Data supplement for Pretzsch et al., Neurobiological Correlates of Change in Adaptive Behavior in Autism. Am J Psychiatry (doi: 10.1176/appi.ajp.21070711)

SUPPLEMENTARY CONTENT

Table of Contents

Supplementary Materials and Methods 2

Participants 2

Cortical reconstruction using FreeSurfer 2

Measures of adaptive functioning using the VABS-II 3

Neuroimaging: statistical analyses 3

Genetics: statistical analyses 5

Gene expression decoding analyses 5

Gene enrichment analyses 6

Protein-protein interaction analyses 6

Polygenic scores 6

Results: Neuroanatomical differences between adaptive outcome groups 7

Supplementary references 9

Supplementary Figures 12

Supplementary Tables 46

Supplementary Materials and Methods

Participants

Participants were included if they or their parents/guardians were able to provide informed written or verbal consent/assent to their participation in this study (after receiving a complete description of the study), and if they had a high-quality structural MRI scan. Autistic participants were included if they had an existing clinical diagnosis of ASD in accordance with DSM-IV/ICD-10 or DSM-5 criteria (1, 2); and Vineland Adaptive Behavior Scale-II (VABS-II) scores for both time points. As up to 70% of autistic individuals present with one or more co-occurring psychiatric conditions (3), all psychiatric comorbidities (except for psychosis and bipolar disorders) were allowed. Moreover, given the high number of autistic individuals who are being prescribed regular medication (30-50% in Europe (4) and 70% in the US (5)), those on stable medication were also included in this study. In contrast, participants with conditions precluding them from safe scanning, such as metal objects in their body, were excluded. The study was approved by national and local ethics review boards at each study site. This included the London-Central and Queen Square Health Research Authority Research Ethics Committee (University of Cambridge and King's College London; ID 13/LO/1156), the UMM University Medical Mannheim Medical Ethics Commission II (Mannheim University; ID 2014-540N-MA), the Radboud University Medical Centre Institute Ensuring Quality and Safety Committee on Research Involving Human Subjects Arnhem-Nijmegen (Radboud University and Utrecht University; ID 2013/455), and the University Campus Bio-Medical Ethics Committee De Roma (Rome University; ID 18/14 PAR ComET CBM). This study was carried out to Good Clinical Practice (ICH GCP) standards.

Cortical reconstruction using FreeSurfer

To ensure consistent image quality, we first performed manual quality control. Specifically, we inspected all scans manually; and excluded scans with visible anomalies or significant movement artefacts from further analyses. Next, in each remaining T1 volume, we used FreeSurfer v6.0 (https://surfer.nmr.mgh.harvard.edu/), which performs automated image quality control and computes models of the cortical surface. In brief, FreeSurfer performs intensity normalization (correction for intensity non-uniformity in MRI data), skull stripping (improves signal to noise ratio by removing the skull), and an image segmentation (multi-step procedure providing the base for subsequent estimation of surface measures; also provides output that is adjusted during manual edits, see below) using a connected-components algorithm. Next, it generates a filled white matter volume for each hemisphere and, by fitting a deformable template, a surface tessellation for each of these volumes. For each subject, this yields a mesh (of triangular elements) for the inner (white-matter) and outer (pial) cortical surface consisting of ~150k vertices (points in each triangular element) per hemisphere. For a more detailed description of these well-validated and fully automated procedures, please refer to e.g., (6-9). To ensure image quality further (through manual and automated steps), each reconstructed surface was inspected visually for reconstruction errors by three independent and experienced researchers (who were blinded to the diagnostic status of each participant). Of the initial 709 individual's scans, the researchers (a) accepted 347 scans (48.9%); (b) rejected 52 scans (7.3%), mostly due to severe (motion) artefacts; and (c) prescribed manual edits for 310 scans (43.7%) in the case of smaller, i.e., local, reconstruction errors. These 310 images were manually edited, (re-)pre-processed, and visually (re)assessed. Of these, 307 surface reconstructions improved; the remaining three were excluded from all further statistical analyses. We also excluded 15 scans due to scanner upgrades/missing demographic information. In total, we excluded 56 out of 416 (13.5%) autistic individuals and 14 out of 293 (4.8%) neurotypicals across sites, i.e., the original dataset contained data from n=709 individuals, and the final dataset contained data from n=639 individuals. This dropout was approximately evenly distributed across sites (Cambridge: 10 (5 ASD, 5 TD)/88=11.4%; King's: 31 (28 ASD, 3 TD)/241=12.9%; Mannheim: 8 (5 ASD, 3 TD)/69=11.6%; Nijmegen: 13 (11 ASD, 2 TD)/184=7.1%; Rome: 0/41=0%; Utrecht: 8 (7 ASD, 1 TD)/86=9.3%). Finally, we selected those autistic individuals with recorded adaptive behaviour scores; and retained a final sample of 483 individuals (204 ASD, 279 TD) Table2. We then computed three vertex-wise cortical morphometric features: cortical volume (CV), cortical thickness (CT), and surface area (SA). CT values were calculated as the closest distance from the grey-white matter boundary to the grey matter-cerebrospinal fluid boundary at each vertex on the tessellated surface (8). Vertex-based estimates of SA were derived as outlined by Winkler et al. (10) and multiplied by CT to compute vertex-wise estimates of CV. For each subject, we also computed total CV, mean CT, and total SA values across the entire brain. To improve our ability to detect population changes, each parameter was smoothed using a 10 mm surface-based smoothing kernel. This kernel was selected based on previous reports that 10 mm may balance the trade-off between reliability and statistical power (11).

Measures of adaptive functioning using the VABS-II

To assess our autistic participants' adaptive behaviour, we interviewed their parents/carers using the VABS-II (12). This semi-structured interview has a high test-retest reliability (Intraclass-correlation coefficient [ICC]=0.8-0.9; (13)); and moderate to high inter-interviewer reliability (ICC=0.7-.76), moderate to high inter-rater reliability (ICC=.71-.81), and excellent internal consistency (split half reliability coefficient: .93-.97) (all based on VABS-II manual). The VABS-II assesses a person's current level of everyday functioning across three domains, each of which is composed of three subdomains. They include communication (expressive, receptive, and written), daily living skills (community, domestic, and personal), and socialization (coping skills, interpersonal relationships, and play and leisure time). We derived age-normed standard scores (mean=100, standard deviation=15) for each domain. Next, we computed the Adaptive Behaviour Composite (ABC) score (i.e., the total degree of impairment across all three domains) both at T1 and T2, as well as the change in between (V_{T2}-V_{T1}). Based on these ABC change scores and recently published estimates of what constitutes a 'Minimal Clinically Important Difference' (MCID) (14), we classified autistic individuals into three adaptive outcome subgroups: those whose scores could be said to meaningfully improve (Increasers; V_{T2}-$V_{T1}\geq 4$), showed no meaningful change (No-changers; $-4<V_{T2}$-$V_{T1}<4$), and those whose scores appeared to have meaningfully declined (Decreasers; $-4\geq V_{T2}$-V_{T1}). In contrast to other related measures (such as the Minimum Detectable Change [MDC], which considers the standard error of measurement, e.g., based on the test-retest reliability), the MCID quantifies the amount of change required to be clinically, rather than statistically, meaningful. As such, the MCID is widely supported (including by the Food and Drug Administration, FDA) (15) as a way to evaluate (treatment) outcomes. To adjust for regression to the mean, we controlled all relevant analyses for baseline ABC scores (FigureS1).

Neuroimaging: statistical analyses

In this study, we examined three cortical features: CV, SA, and CT. We selected these features because a wealth of studies has researched and implicated them in ASD (16-19). Notably, SA and CT are thought to have different (separate) underpinning neurobiological mechanisms (20-22). In contrast, CV is the product of SA and CT, and therefore not an independent measure. Nonetheless, previous studies suggest that the contributions of SA and CT to CV may vary across regions, age, and be atypical in ASD (16, 21, 23, 24). Hence, examining CV may not only help to jointly investigate the product of (atypical) SA and CT, but also provide insights into how these features interact (differently) with each other in ASD.

Analysis 1: ASD subgroup comparison

First, we tested if there were neuroanatomical differences at T1 between groups of individuals that shared an adaptive outcome. To address this aim, we ran GLM regression models computing the parameter estimates for CV/CT (C_i) and SA (SA_i) at each vertex *i*. We included adaptive outcome subgroup, sex, and site as fixed-effect factors; and linear (CV/CT/SA) and quadratic (CV/CT) age (these age terms were selected based on previous literature suggesting that SA follows a linear growth trajectory (25, 26); while CV (27) and CT (25, 26) follow quadratic growth trajectories), IQ, total brain measures (total CV/SA, mean CT), and T1 ABC scores (V_{T1}) as continuous covariates, i.e. $C_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Group} + \beta_4 \text{Sex} + \beta_5 \text{IQ} + \beta_6 \text{Site} + \beta_7 \text{Total Brain} + \beta_8 V_{T1} + \varepsilon_i$ and $SA_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Group} + \beta_3 \text{Sex} + \beta_4 \text{IQ} + \beta_5 \text{Site} + \beta_6 \text{Total Brain} + \beta_7 V_{T1} + \varepsilon_i$, where ε_i is the residual error at vertex *i*. All continuous variables were mean centred across groups to improve the interpretability of the coefficients. We examined between-group differences using the coefficients β_3 (for CV and CT) and β_2 (for SA). We corrected for multiple comparisons across the whole brain using random-field theory (RFT)-based cluster-correction for non-isotropic images with a cluster-defining and p-value significance threshold of $p<.05$ (two-tailed) (28). Additionally, we performed our analyses applying stricter cluster-defining thresholds (.01 and .001) at a cluster p-value threshold of <.05 (two-tailed) to identify those clusters displaying (highly) robust effects. Given our a priori interest into CV, SA, and CT, (and their distinct neurobiological underpinnings) we treated them as separate analyses and did not correct for multiple comparisons across these three features. To display our findings, we mapped our statistical effects onto the FreeSurfer high-resolution template in standard space (fsaverage).

Analysis 1: supplementary analyses

To examine the possibility that our results were distorted by the inclusion or exclusion of several (potentially) confounding factors, we conducted supplementary analyses. Specifically, we repeated all analyses without covarying for total brain measures; with two additional approaches correcting for site effects, including ComBat batch effect harmonization (29) and modelling site as a random effect; while covarying for medication; and, given the role of IQ in ASD (30), with and without correction for IQ.

Also, we extended our primary analyses, which followed a categorical approach based on the (potentially clinically meaningful) MCID-derived cut-off. Specifically, we conducted supplementary analyses using a continuous approach to test if there was an association between adaptive outcome and neuroanatomy independent of the MCID 'cut-off' of four units. To address this aim, we ran GLM regression models computing the parameter estimates for CV/CT (C_i) and SA (SA_i) at each vertex *i*, including sex and site as fixed-effect factors; and continuous change in adaptive behaviour (Change in ABC scores), linear and quadratic age (selected based on previous literature (26, 31)), IQ, total brain measures (total CV/SA, mean CT), and T1 ABC scores (V_{T1}) as continuous variables, i.e. $C_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Change in ABC scores} + \beta_4 \text{Sex} + \beta_5 \text{IQ} + \beta_6 \text{Site} + \beta_7 \text{Total Brain} + \beta_8 V_{T1} + \varepsilon_i$ and $SA_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Change in ABC scores} + \beta_3 \text{Sex} + \beta_4 \text{IQ} + \beta_5 \text{Site} + \beta_6 \text{Total Brain} + \beta_7 V_{T1} + \varepsilon_i$, where ε_i is the residual error at vertex *i*. We then examined the main effect of Change in ABC scores using the coefficients β_3 (for CV and CT) and β_2 (for SA).

We also repeated our analyses to examine the main effect of subgroup across all ASD individuals. To this aim, we recomputed our analyses using the models specified above (but exchanging the continuous change in adaptive behaviour variable for a categorical factor denoting subgroup); and examined the effect of subgroup using the coefficients β_3 (for CV and CT) and β_2 (for SA).

Moreover, our primary analyses harnessed the full age-range in our sample; but we recognize that, given the developmental nature of ASD, the association between neuroanatomy and adaptive outcome may vary across ages and follow-up durations. Therefore, we conducted supplementary analyses controlling for follow-up duration and its interaction with age.

Additionally, given that we observed both age- and sex-differences between the ASD subgroups, we reran our analyses in a subsample of individuals that were matched for age (range 13.5-19 years, n_{total} = 83), sex, and IQ. Our primary results were unchanged across these sensitivity analyses.

Additionally, to disentangle the neuroanatomical correlates of adaptive outcome across age - and also to make our findings more easily comparable to previous research restricted to individual age groups - we conducted supplementary analyses of between-outcome group differences separately within children/adolescents (6-17 years of age) and adults (18-30 years of age). We grouped children and adolescents together to retain balanced sample sizes, i.e., numbers of participants within outcome groups within each age-group. Analyses for the IG vs DG group (the biggest outcome 'contrast') were performed exactly as outlined above in *Analysis 1*.

Also, our primary goal was to examine adaptive behaviour and we relied on a single instrument (VABS-II). To further corroborate an association between neuroanatomy and change in behaviour related/relevant to adaptive behaviour (such as social-communication processing), we conducted supplementary analyses.
To this aim, we also stratified individuals into outcome groups based on i) total Autism Diagnostic Observation Schedule (ADOS) (32) scores, ii) ADOS social affect domain scores, and iii) Social Responsiveness Scale (SRS) (33) scores. We selected these measures because, similarly to the VABS-II, they capture ASD core symptoms, especially in the social-communication domain. For each individual and measure, we computed change scores (between T1 and T2). We then standardized these scores and used a cut-off of 0.5 standard deviations (SD) to group individuals into those whose scores increased (V_{T2}-$V_{T1} \geq 0.5$ SD), did not change (-0.5SD< V_{T2}-V_{T1}<0.5SD), and decreased (V_{T2}-$V_{T1} \leq$-0.5SD). These cut-offs were chosen to balance participant numbers between outcome groups. They are preliminary and future studies should examine (the effect of) various thresholds. Next, for each measure, we tested if there were neuroanatomical differences at T1 between subgroups. ADOS and SRS scores are 'reverse-coded', i.e., an increase in score indicates a 'worsening' in symptoms. Here, we contrasted those individuals who improved in scores with those who deteriorated. To this aim, we ran GLM regression models computing the parameter estimates for CV/CT and SA at each vertex, using the respective outcome subgroups, sex, and site as fixed-effect factors; and linear (CV/CT/SA) and quadratic (CV/CT) age, IQ, total brain measures (total CV/SA, mean CT), and baseline symptom measures (T1 ADOS total, T1 ADOS social affect domain, and T1 SRS score, respectively) as continuous covariates (as shown above in *Analysis 1*).

