

## Exploring the common genetic architecture of autism spectrum disorder using a novel multi-polygenic risk score approach

Zoe Schmilovich<sup>1,2</sup>, Vincent-Raphaël Bourque<sup>3</sup>, Guillaume Huguet<sup>3</sup>, Qin He<sup>4</sup>, Jay P. Ross<sup>1,2</sup>, Martineau Jean-Louis<sup>3</sup>, Zohra Saci<sup>3</sup>, Boris Chaumette<sup>4,5,7</sup>, Patrick A. Dion<sup>2,6,7</sup>, Sébastien Jacquemont<sup>3</sup>, Guy A. Rouleau<sup>1,2,6,7</sup>

1. Department of Human Genetics, McGill University, Montréal, QC, H3A 1B1, Canada
2. Montreal Neurological Institute-Hospital, Montréal, QC, H3A 2B4, Canada
3. CHU Sainte-Justine, Université de Montréal, Montréal, QC, H3T 1C5, Canada
4. Université Paris Cité, Institute of Psychiatry and Neuroscience of Paris (IPNP), INSERM U1266, 75 014 Paris, France
5. GHU Paris Psychiatrie et Neurosciences, 75014 Paris, France
6. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, H3A 1A1, Canada
7. Department of Psychiatry, McGill University, Montréal, QC, H3A 1B1, Canada

Corresponding author: Dr. Guy Rouleau, Montréal Neurological Institute-Hospital, Department of Neurology and Neurosurgery, McGill University, 3801 University St, Montréal, QC, Canada, H3A 2B4, Tel: +1 514 398-2690, Fax: +1 514 398 8248, [guy.rouleau@mcgill.ca](mailto:guy.rouleau@mcgill.ca)

## ABSTRACT

Compared to disorders of similar heritability and contribution of common variants, few genome-wide significant loci have been implicated in autism spectrum disorder (ASD). This undermines the use of polygenic risk scores (PRSs) to investigate the common genetic architecture of ASD. Deconstructing PRS-ASD into its related traits via “developmental deconstruction” could reveal the underlying genetic liabilities of ASD. Using the data of >24k individuals with ASD and >28k of their unaffected family members from the SSC, SPARK, and MSSNG cohorts, we computed the PRSs for ASD and 11 genetically-related traits. We applied an unsupervised learning approach to the ASD-related PRSs to derive “multi-PRSs” that captured their variability in orthogonal dimensions. We found that multi-PRSs captured a similar proportion of genetic risk for ASD in cases versus intrafamilial controls ( $OR_{\text{multi-PRS}}=1.10$ ,  $R^2=0.501\%$ ), compared to PRS-ASD itself ( $OR_{\text{PRS-ASD}}=1.16$ ,  $R^2=0.619\%$ ). While multi-PRS dimensions conferred risk for ASD, they had “mirroring” effects on developmental phenotypes among cases with ASD. We posit that this phenomenon may partially account for the paucity of genome-wide significant loci and the clinical heterogeneity of ASD. This approach can serve as a proxy for PRS-ASD in cases where non-overlapping and well-powered GWAS summary statistics are difficult to obtain, or accounting for heterogeneity in a single dimension is preferable. This approach may also capture the overall liability for a condition (*i.e.*: genetic “P-factor”). Altogether, we present a novel approach to studying the role of inherited, additive, and non-specific genetic risk factors in ASD.

## KEYWORDS

Autism spectrum disorder; polygenic risk score; principal component analysis; general psychopathology factor

## INTRODUCTION

Autism spectrum disorder (ASD) has a high heritability ranging from 0.65 to 0.91<sup>1-3</sup>. Rare *de novo* copy-number variants (CNVs) and single-nucleotide variants (SNVs) in highly constrained genes of large effect sizes have been implicated in ASD, yet they collectively explain <5% of the overall liability for the disorder<sup>4-6</sup>. In contrast, it is estimated that >50% of the liability for ASD resides in common variation<sup>4-9</sup>. Despite its large contribution to genetic risk, the common variant genetic architecture of ASD remains elusive.

Genome-wide association studies (GWASs) have been instrumental in identifying common risk loci in disorders<sup>10</sup>. Despite the substantial increase in the sample size of ASD GWASs over the last decade<sup>11-16</sup>, only five genome-wide significant loci have been implicated in ASD to date<sup>17</sup>. The number of risk loci identified in ASD is markedly lower than in other psychopathologies. For instance, schizophrenia has a similar narrow sense heritability (81%)<sup>18</sup> and an estimated contribution of common genetic variation that is comparable to ASD<sup>19</sup>. However, >100 genome-wide significant loci have been implicated in schizophrenia. One potential explanation for this discrepancy is the difference in study sample sizes between the two disorders: the schizophrenia GWAS includes more cases than the ASD GWAS (*i.e.*: >36k cases in schizophrenia versus 18k in ASD). Some projections on the genetic architecture of these disorders suggest that ~70k cases could suffice to yield 100 genome-wide significant loci in ASD<sup>20</sup>. However, this would still represent a steep improvement curve, and recent analyses do not appear to follow such projections<sup>21</sup>. This questions whether other factors besides solely sample sizes could contribute to the lower number of implicated loci in ASD.

We hypothesize that the complex and heterogeneous genetic architecture of ASD may partially explain the paucity of genome-wide significant hits. The scarcity of ASD GWAS hits may, in turn, limit the accuracy of polygenic risk scores (PRSs), which rely on summary statistics of variant associations to generate individual-level genetic susceptibility scores for a trait<sup>22</sup>.

The phenotypic overlap<sup>23,24</sup> and genetic correlation<sup>17,25,26</sup> between ASD and diverse developmental phenotypes have been well-documented. In fact, it has been suggested that ASD may arise as a result of increased inherited genetic susceptibility for various developmental phenotypes<sup>27</sup>. As such, we sought to study the genetic architecture of ASD via “developmental deconstruction”, *i.e.*: deconstructing the unitary ASD syndrome liability into its contributory developmental phenotypes, both ASD-specific and non-specific factors<sup>28</sup>. Mous and colleagues previously showed that background susceptibilities for ADHD and for motor

coordination that are inherited, associated, but non-specific to ASD (referred to as “BASINS”), may contribute to the additive genetic liability for ASD in the same way that ASD-specific risk factors could contribute to a diagnosis for ASD<sup>27</sup>. More recently, Warrier and colleagues showed that common variants associated with ADHD and educational attainment (*i.e.*: non-ASD-specific risk factors) contribute to several core features of ASD<sup>29</sup>. However, these studies have focused on univariate associations between the polygenic risk of one trait (ASD) and its core symptoms. Previous studies did not account for the fact that many genetic variants are represented – at different degrees of association or weighting – in more than one of these PRSs. These studies also did not account for the effect of these genetic liabilities on developmental phenotypes outside of the core symptoms of ASD. Accounting for the effect of multiple ASD-related PRSs on ASD risk and related developmental phenotypes in a single model is a knowledge gap that remains unexplored.

We posit that modelling ASD and its associated features as a function of multiple ASD-related PRSs via developmental deconstruction could: 1–serve as a proxy for PRS-ASD, and; 2–highlight the heterogeneity underlying the common genetic architecture of ASD.

To do this, we analyzed a sample of 28,307 cases with ASD and 50,953 of their unaffected relatives across three ASD cohorts. First, we computed the PRSs for ASD and 11 of its genetically correlated traits, as reported in the Grove *et al.* ASD GWAS<sup>17</sup>. Then, we applied an unsupervised learning algorithm (principal component analysis, PCA) to construct PRSs that captured the variation of the 11 ASD-related traits across orthogonal principal components (PCs) that we refer to as “multi-PRS” dimensions.

This study provides support for the use of multi-PRS dimensions (constructed from ASD-related traits) to capture the additive, inherited, and non-specific genetic liability for ASD. First, we showed that the multi-PRS dimensions can capture a similar proportion of the inherited genetic liability of ASD risk compared to PRS-ASD itself (0.501% versus 0.619%). Second, we modelled the effect of PRS-ASD and multi-PRS liability on 46 developmental phenotypes among cases with ASD. Our results reveal that the multi-PRS dimensions can capture unique phenotypic differences among the cases with ASD that PRS-ASD cannot. Interestingly, while multi-PRS dimensions increased the risk for ASD, they had “mirroring” effects on core ASD symptoms, developmental features, and the risk for co-occurring disorders. These findings provide support for the clinical and genetic heterogeneity of ASD, which in turn, may partially explain the paucity of reproducible ASD GWAS hits.

## MATERIALS AND METHODS

### Cohorts

In total, the genetic data of 28,307 individuals with a diagnosis of ASD and 50,953 of their unaffected siblings and parents were included in this study using the available data across three family-based ASD cohorts: the Simons Simplex Collection (SSC)<sup>30</sup>, Simons Foundation Powering Autism Research for Knowledge (SPARK)<sup>31</sup>, and MSSNG<sup>32</sup> (Table S1). We excluded 546 parents and siblings with a diagnosis of ASD from the study.

