The Contribution of Autism-Related Common Variants to the Familial Aggregation of Quantitative Autistic Traits

Rachael E. Wagner (wagnerr@wustl.edu)
Washington University School of Medicine in Saint Louis: Washington University in St Louis School of Medicine https://orcid.org/0000-0001-5478-139X
William Howells
Washington University School of Medicine in Saint Louis: Washington University in St Louis School of Medicine
Emma C. Johnson
Washington University School of Medicine in Saint Louis: Washington University in St Louis School of Medicine
Arpana Agrawal
Washington University School of Medicine in Saint Louis: Washington University in St Louis School of Medicine
Jennifer K. Lowe
University of California Los Angeles
Daniel H. Geschwind
University of California Los Angeles
John N. Constantino
Washington University School of Medicine in Saint Louis: Washington University in St Louis School of Medicine

Research Article

Keywords: autism, quantitative autistic traits, polygenic risk scores

Posted Date: March 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-325654/v1

Abstract

Background: Autism spectrum disorder (ASD) polygenic risk scores (PRS), derived from genome-wide association studies (GWAS), have been employed to predict ASD diagnosis in the general population (*Nagelkerke's* R^2=2.45%). Critically, large genetic epidemiological studies have demonstrated that family members of affected individuals exhibit subclinical quantitative autistic traits (QAT), suggesting that diagnosed ASD is the pathological tail of a continuous distribution of QAT. Recent molecular genetic studies have reinforced these findings by uncovering a genetic correlation between ASD and QAT in the general population, emphasizing the importance of employing QAT in GWAS. Yet no study to date has examined the extent to which aggregated effects of common variants contributes to variance in QAT in a familial multiplex sample—one enriched for inherited ASD risk.

Methods: Given the elevation of QAT among unaffected members in multiplex families, we examined the contribution of ASD-PRS to QAT, as measured by the Social Responsiveness Scale (SRS), in 1491 subjects from multiplex families in the Autism Genetic Resource Exchange (AGRE)—with (N=538) and without (N=921) an ASD diagnosis. Using the iPSYCH-ASD-GWAS as our discovery dataset, comprised of cases (N=8,605) and controls (N=19,526), we estimated how much variance in QAT that the ASD-PRS explained in our target dataset, AGRE. We also examined if there were interaction effects of diagnosis and sex.

Results: The ASD-PRS explained 0.34% of the variance in QAT (p=0.017). There was no significant interaction effect of disorder status (p=0.511), but there was a significant interaction of sex (p=0.027) and a significant interaction effect of ASD and sex (p=0.031).

Limitations: The primary limitation of this study is the small size of the discovery GWAS (iPSYCH).

Conclusions: Our results reveal that autism-related polygenic load, as measured via the ASD-PRS, significantly predicts QAT—critically, in both affected and unaffected family members. Further, the association of ASD-PRS with QAT in females was modified by diagnosis, indicating that this relationship was the strongest among affected females.The unique nature of this study's multiplex familial sample enables a novel demonstration that common polygenic variation in ASD contributes to familial QAT, particularly among affected females.

Introduction

The aggregation of autism spectrum disorder (ASD) in families is well established, [1] but ASD's genetic architecture remains poorly understood. Three classes of genetic risk have been identified as contributing to ASD: *de novo* mutations, rare inherited variants, and common polygenic variation. Several models, none of which are mutually exclusive, have been proposed to explain this aggregation. A major gene model hypothesizes that either one highly penetrant rare mutation or a restricted number of moderately to highly penetrant mutations contribute to the familial nature of ASD. [2–4] In contrast, polygenic models assume that many common inherited variants (minor allele frequency > 0.05), each of very small effect,

collectively account for a substantial portion of the variance of total ASD risk and contribute to the disorder's familial aggregation. [5–9] Further, single-nucleotide polymorphism (SNP) data heritability estimates suggest that common variants explain a large proportion of the population-wide variance of ASD [5,7,10] as in many other common, complex, psychiatric disorders. [8,11–13]. In fact, the latest genome-wide association study (GWAS) of autism has revealed five statistically significant loci and a SNP-heritability of ~12%. [7] Some evidence suggests that even highly penetrant *de novo* mutations are operating on a background of high polygenic risk burden for ASD. [14]

Yet characterizing this polygenic risk has been more challenging than the field initially hoped. Given the underwhelming amount of variance explained by GWAS and relatively few genome-wide significant loci uncovered, the field of quantitative genetics developed polygenic risk scores (PRS), which capture the aggregate effect of genetic variants (polygenicity), even those that may not have reached a stringent genome-wide significance threshold. A cumulative index of measured genetic liability to a disorder, PRS quantifies within an individual the aggregate effect of common variants for a given trait, typically calculated as the sum of trait-associated alleles across the genome, weighted by effect size. [15,16]

Both rare and common genetic variation act additively to contribute to ASD risk in simplex (one clinically-affected individual) and multiplex (more than one clinically-affected individual) families, though evidence also indicates that there exist at least partially disparate genetic pathways in simplex versus multiplex families. A growing body of literature suggests that in simplex families, relatively rare, *de novo*, more highly penetrant and deleterious genetic mutations are present, whereas in multiplex families, evidence suggests a more highly polygenic, common variant burden. [10,17,18] Supporting a polygenic (additive genetic) model in multiplex families is the fact that relatives of children with ASD demonstrate elevated quantitative autistic traits. [19–21] In fact, first-degree relatives without an ASD diagnosis in multiplex families exhibit elevated quantitative autistic traits (QAT; subthreshold autistic-like deficits in social interaction and communication, behavioral rigidity, restricted interests and repetitive behaviors not severe enough to meet a relatively arbitrary cut-off for clinical diagnosis) [19,20], yet this pathological shift has not been found in first-degree relatives of probands in simplex families. [19] Family and twin literature supports the theory that clinical diagnosis of ASD is often the pathological tail of a continuous distribution of heritable quantitative autistic traits. [22–24]

Given the clinical and epidemiological support for an additive, polygenic model of ASD, molecular genetic studies have begun to interrogate the relationship between diagnosed ASD and QAT to determine their genetic overlap. Employing a general population sample, Robinson and colleagues [25] found a genetic correlation (SNP-rg) of ~.3 between ASD and a quantitative measure of autistic traits, the Social Communication Disorders Checklist (SCDC). However, the SCDC captures only deficits in social communication and interaction and does not evaluate the other established domain of ASD symptomology: restricted interests and repetitive behaviors. Given this genetic overlap, St. Pourcain and colleagues found in a UK population-based study that ASD-PRS are higher in ASD cases than in pseudo-controls and were associated with variation in SCDC scores, but only at 8 years (PGC-ASD: adjusted R^2max = 0.13%, P_{min} = 0.0042). [26] Bralten and colleagues employed a novel questionnaire of autistic

traits in a Dutch general population sample to evaluate genetic overlap of ASD with each of the individual traits assessed by the questionnaire as well as employ an ASD PRS to predict phenotypic variance in autistic traits. They found small, yet statistically significant overlap with and prediction of some of the autistic traits (R^2 values ranging from .17-.54% for some subscales). [27] Yet to our knowledge, no study to date has employed ASD-PRS in a multiplex familial sample to predict variation in a psychometrically-established valid and reliable phenotype that captures the continuous variation in trait distribution encompassing social communication and interaction and restricted interests and repetitive behaviors. A multiplex familial sample is an ideal sample in which to interrogate this question, as these individuals are more likely to carry risk alleles. To that end, we employed the Social Responsiveness Scale (SRS) [28], a comprehensive measure of QAT, in a multiplex family sample.