Note that, while we focused on the abovementioned clinical measures (VABS-II, ADOS, SRS), future studies should expand our analyses to other assessments of (features related to) adaptive behaviour. Examples may include the Adaptive Behaviour Assessment System (ABAS) (34), the American Association for Mental Deficiency (AAMD) Adaptive Behaviour Scale (35), the Adaptive Social Behavior Inventory (ASBI) (36), or the Diagnostic Adaptive Behavior Scale (DABS) (37) (reviewed in (38)).

For the sake of completeness, we also examined neuroanatomical differences between neurotypicals and all ASD outcome groups combined. To this aim, we conducted supplementary GLM regression models computing the parameter estimates for CV/CT (C_i) and SA (SA_i) at each vertex *i* including group (ASD vs TD), sex and site as fixed-effect factors; and linear and quadratic age (selected based on previous literature (26, 31)), IQ, and total brain measures (total CV/SA, mean CT) as continuous variables, i.e. $C_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Group} + \beta_4 \text{Sex} + \beta_5 \text{IQ} + \beta_6 \text{Site} + \beta_7 \text{Total Brain} + \varepsilon_i$ and $SA_i = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Group} + \beta_3 \text{Sex} + \beta_4 \text{IQ} + \beta_5 \text{Site} + \beta_6 \text{Total Brain} + \varepsilon_i$, where ε_i is the residual error at vertex *i*. We then examined the main effect of Group using the coefficients β_3 (for CV and CT) and β_2 (for SA).

All main (and supplementary) steps of Analysis 1 were conducted using the SurfStat toolbox (https://www.math.mcgill.ca/keith/surfstat/) within MATLAB R2019a (The MathWorks, Inc., MA, US).

Analysis 2: Computation of 'Atypicality Indices' (AIs)
Second, we extended our analyses from the (sub)group-level to the individual level. We examined if, in the regions identified above, an individual's deviation from the neurotypical neurodevelopmental trajectory at T1 predicted that individual's subsequent change in adaptive behaviour. To achieve this, we first used the data from the neurotypicals (without intellectual disability, ID) to create a general linear model of the neurotypical developmental trajectory of each vertex' CV/CT/SA given a subject's age, sex, IQ, site, and overall brain measures (see GLM above). Next, we applied this model to the ASD group. We included autistic individuals with and without ID; however, there were no differences in IQ/the occurrence of ID between subgroups, and we controlled for IQ in our analyses. We then measured how much each autistic individual deviated from the neurotypical trajectory. We centred and normalised these deviances (residuals) based on the neurotypical distribution to express all data in units of standard deviation from the neurotypical mean. We used the residuals to identify neuroanatomical 'outliers' (morphometric values outside the 90% neurotypical Prediction Interval [$PI_{90\%}$] at each vertex). Next, we summarized each individual's outliers in an 'Atypicality Index' (AI). Specifically, we computed each individual's degree of neuroanatomical deviation, i.e., the mean of the standardized residuals outside the $PI_{90\%}$, where deviations in both the positive and the negative direction are classified as atypical. We computed AIs separately for each subject and for each set of regions that differed in a morphometric feature in the previously described subgroup comparisons. We then tested if these regional AIs predicted change in adaptive behavior in ASD (Y) by running separate linear regressions for the different subgroup comparisons. Models were controlled for age, sex, IQ, site, and T1 ABC scores (V_{T1}), i.e., $Y = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 \text{IQ} + \beta_4 \text{Site} + \beta_5 V_{T1} + \beta_6 \text{AI CV} + \beta_7 \text{AI SA} + \beta_8 \text{AI CT}$. In sum, this allowed us to identify regional deviations from the neurotypical neurodevelopmental profile of a particular morphometric feature that predict change in adaptive behaviour in ASD at the individual level. We also recomputed these analyses while performing Bootstrapping (4000 iterations) to increase the accuracy of our estimates.

All main (and supplementary) steps of Analysis 2 were conducted using the SurfStat toolbox (https://www.math.mcgill.ca/keith/surfstat/) within MATLAB R2019a (The MathWorks, Inc., MA, US); and R (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

Genetics: statistical analyses

Gene expression decoding analyses

We examined how our neuroanatomical findings may relate to underlying genomic mechanisms. This analysis (decoding analysis) was based on the Allen Human Brain Atlas (AHBA) (39). The AHBA is the most comprehensive gene-expression atlas currently available, but it is also based on adult donors only. Also, its spatial coverage and resolution are much lower than those of neuroimaging data. Consequently, future studies should repeat our decoding analysis using age-specific high-resolution gene expression datasets once they become available. Here, we leveraged the AHBA to highlight lists of genes that are highly expressed throughout the brain in spatial patterns that are similar to the observed neuroanatomical differences between the adaptive behavioural change subgroups. To this aim, we uploaded the cortical difference maps of the various subgroup contrasts to Neurovault (https://neurovault.org). Next, we used python code embedded within Neurovault and Neurosynth (https://neurosynth.org) to perform a gene expression decoding analysis that statistically tests all 20,787 protein coding genes for spatial gene expression similarity to our imaging maps (40). Specifically, the analysis constructs a linear model for each of the six donor brains in the AHBA, where the slopes encode how similar each gene's spatial expression pattern is to the input imaging map. In line with the input maps, these analyses are restricted to cortical tissue. The slopes are then subjected to a one-sample t-test to identify genes whose spatial expression patterns are consistently (across the donor brains) highly similar to the imaging maps. The resulting list of genes was thresholded at $p<.001$. We chose this 'liberal' threshold as this analysis did not

constitute a hypothesis test per se, but rather was 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.

Following the identification of these genes, we investigated i) their implication in ASD (gene enrichment analysis), ii) their functional roles and biological relevance (protein-protein interaction analysis), and iii) if variation in these genes correlated with neuroanatomical variability in our participants

Gene enrichment analyses

We examined if these identified genes may be implicated in ASD using R code written by M.V.L. (https://github.com/mvlombardo/utils/blob/master/genelistOverlap.R). Specifically, we tested our gene lists for enrichment with different classes of genes known to be associated with ASD at the genetic and transcriptomic level. At the genetic level, this included the 102 rare de novo protein truncating variants identified in the largest exome sequencing study of autism world-wide (41) (ASD dnPTVs). We also included an ASD-related gene list compiled by SFARI (ASD SFARI; categories S, 1, 2, & 3 downloaded on January 14th, 2020, from https://gene.sfari.org/). At the transcriptomic level, we included a list of differentially expressed (upregulated/downregulated) genes identified in autism post-mortem frontal and temporal cortex tissue (ASD DE Downreg, ASD DE Upreg) (42). From the same study, we also included a list of genes differentially expressed in schizophrenia and bipolar disorder (SCZ DE, and BD DE), as these conditions are somewhat genetically correlated with ASD (42). Additionally, we examined genes that are differentially expressed in ASD in specific cell types (ASD: Excitatory, Inhibitory, Microglia, Oligodendrocyte, Astrocyte, Endothelial) (43). Finally, we examined genes from differentially expressed co-expression modules in ASD (ASD CTX Downreg CoExpMods, ASD CTX Upreg CoExpMods) (44).

First, we conducted our enrichment analyses using a background list of 20,787 genes, which are all genes considered in Neurosynth (https://neurosynth.org/). Second, to avoid biasing our findings towards genes expressed in brain, we limited our background list to 16,906 genes, based on real estimates of genes expressed in cortical tissue (45) (FigureS33). Our enrichment analyses yielded enrichment Odds Ratios, hypergeometric p-values, and False Discovery Rate (FDR) q-values. Only those tests with pFDR<.05 were interpreted further.

While other methods (e.g. Family-wise error rate Bonferroni etc.) tackle the multiple comparison problem by controlling the probability of making even one false discovery (and thus reduce power especially in the case of a large number of variables/endpoints), FDR correction aims to control the expected proportion of falsely rejected hypotheses and can thereby increase power (46). Accordingly, FDR correction has been widely used in previous studies conducting genetic enrichment analyses in ASD (47, 48).

Protein-protein interaction analyses

Next, to explore the functional roles and biological relevance of the identified genes, we conducted protein-protein interaction (PPI) analyses to determine if these specific genes highly interact at the protein level. In other words, these PPI analyses assessed if the products of the identified genes are thought to interact physically to support biological processes – and, if so, which processes they are. This PPI analysis was implemented within STRING (https://string-db.org/). Our input gene list for this PPI analysis included the genes from the Increasers vs Decreasers CV and SA enrichment for ASD DE Downreg genes. To estimate evidence for PPI, we used only the seed genes and all default settings (e.g., minimum required interaction score=0.4, using all interaction sources). We also constructed an extended PPI network that included the seed genes plus up to 50 top interactors with these seed genes. From this extended network, we computed Gene Ontology (GO) biological process enrichment results and then selected and categorized terms passing $p_{FDR}<.05$ that were of relevance to the cortical phenotypes (e.g., autophagy, cell cycle, cell death, growth, neurogenesis, and synapse).

Polygenic scores

We also tested if variation in the identified genes correlated with neuroanatomical variability in our participants, i.e., if single nucleotide polymorphisms (SNPs) in the genes identified in the gene-expression analyses above were related to neuroanatomical deviations from the neurotypical profile. To do that, within the LEAP cohort, we generated autism PGS restricted to common genetic variants in the previously identified sets of genes that were expressed in regions that differed neuroanatomically between subgroups.

To this aim, genetic samples collected in LEAP were genotyped using the Infinium OmniExpress-24v1 BeadChip. We excluded participants with a genotyping rate below 95%, heterozygosity above or below 3 SD from the mean, or a mismatch in reported and genetic sex. We further removed single SNPs that deviated from Hardy-Weinberg Equilibrium ($p<1x10^{-6}$) and that had a genotyping rate below 95%. Imputation was conducted on the Michigan Imputation server. Given that the majority of individuals in LEAP were of European ancestry, we used the HRC r1.1 2016 reference panel. We generated principal components and, using the first genetic principal components, reduced their dimensionality to two components (x-umap-spread and y-umap-spread columns) (https://arxiv.org/abs/1802.03426). Density based-clustering on these clusters identified individuals of European ancestry. We only included unrelated individuals of European ancestries in our PGS analyses. Thus, we retained 153 subjects.

PGS were generated in PRSice2 (49) using independent SNPs present in both the training set, i.e., the largest genome-wide association study (GWAS) of autism (50), and the testing dataset, i.e., the individuals that were genotyped in the LEAP cohort. PGS were calculated as weighted averages of the total number of trait-increasing alleles (for a particular phenotype) in an individual. The weights were derived from the GWAS regression betas in the training dataset. We clumped SNPs using an imputation r^2 of 0.1 and a physical distance of 250 kb. We used the latest GWAS of autism (50) to generate PGS. We chose an a priori p-value threshold of $p\leq.1$ as this explained the highest variance in autism (50). This resulted in a total of 50,122 SNPs.

Next, we generated gene-set based PGS, i.e., to increase the signal-to-noise ratio we limited PGS to SNPs in specific sets of genes. Specifically, we restricted our PGS to SNPs in the genes we previously identified in our decoding analysis, i.e., those whose spatial expression patterns were significantly similar to the neuroanatomical subgroup differences. SNPs were mapped to genes based on physical positions of the genes included in the gene sets (hg19). These SNPs were then clumped, restricted to those with an a priori p-value threshold of $p\leq.1$. In total, we generated seven gene set based analyses (the contrast for Increasers vs No-changers yielded a small number of genes, hence these results should be interpreted with caution). Gene numbers were as follows: IG vs DG CV: 97; IG vs DG SA: 363; IG vs DG CT: 42; IG vs NCG CV: 57; IG vs NCG SA: 11; DG vs NCG CV: 33; DG vs NCG SA: 36.

We tested the association between PGS and AIs for each set of regions differing between subgroups using Pearson correlation analyses, controlling for age, IQ, sex, site, and the first five genetic principal components ($p_{FDR}<.05$). We also repeated our analyses using Bootstrapping (4000 iterations). This step was taken to increase the accuracy of our estimates.

Results: Neuroanatomical differences between adaptive outcome groups

Increasers and Decreasers differed in CV in eight clusters (Figure2a). CV was greater in Increasers vs Decreasers in clusters containing parts of the (1) right precuneus cortex, superior parietal cortex, and isthmus-cingulate cortex; (2) right precuneus cortex, superior parietal cortex, and paracentral lobule; (3) right superior parietal cortex, supramarginal gyrus, and inferior parietal cortex; and (4) left entorhinal cortex. In contrast, CV was lower in Increasers vs Decreasers in clusters including the (5) right superior frontal gyrus; (6) right superior temporal gyrus and the banks of the superior temporal sulcus; (7) left lateral occipital cortex and inferior parietal cortex; and (8) right lateral occipital cortex.

Also, Increasers and Decreasers differed in CT in three clusters (Figure2b). CT was greater in Increasers vs Decreasers in clusters including the (1) left insula, entorhinal cortex, inferior temporal gyrus, temporal pole, and superior temporal gyrus; and (2) right precuneus. CT was lower in Increasers vs Decreasers in a cluster encompassing the left lateral occipital cortex and fusiform gyrus.

Moreover, Increasers and Decreasers differed in SA in five clusters (Figure2c). SA was greater in Increasers vs Decreasers in clusters including the (1) right postcentral gyrus, supramarginal gyrus, precentral gyrus, pars opercularis, insula, and rostral middle frontal gyrus; (2) right precuneus cortex, paracentral lobule, and superior parietal cortex; (3) left inferior temporal gyrus, fusiform gyrus, and parahippocampal gyrus; and (4) left supramarginal gyrus and postcentral gyrus. In contrast, SA was lower in Increasers vs Decreasers in a cluster containing the right lingual gyrus and isthmus-cingulate cortex.

