2 DiGeorge syndrome critical region. In the 15q11.2 deletion group, there was a smaller surface area and thicker cortices than the control. The 22q11.2 deletion group displayed reductions in surface area, primarily in the medial occipital and anterior cingulate cortex. Overall, they found that subjects with gene deletions showed worse cognitive performance in assessments than the control group. These results highlight the importance of specific genes in neurodevelopmental disorders and open up future research in this area. Convergence across CNVs, when better understood, will provide new treatment options and enhanced diagnosing of these disorders.

Expanding upon these findings, Li et al (2012) intended to ascertain the exact genes involved in ASD. This review paper presented an overview of 13 genes reported to be associated with ASD through molecular, cytogenetic, and linkage studies. Significant genes reported were:

GABR, NLGN, OXTR, MET, SLC25A12, RELN, and SLC6A4. These results were corroborated by Ma et al, Quartier et al, Lerer et al, Skaar et al, and many others. These genes discussed in this paper and Silva et al contributed as a framework for the genetic analyses in my thesis. The issues of heterogeneity explained in all articles clarified the challenge of classifying a genetic or functional biomarker, while serving as a concise background of current knowledge.

Implementing multiple levels of analysis, I aimed to predict ASD symptoms in participants across four categories of the Autism Diagnostic Observation Schedule (ADOS): Total, Social, Communication, and RRB. Despite the complexity of ASD, if there is a connection between genetic variability and functional connectivity variability it could help improve diagnosis and treatment selection. From this, individualized treatment strategies can be established to enhance the quality of life of an autistic patient. I hypothesized that there would be a correlation between genetic variability and functional connectivity variability. Specific genes found in ASD will possibly be related to distinct ADOS categories (Total, Social, Communication, and RRB). From this association, I can predict a subject's ADOS score from the relationship between their genetics and functional connectivity. The goal of this study is to determine if using a multimodal approach of genetic and neuroimaging data can better capture symptom individual variability than utilizing only one.

II. Methods

To examine functional connectivity, I acquired data from the Autism Brain Imaging Data Exchange (ABIDE), which encompassed 699 ASD participants (Martino et al, 2014). Additional data from the Simons Foundation Autism Research Initiative (SFARI) was also compiled, consisting of 121 ASD participants (Simons Vip Consortium, 2012). This data reported the ADOS symptoms of Total, Social, and RRB. There was no information on ADOS

Communication. For this data, I exclusively used genes associated with 16p11.2 duplications and deletions to match those of the subjects. This consisted of 34 genes, resulting in a 121 by 34 matrix. The statistical analyses mentioned above were applied again to this data.

The brain scans were then preprocessed using fMRI Prep (Esteban et al, 2018), a data preprocessing pipeline that reads in raw fMRI data and produces preprocessed data ready for analysis. Following this, the data were normalized to a Montreal Neurological Institute (MNI) space, which designated the bounds of the brain. Each scans' volumes were scrubbed, with a threshold of 0.25mm, to correct for unwanted motion during the time of the scan. At each time point, a subject's movements were calculated and if they moved greater than 0.25mm, that volume was dropped from their scan. The rest of the volumes were concatenated together. A participant was removed if there was less than 4 minutes of resting state data. This was further processed using CompCor (Behzadi et al, 2007), which corrects for noise in fMRI through the input of the Echo Planar Imaging (EPI) signal and noise regions of interest from the scan. It takes brain activity in regions of no interest, such as the ventricles, to predict activity in areas of interest, like grey matter. All predictions were subtracted from the EPI signal and the remaining areas were kept in the scan. After this removal, filtering was administered to narrow down to a smaller band of blood-oxygen-level-dependent (BOLD) signal activity. Any activity outside of 0.1Hz to 0.01Hz, or between 10 and 100 seconds, was removed. Frequencies outside of this metric were likely to be noise, such as cardiac and respiratory oscillations (Heuvel and Pol, 2010). Once the scans had been fully preprocessed, functional connectivity was extracted using an atlas of 51 regions (Yeo et al, 2011). This was employed to measure the interaction between time series of these areas. A pairwise correlation was calculated across this time series, creating a 51 by 51 functional connectivity matrix.

Connectivity matrices can differ across age (Xie et al, 2020), IQ (Dubois et al, 2018), gender (Sie et al, 2019), and personality (Dubois et al, 2018). Some of these variables can be shared between the ASD and typically developed (TD) populations, such as age (Tecoulesco et al, 2019) and IQ (Leno et al, 2017). But there is still variability specific to ASD (Harlalka et al, 2019). Shared variability must be controlled to conclude which of the differences amongst participants are particular to ASD.

To resolve this issue, I implemented contrastive variational autoencoders (CVAE) (Abid and Zou, 2019) that separate shared features from ASD specific ones in the population to design reconstructed functional connectivity matrices (Aglinskas et al, 2022). A TD twin, that has the same brain as the subject but without ASD, was configured from the shared features. The reconstructed and TD twin matrices were subtracted to build a difference matrix, or diffmat. Through this process, these diffmats exhibited the functional connections altered by ASD. I compared the results of these diffmats to correlation matrices that included both the shared and specific features, referred to as cmats. There were no significant differences between the cmats and diffmats, so I will only report on the results of the latter.

For the genetic component, I obtained gene expression maps from Neurosynth.org. These create a map of where a certain gene is expressed in the brain, based on findings from numerous research studies. They measure the relative intensity of how much that gene is expressed. To decide which genes to use, I researched genes known to be linked to ASD from various studies. I applied the same 13 genes that Li et al (2012) highlighted in their results and the CNV regions described in Silva et al (2022). From these papers, I identified 122 genes connected to ASD that I downloaded into JupyterLab.

I then proceeded to make a correlation matrix of genes and diffmat data to better relate functional connectivity and genetic variability. A gene vector was generated from the gene expression in select Regions of Interest (ROI). The diffmat data were then averaged by row to create the functional connectivity vector. For each region, the gene vector contained the amount of gene expression and the functional connectivity vector consisted of the amount of connectivity change. These two vectors were flattened to be the same size, allowing comparisons to be made between them. Gene expression similarity maps then needed to be established, to determine where the selected genes were expressed in these subjects. ROI expression, as previously concluded by the EPI signal, was subtracted from both the genemaps and diffmats. Next, I correlated the gene vector with the functional connectivity vector, which was constructed from the gene expression similarity maps to construct a more compressed form of data. I added this to a correlation array with the subject and gene information. Altogether, this produced a large correlation matrix of data that I could utilize to perform statistical analyses on (Figure 1).

