Affected Sib-Pair Analyses Identify Signaling Networks Associated With Social Behavioral Deficits in Autism

Mehdi Pirooznia[1,2]*, Tejasvi Niranjan[3], Yun-Ching Chen[1], Ilker Tunc[1], Fernando S. Goes[2], Dimitrios Avramopoulos[2,3], James B. Potash[2], Richard L. Huganir[4], Peter P. Zandi[2,5] and Tao Wang[3]*

[1] Bioinformatics and Computational Biology Core Facility, National Heart Lung and Blood Institute, NIH, Bethesda, MD, United States, [2] Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [3] McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [4] Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [5] Department of Mental Health and Epidemiology, Johns Hopkins University School of Public Health, Baltimore, MD, United States

Autism spectrum disorders (ASDs) are characterized by deficits in three core behavioral domains: reciprocal social interactions, communication, and restricted interests and/ or repetitive behaviors. Several hundreds of risk genes for autism have been identified, however, it remains a challenge to associate these genes with specific core behavioral deficits. In multiplex autism families, affected sibs often show significant differences in severity of individual core phenotypes. We hypothesize that a higher mutation burden contributes to a larger difference in the severity of specific core phenotypes between affected sibs. We tested this hypothesis on social behavioral deficits in autism. We sequenced synaptome genes (n = 1,886) in affected male sib-pairs (n = 274) in families from the Autism Genetics Research Exchange (AGRE) and identified rare (MAF ≤ 1%) and predicted functional variants. We selected affected sib-pairs with a large (≥10; n = 92 pairs) or a small (≤4; n = 108 pairs) difference in total cumulative Autism Diagnostic Interview-Revised (ADI-R) social scores (SOCT_CS). We compared burdens of unshared variants present only in sibs with severe social deficits and found a higher burden in SOCT_CS≥10 compared to SOCT_CS ≤ 4 (SOCT_CS≥10: 705.1 ± 16.2; SOCT_CS ≤ 4, 668.3 ± 9.0; *p* = 0.025). Unshared SOCT_CS≥10 genes only in sibs with severe social deficits are significantly enriched in the SFARI gene set. Network analyses of these genes using InWeb_IM, molecular signatures database (MSigDB), and GeNetMeta identified enrichment for phosphoinositide 3-kinase (PI3K)-AKT-mammalian target of rapamycin (mTOR) (Enrichment Score [eScore] *p* value = 3.36E–07; n = 8 genes) and Nerve growth factor (NGF) (eScore *p* value = 8.94E–07; n = 9 genes) networks. These studies support a key role for these signaling networks in social behavioral deficits and present a novel approach to associate risk genes and signaling networks with core behavioral domains in autism.

Keywords: affected sibs, autism social behavior, network analysis, InWeb_IM, GeNetMeta, synaptome, PI3K-AKT-mTOR, NGF signaling

INTRODUCTION

Autism spectrum disorders (ASDs) are a group of heterogeneous neurodevelopmental disorders characterized by deficits in reciprocal social interactions, communication, and restricted interests and/or repetitive behaviors. ASDs are caused by a combination of environmental risk factors and genetic mutations (Devlin and Scherer, 2012; Huguet et al., 2013). Twin-, family-, and population-based studies indicate that genetic factors contribute to more than half of the risk of developing ASDs (Klei et al., 2012; Gaugler et al., 2014; Sandin et al., 2014). A recent meta-analysis on 6,413 twins including affected twins showed that heritability in families with an autistic patient is 64–91% (Tick et al., 2016). Common variants of small effect and rare variants of large effect could have a substantial impact on risk of developing autism and/or on severity of specific behavioral domains in autism (Huguet et al., 2013).

Genome-wide sequencing of large cohorts of autism patients and their families has generated considerable numbers of sequence variants in recent years (Michaelson et al., 2012; Lim et al., 2013; Yu et al., 2013). Case-control and family-based studies implicate risk alleles by identifying an association of common variants of small effect, a higher mutation load of rare variants of large effect, and/or presence of *de novo* variants in affected probands (De Rubeis and Buxbaum, 2015). Several hundreds of autism risk genes have been implicated using these approaches (Huguet et al., 2013). However, the majority of these studies treat autism as a "single disease" in a "case-control" study design rather than a spectrum of disorders with deficits in core behavioral domains. Diagnosis of ASDs are based on standard, semi-quantitative behavioral tests including Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994) to assess deficits in three core behavioral domains: reciprocal social interaction, communication, and restricted interests and/or repetitive behaviors. Although defects in all three domains are required to make a diagnosis, patients with ASDs often present with significant differences in the severity of specific behavioral domains. Furthermore, deficits in these behavioral domains are shared by other neuropsychiatric disorders. For example, abnormal social behaviors are a key feature of schizophrenia; increased repetitive behaviors or movements are a core phenotype in obsessive-compulsive disorder, and defects in speech and communication are seen in neurodevelopmental syndromes with different underlying genetic causes. Understanding shared mechanisms responsible for specific domains of psychiatric phenotypes in these disorders is crucial to development of medications and interventions for individualized care for patients with these disorders and identification of family members at increased risk for genetically influenced behavioral or psychiatric phenotypes. Due to extensive genetic heterogeneity and phenotypic variability

(Jeste and Geschwind, 2014; De Rubeis and Buxbaum, 2015), it remains a major challenge to associate genetic risk genes and networks with specific behavioral domains in autism.

It has long been noted that in multiplex families, affected sibs with autism often show a significant difference in the severity of one or more behavioral domains (Spiker et al., 2002). We hypothesize that genetic burdens of cumulative risk genes and signaling networks contribute to the differences in severity of specific domains in the affected sibs. A recent WGS study of a cohort of quartet autism families showed that in a large fraction (50–69%) of multiplex families, two affected sibs do not share the same rare penetrant risk alleles (Yuen et al., 2015). These affected sibs with discordant mutations tended to demonstrate more phenotypic variability as compared to those who shared the same risk variants. These results support a strong genetic determinant responsible for differences in severity of specific autism phenotypes between affected sibs.

Genetic studies of affected sib-pairs have been used previously to map genes for rare Mendelian disorders based on the principle of identical-by-descent (IBD) (Zhang and Risch, 1996; Liang et al., 2001; Perdry et al., 2012). It has been shown that an affected sib-pair design based on sharing pathogenic variants allows a tremendous gain of analysis power over a traditional case-control study design to implicate pathogenesis of rare variants (Xing et al., 2006; Sul et al., 2017). Discordant sib-pair designs have also been used in genetic linkage and association studies to increase power of analysis (Boehnke and Langefeld, 1998; Poznik et al., 2006). We explore a novel family-based strategy utilizing affected, phenotypic discordant sib-pairs to identify rare genetic variants of large effect contributing to specific domains in autism (Hu and Steinberg, 2009; Hu et al., 2011; Sacco et al., 2012; Veatch et al., 2014). We hypothesize that (Huguet et al., 2013) affected sibs with severe deficits in specific domains carry a larger burden of risk variants as compared to sibs with mild deficits in the same domain, (Devlin and Scherer, 2012) differences in severity of deficits in specific domains by ADI-R scores between affected sibs correlate with burden of cumulative risk variants, and (Gaugler et al., 2014) sets of autism risk genes associated with specific autism domain(s) in one family may be shared by a fraction of affected families in a study cohort.

