Data supplement for Ecker et al., Interindividual Differences in Cortical Thickness and Their Genomic Underpinnings in Autism Spectrum Disorder. Am J Psychiatry (doi: 10.1176/appi.ajp.2021.20050630).

CONTENTS

Supplementary Material and Methods

Sample Description

All ASD participants had a clinical diagnosis of ASD according to DSM-5 or ICD-10 criteria. We also assessed ASD symptoms using the Autism Diagnostic Observation Schedule (ADOS, (1)), and the Autism Diagnostic Interview-Revised (ADI-R, (2)). However, individuals with a clinical ASD diagnosis who did not reach cut-offs for ASD were not excluded to capture the full (i.e. broader) autism phenotype in the general population. Exclusion criteria included significant hearing or visual impairments, a history of alcohol and/or substance abuse or dependence in the past year, and the presence of MRI contraindications. We also excluded participants without IQ assessments, as well as controls with *t*-scores>70 on the Social Responsiveness Scale-2 (SRS-2, (3)) due to inclusion of individuals with mild intellectual disability (ID). Level of intellectual abilities was assessed using the Wechsler Abbreviated Scales of Intelligence—Second Edition (WASI-II (4)) or — in countries where the WASI is not translated (i.e. The Netherlands, Germany and Italy) — the four-subtest short forms of the German, Dutch or Italian WISC-III/IV (5) for children or WAIS-III/IV (6) for adults. Furthermore, we assessed DSM-5 Domain B symptoms via the Repetitive Behaviours Scale (RBS-R (7)), and the Short Sensory Profiles (SSP (8)). Notably, unlike most questionnaires assessing symptom severity, the scale of the SSP is negatively scored, so that larger values indicate less severe symptoms. To facilitate the interpretation of brain-behaviour correlations, the SSP scores were therefore inversed, so that larger values indicate more severe symptoms. In the ASD sample, the presence of other neurodevelopmental conditions and psychiatric disorders (except for psychosis or bipolar disorder) was allowed. Given the high number of autistic individuals taking regular medication, participants on stable medication were also included (see ST1 for details).

MRI Data Quality Assessments

Structural MRI data was initially available for a total of 709 individuals in the LEAP sample, which was acquired across six European ASD centres of excellence: (i) Institute

of Psychiatry, Psychology and Neuroscience, King's College London (IoPPN/KCL, United Kingdom), (ii) Autism Research Centre, University of Cambridge (UCAM, United Kingdom), (ii) University Medical Centre Utrecht (UMCU, Netherlands), (iv) Radboud University Nijmegen Medical Centre (RUNMC, Netherlands), (v) Central Institute of Mental Health (CIMH, Germany), and (vi) the University Campus Bio-Medico (UCBM) in Rome, Italy. We initially preprocessed all available data using the default pipeline implemented in the FreeSurfer v6.0.0 software (http://surfer.nmr.mgh.harvard.edu/). The resulting surface reconstructions were visually inspected for reconstruction errors and rated by three independent raters, blind to group membership, who had the options to either (1) accept a reconstruction 'as is' (n=347 or 48.8%), (2) reject 'as is' (n=52 or 7.5%) mostly due to severe (motion) artefacts and/or the existence of extra-brain tissue that precluded a successful FreeSurfer reconstruction, or (3) to prescribe manual editing (n=310 or 43.7%) in case of smaller (i.e. 'local') reconstruction errors. Following manual editing, the 310 images were (re)preprocessed and visually (re)assessed. Out of these, 3 surface reconstructions did not improve significantly and were subsequently excluded from the statistical analysis. A further 15 scans were excluded due to scanner upgrades and missing FSIQ, which meant that a total of 70 scans (i.e. 9.9%) were excluded overall. The overall dropout was approximately equally distributed across sites with 10 out of 88 scans (i.e., 11.4%) excluded from Cambridge (n=5 ASD, n=5 controls), 31 out of 241 scans (i.e., 12.9%) excluded from the IoPPN (n=28 ASD, n=3 controls), 8 out of 69 scans (i.e., 11.6%) excluded from Mannheim (n=5 ASD, n=3 controls), 13 out of 184 scans (i.e., 7.1%) excluded from Nijmegen (n=11 ASD, n=2 controls), and 8 out of 86 scans (i.e., 9.3%) excluded from Utrecht (n=7 ASD, n=1 TD). None of the 41 scans from Rome had to be excluded. In terms of diagnostic categories, we excluded a total of 46 out of 353 individuals with ASD (i.e., 13.0%), 10 out of 63 individuals with an intellectual disability (ID) and ASD (i.e., 15.9%), 12 out of 266 TD controls (i.e., 4.5%), and 2 out of 27 TD controls with ID (i.e. 7.4%) across sites. The final (i.e., analyzed) sample thus consisted of 639 individuals (n=360 with ASD, n=279 TD controls).

To assess the influence of MRI data quality on our results, we also examined the Euler number of each of the created FreeSurfer surface reconstructions following manual

editing. The Euler number is a measure of the topological complexity of the reconstructed cortical surface as calculated by the sum of the vertices and faces subtracted by the number of faces (9). Because the Euler number is calculated separately for each hemisphere, we computed the sum of the values across hemispheres here to produce one value per subject. We found that there were no significant differences in the total Euler number between groups (t_{596}=1.73, p<0.1), which means that both groups are matched in terms of the quality of the surface reconstruction SF9a). Moreover, we observed that covarying for the total Euler number in the GLM did not significantly affect the magnitude or pattern of the between-group differences in CT (SF9b). Last, we repeated the analyses of neuroanatomical deviations by including the Euler number as predictor within our neurotypical modelling approach. Including the Euler number did not significantly affect the individual's *tAIs*, the outcomes of the logistic regression, or the spatially distributed pattern of the ASD enrichment mask (SF9c,d). It therefore seems that our results are largely unaffected by the quality of the surface reconstruction, which is why we did not covary for overall reconstruction quality within the main GLM.

Medication

Given the high number of autistic individuals taking regular medication (30-50% in Europe (10, 11) and 70% in US (12)), participants on stable medication were also included. Medication status (based on the ATC classification system, Level-1 ATC code "N") was known for n=531 individuals within our sample. Out of those, n=389 individuals were medication free at the time of study participation. (181 ASD, 208 TD). Overall, n=142 participants (122 ASD, 20 TD) were using at least one and a maximum of three different medications that included antidepressants (34 ASD, 6 TD), antiepileptics (11 ASD, 2 TD), antimigraine preparations (4 ASD, 0 TD), antipsychotics (28 ASD, 0 TD), anxiolytics (2 ASD, 1 TD), drugs used to treat addictive disorders (0 ASD, 1 TD), hypnotics and sedatives (40 ASD, 2 TD), other analgesics and antipyretics (4 ASD, 4 TD), and psychostimulants or other mediation to treat ADHD symptoms (47 ASD, 9 TD). A summary of the different types of medication used within the ASD and TD group is provided in ST1.

To establish the impact of medication status on our results, we initially performed a nested-model comparison to test whether the inclusion of medication status as fixed effect variable significantly improved the model fit. To do so, we subdivided our sample into those with known positive medication status (122 ASD, 20 TD), and those with negative or unknown medication status (238 ASD, 259 TD). However, there was no significant improvement of model fit when covarying for medication status, which means that the main effect of medication status was not significant (SF8a). Furthermore, the main clusters of significant between-group differences in CT remained significant when including medication status as covariate in the GLM (SF8b). To establish whether medication status had an impact on neuroanatomical deviations in CT within the ASD group, we also compared the probability of being a neuroanatomical outlier resulting from the logistic regression analysis of *tAIs* between ASD individuals on (n=122) and off (n=181) medication. However, ASD individuals with medication use were equally likely to be a neuroanatomical outlier than ASD individuals not taking any medication ($t(253)=0.346$, $p=0.729$) (SF8c). Last, to examine whether medication status impacts on neuroanatomical variability within the ASD group, we re-estimated the neurotypical model following the inclusion of medication status as a fixed effect, and statistically compared variability in tAI_s using a Paired Pitman Morgan test for the significance of the difference between two variances. There were no significant differences in tAI_s variability between ASD individuals on and off medication ($t(637)=1.75$, $p=0.1$) (SF8d). It was therefore concluded that medication status had little or no effect on the presented results.

Surface-based smoothing

To establish the effect of the surface-based smoothing kernel on our results, and to substantiate the choice of the employed kernel, we also examined vertex-wise measures of effect sizes associated with the main effect of group, as well as clusters of significant between-group differences across kernels with different full width at half maximum (FWHM). As can be seen in SF4a, vertex-wise measures of effect sizes (assessed via *Cohen's f*) increased across the cortex with increasing FWHM from 5mm, 10mm, 15mm,

to 20mm. The positive relationship between smoothing kernel and effect sizes was also observed on the cluster level, with a larger number of significant clusters being observed as the FWHM increased from 5 to 15 mm (SF4b-e). However, increasing the FWHM beyond 15mm had a detrimental effect on cluster-level effects, which not only depend on the effect sizes at individual vertices, but also on their distribution across the cortex; i.e., the effect size of strong focal effects in small (i.e., isolated) brain regions may be reduced on the cluster-level when applying a large smoothing kernel, whereas small to medium effects in neighboring brain regions might benefit from a larger kernel on the cluster level. We also explored the effects of the smoothing kernel on the Levene's test for homogeneity of variances between groups, which are presented in SF5. Overall, the main brain regions with increased variability in ASD remained stable across kernels, with a larger number of local peaks observed at smaller kernels (5 and 10mm), and fewer but larger peaks at larger kernels (15 and 20mm), as would be expected. Based on these results, we therefore opted for a smoothing kernel of 15mm in the present study, which seems optimally suited to detect between-group differences and variability in CT across groups.

Prediction interval threshold

To establish the effect of the prediction interval (PI) on our results, and to substantiate the choice of the employed PI in our study, we also examined the results of the neurotypical model based on different PIs (80%, 85%, 90%, 95%, 99%). At each PI threshold, we have re-computed the map of neuroanatomical outliers with an enrichment of ASD individuals outside the neurotypical PI, which was subsequently used to derive the individual's total degree of neuroanatomical abnormality (tAI_s), and the accuracy, sensitivity, and specificity of the logistic regression analysis. As can be seen in SF7, the choice of the neurotypical PI had a small effect on the spatially distributed patterns of neuroanatomical outliers across the cortex, with effect sizes starting to decrease beyond a threshold of 90% (left panel of SF7). Changing the PI threshold also had a relatively small effect on the overall accuracy of the model, which ranged from 74.5% at a

neurotypical PI of 85% to 76.8% at a PI of 95% (right panel of SF7). However, as expected, varying the PI significantly impacted on the distribution of the tAIs in the neurotypical controls, were little or no variability was observed at a PI of 99% (i.e., 99% of controls are within the expected range of variability in CT and hence no variability is observed) (middle panel of SF 7). Moreover, at a PI of 95%, the distribution of tAIs within the controls remained 'skwed left' (i.e. non-symmetric), and hence significantly deviated from a Gaussian distribution, which can be problematic for subsequent analysis (e.g. brain-behavior correlations) that rely on normally distributed data. Only at a neurotypical PI equal or lower than 90%, the distribution of *tAIs* was approximately normal, with sufficient variability in both groups, and a prediction accuracy that was comparable to higher PI thresholds. For the purpose of our study, we therefore choose a neurotypical PI of 90%, as it seems particularly well suited for the quantification of neuroanatomical outliers with sufficient variability with both neurotypical controls as well as individuals with ASD.

Site effects

As our study was a multi-centre MRI investigation with parallel recruitment at six European sites, it was important to account for scanner- or site-related confounds that may be unrelated to the main effect of group, but may still have an impact on our findings. Site effects are typically accounted for via (i) the inclusion of site as a fixed effect factor within the GLM, which makes it possible to explicitly model site effects at each cerebral vertex-level. However, a variety of other approaches have been suggested including (ii) the ComBat batch adjustment method prior to the statistical analysis (13), and (iii) including site as a random effect variable within a linear mixed effects model. Here, we therefore also examined the robustness of our results across different approaches for dealing with site effects. We find that there was little difference between the three different techniques overall, and the clusters with a significant between-group difference in CT were stable across approaches (SF3). As shown in SF3c,d, including site as a random effect within a linear mixed effect model had virtually no effect on our results compared to the computationally less intensive fixed effect model. Moreover, applying

ComBat correlation prior to statistically modelling did not affect the spatially distributed pattern of significant differences overall. However, covarying for site directly within the GLM improved effect sizes in the right temporal lobe cluster (in the fixed and random effect model), which did not reach statistical significance when applying ComBat correction (SF3b). Based on these findings, we therefore opted for the inclusion of site as fixed effect factor within the GLM, which also makes our results directly comparable to previous multi-centre neuroimaging studies on ASD by our team and others (e.g. (14)).

Effects of Intellectual Disability (ID)

A number of individuals within our sample had a mild intellectual disability (ID) defined as a full-scale IQ < 70 (N=25 controls and N=53 ASD). The non-ASD individuals with ID were not included in the neurotypical model, but were included in the analysis of between-group differences. To further investigate the potential effects of ID, we also compared ASD and non-ASD individuals with ID to those without. We found that both ASD and non-ASD individuals with ID differed significantly in CT from their non-ID counterparts (see SF15a,c). Overall, however, there was little overlap between the set of brain regions associated with the main effect of ID in non-ASD individuals and the set of brain regions where ASD individual significantly differed from non-ASD individuals. Furthermore, few small clusters remained statistically significant when inter-individual variability in full-scale IQ was accounted for within the GLM (SF15b,d). These findings suggest that while a mild ID might affect measures of CT in various regions across the cortex, these effects can efficiently be accounted for by including full-scale IQ as a continuous covariate in the GLM.

To determine how ID affects the results of the analysis of neuroanatomical outliers, we have also repeated our modelling approach by defining the neurotypical range based on all non-ASD individuals in our sample (i.e., including those with a mild ID). We find that both models (i.e. in-/excluding ID individuals to/from the neurotypical range) resulted in a very similar, spatially-distributed pattern of neuroanatomical outliers, and a very similar distribution of *tAIs* within groups (see SF10). There were also no significant differences

in the overall accuracy ($\chi^2(1)=0.11$, $p=0.75$), sensitivity ($\chi^2(1)=0.33$, $p=0.56$), and specificity ($\chi^2(1)<0.001$, $p=0.999$) between models, which is most likely due to the small number of ID controls within our sample (25 out of 279; 0.8%). Taken together, these findings suggest that the inclusion of non-ASD individuals with ID only had a minimal effect on the presented between-group differences, and the associated patterns of neuroanatomical outliers in the ASD group.

Effects of biological sex

Our sample included both males and females. While a detailed (i.e., thorough) investigation of sex differences in the neuroanatomy of ASD would go beyond the scope of the present study, we also performed a preliminary analysis of group-by-sex interactions to aid the composition of the GLM with regards to included model terms, and also to determine whether males or females have a higher or lower probability of being a neuroanatomical outlier. When extending the GLM described in the main body of the manuscript by a group-by-sex interaction term, we found significant group-by-sex interactions only in one larger cluster located in the right medial orbitofrontal cortex (see SF14a). In all other areas of the cortex, the group-by-sex interaction remained non-significant on the cluster level. In most areas of the cortex, it is therefore sufficient to account for sex effects by including biological sex as a fixed effect factor without the group-by-sex interaction term.

Based on the neurotypical model that included both males and females, we also established that males were no more likely to be a neuroanatomical outlier than females ($t(191)=0.34$, $p=0.72$) utilizing the predictive probabilities resulting from the logistic regression analyses, which used the individual's total degree of neuroanatomical abnormality (tAIs) within the enrichment mask as predictor (SF17). Last, we have established separate models defining the neurotypical range of CT for males and females exclusively. Based on these sex-specific models, we find that the probability of being a neuroanatomical outlier for ASD females (based on an all-female model; N=90 out of 101 ASD females) did not significantly differ from the probability of being an outlier for ASD

males (based on an all-male sample; N=217 out of 259 ASD males) ($\chi^2(1)=1.24$, p=0.26). This is of interest as our results suggest that ASD females are equally likely to be neuroanatomical outliers as ASD males despite the well-documented sex bias in the prevalence of ASD in the general population. There were also no significant differences in the polygenic risk for any of the examined major psychiatric disorders between males and females (i.e., $p>0.05$, two-tailed)(see SF25).

Gene Expression Decoding Analysis

To link our neuroanatomical findings to underlying genomic mechanisms, we leveraged the spatial gene expression data from the Allen Human Brain Atlas (AHBA; (15)) to identify a list of genes with a spatial pattern of expression that resembles the neuroanatomical patterns highlighted by our statistical neuroimaging analyses. To this aim, we initially uploaded (1) the statistical *t-map* associated with the main effect of group (Fig 1b), (2) the *Cohen's f* map representing the effect sizes for the main effect of group (Fig 1d), as well as (3) the χ^2-*outlier* map with a significant enrichment of ASD individuals outside the neurotypical $PI_{90\%}$ (Fig 2e) to the Neurovault server (https://neurovault.org). Next, using python code embedded within Neurovault and Neurosynth (https://neurosynth.org), we performed a gene expression decoding analysis that statistically assesses the spatial correlation between our statistical maps and the pattern of gene expression for each of a total of 20,787 protein coding genes (16). To do so, the six AHBA donor brains are initially co-registered with the MNI atlas (also used by FreeSurfer) using nonlinear registration (transcriptomic alignment). At each sampling site (i.e. probe), a spherical region-of-interest (ROI) is drawn (default radius r=4mm), and the statistical test parameter in each FreeSurfer overlay is averaged within each ROI. This resulted in a spatial vector of values for each donor, which was subsequently correlated with the normalised gene expression data (see SF20). Here, the analysis constructs a linear model for each donor brain, where the slopes encode the spatial correlation between each gene's expression pattern to the statistical neuroimaging map (random effects model). In line with the input maps, these analyses were restricted to cortical tissue. The slopes are then subjected to a one-sample *t-test* to identify genes whose

spatial expression patterns are consistently highly similar to the imaging maps (i.e. across donor brains). The derived list of genes was thresholded at $p<.01$, which resulted in N=546 significant genes for the *t-map*, N=408 significant genes for the *Cohen's f* map, and N=662 significant genes for the χ^2-*outlier* map. We chose this 'liberal' threshold as this analysis did not constitute a hypothesis test *per se*, but rather a selection step aimed at yielding an initial list of genes for the subsequent analyses. Given that both sides of our imaging contrasts were of equal relevance, we considered both positive and negative *t*-statistic values.

