nature genetics

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

Autism spectrum disorder (ASD) risk is influenced by common polygenic and *de novo* variation. We aimed to clarify the influence of polygenic risk for ASD and to identify subgroups of ASD cases, including those with strongly 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 ASD, schizophrenia, and greater educational attainment is over-transmitted to children with ASD. These findings hold independent of proband IQ. We find that polygenic variation contributes additively to risk in ASD cases who carry a strongly 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 the genetic influences on ASD are additive and suggest that they create risk through at least partially distinct etiologic pathways.

Risk for ASD 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 have a strong effect on 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 (for example, seizures or 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 ASD has a different, perhaps surprising relationship with cognition. Polygenic risk for ASD 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 ASD is associated with higher IQ. These findings are difficult to interpret—on average, IQ in ASD is at least 1 s.d. below the population mean[1,15]. Further, ASD 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 ASD and to better understand the subgroups of individuals with ASD for whom polygenic variation contributes to risk. 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) ASD, (ii) schizophrenia, and (iii) years of educational attainment are unambiguously associated with ASD risk, independently 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 deleterious *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 etiological 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 (over-transmitted variants) are nominated for association with that trait—only ascertainment on the basis of a trait in the offspring and the association of the allele with the trait introduce 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, as 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

A full list of authors and affiliations appears at the end of the paper.

Received 28 November 2016; accepted 13 April 2017; published online 15 May 2017; doi:10.1038/ng.3863

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). PRSs are normally distributed in the population, which means that some degree of common variant risk for complex traits like ASD is present in all individuals. As 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' PRSs (mid-parent PRS). This expectation is broken when offspring are specifically ascertained for phenotypic deviation from their parents (**Supplementary Fig. 1**), for example, children who are much taller than their parents or are 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 this expectation, the average offspring PRS for an ASD-associated trait would be greater than the average mid-parent PRS. By comparing ASD individuals' PRSs for various traits against those of their unaffected parents, one can unambiguously associate specific types of common variant risk (for example, polygenic risk for schizophrenia and educational attainment) with ASD 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 and **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 ASD, educational attainment, and schizophrenia for all genotyped family members in the SSC and PGC ASD data sets using a standard approach (Online Methods)[19]. In addition to polygenic risk for ASD itself, we chose to examine polygenic risk for educational attainment and schizophrenia, 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 (GWAS) of each phenotype (**Supplementary Table 3**). The discovery GWAS (including for ASD) did not include any SSC or PGC ASD individuals. Using the calculated PRSs, 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' PRSs for any pair of PRSs and, on the basis of this, no PRS-based evidence of assortative mating (**Supplementary Tables 4** and **5**, and **Supplementary Note**). We also found no evidence that the mothers and fathers of children with ASD differed with regard to the common variant risk that they carried (**Supplementary Table 6** and **Supplementary Note**). Lastly, there was no evidence that mid-parent polygenic risk differed by proband sex (**Supplementary Table 7** and **Supplementary Note**).

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). In brief, for each trio, one averages the parent PRSs 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 s.d. higher than that of their parents.

RESULTS

Polygenic transmission disequilibrium

For ASD, schizophrenia, and educational attainment, we calculated the average pTDT deviation for each affected child in the SSC and PGC ASD cohorts, 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 ASD, educational attainment, and schizophrenia was significantly over-transmitted to affected children ($P < 1 \times 10^{-6}$ for all comparisons), but not to SSC unaffected siblings ($P > 0.05$ for all comparisons; **Fig. 1a**). This means that common polygenic risks defined in GWAS of ASD, schizophrenia, and greater educational attainment are each unambiguously associated with ASD. The results did not change when the SSC and PGC ASD samples were restricted to families of European ancestry (**Supplementary Figs. 2** and **3**, **Supplementary Table 8**, and **Supplementary Note**). We repeated the analysis in the SSC and PGC ASD samples using a PRS for body mass index (BMI)—a polygenic risk category unassociated with ASD[6]—as a negative control (Online Methods). As expected, we found that neither ASD probands nor unaffected siblings over-inherited the 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 was over-transmitted to both male and female probands ($P < 0.05$ for all comparisons; **Fig. 1b**). Each of the three PRSs was significantly and equally over-transmitted to ASD cases with measured IQs in the intellectual disability range (IQ < 70; **Fig. 1c**, Online Methods, and **Supplementary Tables 11** and **12**) when compared with those without intellectual disability. Most probands in the intellectual disability group did 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 the carriers of *de novo* events removed did not change the observed associations (Online Methods and **Supplementary Table 13**).

Common and rare variant additivity

To best interrogate additivity between common polygenic and rare, strongly 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 with ASD[21] (Online Methods). 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 the association of ASD with *de novo* deletions (Online Methods). In the SSC, we found that ASD were strongly associated with *de novo* deletions from two categories: (i) deletions that included a constrained gene predicted to be intolerant of heterozygous

