UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Genetics of Language and Social Behavior in Autism

A thesis submitted in partial satisfaction of the requirements for the degree

Master of Science

in

Biology

by

Mohammad Reza Khorsand

Committee in charge:

Professor Karen Pierce, Chair
Professor Ralph Greenspan, Co-Chair
Professor Eric Courchesne
Professor Nicholas Spitzer

2015

The thesis of Mohammad Reza Khorsand is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2015

Dedication

I dedicate this thesis to my father, Mohammad Khorsand, and Tianci Liu.

Thank you for your love and support.

Be like the promontory against which the waves continually break,

but it stands firm and tames the fury of the water around it.

-Marcus Aurelius

Table of Contents

List of Figures

List of Tables

Acknowledgments

Dr. Pierce & Dr. Courchesne: Thank you for providing me the opportunity to participate in your lab as a graduate student and guiding me throughout my work. You both challenged me to think critically about my research and provided feedback that resulted in this body of work.

Dr. Pramparo: Thank you for helping me learn the different programs necessary for CNV and DGE analyses. You were very patient, and I appreciate that you took time to analyze and explain topics I did not understand.

UCSD Autism Center of Excellence Staff: This body of work, and future ones like it, would not be possible without every member of UCSD ACE. Thank you to everyone—past, present, and future—who is doing their best to help in advancing the understanding of autism.

UCSD Autism Center of Excellence Research Assistants: As a past research assistant, I know that it can be difficult to balance being a full-time student in addition to the 10-12 hour/week commitment as a UCSD ACE research assistant. You have done your part in providing a comfortable testing environment for the participants and their families. Furthermore, another thank you for those of you who went onto other projects that helped the scientists at UCSD ACE.

Study participants and their families: Without your participation, none of this work would be possible.

ABSTRACT OF THE THESIS

The Genetics of Language and Social Behavior in Autism

Mohammad Reza Khorsand

Master of Science in Biology

University of California, San Diego, 2015

Professor Karen Pierce, Chair

Autism, a neurodevelopmental disorder marked with difficulties in one's language ability and social behavior, is currently diagnosed through behavioral analyses diagnostics such as the Autism Diagnostic Observations Schedule. While these types of diagnostics have been reported to be effective, these tests are likely to miss children whom do not demonstrate the stereotypical behaviors of a typical autistic individual like restricted and repetitive behaviors. These children will appear to be typically developing until a certain point in their lives, which follow a decline or stunted development of skills. Given current research regarding the

genetics of autism, I investigated whether genetic variation in autistic children compared to typically developing children could be further stratified to an autistic child's language ability and social behavior. Analyzing both copy number variations and gene expression in autistic children and typically developing children, my results support previous studies in detecting genetic differences (variation) between the two groups. Moreover, my findings suggest that there may be group differences in biological networks related to copy number variations and differentially expressed genes in autistic children with poor versus good language ability and poor versus good social behavior. Genetic variation in children with poor language ability and poor social behavior is closely associated with biological networks related to neuronal development, cytoskeleton, and cell adhesion. In children with better language ability and better social behavior, there is a larger presence of networks associated with immune response and inflammation in addition to the biological networks found in poor language and poor social behavior groups. These findings demonstrate the potential of a biologically based marker for earlier detection of autism risk as well as an indicator for areas of difficulty in the child’s development.

I.

Introduction

A Description of Autism

Autism, a neurodevelopmental disorder typically associated with deficits in language ability and abnormal social behavior, affects approximately 1% of the United States population[1]. Dr. Leo Kanner first described autism in his 1943 publication *Autistic Disturbances of Affective Contact* where he described two key factors about this disorder. Firstly, he demonstrated the presence of autistic language and social behavior characteristics beginning at one's early childhood. Secondly, he described autism as a distinct disorder as opposed to being a variant of intellectual disability, deafness, or schizophrenia. Many of the language characteristics described by Kanner reflect a general trend of a lack or delay in developed language and regular echolalia, an immediate or delayed repetition of vocalization originally made by another person[2]. Kanner described the social behavior characteristics of autistic children as being content when left alone, having little to no regard for the presence of other individuals, and displaying a lack of initiating social interactions e.g. a child will not tell his or her parent about an event unless directly asked[2]. Kanner also described the characteristics of the parents of his case studies. He commented on features such as a parent's late development in speaking, one's rejection of social interaction while working, and another parent's lack of interest in people[2], possibly demonstrating consideration of the heritability of this disorder.

Diagnosing Autism

Autism is currently diagnosed through behavioral analyses such as the Autism Diagnostic Observations Schedule (ADOS)[3], which has measures for assessing social behavior, and language ability and cognitive ability is often assessed through diagnostics like the Mullen Scales of Early Learning (MSEL). However, the specificity and sensitivity of these behavioral analyses during the toddler period (between one and three-years old) can be low. Therefore, researchers are trying to develop better prognostic and diagnostic methods to detect risk for ASD by the third year of age or earlier. One example of an early detection method is Dr. Karen Pierce's One-Year Well-Baby Check-Up Approach, which assesses a child's risk for any developmental delay (e.g. autism, intellectual disability, language delay) at a 12-month doctor's visit[4]. Pierce's findings indicate that her approach can be used for identifying children who are at risk for autism, which allows families to pursue early intervention. In this report, 32 out of 10,479 infants were identified with autism[4]. Yet, Pierce's approach relies on a child's behavioral development. A recent study by Maenner and colleagues demonstrated that a child's age of autism diagnosis may be affected by the child's behavioral development. For example, impairments in nonverbal-communication, repetitive or ritualistic behaviors, and fixed routines were associated with an early autism identification. In contrast, impairments in conversational ability, peer relationships, and unusual speech-patterns were associated with a later autism identification[5].

Potentially, a detectable and reliable autism-risk biological signature (biomarker) can remedy this issue.

A Review of Language and Social Behavior Genetics

Using NCBI's gene database, the following search terms were used to collect a list of genes associated with language: language, language disorder, speech, speech disorder, dyslexia, specific language impairment, aphasia. For social behavior, the following search terms were used: social behavior, social cognition, sociability, social interaction.

A brief enrichment analysis using Thomson Reuters' MetaCore was performed to understand the types of biological networks that could be potentially be altered in language and social behavior in children with autism. For language genes, networks for development, signal transduction, and cell adhesion were prominent—though Metacore only scored two development networks with a false discovery rate less than 10%. For the NCBI-generated social behavior list, signal transduction, cell adhesion, neurophysiological process, and development were the most common. Of these networks, three were signal transduction, and the remaining networks had two networks with a false discovery rate less than 10%. Thus, I expect that copy number variations and differentially expressed genes in autistic individuals—when enriched through MetaCore—would represent these biological networks.

The Genetics of Autism

Many research teams claim to have identified novel candidate genes for autism. Some reported candidate genes fall under the neuroligin and neurexin gene families[6]. These two gene families are of particular interest due to their known role as synaptic genes. Previous studies based on functional magnetic resonance imaging have implicated dysregulations in neural development[7], specifically excessive neural connectivity in the frontal cortex while lacking connectivity to other brain regions[8]. Neuroligins and neurexins, among other genes, have been implicated with autism, though recent reviews indicate that further studies present both supporting and conflicting results[6,9]. Additionally, the effect of copy number variations (CNVs)—alterations in DNA that cause a cell to have abnormal expression of a specific gene or region of DNA—is also of particular interest. Genes related to the following biological processes: neurogenesis and neural development, DNA damage response, cell differentiation, and signaling pathways were observed in CNVs identified from postmortem tissue of autistic individuals and absent from control subjects[10]. A preliminary genetic expression profile generated from identified blood-based biomarkers successfully distinguished between autistic children, control children, language delayed children, and developmental delayed children[1]. Given that autism is likely a complex genetic disorder[11] displaying a spectrum of phenotypes, it would be worthwhile to identify a specific biomarker or a combination of biomarkers that are related to language ability and social behavior.

