High Body Mass Polygenic Risk in Mothers Enhances De Novo Functional Mutations in Epigenetic and Microtubule Gene Pathways in Their Offspring With Autism Spectrum Disorder

Brian Lohman
University of Utah Health Huntsman Cancer Institute
Andrey A Shabalin
The University of Utah School of Medicine
Andrew Farrell
The University of Utah School of Medicine
Gabor T Marth
The University of Utah School of Medicine
Anna R. Docherty
The University of Utah School of Medicine
Hilary Coon (hilary.coon@utah.edu)
University of Utah https://orcid.org/0000-0002-8877-5446

Research

Keywords: autism, maternal BMI, polygenic risk, epigenetic, microtubule

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-917426/v1

Abstract

Background. Autism Spectrum Disorder (ASD) is a neurodevelopmental diagnosis that encompasses deficits in social communication in addition to repetitive and restrictive behaviors and interests. Accumulated evidence implicates over 100 risk genes and suggests possible genetic subtypes. We tested one previously characterized subtype relating to high maternal body mass index (BMI) as an enhancing risk factor in genetically vulnerable offspring.

Methods. Using 1,300 families from the Simons Simplex Collection (SSC), we created an objectively defined subgroup of mothers in the highest quartile of the distribution of derived BMI polygenic risk scores. Polygenic risk for BMI reflects background genetic risk independent of the many environmental modifiers of BMI.

Results. In the ASD offspring of mothers in this highest quartile, we found significant associations with *de novo*, putatively functional variants in genes in pathways related to chromatin state, chromatin structure, histone activity, and microtubule function. These gene pathways represent potential epigenetic vulnerability to alterations in the metabolic prenatal environment and/or alterations in microtubule-related brain development processes. The observed pathway enrichments were maternal-specific, and were not observed in neurotypical offspring. Two-thirds of the 36 genes in the significant epigenetic pathways and over half of the 33 genes in the significant microtubule pathways had existing ASD or neurodevelopmental risk evidence.

Limitations. Though tests and simulations were done to ensure robustness of results, these findings have not been replicated in an external cohort.

Conclusions. Our results suggest that epigenetic modification and/or microtubule deficits may be unique to a subset of ASD probands of mothers at increased genetic metabolic risk, pending external replication. Beyond the current application of these methods, our approach presents a strategy to reveal genetic subsets through polygenic risk stratification across phenotypic domains.

Background

Autism Spectrum Disorder (ASD) is neurodevelopmental diagnosis arising from complex genetic and environmental risks. Established genetic risk factors include common genetic variation in addition to *de novo* and inherited rare genetic events (An et al. 2018; Werling et al. 2018; Sanders et al. 2015; Yuen et al. 2015; Grove et al. 2019; Satterstrom et al., 2020; Schaaf et al., 2020). Despite genetic discoveries, a considerable portion of variation in risk for ASD remains elusive. In addition, even for established risk genes and gene pathways, identifying specific risk mechanisms remains a significant challenge.

One efficient strategy to begin to close these knowledge gaps may be within reach. Using large, well-characterized data resources, we can identify proband and/or family subsets with particular common background genetic risks using polygenic risk scores (PRS). We can then explore enrichment of

functional *de novo* risk variants in genes and gene pathways within these subsets. The identification of subsets using PRS provides an objective method to characterize underlying genetic risk rather than using measures with strong environmental modifiers or subjective self or caregiver assessments.

To test this strategy, we conducted specific analyses of an established ASD subtype arising from prenatal risk of exposure to maternal metabolic alterations. Given the very early onset of autism, environmental *in utero* risks have been a focus of study. In particular, prior work has repeatedly suggested that increased maternal body mass index (BMI) or other metabolic conditions, such as diabetes, before or during pregnancy results in an increased risk of ASD in offspring (Kong et al. 2018; Windham et al. 2019; Varcin, Newnham, and Whitehouse 2019; Andersen et al. 2018; Wang et al. 2016; M. Li et al. 2016; Lyall et al. 2011; Surén et al. 2014; Li et al. 2016; Connolly et al. 2016; Krakowiak et al. 2012; Bilder et al. 2013; Contu & Hawkes, 2017; Hoirish-Clapauch & Nardi, 2019; Rivella & Mattson 2019). However, this widespread maternal risk is inconsistent with the much rarer outcome of ASD; additional interactions with offspring vulnerability are likely also present. Such vulnerability may be manifest in genome-wide evidence implicating variants in genes in epigenetic pathways (Geschwind & State, 205; Pinto et al. 2014; Grayson & Guidotti, 2016) and in other complementary epigenetic evidence (Grafodatskaya et al. 2010; Rangasamy, et al., 2013; Sun et al. 2016; Wisniowiekcka-Kowalnik & Nowakowska, 2019). This evidence suggests that the prenatal environment, among other environmental risks, may directly modify genetic risk in ASD.

Existing work with PRS has predicted psychiatric case status (e.g., Perkins et al. 2019; Misganaw et al., 2019; Grigoroiu-Serbanescu et al., 2019), including ASD status (Jansen et al., 2019). PRS have also been used to explore genetic associations of ASD with other phenotypes (e.g., Weiner et al., 2017). However, rather than using polygenic risk to predict case status or study cross-trait polygenic associations, we used the maternal PRS for high BMI to objectively define a subgroup of mothers with a higher risk of the common underlying genetic mechanisms leading to BMI. We then tested for functional enrichments of *de novo* mutations in the offspring of mothers with this highest genetically-defined risk for obesity (BMI). Using polygenic risk of BMI rather than measured maternal BMI bypasses a number of confounding factors including age, diet, exercise, socioeconomic and other social factors, previous pregnancies and other medical conditions. While tests of de novo variants included all genome-wide variants with functional annotations, we hypothesized that probands of mothers with high BMI polygenic risk would have significant enrichment of *de novo* variants in gene pathways specific to epigenetic mechanisms.

Methods

Participants. We studied participants in the Simons Simplex Collection (SSC), a collection of 1,651 carefully ascertained families made up of both parents, one affected child, and one unaffected child, each with high quality (at least 30x read depth) whole genome sequence (WGS) data. Because the polygenic risk scores essential to our design are sensitive to ancestry effects, we restricted our analysis to only families of European ancestry. This restriction was done by matching the SSC to the European population in the 1,000 Genomes reference resource (The 1000 Genomes Consortium, 2015) using the

Peddy software package (Pedersen & Quinlan 2017), reducing our sample to 1,300 families. Of these, 1,136 probands were male (87.4%).

