Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 November 15.

Published in final edited form as:

Nat Genet. 2017 July ; 49(7): 978–985. doi:10.1038/ng.3863.

Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders

Daniel J. Weiner[1,2,3], **Emilie M. Wigdor**[1,2,3], **Stephan Ripke**[1,2,3,4], **Raymond K. Walters**[1,2], **Jack A. Kosmicki**[1,2,3,5], **Jakob Grove**[6,7,8,9], **Kaitlin E. Samocha**[1,2,3], **Jacqueline Goldstein**[1,2,3], **Aysu Okbay**[10], **Jonas Bybjerg-Grauholm**[7,11], **Thomas Werge**[7,12,13], **David M. Hougaard**[11], **Jacob Taylor**[1,2,3,14], **iPSYCH-Broad Autism Group**[15], **Psychiatric Genomics Consortium Autism Group**[15], **David Skuse**[16], **Bernie Devlin**[17], **Richard Anney**[18], **Stephan J. Sanders**[19], **Somer Bishop**[19], **Preben Bo Mortensen**[6,7,20], **Anders D. Børglum**[6,7,8], **George Davey Smith**[21], **Mark J. Daly**[1,2,3], and **Elise B. Robinson**[1,2,3,*]

[1]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [4]Department of Psychiatry and Psychotherapy, Charité, Campus Mitte, Berlin, Germany [5]Program in Genetics and Genomics, Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, USA [6]Department of Biomedicine (Human Genetics), Aarhus University, Aarhus, Denmark [7]Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark [8]Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark [9]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark [10]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, 1081 HV, The Netherlands [11]Danish Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum

*Please address correspondence to: erob@broadinstitute.org.

Data Availability Statement

The availability of the GWAS summary statistics we used to calculated PRS is detailed in Supplementary Table 3. The SSC genotype data is publically available through application to the Simons Foundation. The *de novo* variant calls from the SSC are available in Iossifov, I., *et al.*.[4] For questions about the PGC ASD genotype data, contact mjdaly@atgu.mgh.harvard.edu.

Author Contributions

D.J.W., E.M.W., S.R., R.K.W., J.A.K., J.G., K.E.S., J.G., A.O., J.B.-G., T.W., D.M.H., R.A., and S.J.S. generated data and/or conducted analyses. E.B.R., M.J.D., D.J.W., S.B., and G.D.S. designed the experiment and tools. D.S., B.D. and J.T. aided in interpretation of the data. E.B.R., M.J.D., P.B.M., and A.D.B. supervised the research. E.B.R., D.J.W., and E.M.W. wrote the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to report.

URLs

pTDT software, https://github.com/ypaialex/ptdt; Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org; Psychiatric Genomics Consortium (PGC), https://www.med.unc.edu/pgc/results-and-downloads

Code Availability Statement

pTDT software available, https://github.com/ypaialex/ptdt

Institut, Copenhagen, Denmark [12]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Copenhagen, Copenhagen, Denmark [13]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark [14]Department of Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts, USA [15]A list of members and affiliations appears at the end of the paper [16]Behavioural Sciences Unit, Institute of Child Health, University College London, London, UK [17]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA [18]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, Wales, UK [19]Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California, 94158, USA [20]National Centre for Register-based Research, University of Aarhus, Aarhus, Denmark [21]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Abstract

Autism spectrum disorder (ASD) risk is influenced by common polygenic and *de novo* variation. We aimed to clarify the influence of polygenic risk for ASDs and to identify subgroups of ASD cases, including those with strong acting *de novo* variants, in which polygenic risk is relevant. Using a novel approach called the polygenic transmission disequilibrium test, and data from 6,454 families with a child with ASD, we show that polygenic risk for ASDs, schizophrenia, and greater educational attainment is over transmitted to children with ASDs. These findings hold independent of proband IQ. We find that polygenic variation contributes additively to risk in ASD cases who carry a strong acting *de novo* variant. Lastly, we show that elements of polygenic risk are independent and differ in their relationship with phenotype. These results confirm that ASDs' genetic influences are additive and suggest they create risk through at least partially distinct etiologic pathways.

Introduction

Risk for autism spectrum disorders (ASDs) is strongly genetically influenced, and reflects several types of genetic variation[1–7]. Common polygenic variation, distributed across the genome, accounts for at least 20% of ASD liability[2, 5, 8, 9]. *De novo* single nucleotide and copy number variants can strongly affect the individuals who carry them[1, 3, 4], but account for less liability at a population level (< 10%)[2]. Over the last several years, the common polygenic and *de novo* influences on ASD risk have been increasingly well characterized, particularly in terms of the distribution of their phenotypic effects. Most consistently, ASD-associated *de novo* variants have been strongly linked to intellectual disability, as well as other indicators of global neurodevelopmental impact (e.g., seizures; motor delay)[1, 10]. Indeed ASD-associated *de novo* mutations that yield protein truncations are far more commonly observed in global developmental delay than in autism itself[11].

Recent studies have suggested that the common polygenic component of ASDs has a different, perhaps surprising, relationship with cognition. Polygenic risk for ASDs has been positively associated with intelligence and educational attainment in several reports[6, 12–14].

In other words, in the general population, greater common variant risk for ASDs is associated with higher IQ. These findings are difficult to interpret – on average, IQ in ASD is at least a standard deviation below the population mean[1, 15]. Further, ASDs share approximately 25% of their common variant influences with schizophrenia, and schizophrenia itself shows a negative genetic correlation with IQ[6, 14]. While genetic correlation analyses are not expected to be transitive, this particularly complex network of common variant associations has led to concerns about confounding resulting from ascertainment and case heterogeneity[16].

Here we attempt to clarify the influence of common variant risk for ASDs, and to better understand the subgroups of ASD individuals for whom polygenic variation is risk contributing. We thus sought to employ a robust family-based design that would be immune to many of the potential confounders that can arise in attempting to construct a well-matched case-control comparison. To advance this analysis we extended the transmission disequilibrium approach to encompass polygenic risk scores; we call the resulting methodology the polygenic transmission disequilibrium test (pTDT). Using pTDT we then show that common polygenic predictors of (i) ASDs, (ii) schizophrenia, and (iii) years of educational attainment are unambiguously associated with ASD risk, independent of the presence of intellectual disability in cases. We find that common polygenic variation still contributes to ASD risk in cases that carry a very strong acting *de novo* event. Lastly, we find that the three aforementioned polygenic risk factors have independent and distinct effects on phenotypic heterogeneity in ASD, suggesting that components of common polygenic variation also behave additively and operate through at least partially distinct etiologic pathways.

Children are expected to inherit half of their parents' risk alleles for a trait. This expectation forms the basis of several commonly used tests of genetic association. The classic transmission disequilibrium test, for example, examines the frequency with which single genetic variants are transmitted from parents to their children[17]. Variants transmitted significantly more than half of the time from unaffected parents to children affected with some trait or condition are nominated for association with that trait – only the ascertainment on a trait in the offspring and the association of the allele with the trait introduces deviation from the 50-50 chance of inheriting either allele from a heterozygous parent. Transmission disequilibrium tests have several convenient properties. First, they are immune to confounding by ancestry. They are also less vulnerable to bias from other potential differences between cases and controls, such as socioeconomic background or other factors commonly related to case ascertainment since the 'controls' are, in effect, the perfectly matched untransmitted chromosomes.

In this study, we extend the transmission disequilibrium test to polygenic risk scores and introduce the polygenic transmission disequilibrium approach. A polygenic risk score (PRS) provides a quantitative measure of an individual's genome-wide common variant predisposition (or "risk") for a trait (**Online Methods: *Polygenic Risk Scoring***). PRS are normally distributed in the population, which means that some degree of common variant risk for complex traits like ASDs is present in all of us. Since a child has a 50% chance of inheriting either allele from a heterozygous parent, it is algebraically defined that the

expected value of a child's PRS for any trait will be equal to the average of their parents' PRS (mid-parent PRS). This expectation is broken when offspring are specifically ascertained for a phenotypic deviation from their parents (Supplementary Figure 1), for example children who are much taller than their parents or children affected with ASD when their parents are not. In the example of ASD, we would then expect the affected child to have received more of their parents ASD risk alleles than expected by chance[9]. From that, the average offspring PRS for an ASD-associated trait would be greater than the average mid-parent PRS. By comparing ASD individuals' PRS for various traits against those of their unaffected parents, one can unambiguously associate specific types of common variant risk (e.g., polygenic risk for schizophrenia, educational attainment) with ASDs, and query the subsets of cases in which those risk factors are most relevant.

We used two independent ASD family cohorts to examine transmission of polygenic risk (**Online Methods: *Sample Description***; Supplementary Table 1). The Simons Simplex Collection (SSC) is a resource of more than 2,500 families with a single child diagnosed with ASD[18]. No other family members to the level of first cousins had an ASD diagnosis. Genotype data were available for the parents, the affected child, and an unaffected sibling for 2,091 SSC families (quad families). An additional 493 SSC families had available data for the parents and the affected child alone (trio families). Most genotyped individuals in SSC were exome sequenced in previous studies (89.0%)[4]. Our independent Psychiatric Genomics Consortium ASD (PGC ASD) sample consisted of 3,870 genotyped parent-child trios from the Psychiatric Genomics Consortium Autism Group (Supplementary Table 2). The PGC ASD cohort described here does not include individuals from the SSC, and the data set does not include exome sequence information.