In the current study, we sequenced the exons of synaptome genes (n = 1,886) in 274 pairs of affected male siblings from Autism Genetics Research Exchange (AGRE). We identified an excess burden of rare deleterious variants in cohorts of sibs showing a large versus small differences in severity of social deficits defined by total cumulative ADI-R social interaction score (SOCT_CS). We performed network-based analyses on these gene sets carrying the excess mutations and identified several neural signaling networks associated with social behavioral deficits in autism.

METHODS

Patients

We surveyed autism pedigrees in the AGRE repository and identified 274 pairs of male affected sibs in multiplex families (www.autismspeaks.org/agre). We selected male affected sibs

---

Abbreviations: AGRE, autism genetics research exchange; ADI-R, autism diagnostic interview-revised; eScore, enrichment score; SOCT_CS, total cumulative score for social interaction; PPI, protein interaction network; NGF, nerve growth factor; PI3K, phosphoinositide 3-kinase; AKT (PKB), AKT serine/threonine kinase (protein kinase B); TSC, tuberous sclerosis complex; mTOR, mammalian target of rapamycin; PTEN, phosphatase and tensin homolog; CN, communal nesting; MSigDB, molecular signatures database.

who have full behavioral evaluations including ADI-Rs and show significant differences in the severity of autism phenotypes, i.e., one sib presents as severe while the other sib presents mild phenotype as defined by the cumulative ADI-R scores. Three-generation pedigrees, DNA samples, developmental histories, and behavioral test scores including ADI-R and ADOS are obtained for all enrolled patients and most relatives in these families. An institutional review board at the Johns Hopkins University has approved this study.

Synaptome Analysis
Sequencing
We surveyed all published proteomics studies and publically available databases that focus on the synapse and identified genes that encode 1,886 synaptic proteins consisting of proteins found in the vesicles (N = 107), in the presynaptic membrane (N = 336), in the presynaptic active zone (N = 209) and in the post-synaptic density, as established in SynaptomeDB (http://metamoodics.org/SynaptomeDB/) (Pirooznia et al., 2012). We utilized an Agilent Sure-Select target enrichment kit to capture 6.7 Mb of targeted genomic sequence for human Synaptome and completed next-generation sequencing for 274 affected sib-pairs and 336 matched normal controls using HiSeq2000 at the high-throughput sequencing core at the Johns Hopkins University (Pirooznia et al., 2016).

Data Processing
Sequence reads were aligned to the human reference genome (UCSC hg19) using BWA aligner (Li and Durbin, 2009) allowing for two mismatches in the 30-base seed. Picard (http://picard.sourceforge.net/) was used to fix any mate pair mismatch and remove reads with identical outer mapping coordinates, which represent likely PCR artifacts. Target coverage for the Agilent Sure Select capture was assessed using Picard's HSmetrics utility. The Genome Analysis Toolkit (Mckenna et al., 2010) was used to generate SNV and small indel calls within the targeted regions. We performed variant calling using GATK's HaplotypeCaller followed by a Variant Recalibration step. SNV clusters, defined as greater than three SNVs per ten bases, and SNVs falling within a called indel region, were masked. Variant Call Format (VCF) files were converted to PLINK file format using VCF tools and custom scripts. PLINK was subsequently used to remove variants with >10% missing calls and variants in Hardy-Weinberg Disequilibrium ($p < 1 \times 10^{-6}$). Principal component analysis (PCA) of the case-control sample was performed using Eigenstrat to assess for potential population stratification and batch effects across the sequencing platforms using common sequenced variants (MAF > 0.05) pruned to be in approximate linkage equilibrium. We inspected the top axes of variation in each PCA component and removed three outlier individuals, with the remaining samples showing appropriate clustering consistent with a European-American sample.

Annotation
Identified variants, including single nucleotide variants (SNVs) and indels, were annotated with ANNOVAR (Wang et al., 2010) using Ensembl release version 63 as the reference assembly. ANNOVAR provides information on gene annotation, amino acid change annotation, dbSNP ids, 1000 Genomes Project allele frequencies, and NHLBI-Exome Sequencing Project (ESP) allele frequencies. For annotation of missense variants, we used SIFT and Polyphen-2 to identify variants of potentially damaging effect. We used default thresholds of SIFT (>0.95) and PolyPhen (>0.85) to classify a SNV as damaging. For indels, we included stopgain, stoploss, frameshift, and splicing insertions as damaging variants. Rare variants, defined as having a population frequency ≤1% were selected for gene burden analyses by using the European-American and ALL frequency estimates of the NHLBI-ESP, and both the European-American and ALL estimates from the 1000 Genomes April 2012 release to exclude variants with allele frequency >1% in any of these external datasets.

Pathway Analysis
We employed GeNets platform (Li et al., 2018) for network and pathway analyses to evaluate the connectivity of genes in our gene-set, based on a network of known susceptibility genes that are interconnected by protein–protein interaction (PPI) using InWeb. It builds "neighborhoods" of genes in a gene list that are more interconnected within the reference network by creating a general classifier to predict membership from networks in the InWeb PPI network. For any candidate gene, the classifier can assign a probability that it belongs to a pathway as defined by the candidate's architectural properties in the overall network. GeNets creates networks of PPIs using evidence of physical interaction from the InWeb database, which contains more than 420,000 high-confidence pair-wise interactions involving 12,793 proteins (Lage et al., 2007; Wagner et al., 2012; Li et al., 2017; Raj et al., 2018). It displays these interaction networks as community structures (also called modular sub-network structure). A module is a set of genes (called nodes) that are more connected to one another than they are to other groups of genes based on a probability score that is calculated based on network metrics using a machine learning algorithm (quack) trained on 853 curated molecular signatures database (MSigDB) pathways with the reference network; and using that same algorithm to "predict" other genes in the network that are not on our gene list, but may belong to the same pathway that is captured by our gene list. The 853 MSigDB gene sets are curated from 1,329 C2:CP gene sets in MSigDB, by calculating pairwise Jaccard index (Intersection over Union and the Jaccard similarity coefficient), and obtaining pathways with pairwise Jaccard index < = 0.5 (Li et al., 2018).

The sub-network analysis ranks genes and predicts candidates based on InWeb PPI patterns found in known pathways, highlights genes that are more connected to one another than they are to other genes in other modules, and segments them based on their similarity to known pathway gene sets. GeNets also employs a within-degree node-label permutation strategy to build random networks similar to the original network and generate empirical distributions to assess the statistical significance of PPI networks. In addition to InWeb PPI network analysis, GeNets performs gene set enrichment analysis on genes within the network. We performed this function on Molecular Signatures Database (MSigDB) curated Gene Sets (C2, containing pathway databases such KEGG, BioCarta, and Reactome) to test for enrichment of these pathways within the

network. The gene set enrichment analysis p value is generated based on a hypergeometric test. We used Bonferroni-corrected P < 0.05 to correct for multiple testing.