Genomic data. All genomic analyses were conducted using whole genome sequence (WGS) data. We combined the two available batches of the Simons Foundation Autism Research Initiative (SFARI) WGS data by converting all genomic coordinates to hg19 with Bioconductor package liftOver using the hg38ToHg19.over.chain chain file provided within the package (https://www.bioconductor.org/packages/devel/workflows/vignettes/liftOver/inst/doc/liftov.html, version 1.8.0). For polygenic risk score (PRS) calculations, variants were extracted and filtered to include only those with minor allele frequency greater than 0.01 using the PLINK software package (Purcell et al., 2007; Chang et al., 2015). Single nucleotide polymorphisms (SNPs) with minor allele frequency discrepancies between batches of over 0.01 were excluded from the PRS calculations. For *de novo* gene enrichment testing in probands and comparison enrichment in unaffected siblings, *de novo* variants in probands and siblings were curated for quality and called using standard GATK (Auwera et al., 2013) and RUFUS (Ostrander et al., 2018) variant calling pipelines, resulting in a total of 216,476 *de novo* putatively functional variants.

Polygenic risk score generation and quartile selection. We obtained genome-wide association study (GWAS) summary statistics from a published study of BMI based on a sample size of 339,224 individuals (Locke et al., 2015). PRS on the SSC participants were then calculated using PRSice2.0 (Choi & O'Reilly, 2019) using p-value thresholds from 0.001 to 1.0. Using regression, PRS were adjusted for sex and for the first 20 principal components of ancestry to account for any residual ancestry stratification. A PRS p-value threshold of 0.2 was selected as optimal by determining the correlations in adult parents between measured BMI and the PRS at each p-value threshold, then selecting the PRS p-value corresponding to the most stable peak (Supplementary Fig. 1). Though sequence batches were carefully combined, we assessed for evidence of residual WGS batch effects by analyzing the distribution of PRS, testing for differences in mean PRS by batch. We observed no significant residual batch differences (Supplementary Fig. 2).

PLACE FIGURE 1 ABOUT HERE

Our PRS selection design is outlined in Fig. 1. We selected mothers in the top quartile of the distribution of maternal PRS for BMI (N = 325). Affected offspring of these selected mothers (and their unaffected siblings as a control test) were then the focus of subsequent GO *de novo* gene pathway enrichment tests (Figure. 2). Similarity of maternal quartile membership was tested varying the PRS p-value threshold from 0.1 to 0.4. Quartile membership was found to be stable (r = 0.7–0.9, Supplementary Fig. 3), demonstrating that quartile membership is robust to our selection of PRS p-value threshold selection.

GO enrichment testing of de novo functional variants. Of the 216,476 curated WGS variants, we selected variants with predicted medium or high functional consequences on gene function using VEP (McLaren et al. 2016). Variants were extracted using the GEMINI database tool (Paila et al. 2013), producing a final

sey of 3,001 putatively functional, genome-wide *de novo* variants. Gene set enrichment analysis was then carried out using the topGO R package (https://bioconductor.org/packages/release/bioc/html/topGO.html). For the probands and siblings defined by the quartile selections described above, we included all genes in the enrichment if they had at least one *de novo* variant that passed our predicted medium to high VEP impact filter. We used annotations provided by the org.Hs.eg.db R package, as suggested by the topGO vignette. We limited our analysis to the Biological Processes ontology, which contains 16,113 GO pathways. We used a node size (minimum number of genes associated with a given GO pathway for it to be included) of 10, which reduced the number of GO pathways we considered to 6,917. We applied a FDR correction to raw p-values generated by topGO (6,917 tests, one for each GO pathway passing the node size filter) with the p.adjust function (base R, https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/p.adjust), referred to hereafter as q-values.

Monte Carlo simulation. To provide a null distribution against which to assess our observed results for enrichment testing, we conducted gene set enrichment analysis (as described above) of *de novo* mutations in randomly selected groups of 325 probands (the same sample size as our selected high maternal BMI PRS quartile). We repeated this process 10,000 times, recording all data provided by topGO during each iteration. We report these null distributions as the mean and two standard deviations of q-values from these 10,000 iterations (multiple testing correction was applied within each iteration) for relevant GO pathways.

Secondary tests to assess specificity of pathway associations. We expected that patterns of GO enrichment that implicate mechanism in ASD would be specific to the probands and not present in the unaffected siblings. Unaffected sibling pathway enrichment was therefore identified using the methods outlined above, and pathways enriched in both probands and unaffected siblings were flagged and omitted, leaving pathways only specific to probands. In addition, our hypothesis focused on enrichment specific to ASD offspring of mothers with high PRS for BMI. It is possible that observed significant pathway enrichment could highlight processes involved in more general metabolic disturbance not specific to high BMI. We therefore also tested for pathway enrichment in probands of mothers in the lowest PRS quartile for BMI and omitted any overlap under the assumption that overlapping pathways likely reflect more general metabolic processes. Specificity of a maternal effect was also part of our hypothesis. We therefore tested for pathway enrichment ranking probands themselves based on their PRS for BMI and selected the top proband quartile. This enrichment would reflect within-proband interaction of background polygenic high BMI risk and *de novo* events, rather than a maternal effect. We therefore omitted any within-proband *de novo* enrichments to retain only those due to maternal risk. Finally, we stratified the sample using PRS for BMI of fathers and tested for GO enrichment in the probands in the top paternal BMI quartile. Any overlapping enrichments were deleted, leaving pathways reflecting only maternal and not paternal specificity of the observed enrichments.

Specificity of stratification on maternal polygenic risk for BMI vs. maternal measured BMI. We have focused on underlying polygenic risk of BMI rather than measured BMI to avoid age effects and other

environmental effects that can substantially confound measured maternal BMI. To test the correspondence of findings using measured BMI rather than underlying polygenic BMI risk, we also ranked mothers by their measured BMI adjusted for age and selected the top quartile for subsequent proband and unaffected sibling *de novo* enrichment tests.

Phenotypic tests. All phenotypic data were provided by the Simons Foundation (https://www.sfari.org/resources/ssc-instruments/). A binomial test was used to determine differences in the sex ratio of probands in the selected highest quartile for maternal BMI PRS compared to other ASD probands. For quantitative phenotypes, comparisons of this quartile to the other probands were done using linear regression models. Quantitative assessments included full scale IQ scores measured using assessments appropriate for age and level of development (Fischbach & Lord 2010), available on 277 cases with high maternal BMI PRS and 830 cases without high maternal BMI PRS. In addition, we tested proband scores on the Social Responsiveness Scale (SRS; Constantino et al., 2003), a quantitative measure of autism severity with demonstrated reliability and validity (Parks, 1983; Bolte et al., 2008; Frazier et al., 2010). We included tests of the total SRS score and the individual domain scores for cognition, awareness, communication, mannerisms, and motivation. We similarly tested two other evidence-based behavioral measures, the Autism Behavior Checklist (ABC; Krug et al., 1980) and the multi-axial Child Behavior Checklist (CBCL; Achenbach, 1999), a measure of global emotional/behavioral problems. All tests were subjected to false discovery rate (FDR) correction for multiple testing.