We calculated common polygenic risk for ASDs, educational attainment (EA) and schizophrenia (SCZ) for all genotyped family members in the SSC and PGC ASD datasets using a standard approach (**Online Methods: *Polygenic Risk Scoring***)[19]. In addition to polygenic risk for ASD itself, we chose to examine polygenic risk for EA and SCZ as those phenotypes have been most strongly linked to ASD in genome-wide genetic correlation studies[6, 14, 20]. To build the polygenic predictors, we used summary statistics from the largest independent genome-wide association studies of each phenotype (Supplementary Table 3). The discovery GWAS (including ASD) did not include any SSC or PGC ASD individuals. Using the calculated polygenic risk scores, we first examined properties of their distribution in the parents. Analyzing the SSC and PGC ASD cohorts separately, we did not observe any consistent correlations between the parents' PRS, for any pair of PRS, and based on that no PRS-based evidence of assortative mating (Supplementary Note 1, Supplementary Tables 4, 5). We also found no evidence that the mothers and fathers of children with ASD differ with regard to the common variant risk that they carry (Supplementary Note 2, Supplementary Table 6). Lastly, there was no evidence that mid-parent polygenic risk differs by proband sex (Supplementary Note 3, Supplementary Table 7).

The pTDT is a t-test asking whether the mean of the offspring PRS distribution is consistent with its parentally-derived expected value (**Online Methods: *pTDT***). In brief, for each trio, one averages the parent PRS for a given trait to generate the mid-parent value, and then

subtracts that value from the proband's PRS for the same trait. To standardize and improve interpretability, we divide the resulting difference by the standard deviation of the (observed) mid-parent distribution. This yields the estimated pTDT deviation, and the pTDT then tests whether the average pTDT deviation across all offspring differs from zero. A mean pTDT deviation of 0.1 for the ASD PRS, for example, would indicate that the offspring's ASD PRS is on average 0.1 standard deviations higher than that of their parents.

Results

Polygenic Transmission Disequilibrium

For each of ASD, SCZ, and EA, we calculated the average pTDT deviation for each affected child in the SSC and PGC ASD, as well as for each unaffected sibling in the SSC. The primary pTDT results are shown in Figure 1a. In both the SSC and PGC ASD samples, polygenic risk for each of ASD, EA, and SCZ was significantly over transmitted to affected children ($P < 1E\text{-}06$ for all comparisons), but not to SSC unaffected siblings ($P > 0.05$ for all comparisons; Figure 1a). This means that common polygenic risk defined in GWAS studies of ASD, SCZ, and greater EA are each unambiguously associated with autism spectrum disorders. The results did not change when the SSC and PGC ASD samples were restricted to families of European ancestry (Supplementary Note 4; Supplementary Table 8, Supplementary Figures 2, 3). We repeated the analysis in the SSC and PGC ASD using a PRS for body mass index (BMI) – a polygenic risk category unassociated with ASD[6] – as a negative control (**Online Methods: *Polygenic Risk Scoring***). As expected, we found that neither ASD probands nor unaffected siblings over inherited BMI PRS ($P > 0.05$ for all comparisons) (Supplementary Table 9).

The degree of polygenic over transmission did not differ between SSC and PGC ASD probands for any comparison ($P > 0.05$ for all comparisons; Supplementary Table 10), so we combined the samples to improve the power of subsequent subgroup analyses. Using the combined data, each PRS is over transmitted to both male and female probands ($P < 0.05$ for all comparisons; Figure 1b). In Figure 1c, we also see that each of the three PRS are significantly, and equally, over transmitted to ASD cases with measured IQs in the intellectual disability range (IQ < 70, (**Online Methods: *pTDT***; Supplementary Tables 11, 12) when compared with those without intellectual disability. Most probands in the intellectual disability group do not have an observed genetic event that might explain their low IQ, for example a *de novo* mutation from an ASD-associated class, and repeating the analysis in the low IQ SSC probands with *de novo* carriers removed did not change the observed associations (**Online Methods: De novo *variant analyses***; Supplementary Table 13).

Common and rare variant additivity

In order to best interrogate additivity between common polygenic and rare, strong acting variation, we defined a group of *de novo* mutations of large effect (Supplementary Table 14). We previously identified a subclass of *de novo* protein truncating variants (contributing PTVs) responsible for almost all of the PTV association to ASDs[21] (**Online Methods: De novo *variant analyses***). As *de novo* copy number variants (CNVs) that delete a gene should

have the same, if not greater, molecular impact as a PTV in that gene, we were motivated to investigate whether we could similarly refine ASDs' association with *de novo* deletions (**Online Methods: De novo *variant analyses***). In the SSC, we found that ASDs were strongly associated with *de novo* deletions from two categories: a) deletions that include a constrained gene predicted to be intolerant of heterozygous loss of function variation (probability of loss of function intolerance (pLI) ≥ 0.9)[22], and b) uncommon, large (≥ 500 kilobase) deletions that do not contain a constrained gene (from here: contributing CNV deletions; Supplementary Figure 4, Supplementary Figure 5).

Together, these refined classes of *de novo* PTVs and deletions form a category of strong acting *de novo* variants (from here: contributing *de novo* variants (CDNVs); odds ratio (OR) = 3.91, P = 6.56E-20; case rate = 9.4%; control rate = 2.6%). Consistent with a hallmark of ASD-associated *de novo* variation, Figure 2a shows that the rate of CDNVs in SSC varies substantially based on the probands' counts of co-occurring adverse neurodevelopmental outcomes, in this case the count of: delayed walking, a history of seizures, or intellectual disability (**Online Methods: De novo *variant analyses***; Supplementary Table 15). ASD probands without any of those comorbid traits were more than 3 times (OR = 3.15; P = 3.88E-10) as likely to carry a CDNV than their unaffected siblings; ASD probands with all three of these comorbid traits were approximately 15 times as likely to carry a CDNV than their unaffected siblings (OR = 15.05; P = 9.08E-10). Because *de novo* events observed in cases and controls differ with regard to their average severity[23], their effect size cannot be directly estimated using the case-control carrier ratio. Using the male:female carrier approach described by De Rubeis *et al.*[3], we estimate that, on average, the CDNVs defined here confer an approximate 20-fold increase in risk for an ASD diagnosis (**Online Methods: De novo *variant analyses***). Their effect size, however, likely varies as the male:female ratio of the carriers declines with increasing neurodevelopmental comorbidities (Supplementary Table 16). In other words, the CDNVs seen in the ASD individuals with multiple comorbid neurodevelopmental traits are likely, on average, more deleterious than those seen in probands with ASD alone.

Shown in Figure 2b, we used pTDT to determine whether polygenic risk for ASD, SCZ, and EA was also over transmitted to CDNV carriers (**Online Methods: De novo *variant analyses***). ASD cases with a CDNV (n = 221 cases) indeed carried more polygenic risk for ASD and SCZ than expected based on parent background (P < 0.05 for both comparisons). There was no evidence of over transmission of EA PRS (P = 0.80), though a larger number of ASD cases with a CDNV will be required to differentiate a true null in this instance from an issue of power. Over transmission of the ASD and SCZ PRS provides clear evidence for additivity between common and rare variation in creating risk for ASDs. We did not see a difference in polygenic over transmission between probands with zero versus one or more co-occurring neurodevelopmental outcomes (P > 0.05 for all comparisons; Supplementary Table 17), further supporting the consistent influence of polygenic risk factors.

Additivity among common polygenic risk factors

In the context of three orthogonal risk distributions, each of which is associated with ASDs, one does not have to maintain an extreme position on any single distribution to carry a

cumulatively uncommon amount of ASD risk (Figure 3a). For example, being in the top 20% of three uncorrelated ASD risk distributions would result in a cumulative amount of risk seen in less than 1% of people in the population (0.2^3=0.008).

Common polygenic risk for ASDs, SCZ, and EA is partially correlated[6]. As shown in Figure 3b, however, we observed little evidence of correlation between the ASD, SCZ, and EA PRS in terms of either a) correlation at the mid-parent level, above the diagonal or b) correlation in the degree to which the scores were transmitted to the probands (pTDT deviation), below the diagonal. The lack of strong associations is likely a function of both ascertainment effects and attenuation due to limited predictive ability of the PRS.

Given limited association between the three PRS, we were able to examine additivity among largely distinct common polygenic risk factors. First, we saw significant evidence that each of the three PRS were independently over transmitted (P < 2E-04 in all SSC + PGC ASD comparisons; **Online Methods:** ***Genetic heterogeneity***; Supplementary Table 18). This means that ASD risk is influenced by elements unique to each of the scores, as well as by elements that are shared among them. Independent influences are further suggested by the scores' relationship with proband IQ (Supplementary Table 19). In Figure 3c, we show the association between each PRS and full scale proband IQ in European ancestry SSC probands, controlling for sex, presence of CDNVs, the other two PRS, and the first 10 principal components of ancestry (**Online Methods:** ***Genetic heterogeneity***). The EA and SCZ PRS are associated with proband IQ in opposite directions, consistent with the patterns observed in the general population (Supplementary Table 19)[14].