Gene Enrichment Analyses

Next, we performed several gene enrichment analyses to establish the biological relevance and functional role of decoded genes. All enrichment testing was performed using the GeneOverlap package in R (10.18129/B9.bioc.GeneOverlap). Specifically, we tested the decoded gene-lists for an enrichment with different gene-sets known to be associated with ASD at the genetic and transcriptomic level. At the genetic level, this included the 102 rare and *de novo* protein truncating variants identified in the largest exome sequencing study of autism world-wide (17), and the GWAS-significant (i.e. common variants) ASD genes provided by Grove et al. (2019) (18). We also included an ASD-related gene list compiled by SFARI (SFARI.ASD.genes; categories S, 1, 2, & 3 downloaded in November 2020 from https://gene.sfari.org/). At the transcriptomic level, we included a list of differentially expressed genes (DEGs) (upregulated/downregulated) in post-mortem cortex tissue in ASD (19), and genes that are differentially expressed in specific cell types in ASD (20). Moreover, we included genes from differentially expressed co-expression modules in ASD that map onto specific biological processes (21)(22). Notably, these gene sets are partially overlapping, and the number of total and intersecting genes across sets are displayed in SF22. Our enrichment tests generated enrichment odds ratios, hypergeometric *p*-values, and FDR-corrected *p*-values using a background total of the 20,787 Neurosynth genes. We also conducted our analyses using a more conservative (i.e. restricted) list of 16,541 background genes based on real estimates of genes expressed in cortical tissue (19) (see SF21). Only comparisons with

p<.05 (FDR corrected) were interpreted further. For reasons of completeness, we also performed a cell-type enrichment analysis for the t-map, *Cohen's f*-map and χ^2-outlier map, as well as an enrichment analysis of ASD-related gene sets-based on the F-map representing the main effect of sensory subgroups. The results of these additional analyses are shown in SF23 and SF24 respectively.

Integration of brain transcriptome data of the developing human brain

In addition to testing for an enrichment of ASD-related genes, we also tested for an enrichment of genes underpinning typical brain development. Here, we used the human brain transcriptome dataset by Kang et al. (2011), which covers transcriptome profiles of 16 different brain regions within a timeframe ranging from embryonic development to late adulthood (23). Co-expression patterns of genes within this dataset have previously been allocated to a total of 29 co-regulated gene modules that are described in detail in (23). To further link our neuroimaging findings to specific genetic mechanisms, we therefore also tested for an enrichment of genes within the 29 co-regulated gene modules, using a background total of 18,675 genes contained within the modules (23). For this purpose, 2D heatmaps representing the time course of gene expression across different brain regions were created based on the module 'eigengene', as implemented in the MAGMA pipeline (24). All presented p-values are FDR corrected for multiple comparisons. Since each module has a distinct spatio-temporal pattern of expression, and represents different biological processes mediating brain development and ageing, we also compared the patterns of differences in CT highlighted by our neuroimaging analyses to the spatio-temporal expression patterns of the 29 modules. For this purpose, 2D heatmaps representing the time course of gene expression across different brain regions were created based on the module 'eigengene', as implemented in the MAGMA pipeline (24).

Genotyping

DNA isolated from blood or saliva were genotyped at 'Centre National de Recherche en Genomique Humaine' (CNRGH) in Paris using the Infinium OmniExpress-24v1 BeadChip (>700K markers) from Illumina. We excluded participants with a genotyping rate <95%, heterozygosity above or below 3SD from the mean, or a mismatch in reported and genetic sex. We further removed SNPs that deviated from Hardy-Weinberg Equilibrium ($p<1x10^{-6}$), and had a genotyping call rate below 95% using PLINK v1.9 (https://www.cog-genomics.org/plink2). Imputation of 17 Mio. SNPs was then performed using the 700k genotyped SNPs on the Michigan Imputation Server (25). Given that the majority of individuals in the sample were of European ancestry, we used the HRC r1.1 (2016) reference panel. A Principal Component Analysis (PCA) of variance standardized relationship matrix was used to evaluate the ancestry of individuals in the LEAP cohort, and to provide components for any covariate adjustments. Here, four genetic components (PC1 to PC4) reflecting population stratification were extracted and utilized in further analyses. To cluster individuals based on ancestry, we further reduced the dimensionality with uniform manifold approximation and projection (UMAP; (26)), reducing the first 8 PCA components to 2 components for better visualization and easier interpretation. Finally, to derive subpopulation clusters, we performed density-based clustering on these clusters (HDBSCAN; (27)), and selected individuals of European genetic ancestries. We only included unrelated individuals of European ancestries in our analyses with usable FreeSurfer reconstructions, leading to a total of N=501 individuals (N=279 ASD, N=222 TD controls).

Gene-Set analysis of quantitative ASD GWAS data in our sample

In order to test to what extent the gene sets identified by the GSEA for the 29 co-expression modules (23) are associated with the ASD genotype, we performed a competitive gene-set analysis using MAGMA v1.07 (28). For this purpose, we initially performed a quantitative genome-wide association study (GWAS) in the N=501 individuals with available MRI and genetic data (see above for genotypic information).

Only SNPs with a MAF>1% were selected. Logistic regression models predicting diagnostic category (i.e. ASD vs. TD) were applied with fixed effects for gender, age, iq, site, and the first four dimensions of the multidimensional scaling results (population stratification) from plinkv1.9. The analysis was implemented using the R package *lme4* (29). The resulting GWAS summary file was subsequently used for gene-set analysis. For gene-set analysis, variants were initially annotated to genes based on NCBI (37.3)-hg19 gene definitions, mapping variants to a gene if they were located in the transcription region of that gene, or within 5 kilobase upstream or 1.5 kilobase downstream of the transcription region. A gene-analysis was then performed to quantify the degree of association each gene has with the ASD phenotype (i.e. diagnosis of ASD vs. control). A detailed description of the approach can be found in (28). Briefly, gene analysis in MAGMA is based on a multiple linear principal component regression, that uses an F-test to compute gene p-values. This model first projects the SNP matrix for a gene onto its principal components (PCs), pruning away PCs with very small eigenvalues, and then utilizing only those PCs are predictors for the phenotype in the linear regression mode. By default, only 0.1% of the variance in the SNP data is pruned away (28). As the GWAS data already accounted for principal components related to ancestry, we did not include these as covariates in the gene analysis. The gene p-values and gene correlations (i.e. LD between genes), are then used to perform the (competitive) gene-set analysis, which tests within a regression framework whether the genes in a gene-set are more strongly associated with the phenotype of interest than other genes, correcting for potential confounds such as gene size, gene density, and potential differences in sample size. Here, we ran the MAGMA gene-set analysis with default parameters, testing the Kang et al. gene modules highlighted by the enrichment analysis described above. For each gene set, MAGMA provides p-values (uncorrected), and following FDR correction. Only the corrected p-value were interpreted.

Polygenic Scores

Gene set-based polygenic scores

To establish the link between the individual's genotype and neuroanatomical phenotype, we initially derived gene set-based polygenic scores (PGS_{set}) across gene sets that were significantly enriched in the χ^2-outlier map as resulting from the gene expression and decoding analysis. This resulted in a polygenic score (PGS_{DGE}) across gene sets M16.up (22), CTX.M9.up, CTX.M19.up, and CTX.M20.up (21), ASD.DEGs.up (19), and M2 and M4 (23), which are differentially expressed in ASD and/or typical brain development. To compare the impact of differentially expressed genes with the impact of ASD risk-genes on the individuals' *tAIs*, we also generated gene-set polygenic scores across the ASD risk-genes with rare and *de novo* variants (ASD.risk.DeNovo) provided by Satterstrom et al. (2020) (17), and the common variants (ASD.risk.common) identified by Grove et al. (2019) (18) ($PGS_{ASD.risk}$). All genes were initially converted to ENTREZ gene identifiers using the gprofiler2 package for R (https://biit.cs.ut.ee/gprofiler). Information on set sizes and gene overlap between sets is provided in SF 21. PGS_{set} were calculated using the PRSet function in PRSice-2 (https://www.prsice.info; (30)) using an additive model based on the independent SNPs present in both the training set (the GWAS summary file), and the testing data (the genotyped LEAP cohort). Here, the GWAS summary statistics was taken from the largest GWAS study to date with more than 18,381 individuals with ASD and 27,969 controls (18). The genes in the gene-sets of interest were matched to their genome boundaries according to human assembly GRCh37-hg19. For the LD based SNP pruning, we clumped only SNPs with a minor allele frequency (MAF) >1% using a window of 1 Mb. Since flanking SNPs not physically located with the gene-set region might also influence functions of the set, a cut-off (proxy threshold) of r^2=0.8 was used for gene-set membership (see https://www.prsice.info/prset_detail/ for details). The PGS_{set} were calculated at a *p*-value threshold of 1, because gene-set PGS containing a small portion of SNPs may be unrepresentative of the whole gene-set. After clumping, we retained 132,424 independent SNPs across the entire genome. Our target genome included 4,356,796 variants that matched a total of 6,645,928 variants included in the base file (i.e.

GWAS summary file). PGS_{set} were adjusted for principal component ancestry using genetic principal components (PC1 to PC4). For the PGS_{set}, a self-contained *p*-value and a competitive *p*-value were provided, which respectively tested the association with the target phenotype (i.e. $Phenotype \sim PRS_{set} + PC1 + ... + PC4 + error$), and the enrichment of signal of the specific gene-set (see ST4). The competitive *p*-value was obtained by comparing the observed gene-set PRS association with the 5k permuted null *p*-value distribution of random gene-sets PRSs.

Genome-wide polygenic scores

As the number of ASD risk genes with common and *de novo* variants was small overall, and only included a total of 1,455 SNPs in our sample (see ST 4), we also computed genome-wide PGS (PGS_{genome}) for ASD (18) and a variety of neuropsychiatric conditions including ADHD (31), schizophrenia (32), major depressive disorder (33), epilepsy (34), and neuroticism (35). Moreover, we computed PGS_{genome} for a variety of general phenotypic traits, such as Body Mass Index (BMI) (36), intelligence (37), educational attainment (38), insomnia (39), and subjective well-being (40). Here, for LD based SNP pruning, only SNPs with a MAF>1% and with an R^2<0.1 in windows of 500kb were selected. PGS_{genome} were also adjusted for principal component ancestry using PC1 to PC4. At the time of the study, PGS_{genome} were available for 434 individuals (202 controls, 231 ASD) in our sample.

Details on set size, the number of variants included, and statistical details for all PGS are provided in ST 4. All PGS were standardized (i.e. mean centered and scaled) prior to further analysis.

Further Analyses

Differences in CT in individuals with 'narrower' definition of ASD

In addition to examining the wider autism phenotype, we also explored between-group differences in a group of individuals with a 'narrower' definition of ASD. Here, the term 'narrow' refers to ASD individuals who, in addition to meeting a clinical diagnosis of ASD

according to DSM-5 or ICD-10 criteria, also met the diagnostic cut-offs for ASD on the ADOS and the ADI-R (in accordance with (41)). This yielded N=199 ASD individuals, which we compared to the N=279 neurotypical controls in terms of their CT. Overall, we found that, while the largest clusters with a significant between-group difference in CT remained stable across subsets, applying stricter inclusion criteria for ASD did not improve effect sizes as would be expected (SF12). Instead, several clusters in the frontal and midline brain structures were no longer significant following the exclusion of ASD individuals not meeting ADOS and ADI-R cut-off. Such a reduction in effect size might partially be related to the smaller sample size in the 'narrow' ASD subset (i.e. N=199 vs. N=360). However, our finding also implies that restricting the ASD group phenotypically to the more severe (i.e. narrower) cases may not necessarily improve the clinical and biological homogeneity within the sample.

Between-group differences in CT within age-stratified subgroups

To further investigate the effects of age on our results, we examined between-group differences in CT within age-stratified subgroups that included (i) children (aged 6-11 years; 60 ASD, 64 TD), (ii) adolescents (aged 12-17 years; 138 ASD, 97 TD), and (ii) adults (aged 18-30 years; 162 ASD, 119 TD) as published by (42). We find no significant differences in CT during early childhood, with frontal-lobe abnormalities emerging during adolescents (see SF13). This is of particular interest as the frontal lobes continue to mature until early adulthood (e.g. (43). During adulthood, CT differences were mostly observed in medial brain structures including anterior and posterior cingulate regions, and the right fusiform gyrus. While these results are well worth reporting, it should be noted, however, that the reduced size of effects within age bins might also be a reflection of the reduced sample size, with children representing the smallest subset (i.e. N=126 children, N=235 adolescents, N=280 adults). The findings thus suggest that even a sample size of N=126 may not be sufficient to reliably detect CT differences in CT.

CT differences across SRS subdomain scores

In addition to modelling group as a binary fixed-effect factor, we modelled ASD symptomatology in cases and controls based on the five (parent-rated) SRS-2 (3) subdomains of (i) social communication, (ii) social motivation, (iii) social awareness, (iv) social cognition, and (v) autistic mannerisms. More specifically, measures of CT were predicted for all individuals in our sample (i.e. cases and controls) using a GLM that included the five SRS subdomains while covarying for age, age^2, sex, site, and CT_0. We then examined the spatially distributed pattern of CT associated with each SRS subdomain. Overall, the results show that different SRS subdomains are linked to distinct, spatially distributed patterns of variability in CT, which together highlight a set of brain regions that is similar to brain regions associated with the main effect of group (e.g. superior temporal regions, dorsolateral and medial prefrontal regions, cingulate regions) (see SF19). Yet, few clusters survived multiple comparison correction; this may be due to the reduced sample size (149 TD, 290 ASD).

Analysis of neuroanatomical outliers

To establish the effects of neuroanatomical outliers on between-group differences in CT, we iteratively removed ASD individuals with the largest percentage of neuroanatomical outliers based on their *tAIs,* and re-computed the main effect of group. More specifically, we removed ASD individuals if they exceeded the 85th, 90th, and 95th upper percentage of the *tAIs* distribution, which corresponded to a value of 19.81%, 21.07%, and 13.54% of vertices within the ASD-outlier mask. Based on the percentiles of distribution, we excluded a total of N=54, N=36, and N=18 ASD individuals respectively. As can be seen in SF16, the main clusters of significant between-group differences remained stable when excluding the upper 5% of outliers in the ASD group, and started to decrease in effect size when excluding more than 10% of individuals in the ASD group, potentially as a direct result of a reduced sample size. Notably, right-hemisphere differences were more sensitive to the removal of outliers than left-hemisphere differences, which mostly remained stable even following the removal of more than 15% of outliers.

Furthermore, we explored the probability of being a neuroanatomical outliers resulting from the logistic regression between (i) individuals taking medication and those without, (ii) between males and females, and (iii) as a function of full-scale IQ and age (see SF18). There was no significant difference in the probability of being a neuroanatomical outlier between individuals taking medication and those without, or between males and females ($p>0.05$, two-tailed). However, the probability of being a neuroanatomical outlier decreased significantly with increasing full-scale IQ (FSIQ) ($r=-0.18$, $t(358)=-3.47$, $p<0.001$), and with increasing age ($r=-0.23$, $t(358)=-4.48$, $p<0.001$).