Results

Overview. Probands who were offspring of mothers in the highest quartile for BMI PRS showed significant enrichment of *de novo* variants in Gene Ontology pathways (GO; The Gene Ontology Consortium, 2019). Significance of pathway enrichment was assessed through extensive simulation analyses. In addition, the resulting significant gene pathways in our study were specific to autism and not present in siblings. They were specific to high maternal BMI, not occurring in probands of mothers in the lowest quartile for BMI. They represented a maternal effect, not occurring when stratification was done either within the probands themselves, or when stratification was done within fathers.

PRS Stratification and de novo enrichment of gene pathways. Tests of GO enrichment in *de novo* variants within probands of mothers in the top quartile of maternal PRS for BMI resulted in 18 significant GO pathways after FDR correction ($q < 0.05$, Fig. 2, Supplementary Table 1). All GO pathway enrichments were more extreme than expected using 10,000 iterations of testing for GO enrichment in randomly selected quartiles of probands (Fig. 2).

PLACE FIGURE 2 ABOUT HERE

Secondary tests of de novo pathway specificity. We found no significant GO pathway enrichment in the unaffected siblings of mothers in the top quartile of PRS for BMI; therefore, no pathways needed to be removed for this test of ASD specificity. Testing for specificity of an effect of high BMI vs. more general

metabolic dysregulation involved assessment of enrichment in probands of mothers in the lowest quartile for BMI PRS. Two GO pathways overlapped our high BMI proband results: cellular component organization, and cellular component organization or biogenesis. These two pathways were removed. To test the hypothesis that significant GO pathways might be due to an interaction with basic metabolic risk in the probands themselves, we assessed for enrichment after stratifying directly on proband PRS for BMI. An additional four significant GO pathways in this stratified top quartile of proband PRS for BMI overlapped with those in the high maternal BMI PRS group: cellular protein modification process, protein modification process, cellular macromolecule metabolic process, and macromolecule modification. We stratified probands by paternal PRS for BMI to test the specificity of a maternal effect as opposed to a more general parental effect. We found no significant enrichment of *de novo* variants in probands of the top quartile of paternal PRS for BMI. Finally, when mothers were stratified by high measured BMI rather than high BMI PRs, we found no significant GO enrichment in the proband offspring in this group. Therefore, no pathways needed to be removed for the test of maternal vs. parental specificity or the test of polygenic vs. measured BMI stratification. For a complete list of GO pathways across maternal and proband quartiles, see Supplementary Tables 1 and 2.

After removing the six pathways described above, there were 12 significant GO pathways specific to probands in the top quartile of maternal PRS for BMI (Table 1). Seven of theses 12 significant pathways included histone modification, covalent chromatin modification, histone lysine methylation, histone methylation, and chromatin organization. In addition, there were two pathways associated with microtubule-based process, and three relatively large pathways reflecting more general processes: transcription, DNA-templated (N of genes = 3603), cytoskeleton organization (N of genes = 1308), and chromosome organization (N of genes = 1193).

Table 1
GO pathways unique to probands of mothers in the top quartile of maternal BMI PRS.

GO ID	Pathway	Annotated	Significant	Expected	p-value	q-value
GO:0006325	chromatin organization	789	32	14.08	1.4e-05	0.0130
GO:0016569	covalent chromatin modification	470	26	8.39	3.7e-07	0.00256
GO:0016570	histone modification	456	25	8.14	7.4e-07	0.00256
GO:0016571	histone methylation	135	11	2.41	3.1e-05	0.0239
GO:0034968	histone lysine methylation	110	11	1.96	4.4e-06	0.00507
GO:0018205	peptidyl-lysine modification	361	18	6.44	9.3e-05	0.0387
GO:0018022	peptidyl-lysine methylation	125	11	2.23	1.5e-05	0.0130
GO:0007017	microtubule-based process	723	32	12.9	2.3e-06	0.00398
GO:0032886	regulation of microtubule-based process	204	13	3.64	8e-05	0.0387
GO:0051276	chromosome organization	1193	40	21.29	9.1e-05	0.0387
GO:0007010	cytoskeleton organization	1308	43	23.34	7.6e-05	0.0387
GO:0006351	Transcription, DNA-templated	3603	92	64.29	0.00013	0.0450

Columns indicate the GO ID, the name of the GO pathway, the number of genes in the GO pathway, the number of genes present in the quartile, the number expected by chance, the p-value for the Fisher's exact test, and the FDR corrected q-value (6,917 tests). Similar/overlapping pathways are grouped, with general pathways listed above more specific pathways. The first seven pathways listed in the table describe epigenetic processes.

PLACE Table 1 ABOUT HERE

Genes in Significant Pathways. To examine the significant GO pathways more closely, we determined the subset of genes within these pathways where *de novo* functional variants specifically occurred in the selected probands. Table 2 shows 62 unique genes among the nine GO pathways with specific functional definitions. There is substantial overlap among these pathways; 11 significant genes are shared by all epigenetic-associated pathways, and 14 significant genes are shared by the two microtubule pathways. Five genes were common to both epigenetic and microtubule pathways: *SETD2, KAT2B, TRIM37, SFPQ*,

and *CHD3*.In addition, 43 of the 62 genes (69%) have existing evidence for ASD or neurodevelopmental risk based on the SFARI Gene score (https://gene.sfari.org/) and/or the published literature.

PLACE TABLE 2 ABOUT HERE

Using STRING (Szklarczyk et al. 2019), an interactive online database tool to assess known physical and functional protein-protein associations, we investigated the types of associations among the implicated genes shown in Table 2. Not surprisingly, we found a highly significant overall number of protein-protein interactions ($p < 1E-16$; Fig. 3). Of these 159 observed connections, 75% had interactions based on experimental evidence. Other interactions were predicted from co-expression, protein homology, and text mining. The figure shows the epigenetic- and microtubule-related clusters of genes, and additionally shows a degree of protein-protein interaction between these clusters. Of the 19 novel genes in Table 2 without existing ASD or neurodevelopmental evidence, all but 10 have protein-protein interactions in the STRING network, suggesting close interactions of half of these novel genes with ASD-associated genes.