Overview of My Study

My research project on autism genetics will address two questions:

Aim 1: What are the genetic characteristics of individuals with autism? More specifically, will the findings of this study validate previous studies and observe genetic alterations that are specific to autistic individuals compared to controls? Previous studies identified differential gene expression and CNVs in individuals with autism compared to typically developing individuals[10]. This part of my research will be conducted using a program called CNVision[12] to simplify the process of detecting and analyzing copy number variations. My analysis will benefit from utilizing more recently updated databases such as the Database of Genomic Variants from The Centre for Applied Genomics[13]. This database specifically contains information related to common genetic variations, which can be used to identify rare or novel CNVs in my subjects. I will also observe the biological processes that are associated with the detected genes using MetaCore, a software suite used for functional analysis[14].

Aim 2: What trends can be observed between autism-related phenotypes of language ability and social behavior and genetic variation? Current research indicates that there is a strong genetic basis for autism[11,15]. Therefore, I will observe if differential gene expression is associated with an individual's language ability or social behavior. I compared gene expression and CNVs in participants from studies at the UCSD Autism Center of Excellence to their scores on the MSEL

and ADOS, which serve as quantifiable measures of language ability and social behavior respectively.

With information from Aims 1 and 2, clinicians and families with a child who is at-risk for autism could plan therapy sessions that stress specific developmental delays e.g. a child who has altered expression of genes related to deficits in language ability would spend more time with a speech therapist. It is my hope that my research can assist in developing a genetic profile or genetic test that could be used by clinicians to assess an individual's risk for specific autistic characteristics as well as overall autism-risk. As discussed by Pierce and colleagues, The One-Year Well-Baby Check-Up Approach may fail to detect children who have late-onset or regressive forms of autism[5]. Presumably, a diagnostic tool that relies on a genetic biomarker will be effective without relying on a child's behavioral development.

II.

Methods

Subject Collection

Subjects in this study were participants of other studies conducted at UCSD Autism Center of Excellence. Therefore, study participant selection and blood sample acquisition and analysis are as previously described[1,10,16].

CNV Analysis

As previously described[10], a total of 210 male subjects (cases=139; controls=71) were genotyped on the Illumina 660W-Quad BeadChip. SNP calls were performed using Illumina's GenomeStudio application. SNP calls were exported to FinalReport format and CNVs were called using CNVision[17], a Perl-based CNV detection pipeline, which runs three CNV detection algorithms (Gnosis[17], QuantiSNP[18], and PennCNV[19]). Following the CNVision pipeline, CNVs used for further analysis were filtered by the following parameters: "good sample" according to CNVision's quality check, >5 SNPs in the CNV, >5Kbp CNV size, and one of the following: 50% overlap in three or two algorithms. Furthermore, in order to determine the type of biological networks that are influenced by CNVs in autistic subjects, rare and common CNVs were analyzed together.

After filtering CNVision', PennCNV was used to identify genes contained within and neighboring the called CNVs. The PennCNV command used for gene identification is the following:

```
perl ./scan_region.pl CNVfile_of_interest.txt refGene.txt -refgene -reflink refLink.txt -
expandleft 5m | ./scan_region.pl CNVfile_of_interest.txt refGene.txt -refgene -reflink
refLink.txt -expandright 5m > CNVfile_of_interest_expand.txt
```

MetaCore GeneGo from Thomas Reuters was used to assess the biological networks associated with the genes contained within the detected CNVs. Gene lists were obtained from the PennCNV output files.

Gene Expression Analysis

As previously described[1], a total of 270 male subjects' (cases=122; control=148) RNA was processed on Illumina HumanHT-12 v4 Expression BeadChip microarray. Raw data from Illumina BeadStudio was log2 transformed and normalized by the lumi package contained in the R Bioconductor package[20].

Differential gene expression analyses were performed using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools Development Team. In order to obtain higher quality data, several quality checks were performed in the differential gene expression analysis. Initially, probes were excluded from the analysis if they were not associated to a known gene or had >50% percent missing or log intensity variation p-value > 0.05. Following probe filtering, standard class comparisons were performed for gene lists by setting the significance threshold of univariate tests to 0.05 and running 10,000 univariate permutation tests. Lastly, differentially expressed genes were determined to be significant if FDR <10%.

The filtered gene lists were then analyzed using MetaCore GeneGo to identify the biological networks that were disrupted between the classes.

Specifications of the system and dependent software used for CNVision and BRB-ArrayTools

CPU: Intel Xeon 1230 v3 (Base frequency: 3.3 GHz; Max Turbo Frequency: 3.7 GHz, 4 cores; 8 threads); RAM: Corsair Vengeance 32GB (4X8GB) DDR3-1600; Operating Systems: Windows 8.1 Pro 64-bit and Windows 7 Home Premium 64-bit; Perl Version for CNVision: Perl-5.8.8 32-bit; Excel for BRB-Array Tools: Microsoft Office 2013 64-bit

III.

Results

CNV Subject Profile

For this analysis, 139 autistic and 71 controls were analyzed using the Illumina 660W Beadchip platform. After CNV filtering, autistic males were then placed into poor and good subgroups for language ability and social behavior (Table 1, 2).

Table 1. Language Subgroups used for CNV analysis comparison. Receptive Language (RL) and Expressive Language (EL) from the MSEL were used to classify subjects. Those individuals with <40 in both RL and EL were grouped into "poor". Any subject who have >40 in either RL or EL fell into "good".

ASD Language Category	N	RL Avg	RL StDev	EL Avg	EL StDev
Poor	71	23.55	9.85	23.25	10.54
Good	38	49.08	9.55	48.24	10.72

Table 2. Social Behavior Subgroups used for CNV analysis comparison. Communicative Social (COSO) total score from the ADOS and Social Domain Standard score from the Vineland were used to cluster subjects, by the K means squared approach, into "poor" and "good" subgroups.

ASD Social Behavior category	N	COSO Avg	COSO StDev	Vineland Soc domain Avg	Vineland Soc domain StDev
Poor	37	16.57	2.36	74.86	11.83
Good	41	10.10	2.15	82.49	10.48

Copy Number Variation Analysis:

Though some of the enriched networks overlap, Controls primarily display CNVs that are associated with cell adhesion and Cases demonstrate enrichment in cytoskeleton and development networks (Figure 1).

To assess the potential variation in genetics between low and good language and low and good social behavior phenotypes, subjects were placed into classes using a predetermined classifier for language and an experimental classifier for social behavior. With respect to language comparisons, CNVs in cases with low language ability demonstrated higher and more prevalent enrichments in development-neurogenesis and cytoskeleton, which is markedly different than cases with good language. These individuals demonstrated a high number of

immune response and inflammation networks (Figure 2). For social behavior comparisons, a trend similar to the language comparison is observed (Figure 3). No immune response or inflammation networks in the poor ASD social behavior group were found to have an FDR less than 10^{-2}. In contrast, approximately half of the enriched networks for the good ASD social behavior group were immune response or inflammation.