RESULTS

Patients

We identified 274 affected male sib-pairs from multiplex families recruited to Autism Genetic Resource Exchange (AGRE). The mean age difference was 2.5 years between the cohorts of younger and older sibs. General developmental milestones including ages of first word and first walk and cumulative ADI-R scores in the three core domains, e.g., reciprocal social interactions, communication, and repetitive behaviors, were comparable between the cohorts of younger and older sibs (**Supplementary Table 1**).

Variant Burdens of Sib-Pairs With Severe Versus Mild Deficits in Social Behaviors

Affected sibs (n = 274) were divided into two cohorts with either severe or mild phenotypes defined by total cumulative ADI-R score in three behavioral domains. No significant difference was found for age of diagnosis and general motor development including age of first walk between these two cohorts (**Table 1**). As expected, the total cumulative ADI-R social scores (SOCT_CS) showed a significant difference between these two cohorts (n = 274; SOCT_CS, severe, 24.77 ± 0.27; mild, 17.18 ± 0.43, mean ± SEM; t-test; p = 1.68E−42) (**Table 1**). A direct comparison of burdens for total cumulative rare (MAF ≤ 0.01), predicted deleterious variants identified no significant difference between these two cohorts (p = 0.96).

Variant Burdens of Sib-Pairs With Large Versus Small Differences in Deficits of Social Behaviors

We next identified subsets of affected sib-pairs with either a large (SOCT_CS≥10, n = 92 pairs; severe, 25.32 ± 0.28; mild, 13.67 ± 0.45; mean ± SEM; t-test; p = 7.18E−52) or a small difference (SOCT_CS ≤ 4; n = 108 pairs; severe, 24.60 ± 0.44; mild, 22.54 ± 0.47; mean ± SEM; t-test; p = 1.82E−03) in total cumulative ADI-R social interaction scores between individual sib-pairs (**Table 2** and **Supplementary Table 2**). We hypothesized that a larger burden of unshared variants would be seen in the severe only sibs in the SOCT_CS≥10 cohort as compared to that in the SOCT_CS ≤ 4 cohort (**Figure 1A**). We thus extracted variants

TABLE 2 | Affected Sib-Pairs with Large versus Small Difference in Social Behavioral Deficits.

	Large difference (SOCT_CS≥10)		Small difference (SOCT_CS ≤ 4)	
	Age (year)	SOCT_CS*	Age (year)	SOCT_CS*
Sib-pairs (n)	92		108	
Severe	8.8	25.32 ± 0.28	9.29	24.60 ± 0.44
Mild	8.51	13.67 ± 0.45	9.27	22.54 ± 0.47
t-test	p = 6.17E−1	p = 7.18E−52	p = 9.71E−1	p = 1.82E−03

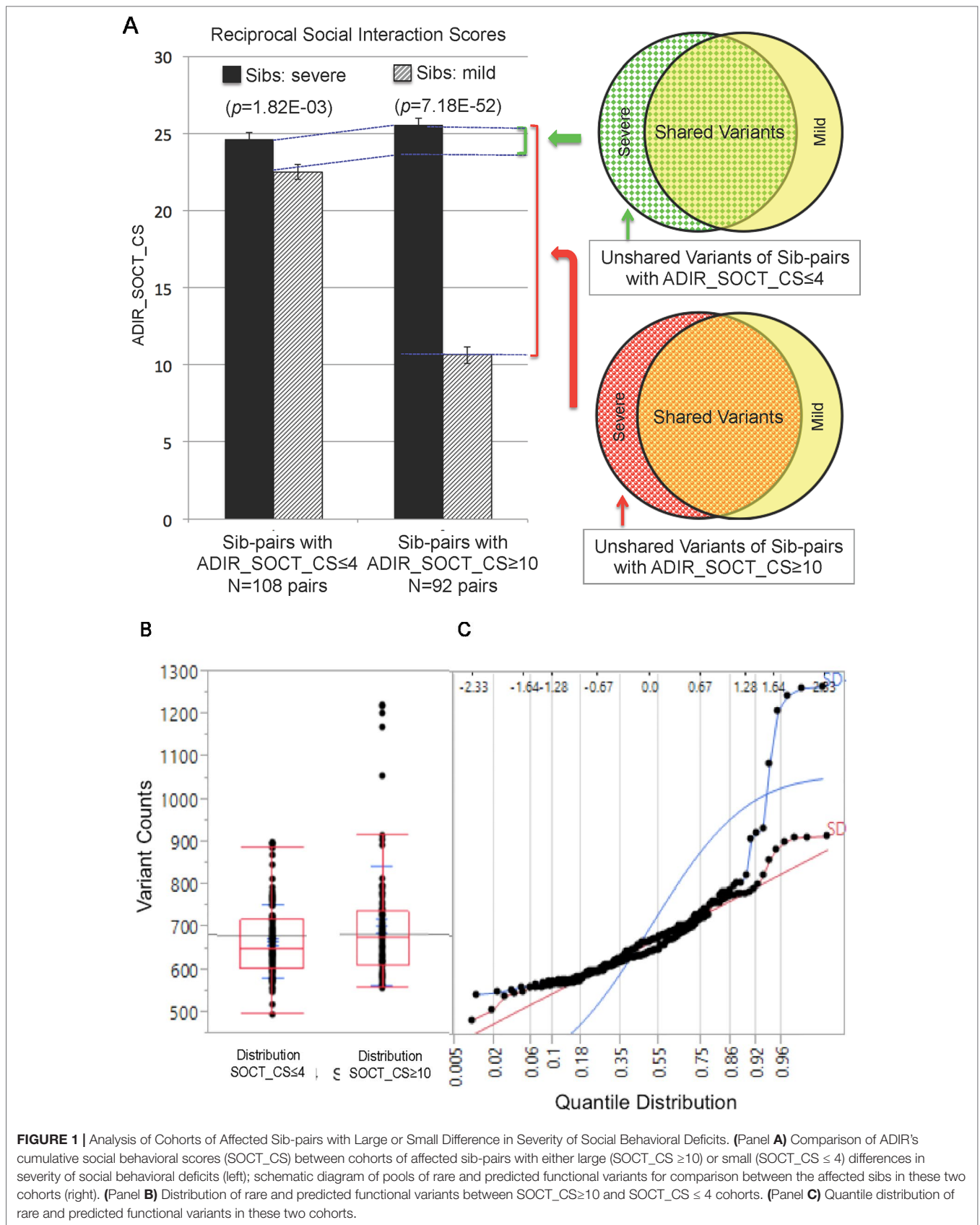*ADI-R's Cumulative Total Social Interaction Score (SOCT_CS).

from each affected sib-pair to identify those unshared variants present only in the sibs with severe social deficits. We compared cumulative allele frequencies and quantile distribution of these sets of unshared variants between the cohorts of SOCT_CS≥10 and SOCT_CS ≤ 4 (**Figures 1B, C**). We observed a higher mutation burden of unshared variants in SOCT_CS≥10 compared to that in SOCT_CS ≤ 4 (SOCT_CS≥10: 705.1 ± 16.2; SOCT_CS ≤ 4, 668.3 ± 9.0; p = 0.025), which is consistent with the prior hypothesis. Furthermore, this set of unshared severe only variants in SOCT_CS≥10 are significantly enriched in the SFARI gene set (http://gene.sfari.org) (hypergeometric p value: 2.7E−14) (**Figure 2** and **Supplementary Table 3**).