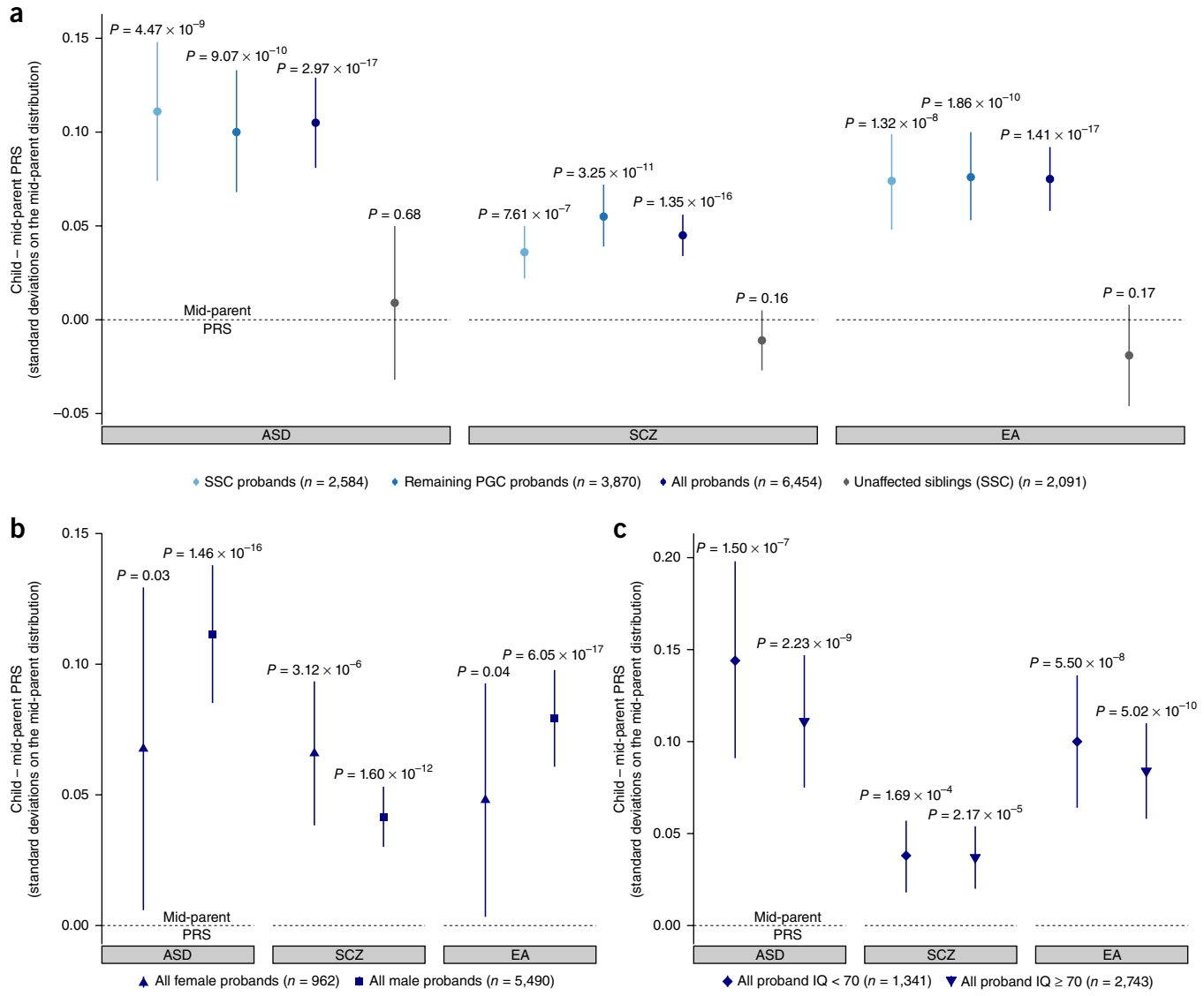

Figure 1 ASD probands over-inherit polygenic risk for ASD, schizophrenia, and greater educational attainment. Transmission disequilibrium is shown in terms of standard deviation on the mid-parent distribution ± 1.96 standard error (95% confidence intervals). *P* values denote the probability that the 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 SSC (n = 2,584), PGC ASD (n = 3,870), and combined (n = 6,454) cohorts. 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. ASD, autism spectrum disorders; SCZ, schizophrenia; EA, educational attainment.

loss-of-function variation (probability of loss-of-function intolerance (pLI) ≥ 0.9)[22] and (ii) uncommon, large (≥500-kb) deletions that did not contain a constrained gene (referred to hereafter as contributing CNV deletions; **Supplementary Figs. 4** and **5**).

Together, these refined classes of *de novo* PTVs and deletions form a category of strongly acting *de novo* variants (referred to hereafter as contributing *de novo* variants (CDNVs); odds ratio (OR) = 3.91, $P = 6.56 \times 10^{-20}$; case rate = 9.4%, control rate = 2.6%). Consistent with a hallmark of ASD-associated *de novo* variation, the rate of CDNVs in SSC varied substantially according to the probands' counts of co-occurring adverse neurodevelopmental outcomes, in this case the count of intellectual disability, a history of seizures, and the presence of delayed walking (**Fig. 2a**, Online Methods, and **Supplementary Table 15**). ASD probands without any of these comorbid traits were more than three times as likely (OR = 3.15; $P = 3.88 \times 10^{-10}$) to carry a CDNV than their unaffected siblings; ASD probands with all three of these comorbid traits were approximately 15 times as likely (OR = 15.05; $P = 9.08 \times 10^{-10}$) to carry a CDNV than their unaffected siblings. 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 estimated that, on average, the CDNVs defined here conferred an approximately 20-fold increase in risk for an ASD diagnosis (Online Methods). Their effect size, however, likely varies, as the male:female ratio of the carriers declined with increasing count of neurodevelopmental comorbidities (**Supplementary Table 16**). In other words, CDNVs seen in individuals with ASD exhibiting multiple comorbid neurodevelopmental traits are likely, on average, more deleterious than those seen in probands with ASD alone.

