RESEARCH   **Open Access**

Characterizing genetic pathways unique to autism spectrum disorder at multiple levels of biological analysis

Lukas S. Schaffer[1,2*], Sophie Breunig[1,2], Jeremy M. Lawrence[1,2], Isabelle F. Foote[1] and Andrew D. Grotzinger[1,2]

Abstract

Background Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by atypical patterns of social functioning and repetitive/restricted behaviors. ASD commonly co-occurs with ADHD and, despite their clinical distinctiveness, the two share considerable genetic overlap. Given their shared genetic liability, it is unclear which genetic pathways increase the likelihood of ASD independently of ADHD.

Methods We applied Genomic Structural Equation Modeling (SEM) to GWAS summary statistics for ASD and childhood-diagnosed ADHD, decomposing the genetic variance for ASD into that which is unique to ASD (*uASD*) and that which is shared with ADHD. We computed genetic correlations between *uASD* and 83 external traits to estimate genetic overlap between *uASD* and other clinically relevant phenotypes. We went on to apply Stratified Genomic SEM to identify classes of genes enriched for *uASD*. Finally, we implemented Transcriptome-Wide SEM (T-SEM) to explore patterns of gene-expression associated with *uASD*.

Results We observed positive genetic correlations between *uASD* and several external traits, most notably those relating to cognitive/educational outcomes and internalizing psychiatric traits. Stratified Genomic SEM showed that heritability for *uASD* was significantly enriched in genes involved in evolutionarily conserved processes, as well as for a histone mark in the germinal matrix. T-SEM revealed 83 unique genes with expression associated with *uASD*, 34 of which were novel with respect to univariate analyses. These genes were overrepresented in skin-related pathologies.

Limitations Our study was limited by summary statistics derived exclusively from individuals of European ancestry. Additionally, using data based on a general ASD diagnosis limits our ability to understand genetic factors contributing to the pronounced clinical heterogeneity in ASD.

Conclusions Our findings delineate the unique genetic underpinnings of ASD that are independent of ADHD at the genome-wide, functional, and gene expression level of analysis. In addition, we identify novel associations previously masked by their diametric effects on ADHD. Collectively, these results provide insight into the processes that make ASD biologically unique.

Keywords Autism spectrum disorder, Attention-deficit/hyperactivity disorder, Genomic SEM, Multivariate genomics, Comorbidity, Neurodevelopmental disorders, Psychiatric genetics

*Correspondence:
Lukas S. Schaffer
Lukas.Schaffer@colorado.edu

[1]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA
[2]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

Schaffer *et al. Molecular Autism*   (2024) 15:46

Page 2 of 11

Introduction

Autism spectrum disorder (ASD) is a heterogenous neurodevelopmental disorder which occurs in >1% of the population [1, 2]. While the phenotypic presentation of ASD is highly variable, it is characterized by two core symptom domains: (i) impairments in social communication and interaction as well as (ii) repetitive, restricted patterns of behavior or interests [3]. Although the etiology of ASD involves an array of contributing factors, extant literature has demonstrated a strong genetic component with heritability estimates from twin and family studies ranging between 64 and 91% [4–6]. A substantial proportion of this heritability (∼12%) is attributable to common genetic variation as evidenced by recent genome-wide association studies (GWASs) of ASD [2, 7].

ASD often presents alongside other neuropsychiatric conditions; the most frequent comorbidity is attention-deficit/hyperactivity disorder (ADHD) [8], a childhood-onset disorder characterized by symptoms of either inattention, hyperactivity, or some combination of the two [3]. Indeed, conservative estimates suggest that one in every three children with ASD will also meet diagnostic criteria for ADHD [9]. Extending beyond complete comorbidity, autistic children are likely to present with features associated with ADHD, and vice versa [10]. Converging evidence indicates that a common genetic architecture partially underlies liability for both disorders and their cardinal symptoms [10, 11]. For example, observations from family-based studies find that ASD and ADHD tend to co-aggregate within families [12], and this co-aggregation is due, in part, to shared additive genetic influences [13, 14]. These findings are corroborated by molecular and statistical genetic studies which have estimated moderate genetic correlations between ASD and ADHD [2, 11, 15], indicating a shared genetic architecture.

These findings illustrate a broad challenge of parsing disorder-specific biological pathways when two phenotypes are genetically and phenotypically correlated. These difficulties necessitate the need for multivariate genomic analyses capable of isolating the genetic variance that is unique to a specific trait. Here, we approach these challenges by leveraging Genomic Structural Equation Modeling (SEM) and its extensions to separate out the genetic variance unique to ASD from that which is shared with ADHD [16–18]. We model GWAS summary statistics for ASD and childhood-diagnosed ADHD using a Cholesky decomposition to derive a unique ASD (henceforth, *uASD*) latent factor, reflecting the residual genetic variance for ASD after removing genetic overlap with ADHD. We then apply downstream analyses to interrogate the genetic architecture of *uASD* at the genome-wide, functional, and gene expression levels of analysis. Collectively, these analyses delineate the biological mechanisms that contribute specifically to the etiology of ASD as opposed to those that may confer a broader spectrum of shared neurodevelopmental susceptibility.

Method

Summary statistics

Summary statistics for ASD were used from the most recent GWAS meta-analysis [2]. Briefly, the original GWAS included 13 076 cases and 22 664 controls from the Danish population-based cohort iPSYCH and 5 305 cases and 5 305 pseudo-controls (i.e., non-transmitted parental alleles) from family-based trio samples from the Psychiatric Genomics Consortium (PGC). Together, the meta-analysis totaled 18 381 ASD cases and 27 969 controls/pseudo-controls. GWAS summary statistics for childhood-diagnosed ADHD were utilized from the GWAS conducted by Rajagopal et al. (2022), which stratified ADHD cases by age of diagnosis. Here, we specifically utilize summary statistics for ADHD diagnosed in childhood due to its higher genetic overlap with ASD compared to persistent or adulthood-diagnosed ADHD [15, 19]. The childhood-diagnosed GWAS included 14 878 cases and 38 303 controls from the iPSYCH cohort. Of note, all subsequent analyses based on these summary statistics capture only common genetic variation indexed by GWAS. The term "genetic variance" is henceforth used to refer to the variance that can be explained by this common genetic variation.

To evaluate genetic overlap with other relevant phenotypes, we leveraged publicly available European-ancestry summary statistics for 85 external traits spanning domains of cognition, psychopathology, health/lifestyle behaviors, interpersonal relations, and physical activity. We also included several traits and diseases related to immune and gastrointestinal dysfunction given their clinical associations with ASD. We used a SNP-based heritability (h^2_{SNP}) Z-statistic cutoff of 4, as recommended by the original linkage disequilibrium score regression (LDSC) developers [20], to limit our pool of external traits to those with interpretable genetic covariance. Based on this cutoff, 83 of our original 85 traits were carried forward for analysis. A comprehensive list of included external traits and relevant characteristics is reported in Supplementary Table 1.