We hypothesized that common variants of small individual effect sizes contribute to the familial aggregation of QAT. To examine this question, we constructed a polygenic risk score (PRS) from the iPSYCH ASD GWAS, comprised of cases (N=8,605) and controls (19,526). [29] We then investigated the contribution of common polygenic variation to QAT in our multiplex family cohort. Our primary question was to determine how much variance in QAT phenotype an ASD-PRS derived from a case-control GWAS could explain in a multiplex familial sample. Secondary analyses were conducted to determine if there were interaction effects of diagnosis and sex in our sample. [30,31]

Methods

Discovery GWAS

This study's discovery dataset is a previously-published GWAS of ASD (iPSYCH) [29] based on a Danish nationwide population-based cohort [30] including individuals born in Denmark between 1981 and 2005 and diagnosed with ASD prior to 2014. Cases comprised subjects with ASD as the only ascertained diagnosis (classified in accordance with the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)) (N=8,605 of 12,371 possible ASD cases; ICD codes F84.0, F84.1, F84.5, F84.8 and/or F84.9), and controls (N=19,526) were a randomly-selected subset of the cohort without diagnosis (ICD F00-F99). iPSYCH was chosen as the source GWAS over the combined iPSYCH/Psychiatric Genomics Consortium (PGC) meta-analysis GWAS to avoid overlap with the AGRE target data which is included in the PGC dataset.

AGRE Cohort

This study's target dataset was drawn from Autism Speaks' Autism Genetic Resource Exchange (AGRE), the largest private repository of genetic and phenotype data of families with ASD. [31] The sampling bias of the collection is inclusion of multiplex families. The original AGRE genomic dataset comprised 2303 individuals with both GWAS and SRS data. As outlined below, we excluded any non-European subjects to ensure genetic homogeneity and reduce risk for population stratification. Individuals with known chromosomal or neurogenetic abnormalities (e.g., DiGeorge, Fragile X), as well as non-verbal subjects

were also excluded. This final AGRE dataset comprised 1491 individuals (206 of whom were parents) in 663 families (ages ranged from 1.2 to 79 years); 1459 subjects contributed to the analysis of ASD diagnosis (we did not have data regarding diagnosis for 32 subjects).

Phenotypes

Social Responsiveness Scale (SRS)

The quantitative trait measure utilized is the Social Responsiveness Scale (SRS) [28]. The SRS is a 65-item measure of reciprocal social behavior, deficits in which are characterized as quantitative autistic traits (QAT). The SRS capitalizes on observations of individuals in naturalistic social contexts, by parent and/or teacher report. Its internal consistency is very high (*alpha* = ~.95), and it distinguishes ASD-affected individuals from controls with a Cohen's *d* effect size of ~2.7 and from individuals with other psychiatric conditions with an effect size of *d*= ~1. [28,32] It characterizes variation in the two Diagnostic and Statistical Manual of Mental disorders, 5th edition (DSM-5) core domains of social communication and interaction and restrictive interests and repetitive behaviors. Reciprocal social behavior, as measured by the SRS, is continuously distributed in the general population and highly heritable throughout the range observed from unaffected to sub-clinically affected to fully ASD-affected individuals

Individuals' scores on the SRS were obtained from the Internet System for Assessing Autistic Children (ISAAC) database (https://www.autismtools.org/). When longitudinal scores were available, the earliest assessment was employed. Total raw teacher report scores were prioritized, with parental report scores used when teacher scores were unavailable, as has been done previously. [21] Typical norming and adjustment by sex was not done in order to stratify by sex and test interactions with sex in downstream correlation analysis and regression models.

ASD Diagnosis

Diagnoses of subjects with ASD were established with both the Autism Diagnostic Interview-Revised (ADI-R) [33]and Autism Diagnostic Observation Schedule (ADOS) [34], which are the current gold standard instruments for diagnosing individuals with ASD. Parents not in the AGRE pedigree dataset were counted as unaffected.

Genotyping, Quality Control, and Imputation

Genotyping, quality control, and imputation for the original AGRE dataset have been described elsewhere. [21] This AGRE dataset consisted of 2303 individuals for whom imputed genotypes comprising 5.8M autosomal variants were available. [35]

Genetic Ancestry Principal Components Analysis

In order to create an ancestrally homogenous sample and reduce false positives generated by population stratification, we performed an ancestral principal components analysis (PCA) using Eigensoft software.

[36] We employed 1000 Genomes Phase 3 (1KG) reference panels (European (EUR), ad-mixed American (AMR), and African (AFR)) and projected the resulting eigenvectors onto AGRE subjects with both SRS and GWAS data (N=2303), resulting in factor scores for the first three principal components (PC1 – PC3). Using standard Eigensoft protocol, this analysis yielded a sample of 1491 subjects of European-descent ancestry. PC1 – PC3 were included in all regression models to control for bias due to subpopulations within EUR ancestry.

Quantification and Statistical Analysis

Calculation of PRS

To calculate the ASD-PRS, we used the SNP effect sizes estimated for common variants from the previously-published iPSYCH GWAS of ASD, [30] to ensure the validity of the PRS by estimating effect sizes in an independent cohort. To reduce the multiple testing burden associated with PRS at multiple thresholds, we used the *p*-value of 0.1 as our *a priori* threshold as it was the most predictive in our independent source dataset, as well as filtering by imputation information < 0.70, resulting in 745,190 SNPs. We then took the intersection of variants from the iPSYCH ASD source GWAS that overlapped with the imputed variants from our AGRE target dataset (465177 SNPs). Next, we reduced the list of intersecting variants to an independent set by performing LD-clumping (r^2 < 0.15 within each 250kb window) using PRSice 2.3.1.e ([39] and a 1000 Genomes Phase 3 EUR reference panel, [38] resulting in N=39,902 SNPs. The PRS was calculated for each individual as the number of effect alleles at each SNP in the AGRE dataset weighted by the betas from the discovery iPSYCH GWAS at the same SNP. The PRS was standardized to mean=0, standard deviation (SD)=1 in order to interpret the regression coefficient as the change in SRS score for one SD increase in PRS.