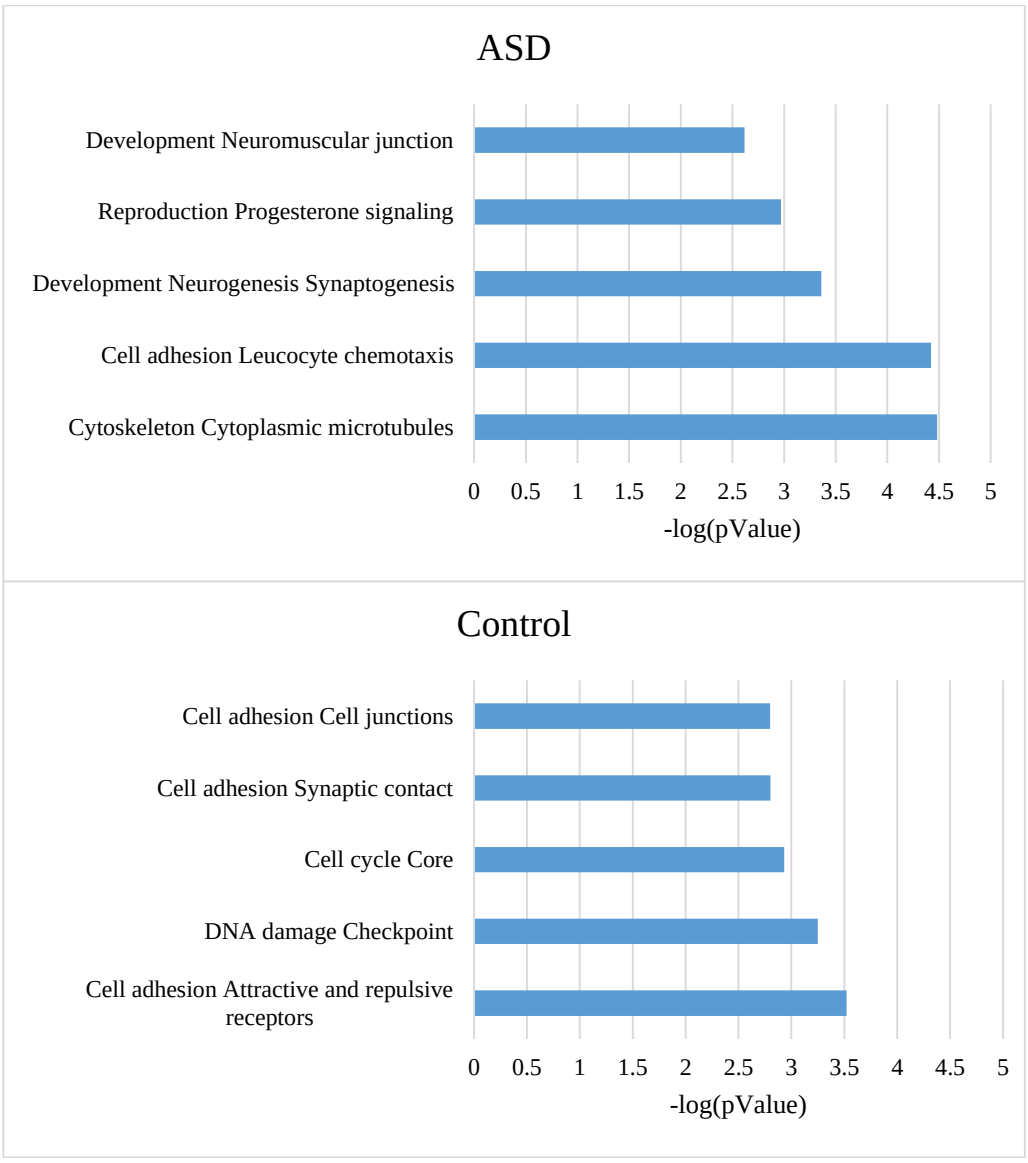

Figure 1. Top Five Biological Network Enrichment from Genes found in CNVs in Cases and Controls. See Supplemental Table 1 for more detailed list of enriched networks.

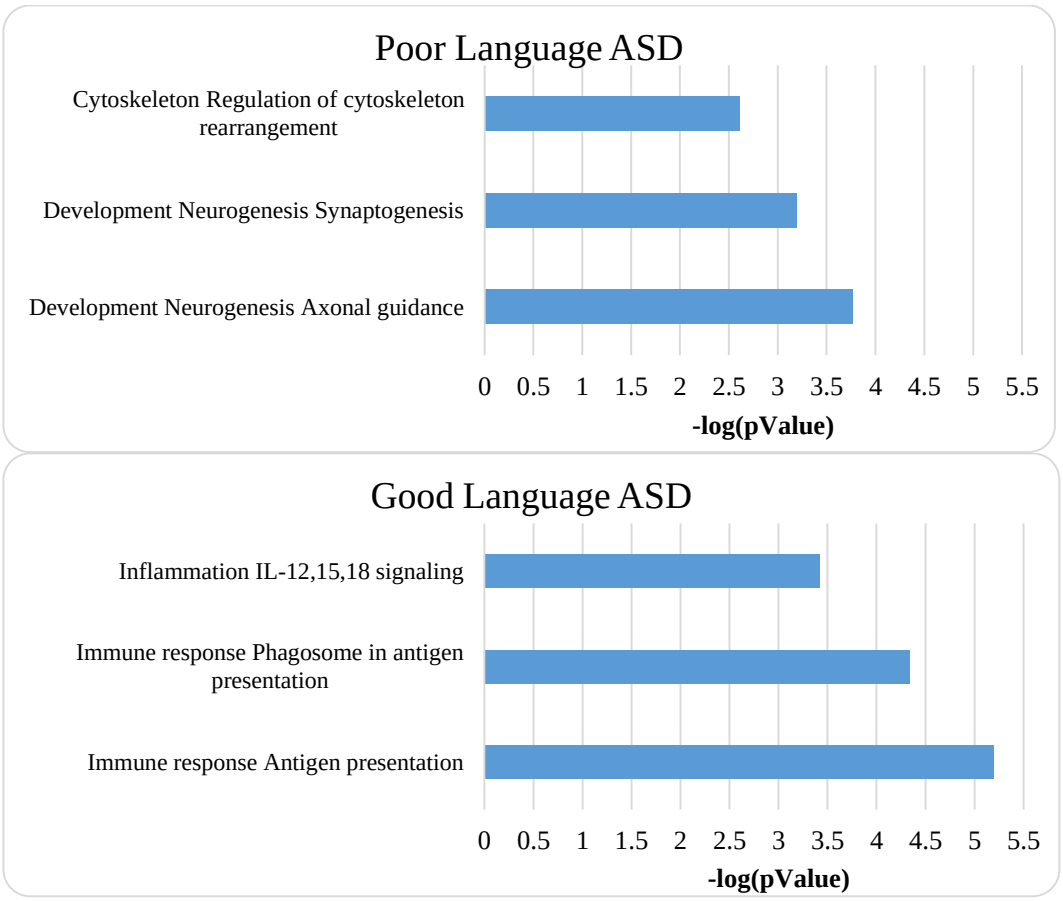

Figure 2. Top Three Biological Network Enrichment from Genes found in CNVs in Cases with Poor and Good Language Ability. See Supplemental Table 2 for more detailed list of enriched networks.