Increasers and No-changers differed in CV, which was larger in the Increasers vs No-changers in a cluster containing the right supramarginal gyrus and postcentral gyrus (Figure2d). Moreover, SA was significantly greater in the No-changers vs Increasers in two clusters containing the (1) right lingual gyrus, isthmus-cingulate

cortex, pericalcarine cortex, and parahippocampal gyrus; and (2) left lingual gyrus, lateral occipital cortex, pericalcarine cortex, cuneus cortex, and isthmus-cingulate cortex (Figure2e). Subgroups did not differ in CT.

Decreasers and No-changers differed in CV in three clusters (Figure2f). CV was greater in Decreasers vs No-changers in clusters including the (1) right posterior-cingulate cortex; (2) left superior frontal and rostral middle frontal gyrus; and (3) right superior frontal gyrus. Subgroups did not differ in CT. However, SA was greater in Decreasers vs No-changers in clusters including the (1) left superior parietal cortex and postcentral gyrus; and (2) left precuneus cortex, lingual gyrus, isthmus-cingulate cortex, and pericalcarine cortex (Figure2g).

Supplementary references

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. DSM-5, 5th edn. In: American Psychiatric Association, editor. Washington, DC: American Psychiatric Association,; 2013.
2. World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. In: World Health Organization, editor. Geneva: World Health Organization; 1992.
3. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry. 2008;47(8):921-9.
4. Wong AY, Hsia Y, Chan EW, Murphy DG, Simonoff E, Buitelaar JK, et al. The variation of psychopharmacological prescription rates for people with autism spectrum disorder (ASD) in 30 countries. Autism Res. 2014;7(5):543-54.
5. Frazier TW, Shattuck PT, Narendorf SC, Cooper BP, Wagner M, Spitznagel EL. 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 Adolesc Psychopharmacol. 2011;21(6):571-9.
6. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 1999;9(2):179-94.
7. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774-81.
8. Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 1999;8(4):272-84.
9. Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, et al. A hybrid approach to the skull stripping problem in MRI. Neuroimage. 2004;22(3):1060-75.
10. Winkler AM, Sabuncu MR, Yeo BTT, Fischl B, Greve DN, Kochunov P, et al. Measuring and comparing brain cortical surface area and other areal quantities. Neuroimage. 2012;61(4):1428-43.
11. Liem F, Merillat S, Bezzola L, Hirsiger S, Philipp M, Madhyastha T, et al. Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. Neuroimage. 2015;108:95-109.
12. Sparrow SS, Balla DA, Cicchetti DV. Vineland II: Vineland adaptive behavior scales. American Guidance Service Publishing 2005.
13. Sparrow SS. Vineland Adaptive Behavior Scales. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011.
14. Chatham CH, Taylor KI, Charman T, Liogier D'ardhuy X, Eule E, Fedele A, et al. Adaptive behavior in autism: Minimal clinically important differences on the Vineland-II. Autism Res. 2018;11(2):270-83.
15. Health USDo, Human Services FDACfDE, Research, Health USDo, Human Services FDACfBE, Research, et al. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79.
16. Ecker C, Ginestet C, Feng Y, Johnston P, Lombardo MV, Lai MC, et al. Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. JAMA Psychiatry. 2013;70(1):59-70.
17. Ecker C, Marquand A, Mourao-Miranda J, Johnston P, Daly EM, Brammer MJ, et al. Describing the brain in autism in five dimensions--magnetic resonance imaging-assisted

diagnosis of autism spectrum disorder using a multiparameter classification approach. J Neurosci. 2010;30(32):10612-23.
18. Ecker C, Murphy D. Neuroimaging in autism--from basic science to translational research. Nat Rev Neurol. 2014;10(2):82-91.
19. Ecker C, Schmeisser MJ, Loth E, Murphy DG. Neuroanatomy and Neuropathology of Autism Spectrum Disorder in Humans. Adv Anat Embryol Cell Biol. 2017;224:27-48.
20. Rakic P. Specification of cerebral cortical areas. Science. 1988;241(4862):170-6.
21. Hazlett HC, Poe MD, Gerig G, Styner M, Chappell C, Smith RG, et al. Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. Arch Gen Psychiatry. 2011;68(5):467-76.
22. Panizzon MS, Fennema-Notestine C, Eyler LT, Jernigan TL, Prom-Wormley E, Neale M, et al. Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. Cerebral Cortex. 2009;19(11):2728-35.
23. Storsve AB, Fjell AM, Tamnes CK, Westlye LT, Overbye K, Aasland HW, et al. Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. J Neurosci. 2014;34(25):8488-98.
24. Im K, Lee JM, Lyttelton O, Kim SH, Evans AC, Kim SI. Brain size and cortical structure in the adult human brain. Cereb Cortex. 2008;18(9):2181-91.
25. Hogstrom LJ, Westlye LT, Walhovd KB, Fjell AM. The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. Cereb Cortex. 2013;23(11):2521-30.
26. Ecker C, Shahidiani A, Feng Y, Daly E, Murphy C, D'Almeida V, et al. The effect of age, diagnosis, and their interaction on vertex-based measures of cortical thickness and surface area in autism spectrum disorder. J Neural Transm (Vienna). 2014;121(9):1157-70.
27. Raznahan A, Shaw P, Lalonde F, Stockman M, Wallace GL, Greenstein D, et al. How does your cortex grow? J Neurosci. 2011;31(19):7174-7.
28. Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC. Detecting changes in nonisotropic images. Hum Brain Mapp. 1999;8(2-3):98-101.
29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.
30. Joseph RM, Fein D. The significance of IQ and differential cognitive abilities for understanding ASD. The neuropsychology of autism. 2011;281:294.
31. Pfefferbaum A, Rohlfing T, Rosenbloom MJ, Chu W, Colrain IM, Sullivan EV. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85 years) measured with atlas-based parcellation of MRI. Neuroimage. 2013;65:176-93.
32. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, DiLavore PC, 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(3):205-23.
33. Constantino JN, Gruber CP. Social responsiveness scale : SRS-2. Second edition ed. Torrane, California: Western Psychological Services; 2012. v, 106 pages p.
34. Oakland T. Adaptive Behavior Assessment System – Second Edition. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer New York; 2011. p. 37-9.
35. Hopp CA, Baron IS. AAMD Adaptive Behavior Scales. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. Cham: Springer International Publishing; 2018. p. 2-4.

36. Hogan AE, Scott KG, Bauer CR. The Adaptive Social Behavior Inventory (ASBI): A new assessment of social competence in high-risk three-year-olds. Journal of Psychoeducational Assessment. 1992;10(3):230-9.
37. Balboni G, Tasse MJ, Schalock RL, Borthwick-Duffy SA, Spreat S, Thissen D, et al. The diagnostic adaptive behavior scale: evaluating its diagnostic sensitivity and specificity. Res Dev Disabil. 2014;35(11):2884-93.
38. Price JA, Morris ZA, Costello S. The Application of Adaptive Behaviour Models: A Systematic Review. Behav Sci (Basel). 2018;8(1).
39. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489(7416):391-9.
40. Gorgolewski K. “Tight fitting genes: finding relations between statistical maps and gene expression patterns” presented at the 20th Annual Meeting of the Organization for Human Brain Mapping, Hamburg, Germany, 8-12 June 2014. 2014.
41. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):568-84 e23.
42. 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(6420).
43. Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. Science. 2019;364(6441):685-9.
44. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature. 2016;540(7633):423-7.
45. Richiardi J, Altmann A, Milazzo AC, Chang C, Chakravarty MM, Banaschewski T, et al. BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. Science. 2015;348(6240):1241-4.
46. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279-84.
47. Pramparo T, Pierce K, Lombardo MV, Carter Barnes C, Marinero S, Ahrens-Barbeau C, et al. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. JAMA Psychiatry. 2015;72(4):386-94.
48. Romero-Garcia R, Warrier V, Bullmore ET, Baron-Cohen S, Bethlehem RAI. Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. Mol Psychiatry. 2018.
49. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience. 2019;8(7).
50. 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(3):431-44.

Supplementary Figures

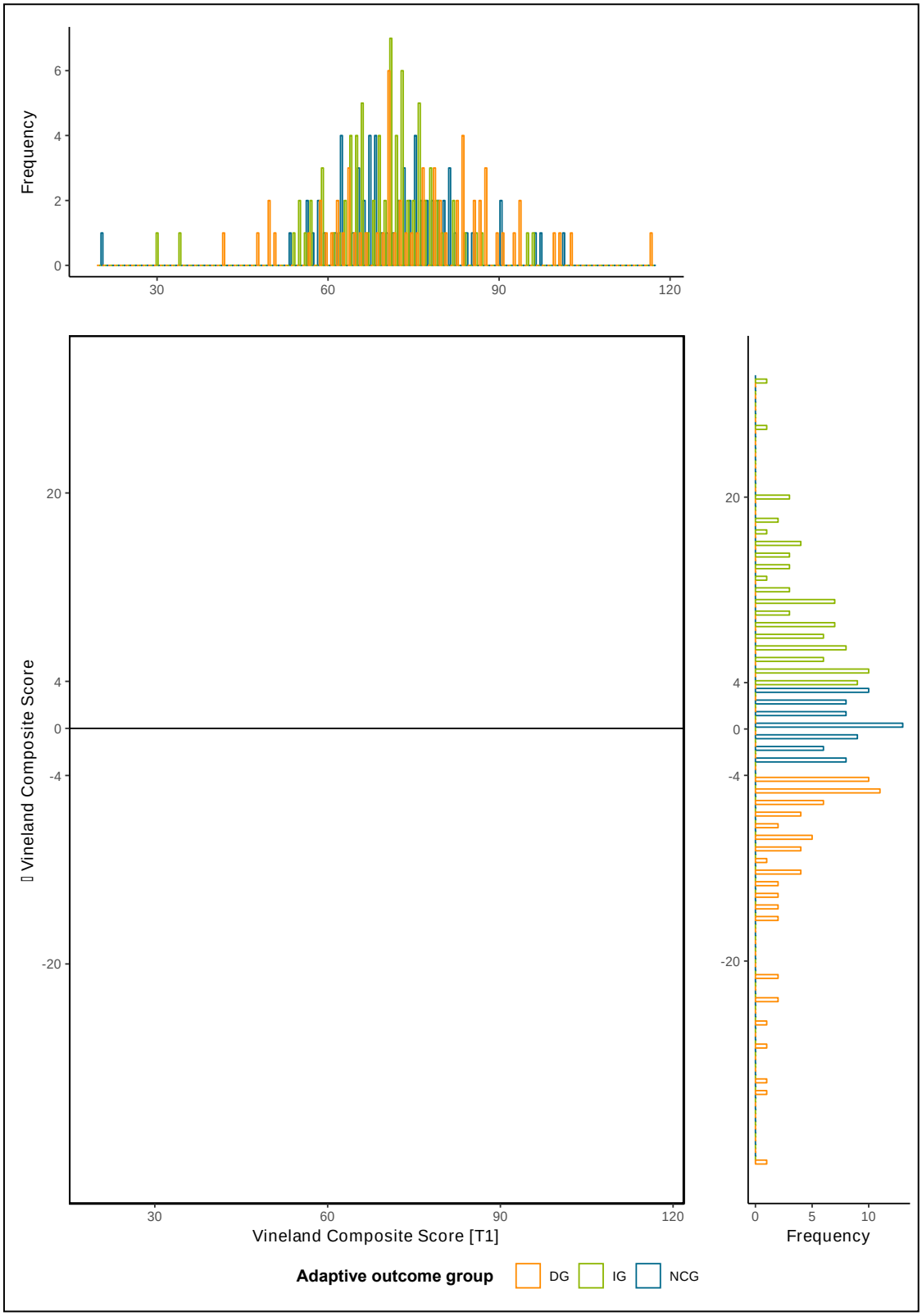

Figure S1 Adaptive behaviour measured using the Vineland-II scale. Middle: Vineland scores at timepoint 1 (T1) plotted against the change in Vineland scores (V_{T2}-V_{T1}) between T1 and timepoint 2 (T2). Changes of 4 units are marked in grey. Top: Histogram of Vineland scores at T1. Right: Histogram of Vineland change scores. Adaptive outcome subgroups are indicated through different colours: decrease group (DG) = orange, increase group (IG) = green, no change group (NCG) = blue.