Computation of Odds Ratios

We grouped subjects into quartiles of PRS scores and modeled ASD diagnosis in a logistic regression model. We estimated odds ratios (OR) for ASD in Quartiles 2 through 4, conducting three separate tests of each quartile, as compared to the first (lowest) quartile (<25% PRS). Using generalized estimating equations (GEE) (as implemented in SAS Proc Genmod), the model was adjusted for PC1 – PC3, sex, and family variance. We also computed a Cochran-Armitage trend test, a chi-squared test for a linear trend across the four quartiles to test for dose response.

Correlation Analyses

As a preliminary step and for purposes of data visualization, unadjusted Pearson correlations between SRS and PRS were computed in the total sample and stratified by sex and affected status. Graphs were produced with R version 3.4.1 Copyright (C) 2017. [41]

Linear Mixed Model Regression

We employed a multiple linear mixed model (LMM) regression analysis because our sample comprised multi-level data in which subjects existed in groups (i.e., families), where the SRS phenotype is expected to correlate within families as well as between families. First, to account for relatedness in the AGRE family sample, we calculated a genetic relatedness matrix (GRM) using Plink 1.90 [37]. The GRM was then entered into all models as a random effect, implemented in SAS Proc Mixed [42], to adjust the residual variance for relatedness. Age was not included in our modeling as a covariate because SRS scores are known to be invariant to age effects [40]. For our primary model estimating the variance of QAT explained by PRS (Model 1), the fixed effects comprised PC1 – PC3, sex, and PRS. We computed three other models in which we tested the interactions of PRS*sex (Model 2), PRS*ASD (Model 3), and PRS*sex*ASD (Model 4). To properly control for confounders in models including 2-way interactions, we also included covariate x environment and covariate x gene interaction terms into the models [43] (see Table 4 for a breakdown of all interactions included in the modeling).

To estimate the variance explained, we fit a series of full and reduced models, as outlined in Table 4, and from these multi-level models, we then calculated a conditional pseudo-R^2 using SAS mixed fit, as traditional R^2 statistics are not available for LMM due to the random effect for GRM .[44] We reduced the full model by eliminating the fixed effect or interaction term of interest and measured the difference in pseudo-R^2 between the full model and this reduced model. This change in R^2 (ΔR^2) serves as our estimate of the variance explained by the fixed effect (PRS) or interaction term.

Datasets

iPSYCH discovery GWAS dataset: https://github.com/mgandal/Shared-molecular-neuropathology-across-major-psychiatric-disorders-parallels-polygenic-overlap/tree/master/raw_data/GWAS

AGRE target dataset: obtained from author JL

1000 Genomes Phase 3 LD reference panel: https://ctg.cncr.nl/software/magma

Publicly-Available Software

IMPUTE2: https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

PLINK: https://www.cog-genomics.org/plink2

SHAPEIT2: https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

EIGENSOFT/smartPCA: https://github.com/chrchang/eigensoft/blob/master/POPGEN/smartpca.info

SAS %mixed_fit macro: http://support.sas.com/resources/papers/proceedings13/255-2013.pdf

Results

The sample's descriptive statistics are outlined in Table 1. The distributions of the PRS and SRS observed in the sample stratified by sex are shown in Figures 1 and 2, respectively; and the distributions of PRS and SRS stratified by affected status are depicted in Figures 3 and 4, respectively. The unadjusted mean PRS scores of the total sample, stratified by sex and diagnostic status are reported in Table 2.

Table 1 AGRE sample characteristics

Group	N*	Mean Age (95% CI)	N Sibs (%)
All	1491	14.41(13.80, 15.02)	1285 (86.18)
Females	521	16.69 (15.56, 17.83)	415 (79.65)
Males	970	13.19 (12.49, 13.88)	870 (89.69)
Unaffected	538	22.80 (21.43, 24.17)	332 (61.71)
ASD	921	9.74 (9.49, 9.99)	921 (100.00)
Females: Unaffected	299	21.84 (20.12, 23.57)	193 (64.55)
Females: ASD	215	9.80 (9.27, 10.34)	215 (100.00)
Males: Unaffected	239	24.00 (21.79, 26.21)	139 (58.16)
Males: ASD	706	9.72 (9.44, 10.00)	706 (100.00)

*Sample characteristics, stratified by sex and ASD; *N=1491 subjects contributed to sex strata; n=1459 subjects contributed to strata with ASD.*

Table 2: Mean PRS Scores

Group	N*	Mean PRS (95% CI)
All	1491	0.00 (-0.05, 0.05)
Females	522	-0.03(-0.11, 0.06)
Males	970	0.01(-0.05, 0.08))
Unaffected	538	-0.08(-0.16, 0.01)
ASD	921	0.05(-0.02, 0.12)
Females: Unaffected	299	-0.07(-0.19, 0.04)
Females: ASD	215	0.05(-0.08, 0.19)
Males: Unaffected	239	-0.08(-0.20, 0.05)
Males: ASD	706	0.05(-0.02, 0.12))

*Mean PRS scores, stratified by sex and ASD; *N=1491 subjects contributed to sex strata; n=1459 subjects contributed to strata with ASD.*

ASD as Outcome

Prior to QAT analyses, we conducted a preliminary set of analyses with ASD diagnosis as our outcome measure to verify foundational hypothesized effects. To determine if odds ratios (OR) for ASD diagnosis increased with greater polygenic load, we separated individuals from the AGRE cohort into quartiles of

PRS risk. Using a logistic regression model (adjusted for sex, PC1 – PC3, and family variance) to test for association between the PRS quartiles and diagnostic status, we found that OR in each PRS quartile increased with greater polygenic risk. As outlined in Table 3, compared to the lowest <25% percentile group, OR for individuals in the 75th percentile were statistically significant, OR=1.450 (95% CI: 1.084, 1.941); *p*=0.012. The test for trend across PRS quartiles was significant, OR=1.235 (95% CI: 1.016, 1.245, *p*=0.023), demonstrating a dose response effect of PRS on risk for ASD (i.e., the OR for the ordinal PRS quartile predictor is significant if there is a linear increase in risk across the four quartiles). As expected, the main effect of sex was large: the OR for males was 4.12 (95% CI: 3.27, 5.19, *p*=5.30E-32).