Genomic structural equation modeling

Prior to analysis, all GWAS summary statistics underwent an identical set of QC filters using the *munge* function in the *GenomicSEM* R package. These filters included restricting analyses to HapMap3 SNPs and removing SNPs with a minor allele frequency (MAF) < 1% and imputation score (INFO) < 0.9 (when available). Once processed, these GWAS summary statistics were used as input for multivariable LDSC using the *ldsc* function

within *GenomicSEM*, which produces the genetic covariance and sampling covariance matrices across included traits. The genetic covariance matrix includes the h^2_{SNP} on the diagonal and the genetic covariances on the off-diagonal. The sampling covariance matrix contains squared standard errors (sampling variances) on the diagonal and the sampling covariances (sampling dependencies) on the off-diagonal that will arise in the presence of participant sample overlap. The sampling covariance matrix is estimated directly from the data using a block jackknife resampling procedure and allows for GWAS with varying degrees of power and sample overlap to be included in the same statistical model.

For binary traits, estimates were converted to the liability scale using the population prevalence and the sum of effective sample size across contributing cohorts [21]. For ASD, the effective sample size was estimated directly from the data. This is because the ASD GWAS used pseudocontrol subjects, which reduces power to detect GWAS associations, such that using the observed sum of effective sample size would produce downwardly biased estimates of heritability [22]. LDSC requires that estimates are produced within a single ancestry group as the LD weights used to estimate the regression model will vary across ancestral populations. Due to limited availability of data in other ancestral groups, GWAS statistics were limited to participants of European ancestry, and the LDSC model was estimated using the 1000 Genomes Phase 3 European LD scores. These scores excluded the major histocompatibility complex (MHC) due to complex LD structures in this region that can bias estimates.

The output from LDSC was used as input to all subsequent analyses in Genomic SEM. We began by fitting a Cholesky decomposition model to our observed variables based on GWAS summary statistics for ASD and childhood-diagnosed ADHD (henceforth, simply referred to as ADHD). Both ASD and ADHD were regressed onto a latent factor, *cADHD*, which represents the genetic variance of ADHD as well as the proportion of genetic variance for ASD that is shared with ADHD. ASD was additionally regressed onto *uASD*, representing the residual genetic variance that is unique to ASD after accounting for genetic covariance with ADHD. By construction, *uASD* and *cADHD* were orthogonal ($r_g = 0$). The genetic residual variances of ASD and ADHD were fixed to 0 so that all common genetic variance in the disorders was explained by the latent factors. Given that the Cholesky decomposition is a fully saturated model that perfectly recapitulates the observed data, we do not report fit indices as they would indicate a perfect model fit by definition. At the genome-wide level, this model was expanded to compute the genetic correlations between *uASD* and each of our pre-selected 83 external traits (see Supplementary Table 2). In interpreting the statistical significance of each genetic correlation, we apply a strict Bonferroni-adjusted significance threshold ($p < 5.9E-4$).

Stratified genomic SEM

We applied Stratified Genomic SEM to identify enrichment for functional annotations (i.e., categories of genes) for *uASD*. We first ran multivariable Stratified LDSC (S-LDSC) to obtain genetic covariance estimates stratified across different functional annotations. These stratified estimates reflect the proportion of variance for ASD and ADHD that are captured by each annotation. To avoid confusion with similar analyses in the genetic field, we highlight that this analysis is not stratified by chromosome. We originally included a total of 168 annotations for analysis. However, 12 of these annotations were excluded due to the model failing to converge ($n=10$) or negative heritability estimates ($n=2$). The remaining 156 were examined to ensure that none required a smoothing of the covariance matrix resulting in a *z*-statistic difference>1.96 (as recommended by Grotzinger et al., 2022) before moving forward. Our final analysis included 49 annotations from the 1000 Genomes Baseline LD Version 2.2 [23], as well as neuronal and brain tissue annotations from DEPICT [24], gnomAD [25], GTEx v8 [26], and the Roadmap Epigenomics Project [27] (see Supplementary Table 3 for full list of annotations). Using the *enrich* function within *GenomicSEM*, we then estimated enrichment for *uASD* within each annotation. The *enrich* function works by first estimating the model using the genome-wide estimates (i.e., from an annotation including all SNPs). The function then fixes the loadings from the genome-wide model, and finally estimates the variances of the factors within each annotation. Fixing the loadings from the genome-wide model serves to produce within-annotation factor variances that are on the same scale, which allows for calculating the enrichment 'ratio-of-ratios.' For the current analyses, the numerator of this ratio reflects the proportion of *uASD* genetic variance explained by an annotation (i.e., the within-annotation genetic variance for *uASD* divided by the total *uASD* genetic variance). The denominator of the ratio reflects the proportional size of the annotation (i.e., the number of SNPs in the annotation divided by the total number of SNPs analyzed across all annotations). The null for this enrichment ratio-of-ratios is 1, where values above 1 index functional annotations that account for a greater proportion of genetic variance in *uASD* than would be expected based solely on the proportional size of that annotation. Given the non-independent nature of functional annotations, we applied the FDR correction for multiple comparisons to the accompanying *p*-values using the *p.adjust* R package.

Transcriptome-wide SEM (T-SEM)

Transcriptome-wide SEM (T-SEM) was applied to identify patterns of gene expression associated with *uASD*. First, FUSION [28] was used to perform univariate transcriptome-wide association studies (TWASs) on both ASD and ADHD. We utilized functional weights for 13 brain tissue types from the Genotype Tissue Expression Project (GTEx v8 [29]), two dorsolateral prefrontal cortex weights from the CommonMind Consortium (CMC [30]), and one set of weights for the prefrontal cortex from PsychEncode [31]. This resulted in 16 total functional weights from which we derived 73 412 genes with imputed expression data across different brain regions and tissues. These univariate FUSION outputs were then input into the *read_fusion* function in Genomic SEM.

The gene expression estimates were then added to the LDSC covariance matrix for ASD and ADHD, and the *userGWAS* function was used to estimate the effect of gene expression on both *uASD* and *cADHD*. Given the scope of the analyses, we specifically focus on the relationship between gene expression and *uASD*. Finally, to identify gene sets with significant enrichment in *uASD*, we used the *WebGestalt* package to conduct an over-representation analysis (ORA [32]), on genes identified by T-SEM. An FDR correction was used in interpreting significance of T-SEM and ORA results. We also carried forth a drug repurposing analysis on these genes using methods outlined by Grotzinger et al. 2023. As these analyses did not produce any findings, no results are reported below.