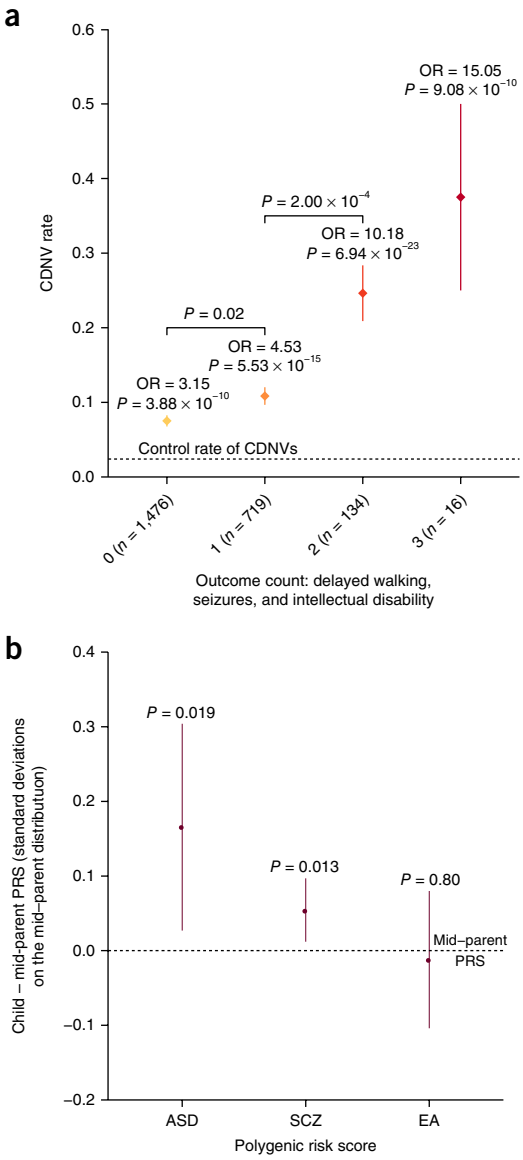

Figure 2 Contributing *de novo* mutations are associated with adverse neurological and developmental outcomes and act additively with polygenic burden to influence ASD risk. (**a**) SSC probands are grouped by their count of the following: delayed walking (≥19 months); presence of seizures; intellectual disability (full-scale IQ < 70) ($n = 1,476$ with no outcomes; $n = 719$ with one outcome; $n = 134$ with two outcomes; $n = 16$ with three outcomes). The CDNV rate is calculated by dividing the count of CDNVs by the count of individuals. The 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 were calculated by Poisson exact test and indicate the probability that there is no difference in CDNV rate between the two noted groups. Error bars, ±1 standard error. (**b**) pTDT analysis for SSC CDNV proband carriers ($n = 221$). Transmission disequilibrium is shown in terms of standard deviation 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).

We used pTDT to determine whether polygenic risk for ASD, schizophrenia, and educational attainment was also over-transmitted to CDNV carriers (**Fig. 2b** and Online Methods). ASD cases with a CDNV ($n = 221$ cases) indeed carried more polygenic risk for ASD and schizophrenia than expected on the basis of parent background ($P < 0.05$ for both comparisons). There was no evidence of over-transmission of the educational attainment PRS ($P = 0.80$), although 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 schizophrenia PRSs provides clear evidence for additivity between common and rare variation in creating risk for ASD. 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 ASD, one does not have to maintain an extreme position on any single distribution to carry a cumulatively uncommon amount of ASD risk (**Fig. 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 ASD, schizophrenia, and educational attainment is partially correlated[6]. However, we observed little evidence of correlation between the ASD, schizophrenia, and educational attainment PRSs in terms of either (i) correlation at the mid-parent level or (ii) correlation in the degree to which the scores were transmitted to the probands (pTDT deviation) (**Fig. 3b**). The lack of strong associations is likely a function of both ascertainment effects and attenuation due to limited predictive ability of the PRSs.

Given the limited association between the three PRSs, we were able to examine additivity among largely distinct common polygenic risk factors. First, we saw significant evidence that each of the three PRSs was independently over-transmitted ($P < 2 \times 10^{-4}$ in all SSC + PGC ASD comparisons; Online Methods and **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' relationships with proband IQ (**Supplementary Table 19**). We determined the association between each PRS and full-scale proband IQ in European-ancestry SSC probands, controlling for sex, presence of CDNVs, the other two PRSs, and the first ten principal components of ancestry (**Fig. 3c** and Online Methods). The educational attainment and schizophrenia PRSs were 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 PRSs are influencing ASD risk through at least partially different processes. Polygenic risk for schizophrenia influences ASD liability in a manner that negatively influences cognition, whereas polygenic risk for educational attainment 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 these associations, particularly the recently published—and counterintuitive—findings from genetic correlation analysis. Using pTDT, we have shown an unambiguous