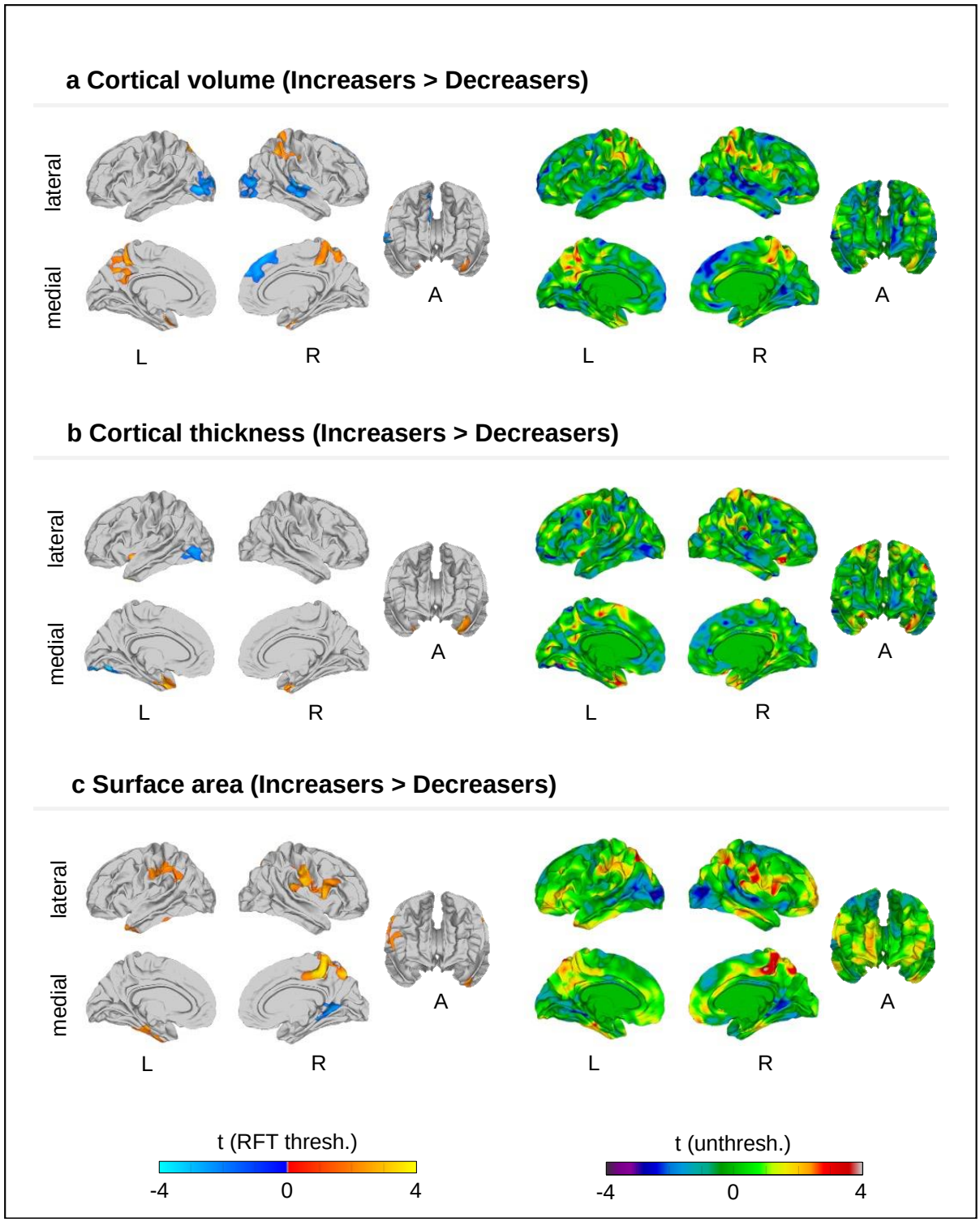

Figure S2 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers. Each row displays random field theory (RFT)-corrected t-values on the left and unthresholded t-values on the right. Abbreviations: A, anterior view; L, left; R, right.

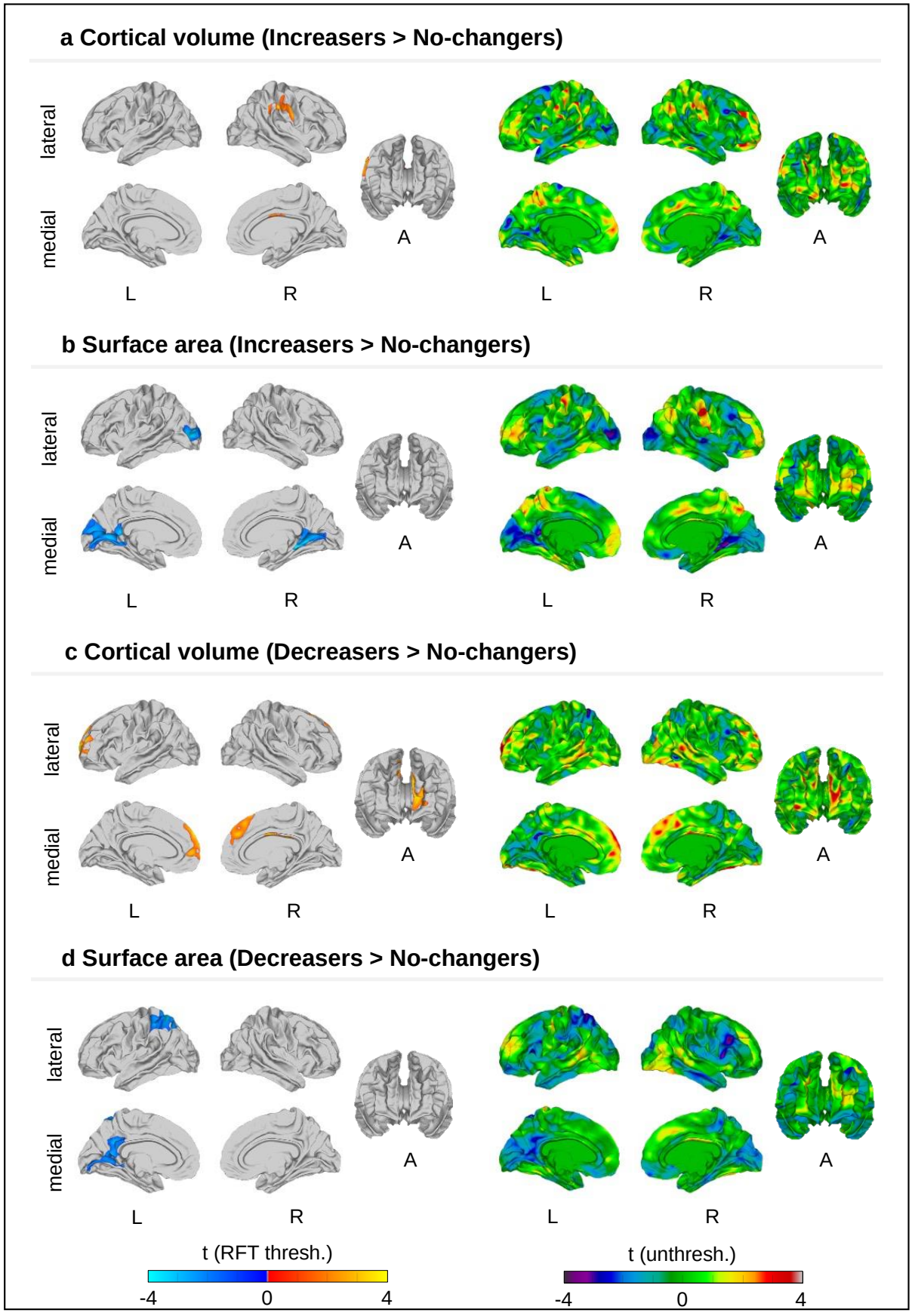

Figure S3 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers, No-changers, and Decreasers. Each row displays random field theory (RFT)-corrected t-values on the left and unthresholded t-values on the right. Abbreviations: A, anterior view; L, left; R, right.

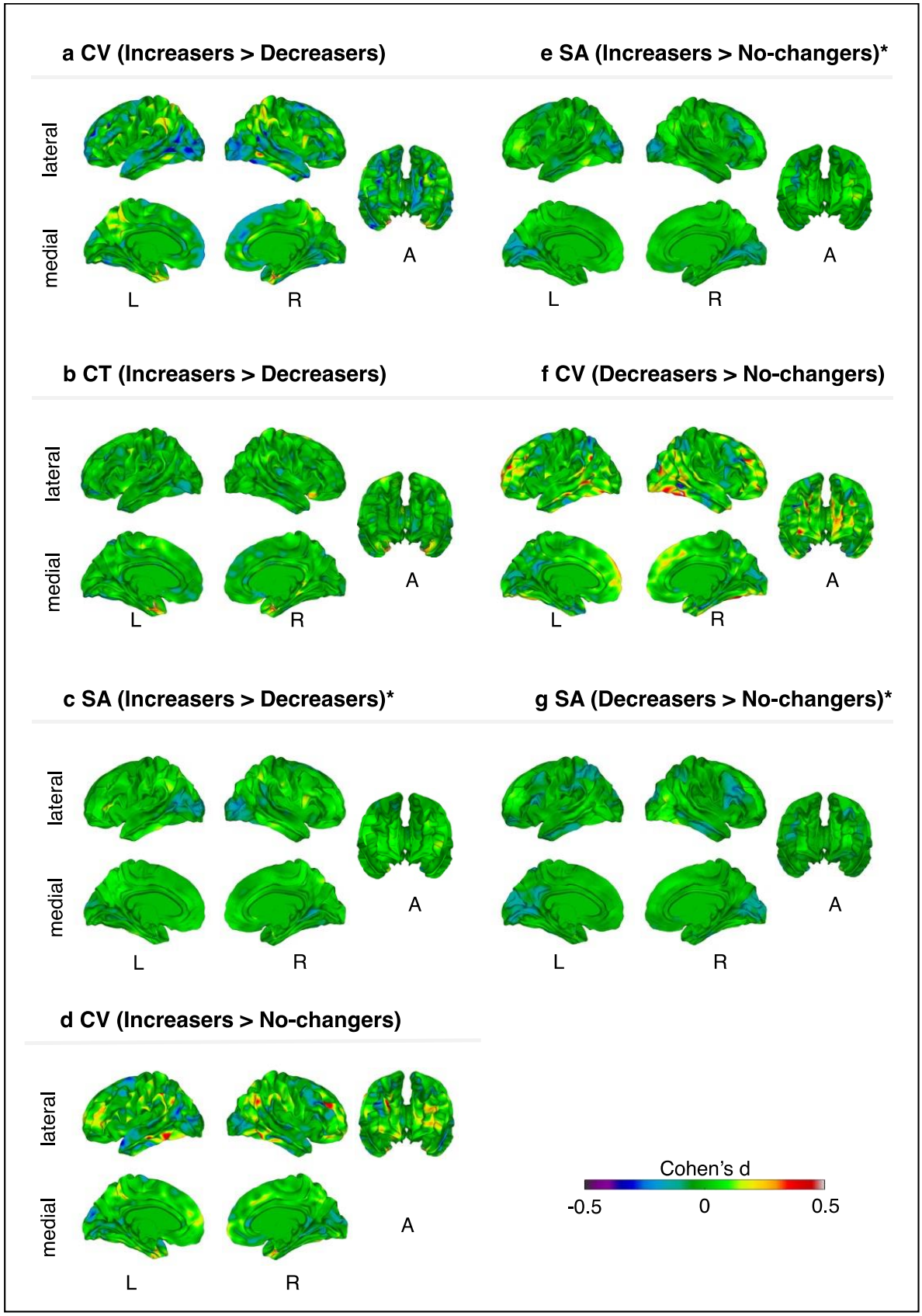

*Figure S4 Effect sizes (Cohen's d) of neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers, No-changers, and Decreasers. Effect sizes are indicated by the colorbar. Positive/negative values indicate a positive/negative effect in terms of the contrasts indicated in the subheadings. *Cohen's d range = [-0.2, 0.2], instead of [-0.5, 0.5]. Abbreviations: A, anterior view; CT, cortical thickness; CV, cortical volume; L, left; R, right; SA, surface area.*

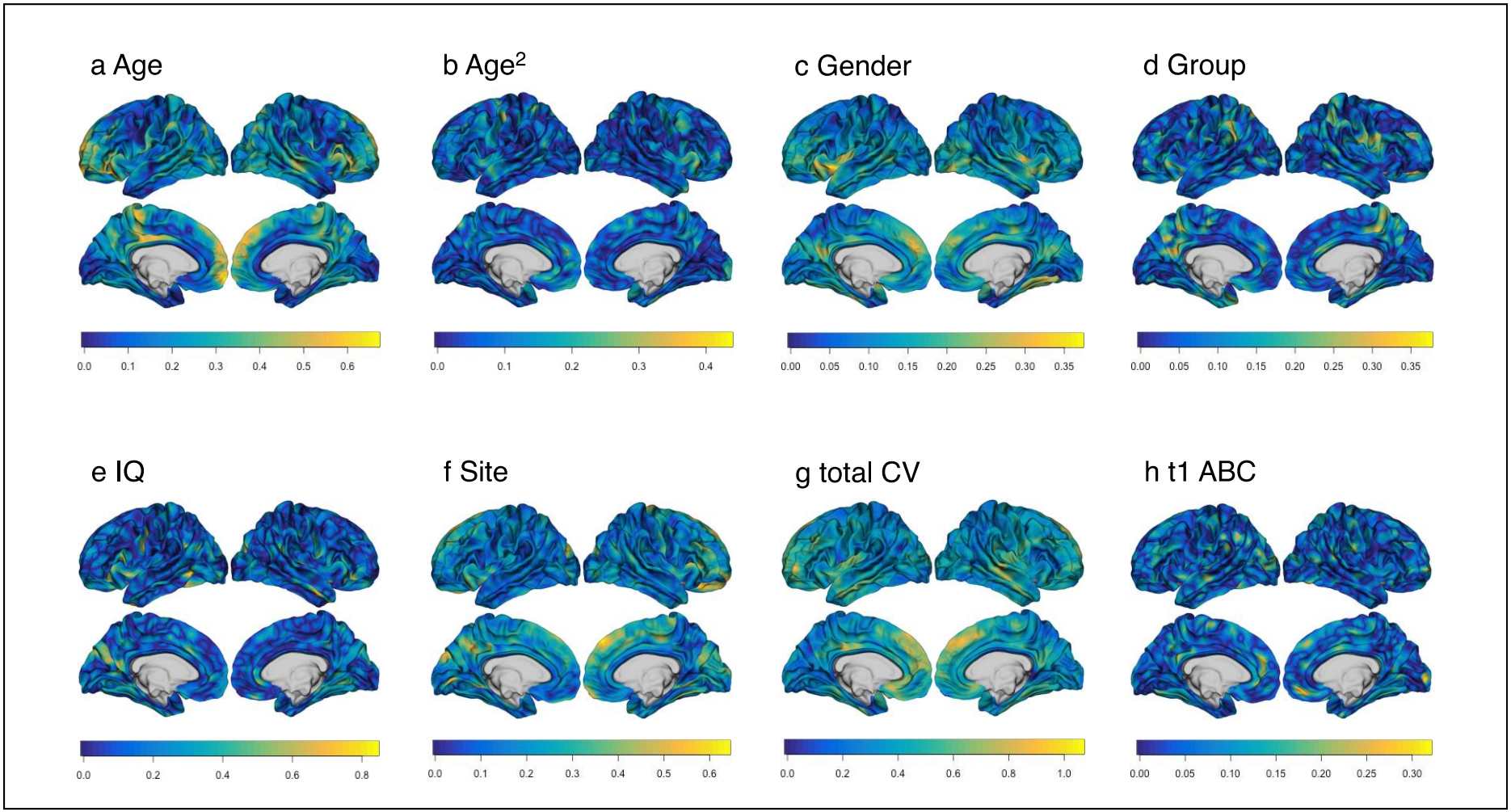

Figure S5 Effect sizes (Cohen's f) of each model term in the Increasers vs Decreasers contrast (cortical volume). Effect sizes are indicated by the colorbar.