Results

Genome-wide results reveal genetic correlates of uASD

ASD and ADHD were moderately genetically correlated ($r_g = 0.45$, $SE = 0.05$). The Cholesky decomposition revealed that *cADHD* explained ~20% of the genetic variance in ASD and *uASD* explained the remaining ASD genetic variance (see Fig. 1 for visual representation of the structural equation model and partitioning of genetic variance). The remaining analyses sought to clarify what genetically differentiates ASD from ADHD by examining associations with the *uASD* factor at multiple levels of analysis. We began by genetically correlating *uASD* with 83 external traits in order to evaluate the extent of genetic overlap with other phenotypes relevant to mental and physical health. By virtue of our statistical definition of *uASD*, these correlations represented associations extending above and beyond those with *cADHD*. A full list of external traits and relevant outputs is available in Supplementary Table 2. Our analyses revealed that *uASD* was significantly correlated with 20/83 external traits. These 20 correlations tended to span four primary phenotypic dimensions – cognition, psychopathology, physical movement, and interpersonal relations – that we review below (and in Fig. 2, panels **A**, **B**, **C**, and **D**).
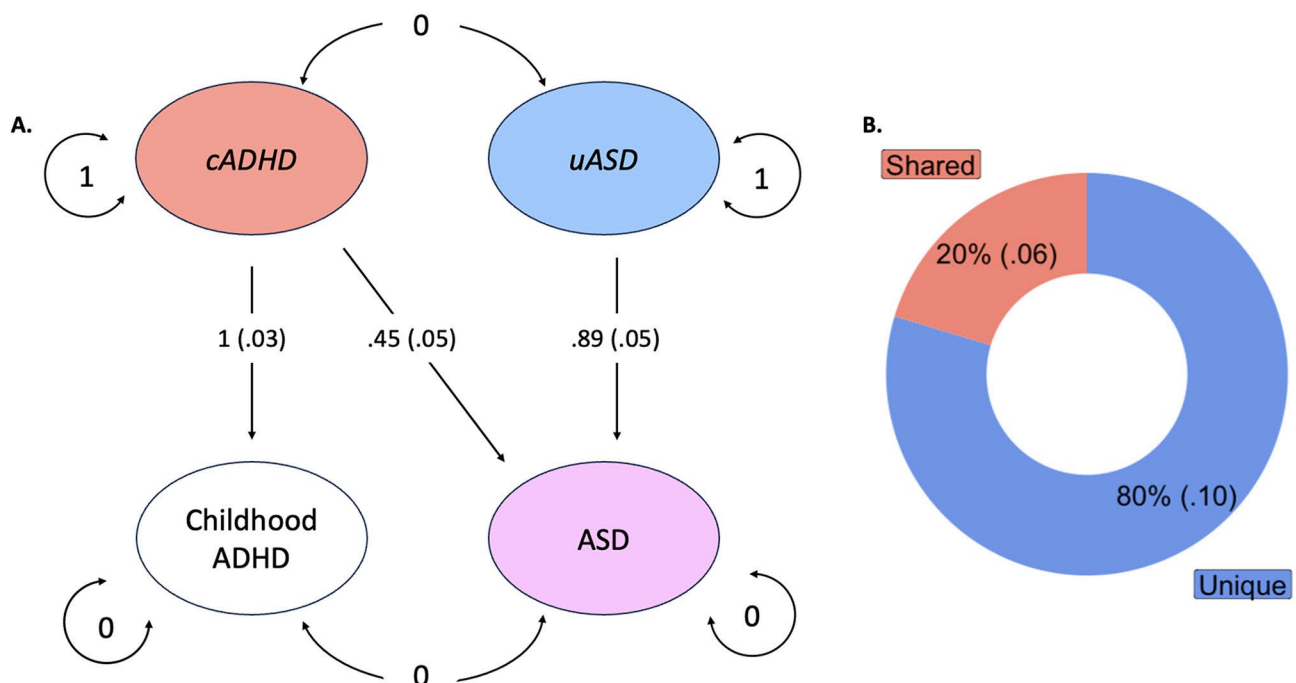

Fig. 1 Genomic structural equation modeling to decompose ASD genetic variance. (**A**) Cholesky decomposition model producing *uASD*, a latent variable encompassing the genetic variance unique to ASD independent of ADHD, and *cADHD*, which captures the residual genetic variance of ASD (i.e., variance shared between ASD and ADHD) and the genetic variance of ADHD. (**B**) Donut plot showing the proportion of residual genetic variance unique to ASD (blue) and shared with childhood ADHD (red). Standard errors for the variance estimates are provided in parentheses
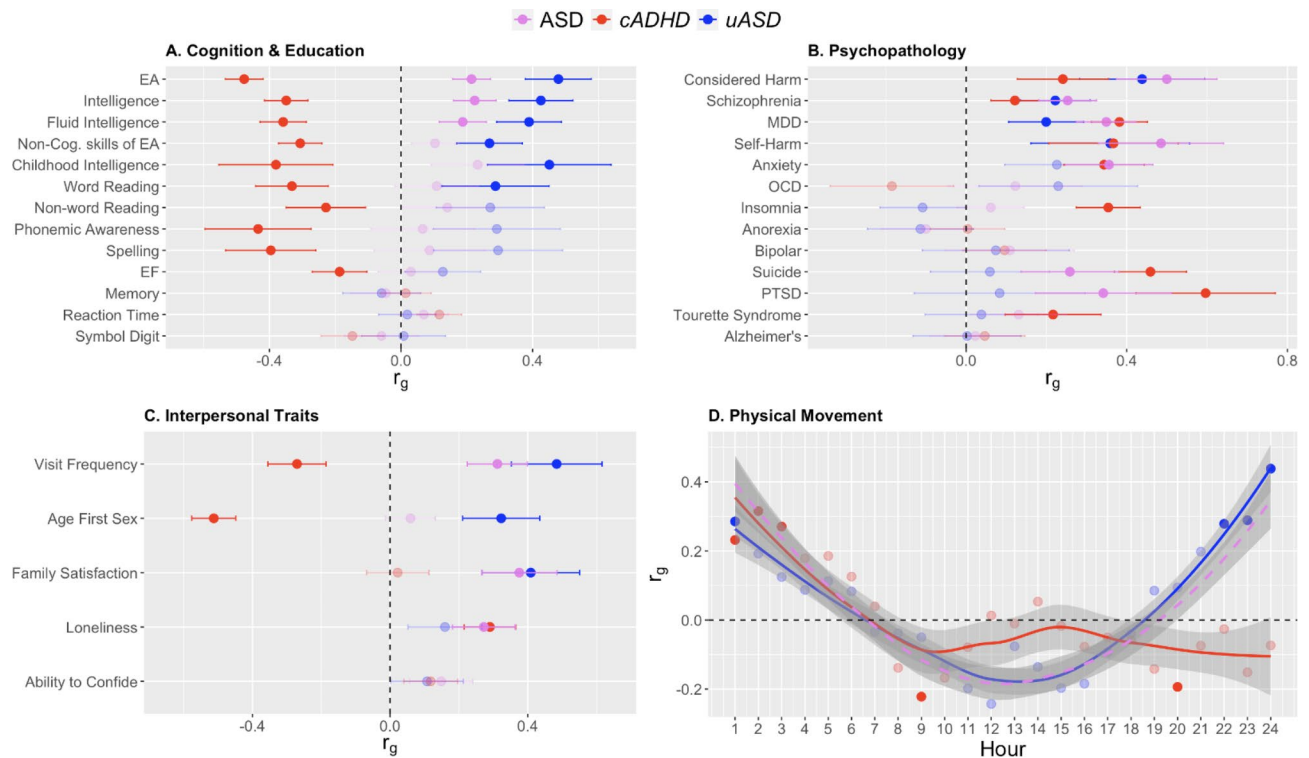
Schaffer *et al. Molecular Autism* (2024) 15:46