In addition to statistical evidence of unique contributions, the phenotypic associations suggest that the three PRS are influencing ASD risk through at least partially different processes. Polygenic risk for SCZ influences ASD liability in a manner that negatively influences cognition. Polygenic risk for EA influences ASD liability in a manner that positively influences cognition. These findings reinforce the idea that ASD heterogeneity is shaped not only by rare variants of strong effect, but also by diverse common variant risk factors acting through multiple biological pathways.

Discussion

Despite longstanding evidence for common polygenic influences on ASD risk, many have questioned those associations, particularly the recently published – and counterintuitive – findings from genetic correlation analysis. Using pTDT, we have shown an unambiguous association between ASD risk and the common polygenic influences on: ASDs themselves, schizophrenia, and greater educational attainment. These effects were evident in affected males and affected females, as well as ASD individuals with and without intellectual disability. Because of the strong correlation between the polygenic influences on educational attainment and intelligence[14], this finding means that, on average, individuals with ASD and intellectual disability have inherited more IQ increasing alleles than their typically developing siblings. That association, which replicated in independent ASD cohorts and held in probands without an ASD-associated *de novo* event, will require further study. With regard to proband phenotype, the finding furthers existing questions of whether an IQ

measured below 70 in someone with an ASD might be different in some ways from an IQ measured below 70 in someone without. With regard to better understanding the mechanisms through which genetic risk is conferred, we need to examine how and when certain amounts of risk are beneficial (e.g., a strong interest in the arts or sciences that increases one's educational attainment), and how and when they are deleterious (e.g., an overwhelming interest in a topic that disrupts healthy and necessary activities). Genetic risk and phenotypic traits relevant to neuropsychiatric disease exist on continua[24], for which effective research and treatment paradigms will likely need to account.

These findings also highlight important differences between the common and rare variant contributors to ASD risk. Strong acting, *de novo* variant risk for ASDs impacts a limited subset of cases. The phenotypic preferences of those types of variants are now well established; they are associated with intellectual disability, seizures, and global neurodevelopmental impact[1]. Common variant risk factors, on the other hand, appear more pervasively influential among ASD cases. Common variant risk appears similarly relevant to ASD individuals with high and low IQ, and with and without a large acting *de novo* mutation. The common polygenic influences also appear comparatively neurologically gentle. They are, in many cases, in fact associated with better educational and cognitive outcomes in the population. These differences strongly suggest that *de novo* and common polygenic variation may confer risk for ASD in different ways. Particularly as common polygenic risk is the more consistent contributor to ASD liability across cases, it will be critically important to take common variation into account in creating animal or stem cell models of ASD.

The pTDT approach can be broadly used to interrogate and clarify polygenic relationships. As a family-based approach, pTDT is immune to ancestral stratification and is less likely to be confounded by other influences on case ascertainment, for example socioeconomic status. While pTDT achieves optimal power comparing offspring with parents, it can be easily adapted to compare probands and unaffected siblings, and its predictive ability will improve through additional methodological development and larger discovery sample sizes. A command line tool to assist with pTDT analysis is publically available at: https://github.com/ypaialex/ptdt.

Online Methods

Sample Description

The analytic cohorts are presented in Supplementary Table 1. The research specific to this study was approved by the Partners Healthcare Institutional Review Board. The Simons Simplex Collection (SSC) is a resource of more than 2,500 families with a child diagnosed with an autism spectrum disorder (ASD)[18]. Informed consent and assent was provided for all subjects. Each member of the family was genotyped on one of the following platforms: Illumina Omni2.5, Illumina 1Mv3 or Illumina 1Mv1[1]. We analyzed 2,091 SSC quads, defined as families with both parents, the proband, and a designated unaffected sibling genotyped, and 493 SSC trios, defined as families with both parents and the proband genotyped. Most SSC families were also whole exome sequenced to detect rare coding

variation (82.7% of quads, an additional 8.0% of quads without sibling sequencing; 91.1% of trios).

The Psychiatric Genomics Consortium Autism Group (PGC ASD) sample consisted of parent-proband trios from the Psychiatric Genomics Consortium Autism Group (Supplementary Table 2). Our PGC ASD analytic sample excluded the SSC families present in the PGC ASD genome-wide association study (GWAS)[20]. The PGC ASD sample described here is accordingly independent from the SSC sample and included 3,870 parent-proband trios. In brief, the PGC ASD data included ASD trio cohorts from: Autism Center of Excellence UCLA (n = 215), Autism Genome Project (n = 2,254), Montreal/Boston Collection (n = 138), Johns Hopkins University (n = 764), and the Children's Hospital of Philadelphia (n = 499)[20]. All genotype data (SSC and PGC ASD) were imputed using the 1000 Genomes reference panel and Ricopili pipeline, which are publicly available and have been reported on extensively[19, 25, 26].

Though trio approaches are broadly immune to confounding through ancestry, we isolated European-only subsets of both the SSC and PGC ASD samples to a) ensure that our primary results did not change in an ancestrally homogeneous subset of the data and b) conduct comparisons across probands or across parents, which might be sensitive to ancestry (as opposed to comparisons between probands and parents in a trio). In the SSC, we first selected probands with parent-reported white non-Hispanic ancestry (n = 1,912). We merged the genotype data from those individuals with the Hapmap III dataset[27] and generated principal components of ancestry using GCTA[28]. Through visual inspection (Supplementary Figure 6), we defined an ancestrally European SSC subcohort, leaving 1,851 probands (and, by extension, 1,851 families). We calculated principal components of ancestry distinctly within the derived SSC European ancestry subcohort and used these as covariates in the non-trio analysis (e.g., genotype to phenotype analyses among probands).

Self-reported ancestry was not available for the PGC ASD samples. To conservatively isolate a European ancestry PGC ASD subcohort, we identified those families in which both parents were of European ancestry. To do so, we merged the PGC ASD parent data with Hapmap III and similarly generated principal components of ancestry using GCTA. By visual inspection, we identified European ancestry individuals, leaving 6,742 of the original 7,740 PGC ASD parents (Supplementary Figure 7). Both parents were of European ancestry in 3,209 families, and those families comprised the European ancestry PGC ASD subcohort. We again calculated principal components of ancestry within the derived PGC ASD European ancestry subcohort for use as analytic covariates in non-trio analyses.

Polygenic Risk Scoring

A polygenic risk score (PRS) provides a quantitative estimate of an individual's genetic predisposition ("risk") for a given phenotype based on common variant genotype data and independent GWAS results for the target phenotype (e.g., schizophrenia or educational attainment)[29]. The score provides a relative, not absolute, measure of risk. For example, individual A ($PRS_{schizophrenia}$ = 8) is at higher estimated genetic risk for schizophrenia than individual B ($PRS_{schizophrenia}$ = 6), but a PRS of 8 or 6 is not independently interpretable.

We calculated polygenic risk for ASD, educational attainment (EA), schizophrenia (SCZ) and body mass index (BMI) for all individuals with available genotype data in the SSC and PGC ASD datasets. To do so, we used summary statistics from the largest available, independent GWAS of each phenotype (Supplementary Table 3). None of the subjects in the SSC or PGC ASD cohort were included in any of the four GWAS discovery samples. We selected SCZ and EA because of their robust, well-replicated associations with ASD risk[6, 14, 20]. We selected BMI as a negative control due to its lack of association with ASD risk[6]. We used ASD summary statistics from a GWAS of a Danish population-based sample of 7,783 cases and 11,359 controls from the first 10 genotyping waves of the iPSYCH-Broad Autism project[24]. The SCZ summary statistics were from the 2014 GWAS of 36,989 cases and 113,075 controls from the Psychiatric Genomics Consortium[19]. The EA summary statistics were from a population-based GWAS of years of schooling (n = 328,917, discovery and replication meta-analysis, excluding 23andMe)[12]. The BMI summary statistics were from a population-based GWAS of body mass index (n = 322,154, European ancestry meta-analysis)[30].

To construct the polygenic risk scores, we first gathered the summary statistics from the GWAS described above. The summary statistics included effect sizes and p-values for each single nucleotide polymorphism (SNP) in the imputed GWAS analysis, typically approximately 10 million markers. We then employed the widely used Ricopili pipeline to generate the PRS[19]. In brief, Ricopili first removes SNPs within 500 kilobases of and correlated ($r^2 \geq 0.1$) with a more significantly associated SNP in the GWAS. We used the 1000 Genomes Reference panel to estimate SNP correlations[25]. This process typically reduces the SNP list to fewer than 300,000 markers. For complex, polygenic traits, only a small fraction of those SNPs remaining pass the genome-wide significance threshold of P = 5.00E-08; the majority of the signal resides in SNPs that do not pass the significance threshold and cannot be specifically identified [6, 12, 19, 28]. In order to maximally capture common, polygenic influence, we therefore relaxed the p-value threshold for SNPs in the PRS until doing so added more noise than signal (threshold options: P ≤ 1, 5E-1, 2E-1, 1E-1, 5E-2, 1E-2, 1E-3, 1E-4, 1E-6, and 5E-8). We identified the optimal p-value threshold as that which explained the most phenotypic variation for each trait. For the ASD PRS, we found that the threshold of P ≤ 0.1 explained the most case-pseudocontrol variance in SSC (as the SSC does not have independent control data, we generated pseudocontrol genotypes from the untransmitted parental alleles)[20, 24]. For the SCZ and EA PRS, we used the threshold identified by analyses accompanying the discovery GWAS as explaining the most variance[12, 19]. In the PGC's 2014 schizophrenia analyses, a p-value threshold of P ≤ 0.05 most commonly explained the most SCZ case-control variance in 40 leave-one-out analyses[19]. For educational attainment, constructing a PRS using a threshold of P ≤ 1 explained the most variance in number years of education achieved in an independent sample[12]. For BMI, constructing a PRS using a threshold of P ≤ 0.2 explained the most variance in phenotypic body mass index among cases in the Simons Simplex Collection (Supplementary Note 5). These thresholds create the strongest polygenic risk scores for ASD, EA, SCZ and BMI, leaving us maximally powered to investigate the relationship between these four traits and ASDs.