### Genetic data quality control

We performed quality control of the genetic data for each platform (SNP genotyping: Illumina 1Mv1, 1Mv3, Omni2.5 for SSC; Illumina Infinium Global Screening Array-24 for SPARK, and Whole-Genome Sequencing (WGS): Complete Genomics and Illumina HiSeq (2000 and X Ten) for MSSNG) separately. Standard quality control filtering criteria were applied to the genetic data<sup>33</sup>. We excluded individuals with genotyping rate <95%, excessive heterozygosity ( $\pm 3$  standard deviations from the mean), sample missingness >0.02, mismatched in reported and genetic sex, and families with Mendelian errors >5%. We removed SNPs with a call rate <98%, a minor allele frequency (MAF) <1%, deviated from Hardy-Weinberg Equilibrium ( $P < 1 \times 10^{-6}$ ), and >10% Mendel error rate.

We used the *mds* parameter from the KING software<sup>34</sup> to infer the population substructure of the samples in the study. To avoid confounders due to ancestry, we restricted our analyses to only individuals with a >90% probability of inferred-European ancestry. Imputation of SNP genotypes was performed using the 1000 Genomes Project, phase 3 (1KGP3) reference panel<sup>35</sup> through the Sanger Imputation Server (<https://www.sanger.ac.uk/tool/sanger-imputation-service/>). The VCFs of the imputed SNP and WGS genotypes were subsequently merged using the *merge* command from the “bcftools” program<sup>36</sup>, such that only the loci that were present across all technologies were retained<sup>36</sup>. The merged imputed files were then converted to PLINK files, and were subsequently filtered to remove SNPs with a poor imputation quality ( $\geq 0.3$ ); more than 2 alleles (multiallelic variants); MAF <5%; call rate <98%, and deviated from Hardy-Weinberg Equilibrium ( $P < 5 \times 10^{-7}$ ). Finally, we computed the top 10 ancestry principal components (PCs) for the final European samples using the “mds” parameter from the KING software<sup>34</sup>.

Following sample and variant-level quality control, 24,549 individuals with ASD and 28,898 unaffected family members were retained in the study. A summary and description of the final samples included in the study are detailed in Table S2.

### *Polygenic risk score (PRS) calculation*

PRSs were constructed using the GWAS summary statistics of ASD and 11 other traits with a reported significant genetic correlation with ASD<sup>17</sup> (Table 1). To avoid sample overlap, custom summary statistics for the ASD GWAS summary statistics were generated to exclude the SSC cohort (obtained through the PGC application for secondary analysis proposal – <https://pgc.unc.edu/for-researchers/data-access-committee/data-access-information/>).

We used PRS-CS<sup>37</sup> to infer the posterior effect sizes of SNPs using GWAS summary statistics and the 1KGP3 European linkage disequilibrium (LD) reference panel<sup>35</sup>. Individual-level polygenic risk scores were then computed from the PRS-CS output summary statistics using the PLINK<sup>38</sup> “score” and “no-mean-imputation” parameters.

To account for subtle differences in population structure among the samples with inferred European ancestry, the PRS of each trait was modelled as a function of the top 10 ancestry principal components (PCs) in a linear regression (*lm* function within the R *stats* package):  $PRS_{trait} \sim PC_1 + PC_2 + PC_3 + \dots + PC_{10}$ . Then, the residuals from each regression model were extracted to represent PRS values that accounted for the underlying effects of ancestry among the samples. Finally, the PRSs for each trait were transformed into z-scores.

### *Correlation between ASD and related psychopathologies*

The genetic correlation ( $r_G$ ) between ASD and its related psychopathologies were computed using the command-line tool LD SCore (LDSC, v1.0.1)<sup>39,40</sup> based on the included GWAS summary statistics (Table 1). The correlation between the PRSs (Figure 2) was computed using the *cor* functions from the “stats” base R package.

### *Reducing PRSs of ASD-related traits into representative principal components*

PCA is a dimensionality reduction technique used to compress multidimensional data into representative principal components (PCs) while retaining the most amount of information. We employed this technique to reduce the PRSs of the 11 ASD-correlated traits (Table 1) into variables that captured the variability of all the traits into single dimensions (PCs). We refer to these “reduced” PC variables – representing an individual’s genetic propensity for all 11

ASD-related traits – as “multi-PRS” dimensions. We used this approach to model the effect of all ASD-related traits in a single regression. Given that the PCs from the PCA are orthogonal, we can include all multi-PRS variables as predictors without violating the regression assumption of the absence of multicollinearity.

To do this, we used the *PCA* function to perform the PCA, and the *get\_pca* function to extract the output from the “factoextra” R package<sup>41</sup>. The results from this analysis are detailed in Figure 3. As a sensitivity analysis, we also ran the PCA within six different subgroups (cases with ASD, intrafamilial controls, and the three separate cohorts) (Figures S4, S5, S6). The sensitivity analyses confirm that the PC loadings discussed in the main analyses, whereby we group all samples together, are not driven by any one of these subgroups.

### Statistical analyses

#### *Effect of PRS dimensions on ASD risk in cases versus their unaffected family members*

The effect of the PRSs on ASD risk in cases with ASD versus their unaffected family members was modelled using a generalized linear mixed-effects (GLME) model using the *glmer* function from the “lme4” R package (see Figure 1 for the analysis workflow). This model accounts for the effects of relatedness among ASD individuals and their intrafamilial controls by including the family identifier as a random effect. We ran 13 separate GLME models with ASD diagnosis as the outcome, with the following predictors:

Model 1 (Naïve model using PRS-ASD as the predictor):

$$ASD \sim PRS_{ASD}$$

Models 2–12 (PRSs of ASD-related traits as predictors):

$$ASD \sim PRS_{ADHD}$$

$$ASD \sim PRS_{MDD}$$

$$ASD \sim PRS_{SWB}$$

$$ASD \sim PRS_{tiredness}$$

$$ASD \sim PRS_{SCZ}$$

$$ASD \sim PRS_{chronotype}$$

$$ASD \sim PRS_{IQ}$$

$$ASD \sim PRS_{EA}$$



$$ASD \sim PRS_{CA}$$

$$ASD \sim PRS_{neuroticism}$$

$$ASD \sim PRS_{depressive}$$

Model 13 (multi-PRS dimensions that reflect the reduced PRSs of the 11 ASD-related traits):

$$ASD \sim PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 + PC11$$

For all the models above, sex was included as a covariate and familial relationship was included as a random effect variable.

The goodness-of-fit of each model was evaluated according to its  $R^2$  (detailed model performance metrics in Table S6). We computed the conditional  $R^2$  of each model with the *r.squaredGLMM* function from the “MuMIn” R package<sup>42</sup>. The conditional  $R^2$  represents the variance explained by the entire model, including both fixed and random effects. In this study, we were interested in the proportion of ASD risk that was captured by the PRS predictors in each model. To do this, we first ran a GLME regression that included only the covariates as predictors. Then, we subtracted the  $R^2$  of the covariate model from the conditional  $R^2$  of each PRS regression model (“*r.squared.adj*” in Table S6). For ease of interpretability, the adjusted  $R^2$  was multiplied by 100% to indicate the percentage of ASD variability that each model captured.

#### *Effect of PRS dimensions on developmental phenotypes in cases with ASD*

To assess the effect of PRS-ASD and the multi-PRS dimensions on developmental phenotypes in cases with ASD, we used either a linear (*lm* function) or a logistic (*glm* function) regression model depending on the continuous or binary phenotypic outcome, respectively. The phenotypes were included if there were data available for  $\geq 5\%$  of the samples with ASD. Given the limited data availability, the number of ASD samples with phenotypic data ranged from 1484 (6.04%) to 21,857 (89.03%). Overall, 46 traits (18 continuous and 28 binary) were included in the analyses (Table S3). The continuous outcomes were standardized (Z-score) by the mean within each of the three cohorts. These developmental phenotypes were grouped into nine categories: core ASD features (3); cognitive ability (3); adaptive functioning (3); developmental features (11); co-occurring disorders (17); language ability (4); family history (2); neurological disorder (1), and; health outcome (1). In total, we ran 92 regression models: two models for each developmental phenotype, using either PRS-ASD or all multi-PRS dimensions as the predictors. All models were adjusted for sex and age. The detailed performance metrics of each model are detailed in Table S8.



All  $P$  values were adjusted by the Benjamini–Hochberg false-discovery rate (FDR) correction for multiple comparisons using the *p.adjust* function from the base R package.

## RESULTS

### Genetic and PRS correlation between ASD and ASD-related traits

The results from the genetic and PRS correlation between ASD and its related traits are detailed in Figure 2. We included all 11 traits that had a reported significant genetic correlation with ASD from the latest ASD GWAS<sup>17</sup>. Ten traits (ADHD, MDD, depressive symptoms, tiredness, neuroticism, college attainment, schizophrenia, intelligence, and educational attainment) had a significant genetic ( $r_G$ ) and PRS correlation with ASD ( $P_{FDR} < 0.05$ ). Two traits (chronotype and subjective well-being) had a significant negative genetic ( $r_G$ ) and PRS correlation with ASD. The genetic and PRS correlations in our study were concordant with those reported in Grove *et al.* This is expected, given that PRSs we computed were derived from the GWAS summary statistics used to compute the  $r_G$  in the ASD GWAS<sup>17</sup>. This finding highlights the shared genetic heterogeneity among all traits included in this study. Indeed, every trait that has a significant genetic correlation with ASD is also correlated with the PRS of another ASD-related trait. The high correlation between ASD and the ASD-related traits supports the use of a dimensionality reduction approach to flatten the variability of the PRSs into orthogonal variables.