References

1. Lord C, Risi S, Lambrecht L, et al.: The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. J Autism Dev Disord 2000; 30:205–223

2. Lord C, Rutter M, Couteur AL: Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994; 24:659–685

3. JN C, CP G: Social Responsiveness Scale. 2nd ed. Lon Angeles, Western Psychological Services, 2012

4. D W: Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II). San Antonio, NCS Pearson, 2011

5. D W: Wechsler Intelligence Scale for Children–Third Edition. San Antonio, Psychological Corporation, 1991

6. D W: Wechsler Adult Intelligence Scale–Fourth Edition. San Antonio, Pearson, 2009

7. Bodfish JW, Symons FJ, Parker DE, et al.: Varieties of Repetitive Behavior in Autism: Comparisons to Mental Retardation. J Autism Dev Disord 2000; 30:237–243

8. Tomchek SD, Huebner RA, Dunn W: Patterns of sensory processing in children with an autism spectrum disorder. Res Autism Spect Dis 2014; 8:1214–1224

9. Dale AM, Fischl B, Sereno MI: Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction. Neuroimage 1999; 9:179–194

10. Wong AYS, Hsia Y, Chan EW, et al.: The Variation of Psychopharmacological Prescription Rates for People With Autism Spectrum Disorder (ASD) in 30 Countries. Autism Res 2014; 7:543–554

11. Freitag CM, Jensen K, Elsuni L, et al.: Group-based cognitive behavioural psychotherapy for children and adolescents with ASD: the randomized, multicentre, controlled SOSTA – net trial. J Child Psychol Psyc 2016; 57:596–605

12. Frazier TW, Shattuck PT, Narendorf SC, et al.: Prevalence and Correlates of Psychotropic Medication Use in Adolescents with an Autism Spectrum Disorder with and without Caregiver-Reported Attention-Deficit/Hyperactivity Disorder. J Child Adol Psychop 2011; 21:571–579

13. Radua J, Vieta E, Shinohara R, et al.: Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. Neuroimage 2020; 218:116956

14. Ecker C, Ginestet C, Feng Y, et al.: Brain Surface Anatomy in Adults With Autism: The Relationship Between Surface Area, Cortical Thickness, and Autistic Symptoms. Jama Psychiat 2013; 70:59–70

15. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al.: An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 2012; 489:391–399

16. Gorgolewski1 KJ, Fox AS, Chang L, et al.: Tight fitting genes: finding relations between statistical maps and gene expression patterns [Internet]Hamburg, Germany, 2014Available from: https://f1000research.com/posters/1097120

17. Satterstrom FK, Kosmicki JA, Wang J, et al.: Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 2020; 180:568-584.e23

18. Grove J, Ripke S, Als TD, et al.: Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 2019; 51:431–444

19. Gandal MJ, Zhang P, Hadjimichael E, et al.: Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 2018; 362:eaat8127

20. Velmeshev D, Schirmer L, Jung D, et al.: Single-cell genomics identifies cell type-specific molecular changes in autism. Science 2019; 364:685–689

21. Parikshak NN, Swarup V, Belgard TG, et al.: Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature 2016; 540:423–427

22. Voineagu I, Wang X, Johnston P, et al.: Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 2011; 474:380–384

23. Kang HJ, Kawasawa YI, Cheng F, et al.: Spatio-temporal transcriptome of the human brain. Nature 2011; 478:483–489

24. Yousaf A, Duketis E, Jarczok T, et al.: Mapping the genetics of neuropsychological traits to the molecular network of the human brain using a data integrative approach. Biorxiv 2018; 336776

25. Das S, Forer L, Schönherr S, et al.: Next-generation genotype imputation service and methods. Nat Genet 2016; 48:1284–1287

26. McInnes L, Healy J, Melville J: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Arxiv 2018;

27. Campello RJGB, Moulavi D, Sander J: Advances in Knowledge Discovery and Data Mining, 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II2013; 160–172

28. Leeuw CA de, Mooij JM, Heskes T, et al.: MAGMA: Generalized Gene-Set Analysis of GWAS Data. Plos Comput Biol 2015; 11:e1004219

29. Bates D, Mächler M, Bolker B, et al.: Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015; 67

30. Choi SW, Mak TS-H, O'Reilly PF: Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc 2020; 15:2759–2772

31. Demontis D, Walters RK, Martin J, et al.: Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 2019; 51:63–75

32. Ripke S, Neale BM, Corvin A, et al.: Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014; 511:421–427

33. Howard DM, Adams MJ, Clarke T-K, et al.: Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 2019; 22:343–352

34. Epilepsies ILAEC on C: Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurology 2014; 13:893–903

35. Nagel M, Jansen PR, Stringer S, et al.: Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet 2018; 50:920–927

36. Graff M, Scott RA, Justice AE, et al.: Genome-wide physical activity interactions in adiposity — A meta-analysis of 200,452 adults. Plos Genet 2017; 13:e1006528

37. Savage JE, Jansen PR, Stringer S, et al.: Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet 2018; 50:912–919

38. Okbay A, Beauchamp JP, Fontana MA, et al.: Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016; 533:539–542

39. Jansen PR, Watanabe K, Stringer S, et al.: Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet 2019; 51:394–403

40. Study LC, Okbay A, Baselmans BML, et al.: Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet 2016; 48:624–633

41. RISI S, LORD C, GOTHAM K, et al.: Combining Information From Multiple Sources in the Diagnosis of Autism Spectrum Disorders. J Am Acad Child Adolesc Psychiatry 2006; 45:1094–1103

42. Charman T, Loth E, Tillmann J, et al.: The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. Mol Autism 2017; 8:27

43. Gogtay N, Giedd JN, Lusk L, et al.: Dynamic mapping of human cortical development during childhood through early adulthood. P Natl Acad Sci Usa 2004; 101:8174–8179

44. Polioudakis D, Torre-Ubieta L de la, Langerman J, et al.: A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron 2019; 103:785–801.e8

Supplementary Tables

TABLE S1. Summary of known types of medication taken by participants in the LEAP sample

	ASD	TD
N	122	20
Antidepressants	34	6
SSRIs	29	5
Tetracyclic (TeCA)	2	1
Tricyclic (TCA)	3	0
Antiepileptics	11	2
Antimigraine preparations	4	0
Antipsychotics	28	0
Aripiprazole	6	0
Clozapine	1	0
Pipamperone	2	0
Quetiapine	1	0
Risperidone	18	1
Anxiolytics	2	1
Drugs used in Addictive Disorder	0	1
Hypnotics & Sedatives	40	2
Hyoscine butylbromide	1	0
Melatonin	38	2
Niaprazine	1	0
Other Analgesics & Antipyretics	4	4
Opioids	1	0
Others	3	4
Psychostimulants & Other drugs used to treat ADHD	47	9
Atomoxetine	3	2
Dexamfetamine	1	0
Methylphenidate hydrochloride	43	7

Note. Participants may have taken up to 3 different types of medication across the listed categories during study participation. Data based on the ATC classification system, Level-1 ATC code "N")

TABLE S2. MRI Acquisition Parameters across sites

Site	Manufacturer	Model	Software Version	Acquisition sequence	Coverage	Slices	Thickness [mm]	Resolution [mm^3]	TR [s]	TE [ms]	FA [°]	FOV
Cambridge	Siemens	Verio	Syngo MR B17	Tfl3d1_ns	256*256	176	1.2	1.1*1.1*1.2	2.3	2.95	9	270
London	GE Medical systems	Discovery mr750	LX MR DV23.1_V02_1317.c	SAG ADNI GO ACC SPGR	256*256	196	1.2	1.1*1.1*1.2	7.31	3.02	11	270
Mannheim	Siemens	TimTrio	Syngo MR B17	MPRAGE ADNI	256*256	176	1.2	1.1*1.1*1.2	2.3	2.93	9	270
Nijmegen	Siemens	Skyra	Syngo MRD13	Tfl3d1_16ns	256*256	176	1.2	1.1*1.1*1.2	2.3	2.93	9	270
Rome	GE Medical systems	Signa HDxt	24/LX/MR HD16.0_V02_1131.a	SAG ADNI GO ACC SPGR	256*256	172	1.2	1.1*1.1*1.2	5.96	1.76	11	270
Utrecht	Philips Medical Systems	Achieva/Ingenia CX	3.2.3, 3.2.3.1	ADNI GO 2	256*256	170	1.2	1.1*1.1*1.2	6.76	3.1	9	270

Note. TR: repetition time; TE: echo time; FA: flip angle; FOV: field of view

TABLE S3. Vertex-wise differences in CT for the main effect of group

Cluster	Regional Labels	Side	BA	Vertices	t_{max}	*p*	Talairach coordinates x y z
ASD > TD controls							
1	rostral anterior cingulate cortex	L	24/33	1153	4.14	3.16E-06	-5 30 8
2	superior temporal gyrus, inferior temporal gyrus, middle temporal gyrus	L	20/21/22/38/41/42	6555	3.54	1.12E-04	-56 -14 -4
3	rostral anterior cingulate cortex	R	24/33	239	3.55	2.31E-04	5 24 -6
4	superior temporal gyrus, banks superior temporal sulcus, supramarginal gyrus, middle temporal gyrus	R	20/21/22/38/41/42	5280	3.78	1.82E-03	63 -35 13
5	lingual gyrus	L	18/19	877	3.01	1.25E-02	-15 -38 -2
6	precuneus cortex, posterior-cingulate cortex, isthmus-cingulate cortex	R	23/31	2674	2.70	1.58E-02	12 -43 31
ASD < TD controls							
7	superior frontal gyrus, caudal middle frontal gyrus, rostral middle frontal gyrus, precentral gyrus	L	4/6/8/9	6334	-4.23	5.23E-05	-31 27 35
8	parahippocampal gyrus, fusiform gyrus	R	18/19/34/37	1128	-3.04	2.22E-03	37 -38 -9
9	temporal pole	L	20/38	503	-3.06	4.19E-03	-25 -3 -28
10	precentral gyrus, postcentral gyrus	R	4/6	4825	-2.80	5.56E-03	38 -9 41

Note. R, right; L, left; BA, approximate Brodmann area(s); Vertices, number of vertices within the cluster; t_{max}, maximum *t*-statistic within cluster; *F*, *F*-statistic within cluster; p, cluster-corrected *p*-value.

TABLE S4. Information on gene set-based and genome-wide polygenic scores

Set ID	Set Name	Publication	N (genes)	N (SNPs)	R^2	p-value	
Gene set-based PGS							*p*-value competitive
1	M16.up	Voineagu (2011)	386	2038	0.0041	0.21	0.21
2	CTX.M9.up	Parikshak (2016)	506	3658	0.0001	0.84	0.84
3	CTX.M19.up	Parikshak (2016)	274	1577	0.0034	0.25	0.25
4	CTX.M20.up	Parikshak (2016)	330	2479	0.0028	0.29	0.28
5	ASD.DEGs.up	Gandal (2018)	696	4197	0.0125	0.02*	0.02
6	M2	Kang (2011)	2744	17993	0.0020	0.38	0.32
7	M4	Kang (2011)	44	238	0.0047	0.17	0.18
8	PGS_{DGE}	*Set ID1 to ID7*	3940	24551	0.0003	0.71	0.66
10	ASD.risk.DeNovo	Satterstrom (2020)	102	1206	0.0025	0.32	0.33
9	ASD.risk.common	Grove (2019)	31	252	0.0005	0.64	0.65
11	$PGS_{ASD.risk}$	*Set ID9 to ID10*	134	1455	0.0008	0.56	0.55
Genome—wide PGS							*p*-value GWAS threshold
12	ASD	Grove (2019)	-	29591	0.0094	0.027	0.04
13	ADHD	Demontis (2019)	-	71033	0.012	0.012	0.17
14	MDD	Howard (2019)	-	24435	0.0295	0.0001	0.02
15	SCZ	Ripke (2014)	-	10848	005	0.1059	0.01
16	Neuroticism	Nagel (2018)	-	17284	0.0262	0.0002	0.03
17	Insomnia	Jansen (2019)	-	62002	0.0028	0.225	0.39
18	Epilepsy	ILAE (2014)	-	59274	0.0055	0.091	0.20
19	SWB	Okbay (2016)	-	4190	0.0019	0.3249	0.01
20	Intelligence	Savage (2018)	-	172222	0.0012	0.4223	0.35
21	BMI	Graff (2017)	-	29797	0.0037	0.166	0.08
22	YearsEdu	Okbay (2016)	-	14468	0.0038	0.159	0.01

Note. N (genes): number of genes per set with ENTREZ identifiers, N (SNPs): number of variants per set, R^2: Variance explained by the PRS, *p*-value: self-contained p-value, p-value competitive: p-value resulting from the permutation procedure, p-value GWAS threshold, ADHD: Attention Deficit Hyperactivity Disorder, MDD: Major Depressive Disorder, SCZ: schizophrenia, SWB: subjective well-being, BMI: Body Mass Index, YearsEdu: years in education, ILAE: International League against Epilepsies

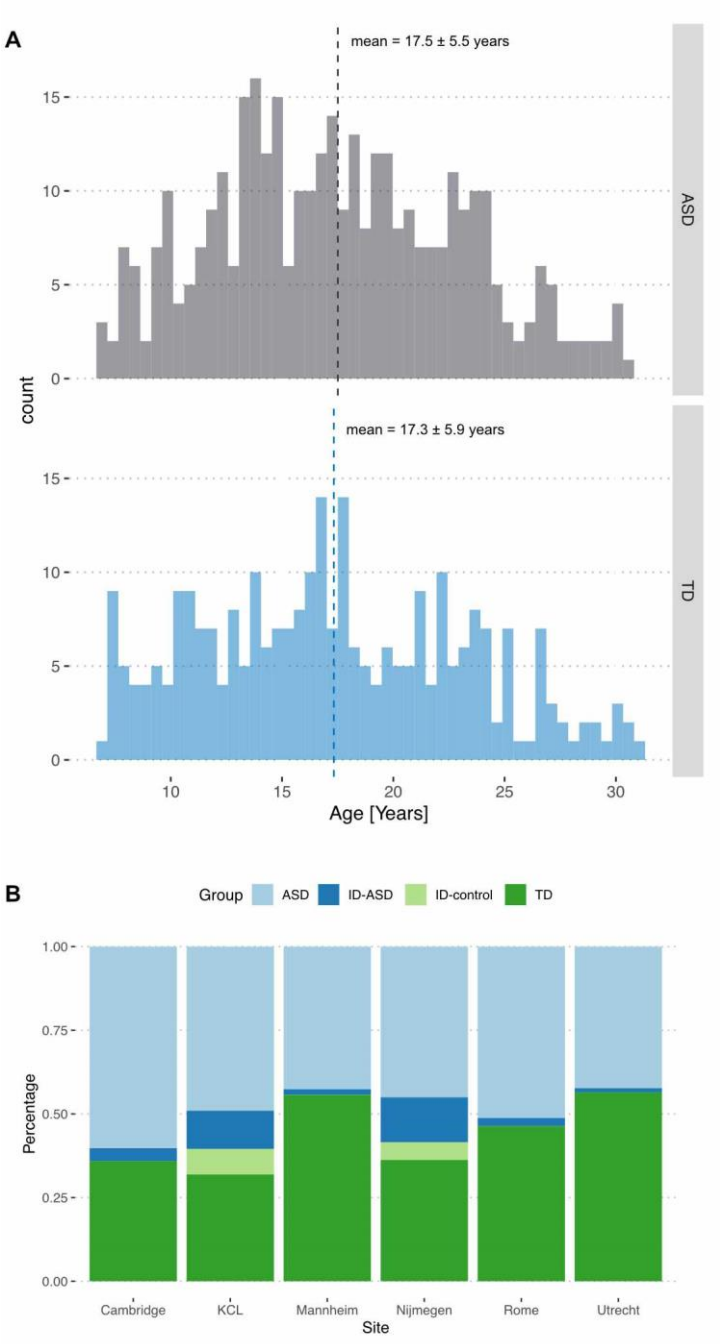
TABLE S5. Short Sensory Profile (8) total and subdomain scores within the ASD group

Short Sensory Profile	ASD (N=264)			
	M	SD	Min	Max
Tactile sensitivity (TAC)	26.77	6.16	9.00	35.00
Taste / smell sensitivity (TSM)	15.23	4.98	4.00	20.00
Movement sensitivity (MOV)	12.19	3.35	3.00	15.00
Under-responsiveness / seeks sensation (USS)	26.71	6.38	9.00	35.00
Auditory filtering (AFL)	17.44	5.34	6.00	30.00
Low energy / weak (LEW)	23.43	7.01	6.00	30.00
Visual / auditory sensitivity (VAS)	18.49	4.98	5.00	25.00
Total score	139.61	27.31	53.00	190.00

Note. ASD: autism spectrum disorder, N: sample size, M: mean, SD: standard deviation; Min: minimum; Max: maximum