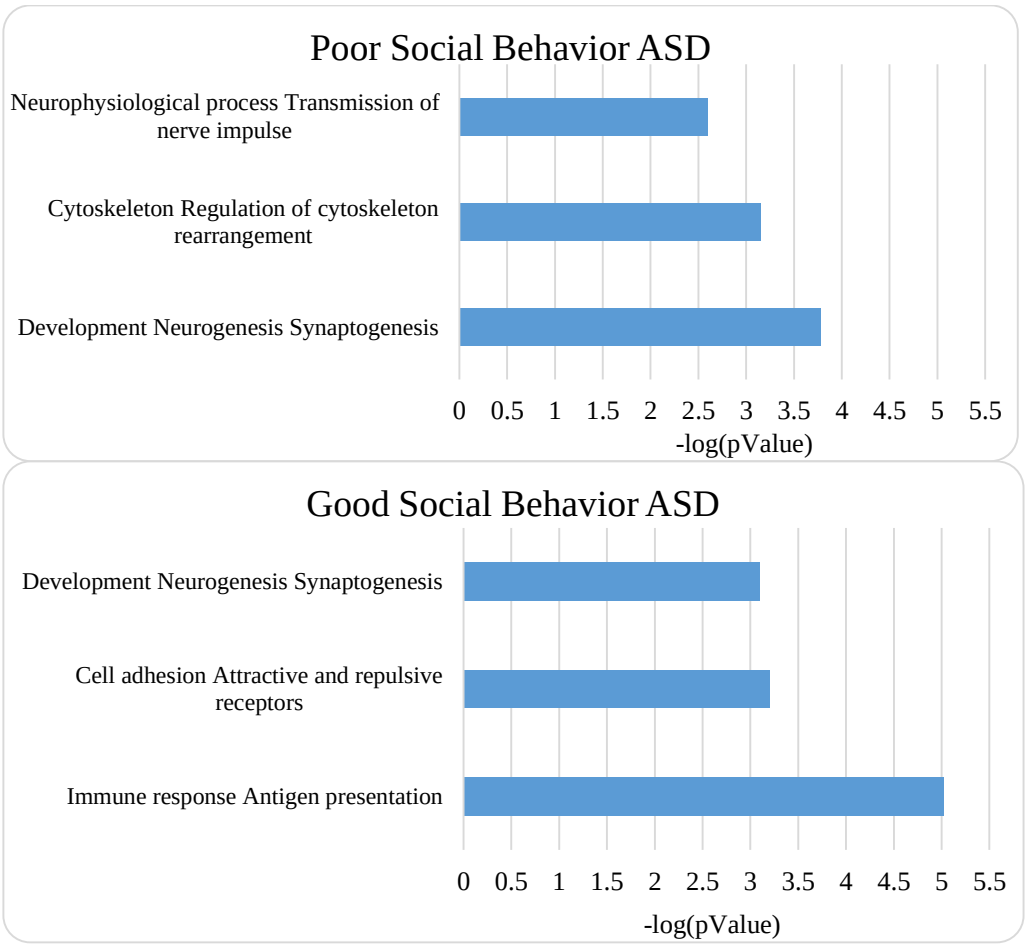

Figure 3. Top Three Biological Network Enrichment from Genes found in CNVs in Cases with Poor or Good Social Behavior. See Supplemental Table 3 for more detailed list of enriched networks.

Gene Expression Analysis

To extend the findings of the CNV analysis to one's gene expression, class comparisons were performed on data acquired from the Illumina HumanHT-12 v4 Expression BeadChip. Cases were separated into subgroups, as previously described. Class comparisons were performed and genes demonstrating FDR ≤ 10% were used for differential gene enrichment analysis.

Subject Profile

For this analysis, 121 cases and 148 controls (all males) were analyzed using the Illumina HT-12v4 Expression Beadchip platform. Class comparisons were conducted as described (Table 5) through BRB-ArrayTools. Autistic subjects were placed into "poor" and "good" subgroups as previously described (Table 3, 4).

Table 3. Language Subgroups used for Differential Gene Expression Class Comparison. Receptive Language (RL) and Expressive Language (EL) from the MSEL were used to classify subjects. Those individuals with <40 in both RL and EL were grouped into "poor". Any subject who had >40 in either RL or EL fell into "good".

ASD Language Category	N	RL Avg	RL StDev	EL Avg	EL StDev
Low	69	22.28	11.37	19.86	12.58
Good	52	44.38	9.35	47.08	9.43

Table 4. Social Behavior Subgroups used for Gene Expression Class Comparison. Communicative Social (COSO) total score from the ADOS and Social Domain Standard score from the Vineland were used to cluster subjects, by the K means squared approach, into "poor" and "good" subgroups.

ASD Social Behavior Category	N	COSO Avg	COSO StDev	Vineland Soc domain Avg	Vine Soc domain StDev
Poor	62	16.48	2.28	75.11	11.36
Good	49	11.22	2.34	87.61	12.84

Overall Analysis:

Class comparison of Cases and Controls demonstrated that 2957 genes were found to be differentially expressed (FDR≤10%). When run through a biological network enrichment program, fifty networks were found to meet an FDR less than

10%. Of these, cell cycle networks and cytoskeleton networks were most present (Table 6, 7).

Cases were then separated into a language ability category and a social behavior category. Differential gene expression analysis was conducted to compare gene expression between subgroups to controls.

Table 5. Class Comparisons and Number of Identified Differentially Expressed Genes (FDR≤10%)

Class Comparison	# DE Genes
Autism X Control	2957
Poor Language ASD X Control (PLAxC)	986
Good Language ASD X Control (GLAxC)	1056
ASD: Good Language X Low Language	0
Poor Social behavior ASD X Control (PSBAxC)	3415
Good Social behavior ASD X Control (GSBAxC)	129
ASD: Poor Social behavior X Good Social behavior	0

Table 6. Sample of Enriched Networks from Differentially Expressed Genes in Cases and Controls. In total, 50 networks were significantly enriched (FDR<10%)

Networks	Subnetworks	pValue	FDR
Proteolysis	Proteolysis in cell cycle and apoptosis	6.59E-09	1.05E-06
Cell adhesion	Platelet aggregation	3.26E-07	1.91E-05
Signal Transduction	Cholecystokinin signaling	3.61E-07	1.91E-05
Cytoskeleton	Regulation of cytoskeleton rearrangement	2.83E-06	1.13E-04
Immune response	Phagosome in antigen presentation	7.60E-06	2.42E-04
Proteolysis	Ubiquitin-proteasomal proteolysis	2.15E-05	5.58E-04
Cytoskeleton	Cytoplasmic microtubules	2.46E-05	5.58E-04

Table 7. Totals of Networks from Case and Control Differential Gene Expression Analysis. Only significant ($p<0.05$; FDR<10%) networks are represented.

Networks	Sum
Cell cycle	8
Cytoskeleton	6
Apoptosis	6
Reproduction	4
Development	4
Signal Transduction	4
Cell adhesion	4
DNA damage	3
Proteolysis	2
Proliferation	2
Transcription	2
Inflammation	2
Translation	1
Cardiac development	1
Protein folding	1
Immune response	1

Language Analysis:

Comparing both poor language and good language cases to controls demonstrated a considerable number of differentially expressed genes (Table 5). Dysregulated genes in the poor language to control comparison were enriched into six signal transduction networks and six cell adhesion networks. The same number of signal transduction networks were observed in the differentially expressed genes found in the good language and control class comparison. However, several differences were observed in the good language and control class comparison. Firstly, a higher number of total enriched networks was present—41 for poor language and control and 55 for good language and control. Secondly, a higher number of inflammation networks (nine compared to five in the PLxControl comparison), development networks (six compared to three in the PLxControl comparison), and immune response (five compared to four in the PLxControl comparison) (Figure 4, Table 7). Initially, a class comparison was run to detect differentially expressed genes in low language and good language cases. The findings of this analysis were that no genes were significantly ($p<0.05$ and FDR<10%) differentially expressed between these two groups.