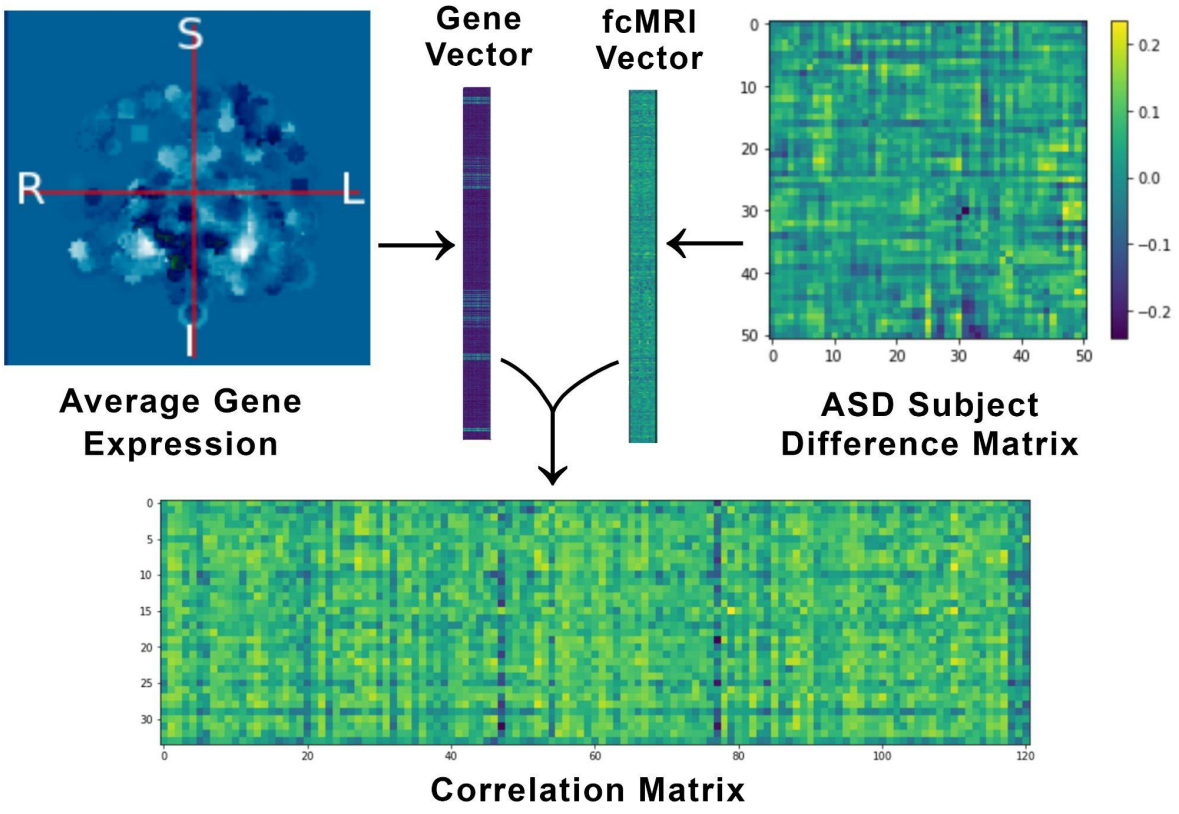


Figure 1: A gene vector was made by the average gene expression, as determined by the genemaps. A functional connectivity (fcMRI) vector was developed from subject diffmat data. Both vectors were correlated to make a correlation matrix. The columns are the correlations for every participant and the rows are the correlations for each gene. Each matrix coordinate corresponds to the correlation of the expression of a specific gene with the functional connectivity data for a certain subject.

Multivariate regression was carried out on the data to predict ASD symptoms and establish which genes have the largest impact on these predictions. For each ADOS measure, I calculated a scatter plot of the predicted score versus the actual score, plotted the correlation coefficients of each gene, and computed the variance explained. Analysis of Variance (ANOVAs) were also run on the data to conclude if behavioral symptoms are predicted by exclusive genes or groups of genes correlated in the gene-fMRI correlation matrix. Following these results, I added more genes from Silva et al and completed the same statistical procedure.

Due to the large amount of genes in the analysis, I applied Principal Components Analysis (PCA) to the data (Lever et al, 2017). The PCA reduced the gene-fMRI correlation matrix to 10 by 699 participants, instead of 122 by 699. The same statistical analysis was carried out, producing data with ten variables instead of 122.

III. Results

3.1 ABIDE Data

From the diffmats in the ABIDE data, none of the ANOVAs were significant and 19 out of 122 genes were significantly correlated with functional connectivity activity from the diffmats. In the ADOS Total category, PRKAB2, CDKL5, STX1A, SYN2, P2RX6, and MVP showed significance in correlating with the diffmats. For ADOS Communication: CDIPT,

YPEL3, STX1A, CLDN5, WBSCR22, KCTD13, and MVP. ADOS Social and RRB had fewer genes each: ZDHHC8, CDKL5, and SYN2 for Social; EN2, NIPA2, BCL9 for RRB. The scatterplots for each metric did not vary from one another, each resulting in an R^2 value around 0.3 (Figure 2). Additionally, there were no significant *P* values: ADOS Total, $R^2 = 0.29$, $P = 0.434$; ADOS Communication, $R^2 = 0.297$, $P = 0.995$; ADOS Social, $R^2 = 0.302$, $P = 0.428$; ADOS RRB, $R^2 = 0.314$, $P = 0.637$. Percent variance explained for these scores ranged from 28.95% to 31.36%. ADOS Total had the least variance explained, at 28.95%, while ADOS RRB had the greatest at 31.36%. ADOS Communication had 29.65%, while ADOS Social was 30.19%.