Network Analysis to Associate Risk Gene Networks With Social Deficits in Autism

Since these unshared, severe only variants were found by comparing affected sibs in the SOCT_CS≥10 cohort, we speculate that genes harboring these variants are enriched in signaling networks that are crucial to social deficits in autism. To identify signaling genes in networks connected to these variants, we performed a network analyses on the gene-set found in sibs with higher scores in the SOCT_CS≥10 cohort using three platforms (Huguet et al., 2013) InWeb_IM, an integrated human PPI network, (Devlin and Scherer, 2012) MSigDB[2], an annotated gene set for Gene Set Enrichment Analysis (GSEA), and (Gaugler et al., 2014) GeNetMeta, a unified web-based platform for network analyses of genetic data. The InWeb and GeNets algorithm scoring system (Lage et al., 2007) calculates a connectivity p value that indicates whether the network was significantly more connected than expected, for a gene set of this size and the global connectivity of its genes, to construct a modular sub-network structure of the underlying genes. The GeNets algorithm builds a general classifier to predict pathway membership from networks in the InWeb PPI network. For any candidate gene, the classifier can assign a probability that it belongs to a pathway as defined

TABLE 1 | Affected Male Sib-Pairs with Severity versus Mild Social Behavioral Deficits.

Affected sibs		Age (Year)	First Walk (Month)	Social behavior (SOCT_CS)
Phenotype	Number	(Mean ± SEM)	(Mean ± SEM)	(Mean ± SEM)
Severe	126	8.53 ± 0.33	12.74 ± 0.21	25.32 ± 0.28
Mild	126	8.15 ± 0.39	12.35 ± 0.15	13.67 ± 0.45
	t-test	0.67	0.12	2.79E−61

FIGURE 1 | Analysis of Cohorts of Affected Sib-pairs with Large or Small Difference in Severity of Social Behavioral Deficits. **(Panel A)** Comparison of ADIR's cumulative social behavioral scores (SOCT_CS) between cohorts of affected sib-pairs with either large (SOCT_CS ≥10) or small (SOCT_CS ≤ 4) differences in severity of social behavioral deficits (left); schematic diagram of pools of rare and predicted functional variants for comparison between the affected sibs in these two cohorts (right). **(Panel B)** Distribution of rare and predicted functional variants between SOCT_CS≥10 and SOCT_CS ≤ 4 cohorts. **(Panel C)** Quantile distribution of rare and predicted functional variants in these two cohorts.
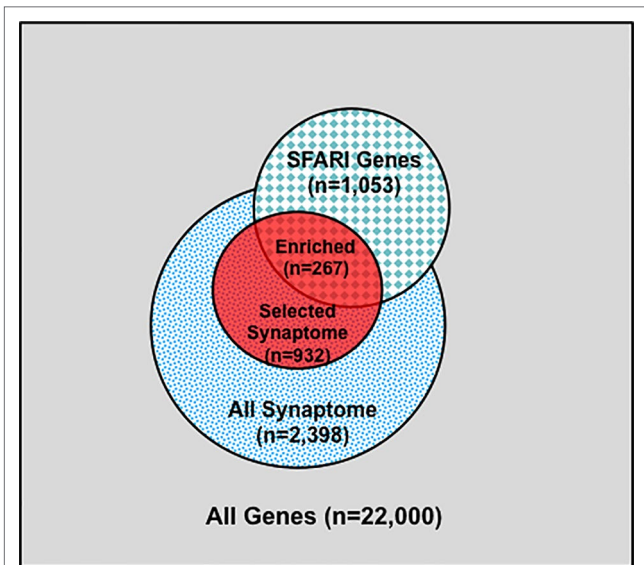
FIGURE 2 | Synaptome Genes with Rare and Predicted Functional Variants are Enriched in SFARI Gene Set. Schematic distribution of total genes (n = 22,000), all synaptome genes (n = 2,398), selected synaptome genes with rare (MAF ≤ 0.01) and predicted functional variants (n = 932), SFARI autism gene set (n = 1,053), and shared genes between the selected synaptome and SFARI gene sets (n = 267).

by the candidate's architectural properties in the overall network. This concept will be used to identify functional modules in gene sets and simplify visualizations. A module is a set of genes (called nodes) that are more connected to one another than they are to other groups of genes.

A total of 932 synaptome genes with rare (MAF ≤ 0.01) and predicted functional variants were identified in ≥1 pairs of affected sibs; 276 genes are shared in ≥3 pairs of affected sibs and 32 in ≥10 pairs (**Figure 2** and **Supplementary Table 2**). The top 250 of 276 genes that were shared in ≥3 families were input into these three analysis platforms to identify connected gene sets and communities. Using a connective p value of 2.00E−3 as a cutoff, communities enriched for the largest gene sets from each of the three analysis platforms were identified (**Table 3**). GeNetMeta identified 166 connected genes that were classified into 10 communities (**Figure 3A**). Community 2 showed the largest enrichment for network genes (**Figure 3B**). InWeb identified 142 connected genes that were classified into nine network communities (**Supplementary Figure 1**). Community 4 showed the largest enrichment for network genes (**Figure 3C**).