PLACE FIGURE 3 ABOUT HERE

136 genes with *de novo* functional variants in the three large GO pathways with general function are shown in Supplementary Table 3; 51 of these overlap with the more specific pathways and are also listed in Table 2. Of the remaining 85 genes, 36 have existing ASD or neurodevelopmental evidence.

Phenotypic correlates of PRS stratification. The top quartile contained 278 males and 47 females, which is not significantly different from the cohort sex ratio (binomial test, $p = 0.4$). Additional tests using linear regression models revealed no statistical differences between the selected proband quartile versus other probands for any of the available phenotypes, including full scale IQ and aspects of autism severity and other behaviors measured using the Social Responsiveness Scale (SRS; Constantino et al., 2003) the Aberrant Behavior Checklist (ABC; Krug et al., 1980), and the Child Behavior Checklist (CBCL; Achenbach, 1999).

As a final assessment of the interpretation of our results, we investigated if the maternal PRS for BMI could be associated with other non-metabolic traits or diagnoses. Tests of PRS for 46 other traits and diagnoses in the mothers of our study cohort showed positive associations with waist-hip ratio ($r = 0.34$, $p = 2.2E-16$) and type 2 diabetes ($r = 0.13$, $p = 2.4E-16$), and a negative association with HDL cholesterol ($r = -0.14$, $p = 5.4E-7$). No other associations were significant (see Supplementary Fig. 5).

Discussion

Ongoing discovery and replication efforts for genetic risks for ASD face substantial challenges due to genetic heterogeneity and complexity. Previous efforts to reduce heterogeneity through subsetting have relied upon phenotypic classification. Our approach offers an alternative method, stratifying instead on underlying polygenic liability, hypothesizing that this aggregation may enhance the signal of *de novo* risk gene pathways in a subgroup having increased homogeneity through shared polygenic background risk.

We have initially applied this strategy to study offspring of mothers with increased BMI as an ASD risk subgroup with previous extensive research evidence. Using molecular resources available in the Simons Simplex Collection (SSC) we characterized maternal polygenic risk of high BMI. Using underlying polygenic risk avoids potential substantial modifying environmental effects that can confound measured BMI. Existing SSC sequencing data additionally allows tests for enrichment of putatively functional *de novo* risk variants in the selected probands of mothers with the highest polygenic BMI risk.

This stratification scheme revealed a significant excess of *de novo* variants involved with chromatin state/structure, histone activity, and microtubule function in probands whose mothers were in the top quartile of genetic risk for high BMI. Significance was assessed using 10,000 simulations of randomly selected quartiles. These pathways were absent in unaffected siblings. Pathways not specific to maternal effects or to specific effects of high BMI as opposed to overall metabolic changes were eliminated. Significant enrichment of pathways was also not exhibited when stratification was done using measured BMI. While prior large epidemiological studies have implicated measured BMI, rather than polygenic risk of BMI, as a prenatal risk factor for ASD, measured BMI is strongly modified by many environmental factors. The specificity of our results to polygenic BMI risk stratification suggests underlying genetic risk, decoupled from potentially strong modifying environmental influences, may have been required to observe the proband *de novo* gene pathway enrichment.

The functional enrichment in the top quartile of maternal PRS for BMI suggests that background maternal genetically-driven metabolic risk may interact with *de novo* variants preferentially in epigenetic and microtubule-associated pathways in offspring to influence ASD risk. While our initial hypothesis predicting the enrichment of epigenetic gene pathways was correct, pathway enrichment relating to microtubule processes was unexpected. However, microtubule dynamics, which are directly connected with neuronal development and function, have also been implicated in autism (Chang et al., 2018; Satterstrom et al., 2019) and other neurodevelopmental disorders, including intellectual disability (Lasser et al., 2018).

We looked specifically at the subset of genes driving the pathway significance, where *de novo* putatively functional variants occurred in the selected proband subset. While we imposed no prior selection of *de novo* variants beyond functional annotations, two-thirds of these implicated genes showed existing ASD evidence, in addition to the specific associations with enriched pathways. Additional published evidence summarized in Table 2 shows that these genes additionally have multiple prior associations with other neurodevelopmental disorders (Kim et al., 2017) and syndromes, with other psychiatric disorders (including schizophrenia and psychosis, anxiety, attention deficit hyperactivity disorder, bipolar disorder, post-traumatic stress disorder, and Alzheimer's disease), and with neurological processes (including memory, neurodevelopment, and neurodegeneration).

Interpretation of our results relies upon the assumption that maternal PRS for BMI is associated with BMI and other metabolic traits and not with other medical or behavioral traits. Our tests of multiple other polygenic risks (Supplementary Fig. 5) confirmed this assumption, producing results consistent with

other findings (Zheng et al., 2017; Docherty et al. 2018; Krapohl et al. 2017). The specificity of PRS for BMI to polygenic liability of other metabolic condisions suggests that results based on stratifying by maternal PRS are indeed driven by maternal genetic risk for metabolic disorders.

The identified subgroup of probands showed no significant differences in any clinically relevant measured phenotype, including sex ratio, overall IQ, quantitative ASD severity as measured with the SRS and with the ABC, aspects of the ASD clinical phenotype as measured by SRS domain scores, or additional emotional/behavioral attributes measured by the CBCL. Therefore, while the quartile in this study shows significant enrichment of specific genetic pathways, obvious phenotypic differences were not apparent.

Limitations

Our results suggest prior findings implicating epigenetic mechanisms and microtubule dysfunction may be augmented by maternal genetic metabolic risk. While our approach included Monte Carlo simulation methods and multiple secondary tests to ensure accuracy of this interpretation, replication in an independent cohort is required to allow results to be interpreted beyond the SSC cohort. In addition, because polygenic risks scores are sensitive to ancestry effects, we confined our analyses to families of European ancestry giving findings that may not be generalizable to individuals of other ancestries.

Conclusions

Our approach uses stratification of a subgroup with shared background polygenic risk as a strategy to reduce heterogeneity, then testing for interaction of this inherited polygenic risk and discrete *de novo* functional risk variants. This approach to use PRS as genetic phenotype to reduce heterogeneity focused investigations of *de novo* pathway enrichment is worthy of additional exploration for ASD and other complex psychiatric phenotypes.