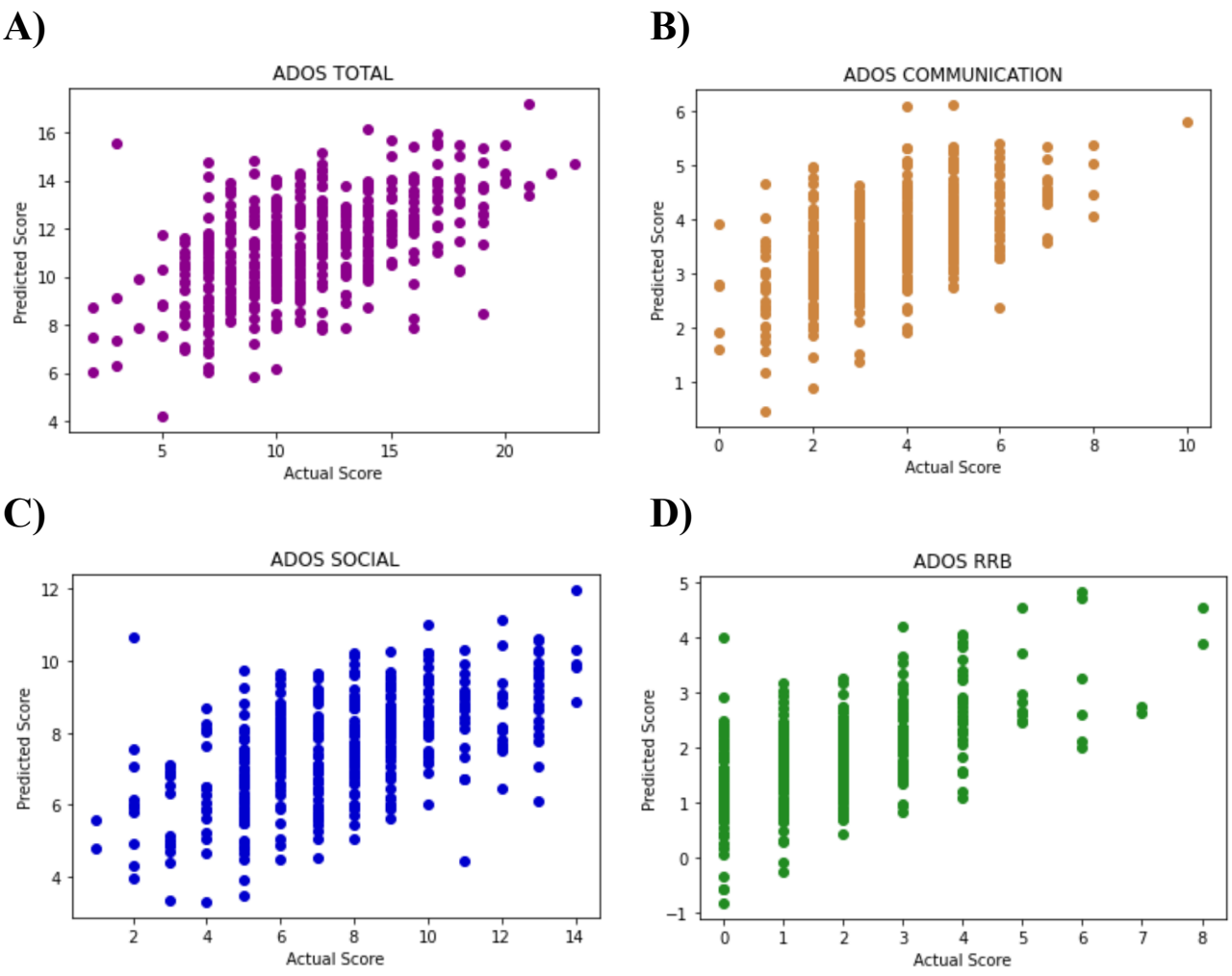
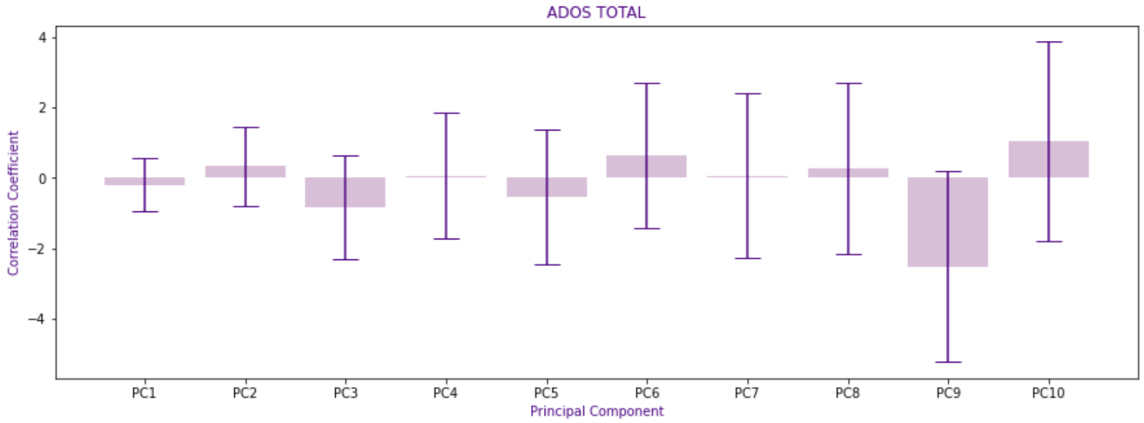


Figure 2 ***(A-D)****: For each ADOS score in the ABIDE data, a scatterplot compared the actual ADOS score from subject data to the predicted score. The predicted score was obtained from the results of the multivariate linear regression performed on the matrix correlating gene and fMRI data.* R^2 *average was 0.3, meaning that 30% of the variability in the ADOS scores is explained by the predictor variable. This low value demonstrates that a small portion of the variability was explained by this model. Overall, the results of the multivariate linear regression analysis on the ABIDE data is not significant.*

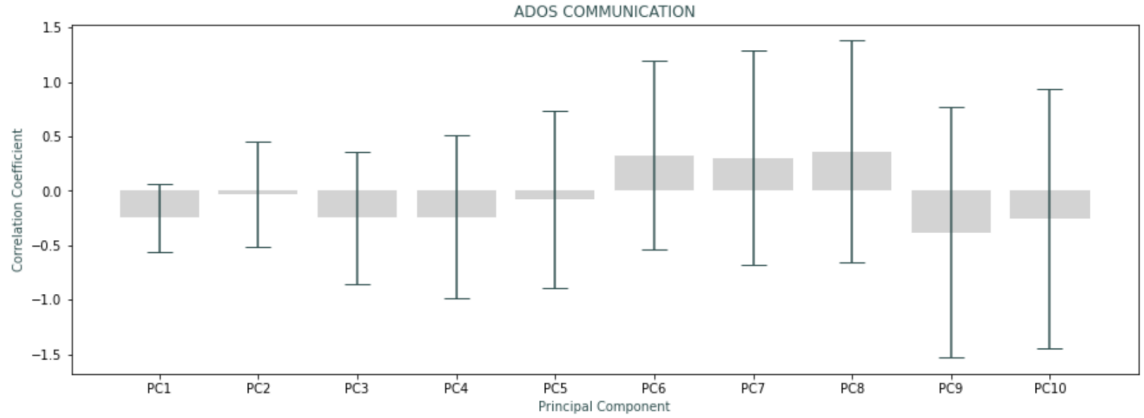
3.2 Principal Components Analysis

None of the ANOVAs performed on the ten principal components were significant (Figure 3). Percent variance demonstrated low values, with all scores between 1.33% and 2.27%. ADOS Social had the most variance explained at 2.27%, and ADOS Communication had the lowest at 1.33%. ADOS Total was 1.61% and ADOS RRB was 1.99%.