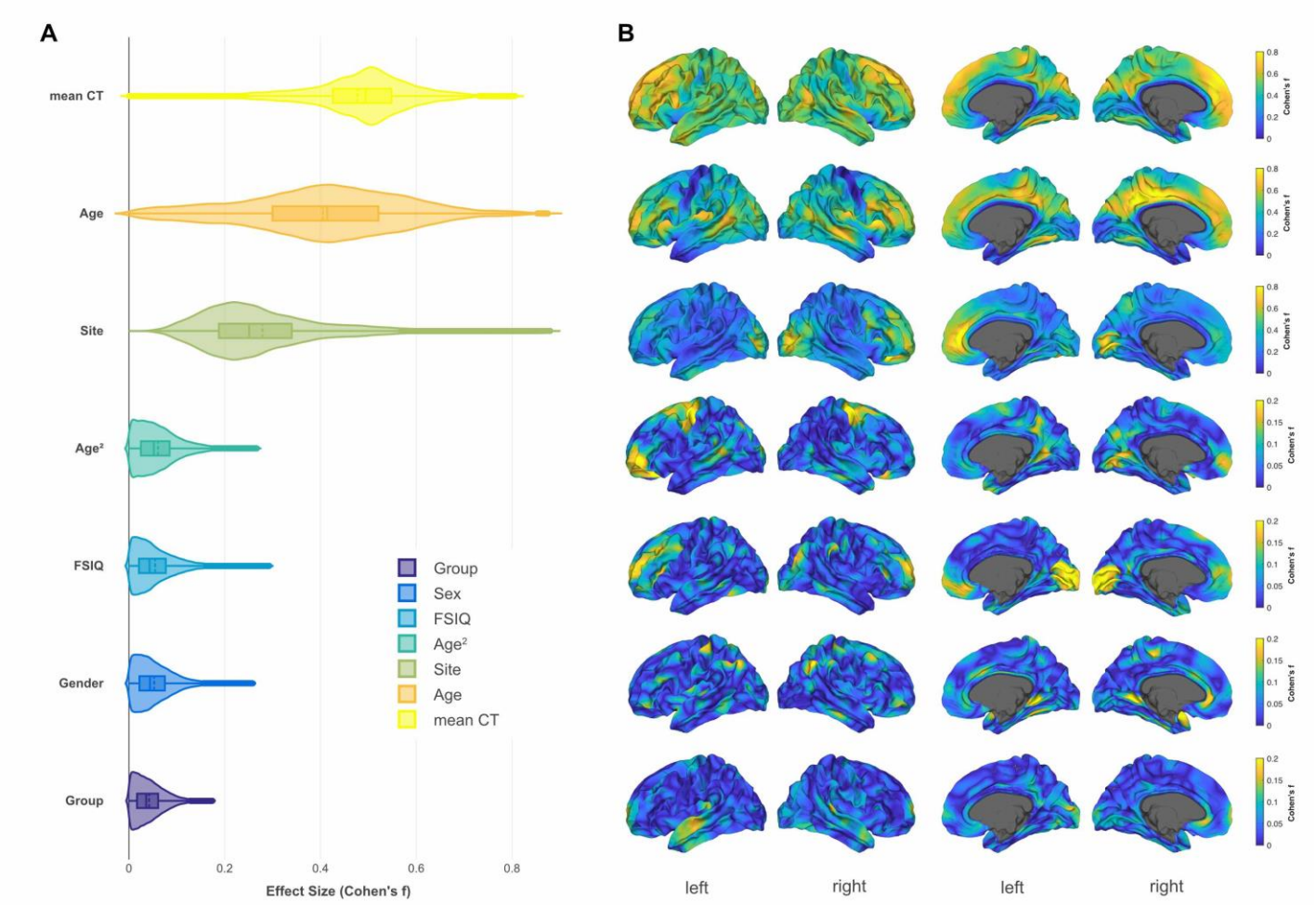
Supplementary Figures

FIGURE S1. Distribution of age and FSIQ across groups and sites



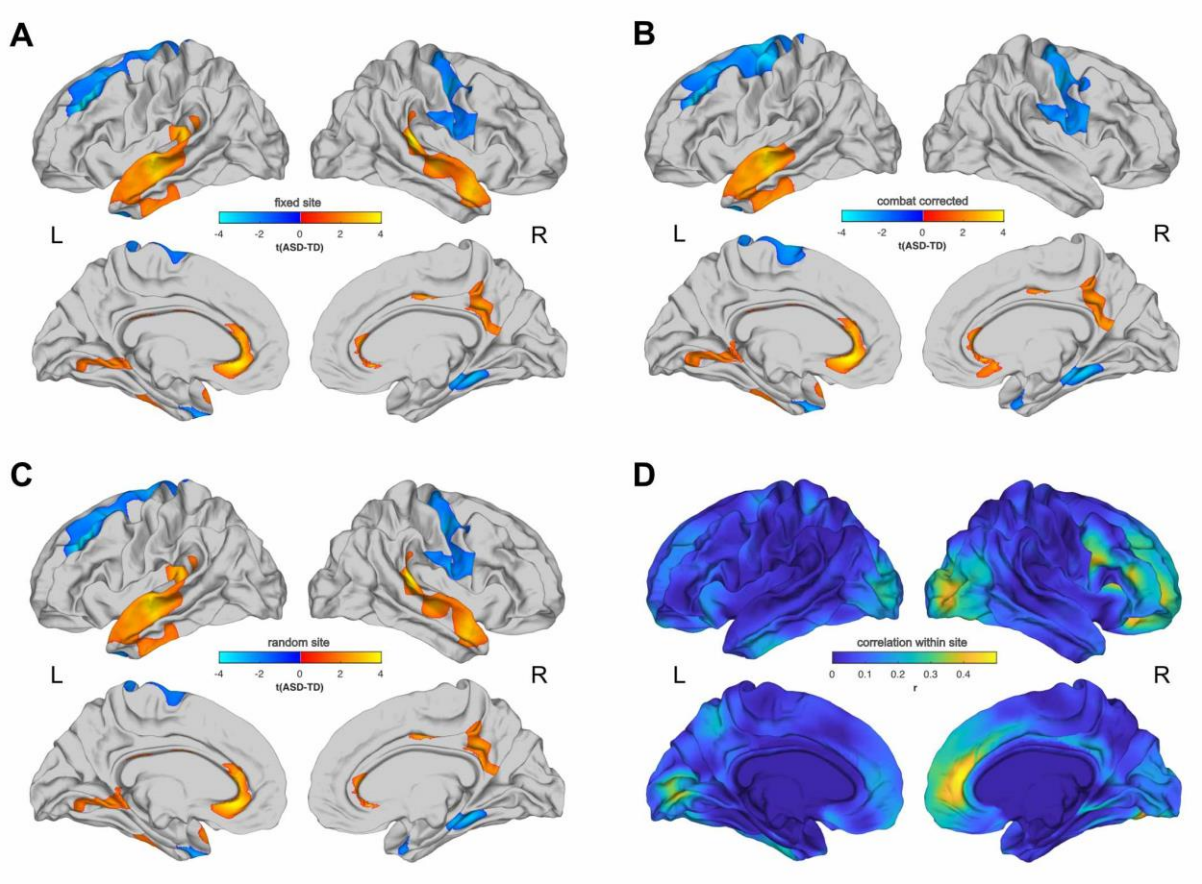
(**A**) Distribution of age in the group of ASD individuals and neurotypical controls. (**B**) Proportion of individuals with and without a mild intellectual disability (ID) in the ASD and non-ASD group across sites.

FIGURE S2. Effect Sizes for individual model terms



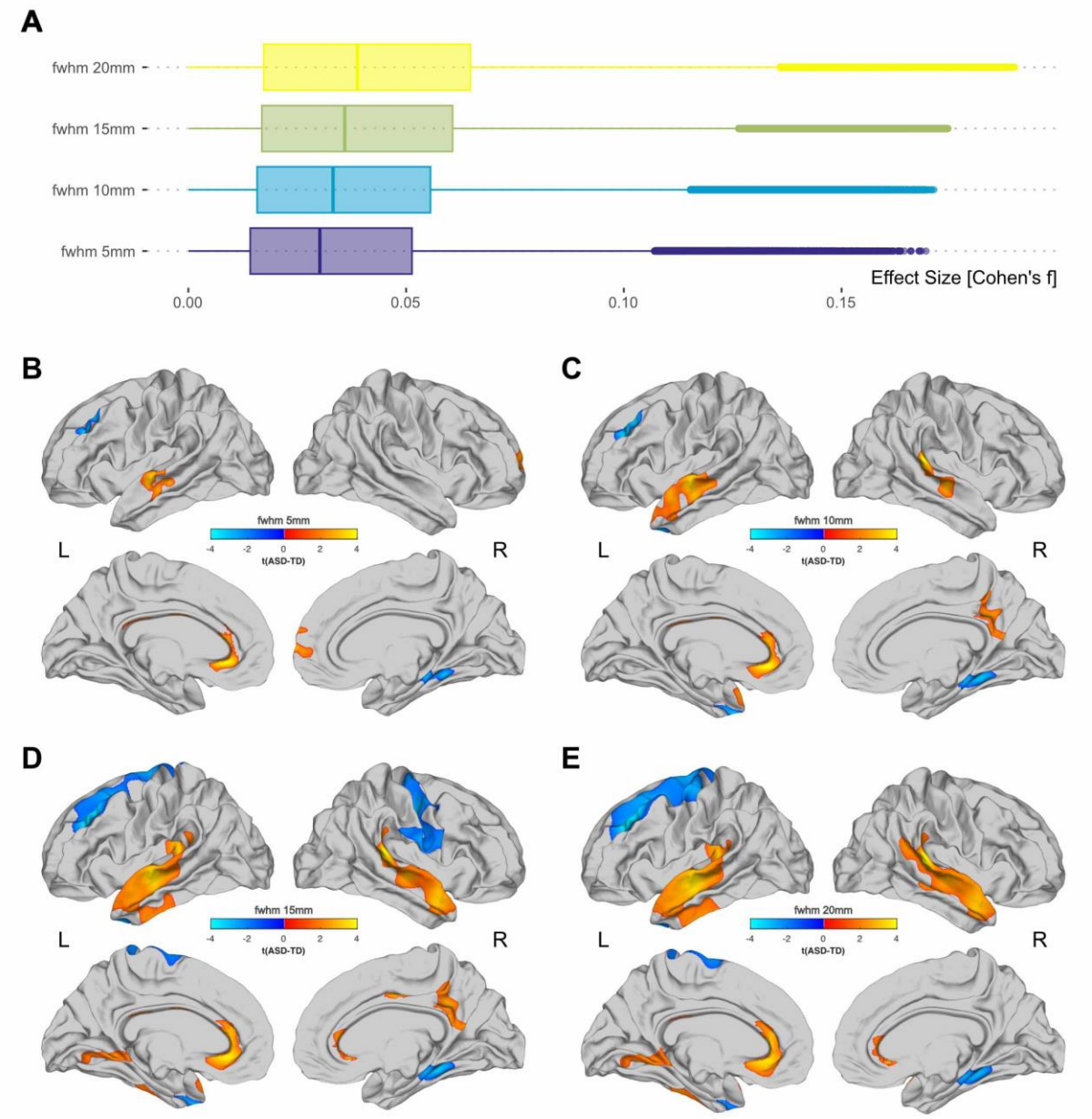
Effect Sizes (Cohen's *f*) for all components of the general linear model. (**A**) Distribution of effects sizes across all vertices on the cortical surface. (**B**) Spatially-distributed patterns of effects of each model term on the cortical surface.

FIGURE S3. Effects of acquisition site



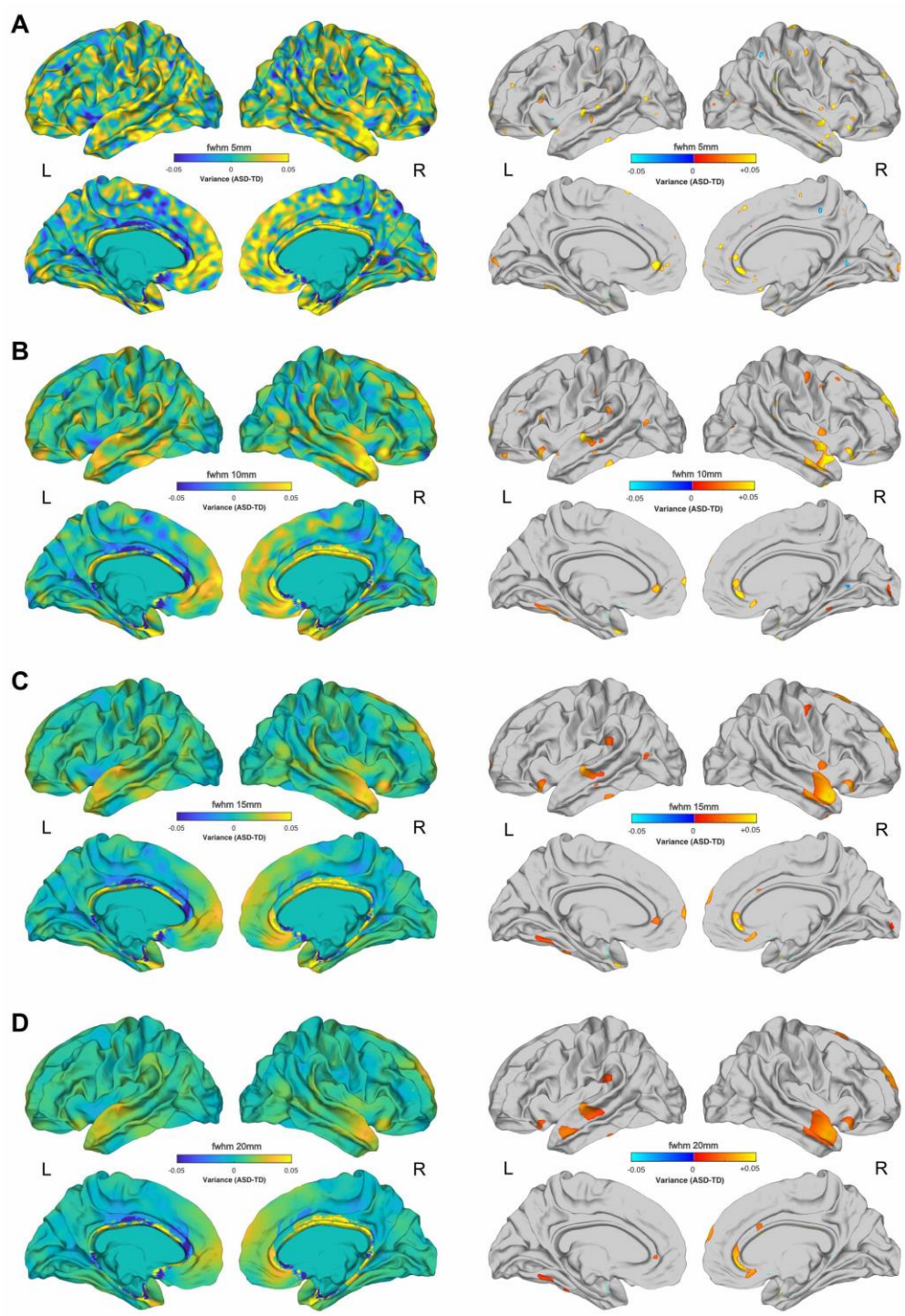
Treatment of site effects. (A) Main effect of group treating site as a fixed effect variable. (B) Main effect of group following ComBat correction for site effects. (C) Main effect of group treating site as a random effect. (D) Correlations within site based on mixed-effects model that treats site as random effect. (A) to (C) displays clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD relative to controls (RFT-based cluster corrected, $p<0.05$, *two-tailed*). L: left hemisphere, R: right hemisphere

FIGURE S4. Effects of smoothing on main effect of group



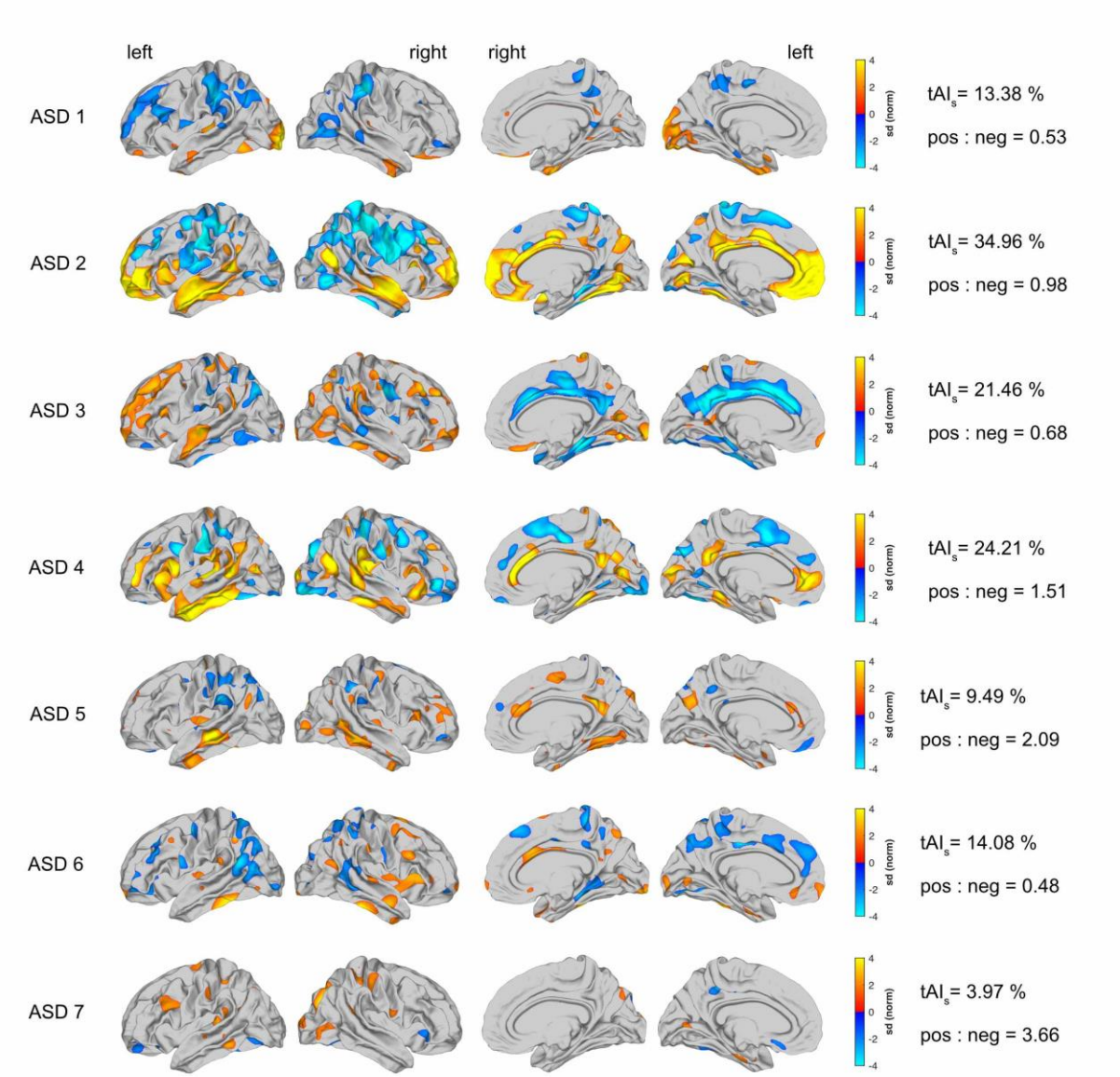
Effect of smoothing kernel on the main effect of group. (A) Vertex-wise effect sizes (Cohen's *f*) associated with the main effect of group across vertices on the cortical surface given a surface-based smoothing filter with a full width at half maximum (FWHM) of 5 mm, 10 mm, 15 mm, and 20 mm. Clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD relative to controls (RFT-based cluster corrected, $p<0.05$, *two-tailed*) given a smoothing kernel of 5 mm (B), 10 mm (C), 15 mm (D), and 20 mm (E). L: left hemisphere, R: right hemisphere