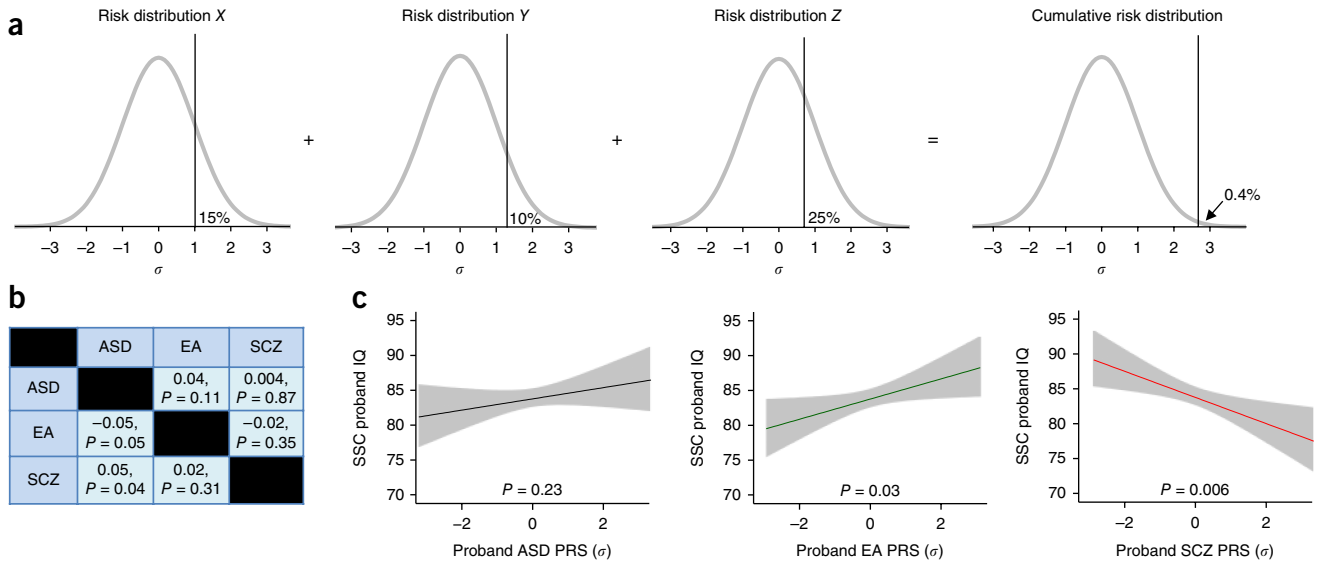

Figure 3 Polygenic risk factors for ASD are partially independent and differ in their relationship with cognition. (**a**) Additivity among orthogonal risk factors can yield high cumulative risk. (**b**) PRSs for ASD, SCZ, and educational attainment are not strongly associated at either the mid-parent level (above the diagonal) or the pTDT deviation level (below the 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 the first ten principal components of parental ancestry. PRSs 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, were calculated from linear regression. We predicted full-scale IQ from each PRS, *z* normalized following residualization for the other two PRSs, CDNV presence/absence, proband sex, and the first ten principal components of proband ancestry. Each panel displays the linear association between full-scale proband IQ and the normalized PRS.

association between ASD risk and the common polygenic influences on ASD 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. This 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 ASD might be different in some ways from an IQ measured below 70 in someone without ASD. 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 (for example, a strong interest in the arts or sciences that increases one's educational attainment) and how and when they are deleterious (for example, 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], which will likely need to be taken into account in effective research and treatment paradigms.

These findings also highlight important differences between the common and rare variant contributors to ASD risk. Strongly acting, *de novo* variant risk for ASD has an impact on a limited subset of cases. The phenotypic preferences of these 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 to be more pervasively influential among ASD cases. Common variant risk appears similarly relevant to ASD individuals with high and low IQ, and those with and without a strongly acting *de novo* mutation. The common polygenic influences also appear comparatively neurologically gentle. In fact, they are in many cases 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 in 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 (see URLs).

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; Ricopili, https://sites.google.com/a/broadinstitute.org/ricopili/home.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

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.B.R. and D.J.W. were funded by National Institute of Mental Health grant 1K01MH099286-01A1 and Brain Behavior Research Foundation (NARSAD) Young Investigator grant 22379. E.M.W. was funded by the Stanley Center for Psychiatric Research at the Broad Institute. A.O. was funded by an 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.J.D.), and the National Institute of Mental Health (5U01MH094432-02 to M.J.D.). This work was also supported by a grant from the Simons Foundation (SFARI 402281 to S.J.S.). 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).

AUTHOR CONTRIBUTIONS

D.J.W., E.M.W., S.R., R.K.W., J.A.K., J.G., K.E.S., J.I.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.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Sanders, S.J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
2. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
3. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
4. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
5. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
6. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
7. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
8. Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* **21**, 4781–4792 (2012).
9. Klei, L. *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* **3**, 9 (2012).
10. World Health Organization. WHO Motor Development Study: windows of achievement for six gross motor development milestones. *Acta Paediatr.* **450**, 86–95 (2006).
11. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
12. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
13. Clarke, T.K. *et al.* Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol. Psychiatry* **21**, 419–425 (2016).
14. Hagenaars, S.P. *et al.* Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (*N*=112 151) and 24 GWAS consortia. *Mol. Psychiatry* **21**, 1624–1632 (2016).
15. Robinson, E.B. *et al.* Autism spectrum disorder severity reflects the average contribution of *de novo* and familial influences. *Proc. Natl. Acad. Sci. USA* **111**, 15161–15165 (2014).
16. Munafo, M.R., Tilling, K., Taylor, A.E., Evans, D.M. & Davey Smith, G. Collider Scope: how selection bias can induce spurious associations. Preprint at *bioRxiv* http://dx.doi.org/10.1101/079707 (2016).
17. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
18. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
19. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
20. Cross-Disorder Group of the Psychiatric Genomic Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
21. Kosmicki, J.A. *et al.* Refining the role of *de novo* protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
22. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
23. Samocha, K.E. *et al.* A framework for the interpretation of *de novo* mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
24. Robinson, E.B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).

Daniel J Weiner[1–3], Emilie M Wigdor[1–3], Stephan Ripke[1–4], Raymond K Walters[1,2], Jack A Kosmicki[1–3,5], Jakob Grove[6–9], Kaitlin E Samocha[1–3], Jacqueline I Goldstein[1–3], Aysu Okbay[10], Jonas Bybjerg-Grauholm[7,11], Thomas Werge[7,12,13], David M Hougaard[11], Jacob Taylor[1–3,14], 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[6–8], George Davey Smith[21], Mark J Daly[1–3] & Elise B Robinson[1–3]

[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, the Netherlands. [11]Danish Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum 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, 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. Correspondence should be addressed to E.B.R. (erob@broadinstitute.org).