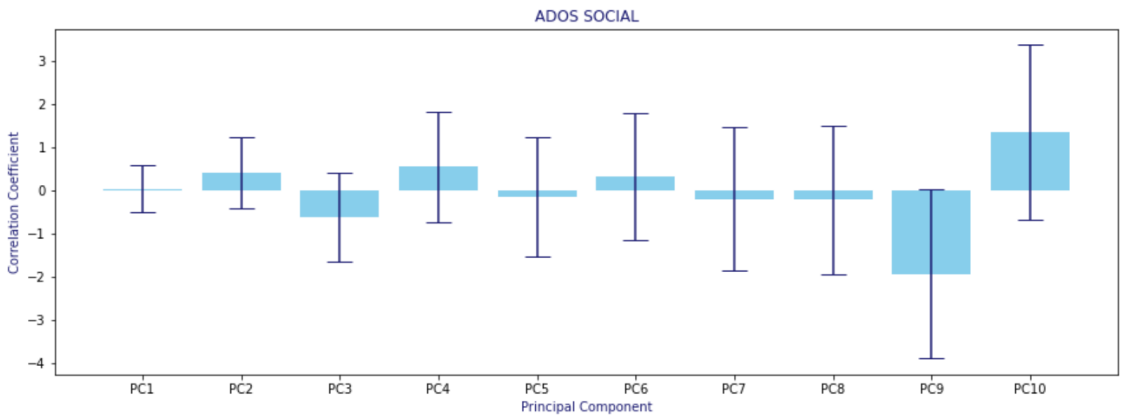
A)



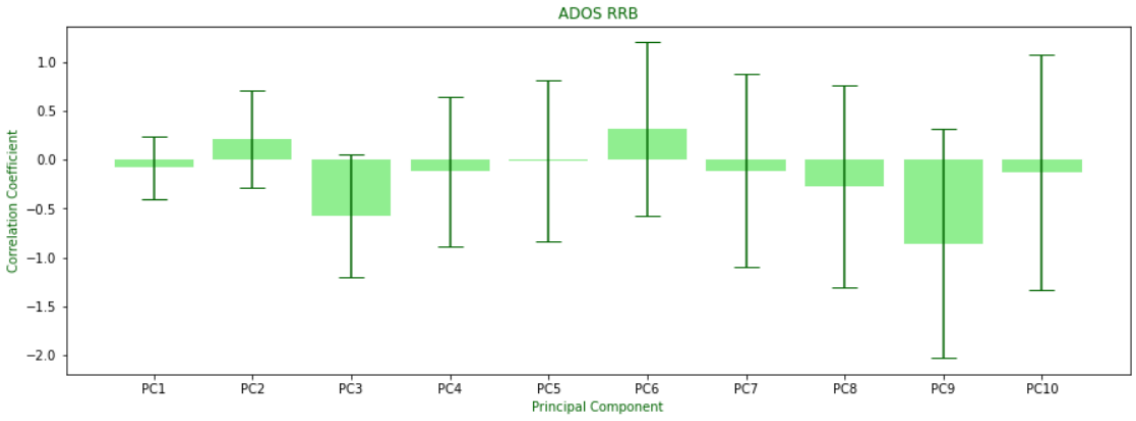
B)



C)



D)



<u>Figure 3</u> ***(A-D)****: Due to the large amount of predictors, a Principal Components Analysis was applied. It made ten principal components (PC) from the initial 122 genes. The correlation coefficient for each ADOS score was measured across these PCs. Correlation coefficient values were determined by the multivariate regression calculated on the data. None of the ANOVAs performed on the PCs were significant. Each produced a low* R^2 *and high P value, indicating that the data was not statistically significant in general.*

The scatterplots for each ADOS score displayed less of a trend compared to the ABIDE data (Figure 4). R^2 values were close to 0 for each metric and there was no significance from the

correlation coefficients: ADOS Total, $R^2 = 0.016$, $P = 0.740$; ADOS Communication, $R^2 = 0.013$, $P = 0.861$; ADOS Social, $R^2 = 0.023$, $P = 0.501$; ADOS RRB, $R^2 = 0.020$, $P = 0.739$.