### PRSs for distinct ASD-related traits can be reduced to representative “multi-PRS” dimensions

Given the correlation between PRS-ASD and its related traits, we assessed whether multi-PRS dimensions, which encompassed the variability of polygenic risk across 11 ASD-related traits, could be used as a proxy to capture the genetic risk for ASD. We used PCA to capture the variability of PRS across all ( $n=11$ ) ASD-related traits into representative “multi-PRS” dimensions (principal components, PCs).

The proportion of PRS variability captured by each PC ranged from 24.7% (PC1) to 3.1% (PC11) (Figure 3). The proportion of polygenic risk for each ASD-related trait that each multi-PRS dimension captures is detailed in Figure 4a. The correlation between PRS-ASD and each of the multi-PRS dimensions is detailed in Table S5.

Multi-PRS PC1 mostly captures the polygenic variability of negative symptom outcomes (i.e.: depressive symptoms, MDD, neuroticism, tiredness), whereas multi-PRS PC2 mostly captures the polygenic variability of higher cognitive ability (i.e.: college and educational attainment,

intelligence). The similarities between the PRSs are represented in Figure 4b, whereby the traits that cluster together have similar contributions to PC1 and PC2.

Overall, these findings suggest that despite their high genetic and PRS overlap (Figure 2), there is a distinct pattern of genetic relatedness between the traits (Figure 4b). Moreover, these findings show that ASD-related PRSs can be transformed into representative – and orthogonal – multi-PRS dimensions that capture a substantial proportion of polygenic variability (Figure 3, Figure 4a). This, in turn, provides support for the potential use of multi-PRS dimensions as a proxy for PRS-ASD to capture the genetic risk for ASD.

#### *Multi-PRS dimensions capture a similar proportion of ASD risk, compared to PRS-ASD itself*

We then modelled the effect of each PRS and the effect of the multi-PRS dimensions (*i.e.*, the PCs representing the variability across polygenic risk for the 11 ASD-related traits) on ASD risk in cases with ASD versus their unaffected family members (Figure 5a). An increase in polygenic risk for ASD, ADHD, MDD, tiredness, neuroticism, depressive symptoms, and schizophrenia significantly increased the risk for ASD diagnosis in probands compared to their unaffected family members. Individuals with ASD had a significant decrease in polygenic risk for subjective well-being and educational attainment in comparison to their unaffected family members. As expected, PRS-ASD had the highest effect on ASD risk (adjusted  $P$  value = 4.45E-52; OR = 1.16 [1.14, 1.18]).

Four of the multi-PRS dimensions (PC1, PC2, PC4, and PC7) significantly increased the risk for ASD ( $P_{\text{FDR}} < 0.05$ ) (Figure 5a). PC4 significantly increased the risk of ASD by 1.10 (adjusted  $P$  value = 2.81E-21; 95% CI of OR = [1.07, 1.11]) and captured 8.9% of variability across all PRSs (Figure 3) – including mostly the polygenic risk for ADHD (64%), MDD (12%), intelligence (7%), and neuroticism (6%) (Figure 4a). PC1 nominally increased the risk of ASD by 1.05 (adjusted  $P$  value = 2.44E-20; 95% CI of OR = [1.04, 1.07]) and captured 24.7% of the variability across all the PRSs (Figure 3), which mostly represented the variability of polygenic risk for depressive symptoms (49%), major depressive disorder (40%), neuroticism (37%), tiredness (36%), educational attainment (27%), and college attainment (25%) (Figure 4a). PC2 also nominally increased the risk for ASD by 1.02 (adjusted  $P$  value = 3.59E-03; 95% CI of OR = [1.01, 1.04]) and captured 16.8% of the variability across all PRSs (Figure 3), which mostly encompassed the variation of PRS for college attainment (46%), educational attainment (45%), and intelligence (23%) (Figure 4a).

In comparison, PRS-ASD increased the risk of ASD by 1.16 (adjusted  $P$  value =  $9.60E-58$ ; 95% CI of OR [1.14, 1.18]). The relative effect size of the significant multi-PRS dimensions, compared to PRS-ASD, ranged from 0.88 to 0.94 (Figure 5b).

We then compared the performance ( $R^2$ ) of the models according to the proportion of genetic risk for ASD that they captured (Figure 4C). The univariate regression model with PRS-ASD as the predictor (naive model) captured the greatest proportion of ASD risk (Conditional  $R^2$  = 0.619%). The multivariate regression model, which included all multi-PRS dimensions as predictors, captured a similar proportion of ASD risk (Conditional  $R^2$  = 0.501%) compared to the naive model. Importantly, given that each multi-PRS dimension is orthogonal, including them all as predictors would not violate any assumption of independence. In contrast, the univariate regression models that used each individual ASD-related PRS as separate predictors captured markedly lower proportions of ASD risk (ranging from 0.002% to 0.37%). The results from the multivariate regression that included PRS-ASD and all ASD-related PRSs as predictors is detailed in Table S7.

#### *Multi-PRS dimensions capture developmental variability among cases with ASD*

We then assessed the effect of PRS-ASD and all multi-PRS dimensions on various developmental phenotypes among the samples with ASD (Figure 6).

Overall, PRS-ASD only had a significant effect on 13 phenotypes. In brief, PRS-ASD influenced one core ASD feature (increased the severity of SCQ); increased non-verbal and full-scale IQ (cognitive ability); decreased the risk for intellectual disability (adaptive functioning); lowered the age of five developmental milestones (developmental features); decreased the risk for separation anxiety (co-occurring disorder) and word delay (language ability); increased the risk for sibling diagnosis of ASD (family history); and reduced the risk for nonfebrile seizures (neurological disorder).

In comparison, the multi-PRS dimensions had a significant effect on a substantial number of developmental phenotypes in ASD (Figure 5, Figure S8). Moreover, the multi-PRS dimensions that conferred significant genetic risk for ASD (Figure 5) had unique effects on core ASD and developmental features (Figure 6). Interestingly, we found a striking “mirroring” pattern in the effects of PC1 and PC2 on the risk for core ASD features, developmental features, and risk for co-occurring disorders. PC1, which mostly captured the PRS variability for negative symptom outcomes (Figure 4a: depressive symptoms, MDD, neuroticism, and tiredness), significantly

increased the severity of core ASD features, the age of developmental milestones, and increased the risk for word delay and familial ASD. In contrast, PC2 – which mostly captured the PRS variability for high cognitive ability (Figure 4a: college and educational attainment, intelligence) – significantly reduced the severity of core ASD features, increased cognitive ability, reduced the age of developmental milestones, and decreased the risk for co-occurring disorders and word delay.

## DISCUSSION

This findings from this study suggest that the paucity of genome-wide significant hits in ASD may be partially driven by the complex and heterogenous genetic architecture of the disorder.

We applied PCA to 11 ASD-related PRSs to construct PRSs that captured the additive, inherited, and non-specific genetic risk for ASD in unique and orthogonal dimensions (referred to as “multi-PRS” dimensions).

First, we found that the multi-PRS dimensions can capture a similar proportion of genetic risk for ASD (0.501%), compared to PRS-ASD itself (0.619%). This suggests that ASD-related PRSs may be used as a proxy for PRS-ASD to capture a similar proportion of genetic risk for the disorder.

Second, we found that the multi-PRS approach can capture unique differences in developmental phenotypes among cases with ASD that PRS-ASD cannot. While multi-PRS dimensions significantly increased the risk for ASD, there was a striking “mirroring” effect between the PC1 and PC2 multi-PRS dimensions on core ASD features, developmental outcomes, and risk for co-occurring disorders. PC1 – which mostly captures the polygenic variability for negative symptom outcomes – increased the risk for ASD and increased the severity of core ASD features, delayed developmental milestones, and risk for co-occurring disorders. Conversely, PC2 – which mostly captures the polygenic variability for higher cognitive ability – also increased the risk for ASD, yet in contrast, reduced the severity of core ASD features, was associated with earlier developmental milestones, and reduced the risk for co-occurring disorders. In other words, our findings highlight the heterogeneity underlying the genetic risk for ASD. While ASD-related PRSs substantially increase the risk for ASD, they have “opposing” effects on core and peripheral ASD phenotypes. This, in turn, could partially explain the low yield of genome-wide significant hits in ASD. We posit that this “mirroring” phenomenon

may dilute the association signal of risk loci in ASD, and could partially explain the diversity observed among individuals with the disorder.