iPSYCH-Broad Autism Group

Marie Bækvad-Hansen[7,11], Ashley Dumont[2], Christine Hansen[7,11], Thomas F Hansen[7,12], Daniel Howrigan[1–3], Manuel Mattheisen[6–8], Jennifer Moran[2], Ole Mors[7,22], Merete Nordentoft[7,23], Bent Nørgaard-Pedersen[7,11], Timothy Poterba[1–3], Jesper Poulsen[7,11] & Christine Stevens[2]

Psychiatric Genomics Consortium Autism Group

Verneri Anttila[1,2], Peter Holmans[24], Hailiang Huang[1,2], Lambertus Klei[17], Phil H Lee[1,2,25], Sarah E Medland[26], Benjamin Neale[1,2], Lauren A Weiss[27,28], Lonnie Zwaigenbaum[29], Timothy W Yu[30], Kerstin Wittemeyer[31], A Jeremy Willsey[27], Ellen M Wijsman[32,33], Thomas H Wassink[34], Regina Waltes[35], Christopher A Walsh[5,30,36–38], Simon Wallace[39], Jacob A S Vorstman[40], Veronica J Vieland[41], Astrid M Vicente[42,43], Herman van Engeland[40], Kathryn Tsang[27,28], Ann P Thompson[44], Peter Szatmari[45], Oscar Svantesson[46], Stacy Steinberg[47], Kari Stefansson[47], Hreinn Stefansson[47], Matthew W State[27], Latha Soorya[48–50], Teimuraz Silagadze[51], Stephen W Scherer[52–54], Gerard D Schellenberg[55], Sven Sandin[56], Evald Saemundsen[56], Guy A Rouleau[57], Bernadette Rogé[58], Kathryn Roeder[59,60], Wendy Roberts[61], Jennifer Reichert[48,49], Abraham Reichenberg[48,49], Karola Rehnström[62], Regina Regan[63,64], Fritz Poustka[35], Christopher S Poultney[48,49], Joseph Piven[65], Dalila Pinto[48,49,66–69], Margaret A Pericak-Vance[70], Milica Pejovic-Milovancevic[71], Marianne G Pedersen[7,20,72], Carsten B Pedersen[7,20,72], Andrew D Paterson[52,54,73], Jeremy R Parr[74,75], Alistair T Pagnamenta[76], Guiomar Oliveira[77,78], John I Nurnberger[79–81], Merete Nordentoft[7,23], Michael T Murtha[82], Susana Mouga[77,78], Ole Mors[7,22], Eric M Morrow[83], Daniel Moreno De Luca[83], Anthony P Monaco[76,84], Nancy Minshew[17], Alison Merikangas[85], William M McMahon[86], Susan G McGrew[87], Manuel Mattheisen[7,8,88], Igor Martsenkovsky[89], Donna M Martin[90], Shrikant M Mane[91], Pall Magnusson[92], Tiago Magalhaes[63,64], Elena Maestrini[93], Jennifer K Lowe[94–96], Catherine Lord[97], Pat Levitt[98], Christa Lese Martin[99], David H Ledbetter[100], Marion Leboyer[101–104], Ann S Le Couteur[74,75], Christine Ladd-Acosta[105], Alexander Kolevzon[48,49,67,69], Sabine M Klauck[106], Suma Jacob[107,108], Bozenna Iliadou[46], Christina M Hultman[46], Irva Hertz-Picciotto[109,110], Robert Hendren[27], Christine S Hansen[7,11], Jonathan L Haines[111], Stephen J Guter[107], Dorothy E Grice[49], Jonathan M Green[112,113], Andrew Green[64,114], Arthur P Goldberg[66,68], Christopher Gillberg[115], John Gilbert[70], Louise Gallagher[85], Christine M Freitag[35], Eric Fombonne[116], Susan E Folstein[117], Bridget Fernandez[118], M Daniele Fallin[119], A Gulhan Ercan-Sencicek[82], Sean Ennis[64,114], Frederico Duque[77,78], Eftichia Duketis[35], Richard Delorme[101,120–122], Silvia De Rubeis[48,49], Maretha V De Jonge[40], Geraldine Dawson[123,124], Michael L Cuccaro[70], Catarina T Correia[42,43], Judith Conroy[64,125], Inês C Conceição[42,43], Andreas G Chiocchetti[35], Patrícia B S Celestino-Soper[80,126], Jillian Casey[64,125], Rita M Cantor[127,128], Cátia Café[77], Sean Brennan[85], Thomas Bourgeron[101,120,121,129], Patrick F Bolton[130,131], Sven Bölte[35,132,133], Nadia Bolshakova[85], Catalina Betancur[134–136], Raphael Bernier[137], Arthur L Beaudet[126], Agatino Battaglia[138], Vanessa H Bal[27], Gillian Baird[139], Anthony J Bailey[39,140], Marie Bækvad-Hansen[7,11], Joel S Bader[141], Elena Bacchelli[93], Evdokia Anagnostou[142], David Amaral[110,143,144], Joana Almeida[77], Joseph D Buxbaum[48,49,66,67,69,145], Aravinda Chakravarti[141], Edwin H Cook[107], Hilary Coon[86], Daniel H Geschwind[94–96,128], Michael Gill[85], Hakon Hakonarson[146,147], Joachim Hallmayer[148], Aarno Palotie[62], Susan Santangelo[149], James S Sutcliffe[111,150] & Dan E Arking[141]