We applied this method to study the well-documented epidemiological risk association between prenatal maternal high BMI and ASD risk. We found significant enrichment of *de novo* variants in the probands of mothers who were at greatest polygenic risk for high BMI. The significant GO pathways predominantly pertain to chromatin state/structure, histone activity, and microtubule function. Broadly, the chromatin and histone pathways are functional pathways that are targets of epigenetic modification, a potential mechanism of interaction between mother and fetus. The microtubule pathways suggest potential alterations in specific early developmental processes critical for normal brain development. Of interest, our results additionally show modest overlap between the epigenetic and microtubule pathways. These enriched gene pathways were unique to ASD probands in this specifically defined subgroup, implicating epigenetic mechanisms and alteration in microtubule function, potentially definable risks that may be amenable to early intervention. Pending replication, our results both advance our understanding of interactions between genetic background and individual variants in ASD cases and serve as a proof of concept for this method of PRS stratification.

Abbreviations

ABD: Autism Behavior Checklist

ASD: Autism Spectrum Disorder

BMI: body mass index

CBCL: Child Behavior Checklist

GWAS: genome-wide association study

PRS: polygenic risk score

SFARI: Simons Foundation Autism Research Initiative

SNP: single nucleotide polymorphism

SRS: Social Responsiveness Scale

SSC: Simons Simplex Collection

WGS: whole genome sequence

Declarations

Ethics approval and consent to participate. This study was approved by the University of Utah (IRB_00006042, Genetics of Autism). All analyses were done using de-identified data available through SFARI-BASE.

Consent for publication. All authors have read this manuscript and consent to its publication. The study hypothesis was pre-registered November 25, 2019 in Open Science Framework (OSF, https://osf.io/k472p).

Competing interests. GTM is co-founder and Chief Scientific Officer of Framshift Labs, Inc. All other authors declare that they have no competing interests.

Authors' contributions. The study design was conceived by HC with significant input from ARD and GTM. Statistical analyses were performed by BL, AAS, ARD, and AF. HC wrote the manuscript with significant contributions from BL, ARD, and GTM. Genomic analyses of sequence data used this study was done by AF, ARA, and GTM.

Funding. The work was supported by SFARI Grant 513631 (HC, GTM, ARD). This funding support included access to molecular and phenotypic data used in the study, and partial support of the authors' time, allowing the analyses to be completed and the work to be written by the authors. The work was

conceived, analyses performed, results interpreted, and findings written up without influence from the funding agency.

Acknowledgements. The authors gratefully acknowledge the use of whole genome data on SSC participants processed by the labs of Drs. Michael E. Talkowsky (Massachusetts General Hospital, Boston, MA) and Stephan Sanders (University of California, San Francisco, CA). The authors further acknowledge input on study design and analyses given by Dr. Aaron Quinlan (University of Utah). Finally, the authors sincerely thank the families whose data are included in the Simons Simplex Collection without whom this research would have been impossible.

Availability of data. All data in this manuscript are available through the Simons Foundation. The dataset supporting the conclusions of this article are available in the Simons Foundation repository, https://www.sfari.org/resource/simons-simplex-collection/

References

1. Achenbach TM. The Child Behavior Checklist and related instruments. In: Maruish ME, editor. The use of psychological testing for treatment planning and outcomes assessment. Lawrence Erlbaum Associates Publishers; 1999. pp. 429–66.
2. An J-Y, Lin D, Zhu L, Werling DM, Dong S, Brand H, Wang HZ, et al. Genome-Wide de Novo Risk Score Implicates Promoter Variation in Autism Spectrum Disorder. Science. 2018;362(6420):eaat6576. https://doi.org/10.1126/science.aat6576.
3. Anney RJL, Kenny EM, O'Dushlaine C, Yaspan BL, Parkhomenka E, Buxbaum JD, Sutcliffe J, et al. Gene-Ontology Enrichment Analysis in Two Independent Family-Based Samples Highlights Biologically Plausible Processes for Autism Spectrum Disorders. Eur J Hum Genet. 2011;19(10):1082–89.
4. Bilder DA, Bakian AV, Viskochil H, Clark EAS, Botts EL, Smith KR, Pimentel R, McMahon WM, Coon H. Maternal Prenatal Weight Gain and Autism Spectrum Disorders *Pediatrics*. 2013, 132 (5): e1276–83.
5. Bishop SL, Farmer C, Thurm A. Measurement of Nonverbal IQ in Autism Spectrum Disorder: Scores in Young Adulthood Compared to Early Childhood. J Aut Dev Disord. 2015;45(4):966–74.
6. Bolte S, Poutka F, Constantino JN. Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). Autism Res. 2008 Dec;1(6):354–63.
7. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigoScience. 2015;4(1):s13741. -015-0047-8..
8. Chang Q, Yang H, Wang M, Wei H, Hu F. Role of microtubule-associated protein in autism spectrum disorder. Neurosci Bull. 2018 Dec;34(6):1119–26.
9. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score Software for Biobank-Scale Data. *GigaScience.* 2019, 8 (7).

10. Connolly N, Anixt J, Manning P, Ping -I, Lin D, Marsolo KA, Bowers K. Maternal Metabolic Risk Factors for Autism Spectrum Disorder—An Analysis of Electronic Medical Records and Linked Birth Data. Autism Research. 2016;9(8):829–37.
11. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, Metzger LM, et al. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the Autism Diagnostic Interview-Revised. J Aut Dev Disord. 2003;33(4):427–33.
12. Contu L, Hawkes CA. A review of the impact of maternal obesity on the cognitive function and mental health of the offspring. *Int J Mol Sci.* 2017 May 19;18(5):pii.E1093.
13. Crawley JN, Heyer W-D, LaSalle JM. Autism and Cancer Share Risk Genes, Pathways, and Drug Targets. Trends Genet. 2016;32(3):139–46.
14. Eicher JD, Gruen JR. Language impairment and dyslexia genes influence language skills in children with autism spectrum disorders. Autism Res. 2015 Apr;8(2):229–34.
15. Farmer A, Korszun A, Owen MJ, Craddock N, Jones L, Jones I, Gray J, et al. Medical Disorders in People with Recurrent Depression. Br J Psychiatry. 2008;192(5):351–55.
16. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron. 2010;68(2):192–95.
17. Frazier TW, Youngstrom E, Sinclair L, Kubu CS, Law P, Rezai A, Constantino JN, Eng C. Autism spectrum disorders as a quantitatively distinct category from typical behavior in a large, clinically ascertained sample. Assessment. 2010. Sep;17(3):308–20.
18. The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. New Eng J Med. 2017;377(1):13–27.
19. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still going strong. Nucleic Acids Res. 2019;47(D1):D330–8.
20. Geschwind DH, State MW. Gene hunting in Autism Spectrum Disorder: on the path to precision medicine. The Lancet Neurology. 2015;14(11):1109–20.
21. Godfrey KM, Reynolds RM, Prescott SL, Nyirenda M, Jaddoe VWV, Eriksson JG, Broekman BFP. Influence of maternal obesity on the long-term health of offspring. Lancet Diabetes Endocrinol. 2017;5(1):53–64.
22. Grafodatskaya D, Chung B, Szatmari P, Weksberg R. Autism Spectrum Disorders and Epigenetics. J Am Acad Child Adolesc Psychiatry. 2010;49(8):794–809.
23. Grayson DR, Guidotti A. Merging data from genetic and epigenetic approaches to better understand autistic spectrum disorder. Epigenetics. 2016 Jan;8(1):85–104.
24. Grigoroiu-Serbanescu M, Giaroli G, Thygesen JH, Shenyan O, Bigdeli TB, Bass NJ, Diaconu CC, et al. Predictive power of the ADHD GWAS 2019 polygenic risk scores in independent samples of bipolar patients with childhood ADHD. *J Affect Disord.* 2019 Nov 23.pii S0165-0327(19)32310-9.
25. Grondhuis SN, Lecavalier L, Arnold LE, Handen BL, Scahill L, McDougle CJ, Aman MG. Differences in verbal and nonverbal IQ test scores in children with Autism Spectrum Disorder. Res Aut Spect Disord.