Next, we excluded SNPs that were poorly imputed in either SSC or PGC ASD (info score < 0.6 in either cohort). The exception to this filtering rule was in SSC-specific analyses, (e.g., analysis of *de novo* variation) where our info threshold of 0.6 was the minimum in the SSC imputation only. For each trait, the polygenic risk scores for individuals in SSC and PGC ASD were then calculated as the product of the GWAS effect size (log odds or beta) at that SNP by the individual's count of reference alleles at that SNP (0, 1 or 2), summed across all remaining markers. We implemented this scoring protocol using the score function in Plink[31]. If an individual was missing genetic data at a SNP in the summary statistics file, Plink calculated the expected score based on the cohort-wide allele frequency.

pTDT

We define pTDT deviation as:

$$\text{pTDT deviation} = \frac{\text{PRS}_c - \text{PRS}_{\text{MP}}}{\text{SD}(\text{PRS}_{\text{MP}})}$$

where PRS_C is the polygenic risk score for the child (proband or unaffected sibling). PRS_{MP} is the mid-parent polygenic risk score:

$$\text{PRS}_{\text{MP}} = \frac{\text{PRS}_{\text{mother}} + \text{PRS}_{\text{father}}}{2}$$

$\text{SD}(\text{PRS}_{MP})$ is the standard deviation of the sample-specific mid-parent PRS. For example, in SSC analysis, the $\text{SD}(\text{PRS}_{MP})$ was the standard deviation of the mid-parent PRS distribution in SSC parents. We chose to standardize the pTDT deviation to improve interpretability and to facilitate comparison between different PRS. We standardized by PRS_{MP} instead of PRS_C because we expect the parent PRS distribution to be a better proxy for the population PRS distribution. The approach can be adapted to PRS from unaffected siblings, but using mid-parent PRS improves statistical power (Supplementary Note 6, Supplementary Table 20).

To evaluate whether the pTDT deviation is significantly different than 0, we defined the pTDT test statistic (t_{pTDT}) as:

$$t_{\text{pTDT}} = \frac{\text{mean(pTDT deviation)}}{\text{SD(pTDT deviation)}/\sqrt{n}}$$

where n is the number of families included in the pTDT. We evaluate t_{pTDT} as a two-sided, one-sample t-test.

We performed pTDT using the ASD, EA, SCZ and BMI PRS described above in four groups: SSC probands (n = 2,584), SSC unaffected siblings (n = 2,091), PGC ASD probands (n = 3,870), and the combination of SSC and PGC ASD probands (n = 6,454). As described in the main text (Figure 1a), each of the ASD, EA, and SCZ PRS were significantly over transmitted from parents to probands, but not to unaffected siblings (P < 1E-06 for all parent

to proband comparisons in either SSC or PGC ASD; P < 1E-15 for all parent to proband comparisons in combined SSC and PGC ASD; P > 0.05 for all unaffected sibling comparisons). In contrast, BMI PRS was not over transmitted to probands (P > 0.05 in both the SSC and PGC ASD). In a complementary permutation test, we randomly assigned case/control labels to the affected and unaffected children in SSC. We then counted the number of times that the simulated difference in ASD PRS between affected and unaffected SSC children exceeded the observed difference (86 of 1,000,000 permutations = 0.0086%), consistent with the primary results (those in Figure 1a).

In our primary associations (Figure 1a, Supplementary Table 8), there were no statistically significant pTDT differences between the SSC and PGC ASD cohorts (Supplementary Table 10). We accordingly combined SSC and PGC ASD for the analyses stratified by sex and IQ to increase statistical power (Figures 1b, 1c). In SSC, full scale IQ (FSIQ; SSC variable: sscfsiq) was derived from a number of scales and available for almost all probands (99.8%). Those scales included but were not limited to the Differential Ability Scales, Second Edition[32]; Mullen Scales of Early Learning[33]; and Wechsler battery[34]. Our full scale SSC IQ estimates were taken from the SSC's 'full scale deviation IQ' variable when it was available, and 'full scale ratio IQ' when it was not[35]. FSIQ was measured heterogeneously across the contributing PGC ASD cohorts (Supplementary Table 2). To accommodate measurement differences, FSIQ was converted to broad groups for PGC consortium-level analyses (for the most part, numeric IQ values from PGC ASD were not available for this analysis). PGC ASD probands were assigned FSIQ levels 1-4 as follows: (1) FSIQ < 35; (2) 35 ≤ FSIQ < 70; (3) 70 ≤ FSIQ < 90; (4) 90 ≤ FSIQ. In PGC ASD, 38.9% of probands had estimated IQ available. We divided each of SSC and PGC ASD by presence/absence of intellectual disability (ID) in the proband so that they could be analyzed together (SSC ID: IQ < 70; PGC ASD ID: IQ = 1 or IQ = 2). We repeated the IQ stratified analyses in SSC and PGC ASD separately, which further suggested no differences between the cohorts, despite the limited FSIQ data available in PGC ASD (Supplementary Tables 11, 12).

De novo variant analyses

We defined a group of *de novo* mutations strongly associated with ASD risk (Supplementary Table 14). We performed this analysis exclusively in SSC; we could not perform the analysis in PGC ASD because only common variant (GWAS) data was available. Our previous work has identified a subclass of *de novo* protein truncating variants (PTVs: frameshift variants, splice acceptor variants, splice donor variants, nonsense variants) that are a primary source of association to ASD[21]. *De novo* PTVs in this class are a) absent from the Exome Aggregation Consortium database, a reference sample of over 60,000 exomes, and b) found within a gene predicted to be intolerant of heterozygous loss of function variation (probability of loss of function intolerance (pLI) ≥ 0.9)[22]. In the SSC, *de novo* PTVs in this class were found in 7.1% of cases and 2.1% of unaffected siblings (P = 4.12E-14). PTVs outside of this contributing class are unassociated with ASD risk (observed in 7.8% of cases and 6.9% of unaffected siblings, P = 0.50) and are not associated with proband IQ (P = 0.76) (Supplementary Figure 8).

As detailed in the main text, *de novo* copy number variants (CNVs) that delete a gene should have the same molecular impact as a PTV in that same gene. Contributing CNV deletions were seen in 2.5% of SSC cases and 0.5% of SSC unaffected siblings (P = 1.80E-08). All other types of *de novo* deletions were observed in 1.7% of cases and 1.4% of unaffected siblings, and were not associated with ASD risk (P = 0.48; Supplementary Figure 5). The associations between CNV categories and ASD risk are presented in the supplement and did not differ substantially when controlling for parental age at birth of child (Supplementary Table 21). Contributing deletions outside of our defined class were not associated with proband IQ (P = 0.34, Supplementary Figure 9). In contrast, *de novo* duplications of constrained genes are not disproportionately associated with ASD risk (P = 0.49; Supplementary Note 7; Supplementary Figure 10).

We identified a subset of *de novo* CNV deletions, primarily those containing loss of function intolerant genes, that account for most of the category's association to ASDs (contributing CNV deletions; Supplementary Figure 4). Together, contributing PTVs and contributing CNV deletions formed a strong acting *de novo* variant category in SSC, known from here as contributing *de novo* variants (CDNVs). Multiple lines of evidence suggest CDNVs are robustly associated with ASD risk. CDNVs were seen in 9.4% of SSC cases (n = 221 of 2,346) that were both sequenced and genotyped and 2.6% of SSC unaffected siblings (n = 45 of 1,736) that were both sequenced and genotyped (P = 6.56E-20). CDNVs were very strongly associated with proband IQ (P = 5.80E-06; controlling for proband sex), while all other *de novo* PTVs and deletions (those not in the CDNV category) were not associated with proband IQ (P = 0.44, controlling for proband sex; Supplementary Figure 11).

As *de novo* variants are on average more severe when observed in ASD cases than in controls[23], we could not estimate the amount of ASD risk conferred by CDNVs directly from the case-unaffected sibling odds ratio. As noted in the main text, we used the male:female CDNV carrier ratio to re-estimate the effect size for the event class, as described in De Rubeis et al., 2014. In brief, the approximately 4:1 male:female ratio among ASD cases suggests a different ASD liability threshold for males and females in the population. Regardless of whether that difference reflects etiology or ascertainment, it results in a direct mathematical relationship between the expected effect size of an event class and the male:female carrier ratio of cases. In the case of CDNVs, a variant class observed twice as frequently in female probands than in male probands (17.4% v 8.5%; P = 5.26E-06; Supplementary Figure 12), we estimate an OR of approximately 20^3. This estimate exceeds that directly suggested by the CDNV case-control excess of 3.63.