FIGURE S5. Effects of smoothing kernel on Levene's test



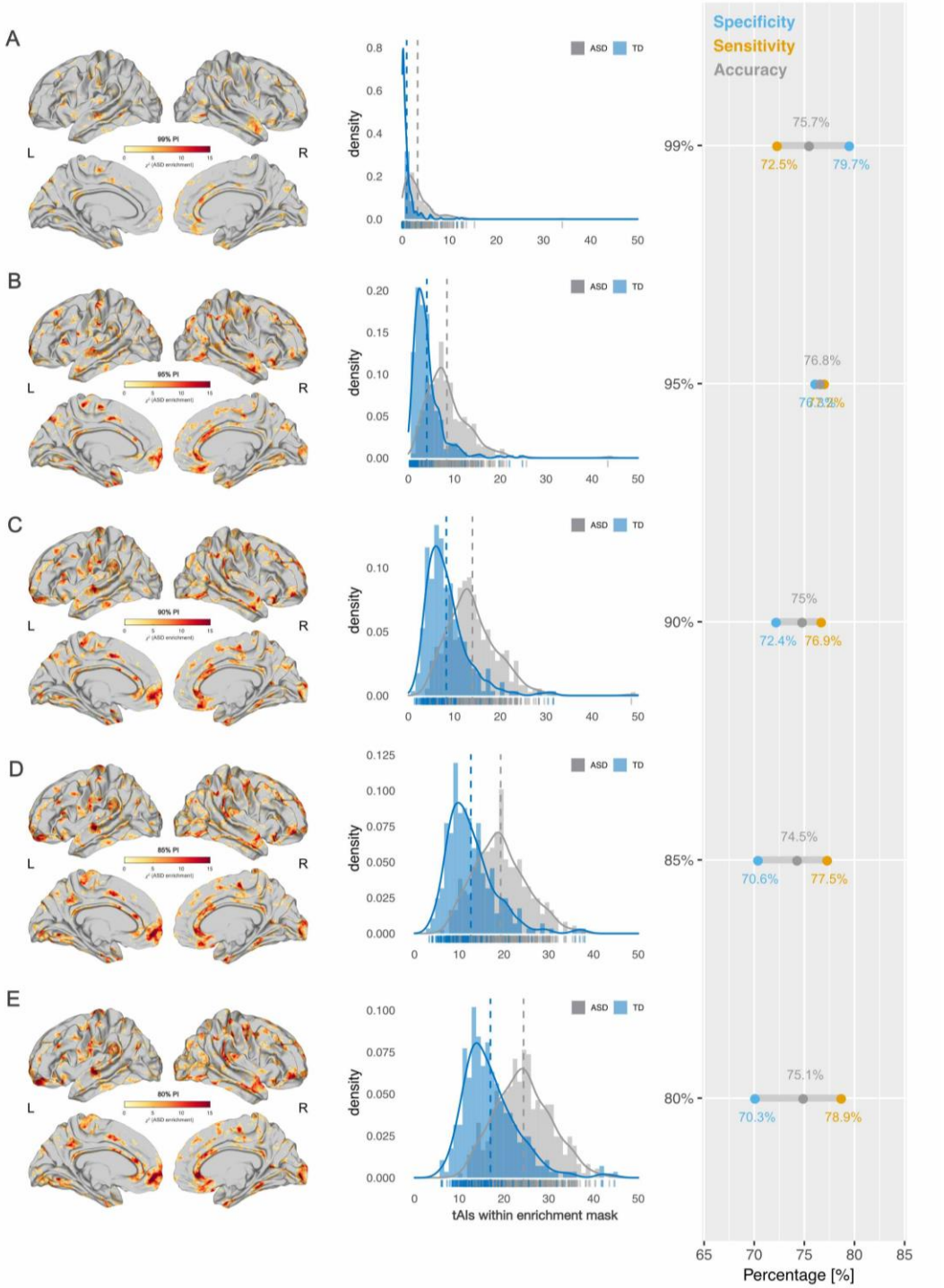
Levene's test for homogeneity of variances in cortical thickness between ASD individuals and controls at different smoothing kernels. The left panel shows the difference in CT variance between ASD individuals and controls using a surface-based smoothing filter with a full width at half maximum (FWHM) of 5 mm (**A**), 10 mm (**B**), 15 mm (**C**), and 20 mm (**D**). The right panel shows vertices with significantly increased variability (red to orange) in the ASD group relative to controls (FDR-corrected, $p<0.05$). L: left hemisphere, R: right hemisphere

FIGURE S6. Interindividual differences in vertex-level deviations



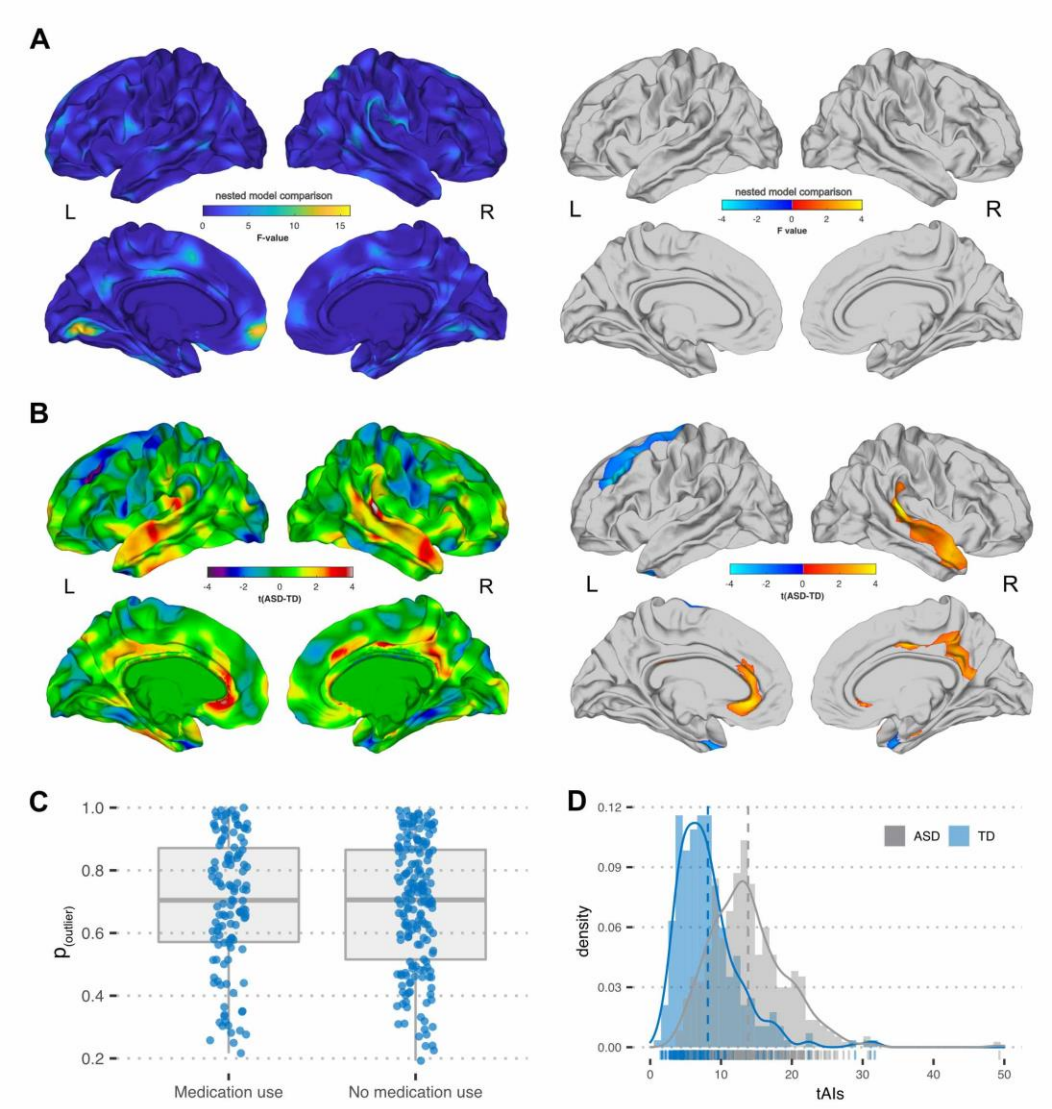
Interindividual differences in vertex-level deviations from the neurotypical range of cortical thickness in seven randomly selected ASD individuals. Colorbars indicate positive (red to yellow) and negative (blue to cyan) deflections in unit standard deviations of the neurotypical distributions. Standardized deviations are thresholded based on whether a value falls inside (grey) or outside the neurotypical 90% Prediction Interval ($PI_{90\%}$) at a given vertex. tAI_s: total degree of neuroanatomical abnormality assessed as the percentage of vertices falling outside the neurotypical $PI_{90\%}$ across the cortex. pos:neg: ratio of positive to negative deviations.

FIGURE S7. Effects of prediction interval thresholds



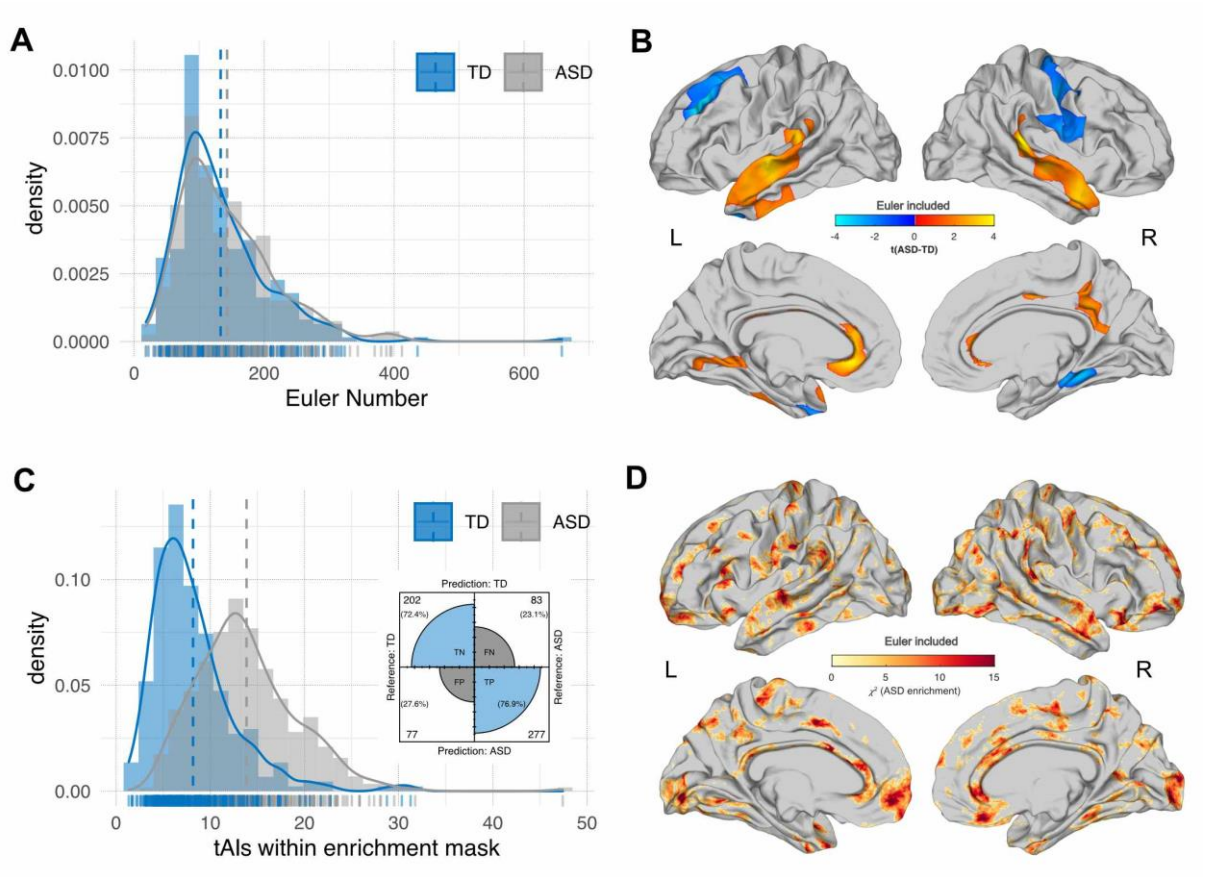
Effects of different Prediction Intervals (PIs). The left panel shows vertices with a significant enrichment of ASD individuals outside the neurotypical PI at different cut-offs of 99% (A), 95% (B), 90% (C), (85%) (D), and 80% (E). The resulting distribution of *tAIs* for each group are shown in the middle panel, and their model characteristics (i.e., accuracy, sensitivity, specificity) are shown in the right panel. L: left hemisphere, R: right hemisphere

FIGURE S8. Effects of medication status on variability in CT



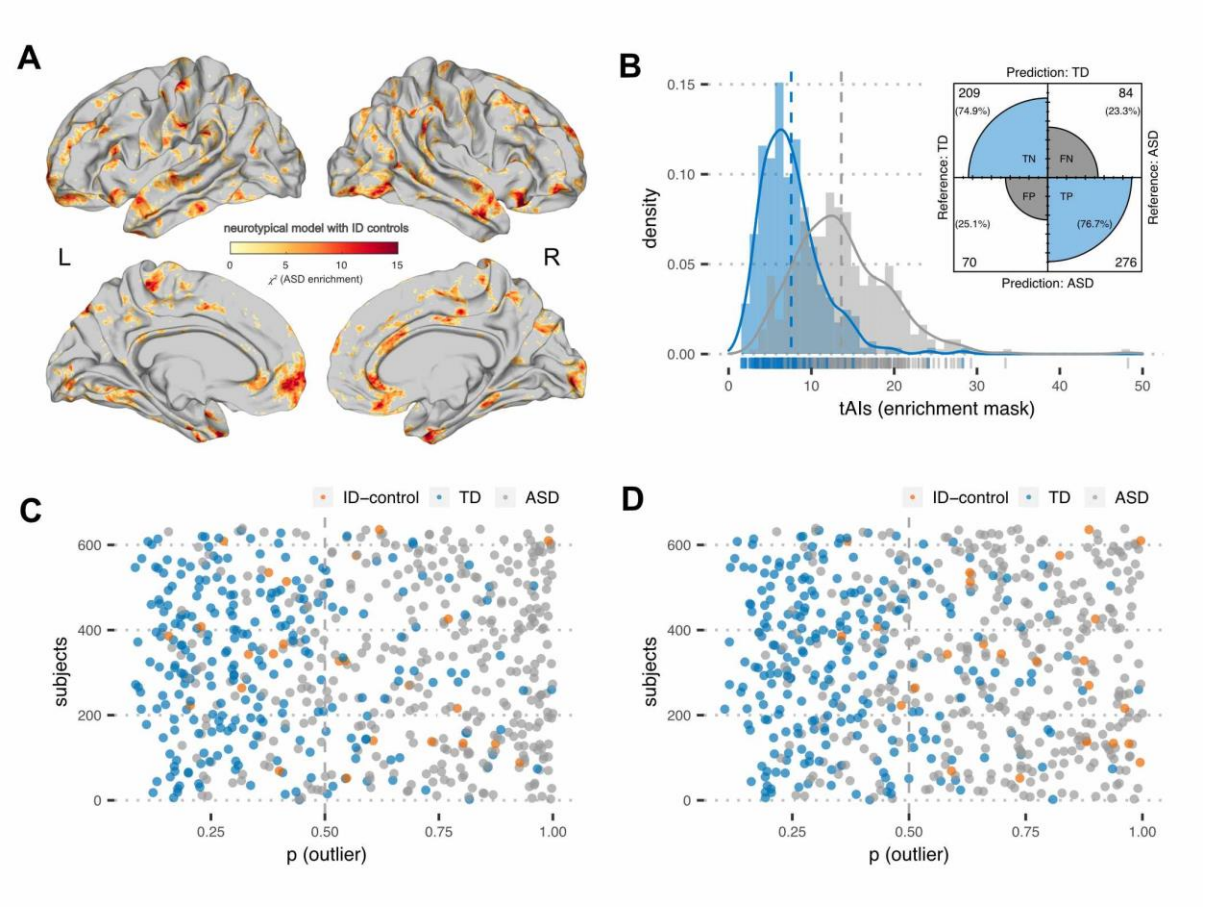
Effect of medication status on variability in CT. (A) *F-statistic* (un-thresholded) resulting from the nested-model comparison for the main effect of medication (left panel) and corrected for multiple comparisons (RFT-based cluster corrected, $p<0.05$, *two-tailed*) (right panel). (B) Main effect of group when including known medication status as covariate in the GLM. The left panel shows *t-test* statistic for the contrast ASD minus control (un-thresholded). The right panel shows clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*). L: left hemisphere, R: right hemisphere. (C) Probability of being a neuroanatomical outlier (*p(outlier)*) for ASD individuals based on the individual's total degree of neuroanatomical abnormality (i.e. *tAIs*) in cortical thickness within the ASD enrichment mask as a function of medication use (non-significant at $p<0.05$). (D) Distribution of *tAIs* within the ASD enrichment mask based on a predictive model that included medication status as predictor.

FIGURE S9. Effects of FreeSurfer surface reconstruction quality



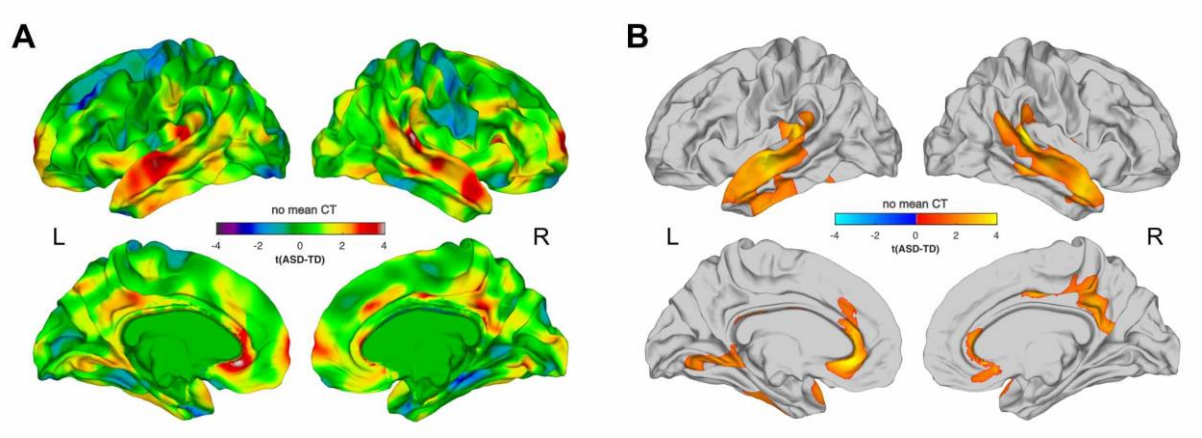
Influence of FreeSurfer reconstruction quality on between-group differences in CT and neuroanatomical outliers. (A) Between-group differences in the Euler Number of the reconstructed cortical surfaces summarized across hemispheres. (B) Main effect of group when including the Euler Number as covariate in the GLM. The figure shows clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*). (C) Distribution of tAIs within groups and prediction accuracies based on a neurotypical model that includes the Euler Number as covariate. (D) Vertices with a significant enrichment of ASD individuals outside the neurotypical 90% Prediction Interval (i.e. ASD enrichment mask) based on neurotypical model that includes the Euler Number as predictor. L: left hemisphere, R: right hemisphere.