Page 5 of 11

Fig. 2 Genetic correlations between *uASD* and external traits. Genetic correlations between *uASD* (blue) and external traits for domains of cognition and education (**A**), psychiatric disorders and symptoms (**B**), and interpersonal traits (**C**). Traits are sorted top to bottom by ascending *p*-value for the *uASD* correlation. (**D**) Genetic correlations between accelerometer-based average total hourly movement within the 24-hour day beginning at midnight (i.e., hour 1) and *uASD* and *cADHD*. Correlations are also shown between external traits and ASD (pink) as well as *cADHD* (red). Error bars represent 95% confidence intervals. Translucent points and error bars represent genetic correlations that did not surpass the Bonferroni-adjusted significance threshold. Panel D depicts a LOESS regression line used to visualize overall trends across individual point estimates. Shaded region around the regression line represents a 95% confidence interval. Dashed pink line represents the LOESS regression line for genetic correlations with ASD. EA = educational attainment, EF = executive function (assessed using trail-making task B; TMTB), MDD = major depressive disorder, OCD = obsessive-compulsive disorder, PTSD = post-traumatic stress disorder, Visit Frequency = frequency of friend and family visits

Cognition

Cognitive-related phenotypes demonstrated the most robust genetic correlations with the latent factor *uASD*, both in statistical significance and magnitude. We observed positive correlations between *uASD* and educational attainment (r_g=0.48, SE=0.05), childhood intelligence ($r_g = 0.45$, SE=0.1), general intelligence ($r_g = 0.42$, SE=0.05), noncognitive skills of educational attainment ($r_g = 0.27$, SE=0.05), word reading ($r_g = 0.29$, SE=0.08), and verbal numerical reasoning ($r_g = 0.39$, SE=0.05). Interestingly, a number of these traits (childhood intelligence, noncognitive skills of educational attainment, and word reading) were not found to be significantly associated with ASD, indicating that *uASD* may be capturing additional genetic variance uniquely related to cognitive and education-related traits. While we observed consistent associations with traits indexing intellectual abilities, we did not observe a statistically significant correlation between *uASD* and the trail-making task B, a well-established measure of executive functioning [33, 34]. Collectively these cognitive results indicate that the genetic

component unique to ASD is specifically associated with more general cognitive processes independent of self-regulatory processes.

Psychopathology

Several psychiatric phenotypes, especially those falling within the internalizing spectrum, were positively and significantly associated with *uASD*. These included major depressive disorder (MDD; $r_g = 0.20$, SE=0.05), self-harm ($r_g = 0.36$, SE=0.10), consideration of self-harm ($r_g = 0.44$, SE=0.08), and sensitivity to environmental stress and adversity ($r_g = 0.16$, SE=0.05). Positive genetic correlations were also observed between *uASD* and schizophrenia ($r_g = 0.22$, SE=0.04) and anxiety ($r_g = 0.22$, SE=0.07), though the correlation with anxiety did not surpass the Bonferroni-adjusted significance threshold. These genetic correlations with psychopathology were similar in directionality and magnitude to those observed for ASD and *cADHD*. This suggests that, despite their high level of genetic overlap, ASD and ADHD each have

unique genetic pathways that link them to other forms of psychopathology.

Given the shared genetic influences across *uASD* and other psychiatric disorders, we went on to implement a follow-up model to examine the magnitude of unique genetic variance in ASD after partialing out genetic overlap with multiple psychiatric disorders. This involved running a multiple regression model with ADHD, MDD, anxiety, and schizophrenia as correlated predictors of ASD. The multiple regression model path diagram is visualized in Supplementary Fig. 1 and its output is provided in Supplementary Table 3. This model revealed that 74% of the residual genetic variance in ASD was unexplained by the other psychiatric disorders. This demonstrates that ASD is not merely a genetic amalgamation of ADHD and other psychiatric disorders, but has a sizable proportion of unique genetic variance distinguishing it from other forms of psychopathology.

Physical movement
We examined the genetic correlations with accelerometer data (i.e., physical movement) in 1-hour increments across a 24-hour period. This accelerometer data can be considered a useful indicator of atypical patterns of movement that may reflect disturbances in sleep and circadian rhythm, wherein these processes have been discussed as transdiagnostic correlates of psychiatric and neurodevelopmental disorders [17, 35, 36]. We observed positive associations between *uASD* and movement at hours 0–1 (i.e., 12:00AM-1:00AM; r_g = 0.29, SE=0.08), 21–22 (i.e., 9:00PM-10:00PM; r_g = 0.28, SE=0.07), 22–23 (r_g = 0.29, SE=0.08), and 23–24 (r_g = 0.44, SE=0.08), indicating that this genetic overlap was restricted to movement during periods of early morning and late night. We also observed overlap between *uASD* and physical inactivity (r_g for hours of moderate exercise=−0.25, SE=0.07) and irritable bowel syndrome (r_g = 0.21, SE=0.06).

Interpersonal relations
A select few traits relating to increased social behavior displayed genetic correlations with *uASD*. Notably, these positively associated traits tended to revolve around social relations with family members, such as family satisfaction (r_g = 0.41, SE=0.07) and the frequency of friend and family visits (r_g = 0.48, SE=0.07). Age of first sexual encounter was also positively genetically correlated with *uASD* (r_g = 0.32, SE=0.06), and this association was not observed with ASD.

Functional results identify uASD Enrichment
Stratified Genomic SEM revealed four significantly enriched functional annotations for *uASD*. Three of these annotations reflected classes of genes implicated in evolutionarily conserved processes, including genes conserved

in primates (p=1.00E-7), and two annotations indexing genes conserved in mammals (p=5.68E-4 and 9.45E-4). We also found significant enrichment for the H3K4me1 histone mark in the germinal matrix (p=6.20E-4), a transient brain region present only during gestational brain development (see Fig. 3 and Supplementary Table 3 for the magnitude of these enrichments). These annotations represent a subset of eight significantly enriched annotations in ASD, which also included genes conserved in vertebrates, as well as enrichment in histone marks in the anterior caudate and fetal male and female brains. The *cADHD* factor captured 26 significantly enriched annotations which encompassed all of the annotations observed in *uASD* and ASD, albeit with smaller point estimates for the evolutionary annotations. Additional significant annotations for *cADHD* were characterized predominantly by markers for genetic modifications in several brain regions and hormonal centers. A full list of annotations and their relative enrichments is provided in Supplementary Table 3.