[22]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [23]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark. [24]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [25]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA. [26]Queensland Institute of Medical Research, Brisbane, Queensland, Australia. [27]Department of Psychiatry, University of California, San Francisco, San Francisco, California, USA. [28]Institute of Human Genetics, University of California, San Francisco, San Francisco, California, USA. [29]Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada. [30]Division of Genetics, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts, USA. [31]School of Education, University of Birmingham, Birmingham, UK. [32]Department of Medicine, University of Washington, Seattle, Washington, USA. [33]Department of Biostatistics, University of Washington, Seattle, Washington, USA. [34]Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa, USA. [35]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, JW Goethe University Frankfurt, Frankfurt am Main, Germany. [36]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA. [37]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA. [38]Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA. [39]Department of Psychiatry, University of Oxford and Warneford Hospital, Oxford, UK. [40]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands. [41]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio, USA. [42]Instituto Nacional de Saœde Dr. Ricardo Jorge, Lisboa, Portugal. [43]Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, Lisboa, Portugal. [44]Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, Ontario, Canada. [45]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [46]Karolinska Institutet, Solna, Sweden. [47]deCODE Genetics, Reykjavik, Iceland. [48]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [49]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [50]Department of Psychiatry, Rush University Medical Center, Chicago, Illinois, USA. [51]Department of Psychiatry and Drug Addiction, Tbilisi State Medical University, Tbilisi, Georgia. [52]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. [53]McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada. [54]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [55]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [56]State Diagnostic and Counseling Centre, Kopavogur, Iceland. [57]Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada. [58]Centre d'Etudes et de Recherches en Psychopathologie, Toulouse University, Toulouse, France. [59]Department of Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [60]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [61]Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario, Canada. [62]Sanger Institute, Hinxton, UK. [63]National Childrens Research Centre, Our Lady's Hospital Crumlin, Dublin, Ireland. [64]Academic Centre on Rare Diseases, University College Dublin, Dublin, Ireland. [65]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. [66]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [67]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [68]The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [69]Friedman Brain

Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [70]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA. [71]Institute of Mental Health and Medical Faculty, University of Belgrade, Belgrade, Serbia. [72]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark. [73]Dalla Lana School of Public Health, Toronto, Ontario, Canada. [74]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK. [75]Institue of Health and Science, Newcastle University, Newcastle upon Tyne, UK. [76]Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, UK. [77]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, Portugal. [78]University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, Faculty of Medicine, University of Coimbra, Coimbra, Portugal. [79]Institute of Psychiatric Research, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana, USA. [80]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [81]Program in Medical Neuroscience, Indiana University School of Medicine, Indianapolis, Indiana, USA. [82]Programs on Neurogenetics, Yale University School of Medicine, New Haven, Connecticut, USA. [83]Department of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island, USA. [84]Tufts University, Boston, Massachusetts, USA. [85]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [86]Department of Psychiatry, University of Utah, Salt Lake City, Utah, USA. [87]Department of Pediatrics, Vanderbilt University, Nashville, Tennessee, USA. [88]Department of Biomedicine–Human Genetics, Aarhus University, Aarhus, Denmark. [89]Department of Child, Adolescent Psychiatry and Medical-Social Rehabilitation, Ukrainian Research Institute of Social Forensic Psychiatry and Drug Abuse, Kyiv, Ukraine. [90]Department of Pediatrics and Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. [91]Yale Center for Genomic Analysis, Yale University School of Medicine, New Haven, Connecticut, USA. [92]Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland. [93]Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy. [94]Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine at the University of California, Los Angeles, Los Angeles, California, USA. [95]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [96]Center for Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [97]Department of Psychiatry, Weill Cornell Medical College, Cornell University, New York, New York, USA. [98]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [99]Autism & Developmental Medicine Institute, Geisinger Health System, Danville, Pennsylvania, USA. [100]Geisinger Health System, Danville, Pennsylvania, USA. [101]FondaMental Foundation, Créteil, France. [102]INSERM U955, Paris, France. [103]Faculté de Médecine, Université Paris Est, Créteil, France. [104]Department of Psychiatry, Henri Mondor–Albert Chenevier Hospital, Assistance Publique–Hôpitaux de Paris, Créteil, France. [105]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [106]Division of Molecular Genome Analysis and Working Group Cancer Genome Research, Deutsches Krebsforschungszentrum, Heidelberg, Germany. [107]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA. [108]Institute of Translational Neuroscience and Department of Psychiatry, University of Minnesota, Minneapolis, Minnesota, USA. [109]Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, California, USA. [110]The MIND Institute, School of Medicine, University of California, Davis, Davis, California, USA. [111]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA. [112]Manchester Academic Health Sciences Centre, Manchester, UK. [113]Institute of Brain, Behaviour, and Mental Health, University of Manchester, Manchester, UK. [114]Centre for Medical Genetics, Our Lady's Hospital Crumlin, Dublin, Ireland. [115]Gillberg Neuropsychiatry Centre, University of Gothenburg, Gothenburg, Sweden. [116]Department of Psychiatry and Institute for Development and Disability, Oregon Health & Science University, Portland, Oregon, USA. [117]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Miller School of Medicine, University of Miami, Miami, Florida, USA. [118]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada. [119]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [120]Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France. [121]Centre National de la Recherche Scientifique, URA 2182, Institut Pasteur, Paris, France. [122]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique–Hôpitaux de Paris, Paris, France. [123]Duke Center for Autism and Brain Developments, Duke University School of Medicine, Durham, North Carolina, USA. [124]Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, North Carolina, USA. [125]Temple Street Children's University Hospital, Dublin, Ireland. [126]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA. [127]Department of Psychiatry, David Geffen School of Medicine at the University of California, Los Angeles, Los Angeles, California, USA. [128]Department of Human Genetics, David Geffen School of Medicine at the University of California, Los Angeles, Los Angeles, California, USA. [129]University Paris Diderot, Sorbonne Paris Cité, Paris, France. [130]Institute of Psychiatry, Kings College London, London, UK. [131]South London and Maudsley Biomedical Research Centre for Mental Health, London, UK. [132]Department of Women's and Children's Health, Center of Neurodevelopmental Disorders, Karolinska Institutet, Stockholm, Sweden. [133]Child and Adolescent Psychiatry, Center for Psychiatry Research, Stockholm County Council, Stockholm, Sweden. [134]INSERM U1130, Paris, France. [135]CNRS UMR 8246, Paris, France. [136]Sorbonne Universités, UPMC Université Paris 6, Neuroscience Paris Seine, Paris, France. [137]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA. [138]Stella Maris Institute for Child and Adolescent Neuropsychiatr, Pisa, Italy. [139]Paediatric Neurodisability, King's Health Partners, Kings College London, London, UK. [140]Mental Health and Addictions Research Unit, University of British Colombia, Vancouver, British Columbia, Canada. [141]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA. [142]Bloorview Research Institute, University of Toronto, Toronto, Ontario, Canada. [143]Department of Psychiatry, School of Medicine, University of California, Davis, Davis, California, USA. [144]Department of Behavioural Sciences, School of Medicine, University of California, Davis, Davis, California, USA. [145]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [146]The Center for Applied Genomics and Division of Human Genetics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. [147]Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [148]Department of Psychiatry, Stanford University, Stanford, California, USA. [149]Maine Medical Center Research Institute, Portland, Maine, USA. [150]Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA.