The findings from this study may also reflect the role of a general psychopathology factor (or “P-factor”) on ASD, which captures the variance across psychiatric symptoms in a shared dimension<sup>43–45</sup>. Our results indicate that multi-PRS dimensions can capture a significant, albeit small, proportion of the inherited genetic liability for ASD. Our proposed multi-PRS approach may represent a genetic “P-factor” that captures the overall liability for a diagnosis of a mental disorder<sup>46,47</sup>. Several studies have aimed to identify a genomic P-factor that captures the general liability for psychopathology. Selzam *et al.* proposed a “polygenic P-factor” by applying PCA to PRSs for eight psychopathology traits and investigating the loadings of each trait on the first PC<sup>46</sup>. They found that this genomic P-factor explained 20-43% of the SNP effects across the disorders. Krapohl *et al.* modelled the effect of multiple PRSs on three developmental outcomes and found that combining multiple PRSs in a model yields better phenotype prediction than single-score predictor models<sup>48</sup>.

These findings also align with the Research Domain Criteria (RDoC), which represents a framework for re-classifying mental disorders based on dimensional behaviours and neurobiological measures<sup>49</sup>. Rather than focusing on binary categories, the RDoC examines the underlying pathophysiology of basic traits along a continuum<sup>50</sup>. Indeed, our findings highlight the benefit of studying the genetic architecture of ASD and its developmental phenotypes via related PRSs. While both the multi-PRS PC1 and PC2 dimensions significantly increase the risk for ASD, they have distinct – and opposing – effects on phenotypic characteristics among cases with ASD. In other words, these findings could not have been elucidated through a solely unitary approach (i.e.: the effect of PRS-ASD on ASD diagnosis alone).

We propose three major use cases for this novel multi-PRS approach. First, constructing a PRS from related traits can serve as a proxy for polygenic risk in studies where the GWAS summary statistics overlap with the individual-level genetic data. Second, this approach can be used to generate a PRS for traits that lack sufficient GWAS statistical power. Finally, the multi-PRS dimensions may account for pleiotropy and heterogeneity in orthogonal dimensions. This overcomes the need for conventional multivariate regression models that include all ASD-related PRSs as predictors and are thus subject to overfitting and multicollinearity.

This study has some limitations. First, while this intrafamilial study allows for the comparison of polygenic risk among affected and unaffected family members with ASD, we expect undiagnosed parents and siblings of probands with ASD to have higher rates of ASD traits as compared to the general population<sup>51</sup>. To account for this, we did not include in the analysis those parents and siblings that also had a diagnosis of ASD. However, further comparisons of phenotypes between individuals with ASD and their relatives were not possible due to limited phenotypic data available for relatives. Second, this study aggregates PRSs across numerous genotyping and sequencing technologies. To ensure that the PRSs across modalities were comparable, we only included loci present across all technologies before constructing the PRSs. We also compared PRSs (Figure S1, S2, S3) and applied PCA across various sensitivity analysis groups (Figure S4, S5, S6) and we found that the correlation structure and the main dimensions of variance of polygenic risk were robust across technologies and cohorts.

Using a developmental deconstruction approach, this study contributes to the proposed role of inherited, additive, and non-ASD-specific genetic risk factors on ASD and its related phenotypes. Our proposed multi-PRS approach highlights the pleiotropy<sup>52</sup> between ASD and its related traits, which increase the risk for ASD and uniquely influence developmental phenotypes among cases with ASD. While the use of PRS for clinical risk assessment at the individual level remains ill-advised<sup>53</sup>, this paper highlights the heterogeneous common genetic architecture of ASD that may hinder GWAS loci discovery.

## **STATEMENTS AND DECLARATIONS**

### ***Funding***

ZS has received funding from Canadian Institutes of Health Research Frederick Banting & Charles Best Canada Graduate Scholarship (FRN 181433) and is supported by the Transforming Autism Care Consortium, a thematic network supported by the Fonds de Recherche Québec-Santé. JPR has received funding from Canadian Institutes of Health Research Frederick Banting & Charles Best Canada Graduate Scholarship (FRN 159279). VRB is supported by a Quebec Research Funds - Health (FRQS) residency training scholarship. BC received a grant from the Fondation Bettencourt Schueller (CCA-INSERM Bettencourt).

### ***Competing Interests***

The authors of this study declare they have no competing interests to disclose.

### ***Ethics approval***

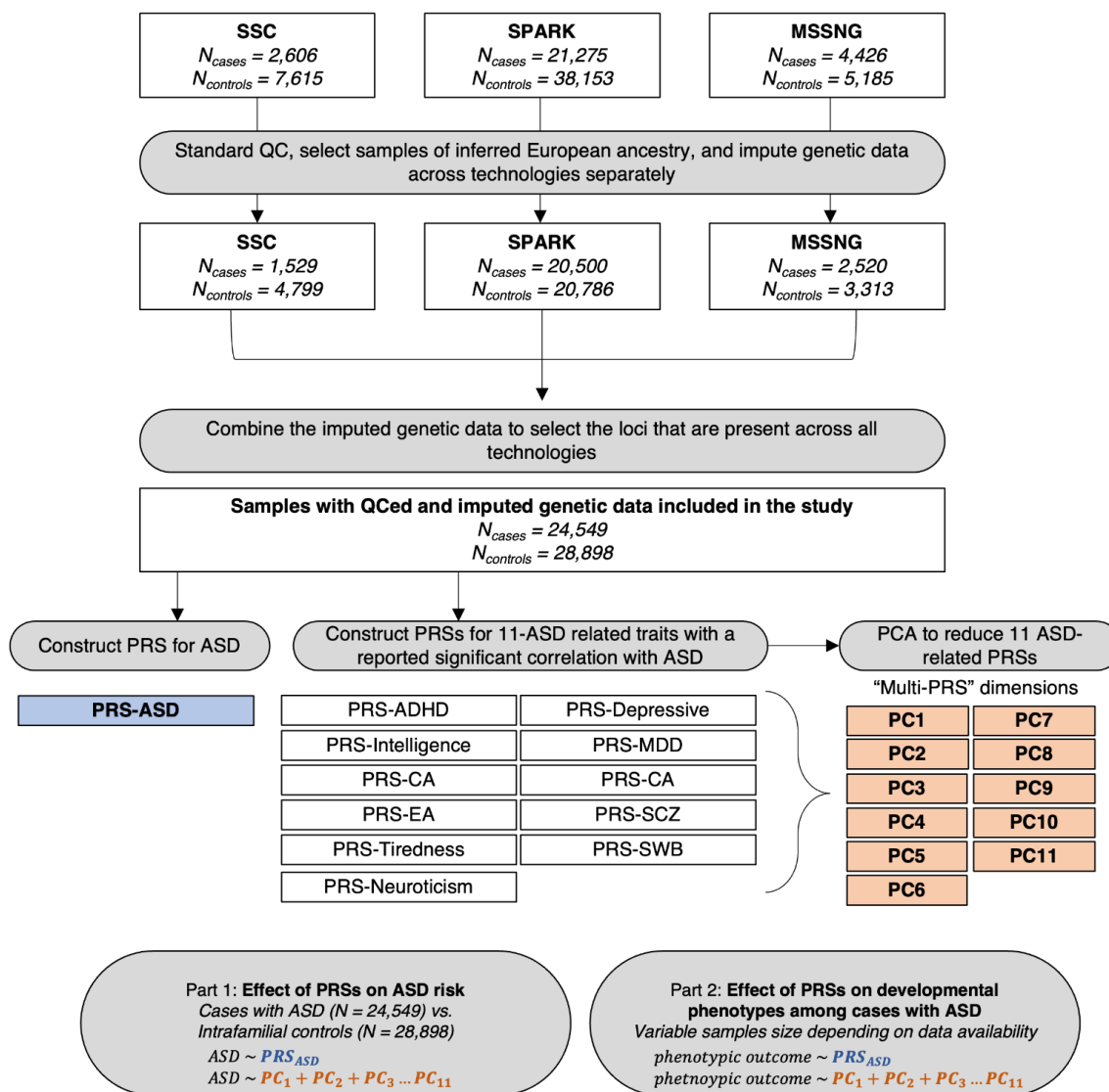
This study was performed in line with the principles of the Declaration of Helsinki.