Table 3: Odds ratios of ASD across quartiles of polygenic load

	<25th Percentile	25th-49th Percentile	50th-75th Percentile	>75th Percentile
Odds Ratio (95% CI), p-value	1.0 (Reference)	1.195 (0.886, 1.611); *p*=0.243	1.269 (0.931, 1.730); *p*=0.132	1.450 (1.084, 1.941); *p*=0.012
N Males ASD	163	167	184	192
N Males Unaffected	67	60	63	49
N Females ASD	49	59	53	54
N Females Unaffected	85	79	65	70

Table 3 outlines the OR, 95% CI, and p-value for affected individuals across quartiles of PRS load, based on n=1459 subjects with a determination of ASD status. Number of subjects per quartile, stratified by sex and ASD status are provided for each quartile.

Association between PRS and SRS

As a preliminary step in our analyses, we calculated descriptive statistics to visualize the data. The unadjusted Pearson correlation coefficient of PRS with SRS for the entire sample was *r* = 0.065 (95%CI: 0.014, 0.115, *p*= 0.013). The correlation between PRS and SRS was higher for females *r* = 0.122 (95%CI: 0.036, 0.206); *p* = 0.005, n=521) than for males *r* = 0.029 (95%CI: -0.034, 0.119, *p* = 0.092 , n=970) (Figure 5). When stratified by ASD status, the correlation for affected individuals (*r* = 0.031 (95% CI: -0.033, 0.108; *p* = 0.342, n=921) was slightly lower than that of unaffected individuals *r* = 0.054 (95% CI: -0.033, 0.137; *p*= 0.214, n=538) (Figure 6). When stratified by ASD status and sex, the correlation of PRS with SRS in unaffected females was *r* = 0.055 (95% CI: -0.059, 0.167, *p* = 0.345, n = 299) and for affected females it was *r* = 0.186 (95% CI: 0.052, 0.311, *p* = .006, n = 215). For unaffected males, the correlation was *r* = 0.055 (95% CI: -0.072, 0.181, *p* = 0.397, n = 239) and for affected males, the correlation was *r* = -0.018 (95% CI: -0.092, 0.056, *p* = 0.634, n = 736) (Figure 7).

Variance in QAT Explained by Common Variation in ASD Risk

In our primary model, Model 1, to estimate the variance explained by PRS, we compared the pseudo-R^2 of a full model (PC1 – PC3, sex, and PRS) to the reduced model (eliminating PRS) (Table 4, Model 1). The variance explained (delta R^2) by the PRS was 0.342% (p=0.023). In Model 2, there was a significant interaction of sex with PRS (i.e., differing slopes in Fig. 5), with the variance explained (delta R^2) of sex*PRS being 0.142% (p=0.027). In Model 3, we examined the interaction of PRS with diagnosis. Here the interaction term was not significant (delta R^2 = 0.012%, p=0.511). In other words, the variance explained by PRS in QAT was was not attributable to individuals with ASD diagnosis alone. Finally, in Model 4, we investigated the interaction of PRS, sex, and diagnosis. The interaction effect was significant (delta R^2 = 0. 13%, p=0.031). Consistent with Figure 7, this 3-way interaction demonstrates that the association between ASD-PRS and QAT was most pronounced in affected females. All modeling estimates are outlined in detail in Table 4.

Table 4: Linear mixed model of fixed and random effects

Full Models			
Model	**Effects in Full Model**	**R^2**	**Full Model *p*-value**
A	PC1 – PC3	0.0015	0.511
B	PC1 – PC3 + sex	0.0989	<1E-15*
C	PC1 – PC3 + sex + **PRS**	0.1023	<1E-15*
D	PC1 – PC3 + sex + ASD + PRS +ASD*(PC1 – PC3) + ASD*sex +PRS*(PC1 – PC3) + **PRS*ASD**	0.5782	<1E-15*
E	PC1 – PC3 + sex + ASD + PRS +ASD*(PC1 – PC3) + ASD*sex +PRS*(PC1 – PC3) + **PRS*sex**	0.5795	<1E-15*
F	PC1 – PC3 + sex + ASD + PRS +ASD*(PC1 – PC3) + ASD*sex +PRS*(PC1 – PC3) + **PRS*sex +PRS*ASD**	0.5796	<1E-15*
G	PC1 – PC3 + sex + ASD + PRS +ASD*(PC1 – PC3) + ASD*sex + PRS*(PC1 – PC3) + **PRS*sex +PRS*ASD + PRS*sex*ASD**	0.5809	<1E-15*

Reduced Models				
Model	**Effect**	**Effect tested in ΔR^2**	**ΔR^2**	**ΔR^2 *p*-value**
Model 1	PRS	C vs. B (full model reduced by PRS)	0.0034	0.017*
Model 2	PRS * Sex	F vs. D (full model reduced by PRS*Sex)	0.0014	0.027*
Model 3	PRS * ASD	F vs. E (full model reduced by PRS*ASD)	0.0001	0.511
Model 4	PRS*Sex*ASD	G vs. F (full model reduced by PRS*Sex*ASD)	0.0013	0.031*

Table 4 details the variance explained (pseudo-R^2) for each model and its accompanying p-value. An asterisk (*) indicates a statistically significant p-value at p<0.05; <1E-15 indicates

smaller value than machine precision.

Discussion

Our results reveal that autism-related polygenic load, as measured via the ASD-PRS, not only significantly predicts ASD but is also related to QAT, as assessed using the SRS—critically, in both affected and unaffected family members. The uniqueness of this study is the application of ASD polygenic risk predictors to a multiplex family sample, enriched for inherited liability to ASD, incorporating both affected and unaffected family members who were uniformly phenotyped using a standardized, comprehensive measure of QAT. These findings corroborate a large body of genetic epidemiology research which has established a biological correspondence between clinically diagnosed ASD and subclinical QAT in family members of an affected individual. Although the variance explained was modest, its magnitude was in keeping with other analyses of social behavioral variation in the general population samples. [25–27] Critically, there was no statistical difference in variance predicted in unaffected vs. affected subjects, suggesting that both reflect the genotype-phenotype association indexed by the iPYSCH PRS. It is likely that the study design resulted in individuals without a diagnosis being enriched for PRS load relative to population controls. Despite the lack of moderation of ASD-PRS by diagnosis, the association of ASD-PRS with QAT was the strongest in affected females, indicating that the links between ASD genetic susceptibility and QAT may be most pronounced in females with an ASD diagnosis.