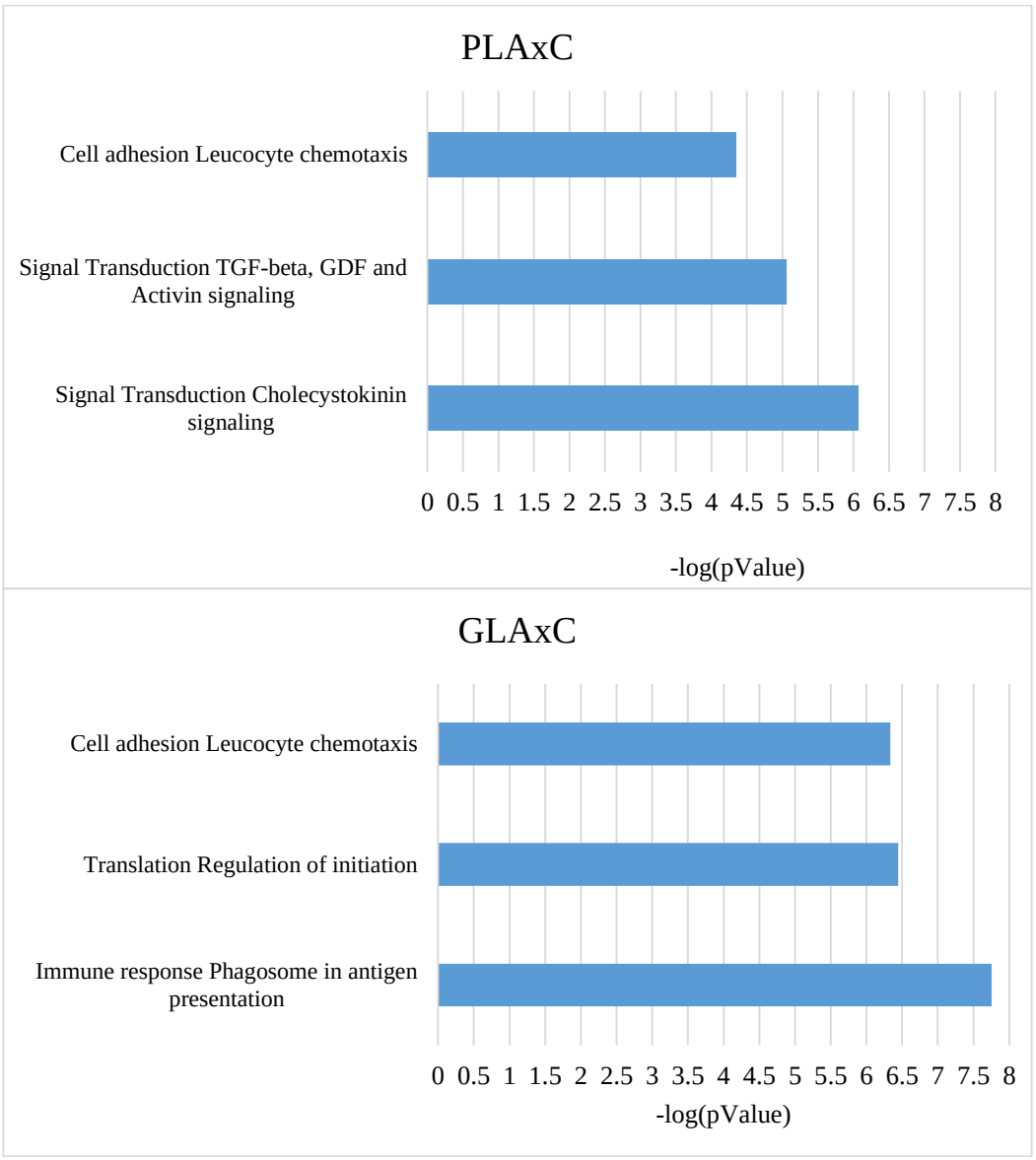

Figure 4. Top Three Biological Network Enrichment of Differentially Expressed Genes in PLAxC and GLAxC Class Comparisons. See Supplemental Table 4 for more detailed list of enriched networks.

Table 7. Top Five Most Prevalent Networks from Language Subgroup and Control Differential Gene Expression Analysis. Only significant (p<0.05; FDR<10%) networks are represented.

PLAxC		GLAxC	
Networks	Sum	Networks	Sum
Signal Transduction	6	Inflammation	9
Cell adhesion	6	Signal Transduction	6
Inflammation	5	Development	6
Reproduction	4	Cell adhesion	6
Apoptosis	4	Apoptosis	6

To compensate for the aforementioned issue, gene lists from the PLxControl and GLxControl analyses were compared through the *Unix comm command to categorize these lists into one of three categories: genes found in both PLxControl and GLxControl gene lists and genes only found in one of these lists (Table 8).

Table 8. Example of Output of the Unix comm Command Used to identify genes that Overlapped Both Case to Control Differential Gene Expression Analyses.

Unique: ASD good to control	Unique: ASD poor to control	Shared: genes that overlap in both lists
March7	March1	March3
AARS	March2	Sept6
ABAT	Sept15	ABCA1
ABCA13	AASDHPPT	ABCA7
ABI1	AATF	ABCC3
ACAD11	AATK	ABCG1

Genes that overlapped both analyses were predominately associated with the following networks: signal transduction, inflammation, cell adhesion, development, and immune response (Table 9).

Table 9. Network Enrichment of Genes that Overlapped the PLAxC and GLAxC Analyses. Only significant (p<0.05; FDR<10%) networks are represented.

Networks	Sum
Signal Transduction	9
Inflammation	9
Cell adhesion	7
Development	6
Immune response	5
Reproduction	4
Apoptosis	4
Cell cycle	3
Cardiac development	2
Cytoskeleton	2
Transcription	2
Neurophysiological process	2
Translation	1
Proliferation	1
Muscle contraction	1
Total	**58**

In the poor language and control class comparison, only one network, signal transduction, was observed with $p<0.05$ and FDR<10%. In contrast, genes only found in the GLxControl class comparison were associated with: inflammation, reproduction, apoptosis, cell adhesion, signal transduction, cytoskeleton, immune response, development, translation, and proteolysis (Table 10).

Table 10. Network Enrichment of Differentially Expressed Genes Only Found in One of the Two Case Language-Subgroup and Control Comparisons. Only significant ($p<0.05$; FDR<10%) networks are represented.

Poor Language and Controls		Good Language and Controls	
Network	**Sum**	**Networks**	**Sum**
Signal Transduction	1	Inflammation	3
Total	**1**	Reproduction	2
		Apoptosis	2
		Cell adhesion	2
		Signal Transduction	2
		Cytoskeleton	2
		Immune response	2
		Development	1
		Translation	1
		Proteolysis	1
		Total	**18**

Social Behavior Analysis:

The number of differentially expressed genes found from the poor social behavior cases and control class comparison is much higher than those found in the good social behavior cases and control class comparison (Table 5). Dysregulated genes from the poor social behavior and control analysis enriched into 74 significant networks. Among those networks, cytoskeleton, and cell adhesion were commonly present. Dysregulated genes from the good social behavior and control analysis enriched into six networks; inflammation, transcription-chromatin modification, and immune response networks were observed as the highest enriched networks for this group (Figure 5). Like in the language ability gene

expression analysis, a class comparison of poor social behavior and good social behavior cases did not yield significant differentially expressed genes.

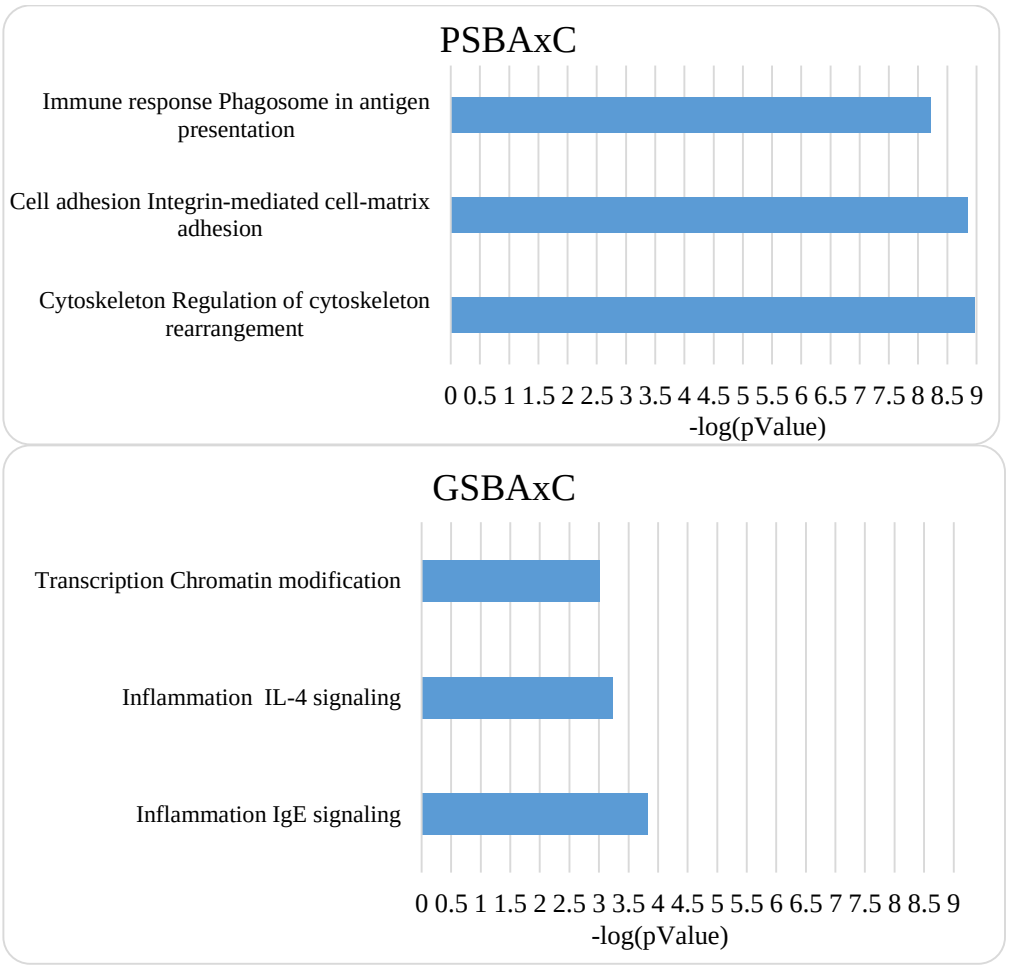

Figure 5. Top Three Biological Network Enrichment of Differentially Expressed Genes in PSBAxC and PSBAxC Class Comparisons. See Supplemental Table 5 for more detailed list of enriched networks.