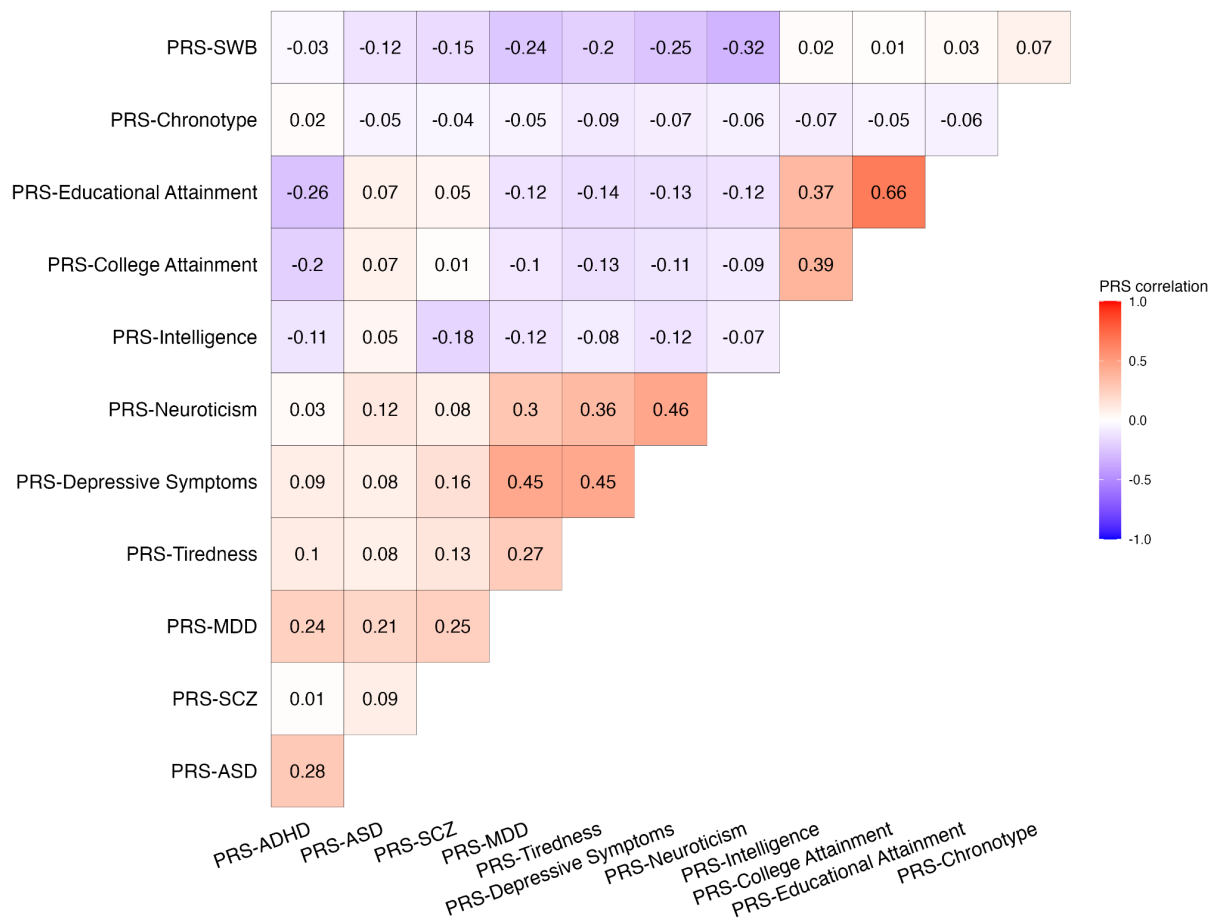
## TABLES AND FIGURES

Trait	GWAS	N	Observed scale heritability (h <sup>2</sup> )	SNP correlation (r <sub>G</sub> ) with ASD	P value for r <sub>G</sub>
ASD*	Grove et al. (2019) <sup>17</sup>	46,350	0.180	-	-
Intelligence	Sniekers et al. (2017) <sup>54</sup>	78,308	0.189	0.219	1.80E-06
College attainment	Davies et al. (2016) <sup>55</sup>	111,114	0.156	0.254	6.00E-11
Educational attainment	Okbay et al. (2016) <sup>56</sup>	405,072	0.124	0.193	3.75E-09
Self-reported tiredness	Deary et al. (2018) <sup>57</sup>	108,976	0.065	0.370	2.51E-11
Neuroticism	Okbay et al. (2016) <sup>58</sup>	170,911	0.090	0.291	1.48E-06
ADHD	Demontis et al. (2019) <sup>59</sup>	20,183	0.637	0.435	6.49E-21
Depressive symptoms	Okbay et al. (2016) <sup>58</sup>	180,866	0.041	0.364	5.73E-10
Major depressive disorder (MDD)	Wray et al. (2018) <sup>60</sup>	173,005 (excluding 23andMe)	0.074	0.441	1.65E-26
Schizophrenia	Ripke et al. (2014) <sup>61</sup>	36,989	0.905	0.225	1.49E-07
Subjective well-being (SWB)	Okbay et al. (2016) <sup>58</sup>	298,420	0.025	-0.481	9.86E-16
Chronotype	Jones et al. (2016) <sup>62</sup>	127,898	0.100	-0.218	1.68E-06

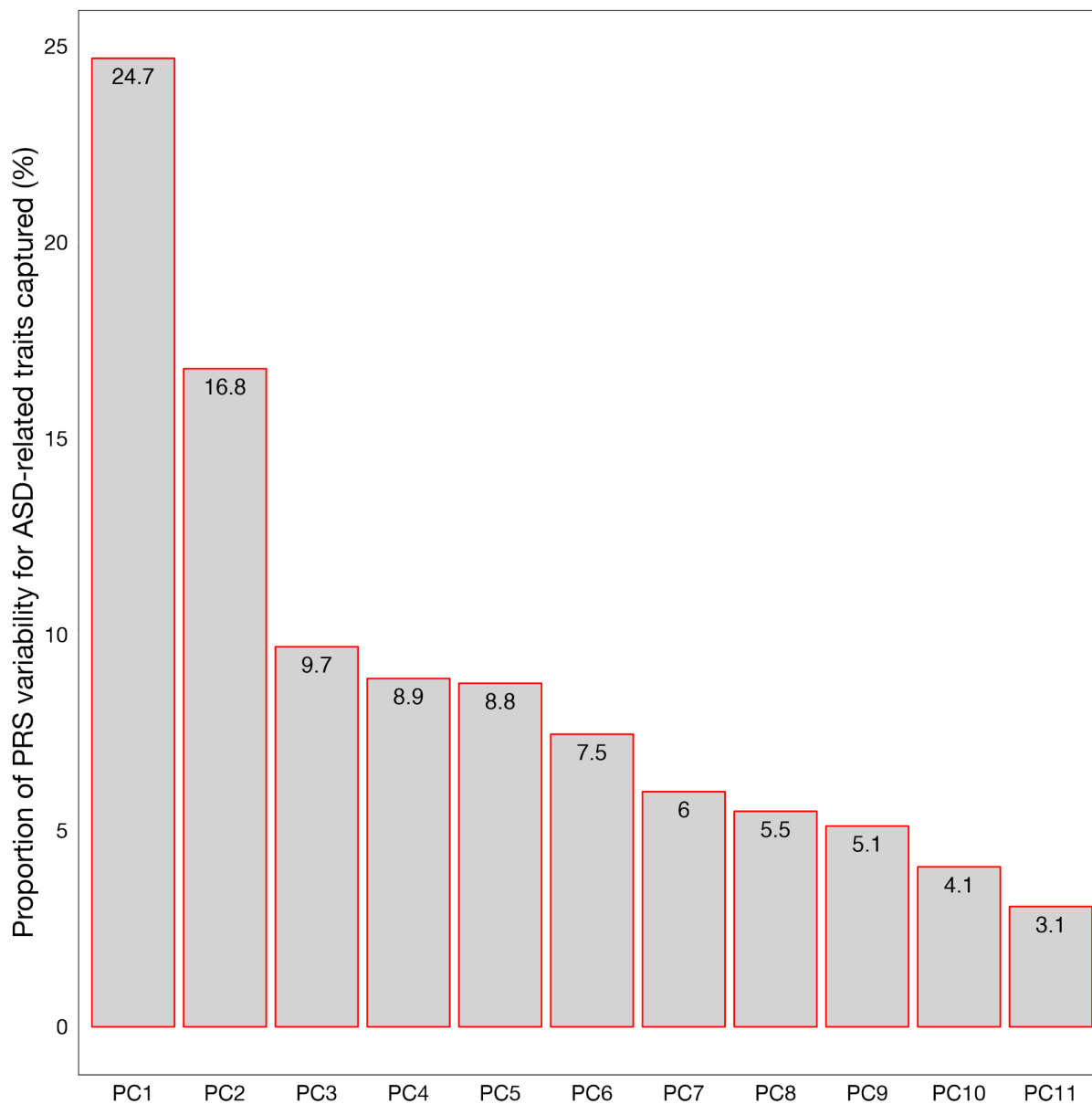
**Table 1. PRSs included in this study and their genetic correlation with ASD.** ASD-related traits were included in the study based on previously-reported significant genetic correlations with ASD in Grove *et al.* (2019)<sup>17</sup>. The GWAS summary statistics that were used to compute PRSs were the same as those used in the Grove *et al.* ASD GWAS. College attainment represents the attainment of a college or university degree. Educational attainment represents the years of education. Chronotype represents the circadian preference for sleep timing. *N* represents the total samples included in the GWAS. The SNP correlation (r<sub>G</sub>) reflects the genetic correlation between ASD and the traits of interest. The observed heritability (h<sup>2</sup>) and r<sub>G</sub> metrics were computed using the LDSC software.