Our results emphasize the importance of considering and incorporating trait variation in (a) the general population and (b) unaffected members of ASD-affected families to improve understanding of the genetic architecture of ASD and risk prediction. Given the mounting evidence that clinically diagnosed ASD is the extreme expression of one or more heritable quantitative traits (at least as it pertains to common variant genetic risk), a failure to capture subclinical variation in QAT and related phenotypes in individuals classified as controls may confound efforts to identify genetic variants associated with ASD. Moreover, greater specification of polygenic risk among individuals with rare genetic disorders may improve efforts to predict variation in expression of deleterious variants and thereby disease severity in affected individuals. [45]

In light of these complexities, moving away from categorical classification of disease and employing *quantitative* approaches to understanding the origins of ASD may reveal a more complete picture of the mechanisms of inheritance in ASD. Uncovering the specific polygenic contributors to ASD is already beginning to reveal a molecular basis for cross-disorder aggregation in families. Tracing the effects of polygenic risk to specific early neurodevelopmental precursors of ASD may identify profiles of allelic variation that map to disparate developmental liabilities, for which specific combinations or permutations give rise to ASD [46], and according to which an individual might be typed for personalized approaches to prevention or therapy. Furthermore, the inclusion of SNPs which are primarily associated with other disorders and traits can actually improve the predictive ability of the ASD-PRS (e.g., SNPs implicated in

schizophrenia, depression, IQ). [7] Therefore, characterizing patients and controls with respect to inherited traits that index genetic liability to ASD—whether or not those traits are specific to ASD—is expected to encompass more of the variance of the condition than can be identified by simply contrasting the genetic profiles of individuals with and without the diagnosis of ASD. Finally, quantitative approaches may also provide a means by which to classify more homogeneous subgroups of patients reflecting underlying biology, thereby allowing the pursuit of underlying neural mechanisms and pathways to be conducted with higher precision.

An important future direction suggested by these results is to determine whether the predictive power of polygenic risk signals for autism might be enhanced by including cohorts of unaffected individuals quantitatively phenotyped for variation in QAT in GWAS discovery sets. A recently-identified nuance of quantitative phenotyping in autism is that the measurements may be significantly less heritable as the threshold for clinical-level affectation is approached: quantitative trait scores in the general population exhibit heritability estimates on the order of 0.85 compared to 0.25 for quantitative measurements of symptom severity (across the wide range of affectation) among clinically-diagnosed individuals. [47] A possible implication, then, would be a paradoxical *improvement* in statistical power for genotype-phenotype association in quantitative trait analyses when sampling from populations *unaffected* by clinical-level aggregation for the trait of interest.

Limitations

As with all complex diseases and traits, a limitation of this study is the proportion of variance explained by the discovery GWAS, likely due at least in part to the small sample size. The largest ASD GWAS (iPSYCH-PGC) is still relatively small (N = 18,381 individuals with ASD and 27,969 control) in comparison to GWAS of other traits, diseases, and psychiatric disorders (e.g., schizophrenia with N = 36,989 cases and 113,075 controls); we intentionally limited the total available discovery set by exclusion of the PGC sample due to its inclusion of some AGRE subjects and the consequent inability to specify with certainty a PGC subset devoid of overlap with the families in our AGRE replication cohort. As ASD GWAS sample sizes increase, the predictive power of PRSs is expected to improve, and the estimation of common variant effect sizes to become more precise. Future studies using a PRS based on weights from a larger GWAS sample are likely to capture even more of the variance in QAT and related, overlapping traits.

Conclusions

In conclusion, we demonstrated for the first time that a PRS for clinical ASD is associated with variation in QAT above and below the threshold for diagnosis, using a quantitative phenotype that is known to specifically index familial liability for autism. This was observed in a sample enriched for the inheritance of clinical autistic syndromes. Our study provides a critical replication of an observed polygenic risk signal for ASD in an independent sample, and supports the contribution of common polygenic risk to the aggregation of clinical and sub-clinical autistic traits in the population. These findings underscore the importance of employing quantitative approaches in future genetic studies of ASD and other

neuropsychiatric disorders, specifically incorporating the measurement of quantitative variation in behavioral traits that exhibit genetic overlap with autism, not only among probands, but among unaffected relatives and controls. Given the modest amount of variance explained in QAT by ASD-PRS, it is possible that future GWAS of QAT will reveal novel signals which will enhance our understanding of the genetic architecture and underlying biology of ASD and related phenotypes.

Abbreviations

1KG—1000 Genomes Phase 3

ADI-R—Autism Diagnostic Interview-Revised

ADOS—Autism Diagnostic Observation Schedule

AFR—African

AGRE—Autism Genetic Resource Exchange

AMR—ad-mixed American

ASD—autism spectrum disorder

EUR—European

GRM—genetic relatedness matrix

GWAS—genome-wide association study

HWE—Hardy Weinberg Equilibrium

ICD—International Statistical Classification of Diseases and Related Health Problems

ISAAC—Internet System for Assessing Autistic Children

LMM—Linear Mixed Model

MAF—minor allele frequency

MZ—monozygotic

OLS—ordinary least squares regression

OR—odds ratio

PCA—principal components analysis

PGC—Psychiatric Genomic Consortium

QAT—quantitative autistic traits

SCDC—Social Communication Disorders Checklist

SD—standard deviation

SNP—single nucleotide polymorphism

SRS—Social Responsiveness Scale

UCLA—University of California Los Angeles

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review boards of the Autism Genetic Resource Exchange (AGRE), the University of California, Los Angeles (UCLA), and Washington University in St. Louis. All subjects provided informed consent for their data to be employed in this research.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from AGRE, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of AGRE.

Competing Interests

JNC receives royalties from Western Psychological Services for the commercial distribution of the Social Responsiveness Scale; however, no royalties were generated from research implementation of the SRS in the AGRE registry. All other authors report no financial relationships with commercial interests.

Funding

Supported by ACE Network grants R01 MH081754 and R01 MH100027 and ACE Center grants P50 HD055784 and R01 MH071425 to Dr. Geschwind; by grant R01 HD042541 and IDDRC Center grant P30 HD062171 to Dr. Constantino. The Autism Genetic Resource Exchange is a program of Autism Speaks and was supported by NIMH grant U24 MH081810.

Authors' Contributions

JNC conceptualized the study and contributed to sample accrual. JKL and DHG participated in sample assembly, advised analyses, and secured grant funding. JNC and REW determined statistical design and analyses, and AA and ECJ contributed substantial statistical and procedural advice. WH conducted the data processing and statistical analyses. JNC, REW, AA, and ECJ analyzed and interpreted the data. REW drafted the manuscript, AA, JKL, ECJ, and JNC critically revised the manuscript, and REW finalized the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We gratefully acknowledge the parents and families participating in AGRE for their participation in this research effort. We also acknowledge the assistance with data acquisition provided by Yi Zhang and Joshua Page.