FIGURE S10. Neurotypical model including controls with mild ID



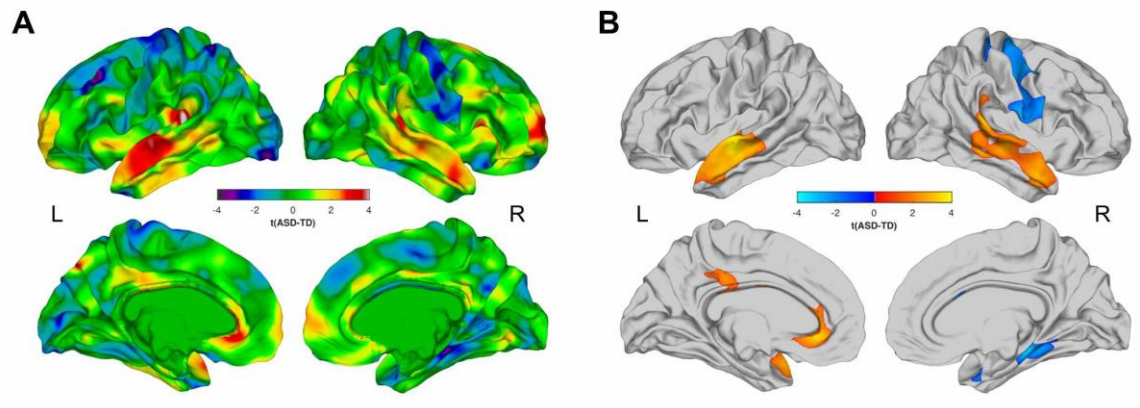
Neurotypical model including controls with mild intellectual disability (ID). (A) Vertices with a significant enrichment of ASD outside the neurotypical $PI_{90\%}$ when including neurotypical individuals with a mild ID. (B) Resulting distribution of tAIs and model accuracy. Probability of being a neuroanatomical outlier for ASD and non-ASD individuals with a mild ID based on a neurotypical model including ID controls (C), and based on a neurotypical model not including ID controls (D).

FIGURE S11. Between-group in CT without covarying for mean CT



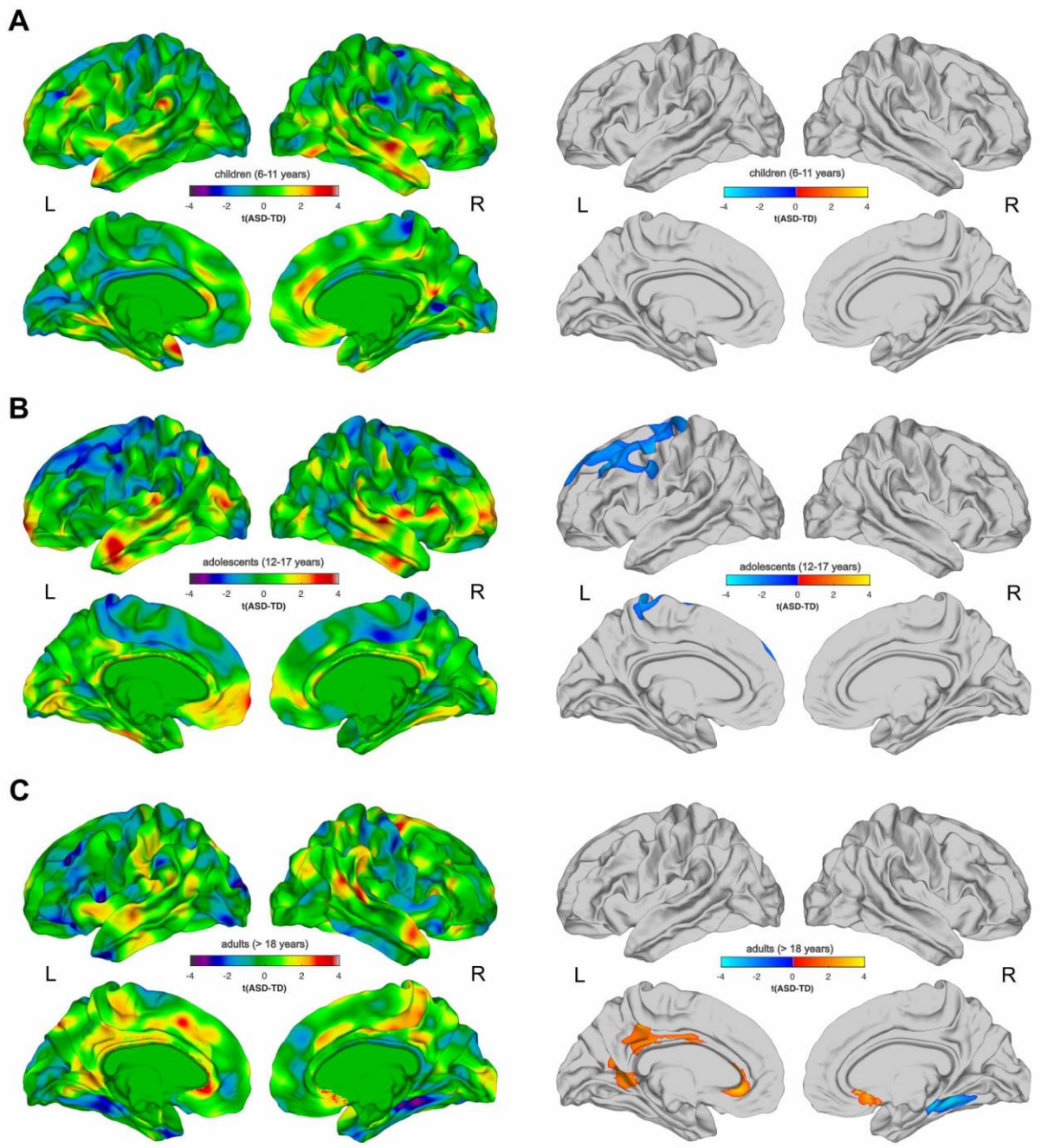
Vertex-wise differences in cortical thickness in ASD individuals without covarying for mean cortical thickness across the cortex. (A) *t-test* statistic for the contrast ASD minus control (un-thresholded). (B) Clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*). L: left hemisphere, R: right hemisphere

FIGURE S12. CT differences in 'narrow' ASD individuals



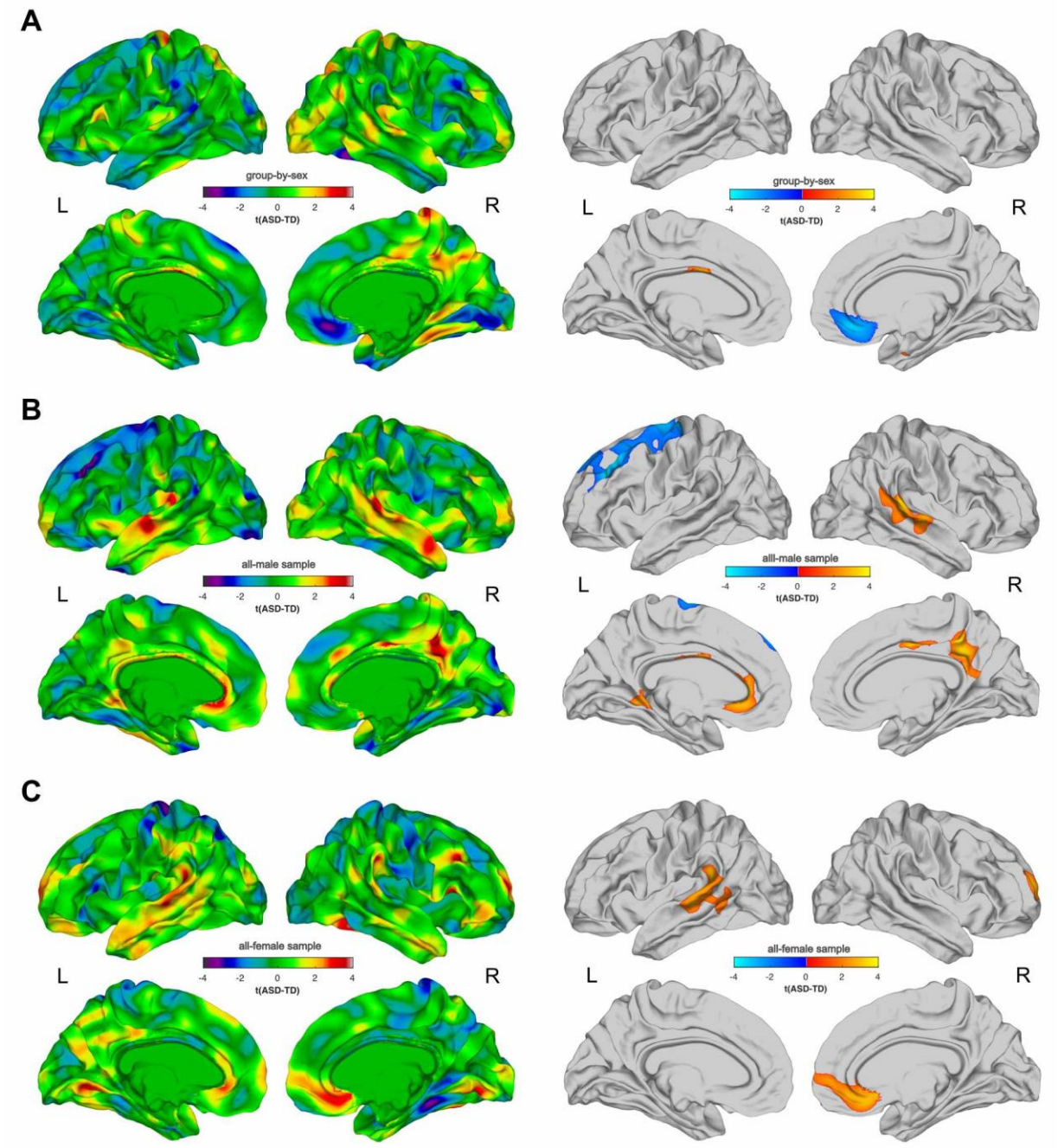
Vertex-wise differences in cortical thickness in individuals meeting gold-standard diagnostic criteria for ASD relative to controls as defined by (41). (A) T-*test* statistic for the contrast ASD minus control (un-thresholded); L: left hemisphere, R: right hemisphere. (B) Clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*). The group comparison was based on N=199 ASD individuals, which were compared to N=279 neurotypical controls. L: left hemisphere, R: right hemisphere

FIGURE S13. Between-group differences in CT within age groups



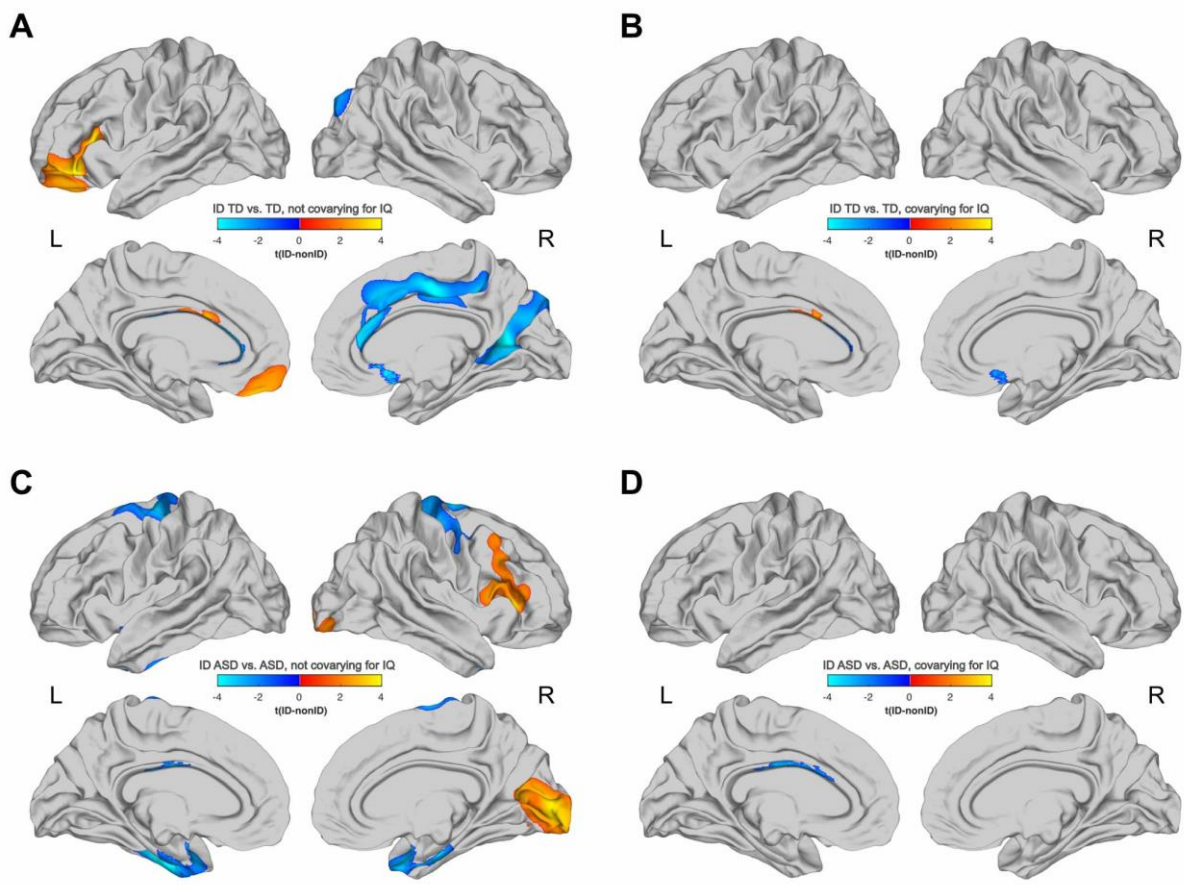
Differences in cortical thickness between ASD individuals and neurotypical controls in (A) children aged 6 to 11 years, (B) adolescents aged 12 to 17 years, and in (C) adults older than 18 years. The left panel shows *t-test* statistic for the contrast ASD minus control (un-thresholded); L: left hemisphere, R: right hemisphere. The right panel shows clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*).

FIGURE S14. Between-group CT differences in males and females



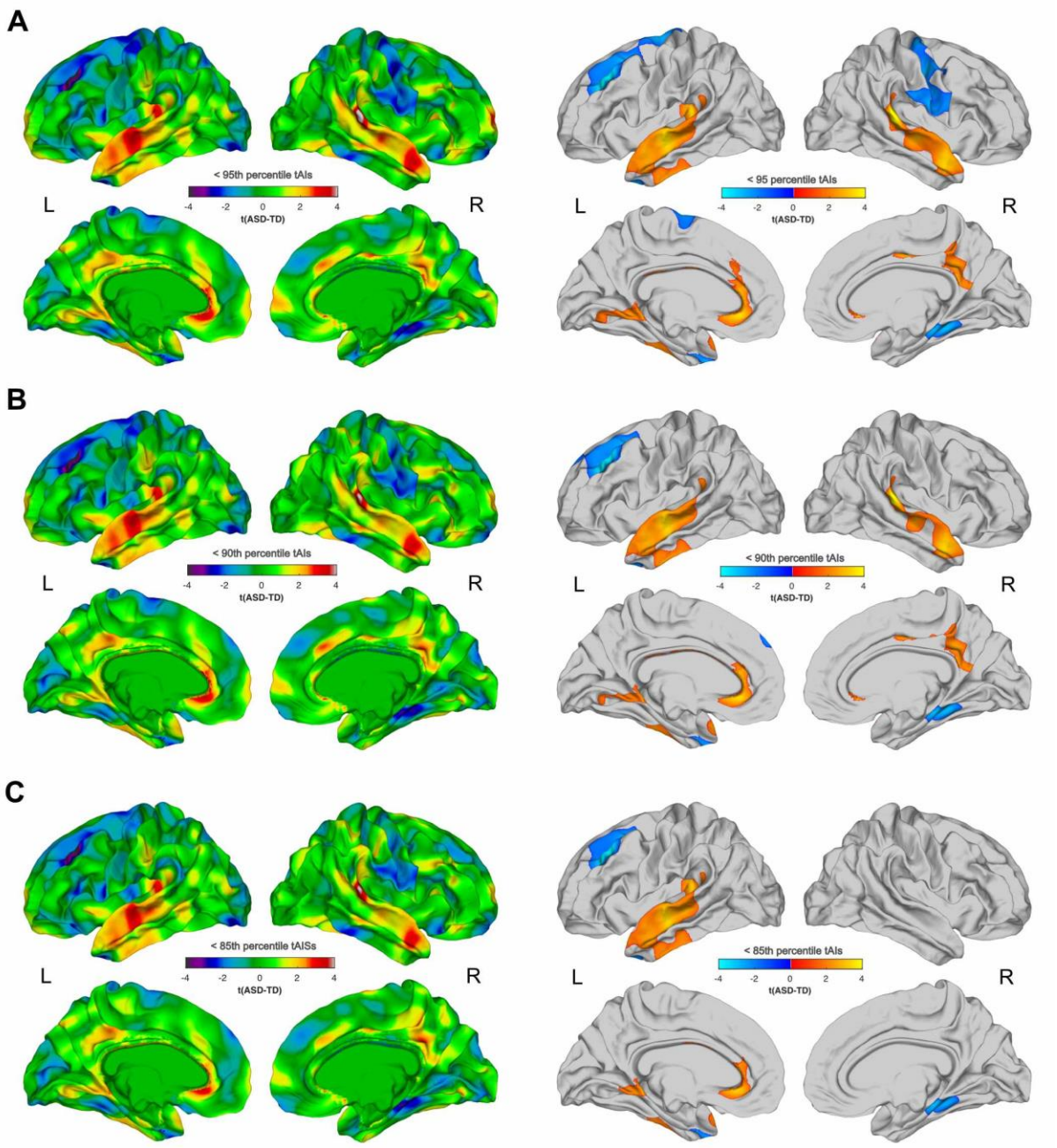
Group-by-sex interactions and differences in CT in males and females with ASD. (**A**) Brain regions where biological sex significantly modulates neuroanatomical variability in CT as examined via extending the GLM by a group-by-sex interaction term. (**B**) Vertex-wise differences in CT between male ASD individuals relative to male neurotypical controls. (**C**) Vertex-wise differences in CT between female ASD individuals relative to female neurotypical controls. The left panel shows *t-test* statistic for the contrast ASD minus control (un-thresholded); L: left hemisphere, R: right hemisphere. For (**B**) and (**C**), the right panel shows clusters with significantly increased (orange to yellow) and decreased (blue to cyan) CT in ASD (RFT-based cluster corrected, $p<0.05$, *two-tailed*). Group comparisons within sexes were based on N=259 males with ASD and N=178 male controls, and N=101 females with ASD and N=101 female controls.