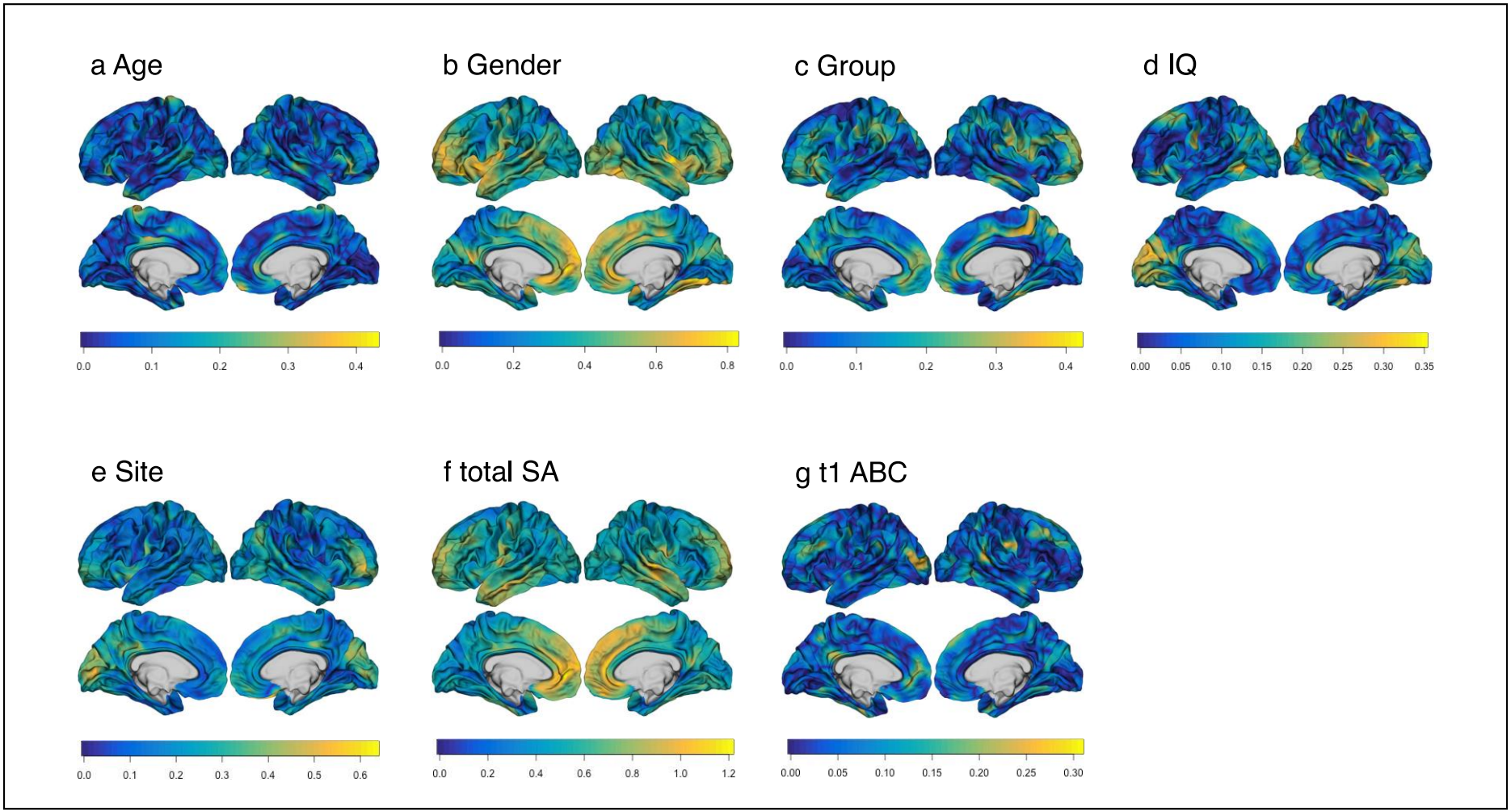

Figure S6 Effect sizes (Cohen's f) of each model term in the Increasers vs Decreasers contrast (surface area). Effect sizes are indicated by the colorbar.

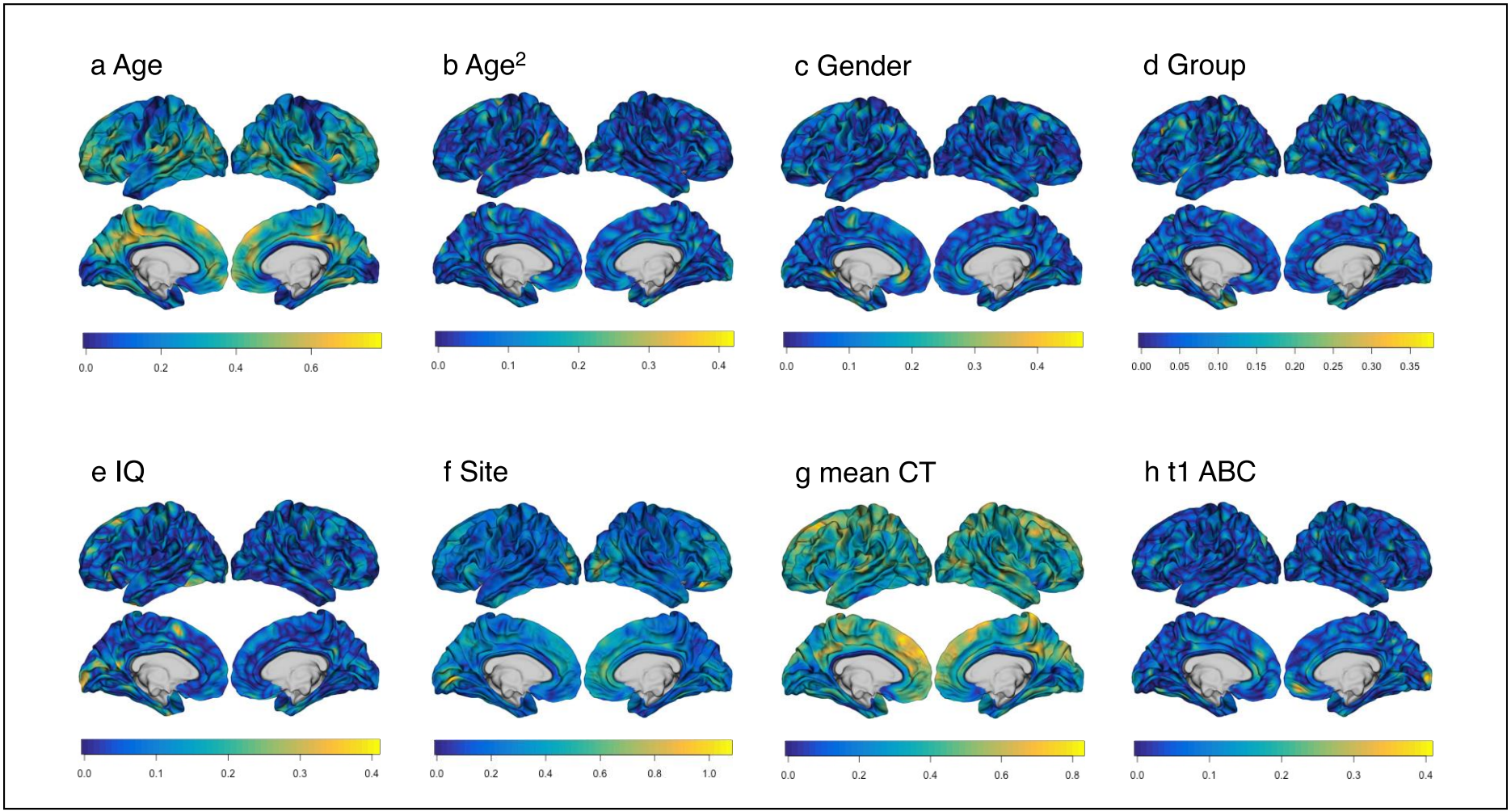

Figure S7 Effect sizes (Cohen's f) of each model term in the Increasers vs Decreasers contrast (cortical thickness). Effect sizes are indicated by the colorbar.

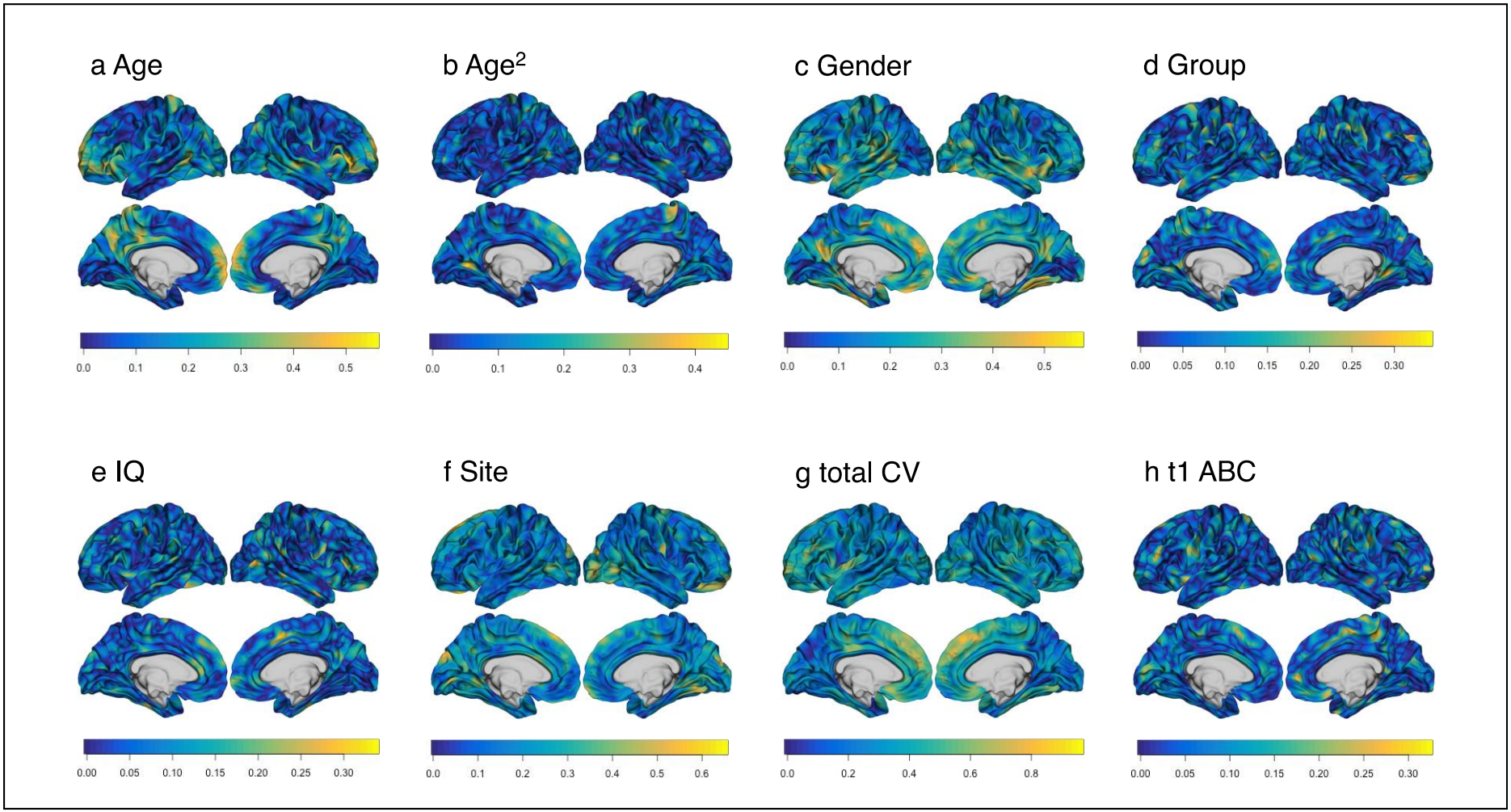

Figure S8 Effect sizes (Cohen's f) of each model term in the Increasers vs No-changers contrast (cortical volume). Effect sizes are indicated by the colorbar.