We next examined the rate of CDNVs in SSC probands based on the number of adverse proband co-occurring neurological and developmental outcomes (Figure 2a). Previous studies have demonstrated that *de novo* PTVs and *de novo* CNVs are both associated with intellectual disability (IQ < 70) in ASD probands and positively associated with history of seizures[1, 15]. Motor delays are an additional neurodevelopmental co-morbitidy associated with autism spectrum disorder[36]. Here, we defined motor delay for SSC probands as walking unaided at or after 19 months, the age by which the great majority of children have begun to walk[37]. As hypothesized, motor delay, seizures, and low IQ were independently associated with CDNV rate in SSC probands after controlling for proband sex

(Supplementary Table 15). CDNV rate was positively associated with the number of these adverse outcomes experienced by probands (Figure 2a, Supplementary Table 16). The decreasing male:female ratio among CDNV carriers as count of co-occurring neurodevelopmental outcomes increases suggests that with increased count of neurodevelopmental comorbidities comes, on average, increasingly severe *de novo* events (Supplementary Note 8, Supplementary Table 16). Multiple lines of evidence, including associations with IQ, male:female carrier ratio, and adverse co-occurring neurodevelopmental outcomes, suggest that CDNVs are strongly associated with ASD risk.

Using pTDT, we evaluated polygenic transmission in probands carrying at least one CDNV (Figure 2b, Supplementary Table 13). As this CDNV analysis is specific to SSC, the polygenic risk scores generated for the CDNV analysis used info score cutoffs from SSC imputation in order to increase the number of well-imputed SNPs retained for PRS. We restricted our analysis to those families with genotyped parents and probands with both genotype and exome sequence data (n = 2,346; n = 221 with CDNV). The cohort of probands with a CDNV is too small to determine whether the difference in transmission between probands that do and do not carry a CDNV is statistically significant. We also analyzed whether polygenic over transmission was seen in a more broadly defined set of *de novo* events (Supplementary Note 9, Supplementary Table 13).

Genetic heterogeneity and proband phenotype

We analyzed whether polygenic risk for ASD, EA and SCZ were each independently over transmitted to ASD cases. For each of SSC (n = 2,584 trios), PGC ASD (n = 3,870 trios) and SSC & PGC ASD combined (n = 6,454 trios), we performed a single logistic regression predicting proband/mid-parent status from polygenic risk for 1) ASD, 2) EA and 3) SCZ. We confirmed that the over transmission acts independently in each PRS (P < 2E-04 for all PRS in SSC + PGC ASD cohort; Supplementary Table 18).

To calculate the relationship between proband polygenic risk and ASD IQ (Figure 3c), we performed three separate linear regressions predicting full-scale proband IQ from the three PRS. The three ASD IQ - PRS associations are from three linear regressions predicting full-scale proband IQ from the residualized and z-normed PRS. The other two PRS, CDNV presence/absence, proband sex, and the first 10 principal components of proband ancestry were regressed out of each PRS before analysis. These associations were performed in European ancestry Simons Simplex Collection probands (n = 1,674). The results were consistent with previously published association between the three polygenic risk scores and IQ in the general population (Supplementary Table 19)[14]. This relationship between polygenic risk and proband IQ holds when using mid-parent PRS and controlling for proband CDNV status (Supplementary Note 10, Supplementary Table 22).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Hyman and R. Hosking for their thoughtful comments. We also thank A. Pai for his help in the development of the pTDT analytic software. E. Robinson and D. Weiner were funded by National Institute of Mental Health grant 1K01MH099286-01A1 and Brain Behavior Research Foundation (NARSAD) Young Investigator grant 22379. E. Wigdor was funded by the Stanley Center for Psychiatric Research at the Broad Institute. A. Okbay was funded by ERC Consolidator Grant (647648 EdGe). We thank the families who took part in the Simons Simplex Collection study and the clinicians who collected data at each of the study sites. The iPSYCH project is funded by the Lundbeck Foundation and the universities and university hospitals of Aarhus and Copenhagen. Genotyping of iPSYCH and PGC samples was supported by grants from the Stanley Foundation, the Simons Foundation (SFARI 311789 to M. Daly) and the National Institute of Mental Health (5U01MH094432-02 to M. Daly). This work was also supported by a grant from the Simons Foundation (SFARI 402281 to S. Sanders). The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found on the ExAC website (see URLs).

References

1. Sanders SJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015; 87:1215–1233. [PubMed: 26402605]
2. Gaugler T, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014; 46:881–885. [PubMed: 25038753]
3. De Rubeis S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–215. [PubMed: 25363760]
4. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]
5. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47:291–295. [PubMed: 25642630]
6. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015; 47:1236–1241. [PubMed: 26414676]
7. Krumm N, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015; 47:582–588. [PubMed: 25961944]
8. Anney R, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. Human molecular genetics. 2012; 21:4781–4792. [PubMed: 22843504]
9. Klei L, et al. Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism. 2012; 3:9. [PubMed: 23067556]
10. Onis M. WHO Motor Development Study: Windows of achievement for six gross motor development milestones. Acta Paediatrica. 2006; 95:86–95.
11. Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015; 519:223–228. [PubMed: 25533962]
12. Okbay A, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016; 533:539–542. [PubMed: 27225129]
13. Clarke TK, et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. Mol Psychiatry. 2015
14. Hagenaars SP, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. Molecular psychiatry. 2016
15. Robinson EB, et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:15161–15165. [PubMed: 25288738]
16. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider Scope: How selection bias can induce spurious associations. bioRxiv. 2016
17. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American journal of human genetics. 1993; 52:506–516. [PubMed: 8447318]
18. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron. 2010; 68:192–195. [PubMed: 20955926]

19. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. [PubMed: 25056061]
20. Cross-Disorder Group of the Psychiatric Genomics, C. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature genetics. 2013; 45:984–994. [PubMed: 23933821]
21. Kosmicki JA, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nature genetics. 2017
22. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]
23. Samocha KE, et al. A framework for the interpretation of de novo mutation in human disease. Nature genetics. 2014; 46:944–950. [PubMed: 25086666]
24. Robinson EB, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nature genetics. 2016; 48:552–555. [PubMed: 26998691]
25. Genomes Project, C. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
26. Ricopili. Rapid Imputation Consortium Pipeline.
27. International HapMap, C. Integrating common and rare genetic variation in diverse human populations. Nature. 2010; 467:52–58. [PubMed: 20811451]
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011; 88:76–82. [PubMed: 21167468]
29. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome research. 2007; 17:1520–1528. [PubMed: 17785532]
30. Locke AE, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518:197–206. [PubMed: 25673413]
31. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015; 4:7. [PubMed: 25722852]
32. Elliott, C. Differential Ability Scales. The Psychological Corporation; San Antonio, TX: 2007.
33. Mullen, E. Mullen Scales of Early Learning. American Guidance Service; Circle Pines, MN: 1995.
34. Wechsler, D. Wechsler abbreviated scale of intelligence. Psychological Corporation; San Antonio, TX: 1999.
35. Chaste P, et al. A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? Biological psychiatry. 2015; 77:775–784. [PubMed: 25534755]
36. Provost B, Lopez BR, Heimerl S. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. Journal of autism and developmental disorders. 2007; 37:321–328. [PubMed: 16868847]
37. World Health Organization, T. WHO Motor Development Study: windows of achievement for six gross motor development milestones. Acta paediatrica. 2006; 450:86–95. [PubMed: 16817682]

Appendix

iPSYCH-Broad Autism Group

Marie Bækvad-Hansen[1,2], Ashley Dumont[3], Christine Hansen[1,2], Thomas F. Hansen[1,4], Daniel Howrigan[3,5,6], Manuel Mattheisen[1,7,8], Jennifer Moran[3], Ole Mors[1,9], Merete Nordentoft[1,10], Bent Nørgaard-Pedersen[1,2], Timothy Poterba[3,5,6], Jesper Poulsen[1,2], and Christine Stevens[3]

[1]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark.

[2]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark.

[3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[4]Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark.

[5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

[6]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[7]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark.

[8]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark.

[9]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark.

[10]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark.