2018;49(May):47–55.

26. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J, et al. Identification of common genetic risk variants for Autism Spectrum Disorder. Nat Genet. 2019;51(3):431–44.
27. Hioiisch-Clapauch S, Nardi AE. Autism spectrum disorders: let's talk about glucose? Trasl Psychiat. 2019 Jan;31(1):51. 9(.
28. Jansen AG, Dieleman GC, Jansen PR, Verhulst FC, Posthuma D, Polderman TJC. Psychiatric polygenic risk scores as predictor for attention deficit/hyperactivity disorder and autism spectrum disorder in a clinical child and adolescent sample. *Behav Genet.* 2019 Jul 25.
29. Kim J-H, Lee JH, Lee I-S, Lee SB, Cho KS. Histone lysine methylation and neurodevelopmental disorders. Int J Mol Sci. 2017;Jul;18(7):1404.
30. King JC. Maternal obesity, metabolism, and pregnancy outcomes. Ann Rev Nutrition. 2006;26(1):271–91.
31. Krakowiak P, Walker CK, Bremer AA, Baker AS, Ozonoff S, Hansen RL, Hertz-Picciotto I. Maternal metabolic conditions and risk for autism and other neurodevelopmental disorders. Pediatrics. 2012;129(5):e1121–28.
32. Krug DA, Arick J, Almond P. Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior.J Child Psychol Psychiatry. 1980 Jul; 21(3):221–9.
33. Lasser M, Tiber J, Lowery LA. The role of the microtubule cytoskeleton in neurodevelopmental disorders. Front Cell Neurosci. 2018 Jun;14:12:165.
34. Li Y-M, Ou J-J, Liu L, Zhang D, Zhao J-P, Tang S-Y. Association Between Maternal Obesity and Autism Spectrum Disorder in Offspring: A Meta-Analysis. J Aut Dev Disord. 2016;46(1):95–102.
35. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, et al. Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology. *Nature.* 518 (7538): 197–206. https://doi.org/10.1038/nature14177.
36. Loomes R, Hull L, Polmear W, Locke M. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2017;56(6):466–74. https://doi.org/10.1016/j.jaac.2017.03.013.
37. Lyall D, Pauls DL, Santangelo S, Spiegelman D, Ascherio A. Maternal Early Life Factors Associated with Hormone Levels and the Risk of Having a Child with an Autism Spectrum Disorder in the Nurses Health Study II. J Aut Dev Disord. 2011;41(5):618–27.
38. Misganaw B, Guffanti G, Lori A, Abu-Amara D, Flory JD, Meuller SBPBC. S, et al. Polygenic risk associated with post-traumatic stress disorder onset and severity. Transl Psychiatry. 2019 Jun;7(1):165. 9(.
39. Ostrander BEP, Butterfield RJ, Pedersen BS, Farrell AJ, Layer RM, Ward A, Miller C, et al. Whole-genome analysis for effective clinical diagnosis and gene discovery in early infantile epileptic encephalopathy. *Genome Med* 2018 Aug;3(22).

40. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. PLOS Comput Biol. 2013;9(7):e1003153.
41. Parks SL, The Assessment of Autistic Children. A Selective Review of Available Instruments. J Autism Dev Disord. 1983 Sep;13(3):255–67.
42. Pedersen BS, Quinlan AR. 2017. "Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum Genet.* 2017, 100 (3): 406–13.
43. Perkins DO, Olde Loohuis L, Barbee J, Ford J, Jeffries CD, Addington J, Bearden CE, et al. Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. Am J Psychiatry. 2019;Nov12:appiajp201918060721.
44. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am J Hum Genet. 2014;94(5):677–94.
45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81(3):559–75.
46. Rangasamy S, D'Mello SR, Narayanan V, Epigenetics. Autism Spectrum, and Neurodevelopmental Disorders. Neurotherapeutics. 2013;10(4):742–56.
47. Rivelli A, Mattson MP. Intergenerational metabolic syndrome and neuronal network hyerexcitability in autism. Trends Neurosci. 2019 Oct;42(10):709–26.
48. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, Peng M, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180(3)(Feb 6):568–84.
49. Satterstom FK, Walters RK, Singh T, Wigdor EM, Lescal F, Demontins D, Kosmicki JA, et al. Autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) have a similar burden of rare protein-truncating variants. Nat Neurosci. 2019 Dec;22(12):1961–5.
50. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215–33.
51. Schaaf CP, Betancur C, Yuen RKC, Parr JR, Skuse DH, Gallagher L, Bernier RA, et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat Rev Genet.* 2020, Apr21.
52. Sun W, Poschmann J, Cruz-Herrera del Rosario R, Parikshak NN, Hajan HS, Kumar V, Ramasamy R, et al. Histone Acetylome-Wide Association Study of Autism Spectrum Disorder. Cell. 2016;167(5):1385–97.e11.
53. Surén P, Gunnes N, Roth C, Bresnahan M, Hornig M, Hirtz D, Lie KK, et al. Parental Obesity and Risk of Autism Spectrum Disorder. Pediatrics. 2014;133(5):e1128–38.
54. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. Nucleic Acids Res. 2019;47(D1):D607–13.