MsigDB identified 142 connected genes that were classified into eight network communities (**Supplementary Figure 2**). Community 1 showed the largest enrichment for network genes (**Figure 3D**). A total of 20 seed genes were identified in the communities that showed the largest enrichment from all three platforms (**Table 3**).
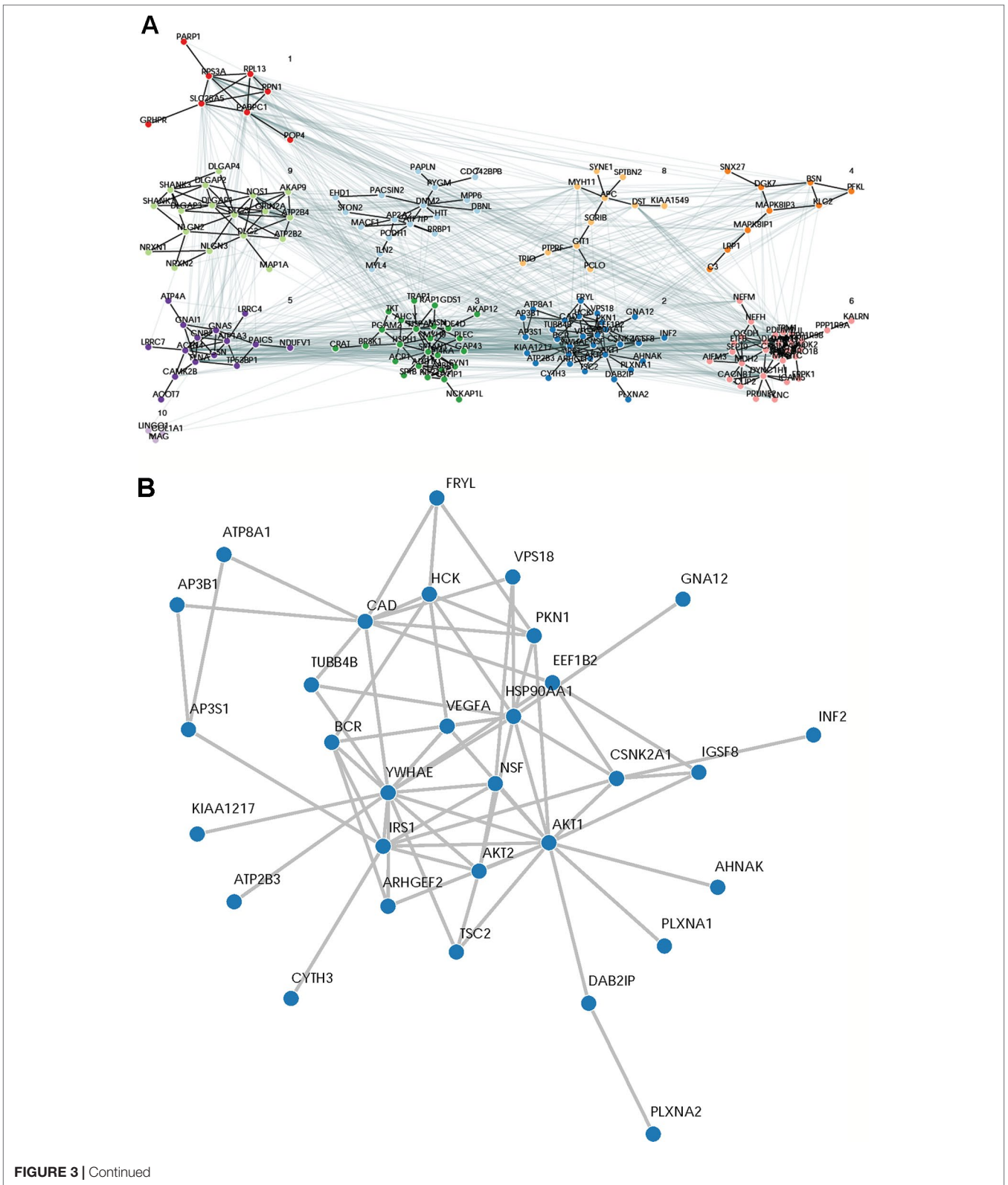
The sub-network analysis ranks genes and predicts candidates based on InWeb PPI patterns found in known pathways, highlights genes that are more connected to one another than they are to other genes in other modules, and segments them based on their similarity to known pathway gene sets. For significance, the algorithm calculates the density of the network as defined by density = (# of edges/# possible edges) and compute the density for randomly sampled gene sets and its empirical determined p value. It computes p values for network overall and also by module to test connectivity of these sub-network. Finally, gene set enrichment will be conducted using a Bonferroni-adjusted hypergeometric test on MSigDB canonical pathways. The three analysis platforms identified two top neural signaling networks based on significance, size of connected genes, and their biological relevance (**Table 4**). Eight genes were connected to the phosphoinositide 3-kinase (PI3K)-AKT-tuberous sclerosis complex (TSC)2-mammalian target of rapamycin (mTOR) network with the most significant Enrichment Score (eScore) P value of 3.36E−07. Variants in these eight genes were identified in 68 of 126 patients in the SOCT_CS≥10 cohort. Nine genes were connected to the Nerve growth factor (NGF)-PC12 signaling network with the most significant eScore P value of 8.94E−07. Variants in these nine genes were identified in 50 of 126 patients in the SOCT_CS≥10 cohort. Additional signaling networks of potential importance in autism social behavioral deficits include AVB3-integrin (n = 7 genes), PI3K-ERBB2/4 (n = 4 genes), Sema3A-PKA (n = 3 genes), and Reelin (n = 3 genes) (**Table 5**).

DISCUSSION

We tested a strategy using affected sib-pairs to identify rare genetic variants and signaling networks associated with specific behavioral domains of autism (Hu and Steinberg, 2009; Hu et al., 2011; Sacco et al., 2012; Veatch et al., 2014). This study design explores unshared rather than shared variants and differences in the severity of core behavioral domains between the affected sibs from same-proband families. It is based on the assumptions that: 1) the cumulative rare variants of large effect contribute

TABLE 3 | Network Analyses Identify Network Communities Connected to SOCT_CS≥10 Geneset.

Network analysis platform	Input geneset (n)	Connected geneset (n)	Connected network community					Shared Genes (n)
			Total number	Largest geneset	Community size (n)	Enriched genes (n)	Connectivity cutoff (p)	
InWeb	250	142	9	4	26	24	2.00E−03	20
MSigDB	250	137	8	1	27	38	2.00E−03	20
GeNetMeta	250	166	10	2	31	25	2.00E−03	20

FIGURE 3 | Continued

FIGURE 3 | Network Analysis Identified Network Communities Connected to the SOCI_CS≥10 Gene Set. Top 250 synaptome genes that carry rare and predicted functional variants and are shared for ≥3 affected sib-pairs were input separately into the following three analysis platforms to identify connected gene sets and network communities. **(Panel A)** GetNetMeta analysis identified 10 network communities connected to SOCT_CS≥10 geneset. **(Panel B)** SOCT_CS≥10 genes are enriched in community 2 from GeNetsMeta analysis. **(Panel C)** SOCT_CS≥10 genes are enriched in community 4 from InWeb analysis. **(Panel D)** SOCT_CS≥10 genes are enriched in community 1 from MSigDB analysis.

TABLE 4 | Two Top Signaling Networks Identified from Gene Communities Connected to SOCT_CS≥10 Geneset.