Psychiatric Genomics Consortium Autism Group

Verneri Anttila[1,2], Peter Holmans[3], Hailiang Huang[1,2], Lambertus Klei[4], Phil H. Lee[1,2,5], Sarah E. Medland[6], Benjamin Neale[1,2], Lauren A. Weiss[7,8], Lonnie Zwaigenbaum[9], Timothy W. Yu[10], Kerstin Wittemeyer[11], A. Jeremy Willsey[7], Ellen M. Wijsman[12,13], Thomas H. Wassink[14], Regina Waltes[15], Christopher A. Walsh[10,16,17,18,19], Simon Wallace[20], Jacob A. S. Vorstman[21], Veronica J. Vieland[22], Astrid M. Vicente[23,24], Herman van Engeland[21], Kathryn Tsang[7,8], Ann P. Thompson[25], Peter Szatmari[26], Oscar Svantesson[27], Stacy Steinberg[28], Kari Stefansson[28], Hreinn Stefansson[28], Matthew W. State[7], Latha Soorya[29,30,31], Teimuraz Silagadze[32], Stephen W. Scherer[33,34,35], Gerard D. Schellenberg[36], Sven Sandin[27], Evald Saemundsen[37], Guy A. Rouleau[38], Bernadette Rogé[39], Kathryn Roeder[40,41], Wendy Roberts[42], Jennifer Reichert[29,30], Abraham Reichenberg[29,30], Karola Rehnström[43], Regina Regan[44,45], Fritz Poustka[15], Christopher S. Poultney[29,30], Joseph Piven[46], Dalila Pinto[29,30,47,48,49,50], Margaret A. Pericak-Vance[51], Milica Pejovic-Milovancevic[52], Marianne G. Pedersen[53,54,55], Carsten B. Pedersen[53,54,55], Andrew D. Paterson[33,35,56], Jeremy R. Parr[57,58], Alistair T. Pagnamenta[59], Guiomar Oliveira[60,61], John I. Nurnberger[62,63,64], Merete Nordentoft[53,65], Michael T. Murtha[66], Susana Mouga[60,61], Ole Mors[53,67], Eric M. Morrow[68], Daniel Moreno De Luca[68], Anthony P. Monaco[59,69], Nancy Minshew[4], Alison Merikangas[70], William M. McMahon[71], Susan G. McGrew[72], Manuel Mattheisen[53,73,74], Igor Martsenkovsky[75], Donna M. Martin[76], Shrikant M. Mane[77], Pall Magnusson[78], Tiago Magalhaes[44,45], Elena Maestrini[79], Jennifer K. Lowe[80,81,82], Catherine Lord[83], Pat Levitt[84], Christa Lese Martin[85], David H. Ledbetter[86], Marion Leboyer[87,88,89,90], Ann S. Le Couteur[57,58], Christine Ladd-Acosta[91], Alexander Kolevzon[29,30,48,50], Sabine M. Klauck[92], Suma Jacob[93,94], Bozenna Iliadou[27], Christina M.

Hultman[27], Irva Hertz-Picciotto[95,96], Robert Hendren[7], Christine S. Hansen[53,97], Jonathan L. Haines[98], Stephen J. Guter[93], Dorothy E. Grice[30], Jonathan M. Green[99,100], Andrew Green[45,101], Arthur P. Goldberg[47,49], Christopher Gillberg[102], John Gilbert[51], Louise Gallagher[70], Christine M. Freitag[15], Eric Fombonne[103], Susan E. Folstein[104], Bridget Fernandez[105], M. Daniele Fallin[106], A. Gulhan Ercan-Sencicek[66], Sean Ennis[45,101], Frederico Duque[60,61], Eftichia Duketis[15], Richard Delorme[87,107,108,109], Silvia De Rubeis[29,30], Maretha V. De Jonge[21], Geraldine Dawson[110,111], Michael L. Cuccaro[51], Catarina T. Correia[23,24], Judith Conroy[45,112], Inês C. Conceição[23,24], Andreas G. Chiocchetti[15], Patrícia B. S. Celestino-Soper[63,113], Jillian Casey[45,112], Rita M. Cantor[114,115], Cátia Café[60], Sean Brennan[70], Thomas Bourgeron[87,107,108,116], Patrick F. Bolton[117,118], Sven Bölte[15,119,120], Nadia Bolshakova[70], Catalina Betancur[121,122,123], Raphael Bernier[124], Arthur L. Beaudet[113], Agatino Battaglia[125], Vanessa H. Bal[7], Gillian Baird[126], Anthony J. Bailey[20,127], Marie Bækvad-Hansen[53,97], Joel S. Bader[128], Elena Bacchelli[79], Evdokia Anagnostou[129], David Amaral[96,130,131], Joana Almeida[60], Joseph D. Buxbaum[29,30,47,48,50,132], Aravinda Chakravarti[128], Edwin H. Cook[93], Hilary Coon[71], Daniel H. Geschwind[80,81,82,115], Michael Gill[70], Hakon Hakonarson[133,134], Joachim Hallmayer[135], Aarno Palotie[43], Susan Santangelo[136], James S. Sutcliffe[98,137], and Dan E. Arking[128]

[1]Analytic and Translational Genetics Unit, Dept. of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

[2]Stanley Center for Psychiatric Research and Program in Medical and Population Genetic, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.

[3]MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, CF24 4HQ, UK.

[4]Dept. of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.

[5]Dept. of Psychiatry, Harvard Medical School, Boston, MA 02115, USA.

[6]Queensland Institute of Medical Research, Brisbane, QLD, 4006, Australia.

[7]Dept. of Psychiatry, University of California San Francisco, San Francisco, CA 94143, USA.

[8]Inst. Human Genetics, University of California San Francisco, San Francisco, CA 94143, USA.

[9]Dept. of Pediatrics, University of Alberta, Edmonton, AB, T6G 1C9, Canada.

[10]Division of Genetics, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115, USA.

[11]School of Education, University of Birmingham, Birmingham, B15 2TT, UK.

[12]Dept. of Medicine, University of Washington, Seattle, WA 98195, USA.

[13]Dept. of Biostatistics, University of Washington, Seattle, WA 98195, USA.

[14]Dept. of Psychiatry, Carver College of Medicine, Iowa City, IA 52242, USA.

[15]Dept. of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, JW Goethe University Frankfurt, Frankfurt am Main, 60528, Germany.

[16]Program in Genetics and Genomics, Harvard Medical School, Boston, MA 02115, USA.

[17]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

[18]Dept. of Pediatrics, Harvard Medical School, Boston, MA 02115, USA.

[19]Dept. of Neurology, Harvard Medical School, Boston, MA 02115, USA.

[20]Dept. of Psychiatry, University of Oxford and Warneford Hospital, Oxford, OX3 7JX, UK.

[21]Dept. of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, 3584 CG, The Netherlands.

[22]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA.

[23]Instituto Nacional de Saœde Dr. Ricardo Jorge, Lisboa, 1600, Portugal.

[24]Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, Lisboa, 1649, Portugal.

[25]Dept. of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, ON, L8S 4L8, Canada.

[26]Dept. of Psychiatry, University of Toronto, ON, M5T 1R8, Canada.

[27]Karolinska Institutet, Solna, SE-171 77, Sweden.

[28]deCODE Genetics, Reykjavik, IS-101, Iceland.

[29]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[30]Dept. of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[31]Dept. of Psychiatry, Rush University Medical Center, Chicago, IL 60612, USA.

[32]Dept. of Psychiatry and Drug Addiction, Tbilisi State Medical University, Tbilisi, 0186, Georgia.

[33]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, M5G 1L4, Canada.

[34]McLaughlin Centre, University of Toronto, Toronto, ON, M5G 0A4, Canada.

[35]Dept. of Molecular Genetics, University of Toronto, Toronto, ON, M5S 1A8, Canada.

[36]Dept. of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19102, USA.

[37]State Diagnostic and Counseling Centre, Kopavogur, IS-201, Iceland.

[38]Montreal Neurological Institute, Dept of Neurology and Neurosurgery, McGill University, Montreal, QC, H3A 2B4, Canada.

[39]Centre d'Etudes et de Recherches en Psychopathologie, Toulouse University, Toulouse, 31058, France.

[40]Dept. of Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

[41]Dept. of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

[42]Autism Research Unit, The Hospital for Sick Children, Toronto, ON, M5G 1L4, Canada.

[43]Sanger Institute, Hinxton, CB10 1SA, UK.

[44]National Childrens Research Centre, Our Lady's Hospital Crumlin, Dublin, D12, Ireland.

[45]Academic Centre on Rare Diseases, University College Dublin, Dublin, D4, Ireland.

[46]University of North Carolina, Chapel Hill, NC 27599, USA.

[47]Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[48]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[49]The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[50]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[51]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL 33101 , USA.

[52]Institute of Mental Health and Medical Faculty, University of Belgrade, Belgrade, 11 000, Serbia.

[53]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark.

[54]National Centre for Register-based Research, Aarhus University, Aarhus, Denmark.

[55]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark.

[56]Dalla Lana School of Public Health, Toronto, ON, M5T 3M7, Canada.

[57]Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, NE2 4HH, UK.

[58]Institue of Health and Science, Newcastle University, Newcastle Upon Tyne, NE2 4AX, UK.

[59]Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, OX3 7BN, UK.

[60]Unidade de Neurodesenvolvimento e Autismo do Serviço do Centro de Desenvolvimento da Criança and Centro de Investigação e Formação Clinica, Pediatric Hospital, Centro Hospitalar e Universitário de Coimbra, Coimbra, 3041-80, Portugal.

[61]University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, Faculty of Medicine, University of Coimbra, Coimbra, 3041-80, Portugal.

[62]Institute of Psychiatric Research, Dept. of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

[63]Dept. of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

[64]Program in Medical Neuroscience, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

[65]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark.

[66]Programs on Neurogenetics, Yale University School of Medicine, New Haven, CT 06520, USA.

[67]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark.

[68]Dept. of Psychiatry and Human Behaviour, Brown University, Providence, RI 02912, USA.

[69]Tufts University, Boston, MA 02155, USA.

[70]Dept. of Psychiatry, Trinity College Dublin, Dublin, D8, Ireland.

[71]Dept. of Psychiatry, University of Utah, Salt Lake City, UT 84108, USA.

[72]Dept. of Pediatrics, Vanderbilt University, Nashville, TN 37232, USA.

[73]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK-8000, Denmark.

[74]Dept. of Biomedicine - Human Genetics, Aarhus University, Aarhus, DK-8000, Denmark.

[75]Dept. of Child, Adolescent Psychiatry and Medical-Social Rehabilitation, Ukrainian Research Institute of Social Forensic Psychiatry and Drug Abuse, Kyiv, 04080, Ukraine.