References

1. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The Familial Risk of Autism. JAMA. 2014;311:1770–7.
2. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43:585.
3. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Hum Mol Genet. 2011;20:3366–75.
4. Torre-Ubieta L, de la, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22:345.
5. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46:881–5.
6. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey JA, et al. Common genetic variants, acting additively, are a major source of risk for autism. Molecular autism. 2012;3:9.
7. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:1–14.
8. Consortium T, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360:eaap8757.
9. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet. 2012;21:4781–92.
10. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey A, et al. Common genetic variants, acting additively, are a major source of risk for autism. Molecular autism. 2012;3:9.
11. Ripke S, Neale BM, Corvin A, Walters JT, Farh K-H, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
12. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression.

Nat Genet. 2018;50:668–81.

13. Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. Molecular Psychiatry. 2018;1–12.
14. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017;49:ng.3863.
15. Bogdan R, Baranger D, Agrawal A. Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. Annu Rev Clin Psychol. 2018;14:1–39.
16. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. Nature Neuroscience. 2016;19.
17. Bernier R, Gerdts J, Munson J, Dawson G, Estes A. Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. Autism Res. 2011;5:13–20.
18. Connolly S, Anney R, Gallagher L, Heron EA. Evidence of Assortative Mating in Autism Spectrum Disorder. Biological Psychiatry. 2019.
19. Virkud YV, Todd RD, Abbacchi AM, Zhang Y, Constantino JN. Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. American journal of medical genetics Part B Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics. 2009;150B:328–34.
20. Lyall K, Constantino JN, Weisskopf MG, Roberts AL, Ascherio A, Santangelo SL. Parental Social Responsiveness and Risk of Autism Spectrum Disorder in Offspring. JAMA Psychiatry. 2014;71:936–42.
21. Lowe JK, Werling DM, Constantino JN, Cantor RM, Geschwind DH. Social Responsiveness, an Autism Endophenotype: Genomewide Significant Linkage to Two Regions on Chromosome 8. Am J Psychiatry. 2015;172:266–75.
22. Robinson EB, Koenen KC, McCormick MC, Munir K, Hallett V, Happé F, et al. Evidence That Autistic Traits Show the Same Etiology in the General Population and at the Quantitative Extremes (5%, 2.5%, and 1%). Arch Gen Psychiatry. 2011;68:1113–21.
23. Constantino JN, Todd RD. Autistic Traits in the General Population: A Twin Study. Arch Gen Psychiatry. 2003;60:524–30.
24. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. Biological psychiatry. 2005;57.
25. Robinson EB, Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet. 2016;48:552–5.
26. Pourcain B, Whitehouse AJ, Ang WQ, Warrington NM, Glessner JT, Wang K, et al. Common variation contributes to the genetic architecture of social communication traits. Mol Autism. 2013;4:34.

27. Bralten J, Hulzen K van, Galesloot M, Vasquez T, Kiemeney AA. L, et al. Autism spectrum disorders and autistic traits share genetics and biology. Mol Psychiatry. 2017;23:1205.
28. Constantino JN, Gruber CP. Social Responsiveness Scale, Second Edition (SRS-2). Western Psychological Services. 2012.
29. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science. 2018;359:693–7.
30. Pedersen C, Bybjerg-Grauholm J, Pedersen M, Grove J, Agerbo E, Bækvad-Hansen M, et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatr. 2018;23:6–14.
31. Lajonchere CM, Consortium A. Changing the Landscape of Autism Research: The Autism Genetic Resource Exchange. Neuron. 2010;68:187–91.
32. Constantino JN, Przybeck T, Friesen D, Todd RD. Reciprocal Social Behavior in Children With and Without Pervasive Developmental Disorders. Journal of Developmental & Behavioral Pediatrics. 2000;21.
33. Rutter M, LeCouteur A, Lord C. ADI-R Autism Diagnostic interview Revised. Manual.. Western Psychological Services. 2003.
34. Lord C, Rutter M, DiLavore PC, Risi S. ADOS: Autism Diagnostic Observation Schedule Manual.. Western Psychological Services. 2001.
35. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14. 1 associate with autism spectrum disorders. Nature. 2009;459:528.
36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904–9.
37. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007;81:559–75.
38. Genomes T 1000. A global reference for human genetic variation. Nature. 2015;68–74.
39. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31:1466–8.
40. Wagner RE, Zhang Y, Gray T, Abbacchi A, Cormier D, Todorov A, et al. Autism-Related Variation in Reciprocal Social Behavior: A Longitudinal Study. Child Dev. 2019;90:441–51.
41. Team CR. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: https://www.r-project.org/.
42. Inc. SI. SAS/STAT® 14.1 User's Guide. Cary: SAS Institute Inc.; 2015.
43. Keller MC. Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. Biol Psychiat. 2014;75:18–24.

44. Singer J, Willett J. Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford UP; 2003.
45. Finucane B, Challman TD, Martin C, Ledbetter DH. Shift happens: family background influences clinical variability in genetic neurodevelopmental disorders. Genet Med. 2016;18:302–4.
46. Constantino JN. Early behavioral indices of inherited liability to autism. Pediatr Res. 2018;85:127–33.
47. Castelbaum L, Sylvester CM, Zhang Y, Yu Q, Constantino JN. On the Nature of Monozygotic Twin Concordance and Discordance for Autistic Trait Severity: A Quantitative Analysis. Behav Genet. 2019;1–10.