NEURAL SIGNALING NETWORK	ANALYSIS PLATFORM	eSCORE	OVERLAPING GENES#
PI3K_AKT_MTOR SIGNALING			
PID_PI3KCI_AKT_PATHWAY	InWeb, MSigDB, GeNetsMeta	2.39E-07	AKT1, AKT2, HSP90AA1, YWHAE
ST_PHOSPHOINOSITIDE_3_KINASE_PATHWAY	InWeb, MSigDB, GeNetsMeta	3.01E-07	AKT1, AKT2, CYTH3, YWHAE
REACTOME_PI3K_AKT_ACTIVATION	InWeb, MSigDB, GeNetsMeta	3.36E-07	AKT1, AKT2, IRS1, TSC2
REACTOME_AKT_PHOSPHORYLATES_TARGETS_CYTOSOL	InWeb, MSigDB, GeNetsMeta	7.35E-07	AKT1, AKT2, TSC2
KEGG_MTOR_SIGNALING_PATHWAY	InWeb, MSigDB, GeNetsMeta	1.21E-06	AKT1, AKT2, TSC2, VEGFA
PID_MTOR_4PATHWAY	InWeb, MSigDB, GeNetsMeta	3.80E-06	AKT1, IRS1, TSC2, YWHAE
REACTOME_PIP3_ACTIVATES_AKT_SIGNALING	InWeb, MSigDB, GeNetsMeta	1.20E-05	AKT1, AKT2, TSC2
NGF-PC12-NEURAL DIFFERENTIATION			
REACTOME_SIGNALLING_BY_NGF	InWeb, MSigDB, GeNetsMeta	8.94E-07	AKT1, AKT2, ARHGEF2, IRS1, TSC2, YWHAE
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	InWeb, MSigDB	2.02E-05	AKT1, AKT2, IRS1, YWHAE
ST_DIFFERENTIATION_PATHWAY_IN_PC12_CELLS	InWeb, MSigDB	2.67E-05	AKT1, MAPK8IP1, MAPK8IP3
REACTOME_NGF_SIGNALLING_VIA_TRKA	InWeb, MSigDB	2.80E-05	AKT1, AKT2, IRS1, TSC2

#AKT1, AKT Serine/Threonine Kinase 1; AKT2, AKT Serine/Threonine Kinase 2; HSP90AA1, Heat Shock Protein 90 Alpha Family Class A Member 1; YWHAE, Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Epsilon; CYTH3, Cytohesin 3; IRS1, Insulin Receptor Substrate 1; TSC2, TSC Complex Subunit 2; ARHGEF2, Rho/Rac Guanine Nucleotide Exchange Factor 2; MAPK8IP1, Mitogen-Activated Protein Kinase 8 Interacting Protein 1; MAPK8IP3, Mitogen-Activated Protein Kinase 8 Interacting Protein 3.

TABLE 5 | Distribution of Shared Genes in Neural Networks from Top Connected Communities.

Gene#	Patients No.	PI3A-AKT-mTOR	NGF-PC12	AVB3-INTEGRIN	PI3K-ERBB2/4	SEMA3A-PKA	REELIN
AKAP9	3						
AKT1	3	+	+	+	+		+
AKT2	3	+	+	+	+		
ARHGEF2	3		+				+
BCR	4						
CYTH3	19	+					
GNA12	3						
HSP90AA1	3	+	+	+		+	
IRS1	17	+	+	+	+		
LRP1	5						
MAPK8IP1	10		+	+			+
MAPK8IP3	3		+	+			
PFKL	4						
PKN1	4						
PLXNA1	13					+	
PLXNA2	8					+	
TSC2	4	+	+		+		
TUBB4B	4						
VEGFA	9	+		+			
YWHAE	4	+	+				
No. of Genes		8	9	7	4	3	3
No. of Patients		62	50	47	27	24	16
Total Patients	126	126	126	126	126	126	126

AKAP9, A-Kinase Anchoring Protein 9; AKT1, AKT Serine/Threonine Kinase 1; AKT2, AKT Serine/Threonine Kinase 2; ARHGEF2, Rho/Rac Guanine Nucleotide Exchange Factor 2; BCR (Breakpoint Cluster Region); CYTH3, Cytohesin 3; GNA12 (G Protein Subunit Alpha 12); HSP90AA1, Heat Shock Protein 90 Alpha Family Class A Member 1; IRS1 (Insulin Receptor Substrate 1); LRP1 (LDL receptor Related Protein 1): MAPK8IP1, Mitogen-Activated Protein; Kinase 8 Interacting Protein 1; MAPK8IP3, Mitogen-Activated Protein Kinase 8 Interacting Protein 3; PFKL, Phosphofructokinase, Liver Type; PKN1, Protein Kinase N1; PLXNA1, Plexin A1; PLXNA2, Plexin A2; TSC2, TSC Complex Subunit 2; TUBB4B, Tubulin Beta 4B; VEGFA, Vascular Endothelial Growth Factor A; YWHAE, Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Epsilon. +presence of corresponding genes/proteins in the indicated signaling networks.

quantitatively to the observed differences in severity of behavioral domains in autism, and 2) genes carrying these risk variants are clustered in genetic signaling networks associated with deficits in the respective behavioral domains.

Affected sib-pair analysis carries several advantages over traditional designs that study patients from unrelated families.

First, affected sibs share a similar environment during prenatal course, infancy, and early childhood. These periods are critical for early brain development and are highly susceptible to the pathogenesis of autism. This approach is expected to minimize environmental influences confounding behavioral phenotypes of autism (Robinson et al., 2014). Genetic factors likely play

a bigger role in the observed differences in severity between affected sibs, then between unrelated patients. Second, affected sibs from same families share ≥50% of their genome and presumably, genetic mutations underlying the differences in the severity of specific behavioral domains reside in the unshared portion of genome. Interestingly, a recent WGS study of a cohort of quartet autism families showed that in a large fraction (50–69%) of multiplex ASD families, two affected sibs do not share the same rare penetrant ASD risk alleles (Yuen et al., 2015). Affected sibs with discordant mutations tended to demonstrate more phenotypic variability as compared to those who shared the same risk variants. These results support a strong genetic determinant that is responsible for the difference in severity of the autism phenotype between affected sibs. Third, patients with ASDs manifest a spectrum from mild to severe phenotypes in domains defined by standard behavioral tests. Standard test scores are usually available for patients with a confirmed diagnosis, but not for unaffected relatives. Furthermore, affected sibs enrolled in the AGRE repository were evaluated by the same psychologists using identical sets of behavioral tests, such as the ADI-R, to reduce subjective variations in test scoring. Taken together, our approach effectively enriches genes that regulate autism social behaviors by exploring unshared variants present in severe only sibs of a cohort of affected sib-pairs with large differences in ADI-R social behavioral scores.

It has been suggested that autism-associated mutations cause disturbances in convergent pathways and networks leading to a shared phenotype (Geschwind, 2008). Discovering these common pathways requires a comprehensive, network-based analysis of causal and risk genes (Hormozdiari et al., 2015; Parikshak et al., 2015; Oron and Elliott, 2017). PPI network analysis identifies groups of proteins that physically interact with each other using different databases that curate experimentally validated or predicted PPIs (Parikshak et al., 2015). PPI identifies hubs or highly interconnected proteins, which could be central in the disease-related pathways. The Gene Ontology project provides a unifying description of genes and their biological roles (Ashburner et al., 2000). Using pathway enrichment tools, gene ontology analyses prove to be a highly effective approach to identifying gene networks central to physiological states and disease pathogenesis (Chowdhury and Sarkar, 2015). Network-based analysis has proven to be a powerful approach to identifying risk genes and disease mechanisms in psychiatric disorders with a large genetic contribution.