[76]Dept. of Pediatrics and Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA.

[77]Yale Center for Genomic Analysis, Yale University School of Medicine, New Haven, CT 06516, USA.

[78]Dept. of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, IS-101, Iceland.

[79]Dept. of Pharmacy and Biotechnology, University of Bologna, Bologna, 40126, Italy.

[80]Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA 90095, USA.

[81]Program in Neurogenetics, Dept. of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[82]Center for Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[83]Dept. of Psychiatry, Weill Cornell Medical College, Cornell University, New York, NY 10065, USA.

[84]Dept. of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA 90027, USA.

[85]Autism & Developmental Medicine Institute, Geisinger Health System, Danville, PA 17837, USA.

[86]Chief Scientific Officer, Geisinger Health System, Danville, PA 17837, USA.

[87]FondaMental Foundation, Créteil, 94000, France.

[88]INSERM U955, Paris, 94010, France.

[89]Faculté de Médecine, Université Paris Est, Créteil, 94000, France.

[90]Dept. of Psychiatry, Henri Mondor-Albert Chenevier Hospital, Assistance Publique – Hôpitaux de Paris, Créteil, 94000, France.

[91]Dept. of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.

[92]Division of Molecular Genome Analysis and Working Group Cancer Genome Research, Deutsches Krebsforschungszentrum, Heidelberg, D-69120, Germany.

[93]Institute for Juvenile Research, Dept. of Psychiatry, University of Illinois at Chicago, Chicago, IL 60612, USA.

[94]Institute of Translational Neuroscience and Dept. of Psychiatry, University of Minnesota, Minneapolis, MN 55454 , USA.

[95]Dept. of Public Health Sciences, School of Medicine, University of California Davis, Davis, CA 95616, USA.

[96]The MIND Institute, School of Medicine, University of California Davis, Davis, CA 95817, USA.

[97]Center for Neonatal Screening, Dept. for Congenital Disorders, Statens Serum Institut, Copenhagen, DK-2300, Denmark.

[98]Dept. of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN 37232, USA.

[99]Manchester Academic Health Sciences Centre, Manchester, M13 9NT, UK.

[100]Institute of Brain, Behaviour, and Mental Health, University of Manchester, Manchester, M13 9PT, UK.

[101]Centre for Medical Genetics, Our Lady's Hospital Crumlin, Dublin, D12, Ireland.

[102]Gillberg Neuropsychiatry Centre, University of Gothenburg, Gothenburg, S-405 30, Sweden.

[103]Dept. of Psychiatry and Institute for Development and Disability, Oregon Health & Science University, Portland, OR 97239, USA.

[104]Division of Child and Adolescent Psychiatry, Dept. of Psychiatry, Miller School of Medicine, University of Miami, Miami, FL 33136, USA.

[105]Memorial University of Newfoundland, St. John’s, NL, A1B 3X9, Canada.

[106]Dept. of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.

[107]Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, 75015, France.

[108]Centre National de la Recherche Scientifique URA 2182 Institut Pasteur, Paris, 75724, France.

[109]Dept. of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique – Hôpitaux de Paris, Paris, 75019, France.

[110]Duke Center for Autism and Brain Developments, Duke University School of Medicine, Durham, NC 27705 , USA.

[111]Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, NC 27708, USA.

[112]Temple Street Children's University Hospital, Dublin, D1, Ireland.

[113]Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

[114]Dept. of Psychiatry, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA 90095, USA.

[115]Dept. of Human Genetics, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA 90095, USA.

[116]University Paris Diderot, Sorbonne Paris Cité, Paris, 75013, France.

[117]Institute of Psychiatry, Kings College London, London, SE5 8AF, UK.

[118]South London & Maudsley Biomedical Research Centre for Mental Health, London, SE5 8AF, UK.

[119]Dept. of Women's and Children's Health, Center of Neurodevelopmental Disorders, Karolinska Institutet, Stockholm, SE-113 30, Sweden.

[120]Child and Adolescent Psychiatry, Center for Psychiatry Research, Stockholm County Council, Stockholm, SE-171 77, Sweden.

[121]INSERM U1130, Paris, 75005, France.

[122]CNRS UMR 8246, Paris, 75005, France.

[123]Sorbonne Universités, UPMC Univ Paris 6, Neuroscience Paris Seine, Paris, 75005, France.

[124]Dept. of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195 , USA.

[125]Stella Maris Institute for Child and Adolescent Neuropsychiatr, Pisa, 56018, Italy.

[126]Paediatric Neurodisability, King's Health Partners, Kings College London, London, SE1 7EH, UK.

[127]Mental Health and Addictions Research Unit, University of British Colombia, Vancouver, BC, V5Z 4H4, Canada.

[128]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21218, USA.

[129]Bloorview Research Institute, University of Toronto, Toronto, ON, M4G 1R8, Canada.

[130]Dept. of Psychiatry, School of Medicine, University of California Davis, Davis, CA 95817, USA.

[131]Dept. of Behavioural Sciences, School of Medicine, University of California Davis, Davis, CA 95817, USA.

[132]Dept. of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[133]The Center for Applied Genomics and Division of Human Genetics, ChildrenÕs Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA.

[134]Dept of Pediatrics, University of Pennsylvania, Philadelphia, PA 19104, USA.

[135]Dept. of Psychiatry, Stanford University, Stanford, CA 94305, USA.

[136]Maine Medical Center Research Institute, Portland, ME 04074, USA.

[137]Center for Human Genetics Research, Vanderbilt University, Nashville, TN 37232, USA.

[†]Please address correspondence to: pgc.autism@gmail.com

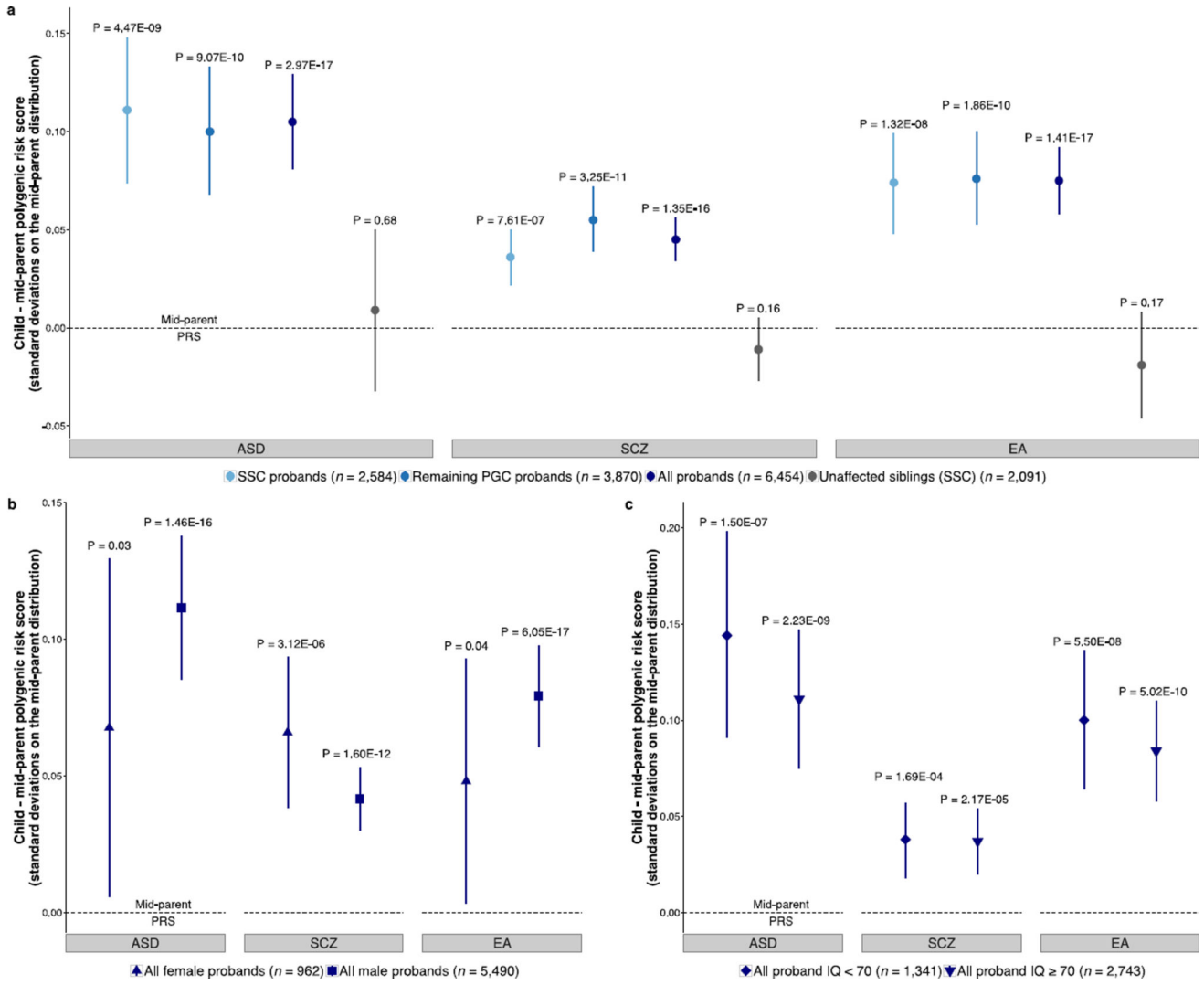

Figure 1.