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 SSC is a resource of more than 2,500 families with a child diagnosed with an ASD[18]. Informed consent and assent were 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 (ref. 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 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 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 the Autism Center of Excellence at UCLA (n = 215), the Autism Genome Project (n = 2,254), the 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 Project reference panel and Ricopili pipeline, which are publicly available and have been reported on extensively[19,25].

Although trio approaches are broadly immune to confounding through ancestry, we isolated European-only subsets of both the SSC and PGC ASD samples to (i) ensure that our primary results did not change in an ancestrally homogeneous subset of the data and (ii) conduct comparisons across probands or across parents, which might be sensitive to ancestry (as opposed to comparisons between the proband 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 these individuals with the HapMap 3 data set[26] and generated principal components of ancestry using GCTA[27]. Through visual inspection (**Supplementary Fig. 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 (for example, 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 families in which both parents were of European ancestry. To do so, we merged the PGC ASD parent data with HapMap 3 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 Fig. 7**). Both parents were of European ancestry in 3,209 families, and these 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 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 (for example, schizophrenia or educational attainment)[28]. 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, schizophrenia, and BMI for all individuals with available genotype data in the SSC and PGC ASD data sets. 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 schizophrenia and educational attainment because of their robust, well-replicated associations with ASD risk[6,14,20]. We selected BMI as a negative control because of 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 ten genotyping waves of the iPSYCH-Broad Autism project[24]. The schizophrenia summary statistics were from the 2014 GWAS of 36,989 cases and 113,075 controls from the Psychiatric Genomics Consortium[19]. The educational attainment 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 BMI (n = 322,154, European-ancestry meta-analysis)[29].

To construct the PRSs, we first gathered the summary statistics from the GWAS described above. The summary statistics included effect sizes and P values for each 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 kb of and correlated with ($r^2 \geq 0.1$) a more significantly associated SNP in the GWAS. We used the 1000 Genomes Project 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 the SNPs remaining pass the genome-wide significance threshold of $P = 5.00 \times 10^{-8}$; the majority of the signal resides in SNPs that do not pass the significance threshold and cannot be specifically identified[6,12,19,27]. 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: $P \leq 1$, 5×10^{-1}, 2×10^{-1}, 1×10^{-1}, 5×10^{-2}, 1×10^{-2}, 1×10^{-3}, 1×10^{-4}, 1×10^{-6}, and 5×10^{-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 $P \leq 0.1$ 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 schizophrenia and educational attainment PRSs, we used the thresholds 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 $P \leq 0.05$ most commonly explained the most schizophrenia case–control variance in 40 leave-one-out analyses[19]. For educational attainment, constructing a PRS using a threshold of $P \leq 1$ explained the most variance in the number of years of education achieved in an independent sample[12]. For BMI, constructing a PRS using a threshold of $P \leq 0.2$ explained the most variance in phenotypic BMI among cases in the SSC (**Supplementary Note**). These thresholds created the strongest PRSs for ASD, educational attainment, schizophrenia, and BMI, leaving us maximally powered to investigate the relationship between these four traits and ASD.

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 (for example, analysis of *de novo* variation) where our info threshold of 0.6 was the minimum in the SSC imputation only. For each trait, the PRS for individuals in SSC and PGC ASD was then calculated as the product of the GWAS effect size (log odds ratio or β value) at that SNP and 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[30]. If an individual was missing genetic data at a SNP in the summary statistics file, PLINK calculated the expected score on the basis of the cohort-wide allele frequency.

pTDT. We define pTDT deviation as

$$\text{pTDT deviation} = \frac{PRS_c - PRS_{MP}}{SD(PRS_{MP})}$$

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

$$PRS_{MP} = \frac{PRS_{mother} + PRS_{father}}{2}$$

$SD(PRS_{MP})$ is the standard deviation of the sample-specific mid-parent PRS. For example, in SSC analysis, $SD(PRS_{MP})$ 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 PRSs. 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 PRSs from unaffected siblings, but using mid-parent PRSs improves statistical power (**Supplementary Table 20** and **Supplementary Note**).

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

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

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

We performed pTDT using the ASD, educational attainment, schizophrenia, and BMI PRSs 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 (**Fig. 1a**), each of the ASD, educational attainment, and schizophrenia PRSs was significantly over-transmitted from parents to probands, but not to unaffected siblings ($P < 1 \times 10^{-6}$ for all parent to proband comparisons in either SSC or PGC ASD; $P < 1 \times 10^{-15}$ for all parent to proband comparisons in combined SSC and PGC ASD; $P > 0.05$ for all unaffected sibling comparisons). In contrast, the 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 and control labels to the affected and unaffected children in SSC. We then counted the number of times that the simulated difference in the 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 **Fig. 1a**).