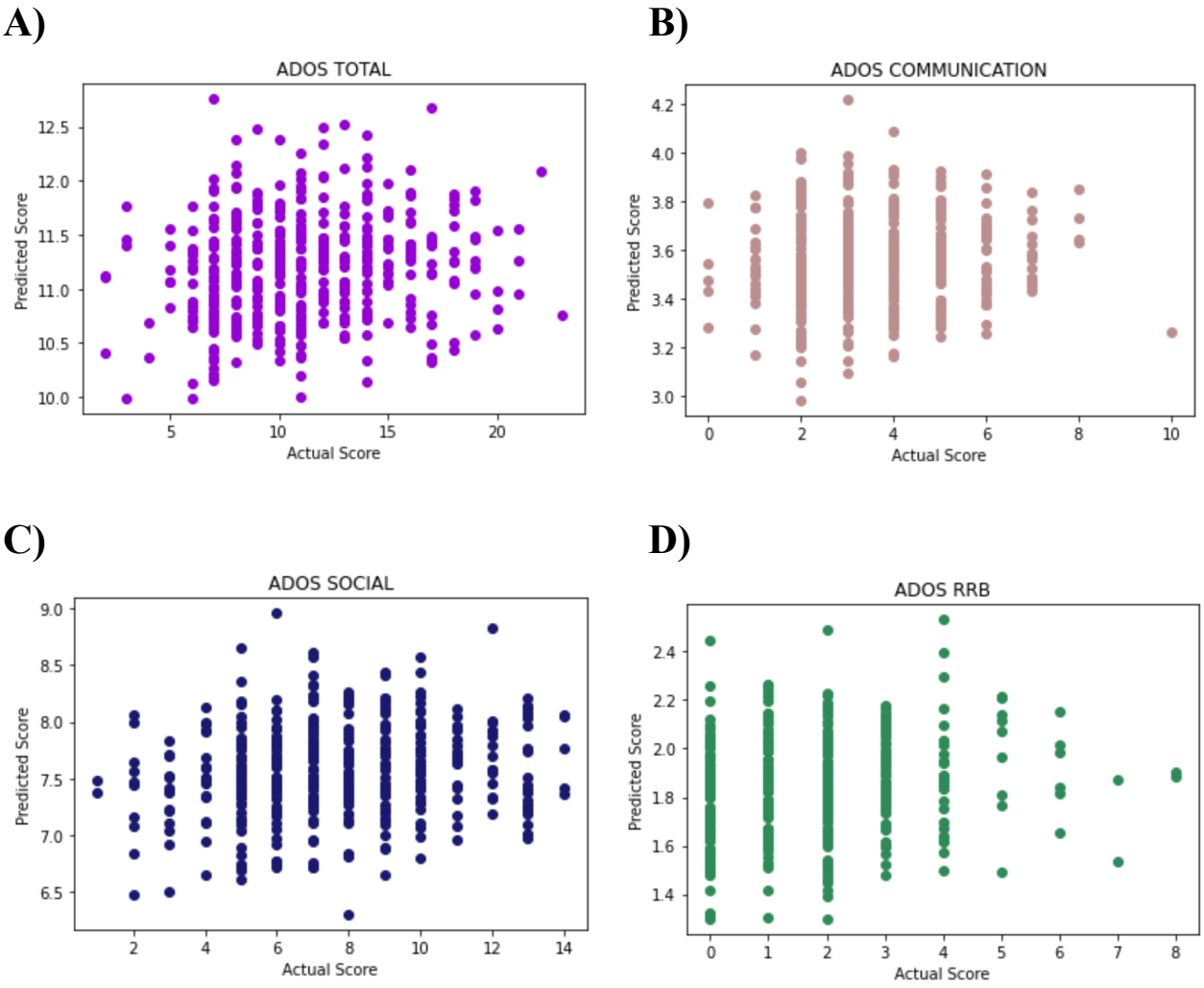


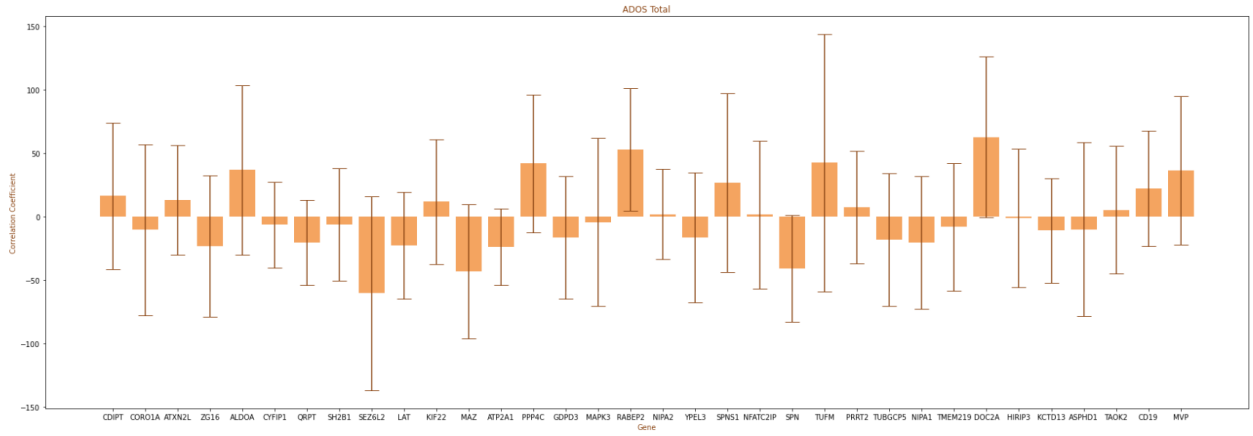
Figure 4 ***(A-D)****: A scatter plot developed from the multivariate linear regression calculated on the Principal Components Analysis. It was created from the actual ADOS scores of each subject and the predicted score, which came from the results of the multivariate linear regression operated on the matrix correlating gene and fMRI data. The average of* R^2 *was 0.018, lower than that of the ABIDE and SFARI data. Equating out to 1.8%, this model explains a limited portion of the variability. Along with the results*

of the correlation coefficient graph, these scatterplots reinforce that the multivariate regression analysis on the Principal Components Analysis of the ABIDE data is not significant.

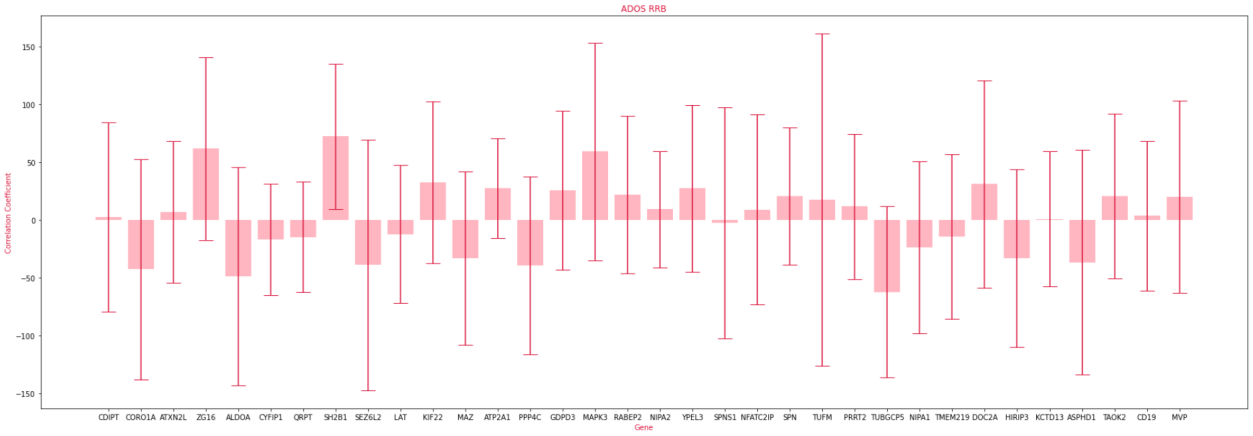
3.3 SFARI Data

The SFARI data delivered more definitive results, with 7 of 34 genes correlated (Figure 5). But once again, none of the ANOVAs operated on the data were significant. ADOS Total and ADOS RRB only had one correlated gene each: RABEP2 for Total and SH2B1 for RRB. ADOS Social had five genes with significance: ATP2A1, PPP4C, RABEP2, SPN, and DOC2A.