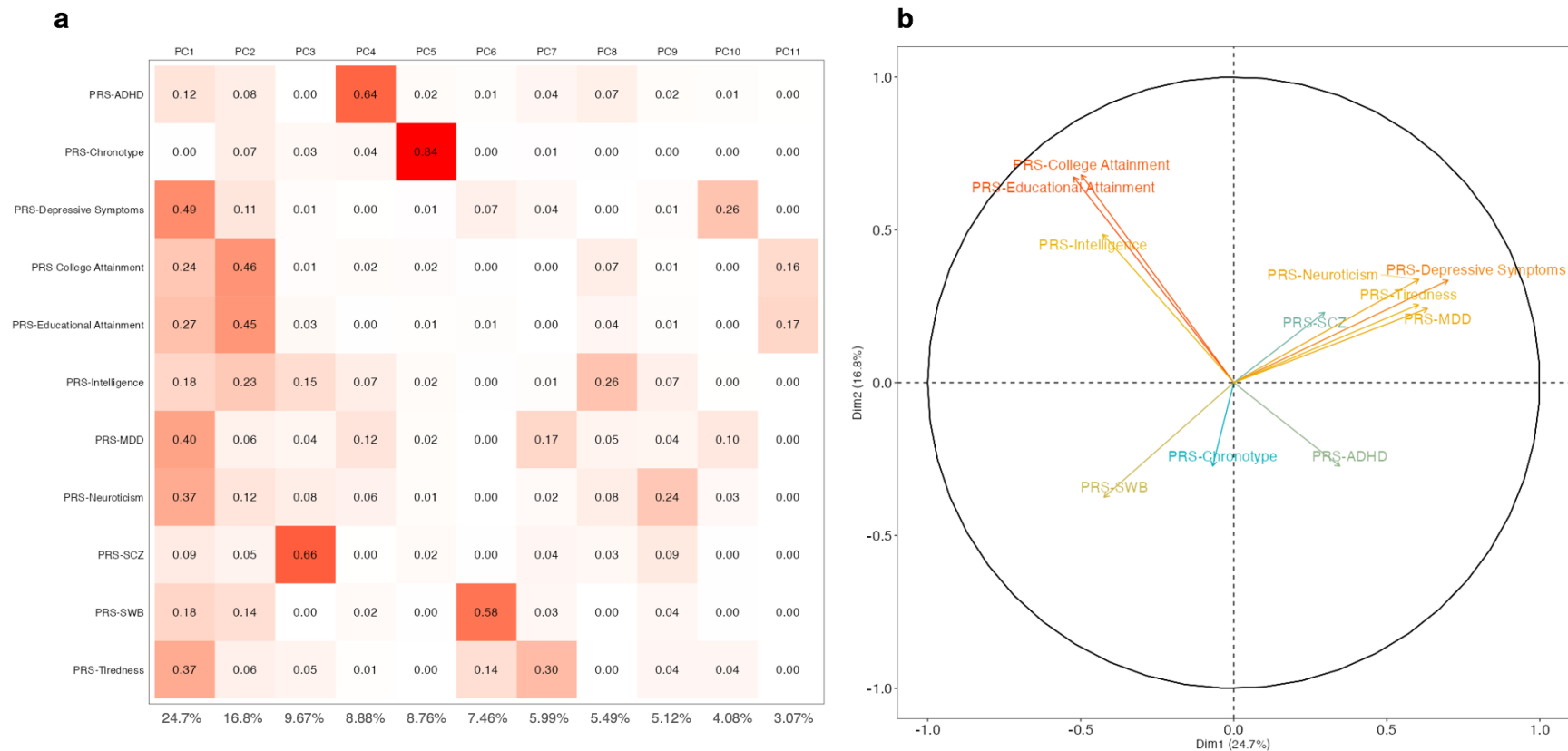
**Figure 1. Analysis flowchart.** The genetic data of three family-based ASD cohorts were included in the study. Following standard genetic and sample-level QC, 24,549 cases with ASD and 28,898 of their unaffected family members (intrafamilial controls) were included in the study. Twelve separate polygenic risk scores (PRSs) were constructed using the GWAS summary statistics of ASD and the 11 traits with a reported significant genetic correlation with ASD. PCA was applied to the 11 ASD-related traits to create “multi-PRS” dimensions (PCs). The first part of the study assessed the effect of the PRS dimensions on ASD risk in cases with ASD versus their unaffected family members. The second part of the study assesses the effect of the PRS dimensions on various developmental phenotypes among cases with ASD.



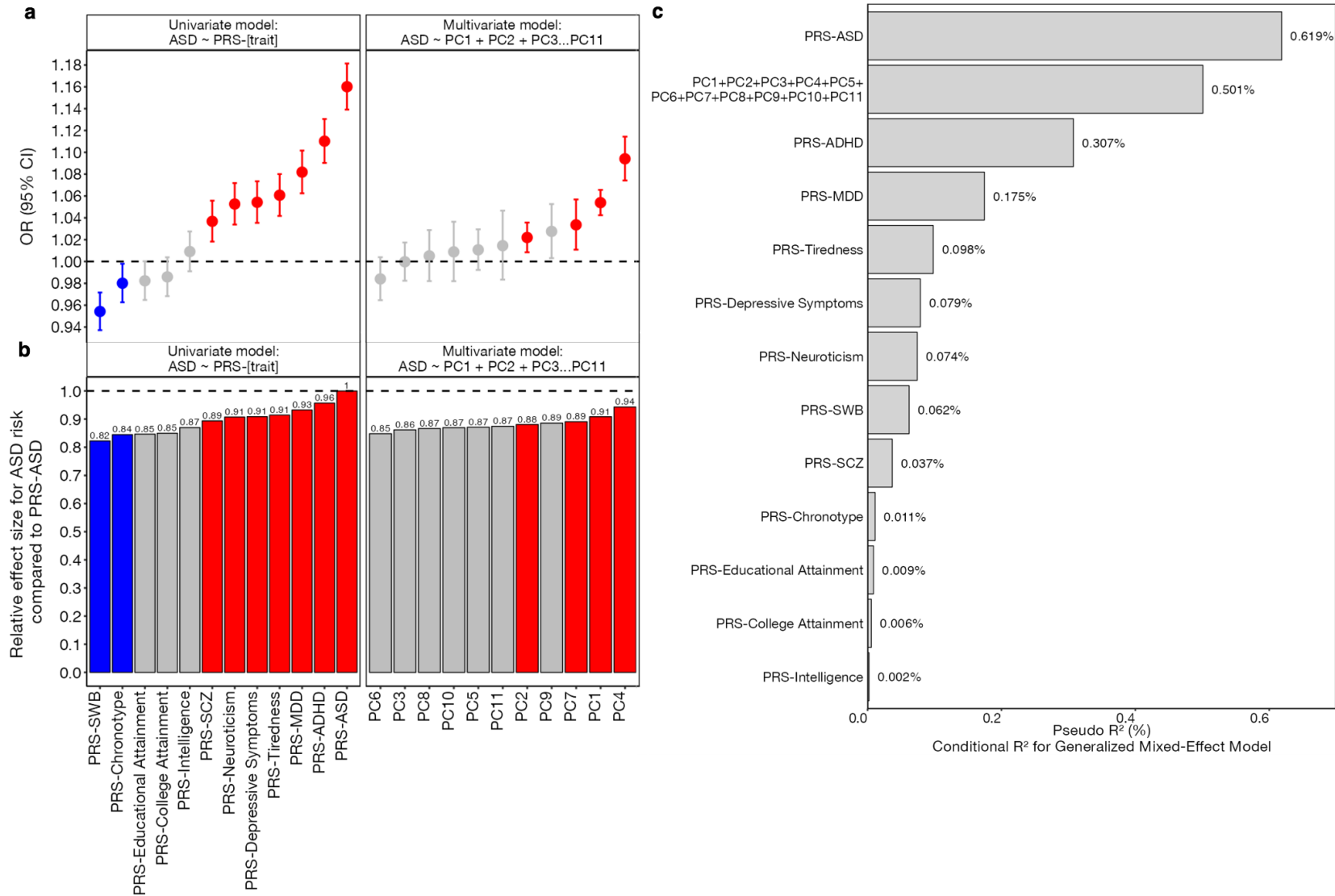
**Figure 2. Genetic relationship between ASD and its related traits.** The correlation between PRS-ASD and the PRS of its 11 related traits.



**Figure 3. The proportion of PRS variability captured by each multi-PRS dimension.** PCA was applied to the PRSs of the 11 ASD-related traits. Each resulting principal component (PC) represents the proportion (%) of PRS variability captured across all traits.