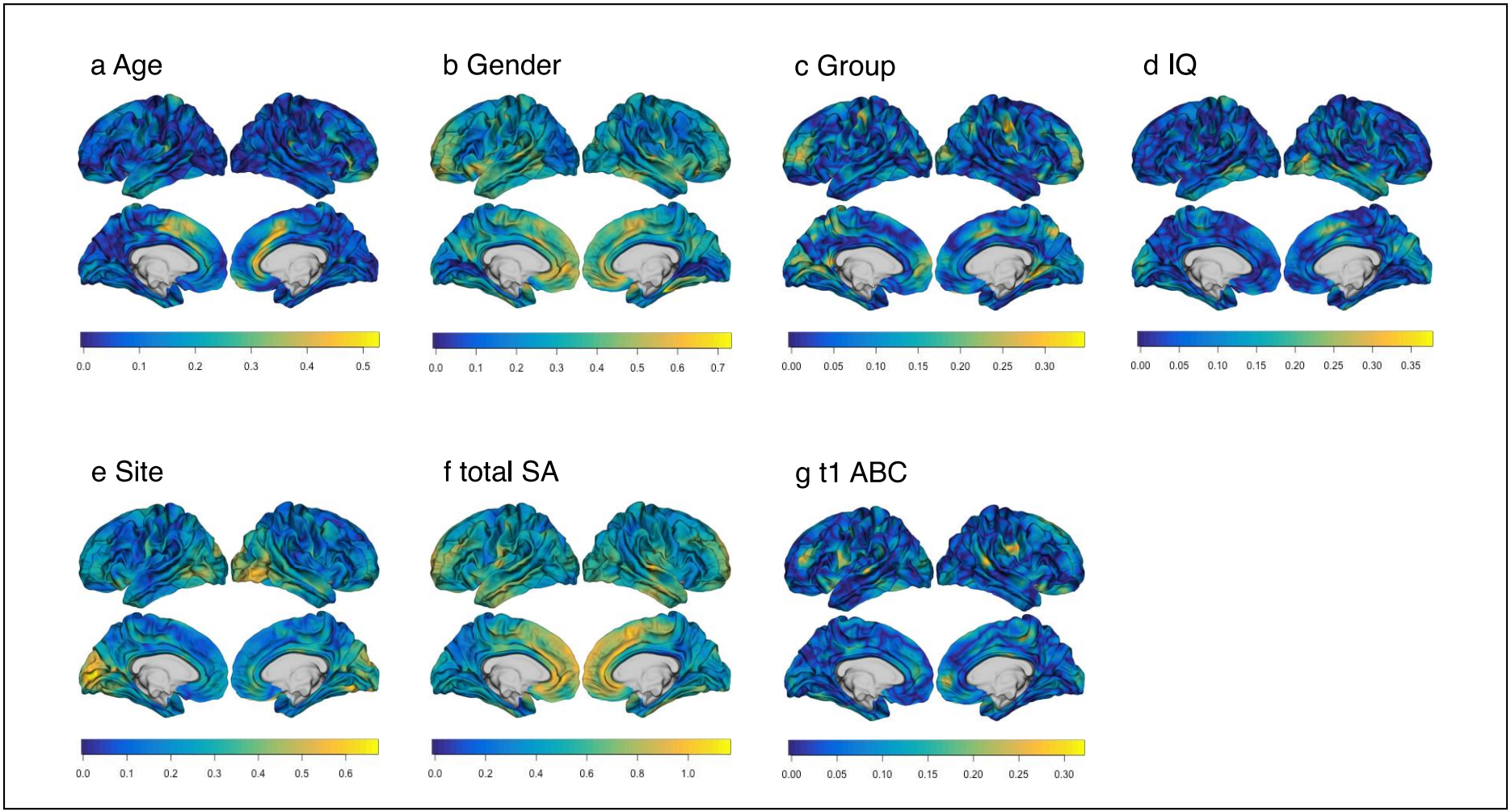

Figure S9 Effect sizes (Cohen's f) of each model term in the Increasers vs No-changers contrast (surface area). Effect sizes are indicated by the colorbar.

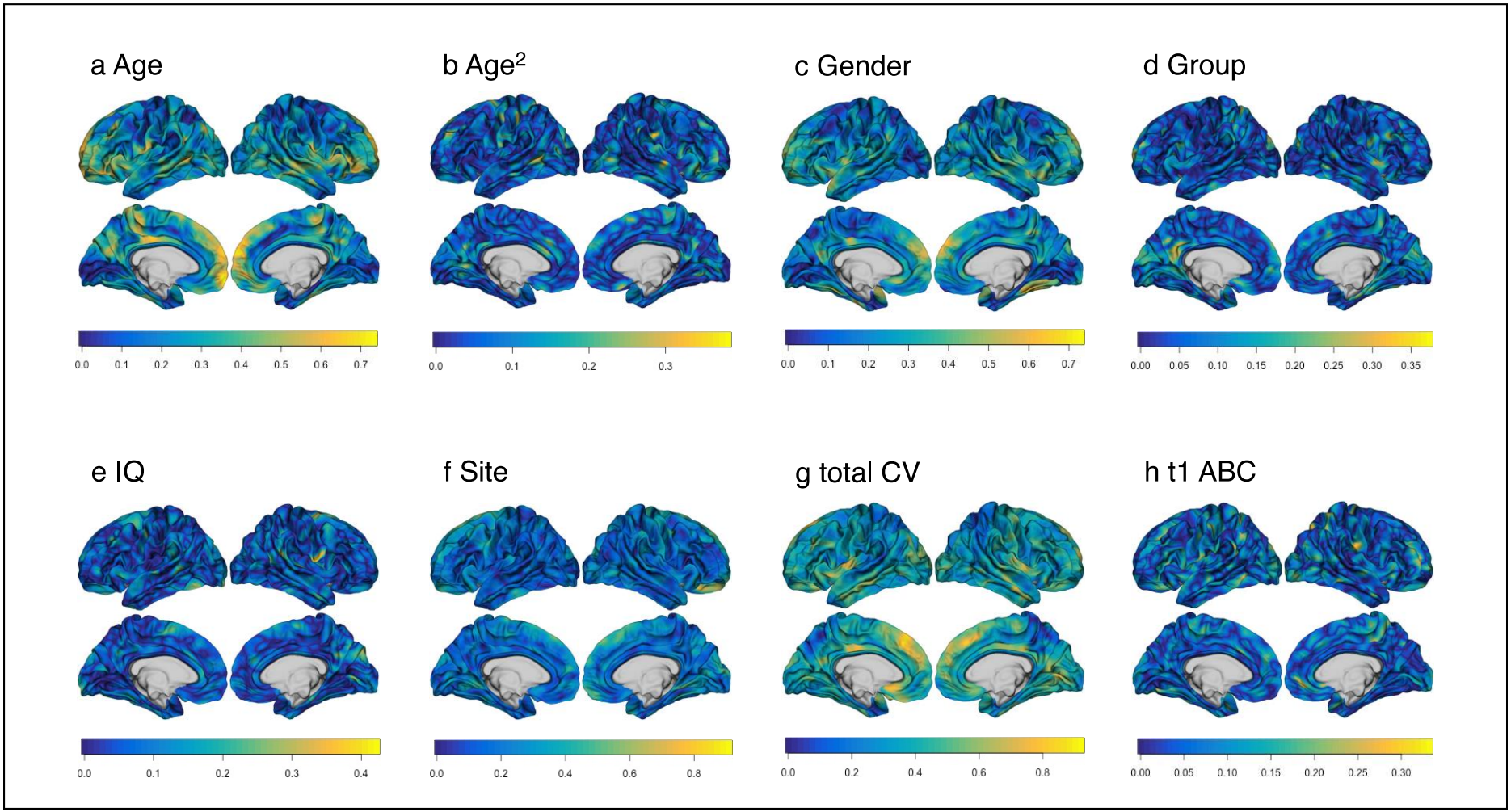

Figure S10 Effect sizes (Cohen's f) of each model term in the Decreasers vs No-changers contrast (cortical volume). Effect sizes are indicated by the colorbar.

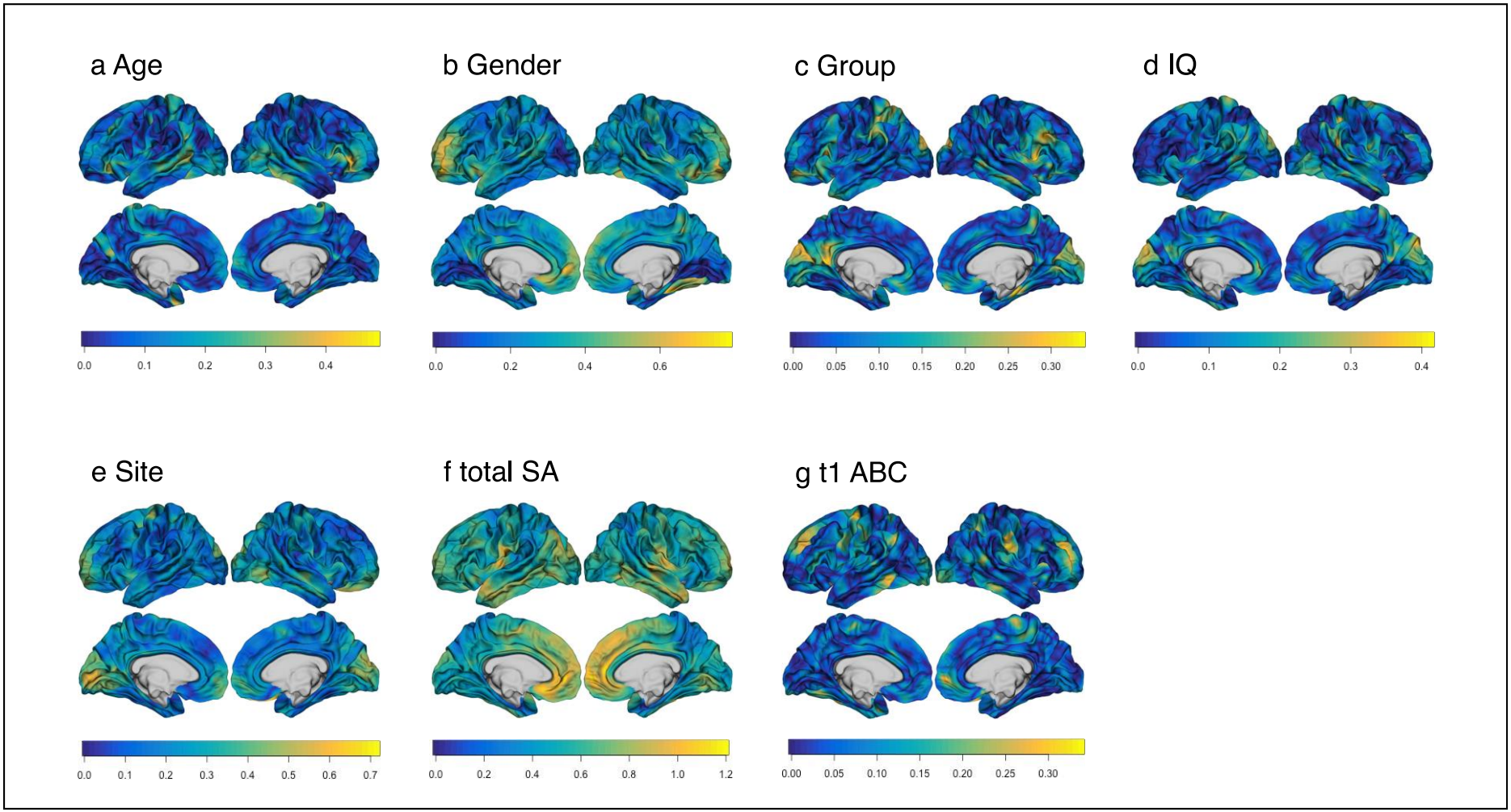

Figure S11 Effect sizes (Cohen's f) of each model term in the Decreasers vs No-changers contrast (surface area). Effect sizes are indicated by the colorbar.

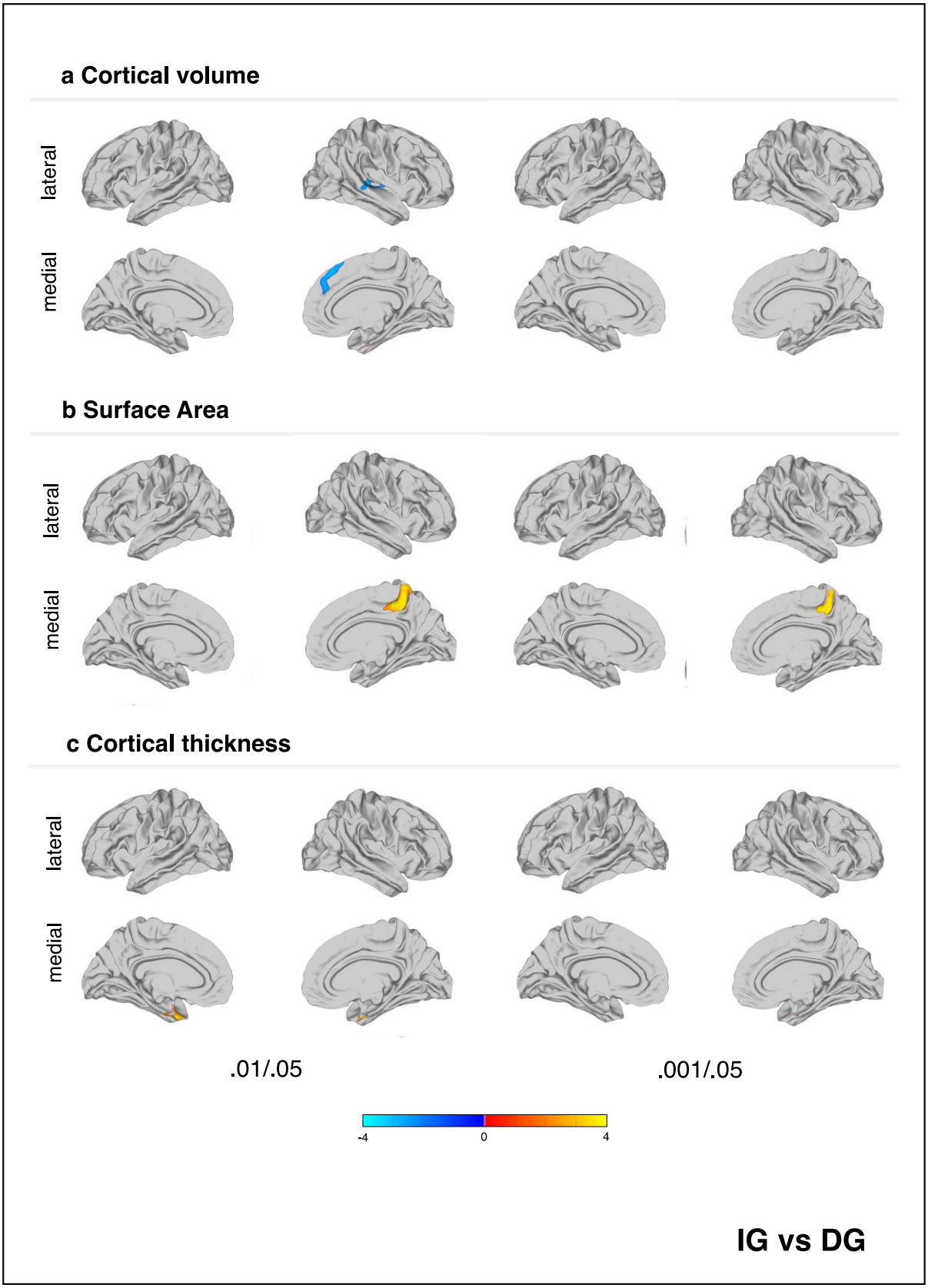

Figure S12 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers using varying statistical thresholds. Each row displays random field theory (RFT)-corrected t-values using a cluster-defining threshold of 0.01 and a cluster p-value threshold of 0.05 on the left; and a cluster-defining threshold of 0.001 and a cluster p-value threshold of 0.05 on the right. Abbreviations: L, left; R, right.