T-SEM uncovers 83 genes Associated with uASD
We obtained 73 412 gene expression estimates for *uASD* (many of which reflect expression levels for the same gene in different tissues). T-SEM revealed 278 significant expression estimates across 83 unique gene IDs, many of which were clustered on chromosomes 8 and 17. These results are visualized as a Miami plot in Fig. 4. The most highly significant estimate corresponded with the downregulation of *PINX1* (z = -5.79, p=6.89E-9), a potent inhibitor of telomerase [37]. The univariate TWAS of ASD revealed 231 significant expression estimates across 69 genes at the same significance threshold used for the *uASD* T-SEM analysis (p<9.37E-5). The *uASD* T-SEM revealed 34 novel genes relative to the ASD univariate TWAS. Despite subtracting out shared genetic variance with ADHD, novel genes can arise in this model for genes with particularly discordant effects across ASD and ADHD.

To identify potential biological pathways implicated in *uASD*, we applied ORA to the significant genes identified by T-SEM. The analysis revealed two gene sets associated with the *uASD* genes; both of which were implicated in skin-related pathologies (*WebGestalt* output provided in Supplementary Table 6). The first set relates to bacterial skin diseases (enrichment ratio=74.53, p=6.55E-9). The second set corresponded to erythema (redness of the skin, often manifesting as a rash) and included all of the genes in the gene list for bacterial skin diseases, with the exception of *FAM167A* (enrichment ratio=43.27, p=2.20E-6).
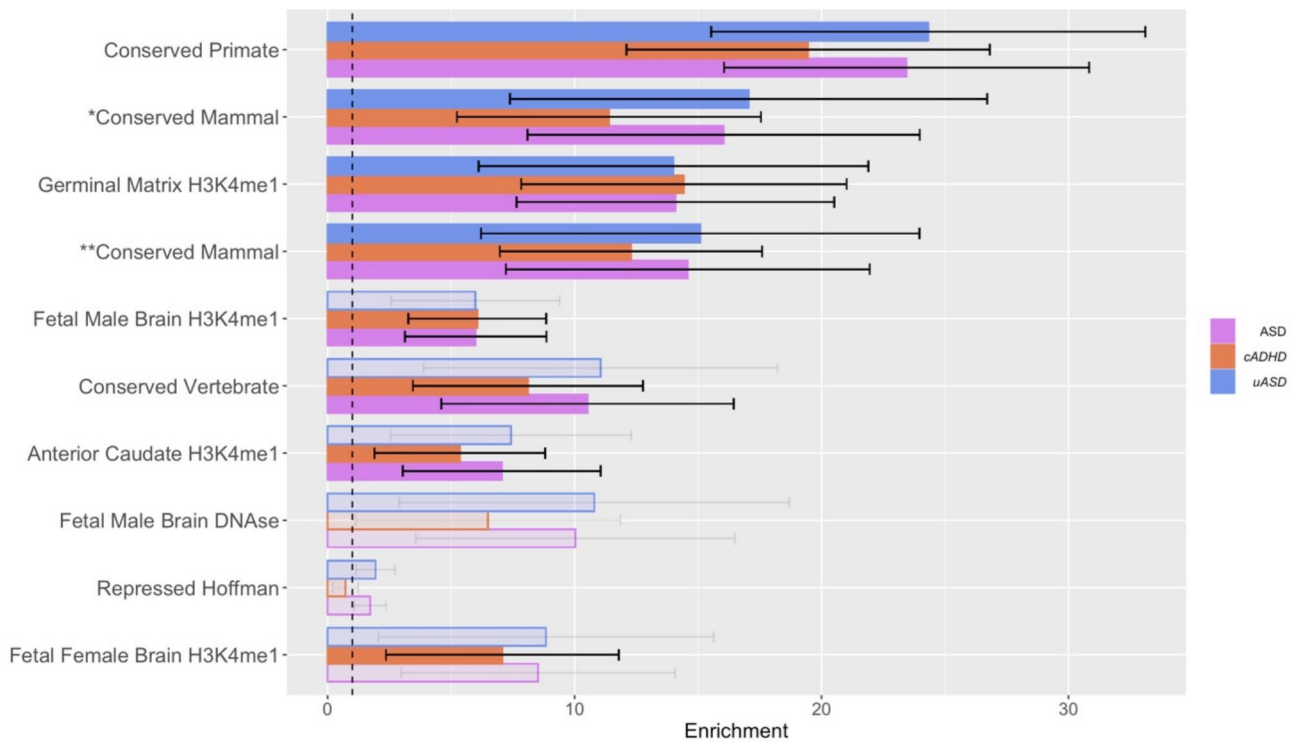
Fig. 3 Genetic enrichment of *uASD* for functional annotations. Functional annotations are arranged by ascending *p*-value. Enrichment is measured as the ratio of the proportion of genome-wide relative risk represented by the size of that annotation relative to the entire genome. Null enrichment value is 1.0 (visualized by dashed vertical line), in which the genetic variance captured by that annotation is proportional to the expected genetic variance based on annotation size. Significant enrichments at an FDR threshold are represented by solid blue bars, and error bars represent the 95% CI around the enrichment estimate. For visualization purposes, enrichment values are also provided for ASD (pink) and *cADHD* (red). Single asterisk (*) indicates conserved mammalian genes defined by GERP score. Double asterisk (**) indicates conserved mammalian genes defined by Lindblad-Toh et al. (2011)
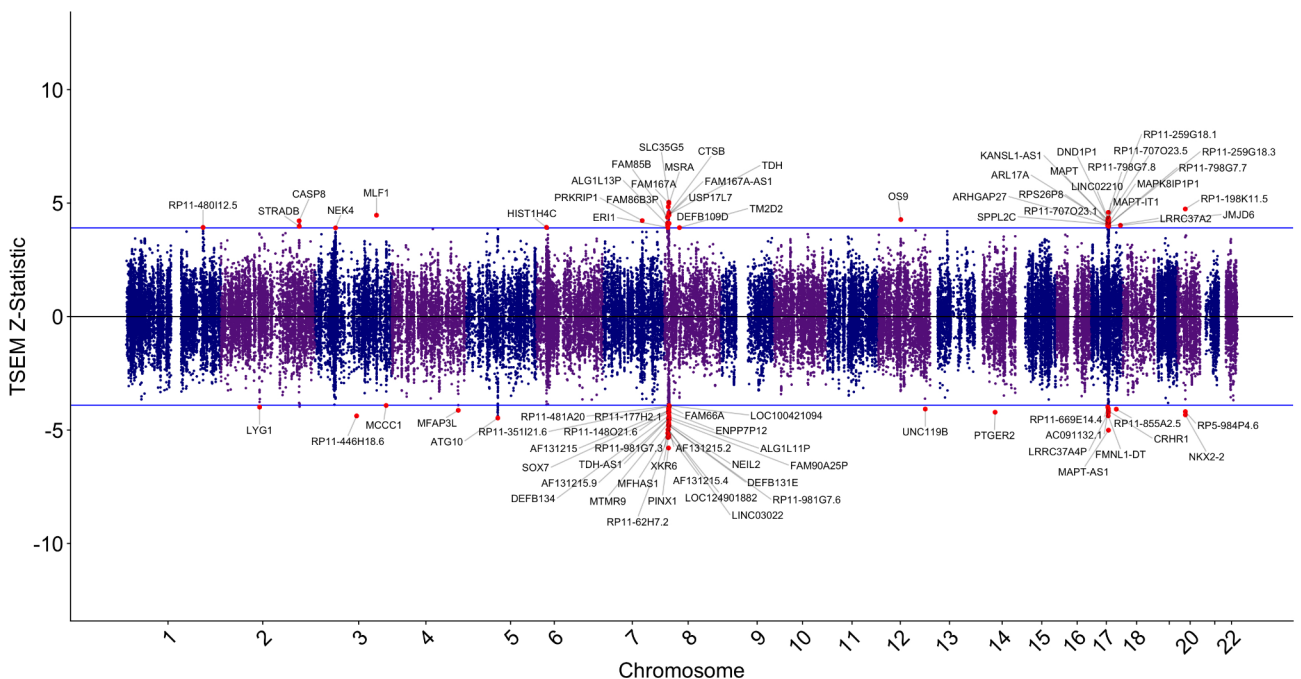