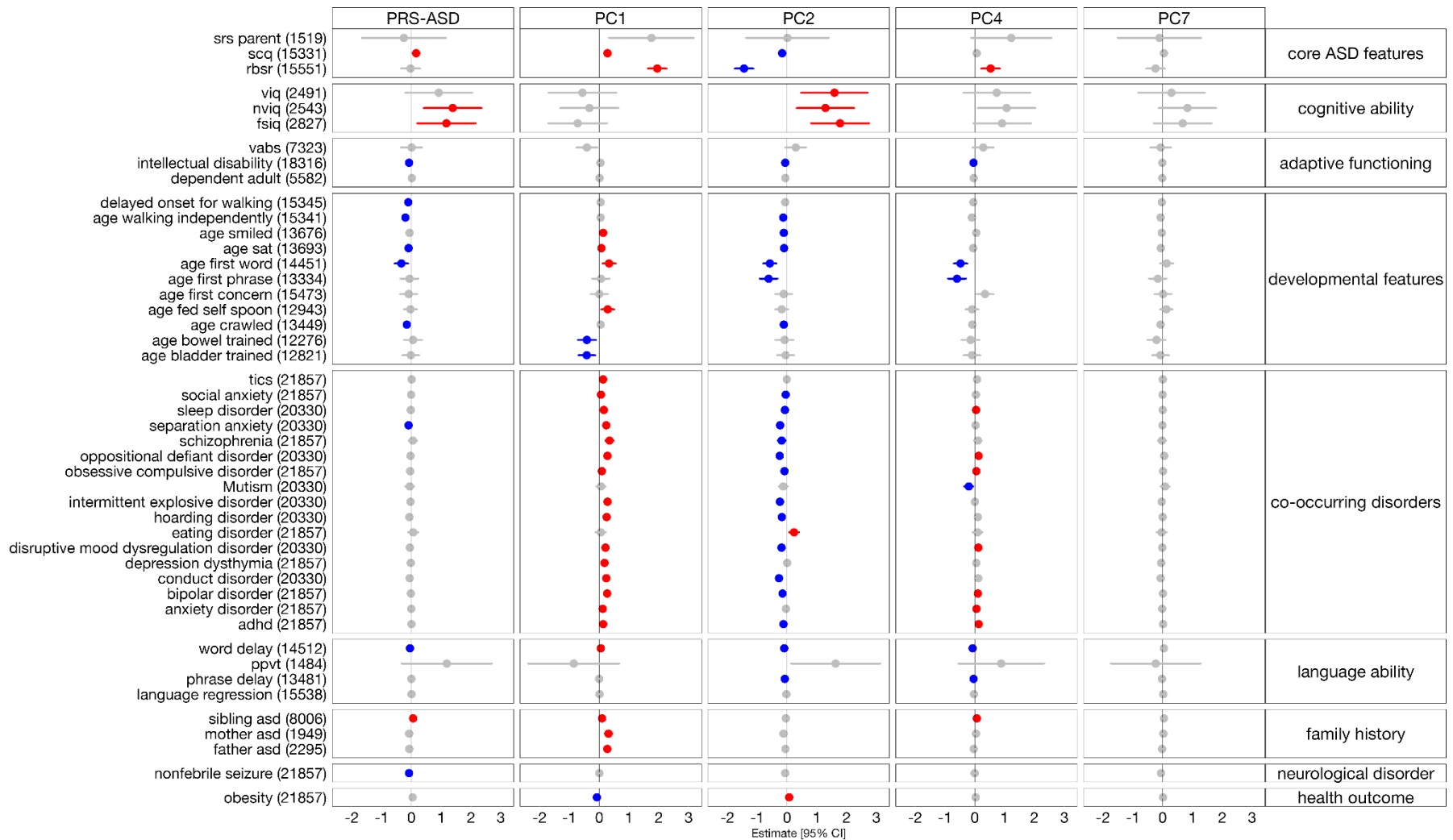
**Figure 4. Reducing the 11 ASD-related PRSs into representative principal components (PCs), herein referred to as “multi-PRS” dimensions. a)** The proportion of PRS variance that is captured in each multi-PRS PC dimension. The percentage of explained variance that each multi-PRS PC dimension captures is shown below each column. **b)** A loading plot representing the influence of each trait on PC1 (Dim.1) and PC2 (Dim.2). The traits that cluster together (similar contributions to PC1/PC2) have relatively similar PRSs.



**Figure 5. Assessing the effect of different PRS modalities on ASD risk.** a) The effect of PRS-ASD, the PRS of ASD-related traits, and multi-PRS dimensions on ASD risk in cases versus their unaffected family members. The effect of PRS-ASD and PRS

of ASD-related traits were modelled separately in univariate generalized linear mixed-effects models. Given that the multi-PRS dimensions were orthogonal, they were included as additive predictors in a single multivariate generalized linear mixed-effects model. Relatedness was included as a random effect in all models. Seven PRSs significantly increased the risk for ASD, and two PRSs significantly decreased the risk for ASD. Four of the multi-PRS dimensions significantly increased the risk for ASD. **b)** The relative odds ratio (OR) of each PRS and multi-PRS dimension on ASD risk, in comparison to PRS-ASD. The relative effect is represented as the proportion of OR relative to the effect of PRS-ASD (OR = 1.16). Red and blue points denote a significant (adjusted  $P$  value <0.05) positive or negative effect on ASD risk, respectively. All  $P$  values were adjusted for FDR correction for multiple comparisons. **c)** The proportion of ASD risk (conditional  $R^2$ ) that each regression model captures. The  $R^2$  represents the percentage of ASD risk that is captured by the PRS dimensions themselves. The multi-PRS model captures a similar proportion of ASD risk compared to the naive PRS-ASD model itself (0.501% versus 0.619%). The effects due to covariates and familial random effects were subtracted from this estimate. This suggests that the multi-PRS dimensions can be used as a proxy for PRS-ASD itself.





**Figure 6. The effect of PRS-ASD and the multi-PRS dimensions on developmental phenotypes in cases with ASD.** The results of 92 regressions are shown (46 outcomes with either PRS-ASD or all 11 multi-PRS dimensions as the predictors). While all 11 multi-PRS dimensions

were included in the model, only those that had a significant effect on ASD risk (Figure 4a) are shown. Red and blue dots represent a significant positive and negative effect of the PRS predictor variable on the developmental phenotype, respectively. All *P* values were adjusted for FDR correction. The brackets denote the number of cases with ASD that were included in the model. The detailed statistical output of each model is detailed in Table S8. All models were adjusted for age and sex. Three and 13 comparisons were considered for the FDR correction of the PRS-ASD and multi-PRS models, respectively. SRS: Social Responsiveness Scale, SCQ: Social Communication Questionnaire, RBSR: Repetitive Behaviour Scale, VIQ: Verbal IQ, NVIQ: Nonverbal IQ, FSIQ: Full-scale IQ, VABS: Vineland Adaptive Behaviour Scale, ADHD: Attention Deficit/Hyperactivity Disorder, PPVT: Peabody Picture Vocabulary Task.

## REFERENCES

1. Bai, D., Yip, B.H.K., Windham, G.C., Sourander, A., Francis, R., Yoffe, R., Glasson, E., Mahjani, B., Suominen, A., Leonard, H., et al. (2019). Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry* 76, 1035–1043. 10.1001/jamapsychiatry.2019.1411.
2. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C.M., and Reichenberg, A. (2014). The familial risk of autism. *JAMA* 311, 1770–1777. 10.1001/jama.2014.4144.
3. Tick, B., Bolton, P., Happé, F., Rutter, M., and Rijdsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J. Child Psychol. Psychiatry* 57, 585–595. 10.1111/jcpp.12499.
4. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional, and chromatin genes disrupted in autism. *Nature* 515, 209–215. 10.1038/nature13772.
5. Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., et al. (2014). Most genetic risk for autism resides with common variation. *Nat. Genet.* 46, 881–885. 10.1038/ng.3039.
6. Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221. 10.1038/nature13908.
7. de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* 22, 345–361. 10.1038/nm.4071.
8. Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., et al. (2018). Analysis of shared heritability in common disorders of the brain. *Science* 360. 10.1126/science.aap8757.
9. Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.-X., et al. (2015). Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* 47, 582–588. 10.1038/ng.3303.
10. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five Years of GWAS Discovery. *Am. J. Hum. Genet.* 90, 7–24. 10.1016/j.ajhg.2011.11.029.
11. Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bölte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* 21, 4781–4792. 10.1093/hmg/dds301.
12. Anney, R.J.L., Ripke, S., Anttila, V., Grove, J., Holmans, P., Huang, H., Klei, L., Lee, P.H., Medland, S.E., Neale, B., et al. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* 8, 21. 10.1186/s13229-017-0137-9.