Differential genes that were shared between the two social behavior-class comparisons were enriched into inflammation, muscle contraction, and signal transduction pathways (Table 11). In total, 25 networks were deemed significant. No significant networks were observed in genes unique to the GSBAxC class comparison, in contrast to the PSBAxC genes that strongly enriched for networks associated with cytoskeleton, cell adhesion, and proteolysis (Table 12, 13).

Table 11. Top Six Prevalent Network Enrichments from Genes that overlapped PSBAxC and GSBAxC Class Comparisons. Only significant ($p<0.05$; FDR<10%) networks are represented.

Row Labels	Sum
Inflammation	6
Signal Transduction	3
Apoptosis	3
Immune response	3
Reproduction	2
Development	2

Table 12. Sample of Network Enrichment of Unique Genes found in PSBAxC and GSBAxC Class Comparisons.

Poor Social Behavior Only				Good Social Behavior Only			
Networks	Subnetwork	pValue	FDR	Networks	Subnetwork	pValue	FDR
Cytoskeleton	Regulation of cytoskeleton rearrangement	3.29E-10	5.20E-08	Autophagy	Autophagy	6.00E-02	2.42E-01
Cell adhesion	Integrin-mediated cell-matrix adhesion	2.81E-09	2.22E-07	Translation	Selenium pathway	1.04E-01	2.42E-01
Proteolysis	Proteolysis in cell cycle and apoptosis	3.52E-08	1.81E-06	Protein folding	Folding in normal condition	1.26E-01	2.42E-01
Cell adhesion	Leucocyte chemotaxis	4.77E-08	1.81E-06	Signal transduction	Androgen receptor nuclear signaling	1.33E-01	2.42E-01
Immune response	Phagosome in antigen presentation	5.73E-08	1.81E-06	Transcription	Chromatin modification	1.35E-01	2.42E-01

Table 13. Network Enrichment of Differentially Expressed Genes Only Found in PSBAxC Class Comparison. GSBAxC networks not shown as none of the networks were deemed significant ($p<0.05$ FDR<10%)

Networks	Sum
Cell cycle	9
Cell adhesion	8
Signal Transduction	8
Inflammation	7
Apoptosis	5
Development	5
Cytoskeleton	5
Immune response	5
Transcription	3
Reproduction	3
Proteolysis	2
Translation	2
DNA damage	2
Proliferation	1
Total	**65**

Gene and Biological Network Validation

The findings of this study were compared to gene lists from the SFARI database[21]. Direct gene list-to-gene list comparisons were performed using the Unix comm command, as described in in the DE language analysis (Table 8). These comparisons demonstrated some overlap in specific genes. Of more interest is the comparison of gene networks between the list provided by SFARI's database and the genes identified in the present study.

Metacore enrichment was performed on the following iterations of the SFARI list: all genes, syndromic only, nonsyndromic only. The following discussion reflects only those networks with an FDR less than 10% (Table 14).

Table 14. Network Enrichment of Autism Genes from SFARI Database. Number of genes in each list is in parenthesis.

All Genes (380)		Syndromic (46)		Nonsyndromic (334)	
Networks	**Sum**	**Networks**	**Sum**	**Networks**	**Sum**
Development	7	Cardiac development	2	Development	7
Cardiac development	4	Cell adhesion	1	Signal transduction	4
Reproduction	4	Development	1	Reproduction	4
Cell adhesion	4	Transcription	1	Cell adhesion	4
Signal transduction	4			Cardiac development	3
Neurophysiological process	3			Neurophysiological process	3
Transcription	2			Transport	1
Apoptosis	1			Apoptosis	1
Transport	1				
Inflammation	1				
Muscle contraction	1				

In the three aforementioned lists, immune and inflammation networks were poorly enriched in favor of networks associated with cell adhesion, development, signal transduction, and transcription. These networks were seen in the overall autism gene enrichment for both CNV and differential gene expression analyses conducted in this study. However, when observing subgroups, enrichment of immune and inflammation networks were closely related to good performance on

the MSEL and the ADOS—proxies for good language and good social behavior respectively. The lack of enrichment for immune response and inflammation networks in the SFARI genes is most likely due to the relatively low number of genes associated with the two aforementioned networks compared to genes associated with development, cell adhesion, signal transduction, and others.

IV.

Discussion

The intent of this study was to characterize the genetics of subjects with autism with respect to their performance on the MESL and the ADOS, representing language ability and social behavior respectively. Through pathway analysis of genes found in the CNV analysis and differential gene expression analysis, this work demonstrates that it is possible that a biomarker may be attributed to the phenotypic presentations of language ability and social behavior in autistic children. These findings serve as an accessory to recent studies demonstrating potential biomarkers for at-risk males with autism[4], the neuroanatomical bases for good and poor language outcome in autism[22], and others in the same vein. Additionally, the findings of this study represent the need for researching the genetics of autism with respect to biological networks as opposed to looking for "the autism gene(s)".

1. Poor and Good Language Ability and Social Behavior Outcomes may be Associated with One's Genetics

In both analyses, a trend is present where children with poor language ability and poor social behavior were associated with biological network enrichments for development, cytoskeleton, cell adhesion, and neurophysiological processes. The good language and good social behavior groups demonstrated these networks to a lower degree. Yet, they also demonstrated a greater number of enrichments of immune response and inflammation networks, though the findings for the DGE good social behavior only enrichment (Figure 5) were underwhelming. This may be due to the clustering method used to differentiate between "poor" and "good" social

behavior groups; this can be observed in the high enrichments in the PSBAxC compared to the low number of enrichments in the GSBAxC (Figure 5, Supplementary Table 5). Furthermore, these findings should be treated as exploratory, as number of subjects in each analysis were not equal throughout all of the comparisons (e.g. overall autism and cases CNV comparison).

2. Comparison to Biological Network Enrichment of the NCIB-generated Language and Social Behavior Gene Lists

In this comparison, immune response and inflammation networks were more enriched and in higher quantity in good language and good social behavior subgroups when compared to their poor counterparts. Additionally, these two networks were not well enriched in the NCBI gene lists. This may represent the possibility that the genes that influence language ability and social behavior in other disorders may be the same ones as those that are influenced in autism. In the same vein, this could be further evidence for treating autism as a network-based disorder as opposed to a disorder based on a unique gene or specific set of genes. Using the binary indicators of “good” and “poor” for language ability and social behavior, there are up to four phenotype-permutations the same umbrella disorder known as autism—good/good, good/poor, poor/good, and poor/poor. The absence of highly enriched immune response and inflammation networks in poor language and poor social behavior autistic children and the decreased number of typically dysregulated networks in the good outcomes autistic children may represent the genetic variation that results in the heterogeneity of phenotypes in autism.