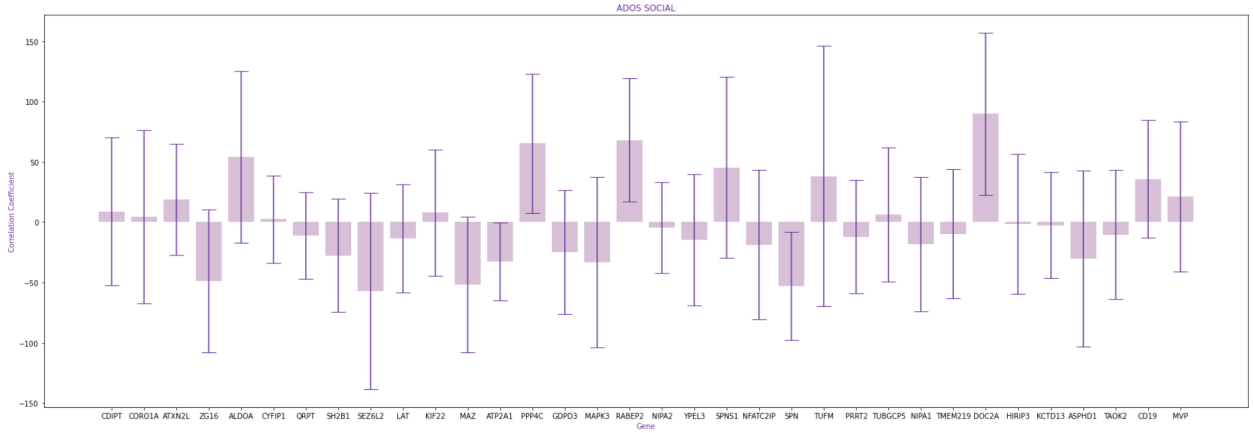
A)



B)

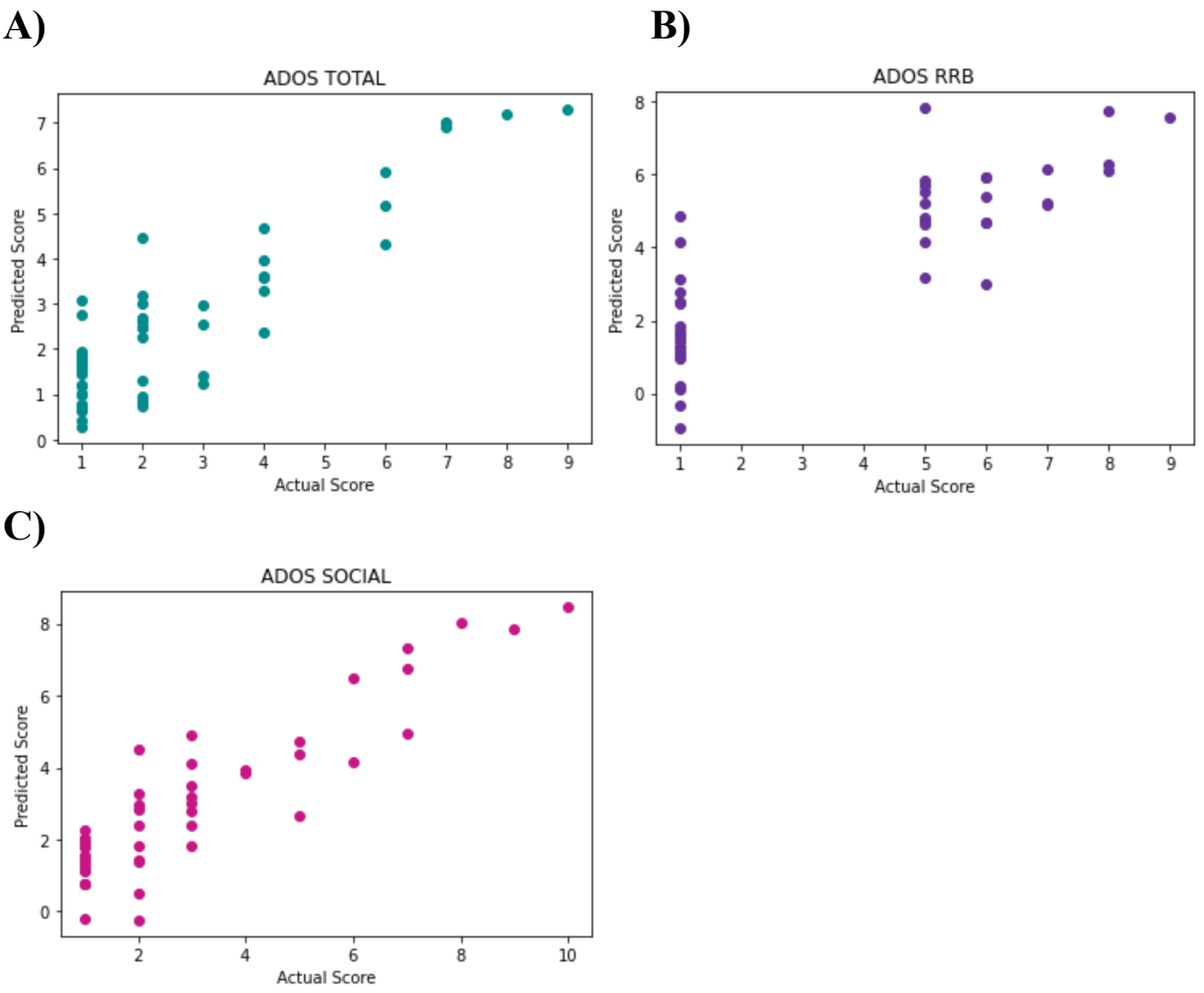


C)



*<u>Figure 5</u> **(A-C)**: Correlation coefficients were calculated with 34 genes in the SFARI data, as determined by multivariate regression analysis. Seven out of 34 genes were significant across the three ADOS scores, ADOS Social had the highest amount of significant genes. The average of R^2 was 0.79, the highest of any previous data. However, the analyses executed on the SFARI data were not significant after correcting for multiple comparisons.*

R^2 values were higher and *P* values were lower than previous datasets (Figure 6): ADOS Total, $R^2 = 0.792$, $P = 0.107$; ADOS RRB, $R^2 = 0.749$, $P = 0.237$; and ADOS Social, $R^2 = 0.816$, $P = 0.059$. Percent variance explained encompassed 74.911% to 81.61%: ADOS Total = 79.23%, ADOS RRB = 74.91%, and ADOS Social = 81.61%. A high percent variance without statistically significant data connotes that there may be too many predictors. The high amount of predictors can cause the model to be unable to identify the specific contribution of the predictor variable to the outcome measure. The adjusted R^2 on this data averages to be 0.33: ADOS Total, $R^2 = 0.351$; ADOS RRB, $R^2 = 0.216$; and ADOS Social, $R^2 = 0.425$. Considering that the adjusted R^2 attempts to correct for any overestimation from R^2, these results suggest that there may be too many predictors in this data.