Figures

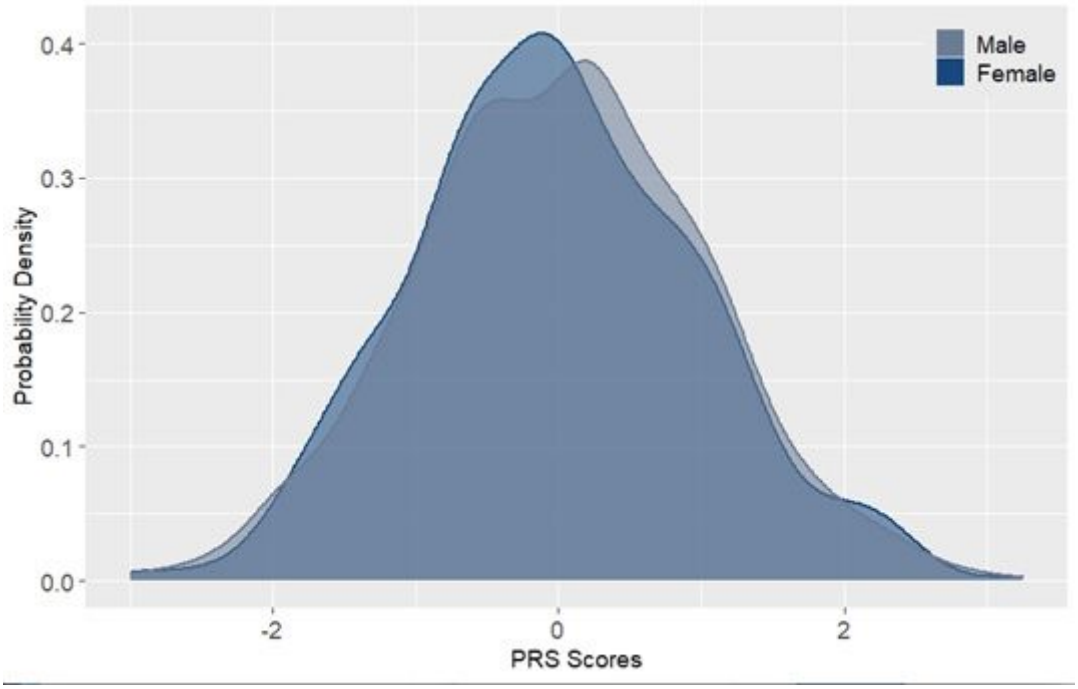

Figure 1

Distribution of relative frequency of PRS scores, stratified by sex

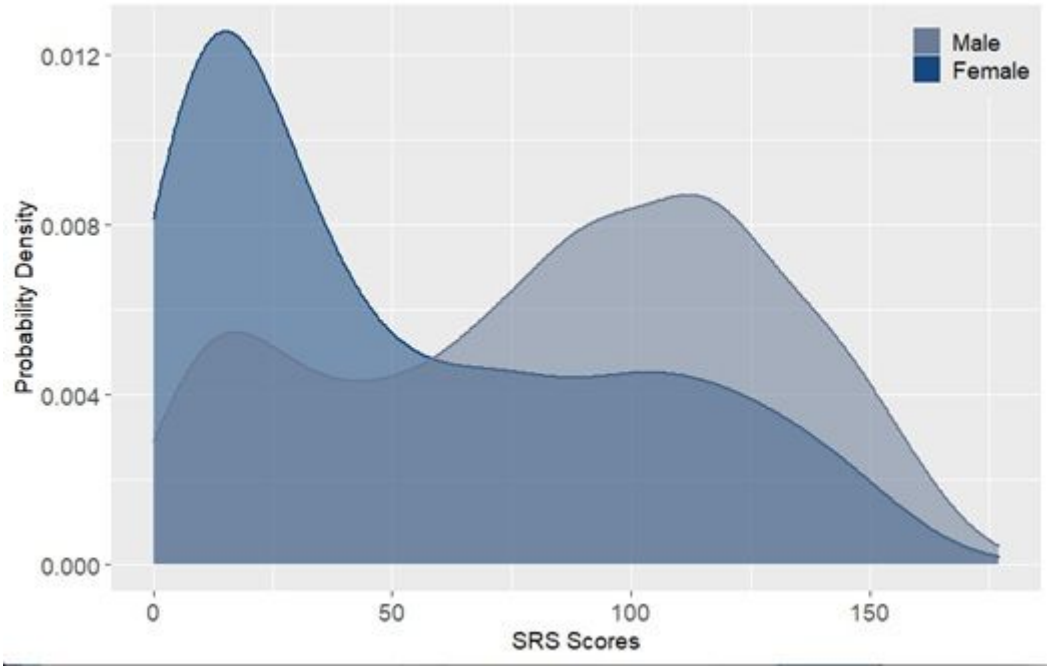

Figure 2

Distribution of relative frequency of SRS scores, stratified by sex

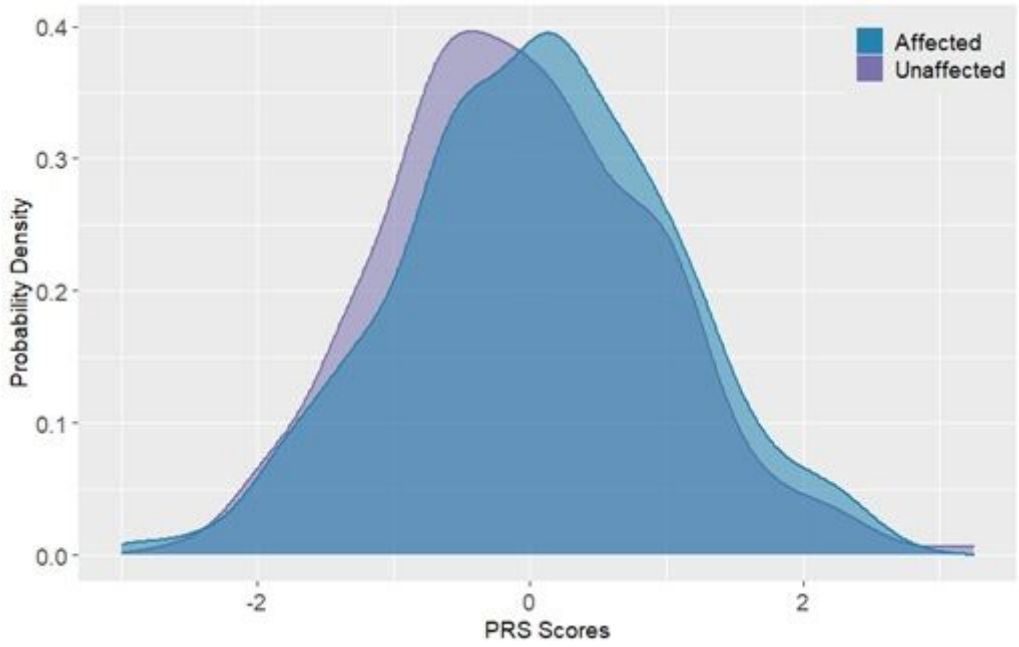

Figure 3

Distribution of relative frequency of PRS scores, stratified by ASD status

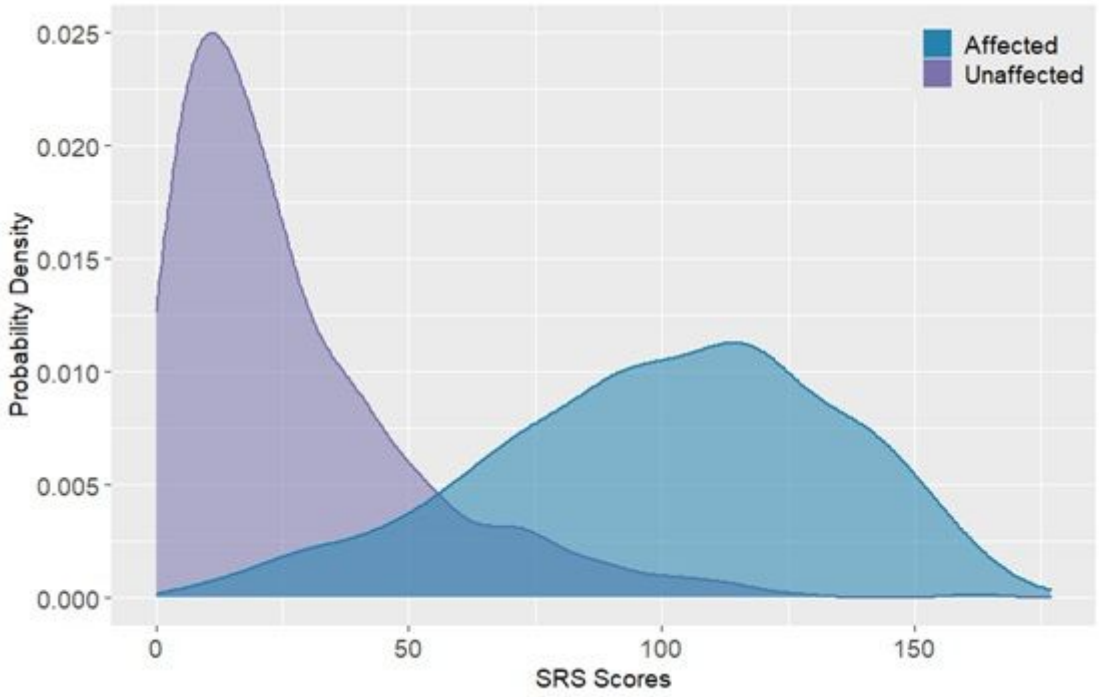