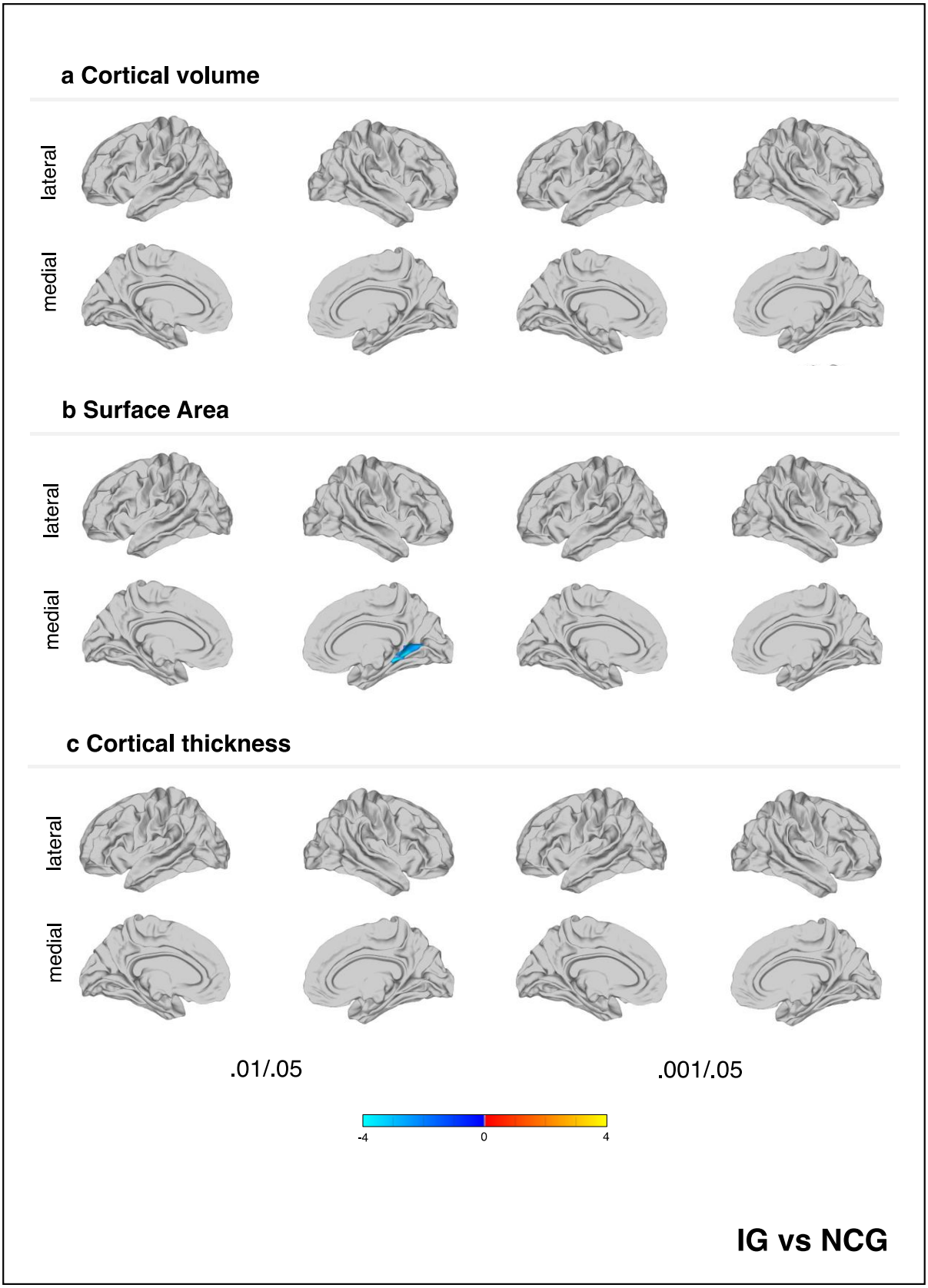

Figure S13 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and No-changers. Each row displays random field theory (RFT)-corrected t-values using a cluster-defining threshold of 0.01 and a cluster p-value threshold of 0.05 on the left; and a cluster-defining threshold of 0.001 and a cluster p-value threshold of 0.05 on the right. Abbreviations: L, left; R, right.

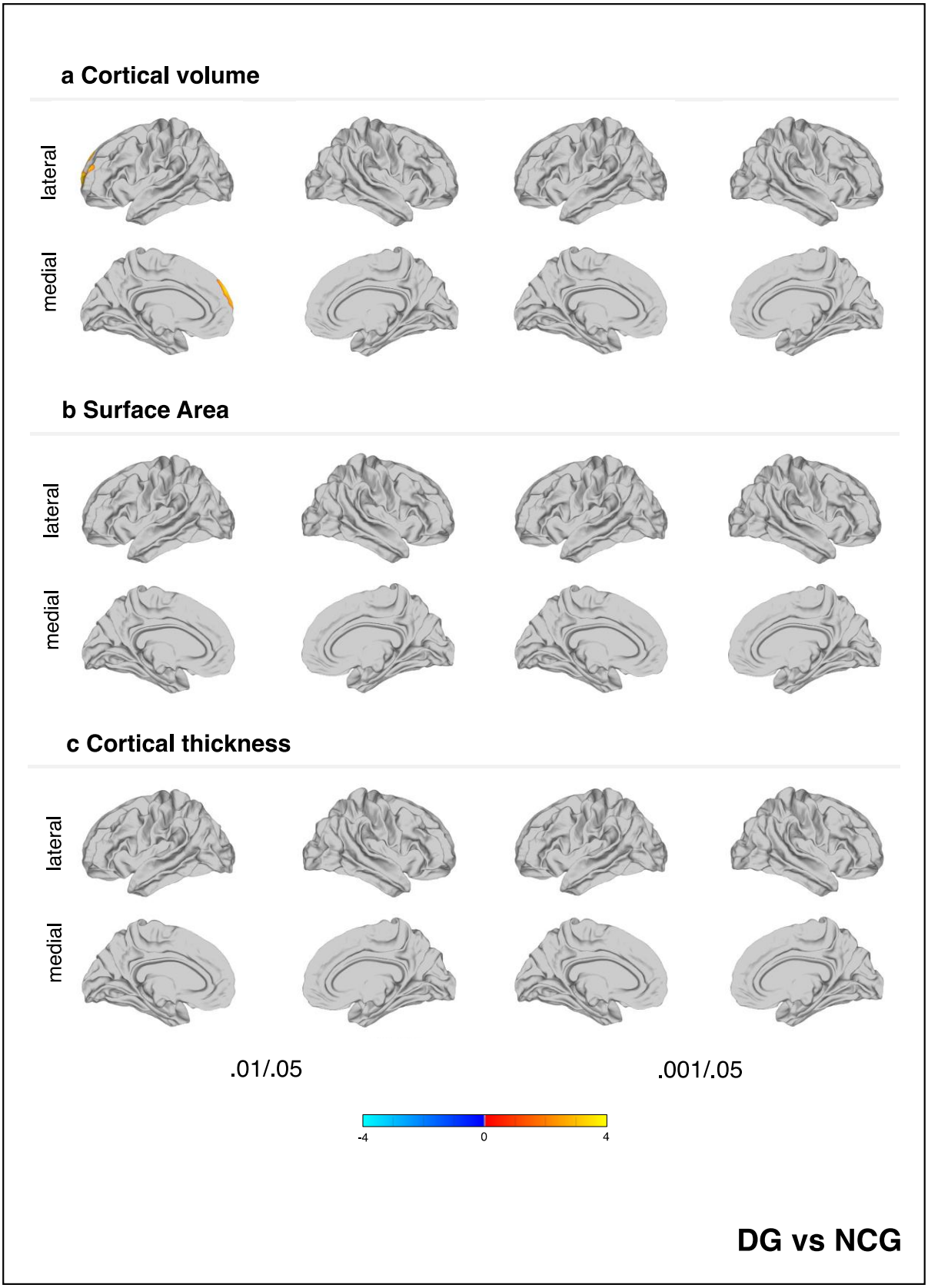

Figure S14 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Decreasers and No-changers. Each row displays random field theory (RFT)-corrected t-values using a cluster-defining threshold of 0.01 and a cluster p-value threshold of 0.05 on the left; and a cluster-defining threshold of 0.001 and a cluster p-value threshold of 0.05 on the right. Abbreviations: L, left; R, right.

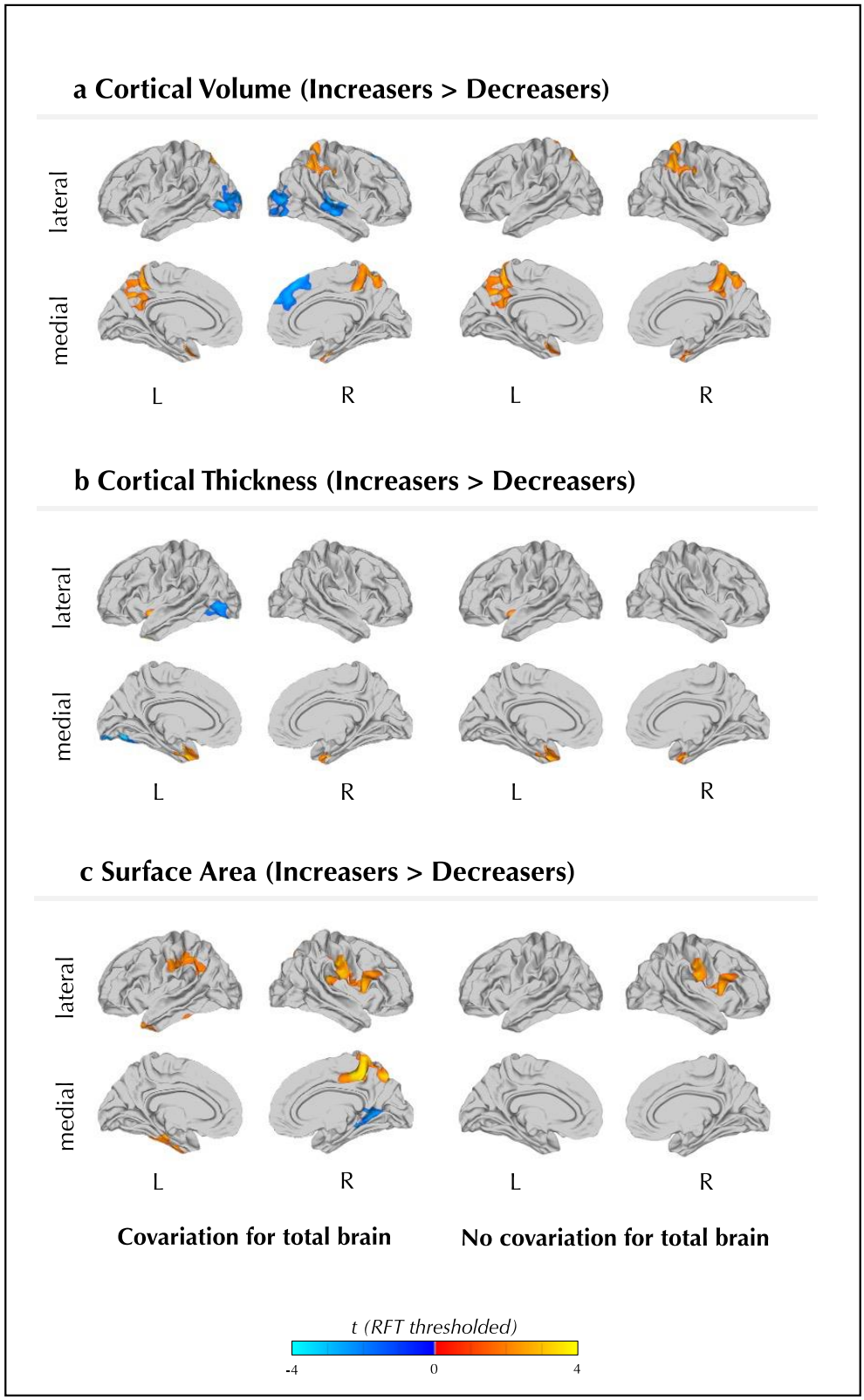

Figure S15 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers while covarying (left column) and not covarying (right column) for total brain. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