55. The 1000 Genomes Consortium. A global reference for human genetic variation. *Nature* 2015 526;68–74.
56. van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols Bioinformatics*, 2013, 43 (1): 11.10.1–11.10.33.
57. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, Turner TN, et al. De Novo Genic Mutations among a Chinese Autism Spectrum Disorder Cohort. Nat Comm. 2016;7(1):1–10.
58. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, et al. Polygenic Transmission Disequilibrium Confirms That Common and Rare Variation Act Additively to Create Risk for Autism Spectrum Disorders. Nat Genet. 2017;49(7):978–85.
59. Werling DM, Brand H, An J-Y, Stone MR, Zhu L, Glessner JT, Collins RL, et al. An Analytical Framework for whole-genome sequence association studies and its implications for Autism Spectrum Disorder. Nat Genet. 2018;50(5):727–36.
60. Wisniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder: current evidence in the field. J Appl Genet. 2019 Feb;60(1):37–47.
61. Yuen RKC, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, Chrysler C, et al. Whole-genome sequencing of quartet families with Autism Spectrum Disorder. Nat Med. 2015;21(2):185–91.
62. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272–9.

Tables

Due to technical limitations, table 2 is only available as a download in the Supplemental Files section.

Figures

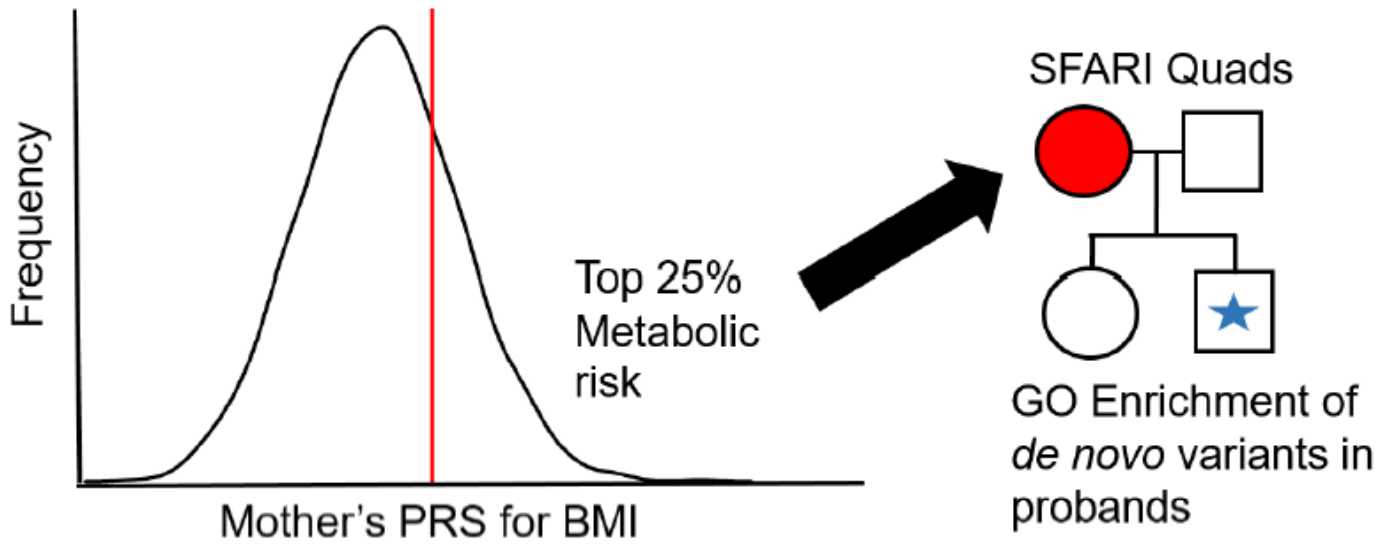

Figure 1

Visual representation of the study design. We ranked mothers based their polygenic risk score (PRS) for BMI, selected the top 25% (marked with shaded red circle), performed GO enrichment in their affected offspring (probands; marked with a blue star). We compared enrichment to a null distribution of 10,000 randomly-selected quartiles to determine significance. We additionally compared enrichment in probands to enrichment in unaffected siblings to determine specificity to ASD risk.