3. Combining Genetics, Clinical Diagnostics, and Biometric Measurements may Generate a Reliable Diagnostic for Early Autism At-Risk Assessment

The findings of this study demonstrate a possible association between a child's genetics and language ability and social behavior phenotypes. Furthermore, there are several studies that demonstrate other features may be associated to autism and, more specifically, autism subgroupings.

Previously, an eye-gaze study was conducted to determine if children with autism displayed any preference for geometrical shapes compared to controls and contrast groups[1]. This study demonstrated that there were two subgroupings of autism: those who were geometric responders (>69% time fixated on geometric patterns) and those who were social responders. This test demonstrated that the preference for geometric patterns may be a signature of early at-risk assessment for autism. However, the social responders would be missed by this test alone and would require further diagnostic evaluation through measures such as the MSEL and the ADOS. A more recent study demonstrated a new metric, neural imaging, as a marker for language ability in autistic children[2]. The aforementioned study demonstrated that clinical measurements in combination to left hemisphere superior temporal cortex activation outperformed all other classifiers that were tested, i.e. ADOS alone, clinical intake measures alone, and fMRI alone, for prediction of good versus poor language outcome.

Recently, a study was conducted to predict diagnosing autism using differential gene expression[16]. In this study, biological networks were identified in a discovery

set and then tested the predictors to a replication set. The study demonstrated a high number of ribosomal/translation genes and immune response genes associated with both sets of subjects and the signature performed well on both the discovery and replication sets. The aforementioned findings resemble the findings described in this study, where immune response, inflammation, and translation genes were associated with CNVs found in autistic subjects and differentially expressed genes between autistic and control subjects (using class comparison).

V.

Limitations

The CNV analysis is limited by the unequal number of subjects whom were analyzed. Therefore, comparing cases to controls for an overall observation of biological network enrichment can only be speculative. Additionally, the classification of good and poor social behavior was done by k means clustering and then validated using k-nearest neighbors. While the validation was good (~93% true positive and 95% true negative), the designation of "good" and "poor" are only representative of the sample that was analyzed. In effect, the full range of social behavior—starting at what could be called true good (lowest score) and true poor (highest score)—could not be represented in subjects who were 36 months old or below at time of ADOS and at-risk for autism. Subjects who are at-risk for autism will be seen as such on the ADOS, given that the diagnostic assess behaviors during a play-simulation with the examiner. At best, the classifiers used in this study for social behavior would be better designated as "poor" and "better-than-poor". A clustering method using more clinical measures or more scores from the ADOS, Vineland, etc. may generate more representative groupings for "poor" and "good" social behavior.

VI.

Conclusion

There appears to be a trend between the type of biological networks that are influenced by CNVs and DGE in autistic children and their language ability and social behavior outcomes. In this study, it was observed that children with good language and good social behavior had genetic variation more closely associated with immune response and inflammation networks which were not as highly represented to their poor language and poor social behavior counterparts. A diagnostic that combines genetics, clinical diagnostics, and biometric measurements may result in a specific and sensitive biomarker for autism.

References

1. Glatt S.J., Tsuang M.T., Winn M., Chandler S.D., Collins M., Lopez L., Weinfeld M., Carter C., Schork N., Pierce K., Courchesne E., Blood-Based Gene Expression Signatures of Infants and Toddlers With Autism. *Journal of the American Academy of Child & Adolescent Psychiatry* **51**, 934-944 (2012)

2. Kanner L. Autistic disturbances of affective contact. *Nervous Child* **2**, 217-250 (1943)

3. Jones R.M., Lord C., Diagnosing autism in neurobiological research studies. *Behavioural Brain Research* **251**, 113-124 (2013)

4. Pierce K., Carter C., Weinfeld M., Desmond J., Hazin R., Bjork R., Gallagher N,. Detecting, Studying, and Treating Autism Early: The One-Year Well-Baby Check-Up Approach. *Journal of Pediatrics* **159**, 458-465 (2011)

5. Maenner M.J., Schieve L.A., Rice C.E., Cunniff C., Giarelli E., Kirby R.S., Lee L., Nicholas J.S., Wingate M.S., Durkin M.S., Frequency and Pattern of Documented Diagnostic Features and the Age of Autism Identification. *Journal of the American Academy of Child 7 Adolescent Psychiatry* **52**, 401-422 (2013)

6. Perisco A.M., Napolioni V., Autism genetics, *Behavioural Brain Research* **251**, 95-112 (2013)

7. Dinstein I., Pierce K., Eyler L., Solso S., Malach R., Behrmann M., Courchesne E., Disrupted Neural Synchronization in Toddlers with Autism, *Neuron* **70**, 1218-1225 (2011)

8. Courchesne E., Redcay E., Morgan J.T., Kennedy D.P., Autism at the beginning: Microstructural and growth abnormalities underlying the cognitive and behavioral phenotype of autism, *Development and Psychopathology* **17**, 577-597 (2005)

9. Murdoch J.D., State M.W., Recent developments in the genetics of autism spectrum disorders, *Current Opinion in Genetics & Development* **23**, 310-315 (2013)

10. Chow M.L., Pramparo T., Winn M.E., Carter-Barnes C., Li, H., Weiss L., Fan J., Murray S., April C., Belinson H., Fu, X., Wynshaw-Boris A., Schork, N.J., Courchesne E., Age-Dependent Brain Gene Expression and Copy Number Anomalies in Autism Suggest Distinct Pathological Processes at Young Versus Mature Ages. PLoS Genetics **8**, 1-14 (2012)

11. Geschwind D.H., Genetics of autism spectrum disorders. *Trends in Cognitive Sciences* **15**, 409-416 (2011)

12. http://futo.cs.yale.edu/mw/index.php/CNVision

13. MacDonald J.R., Ziman R., Yuen R.K., Feuk L., Scherer S.W., The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research* **42,** 986-992 (2014)

14. http://thomsonreuters.com/metacore/

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

16. Pramparo T., Pierce K., Lombardo M., Carter Barnes C., Marinero S., Ahrens-Barbeau C., Murray S.S., Lopez L., Xu R., Courchesne E., Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices. *Jama Psychiatry* **72,** 386-394 (2015)

17. Sanders S.J., Ercan-Sencicek A.G., Hus V., Luo R., Murtha M.T.,… State M.W., Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* **70**, 863–885 (2011)

18. Colella S., Yau C., Taylor J.M., Mirza G., Butler H., Clouston P., Bassett A.S., Seller A., Holmes C.C., Ragoussis J., QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research* **35**, 2013-2025 (2007)

19. Wang K., Li M., Hadley D., Liu R., Glessner J., Grant S.F.A., Hakonarson H., Bucan M., PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* **17**, 1665-1647 (2007)

20. Du P., Kibbe W.A., Lin S.M., *lumi*: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548 (2008)

21. Basu S.N., Kollu R., Banerjee-Basu S., AutDB: a gene references resource for autism research. *Nucleic Acids Research* **37**, D832-D836 (2009)

22. Lombardo M.V., Pierce K., Eyler L.T., Carter Barnes C., Ahrens-Barbeau C., Solso S., Campbell K., Courchesne E., Different Functional Neural Substrates for Good and Poor Language Outcome in Autism. *Neuron* **86**, 567-577 (2015)

Appendix: Supplementary Tables

Supplemental Table 1. Biological Network Enrichment from Genes found in CNVs in Cases and Controls.