Fig. 4 Miami plot of gene expression hits for *uASD*. Upper and lower bounds represent the FDR-adjusted significance threshold (*p* < 9.37E-5). Genes surpassing the upper bound are upregulated and those below the lower bound are downregulated. Significant hits are colored red and labeled with gene ID names

Discussion

The present study leveraged Genomic SEM to dissect the genetic architecture specific to ASD after accounting for shared genetic variance with another neurodevelopmental disorder, ADHD diagnosed in childhood. These ADHD summary statistics evince the highest level of genetic overlap with ASD across the psychiatric space, thereby providing a stringent benchmark for indexing the genetic variance unique to ASD. We find that the majority of genetic variance in ASD, as well as genetic overlap with other clinically relevant traits, is unique from ADHD despite their high levels of genetic overlap [17]. Implementing genome-wide, functional, and gene-expression analyses, we investigated the unique genetic variance at increasing levels of biological granularity. At each level, we interrogated this specific variance and identified distinct biological pathways with specific relevance to ASD.

Genome-wide level

Partitioning the genetic variance unique to ASD, we find a plethora of genetic correlations with cognitive, psychiatric, and other behavioral traits. Notably, we find the strongest correlations between *uASD* and traits related to cognition and education. While the phenotypic literature surrounding ASD and cognitive abilities is mixed [38–40], our findings that *uASD* has a positive genetic association with intelligence and educational attainment corroborate prior genetic studies of ASD [2, 41–43]. Interestingly, we did not see genetic associations with executive function despite meta-analytic research demonstrating ubiquitous executive functioning deficits in ASD [44, 45]. Cognitive traits showed some of the more divergent patterns of association between *uASD* and *cADHD*, with the latter having large and negative genetic associations with the aforementioned cognitive traits. For many cognitive traits, the magnitude of the correlations seen with *uASD* surpassed those seen within ASD broadly. Thus, it appears the genetic variance unique to ASD may have opposing effects to the variance shared between ASD and ADHD, with the unique component driving correlations in the positive direction.

Contrary to cognitive traits, we observed general convergence in the directionality of genetic associations with psychiatric traits, especially in those relating to mood or anxiety disturbances. We therefore conclude that the genetic relationships between ASD and many psychiatric (especially internalizing) phenotypes are not driven solely by the genetic similarity between ASD and ADHD. We also show that ASD is not a simple conglomerate of the genetic components of ADHD and other psychiatric disorders, but rather a genetically distinct construct within the psychiatric space.

Functional genomic level

Functional analyses revealed that the genetic variance unique to ASD was concentrated in evolutionarily conserved genes and the H3K4me1 histone mark in the germinal matrix, a transitory brain region present during the prenatal period which serves as a hub for neural progenitor cells [46]. The enrichment of the H3K4me1 histone mark, indicative of active enhancer elements, reinforces the importance of early epigenetic modifications underlying neurodevelopmental processes in the etiology of ASD [47]. Moreover, our observation of this unique enrichment underscores the value of employing multivariate genomic analyses to dissect disorder-specific biological pathways, facilitating a deeper understanding of the molecular mechanisms underlying ASD.

Transcriptome-wide level

Transcriptome-wide analyses revealed 83 unique genes with differential expression linked with ASD independent of ADHD. Notably, we find several novel genes unique to ASD, reflecting genes with highly discordant effects across ASD and ADHD. Focusing on genes with expression associated with *uASD*, we find overlap with gene sets implicated in two classes of skin-related pathologies: bacterial skin disease and erythema (non-specific reddening of the skin). There are well-documented links between ASD and various immune-mediated conditions, with medical research focusing on the association with atopic dermatitis (i.e., eczema) [48–51]. Dysfunction of the immune system and inflammatory processes have been hypothesized to contribute jointly to ASD and skin disorders such as atopic dermatitis [52]; however, there is a dearth of literature apart from the current study which has explicitly identified gene clusters implicated in both syndromes. We note that our genetic correlation analyses do not provide ample support for pervasive overlap between *uASD*, skin pathologies, and inflammatory processes, indicating that the overrepresentation of genes implicated in skin pathology may represent the sharing of more circumscribed physiological systems undetected by genome-wide correlations.

Limitations

The current analyses were restricted to GWAS summary statistics derived exclusively from individuals of European ancestry [53], which hinders the generalizability of our findings due to differences in allele frequency and LD structure across ancestrally diverse populations [54]. Efforts to broaden representation in GWAS analyses are crucial for extending genetic insights to other ancestral groups [55]. Similarly, the ASD GWAS summary statistics used in these analyses pooled ASD cases across all clinical subgroups. While necessary to conduct well-powered analyses, this approach restricts our capability

to investigate the phenotypic heterogeneity inherent to ASD and to evaluate the invariance of our model across different axes of genetic and clinical variation. For example, autistic females tend to be report more internalizing symptoms and express fewer repetitive behaviors than males [56–58]. Similar divergences are observed at the genetic level for later-diagnosed ASD relative to ASD diagnosed earlier in life, with the former showing stronger genetic overlap with internalizing disorders and symptoms [59]. Unfortunately, we are restricted by the current limitations of stratified ASD summary data, which is relatively underpowered and/or yields negative heritability estimates [60], making them unsuitable for use in a multivariate genomic framework.