FIGURE S15. CT differences between individuals with and without ID



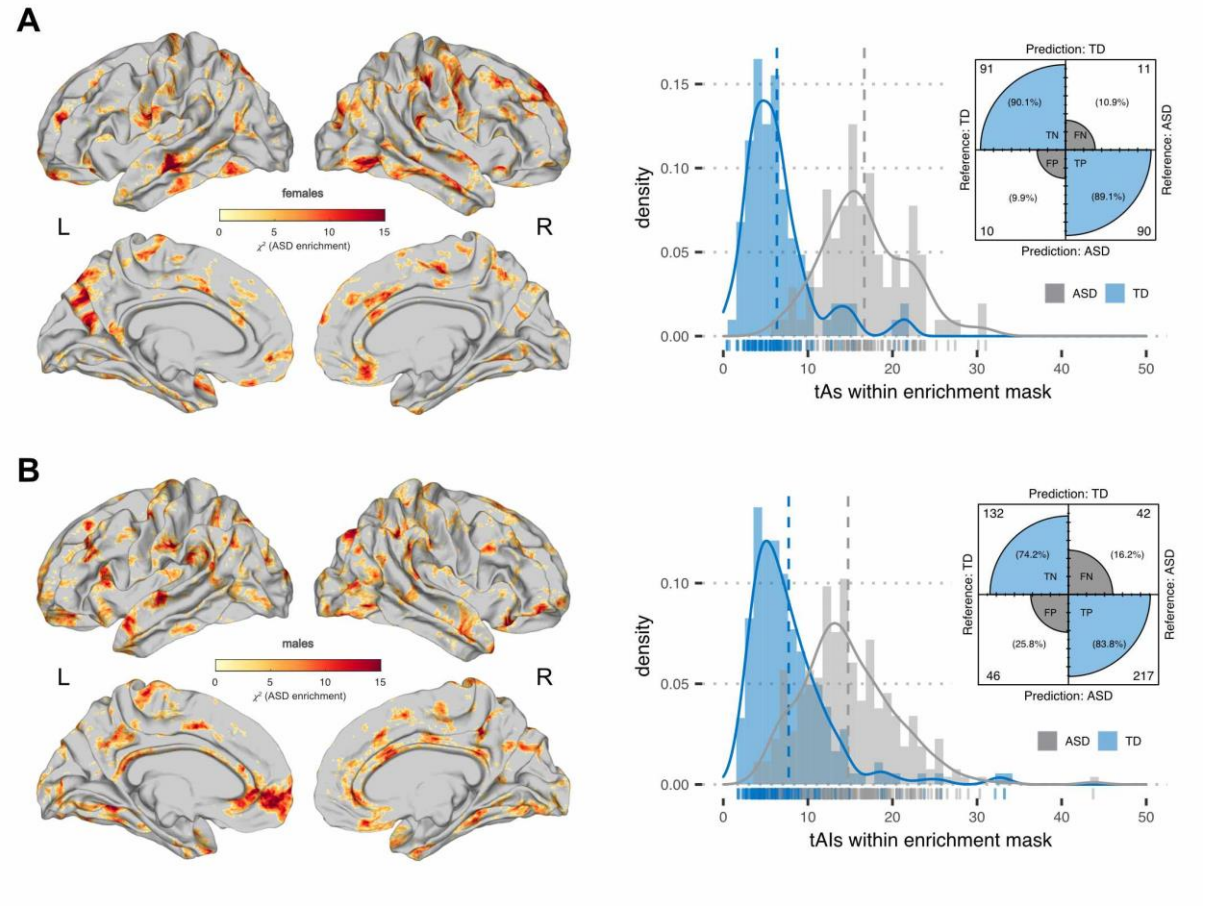
Differences in cortical thickness between individuals with and without mild intellectual disability (ID) and those without while controlling for age, site, sex, and mean cortical thickness. Main effect of ID in non-ASD individuals without covarying for full-scale IQ (FSIQ) (**A**), and when controlling for FSIQ (**B**). Main effect of ID in ASD individuals without covarying for full-scale IQ (FSIQ) (**C**), and when controlling for FSIQ (**D**). The orange to yellow color scale indicates regions with increased CT in the ID group relative to the non-ID group, while the blue to cyan color scale indicates clusters with decreased CT in the ID group relative to the non-ID group. All results are RFT-based cluster corrected at $p<0.05$, two-tailed.

FIGURE S16. Effects of neuroanatomical outliers on main effect of group



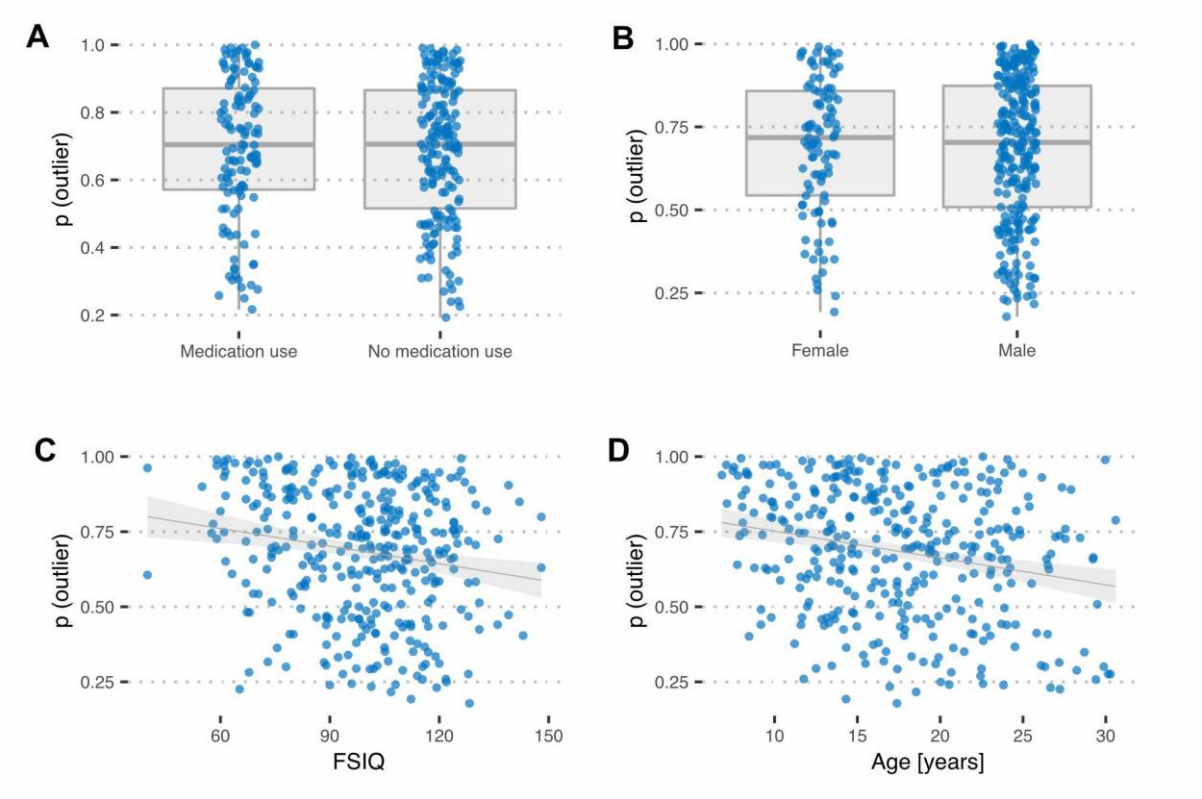
Between-group differences in cortical thickness after the iterative removal of neuroanatomical outliers within the ASD group. Panels show main effect of group following the exclusion of individuals exceeding the 95th (A), 90th (B), and 85th (C) upper percentile of the *tAIs* distribution, which corresponded to a value of 13.54%, 21.07%, and 19.81% of vertices outside the neurotypical 90% prediction interval. Based on these percentiles, a total of N=18, N=36, and N=54 ASD individuals were removed from the analysis in each iteration, respectively. The resulting maps are corrected for multiple comparisons using a Random Field Theory (RFT)-based cluster threshold of $p<0.05$ (*two-tailed*). The blue to cyan color scale indicates clusters with decreased CT in ASD relative to controls, and the orange to yellow color scale indicates clusters with increased CT in ASD relative to controls.

FIGURE S17. Neuroanatomical outliers by biological sex



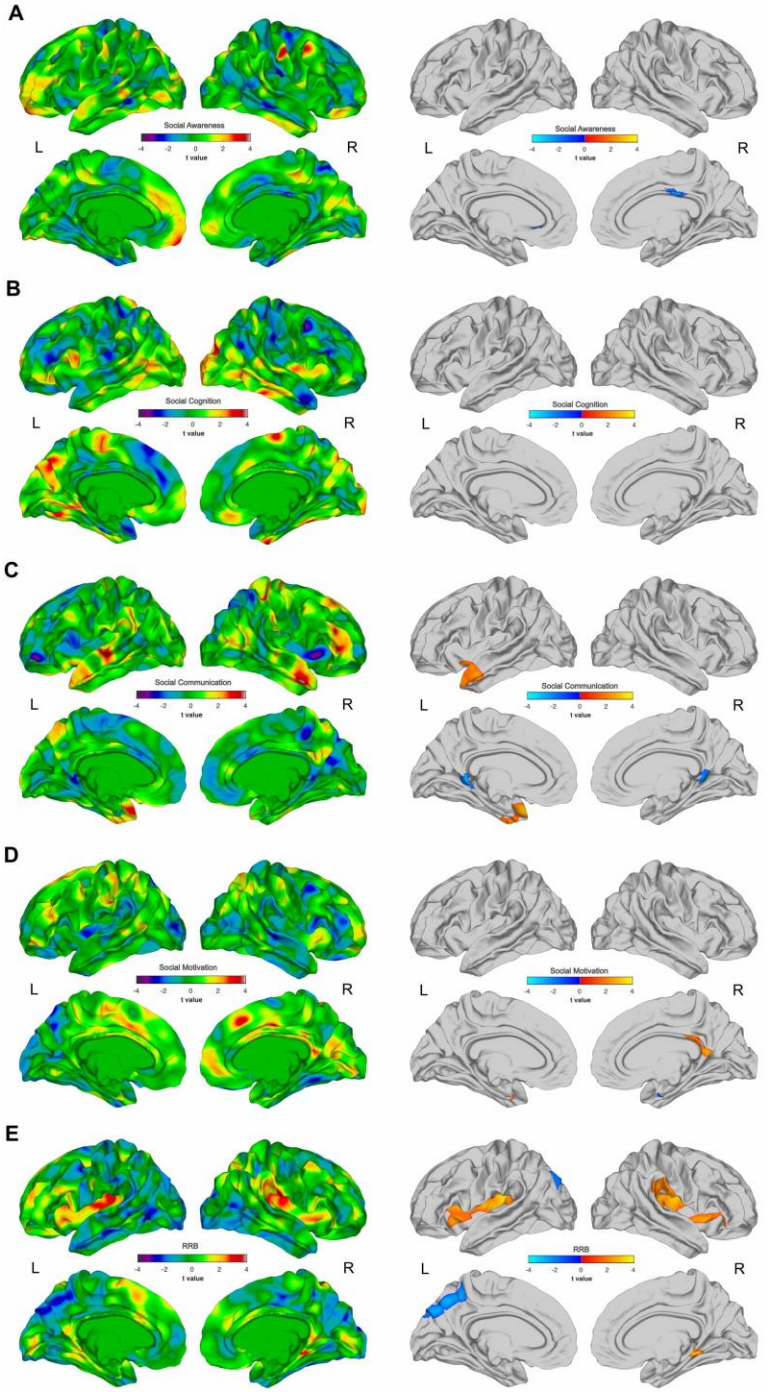
ASD enrichment mask and distribution of *tAIs* when estimating neurotypical variability in cortical thickness within females and males exclusively. The chi-square enrichment mask of neuroanatomical outliers based on the 90% neurotypical prediction interval for the all-female model (left panel), and the distribution of *tAIs* in females (right panel) are shown in (**A**). The chi-square enrichment mask of neuroanatomical outliers based on the 90% prediction interval for the all-male model (left panel), and the distribution of *tAIs* in males (right panel) are shown in (**B**). Based on the results within sexes, females with ASD were not more likely to be neuroanatomical outliers than males with ASD relative to their respective female (male) reference group (90 out of 202 for females vs. 101 out of 259 for males, chi-square(1)=1.24, p=0.265).

FIGURE S18. Neuroanatomical outliers by demographics



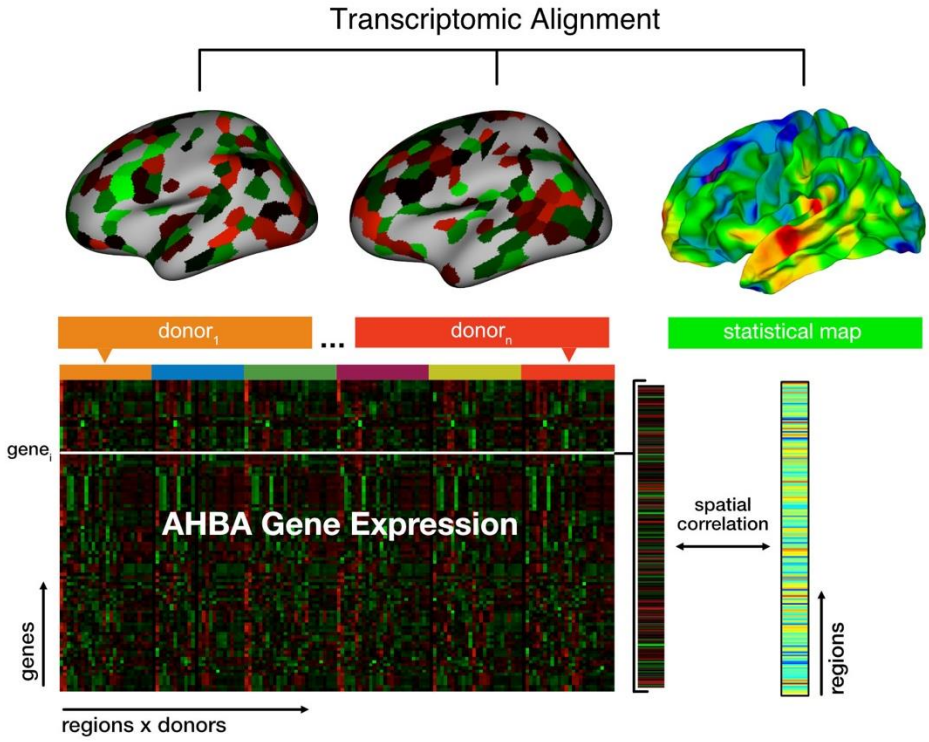
Probability of being a neuroanatomical outlier (*p(outlier)*) for ASD individuals based on the individual's total degree of neuroanatomical abnormality (i.e. *tAIs*) in cortical thickness within the ASD enrichment mask as a function of medication use (**A**), biological sex (**B**), full-scale IQ (FSIQ) (**C**), and age (**D**). There was no significant difference between individuals taking medication and those without, or between males and females ($p>0.05$, two-tailed). However, the probability of being a neuroanatomical outlier decreased significantly with increasing full-scale IQ (FSIQ) ($r=-0.18$, $t(358)=-3.47$, $p<0.001$) and with increasing age ($r=-0.23$, $t(358)=-4.48$, $p<0.001$).

FIGURE S19. Variability in CT across SRS-2 subdomains



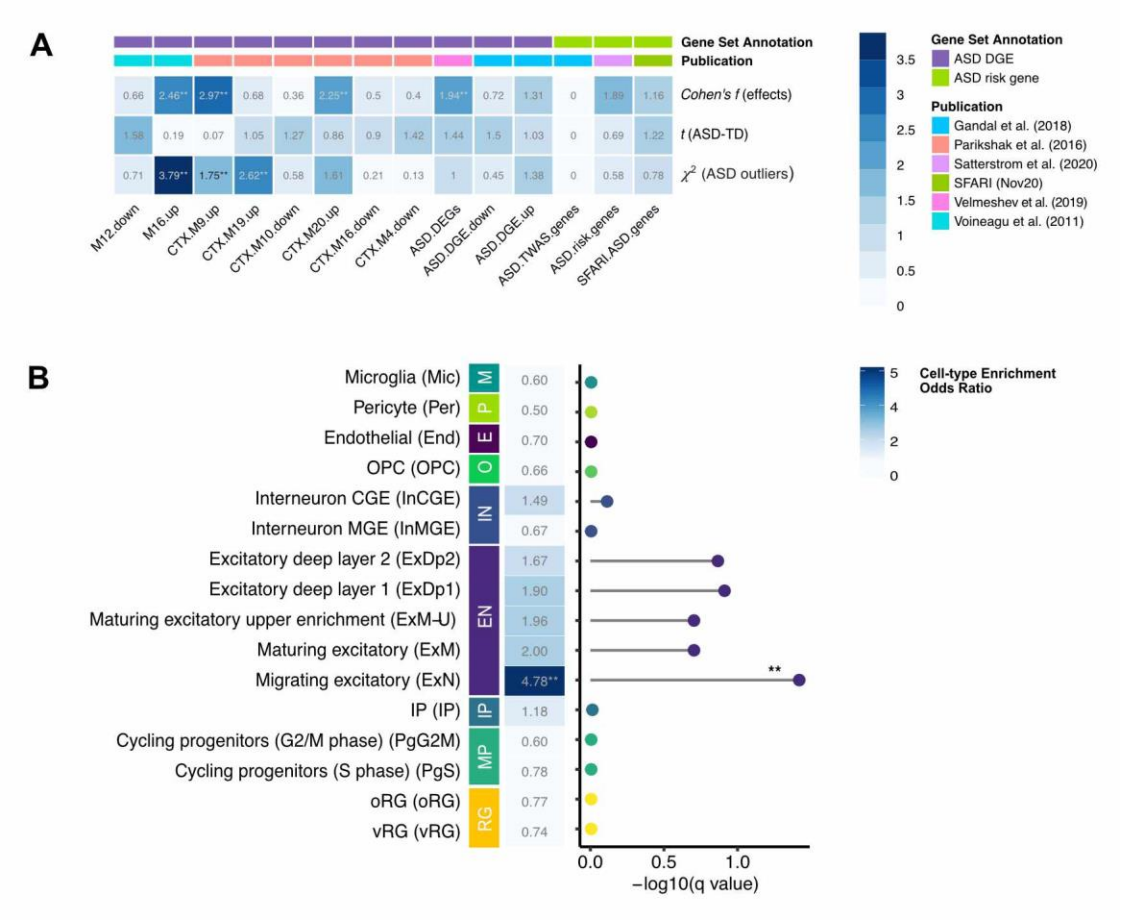
Variability in cortical thickness associated with the five SRS-2 subdomains (parent-rated) across groups while covarying for age (linear and quadratic), sex, full-scale IQ, site, and mean CT, based on N=290 ASD individuals and N=149 neurotypical controls for which parent-rated SRS-2 scores were available. Panels show the main effect of SRS-2 subdomains for the domains of (A) social awareness, (B) social cognition, (C) social communication, (D) social motivation, and (E) restricted/repetitive behavior (RRB). All results are corrected for multiple comparisons using a RFT-based cluster threshold of $p<0.05$, two-tailed. Orange to yellow clusters show increases in CT with increasing SRS-2 scores, and blue to cyan clusters show decreases in CT with increasing SRS-2 scores.