13. Devlin, B., Melhem, N., and Roeder, K. (2011). Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res.* 1380, 78–84. 10.1016/j.brainres.2010.11.026.
14. Ma, D., Salyakina, D., Jaworski, J.M., Konidari, I., Whitehead, P.L., Andersen, A.N., Hoffman, J.D., Slifer, S.H., Hedges, D.J., Cukier, H.N., et al. (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.* 73, 263–273. 10.1111/j.1469-1809.2009.00523.x.
15. Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M.A., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528–533. 10.1038/nature07999.
16. Weiss, L.A., Arking, D.E., Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly, M.J., and Chakravarti, A. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461, 802–808. 10.1038/nature08490.
17. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444. 10.1038/s41588-019-0344-8.
18. Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–1192. 10.1001/archpsyc.60.12.1187.
19. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am. J. Hum. Genet.* 99, 139–153. 10.1016/j.ajhg.2016.05.013.
20. Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., and Matsui, S. (2018). Sample Size for Successful Genome-Wide Association Study of Major Depressive Disorder. *Front. Genet.* 9.
21. Matoba, N., Liang, D., Sun, H., Aygün, N., McAfee, J.C., Davis, J.E., Raffield, L.M., Qian, H., Piven, J., Li, Y., et al. (2020). Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl. Psychiatry* 10, 1–14. 10.1038/s41398-020-00953-9.
22. Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12, 44. 10.1186/s13073-020-00742-5.
23. Matson, J.L., and Goldin, R.L. (2013). Comorbidity and autism: Trends, topics and future directions. *Res. Autism Spectr. Disord.* 7, 1228–1233. 10.1016/j.rasd.2013.07.003.
24. Romero, M., Aguilar, J.M., Del-Rey-Mejías, Á., Mayoral, F., Rapado, M., Peciña, M., Barbancho, M.A., Ruiz-Veguilla, M., and Lara, J.P. (2016). Psychiatric comorbidities in autism spectrum disorder: A comparative study between DSM-IV-TR and DSM-5 diagnosis. *Int. J. Clin. Health Psychol. IJCHP* 16, 266–275. 10.1016/j.ijchp.2016.03.001.

25. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* *45*, 984–994. [10.1038/ng.2711](https://doi.org/10.1038/ng.2711).
26. Kushki, A., Anagnostou, E., Hammill, C., Duez, P., Brian, J., Iaboni, A., Schachar, R., Crosbie, J., Arnold, P., and Lerch, J.P. (2019). Examining overlap and homogeneity in ASD, ADHD, and OCD: a data-driven, diagnosis-agnostic approach. *Transl. Psychiatry* *9*, 1–11. [10.1038/s41398-019-0631-2](https://doi.org/10.1038/s41398-019-0631-2).
27. Mous, S.E., Jiang, A., Agrawal, A., and Constantino, J.N. (2017). Attention and motor deficits index non-specific background liabilities that predict autism recurrence in siblings. *J. Neurodev. Disord.* *9*, 32. [10.1186/s11689-017-9212-y](https://doi.org/10.1186/s11689-017-9212-y).
28. Constantino, J.N. (2018). Deconstructing autism: from unitary syndrome to contributory developmental endophenotypes. *Int. Rev. Psychiatry* *30*, 18–24. [10.1080/09540261.2018.1433133](https://doi.org/10.1080/09540261.2018.1433133).
29. Warriar, V., Zhang, X., Reed, P., Havdahl, A., Moore, T.M., Cliquet, F., Leblond, C.S., Rolland, T., Rosengren, A., Rowitch, D.H., et al. (2022). Genetic correlates of phenotypic heterogeneity in autism. *Nat. Genet.* *54*, 1293–1304. [10.1038/s41588-022-01072-5](https://doi.org/10.1038/s41588-022-01072-5).
30. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron* *68*, 192–195. [10.1016/j.neuron.2010.10.006](https://doi.org/10.1016/j.neuron.2010.10.006).
31. Feliciano, P., Daniels, A.M., Green Snyder, L., Beaumont, A., Camba, A., Esler, A., Gulsrud, A.G., Mason, A., Gutierrez, A., Nicholson, A., et al. (2018). SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* *97*, 488–493. [10.1016/j.neuron.2018.01.015](https://doi.org/10.1016/j.neuron.2018.01.015).
32. Yuen, R.K., Merico, D., Bookman, M., Howe, J.L., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* *20*, 602–611. [10.1038/nn.4524](https://doi.org/10.1038/nn.4524).
33. Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E.M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* *27*. [10.1002/mpr.1608](https://doi.org/10.1002/mpr.1608).
34. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinforma. Oxf. Engl.* *26*, 2867–2873. [10.1093/bioinformatics/btq559](https://doi.org/10.1093/bioinformatics/btq559).
35. McVean, G.A., Altshuler (Co-Chair), D.M., Durbin (Co-Chair), R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* *491*, 56–65. [10.1038/nature11632](https://doi.org/10.1038/nature11632).
36. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* *27*, 2987–2993. [10.1093/bioinformatics/btr509](https://doi.org/10.1093/bioinformatics/btr509).

37. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* *10*, 1776. [10.1038/s41467-019-09718-5](https://doi.org/10.1038/s41467-019-09718-5).
38. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559–575. [10.1086/519795](https://doi.org/10.1086/519795).
39. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295. [10.1038/ng.3211](https://doi.org/10.1038/ng.3211).
40. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* *47*, 1236–1241. [10.1038/ng.3406](https://doi.org/10.1038/ng.3406).
41. Kassambara, A., and Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*.
42. Bartoń, K. (2022). *MuMIn: Multi-Model Inference*.
43. Caspi, A., and Moffitt, T.E. (2018). All for One and One for All: Mental Disorders in One Dimension. *Am. J. Psychiatry* *175*, 831–844. [10.1176/appi.ajp.2018.17121383](https://doi.org/10.1176/appi.ajp.2018.17121383).
44. Caspi, A., Houts, R.M., Belsky, D.W., Goldman-Mellor, S.J., Harrington, H., Israel, S., Meier, M.H., Ramrakha, S., Shalev, I., Poulton, R., et al. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin. Psychol. Sci. J. Assoc. Psychol. Sci.* *2*, 119–137. [10.1177/2167702613497473](https://doi.org/10.1177/2167702613497473).
45. Lahey, B.B., Krueger, R.F., Rathouz, P.J., Waldman, I.D., and Zald, D.H. (2017). A Hierarchical Causal Taxonomy of Psychopathology across the Life Span. *Psychol. Bull.* *143*, 142–186. [10.1037/bul0000069](https://doi.org/10.1037/bul0000069).
46. Selzam, S., Coleman, J.R.I., Caspi, A., Moffitt, T.E., and Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Transl. Psychiatry* *8*, 1–9. [10.1038/s41398-018-0217-4](https://doi.org/10.1038/s41398-018-0217-4).
47. Sprooten, E., Franke, B., and Greven, C.U. (2021). The P-factor and its genomic and neural equivalents: an integrated perspective. *Mol. Psychiatry*, 1–11. [10.1038/s41380-021-01031-2](https://doi.org/10.1038/s41380-021-01031-2).
48. Krapohl, E., Patel, H., Newhouse, S., Curtis, C.J., von Stumm, S., Dale, P.S., Zabaneh, D., Breen, G., O'Reilly, P.F., and Plomin, R. (2018). Multi-polygenic score approach to trait prediction. *Mol. Psychiatry* *23*, 1368–1374. [10.1038/mp.2017.163](https://doi.org/10.1038/mp.2017.163).
49. Insel, T.R., and Wang, P.S. (2010). Rethinking Mental Illness. *JAMA* *303*, 1970–1971. [10.1001/jama.2010.555](https://doi.org/10.1001/jama.2010.555).

50. Cuthbert, B.N. (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry* 13, 28–35. 10.1002/wps.20087.
51. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J., Okbay, A., Bybjerg-Grauholm, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* 49, 978. 10.1038/ng.3863.
52. Visscher, P.M., and Yang, J. (2016). A plethora of pleiotropy across complex traits. *Nat. Genet.* 48, 707–708. 10.1038/ng.3604.
53. Ding, Y., Hou, K., Burch, K.S., Lapinska, S., Privé, F., Vilhjálmsson, B., Sankararaman, S., and Pasaniuc, B. (2022). Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nat. Genet.* 54, 30–39. 10.1038/s41588-021-00961-5.
54. Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R., Krapohl, E., Taskesen, E., Hammerschlag, A.R., Okbay, A., Zabaneh, D., et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* 49, 1107–1112. 10.1038/ng.3869.
55. Davies, G., Marioni, R.E., Liewald, D.C., Hill, W.D., Hagenaars, S.P., Harris, S.E., Ritchie, S.J., Luciano, M., Fawns-Ritchie, C., Lyall, D., et al. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol. Psychiatry* 21, 758–767. 10.1038/mp.2016.45.
56. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533, 539–542. 10.1038/nature17671.
57. Deary, V., Hagenaars, S.P., Harris, S.E., Hill, W.D., Davies, G., Liewald, D.C.M., International Consortium for Blood Pressure GWAS, CHARGE Consortium Aging and Longevity Group, CHARGE Consortium Inflammation Group, McIntosh, A.M., et al. (2018). Genetic contributions to self-reported tiredness. *Mol. Psychiatry* 23, 609–620. 10.1038/mp.2017.5.
58. Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* 48, 624–633. 10.1038/ng.3552.
59. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75. 10.1038/s41588-018-0269-7.
60. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50, 668–681. 10.1038/s41588-018-0090-3.



61. Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. [10.1038/nature13595](https://doi.org/10.1038/nature13595).
62. Jones, S.E., Tyrrell, J., Wood, A.R., Beaumont, R.N., Ruth, K.S., Tuke, M.A., Yaghootkar, H., Hu, Y., Teder-Laving, M., Hayward, C., et al. (2016). Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. *PLoS Genet.* 12, e1006125. [10.1371/journal.pgen.1006125](https://doi.org/10.1371/journal.pgen.1006125).