We are also limited in our ability to account for other sources of clinical heterogeneity, such as co-occurring cognitive impairments or psychopathologies, in the sample. However, prior work demonstrates that the common genetic architecture of ASD is largely consistent across groups with and without intellectual disability, with heritability differences speculatively driven by the presence of rare, pathogenic variants in groups with intellectual disability [2, 59, 61, 62]. We cannot completely rule out confounding from co-occurring psychopathology in our original GWAS sample. However, our multiple regression analysis revealed that associated internalizing psychopathologies explain only a marginal fraction of the genetic variance unique to ASD. Additionally, the high rates of internalizing concerns among autistic individuals suggest that these symptoms may be somewhat inherent to ASD phenomenology rather than an external confound [63]. A final consideration of our work is that many behavioral and psychiatric traits, including ASD, show discrepancies between family- and SNP-based estimates of heritability [64]. A substantial component of ASD's missing heritability is thought to be driven by rare variants, structural variants, de novo mutations, or other genomic variations typically undetected by GWASs [62, 65, 66]. It is therefore important to recognize that we are only examining a subset of the genetic factors thought to independently contribute to ASD etiology.

Future directions
Our work highlights several avenues for future research in the field. The model we utilized here represents a generalizable framework for partitioning the genetic variance shared between correlated traits into shared and unique components. As well-powered and subgroup-stratified summary statistics become increasingly available, extensions of this model can be applied to parse the genetic heterogeneity observed in ASD. For example, the same analytic approach can be used to isolate the genetic variance unique to late-diagnosed autism independent of early-diagnosed autism. Such work provides

the opportunity to formally assess whether clinical distinctions within ASD are mirrored by disparate genetic pathways. Finally, although the current study focused on common genetic variation, future lines of research should also work to characterize the shared and unique effects of alternative forms of genetic variation on neurodevelopmental outcomes.

Conclusions
Taken together, we provide insights across multiple levels of biology that characterize the genetic signature unique to ASD. Relative to ADHD, we find evidence for divergent patterns of relationships with a range of clinically relevant correlates (e.g., cognition) along with unique patterns of functional enrichment and gene expression that implicate neurobiological processes and disease states linked to ASD in the extant literature. While ASD has often been discussed as unique within the psychiatric space, the current findings clarify and characterize the biological substrata that differentiate this complex neuropsychiatric condition.

Abbreviations
ASD   Autism spectrum disorder
ADHD   Attention-deficit/hyperactivity disorder
SEM   Structural equation modeling
T-SEM   Transcriptome-wide structural equation modeling
LD   Linkage disequilibrium
LDSC   Linkage disequilibrium score regression

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13229-024-00624-2.

Supplementary Material 1

Supplementary Material 2

Author contributions
LSS and ADG conceptualized the study; LSS led data analysis; LSS and ADG interpreted results; LSS composed the manuscript; ADG supervised the composition of the manuscript; SB, JML, IFF, and ADG compiled and processed summary statistics; SB and JML provided substantial computational support; all authors contributed to visualizations; all authors read and provided feedback on the final manuscript.

Data availability
GWAS summary statistics for ASD are available for download at https://figshare.com/articles/dataset/asd2019/14671989. GWAS summary statistics for childhood-diagnosed ADHD are available for download at https://ipsych.dk/en/research/downloads/. The sources for publicly available external trait summary statistics are provided in Supplementary Table 1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Preregistration
These analyses were not preregistered.

Competing interests
The authors declare no competing interests.

References

1. Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, et al. The epidemiology of autism spectrum disorders. Annu Rev Public Health. 2007;28:235–58.
2. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
3. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596
4. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016;57:585–95.
5. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of Autism Spectrum Disorder. JAMA. 2017;318:1182–4.
6. Xie S, Karlsson H, Dalman C, Widman L, Rai D, Gardner RM, et al. The familial risk of Autism Spectrum disorder with and without intellectual disability. Autism Res. 2020;13:2242–50.
7. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46:881–5.
8. Khachadourian V, Mahjani B, Sandin S, Kolevzon A, Buxbaum JD, Reichenberg A, et al. Comorbidities in autism spectrum disorder and their etiologies. Transl Psychiatry. 2023;13:71.
9. Stevens T, Peng L, Barnard-Brak L. The comorbidity of ADHD in children diagnosed with autism spectrum disorder. Res Autism Spectr Disord. 2016;31:11–8.
10. Ronald A, Larsson H, Anckarsäter H, Lichtenstein P. Symptoms of autism and ADHD: a Swedish twin study examining their overlap. J Abnorm Psychol. 2014;123:440–51.
11. Peyre H, Schoeler T, Liu C, Williams CM, Hoertel N, Havdahl A, et al. Combining multivariate genomic approaches to elucidate the comorbidity between autism spectrum disorder and attention deficit hyperactivity disorder. J Child Psychol Psychiatry. 2021;62:1285–96.
12. Ghirardi L, Brikell I, Kuja-Halkola R, Freitag CM, Franke B, Asherson P, et al. The familial co-aggregation of ASD and ADHD: a register-based cohort study. Mol Psychiatry. 2018;23:257–62.
13. Reiersen AM, Constantino JN, Grimmer M, Martin NG, Todd RD. Evidence for shared genetic influences on self-reported ADHD and autistic symptoms in young adult Australian twins. Twin Res Hum Genet. 2008;11:579–85.
14. Ronald A, Simonoff E, Kuntsi J, Asherson P, Plomin R. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. J Child Psychol Psychiatry. 2008;49:535–42.
15. Rajagopal VM, Duan J, Vilar-Ribó L, Grove J, Zayats T, Ramos-Quiroga JA, et al. Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder. Nat Genet. 2022;54:1117–24.
16. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav. 2019;3:513–25.
17. Grotzinger AD, Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewis CM, et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. Nat Genet. 2022;54:548–59.
18. Grotzinger AD, Singh K, Miller-Fleming TW, Lam M, Mallard TT, Chen Y, et al. Transcriptome-wide structural equation modeling of 13 Major Psychiatric disorders for Cross-disorder Risk and Drug Repurposing. JAMA Psychiatry. 2023;80:811–21.
19. Breunig S, Lawrence JM, Foote IF, Gebhardt HJ, Willcutt EG, Grotzinger AD. Examining differences in the Genetic and Functional Architecture of Attention-Deficit/Hyperactivity disorder diagnosed in Childhood and Adulthood. Biol Psychiatry Glob Open Sci. 2024;4:100307.
20. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
21. Grotzinger AD, Fuente J, de la, Privé F, Nivard MG, Tucker-Drob EM. Pervasive Downward Bias in estimates of liability-scale heritability in genome-wide Association Study Meta-analysis: a simple solution. Biol Psychiatry. 2023;93:29–36.
22. Peyrot WJ, Robinson MR, Penninx BWJH, Wray NR. Exploring boundaries for the genetic consequences of assortative mating for Psychiatric traits. JAMA Psychiatry. 2016;73:1189–95.
23. Belsare S, Levy-Sakin M, Mostovoy Y, Durinck S, Chaudhuri S, Xiao M, et al. Evaluating the quality of the 1000 genomes project data. BMC Genomics. 2019;20:620.
24. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890.
25. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
26. GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
27. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.
28. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.
29. GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45:580–5.
30. Hoffman GE, Bendl J, Voloudakis G, Montgomery KS, Sloofman L, Wang Y-C, et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. Sci Data. 2019;6:180.
31. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science. 2018;362. https://doi.org/10.1126/science.aat8127
32. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47:W199–205.
33. Kortte KB, Horner MD, Windham WK. The trail making test, part B: cognitive flexibility or ability to maintain set? Appl Neuropsychol. 2002;9:106–9.
34. MacPherson SE, Allerhand M, Cox SR, Deary IJ. Individual differences in cognitive processes underlying trail making Test-B performance in old age: the Lothian Birth Cohort 1936. Intelligence. 2019;75:23–32.
35. Robinson-Shelton A, Malow BA. Sleep disturbances in Neurodevelopmental disorders. Curr Psychiatry Rep. 2016;18:6.
36. Mijnster T, Boersma GJ, van Veen MM, Liemburg E, Cath D, Pijnenborg GHM, et al. Sleep disorders in a naturalistic cohort of Dutch psychiatric outpatients: prevalence rates and associations with psychopathology symptom severity and well-being. J Sleep Res. 2024;33:e14009.
37. Zhou XZ, Lu KP. The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. Cell. 2001;107:347–59.
38. Joseph RM, Tager-Flusberg H, Lord C. Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. J Child Psychol Psychiatry. 2002;43:807–21.
39. Frith U, Hill EL, Tager-Flusberg H, Joseph RM. Identifying neurocognitive phenotypes in autism. Philos Trans R Soc Lond B Biol Sci. 2003;358:303–14.
40. Wolff N, Stroth S, Kamp-Becker I, Roepke S, Roessner V. Autism Spectrum Disorder and IQ - A Complex Interplay. Front Psychiatry. 2022;13:856084.
41. Clarke T-K, Lupton MK, Fernandez-Pujals AM, Starr J, Davies G, Cox S, et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. Mol Psychiatry. 2016;21:419–25.