Cases				Control			
Network	**Subnetworks**	**pValue**	**FDR**	**Network**	**Subnetworks**	**pValue**	**FDR**
Cytoskeleton	Cytoplasmic microtubules	3.31E-05	2.99E-03	DNA damage	Checkpoint	4.14E-04	5.00E-02
Cell adhesion	Leucocyte chemotaxis	3.78E-05	2.99E-03	Cell adhesion	Synaptic contact	7.80E-04	5.00E-02
Development	Neurogenesis Synaptogenesis	4.35E-04	2.29E-02	Cell adhesion	Amyloid proteins	1.30E-03	5.00E-02
Reproduction	Progesterone signaling	1.07E-03	4.23E-02	Cell cycle	S phase	1.72E-03	5.00E-02
Development	Neuromuscular junction	2.42E-03	7.65E-02	Cell adhesion	Attractive and repulsive receptors	1.74E-03	5.00E-02
Cell cycle	S phase	6.41E-03	1.69E-01	Development	Neurogenesis Synaptogenesis	2.17E-03	5.21E-02
Cytoskeleton	Spindle microtubules	9.11E-03	2.01E-01	Reproduction	FSH-beta signaling pathway	2.92E-03	5.74E-02
Cytoskeleton	Regulation of cytoskeleton rearrangement	1.24E-02	2.01E-01	Cell adhesion	Cell junctions	3.19E-03	5.74E-02
Cell adhesion	Synaptic contact	1.32E-02	2.01E-01	Cell cycle	Core	3.98E-03	6.36E-02
Cell adhesion	Attractive and repulsive receptors	1.56E-02	2.01E-01	Neurophysiological process	Transmission of nerve impulse	7.50E-03	1.08E-01

Supplemental Table 2. Biological Network Enrichment from Genes found in CNVs in Cases with Poor and Good Language Ability.

Poor Language Cases				Good Language Cases			
Network	**Subnetwork**	**pValue**	**FDR**	**Network**	**Subnetwork**	**pValue**	**FDR**
Development	Neurogenesis Axonal guidance	1.72E-04	2.20E-02	Immune response	Antigen presentation	6.32E-06	7.77E-04
Development	Neurogenesis Synaptogenesis	6.41E-04	4.10E-02	Immune response	Phagosome in antigen presentation	4.54E-05	2.79E-03
Cytoskeleton	Regulation of cytoskeleton rearrangement	2.47E-03	6.72E-02	Inflammation	IL-12,15,18 signaling	3.81E-04	1.49E-02
Cell adhesion	Synaptic contact	2.58E-03	6.72E-02	Immune response	TCR signaling	4.84E-04	1.49E-02
Neurophysiological process	Transmission of nerve impulse	2.63E-03	6.72E-02	Immune response	Phagocytosis	2.34E-03	4.90E-02
Immune response	Antigen presentation	4.36E-03	9.30E-02	Cell adhesion	Platelet-endothelium-leucocyte interactions	2.39E-03	4.90E-02
				Immune response	BCR pathway	3.24E-03	5.69E-02
				Cell cycle	G1-S Growth factor regulation	4.51E-03	6.93E-02
				Inflammation	NK cell cytotoxicity	7.75E-03	9.17E-02
				Signal transduction	ERBB-family signaling	8.12E-03	9.17E-02
				Inflammation	Amphoterin signaling	8.20E-03	9.17E-02

Supplemental Table 3. Biological Network Enrichment from Genes found in CNVs in Cases with Poor and Good Social Behavior.

Poor Social Behavior Cases				Good Social Behavior Cases			
Network	**Subnetwork**	**pValue**	**FDR**	**Network**	**Subnetwork**	**pValue**	**FDR**
Development	Neurogenesis Synaptogenesis	1.63E-04	1.92E-02	Immune response	Antigen presentation	9.51E-06	1.30E-03
Cytoskeleton	Regulation of cytoskeleton rearrangement	7.02E-04	4.14E-02	Cell adhesion	Attractive and repulsive receptors	6.23E-04	3.66E-02
Neurophysiological process	Transmission of nerve impulse	2.49E-03	9.80E-02	Development	Neurogenesis Synaptogenesis	8.00E-04	3.66E-02
				Inflammation	MIF signaling	1.30E-03	4.45E-02
				Development	Regulation of angiogenesis	1.68E-03	4.62E-02
				Inflammation	IL-12,15,18 signaling	2.13E-03	4.79E-02
				Cytoskeleton	Regulation of cytoskeleton rearrangement	3.00E-03	4.79E-02
				Cell adhesion	Synaptic contact	3.13E-03	4.79E-02
				Inflammation	Protein C signaling	3.14E-03	4.79E-02
				Inflammation	Amphoterin signaling	5.40E-03	7.19E-02
				Development	Neurogenesis Axonal guidance	6.18E-03	7.19E-02
				Immune response	TCR signaling	6.30E-03	7.19E-02

Supplemental Table 4. Comparison of Enriched Networks from Dysregulated Genes from Case Poor Language and Control and Case Good Language and Control Class Comparisons.

Poor Language Cases				Good Language Cases			
Network	**Subnetwork**	**pValue**	**FDR**	**Network**	**Subnetwork**	**pValue**	**FDR**
Signal Transduction	Cholecystokinin signaling	8.54E-07	1.36E-04	Immune response	Phagosome in antigen presentation	1.76E-08	2.77E-06
Signal Transduction	TGF-beta, GDF and Activin signaling	8.77E-06	6.97E-04	Translation	Regulation of initiation	3.60E-07	2.42E-05
Cell adhesion	Leucocyte chemotaxis	4.50E-05	2.38E-03	Cell adhesion	Leucocyte chemotaxis	4.63E-07	2.42E-05
Cell adhesion	Platelet aggregation	1.25E-04	4.95E-03	Signal Transduction	Cholecystokinin signaling	1.34E-06	5.25E-05
Translation	Regulation of initiation	5.00E-04	1.33E-02	Cell cycle	G1-S Growth factor regulation	5.69E-06	1.79E-04
Cell adhesion	Integrin-mediated cell-matrix adhesion	5.79E-04	1.33E-02	Cell adhesion	Platelet aggregation	1.51E-05	3.96E-04
Apoptosis	Anti-Apoptosis mediated by external signals by Estrogen	5.85E-04	1.33E-02	Reproduction	Progesterone signaling	3.05E-05	6.85E-04
Cell adhesion	Integrin priming	8.69E-04	1.66E-02	Cell cycle	G1-S Interleukin regulation	7.93E-05	1.56E-03
Cytoskeleton	Regulation of cytoskeleton rearrangement	9.38E-04	1.66E-02	Cell adhesion	Integrin priming	1.32E-04	2.29E-03
Reproduction	FSH-beta signaling pathway	1.08E-03	1.71E-02	Apoptosis	Anti-apoptosis mediated by external signals via NF-kB	1.47E-04	2.30E-03

Supplemental Table 5. Comparison of Enriched Networks from Dysregulated Genes from PSBAxC and GSBAxC Class Comparisons.

Poor Social Behavior Cases				Good Social Behavior Cases			
Networks	**Subnetworks**	**pValue**	**FDR**	**Networks**	**Subnetwork**	**pValue**	**FDR**
Cytoskeleton	Regulation of cytoskeleton rearrangement	1.08E-09	1.10E-07	Inflammation	IgE signaling	1.51E-04	1.89E-02
Cell adhesion	Integrin-mediated cell-matrix adhesion	1.39E-09	1.10E-07	Inflammation	IL-4 signaling	5.85E-04	3.65E-02
Immune response	Phagosome in antigen presentation	5.98E-09	2.85E-07	Transcription	Chromatin modification	9.51E-04	3.96E-02
Proteolysis	Proteolysis in cell cycle and apoptosis	8.60E-09	2.85E-07	Muscle contraction	Relaxin signaling	1.48E-03	4.62E-02
Cell adhesion	Leucocyte chemotaxis	8.95E-09	2.85E-07	Immune response	Phagosome in antigen presentation	3.08E-03	7.54E-02
Signal Transduction	Cholecystokinin signaling	3.60E-08	9.54E-07	Signal Transduction	Cholecystokinin signaling	3.62E-03	7.54E-02
Cell adhesion	Platelet aggregation	1.95E-07	4.33E-06	Cell cycle	G2-M	7.49E-03	1.24E-01
Cytoskeleton	Actin filaments	2.18E-07	4.33E-06	Development	ERK5 in cell proliferation and neuronal survival	8.98E-03	1.24E-01
Proteolysis	Ubiquitin-proteasomal proteolysis	4.30E-07	7.59E-06	Immune response	Phagocytosis	1.02E-02	1.24E-01
				Development	Regulation of angiogenesis	1.04E-02	1.24E-01