In our primary associations (**Fig. 1a** and **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 (**Fig. 1b**,**c**). In SSC, full-scale IQ (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[31]; the Mullen Scales of Early Learning[32]; and the Wechsler battery[33]. 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[34]. Full-scale IQ was measured heterogeneously across the contributing PGC ASD cohorts (**Supplementary Table 2**). To accommodate measurement differences, full-scale IQ 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 full-scale IQ levels 1–4 as follows: 1, full-scale IQ < 35; 2, 35 ≤ full-scale IQ < 70; 3, 70 ≤ full-scale IQ < 90; 4, 90 ≤ full-scale IQ. In PGC ASD, 38.9% of probands had estimated IQ available. We categorized probands in SSC and PGC ASD separately by presence/absence of intellectual disability so that the data sets 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 full-scale IQ data available in PGC ASD (**Supplementary Tables 11** and **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 were available. Our previous work has identified a subclass of *de novo* PTVs (frameshift variants, splice-acceptor variants, splice-donor variants, nonsense variants) that are a primary source of association with ASD[21]. *De novo* PTVs in this class are (i) absent from the Exome Aggregation Consortium database, a reference sample of over 60,000 exomes, and (ii) 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 ($P = 4.12 \times 10^{-14}$). PTVs outside of this contributing class were unassociated with ASD risk (observed in 7.8% of cases and 6.9% of unaffected siblings, $P = 0.50$) and were not associated with proband IQ ($P = 0.76$) (**Supplementary Fig. 8**).

As detailed in the main text, *de novo* 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 ($P = 1.80 \times 10^{-8}$). 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 Fig. 5**). The associations between CNV categories and ASD risk 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 Fig. 9**). In contrast, *de novo* duplications of constrained genes were not disproportionately associated with ASD risk ($P = 0.49$; **Supplementary Fig. 10** and **Supplementary Note**).

We identified a subset of *de novo* CNV deletions, primarily those containing loss-of-function-intolerant genes, that accounted for most of the category's association with ASD (contributing CNV deletions; **Supplementary Fig. 4**). Together, contributing PTVs and contributing CNV deletions formed a strongly acting *de novo* variant category in SSC, described herein as contributing *de novo* variants (CDNVs). Multiple lines of evidence suggest that 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 ($P = 6.56 \times 10^{-20}$). CDNVs were very strongly associated with proband IQ ($P = 5.80 \times 10^{-6}$; 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 Fig. 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.*[3]. 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 as in male probands (17.4% versus 8.5%, $P = 5.26 \times 10^{-6}$; **Supplementary Fig. 12**), we estimate an odds ratio of approximately 20 (ref. 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 (**Fig. 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 are positively associated with history of seizures[1,15]. Motor delays are an additional neurodevelopmental comorbitidy associated with ASD[35]. 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[10]. As hypothesized, motor delay, seizures, and low IQ were independently associated with CDNV rate in SSC probands after controlling for proband sex (**Supplementary Ta ble 15**). CDNV rate was positively associated with the number of these adverse outcomes experienced by probands (**Fig. 2a** and **Supplementary Table 16**). The decreasing male:female ratio among CDNV carriers as the count of co-occurring neurodevelopmental outcomes increases suggests that with an increased count of neurodevelopmental comorbidities comes, on average, increasingly severe *de novo* events (**Supplementary Table 16** and **Supplementary Note**). 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 (**Fig. 2b** and **Supplementary Table 13**). As this CDNV analysis is specific to SSC, the PRSs generated for the CDNV analysis used info score cutoffs from SSC imputation to increase the number of well-imputed SNPs retained for the PRSs. We restricted our analysis to families with genotyped parents and probands with both genotype and exome sequence data (n = 2,346; n = 221 with CDNVs). The cohort of probands with a CDNV

doi:10.1038/ng.3863

was too small to determine whether the difference in transmission between probands that did and did not carry a CDNV was statistically significant. We also analyzed whether polygenic over-transmission was seen in a more broadly defined set of *de novo* events (**Supplementary Table 13** and **Supplementary Note**).

Genetic heterogeneity and proband phenotype. We analyzed whether polygenic risk for ASD, educational attainment, and schizophrenia were each independently over-transmitted to ASD cases. For SSC (n = 2,584 trios), PGC ASD (n = 3,870 trios), and SSC and PGC ASD combined (n = 6,454 trios), we performed a single logistic regression predicting proband/mid-parent status from polygenic risk for (i) ASD, (ii) educational attainment, and (iii) schizophrenia. We confirmed that the over-transmission acted independently in each PRS ($P < 2 \times 10^{-4}$ for all PRSs in the combined SSC and PGC ASD cohort; **Supplementary Table 18**).

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

Code availability. pTDT software is available at https://github.com/ypaialex/ptdt.

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

25. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
26. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
27. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
28. Wray, N.R., Goddard, M.E. & Visscher, P.M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17**, 1520–1528 (2007).
29. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
30. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
31. Elliott, C. *Differential Ability Scales* (The Psychological Corporation, 2007).
32. Mullen, E. *Mullen Scales of Early Learning* (American Guidance Service, 1995).
33. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence* (Psychological Corporation, 1999).
34. 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? *Biol. Psychiatry* **77**, 775–784 (2015).
35. Provost, B., Lopez, B.R. & Heimerl, S. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. *J. Autism Dev. Disord.* **37**, 321–328 (2007).