Two key neural signaling networks, PI3K-AKT-tuberous sclerosis complex (TSC)2-mTOR and NGF-signaling, show significant enrichment for genes harboring unshared variants in SOCT_CS≥10 in connected communities from three analysis platforms in this study. PI3K activates protein kinase B (PKB or AKT), a serine/threonine-specific protein kinase that regulates many aspects of cell physiology including activation of mTOR signaling. In the developing brain, activation of AKT/mTOR signaling is essential for neuronal development, synaptic formation, and plasticity. Increased activity in PI3K-AKT-TSC2-mTOR signaling has been implicated in the syndromic forms of autism including tuberous sclerosis, phosphatase and tensin homolog (PTEN)-related disorders, neurofibromatosis type I, and fragile X syndrome. Inhibition of this increased activity has been shown to improve autism-related symptoms in mouse models of PTEN and TSC1. Using multiple network analysis platforms, our study further implicates a role of this signaling network in the pathogenesis of nonsyndromic autism and particularly in its contribution to social deficits.

Recent studies have implicated NGF signaling in autism core behavioral deficits. Genetic analyses of heritable quantitative traits that correlate with autism identified an association of NGF locus with nonverbal communication in a large cohort of patients (Lu et al., 2013). One study discovered an association of several SNPs in NTRK1 with autism behavioral traits as measured by empathy quotient and autism spectrum quotient (Chakrabarti et al., 2009). Another study on differential alternative splicing in the blood samples from 2- to 4-year-old boys with autism showed a significant difference for several NGF signaling genes including NGF receptor (Stamova et al., 2013). Very interestingly, an animal study on communal nesting (CN), a highly stimulating early social enrichment for rodents, showed that CN results in significant differences in social behaviors later in life and is associated with higher NGF and BDNF levels in the brain of adult mice (Branchi et al., 2006). Additional signaling networks that are also implicated in social behavioral deficits in autism include AVB3-INTEGRIN (Schuch et al., 2014; Dohn et al., 2017; Gabriele et al., 2019), PI3K-ERBB2/4 (Pinto et al., 2010), SEMA3A-PKA, and REELIN signaling (De Rubeis et al., 2014; Lammert et al., 2017; Stessman et al., 2017).

Together, our studies support that multiple signaling networks are involved in the risk and pathogenesis of social behavioral deficits in autism. Analysis of affected sib-pairs shall be a valuable approach to systematically identifying signaling networks crucial to development of the core behaviors in autism spectrum disorders. Comprehensive analyses of genetic variants from whole genome sequencing data, and ADI-R behavioral scores in the three-core domains in an independent cohort of affected sibs with autism, should help to systematically characterize domain-specific and/or overlapping roles for key signaling networks in autism spectrum disorders.

DATA AVAILABILITY STATEMENT

All datasets for this study are included in the article/the **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Johns Hopkins University, IRB-2. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin to Autism Genetics Research Exchange (AGRE).

AUTHOR CONTRIBUTIONS

MP and TW designed the study. MP, TN, YC-C, IT, and TW performed experiments and/or analyzed data. FG, DA, JP, RH, and PZ contributed reagents and/or data analysis. MP, PZ, and TW wrote the manuscript. All authors read and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01186/full#supplementary-material

REFERENCES

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Boehnke, M., and Langefeld, C. (1998). Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am. J. Hum. Genet.* 62, 950–961. doi: 10.1086/301787

Branchi, I., D'Andrea, I., Fiore, M., Di Fausto, V., Aloe, L., and Alleva, E. (2006). Early social enrichment shapes social behavior and nerve growth factor and brain-derived neurotrophic factor levels in the adult mouse brain. *Biol. Psychiatry* 60, 690–696. doi: 10.1016/j.biopsych.2006.01.005

Chakrabarti, B., Dudbridge, F., Kent, L., Wheelwright, S., Hill-Cawthorne, G., Allison, C., et al. (2009). Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and Asperger syndrome. *Autism Res.* 2, 157–177. doi: 10.1002/aur.80

Chowdhury, S., and Sarkar, R. (2015). Comparison of human cell signaling pathway databases—Evolution, drawbacks and challenges. *Database (oxford)* bau 126. doi: 10.1093/database/bau126

De Rubeis, S., and Buxbaum, J. (2015). Genetics and genomics of autism spectrum disorder: embracing complexity. *Hum. Mol. Genet.* 24, R24–R31. doi: 10.1093/hmg/ddv273

De Rubeis, S., He, X., Goldberg, A., Poultney, C., Samocha, K., and Cicek, A. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 505, 209–215. doi: 10.1038/nature13772

Devlin, B., and Scherer, S. (2012). Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* 22, 229–237. doi: 10.1016/j.gde.2012.03.002

Dohn, M., Kooker, C., Bastarache, L., Jessen, T., Rinaldi, C., Varney, S., et al. (2017). The gain-of-function integrin 3 Pro33 variant alters the serotonin system in the mouse brain. *J. Neurosci.* 37, 11271–11284. doi: 10.1523/JNEUROSCI.1482-17.2017

Gabriele, S., Canali, M., Lintas, C., Sacco, R., Tirindelli, M., Ricciardello, A., et al. (2019). Evidence that ITGB3 promoter variants increase serotonin blood levels by regulating platelet serotonin transporter trafficking. *Hum. Mol. Genet.* 28, 1153–1161. doi: 10.1093/hmg/ddy421

Gaugler, T., Klei, L., Sanders, S., Bodea, C., Goldberg, A., Lee, A., et al. (2014). Most genetic risk for autism resides with common variation. *Nat. Genet.* 46, 881–885. doi: 10.1038/ng.3039

Geschwind, D. (2008). Autism: Many genes, common pathways?. *Cell* 135, 391–395. doi: 10.1016/j.cell.2008.10.016

Hormozdiari, F., Penn, O., Borenstein, E., and Eichler, E. (2015). The discovery of integrated gene networks for autism and related disorders. *Genome Res.* 25, 142–154. doi: 10.1101/gr.178855.114

Hu, V., and Steinberg, M. (2009). Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Res.* 2, 67–77. doi: 10.1002/aur.72

Hu, V., Addington, A., and Hyman, A. (2011). Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association

analyses of published GWAS data. *PloS One* 6, e19067. doi: 10.1371/journal.pone.0019067

Huguet, G., Ey, E., and Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. *Ann. Rev. Genomics Hum. Genet.* 14, 191–213. doi: 10.1146/annurev-genom-091212-153431

Huguet, G., Ey, E., and Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. *Annu. Rev. Genomics Hum. Genet.* 14, 191–213. doi: 10.1146/annurev-genom-091212-153431

Jeste, S., and Geschwind, D. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat. Rev. Neurol.* 10, 74–81. doi: 10.1038/nrneurol.2013.278

Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., et al. (2012). Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* 3, 9. doi: 10.1186/2040-2392-3-9