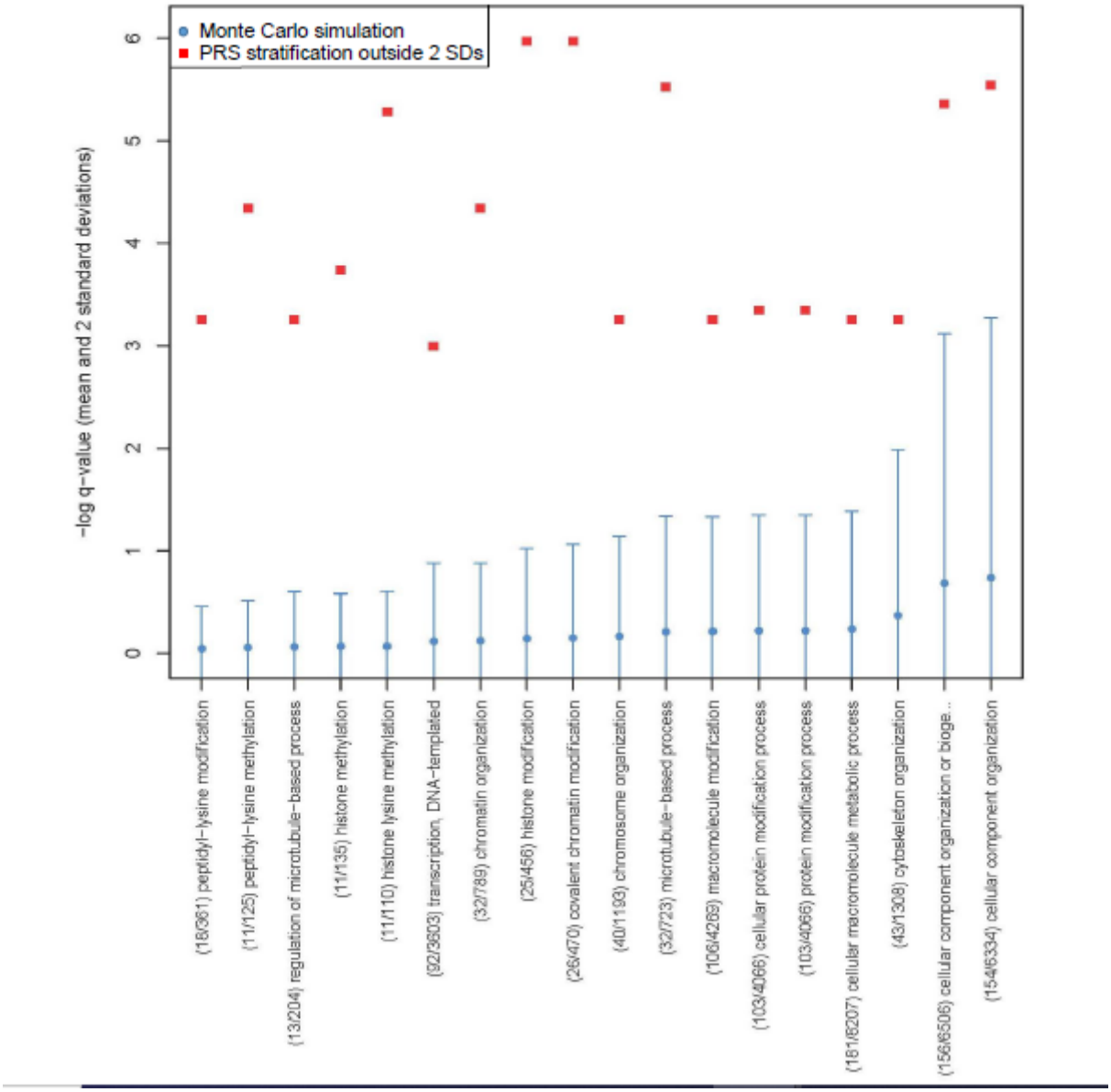

Figure 2

GO enrichment of deleterious de novo variants within probands of top BMI PRS quartile mothers. Only probands of European ancestry were tested. Red squares indicate GO enrichments in these selected probands. Blue squares and lines indicate the mean and two standard deviations of 10,000 random samplings of 325 SSC probands. The total number annotated genes followed by number of genes observed with medium or high impact de novo variants in the top quartile of maternal PRS for BMI are given adjacent to each GO pathway name along the x-axis.

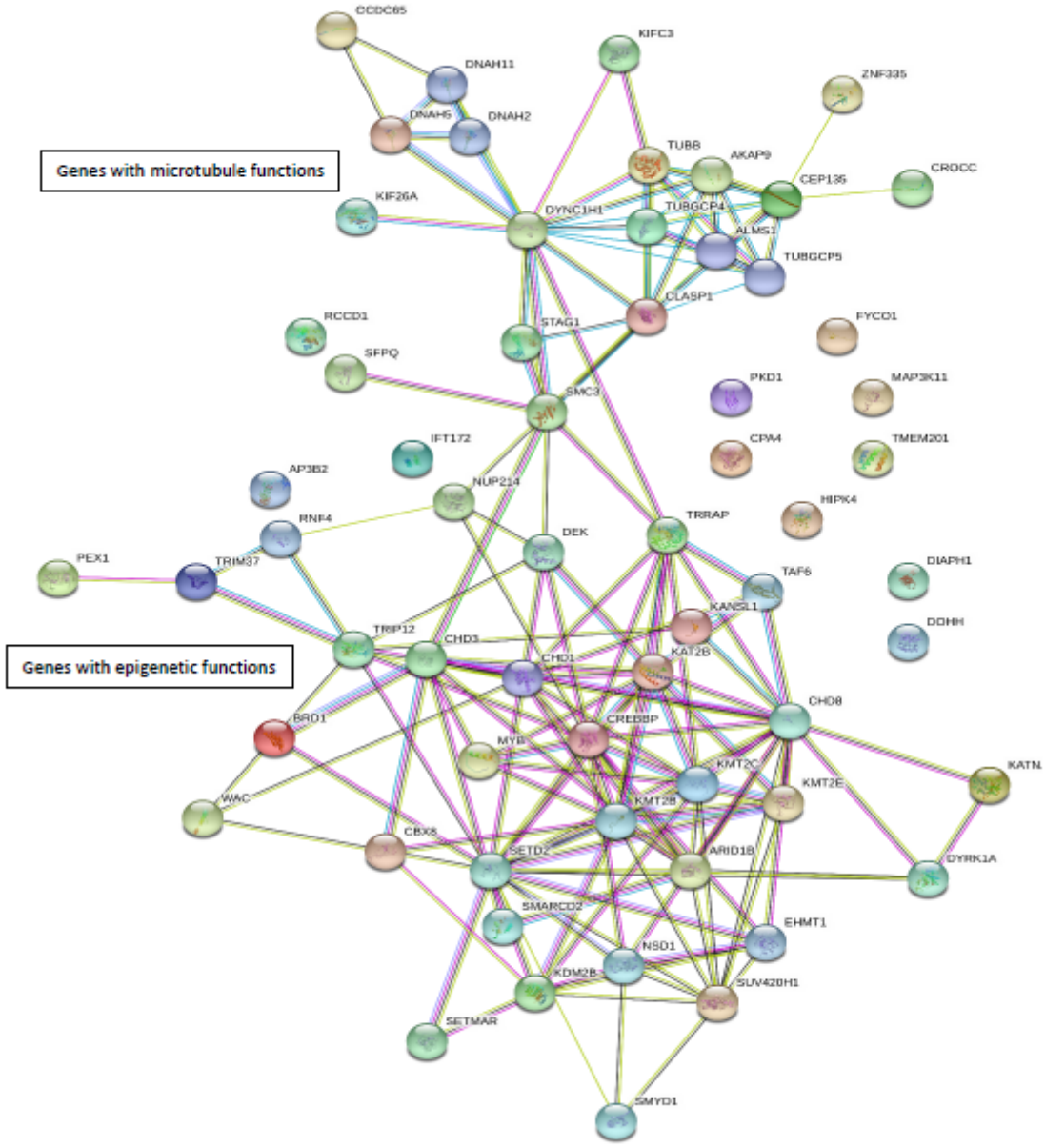

Figure 3

Protein-protein interactions of genes with deleterious de novo variants in 9 significant GO pathways. The analysis applied the STRING (Szklarczyk et al., 2019) software tool to genes in the first 9 significant GO pathways listed in Table 2. Red lines indicate the presence of fusion evidence, green lines indicate neighborhood evidence, blue lines indicate co-occurrence evidence, purple lines indicate experimental evidence, yellow lines indicate text mining evidence, light blue lines indicate database evidence, and black lines indicate co-expression evidence.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BMIPRSsupplementsubmit.docx

- Table2submit.xlsx