Figure 4

Distribution of relative frequency of SRS scores, stratified by ASD status

Figure 5

Correlation of PRS with SRS scores, stratified by sex

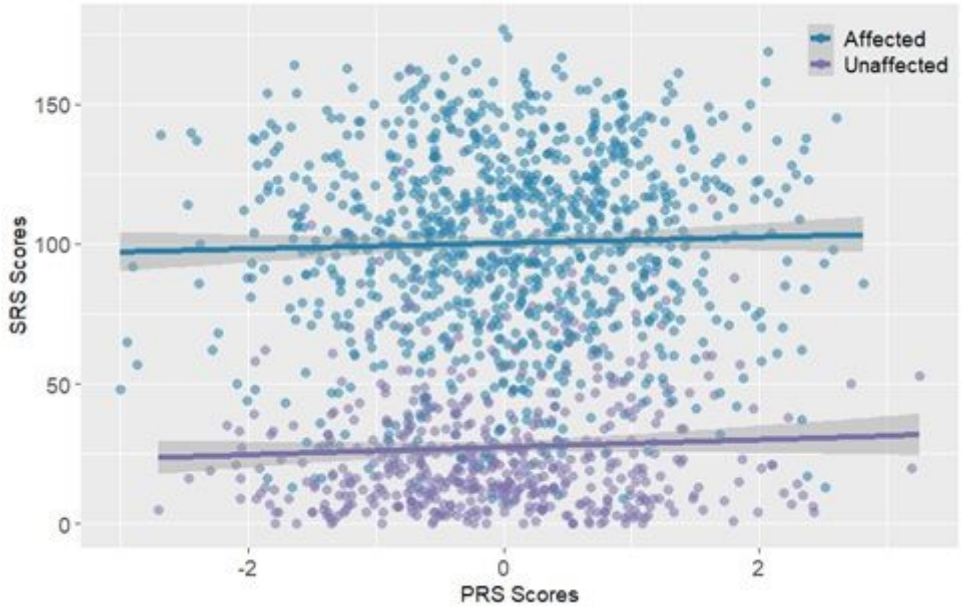

Figure 6

Correlation of PRS with SRS scores, stratified by ASD status

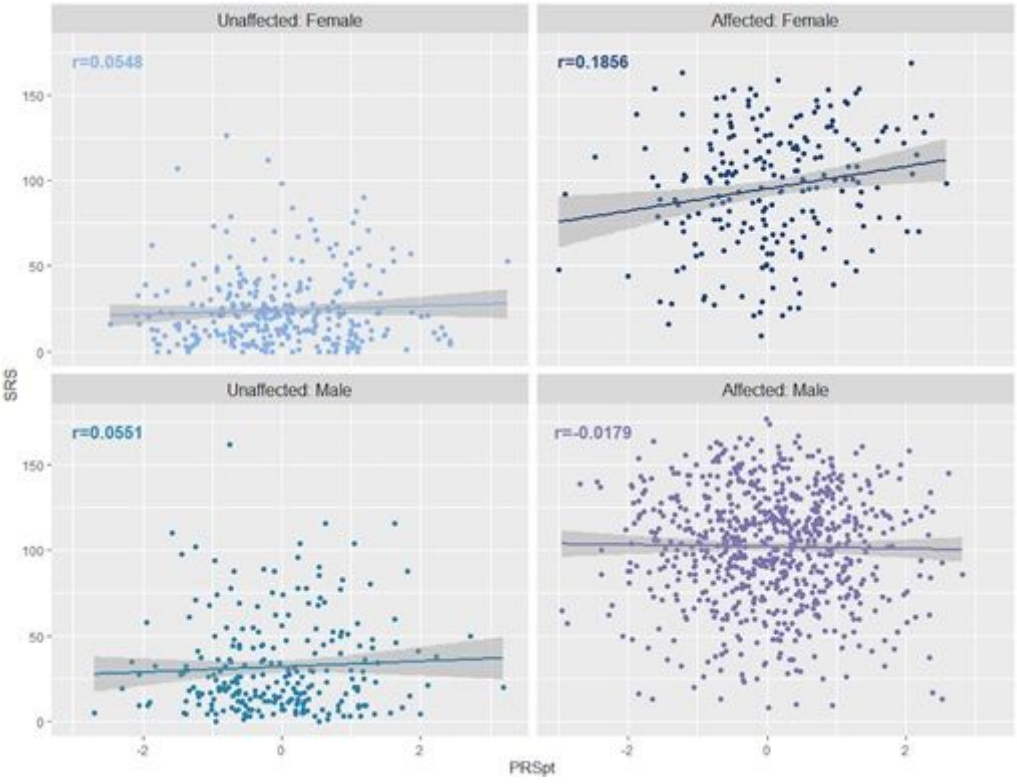

Figure 7

Correlation of PRS with SRS scores, stratified by sex and ASD status