Lage, K., Karlberg, E., Størling, Z., Olason, P., Pedersen, A., Rigina, O., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* 25, 309–316. doi: 10.1038/nbt1295

Lammert, D., Middleton, F., Pan, J., Olson, E., and Howell, B. (2017). The de novo autism spectrum disorder RELN R2290C mutation reduces Reelin secretion and increases protein disulfide isomerase expression. *J. Neurochem.* 142, 89–102. doi: 10.1111/jnc.14045

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, T., Wernersson, R., Hansen, R., Horn, H., Mercer, J., Slodkowicz, G., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14, 61–64. doi: 10.1038/nmeth.4083

Li, T., Kim, A., Rosenbluh, J., Horn, H., Greenfeld, L., An, D., et al. (2018). GeNets: a unified web platform for network-based genomic analyses. *Nat. Methods* 15, 543–546. doi: 10.1038/s41592-018-0039-6

Liang, K., Chiu, Y., Beaty, T., and Wjst, M. (2001). Multipoint analysis using affected sib pairs: incorporating linkage evidence from unlinked regions. *Genet. Epidemiol.* 21, 105–122. doi: 10.1002/gepi.1021

Lim, E., Raychaudhuri, S., Sanders, S., Stevens, C., Sabo, A., MacArthur, D., et al. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77, 235–242. doi: 10.1016/j.neuron.2012.12.029

Lord, C., Rutter, M., and LeCouteur, A. (1994). The Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Aut Dev. Dis.* 24, 659–685. doi: 10.1007/BF02172145

Lu, A., Yoon, J., Geschwind, D., and Cantor, R. (2013). QTL replication and targeted association highlight the nerve growth factor gene for nonverbal communication deficits in autism spectrum disorders. *Mol. Psychiatry* 18, 226–235. doi: 10.1038/mp.2011.155

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce Framework for analysis next-generation DNA sequencing Data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Michaelson, J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., et al. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151, 1431–1442. doi: 10.1016/j.cell.2012.11.019

Oron, O., and Elliott, E. (2017). Delineating the common biological pathways perturbed by ASD's genetic etiology: lessons from network-based studies. *Int. J. Mol. Sci.* 18, 828. doi: 10.3390/ijms18040828

Parikshak, N., Gandal, M., and Geschwind, D. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* 16, 441–458. doi: 10.1038/nrg3934

Perdry, H., Müller-Myhsok, B., and Clerget-Darpoux, F. (2012). Using affected sib-pairs to uncover Rare Disease Variants. *Hum. Heredity* 74, 129–141. doi: 10.1159/000346788

Pinto, D., Pagnamenta, A., Klei, L., Anney, R., Merico, D., and Regan, R. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368–372. doi: 10.1038/nature09146

Pirooznia, M., Wang, T., Pirooznia, M., and Wang, T. (2019) Psyanptome_ASD.vcf.zip; doi: 10.6084/m9.figshare.9248981

Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Huganir, R., et al. (2012). SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* 28, 897–899. doi: 10.1093/bioinformatics/bts040

Pirooznia, M., Wang, T., Avramopoulos, D., Potash, J., Zandi, P., and Goes, F. (2016). High-throughput sequencing of the synaptome in major depressive disorder. *Mol. Psychiatry* 21, 650–655. doi: 10.1038/mp.2015.98

Poznik, G., Adamska, K., Xu, X., Krolewski, A., and Rogus, J. (2006). A Novel Framework for Sib Pair Linkage Analysis. *Am. J. Hum. Genet.* 78, 222–230. doi: 10.1086/499827

Raj, T., Li, Y., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., et al. (2018). Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* 50, 1584–1592. doi: 10.1038/s41588-018-0238-1

Robinson, E., Samocha, K., Kosmicki, J., McGrath, L., Neale, B., Perlis, R., et al. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15161–15165. doi: 10.1073/pnas.1409204111

Sacco, R., Lenti, C., Saccani, M., Curatolo, P., Manzi, B., Bravaccio, C., et al. (2012). Cluster analysis of autistic patients based on principal pathogenetic components. *Autism Res.* 5, 137–147. doi: 10.1002/aur.1226

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., and Reichenberg, A. (2014). The familial risk of autism. *JAMA* 311, 1770–1777. doi: 10.1001/jama.2014.4144

Schuch, J., Muller, D., Endres, R., Bosa, C., Longo, D., Schuler-Faccini, L., et al. (2014). The role of β3 integrin gene variants in Autism Spectrum Disorders — Diagnosis and symptomatology. *Gene* 553, 24–30. doi: 10.1016/j.gene.2014.09.058

Spiker, D., Lotspeich, L., Dimiceli, S., Myers, R., and XXXN., R. (2002). Behavioral phenotypic variation in autism multiplex families: evidence for a continuous severity gradient. *Am. J. Med. Genet.* 114, 129–136. doi: 10.1002/ajmg.10188

Stamova, B., Tian, Y., Nordahl, C., Shen, M., Rogers, S., Amaral, D., et al. (2013). Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. *Mol. Autism* 4, 30. doi: 10.1186/2040-2392-4-30

Stessman, H., Xiong, B., Coe, B., Wang, T., Hoekzema, K., and Fenckova, M. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* 49, 515–526. doi: 10.1038/ng.3792

Sul, J., Cade, B., Cho, M., Qiao, D., Silverman, E., Redline, S., et al. (2017). Increasing generality and power of rare-variant tests by utilizing extended pedigrees. *Am. J. Hum. Genet.* 99, 846–859. doi: 10.1016/j.ajhg.2016.08.015

Tick, B., Bolton, P., Happé, F., Rutter, M., and Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J. Child Psychol. Psychiatry* 57, 585–595. doi: 10.1111/jcpp.12499

Veatch, O., Veenstra-VanderWeele, J., Potter, M., Pericak-Vance, A., and Haines, J. (2014). Genetically meaningful phenotypic subgroups in autism spectrum disorders. *Genes Brain Behav.* 13, 276–285. doi: 10.1111/gbb.12117

Wagner, G., Kin, K., and Lynch, V. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131, 281–285. doi: 10.1007/s12064-012-0162-3

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi: 10.1093/nar/gkq603

Xing, C., Sinha, R., Xing, G., Lu, Q., and Elston, R. (2006). The Affected-/ Discordant-sib-pair design can guarantee validity of multipoint model-free linkage analysis of incomplete pedigrees when there is marker-marker disequilibrium. *Am. J. Hum. Genet.* 79, 396–401. doi: 10.1086/506331

Yu, T., Chahrour, M., Coulter, M., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., et al. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77, 259–273. doi: 10.1016/j.neuron.2012.11.002

Yuen, R., Thiruvahindrapuram, B., Merico, D., Walker, K., Tammimies, K., Hoang, N., et al. (2015). Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* 21, 185–191. doi: 10.1038/nm.3792

Zhang, H., and Risch, N. (1996). Mapping quantitative-trait loci in humans by use of extreme concordant sib pairs: Selected Sampling by Parental Phenotypes. *Am. J. Hum. Genet.* 59, 951–957.