ASD probands over inherit polygenic risk for ASD, schizophrenia, and greater educational attainment. Transmission disequilibrium is shown in terms of standard deviations on the mid-parent distribution ± 1.96 standard error (95% confidence intervals). P-values denote the probability that the mean of the pTDT deviation distribution is 0 (two-sided, one-sample t-test). (**a**) ASD probands over inherit ASD associated polygenic risk in the Simon Simplex Collection (SSC, *n* = 2,584), Psychiatric Genomics Consortium Autism Group (PGC ASD, *n* = 3,870), and combined cohorts (*n* = 6,454). Unaffected siblings in SSC (*n* = 2,091) do not over inherit ASD associated polygenic risk. (**b**) Both male (*n* = 5,490) and female (*n* = 962) probands over inherit ASD associated polygenic risk in the SSC+PGC ASD combined cohort. (**c**) ASD probands with (*n* = 1,341) and without (*n* = 2,743) intellectual disability (full-scale IQ < 70) over inherit ASD associated polygenic risk in the SSC+PGC ASD combined cohort.

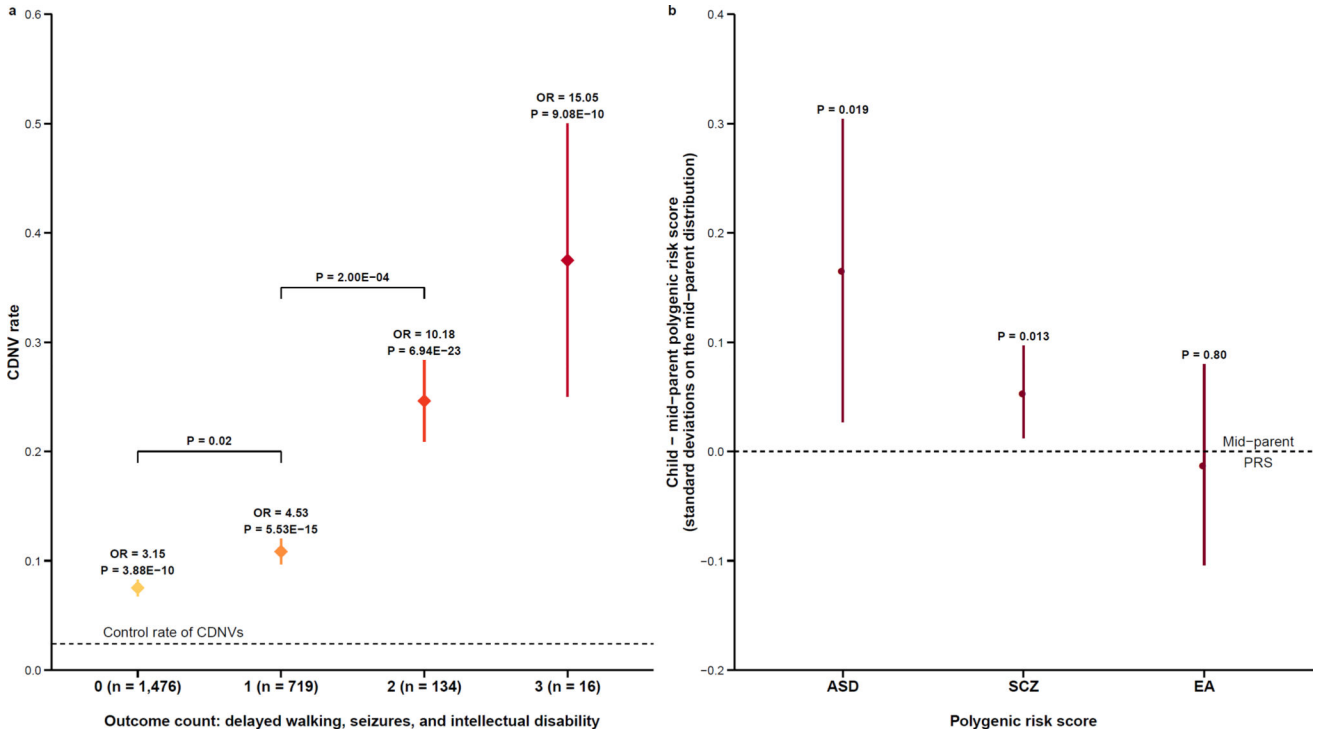

Figure 2.

Contributing *de novo* mutations are associated with adverse neurological and developmental outcomes and act additively with polygenic burden to influence ASD risk. (**a**) Simons Simplex Collection (SSC) probands are grouped by their count of the following: delayed walking (≥ 19 months); presence of seizures; intellectual disability (full-scale IQ < 70) (*n* = 1,476 with no outcomes; *n* = 719 with 1 outcome; *n* = 134 with 2 outcomes; *n* = 16 with 3 outcomes). Contributing *de novo* variant (CDNV) rate is calculated by dividing the count of CDNVs by the count of individuals. The odds ratio (OR) was calculated via Poisson regression predicting CDNV count from case/control status for all controls (*n* = 1,736) and cases in the outcome category, controlling for maternal and paternal age at birth of the child. P-values above each diamond are from the Poisson regression and indicate the probability that the CDNV rate in cases is not different from the CDNV rate in controls. P-values between the diamonds are calculated from Poisson exact test and indicate the probability that there is no difference in CDNV rate between the two noted groups. Error bars are ± 1 standard error. (**b**) pTDT analysis for SSC CDNV proband carriers (*n* = 221). Transmission disequilibrium is shown in terms of standard deviations on mid-parent distribution ± 1.96 standard error (95% confidence intervals). P-values denote the probability that the mean of the pTDT deviation distribution is 0 (two-sided, one-sample t-test).

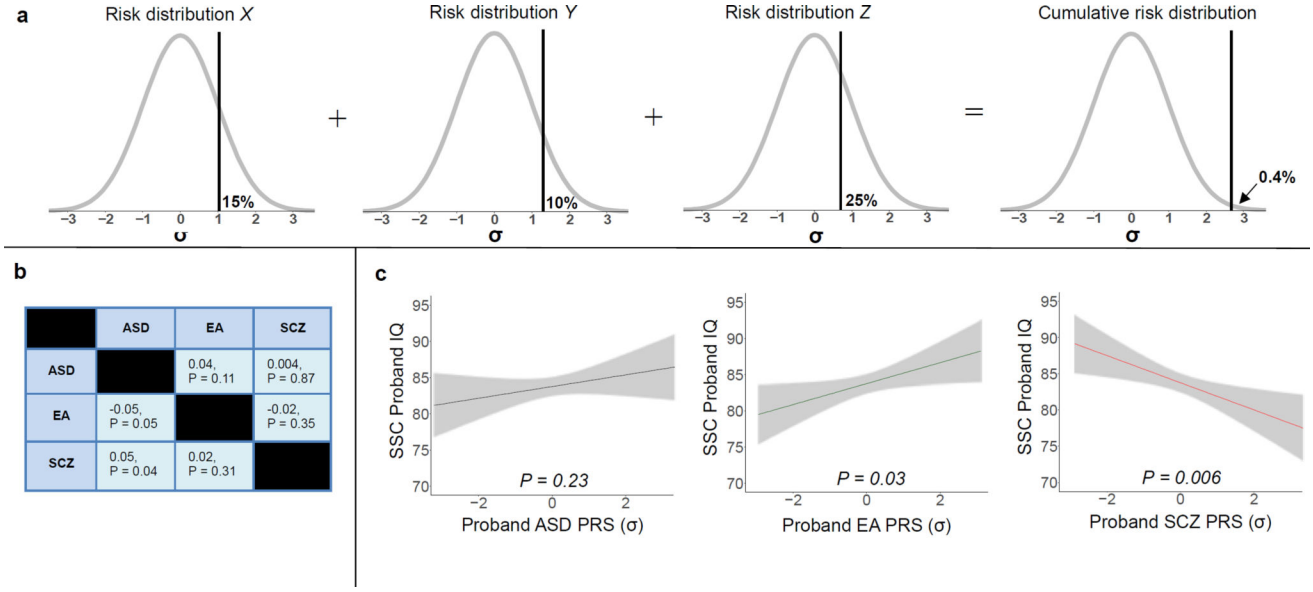

Figure 3.

(**a**) Additivity among orthogonal risk factors can yield high cumulative risk. (**b**) Polygenic risk scores (PRS) for ASDs, schizophrenia (SCZ), and educational attainment (EA) are not strongly associated at either the mid-parent level (above diagonal) or the pTDT deviation level (below diagonal). The table contains Pearson correlation coefficients and associated p-values indicating the probability with which the true correlation is 0. Mid-parent correlations are controlled for first 10 principal components of parental ancestry. PRS are from European ancestry SSC families (n = 1,851). (**c**) Polygenic risk factors for ASD exhibit independent, distinct effects on IQ in European ancestry SSC probands (n = 1,674). P-values, which estimate the probability of no association between each PRS and IQ, are calculated from linear regression. We predicted full-scale IQ from each PRS, z-normalized following residualization for the other two PRS, CDNV presence/absence, proband sex, and the first 10 principal components of proband ancestry. Each panel displays the linear association between full-scale proband IQ and the normalized PRS.