<u>Figure 6</u> ***(A-C)****: In the SFARI data, a scatterplot was constructed that compared the actual ADOS score, acquired from participant data, to the predicted score, determined from the multivariate linear regression carried out on the matrix correlating gene and fMRI data.* R^2 *average was the highest and the P value was the lowest of any data. The high* R^2 *value indicates that 79% of variability from this data was explained by the model. Although each ADOS score produced promising results, only ADOS Social was significant.*

IV. Discussion

This study aimed to determine the connection between genetic variability and functional connectivity variability. From my results, there was no significant correlation found. Several genes were correlated to subject functional connectivity in the ABIDE and SFARI data, but there were no significant genes across every ADOS score. In the ABIDE data, three genes were significant for two ADOS metrics. STX1A and MVP were significant for ADOS Total and Communication, while CDKL5 was significant in both ADOS Total and ADOS Social. Despite several correlated genes, this data generated low R^2 values and high *P* values. Percent variance explained remained less than 32%, implying that the actual ADOS scores are not well explained by the predicted scores.

The PCA performed on this data also produced non-significant results. For every principal component across every ADOS score, none were significant. The aim of applying a PCA to the data was to reduce the number of variables, 122 genes, and the dimensions used in the analysis. Although this did lessen the amount of variables, it ultimately displayed no significant findings. This indicates that the simplification of the PCA was not able to capture the complex relationship between the genes and fMRI.

Although the ANOVAs run on the SFARI data were not significant, it yielded more promising results. RABEP2 was the only gene correlated to more than one ADOS score, in ADOS Total and ADOS Social. Out of each metric, ADOS Social had the most genes correlated and was statistically significant. R^2 values were considerably higher in this data, averaging 0.79. As was the percent variance explained, which had a mean of 79%. Although the prior data was not significant, the results of ADOS Social in this data demonstrated that the correlation between gene expression and functional connectivity predicts individual variation in behavioral symptoms.

These results emphasize the importance of further data collection through additional research into the genes found to be correlated from this evidence. Narrowing down each gene's role in ASD through their connection to functional connectivity could help better understand these individual differences. Matching gene expression and functional connectivity within the same participant would produce more accurate and potentially more significant results. Combined with a selective search of specific genes, this would eliminate the problem of heterogeneity. Using the same subjects' functional connectivity data and gene expression of the aforementioned genes would generate improved findings, proving that additional data collection must be made.

These findings have produced a select few genes that may be important in the link to functional connectivity variability. Future research can expand on these results by investigating the interplay between genetics and the brain through these genes. Specially focusing on PRKAB2, CDKL5, STX1A, SYN2, P2RX6, MVP, CDIPT, YPEL3, CLDN5, WBSCR22, KCTD13, ZDHHC8, EN2, NIPA2, BCL9, RABEP2, SH2B1, ATP2A1, PPP4C, RABEP2, SPN, and DOC2A. Applying these genes to a larger sample size, such as 3000, would be a better, more targeted approach to establishing their function in ASD. I employed a sample size of 699 in the ABIDE data and 121 in the SFARI data, which is far too small to answer my research question fully. Further analyses carried out on these genes should utilize a much larger sample size to better grasp the relationship between genetics and functional connectivity variability.

Previous research has uncovered connections between these genes and ASD. Two studies from Nakamura et al report that STX1A has elevated expression in ASD. Their first article from 2008 compared subjects with high functioning ASD and TD controls, discovering that those with ASD had a greater expression of STX1A. They imply that STX1A plays a role in the disruption

of serotonin synthesis during an early childhood stage of neurodevelopment. In 2011, these same researchers conducted a replication study further investigating STX1A in ASD. Shifting to post-mortem brains, they determined that there was decreased expression of STX1A in the anterior cingulate gyrus (ACG) region in ASD brains compared to the control group. This work supports my conclusion of STX1A's involvement in ASD and the ACG, an area associated with emotion and behavior regulation. ADOS Communication measures emotional regulation, and this metric is one that I found STX1A to be significant in.

KCTD13 likewise has prior research supporting its involvement in ASD. The KCTD gene family has evidence advocating for its relationship with various psychiatric disorders. KCTD13 is implicated in ASD and schizophrenia, while KCTD12 is linked to bipolar disorder and KCTD17 in movement disorders (Teng et al, 2019). KCTD13 is also a major proponent leading to phenotypes associated with the 16p11.2 CNV that contributes to the development of ASD (Golzio et al, 2012). This specific CNV is highly suspected to be unique to ASD, opening up research in this area. Dysregulation of the KCTD13-Cul3RhoA pathway in layer 4 of the inner cortical plate is a driver for the deletion of 16p11.2 CNV (Lin et al, 2015). Further research supports KCTD13 in 16p11.2 deletion (Madison et al, 2020). As I centered my genes around this CNV, several articles highlighted the importance of 16p11.2. Applying my results to these studies, future research targeting the 16p11.2 CNV could advance current knowledge of ASD.

EN2 is another gene that has a lot of backing evidence for its role in ASD. Choi et al (2014) operated a post-mortem analysis inspecting EN2 mRNA levels in ASD, discovering increased EN2 levels in ASD brains. A small study of ten patients with ASD identified a heterozygous variant within the EN2 gene in two patients (Hnoonual et al, 2016). Another article investigating ASD in the Chinese Han population linked EN2 to the predisposition of the

disorder (Wang et al, 2008). These papers examining human models of ASD demonstrate a clear connection between EN2 and ASD. Several knockout mice studies have further cemented EN2's involvement in ASD. EN2 knockout mice display neurobehavioral and neurochemical alterations similar to humans with ASD, seen through impairments in spatial learning, memory, and locomotor activity tasks (Cheh et al, 2006). These mice also show deficits in social interactions, through lack of socialization as an adult, as well as immobility in tests that measure prepulse inhibition and grip strength (Breilmaier et al, 2012). Although these human and rodent studies implicate EN2 in all aspects of ASD, the impairments in motor activity align with my findings of EN2's significance in ADOS RRB. Future investigation into RRB and EN2 can demystify this component of ASD.

Prior work has illuminated the potential of STX1A, KCTD13, and EN2 in ASD. Further research using gene expression maps of these genes with a larger sample size could unveil the secrets behind heterogeneity. The connection of genetics and functional connectivity with the genes now known to be implicated would better elucidate this problem of individual differences. These findings will be critical because they can aid in diagnosis, predict future outcomes, and assist in treatment. Discovering a genetic and functional biomarker pattern displayed in ASD would considerably facilitate this goal. Previous work, such as that of Silva et al (2022) and Li et al (2012), has looked into uncovering a genetic or functional biomarker but has rarely explored the relationship between the two. Revealing this connection is one step closer to discovering how the brain mediates genetics and behavior. This will highlight the underlying mechanisms involved and provide improved treatment from it. Ultimately, these results will help better understand ASD heterogeneity by connecting another facet of ASD to previously established findings.