42. Rao S, Baranova A, Yao Y, Wang J, Zhang F. Genetic relationships between Attention-Deficit/Hyperactivity disorder, Autism Spectrum Disorder, and intelligence. Neuropsychobiology. 2022;81:484–96.

43. Hatoum AS, Morrison CL, Mitchell EC, Lam M, Benca-Bachman CE, Reineberg AE, et al. Genome-wide Association Study shows that executive functioning is influenced by GABAergic processes and is a Neurocognitive Genetic Correlate of Psychiatric disorders. Biol Psychiatry. 2023;93:59–70.

44. Demetriou EA, Lampit A, Quintana DS, Naismith SL, Song YJC, Pye JE, et al. Autism spectrum disorders: a meta-analysis of executive function. Mol Psychiatry. 2018;23:1198–204.

45. Hemmers J, Baethge C, Vogeley K, Falter-Wagner CM. Are executive dysfunctions relevant for the Autism-Specific Cognitive Profile? Front Psychiatry. 2022;13:886588.

46. Raets MMA, Dudink J, Govaert P. Neonatal disorders of germinal matrix. J Matern Fetal Neonatal Med. 2015;28(Suppl 1):2286–90.

47. Yoon SH, Choi J, Lee WJ, Do JT. Genetic and Epigenetic Etiology Underlying Autism Spectrum Disorder. J Clin Med Res. 2020;9. https://doi.org/10.3390/jcm9040966

48. Billeci L, Tonacci A, Tartarisco G, Ruta L, Pioggia G, Gangemi S. Association between atopic dermatitis and Autism Spectrum disorders: a systematic review. Am J Clin Dermatol. 2015;16:371–88.

49. Zerbo O, Leong A, Barcellos L, Bernal P, Fireman B, Croen LA. Immune mediated conditions in autism spectrum disorders. Brain Behav Immun. 2015;46:232–6.

50. Jameson C, Boulton KA, Silove N, Guastella AJ. Eczema and related atopic diseases are associated with increased symptom severity in children with autism spectrum disorder. Transl Psychiatry. 2022;12:415.

51. Wang B, Li Y-D, Wang Z-Y, Zhao J-Q, Zhang G-Q, Man M-Q. Alterations in Epidermal Biophysical Properties in Autistic Children. Skin Pharmacol Physiol. 2023;36:160–4.

52. Man M-Q, Yang S, Mauro TM, Zhang G, Zhu T. Link between the skin and autism spectrum disorder. Front Psychiatry. 2023;14:1265472.

53. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in Human Genetic studies. Cell. 2019;177:26–31.

54. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–91.

55. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. Nat Med. 2022;28:243–50.

56. Mandy W, Chilvers R, Chowdhury U, Salter G, Seigal A, Skuse D. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. J Autism Dev Disord. 2012;42:1304–13.

57. Solomon M, Miller M, Taylor SL, Hinshaw SP, Carter CS. Autism symptoms and internalizing psychopathology in girls and boys with autism spectrum disorders. J Autism Dev Disord. 2012;42:48–59.

58. Szatmari P, Liu X-Q, Goldberg J, Zwaigenbaum L, Paterson AD, Woodbury-Smith M, et al. Sex differences in repetitive stereotyped behaviors in autism: implications for genetic liability. Am J Med Genet B Neuropsychiatr Genet. 2012;159B:5–12.

59. Zhang X, Grove J, Gu Y, Buus CK, Nielsen LK, Neufeld SAS et al. An axis of genetic heterogeneity in autism is indexed by age at diagnosis and is associated with varying developmental and mental health profiles. medRxiv. 2024;2024.07.31.24311279.

60. Schwaba T, Mallard TT, Maihofer AX, Rhemtulla M, Lee PH, Smoller JW et al. Comparison of the multivariate genetic architecture of eight major psychiatric disorders across sex. bioRxiv. 2023 [cited 2024 Sep 9]. p. 2023.05.25.23290545. https://www.medrxiv.org/content/https://doi.org/10.1101/2023.05.25.23290545v1.abstract

61. Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet. 2017;49:504–10.

62. Antaki D, Guevara J, Maihofer AX, Klein M, Gujral M, Grove J, et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat Genet. 2022;54:1284–92.

63. Kerns CM, Kendall PC. The presentation and classification of anxiety in autism spectrum disorder. Clin Psychol (New York). 2012;19:323–47.

64. Maher B. Personal genomes: The case of the missing heritability. Nature Publishing Group UK. 2008 [cited 2024 Apr 26]. https://doi.org/10.1038/456018a

65. Mollon J, Almasy L, Jacquemont S, Glahn DC. The contribution of copy number variants to psychiatric symptoms and cognitive ability. Mol Psychiatry. 2023; https://doi.org/10.1038/s41380-023-01978-4

66. More RP, Warrier V, Brunel H, Buckingham C, Smith P, Allison C, et al. Identifying rare genetic variants in 21 highly multiplex autism families: the role of diagnosis and autistic traits. Mol Psychiatry. 2023;28:2148–57.

Publisher's note