FIGURE S20. Gene Expression and Decoding Analysis



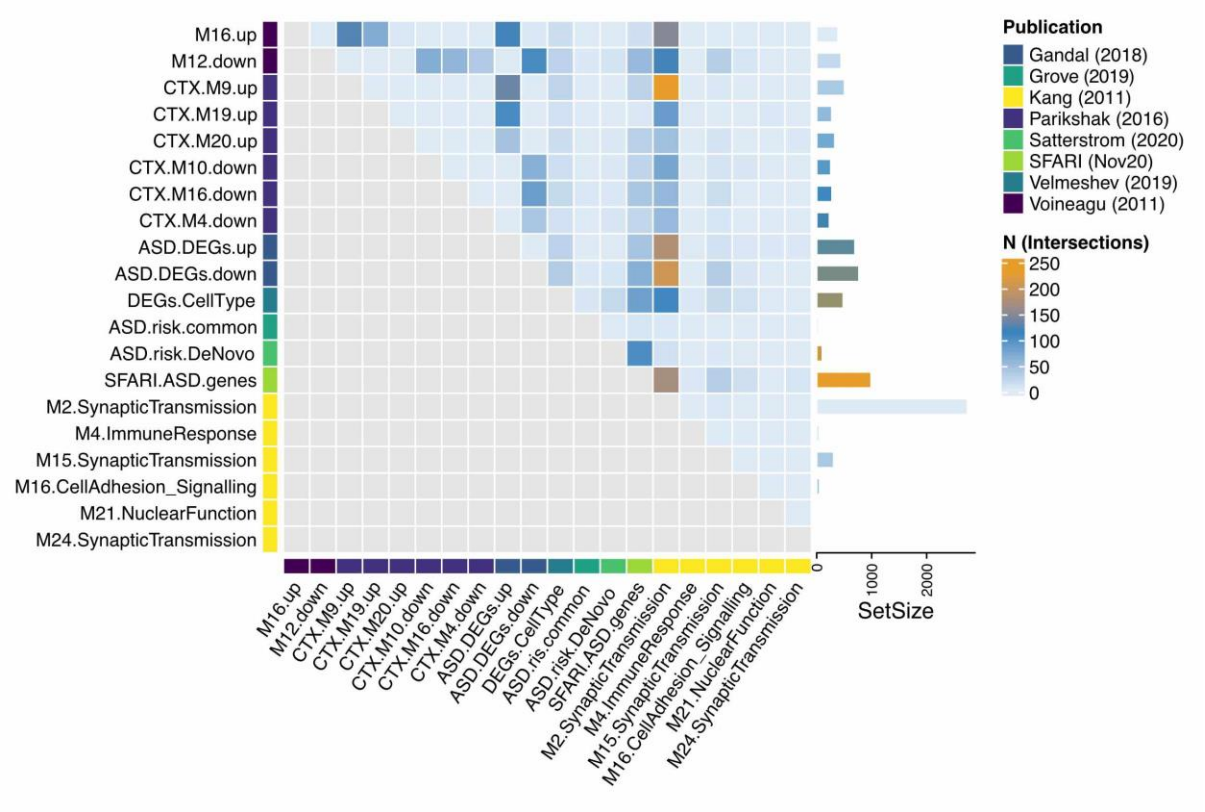
Schematic illustration of the gene expression and decoding analysis using the Allen Human Brain Atlas (AHBA (15)). Here, the AHBA donor brains are initially co-registered with the MNI atlas (also used by FreeSurfer) (i.e. transcriptomic alignment). At each sampling site (i.e. probe), a spherical region-of-interest (ROI) is drawn, and the statistics test parameter in each FreeSurfer overlay is averaged within each ROI (16). This resulted in a spatial vector of values for each donor, which was subsequently correlated with the normalised gene expression in order to identify genes whose spatial expression patterns are consistently highly similar to the imaging maps (i.e. across donor brains). *Note*. Expression data was created for illustration purposes only.

FIGURE S21. Gene Set Enrichment Analysis



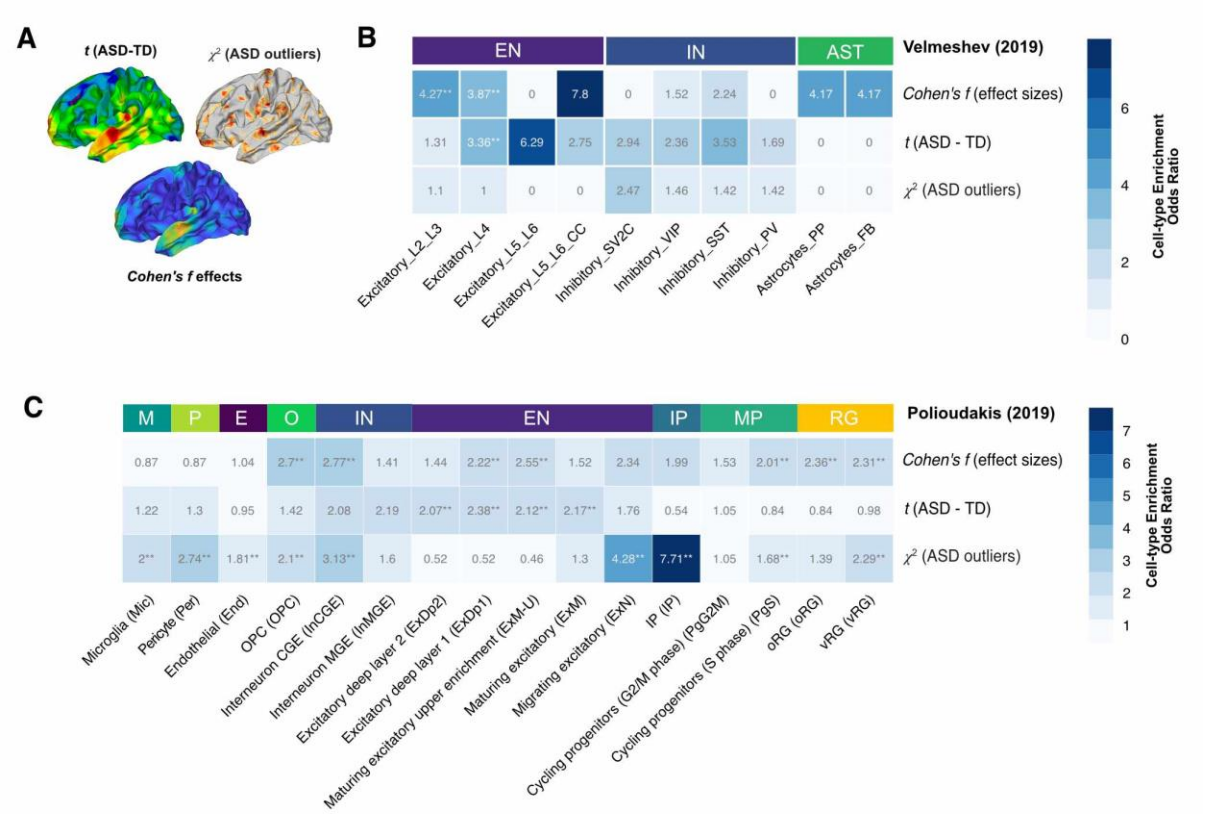
Gene Enrichment Analysis (GEA) given a background total of 16,541 genes (19). (A) GSEA for the *Cohen's f*-map, *t*-map, and chi-square outlier map (FDR-corrected p-value < 0.05), (B) GEA for the *F*-map associated with the main effect of sensory subgroup based on gene sets for different cell types (44)

FIGURE S22. Overlap between gene sets



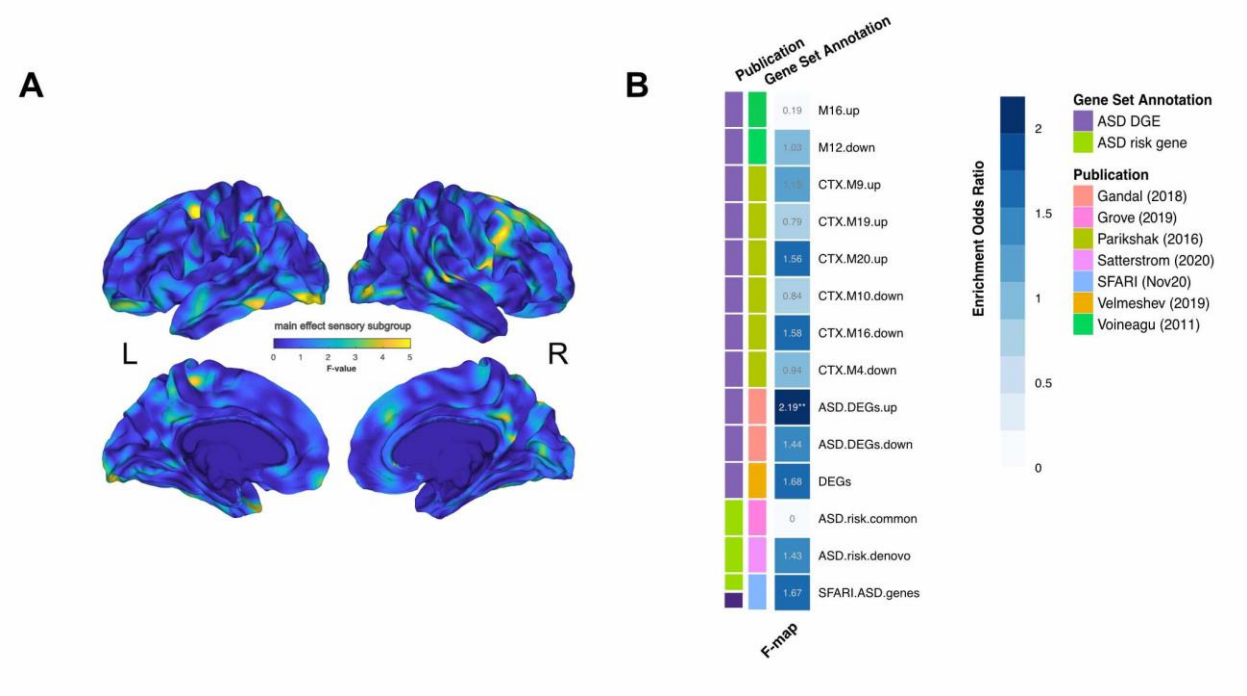
Overlap between the gene sets examined used in the main gene set enrichment analyses. The colorscale indicates the number of intersecting genes for each set pair. The barplot in the right panel shows the total number of genes per set.

FIGURE S23. Cell-type enrichment



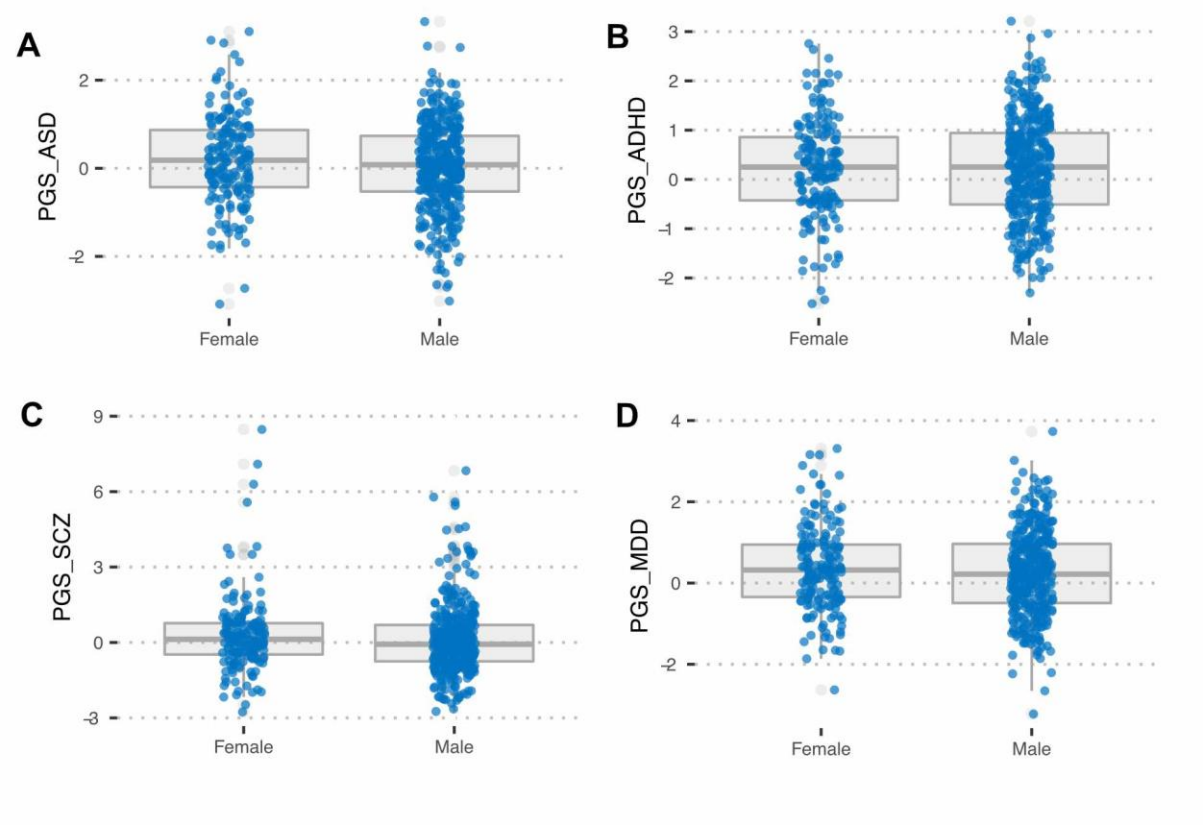
Cell-type enrichment for (A) the *t-map* of statistical between-group differences in CT (see Figure 1a), the *Cohen's f* effect size map associated with the main effect of group (see Figure 1d), and the χ^2-map of neuroanatomical ASD outliers (see Figure 2e). (B) Significant Odds-ratios (OR) for cell-specific gene-sets provided by Velmeshev et al. (2019), and (C) for genes provided by Polioudakis et al. (2019). EN: excitatory neuron, IN: interneuron, AST: astrocytes, M: microglia, P: pericyte, E: endothelial cell, O: oligodendrocyte precursor, IN: interneuron, EN: excitatory neuron, IP: intermediate progenitor, MP: mitotic progenitor, RG: radial glia, *: FDR-corrected *p*-value < 0.05, **: FDR-corrected *p*-value < 0.01.

FIGURE S24. Gene enrichment sensory subgroups



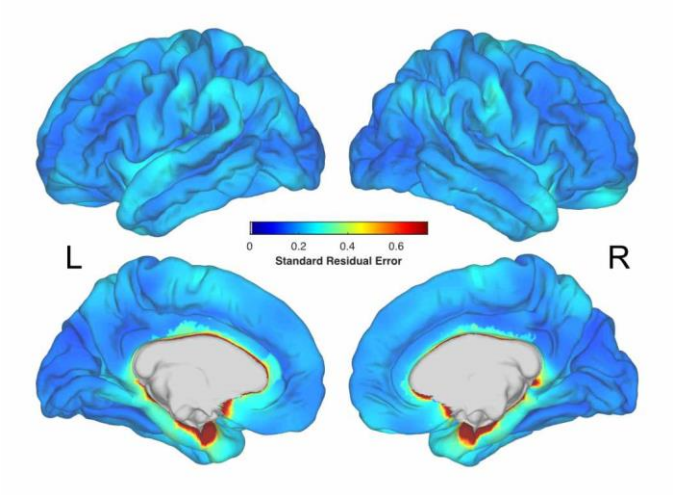
Enrichment analysis of ASD-related gene sets for the F-map associated with the main effect of sensory subgroup (A). (B) Significant Odds-ratios (OR) at an FDR-rate *p*-value <0.05 resulting from the gene set enrichment analyses for genes expressed in the different output maps. Gene sets were subdivided into sets with differential gene expression in ASD (ASD DGE), and sets representing ASD risk genes that contain either common variants (ASD.risk.common) or rare *de novo* variants (ASD.risk.denovo). Gene sets are annotated and labelled based on their original publication. up: upregulated expression in ASD, down: down-regulated expression in ASD, CTX: cortex, DEG: differentially expressed gene, *: FDR-corrected *p*-value < 0.05, **: FDR-corrected *p*-value < 0.01.

FIGURE S25. Polygenic risk for major psychiatric disorders by sex



Polygenic scores (PGS) for major psychiatric disorders by sex. (A) PGS for Autism Spectrum Disorder (ASD), (B) PGS for Attention Deficit Hyperactivity Disorder (ADHD), (C) PGS for Schizophrenia (SCZ), and (D) PGS for Major Depressive Disorder (MDD). There were no significant differences in the polygenic risk for any of these major psychiatric disorders between males and females (i.e., $p>0.05$, two-tailed)..

FIGURE S26. Model Standard Residual Error



Standard residual error associated with the neurotypical model at each vertex on the cortical surface. L: left hemisphere, R: right hemisphere