References

Abid, A., & Zou, J. (2019). Contrastive Variational Autoencoder Enhances Salient Features.

Aglinskas, A., Hartshorne, J., & Anzellotti, S. (2022). Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science, 376*(6597), 1070-1074.

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

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101.

Chaste, P., & Leboyer, M. (2012). Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in clinical neuroscience*, *14*(3), 281–292.

Cheh, M. A., Millonig, J. H., Roselli, L. M., Ming, X., Kamdar, S., & Wagner, G. C. (2006). En2 knockout mice display neurobehavioral and neurochemical alterations relevant to autism spectrum disorder. *Brain research*, *1116*(1), 166–176.

Choi, J., Ababon, M. R., Soliman, M., Matteson, P. G., & Millonig, J. H. (2014). Autism associated gene, engrailed2, and flanking gene levels are altered in post-mortem cerebellum. *PloS one*, *9*(2), e87208.

Dubois, J., Galdi, P., Han, Y., Paul, L. K., & Adolphs, R. (2018). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality neuroscience*, *1*, e6.

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *373*(1756), 20170284.

Drysdale, A. T., Grosenick, L., Downar, J., & Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*, *23*(1), 28–38.

Esteban, O., Markiewicz, C.J., Blair, R.W. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* 16, 111–116 (2019).

Georgiades, S., Szatmari, P., & Boyle, M. (2013). Importance of studying heterogeneity in autism. *Neuropsychiatry, 3*(2), 123-125.

Golzio, C., Willer, J., Talkowski, M. E., & Katsanis, N. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, *485*(7398), 363–367.

Harlalka, V., Bapi, R. S., Vinod, P. K., & Roy, D. (2019). Atypical Flexibility in Dynamic Functional Connectivity Quantifies the Severity in Autism Spectrum Disorder. *Frontiers in human neuroscience*, *13*, 6.

Heuvel, M., & Pol, H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519-534.

Hnoonual, A., Sripo, T., & Limprasert, P. (2016). Whole-exome sequencing identifies a novel heterozygous missense variant of the EN2 gene in two unrelated patients with autism spectrum disorder. *Psychiatric genetics*, *26*(6), 297–301.

Leno, V., Chandler, S., White, P., Pickles, A., Baird, G., Hobson, C., Smith, A. B., Charman, T., Rubia, K., & Simonoff, E. (2018). Testing the specificity of executive functioning impairments in adolescents with ADHD, ODD/CD and ASD. *European child & adolescent psychiatry*, *27*(7), 899–908.

Lerer, E., Levi, S., & Salomon, S., (2008). Association between the oxytocin receptor (OXTR) gene and autism: relationship to Vineland Adaptive Behavior Scales and cognition. *Mol Psychiatry* 13, 980–988.

Lever, J., Krzywinski, M. & Altman, N. Principal component analysis. *Nat Methods* 14, 641–642 (2017).

Li, X., Zou, H., Brown, T., (2012). Genes associated with autism spectrum disorder. *Brain Research Bulletin,* 88(6), 543-552.

Lin, G. N., Corominas, R., Lemmens, & Iakoucheva, L. M. (2015). Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron*, *85*(4), 742–754.

Lombardo, M.V., Lai, MC. & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol Psychiatry* 24, 1435–1450.

Ma, D., Salyakina, D., Jaworski, J. M., Konidari (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Annals of human genetics*, *73*(Pt 3), 263–273.

Madison, J. M., Duong, K., Vieux, E. F., & Cottrell, J. R. (2020). Regulation of purine metabolism connects KCTD13 to a metabolic disorder with autistic features. *iScience*, *24*(1), 101935.

Martino, A., Yan, C. G., Li, Q., Denio, E., Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, *19*(6), 659–667.

Nakamura, K., Anitha, A., Yamada, K., Tsujii, M., & Mori, N. (2008). Genetic and expression analyses reveal elevated expression of syntaxin 1A (STX1A) in high functioning autism. *The international journal of neuropsychopharmacology*, *11*(8), 1073–1084.

Nakamura, K., Iwata, Y., Anitha, A., & Mori, N. (2011). Replication study of Japanese cohorts supports the role of STX1A in autism susceptibility. *Progress in neuro-psychopharmacology & biological psychiatry*, *35*(2), 454–458.

Quartier, A., Courraud, J., Thi Ha, T., & Piton, A., (2019). Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Human mutation*, *40*(11), 2021–2032.

Sie, J. H., Chen, Y. H., Shiau, Y. H., & Chu, W. C. (2019). Gender- and Age-Specific Differences in Resting-State Functional Connectivity of the Central Autonomic Network in Adulthood. *Frontiers in human neuroscience*, *13*, 369.

Silva, A., Ehrhart, F., Ulfarsson, M., (2022). Neuroimaging Findings in Neurodevelopmental Copy Number Variants: Identifying Molecular Pathways to Convergent Phenotypes. *Biological Psychiatry* 92, 341-361.

Simons Vip Consortium (2012). Simons Variation in Individuals Project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron*, *73*(6), 1063–1067.

Skaar, D., Shao, Y., Haines, J., (2005). Analysis of the RELN gene as a genetic risk factor for autism. *Mol Psychiatry* 10, 563–571.

Tecoulesco, L., Fein, D., & Naigles, L. R. (2021). What categorical induction variability reveals about typical and atypical development. *Journal of child language*, *48*(3), 515–540.

Teng, X., Aouacheria, A., Lionnard, L., Metz, K. A., Soane, L., Kamiya, A., & Hardwick, J. M. (2019). KCTD: A new gene family involved in neurodevelopmental and neuropsychiatric disorders. *CNS neuroscience & therapeutics*, *25*(7), 887–902.

Wang, L., Jia, M., Yue, W., & Zhang, D. (2008). Association of the ENGRAILED 2 (EN2) gene with autism in Chinese Han population. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, *147B*(4), 434–438.

Xie, W., Peng, C. K., Shen, J., Lin, C. P., Tsai, S. J., Wang, S., Chu, Q., & Yang, A. C. (2020). Age-related changes in the association of resting-state fMRI signal variability and global functional connectivity in non-demented healthy people. *Psychiatry research*, *291*, 113257.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*(3), 1125–1165.