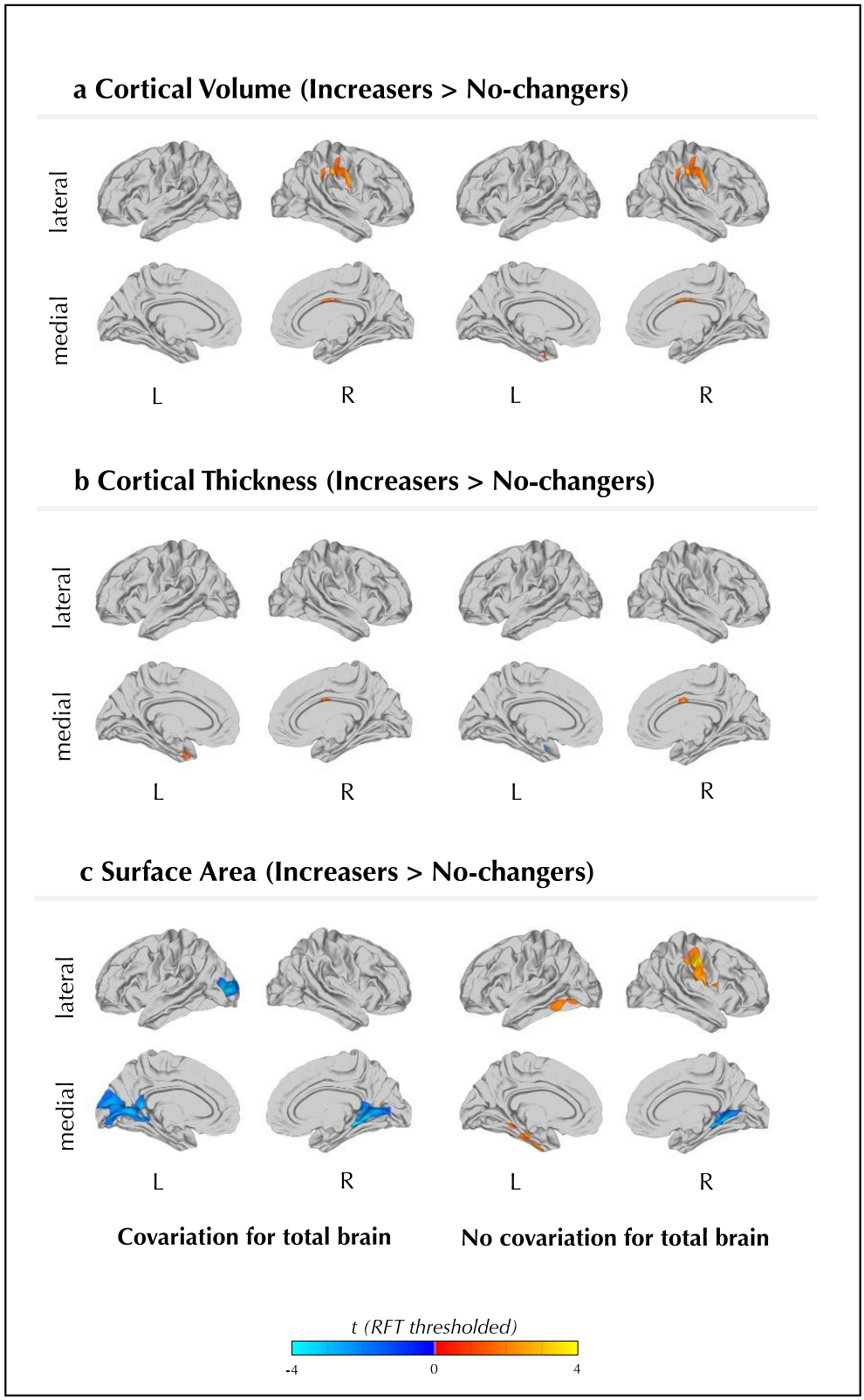

Figure S16 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and No-changers while covarying (left column) and not covarying (right column) for total brain. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

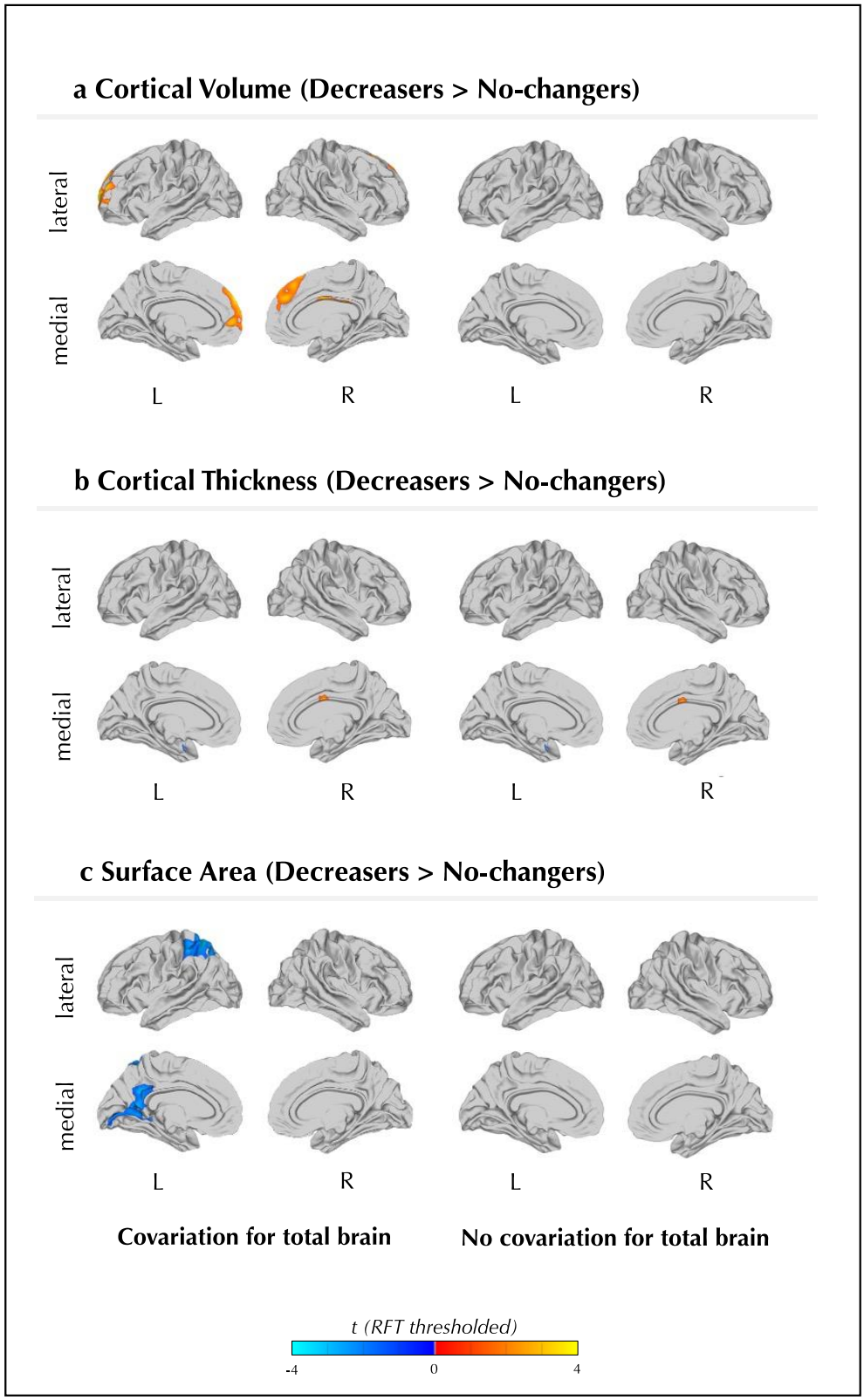

Figure S17 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Decreasers and No-changers while covarying (left column) and not covarying (right column) for total brain. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

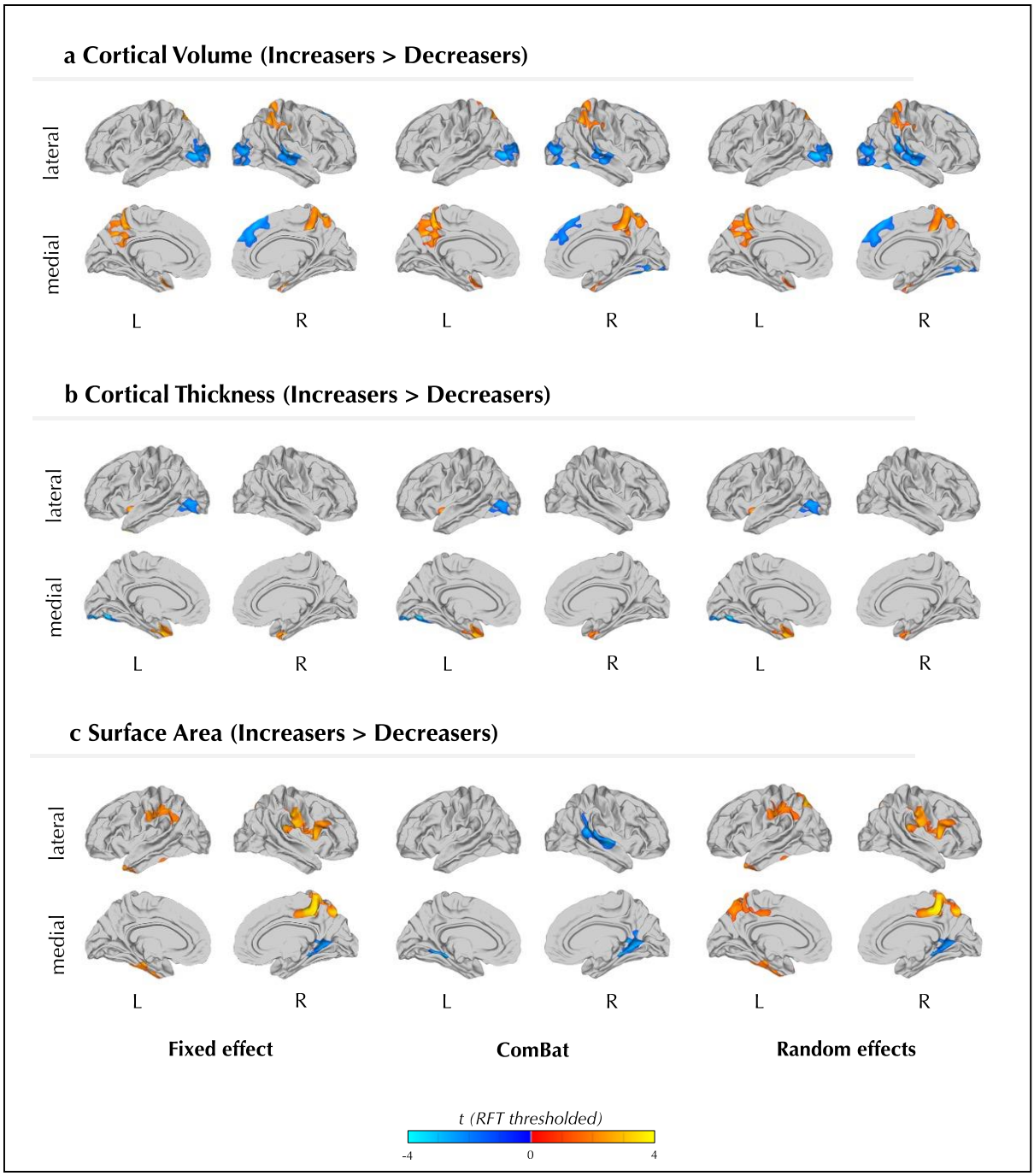

Figure S18 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers controlling for site using fixed-effect modelling (left column), ComBat batch harmonization (middle column), and random effect modelling (right column). Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

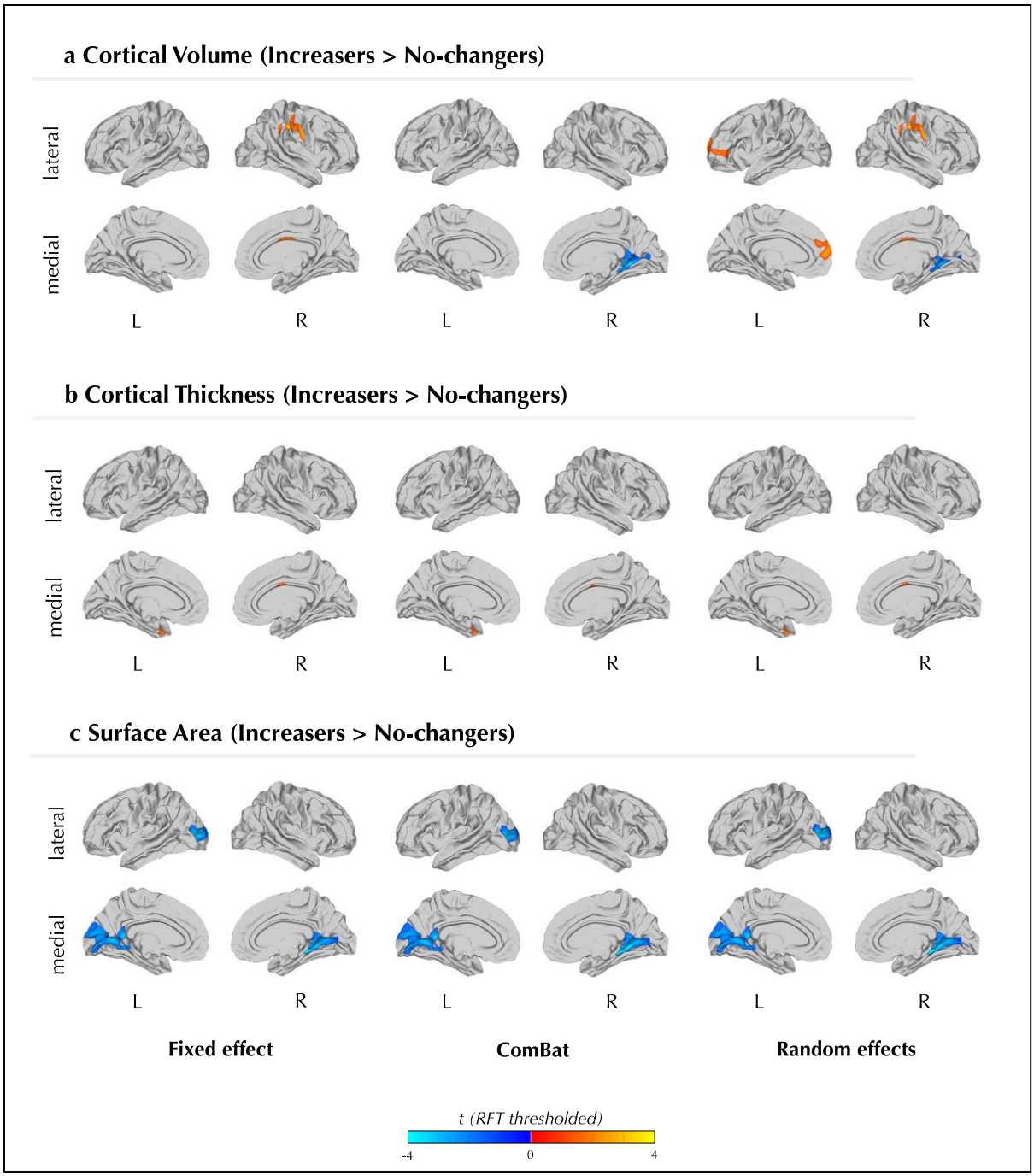

Figure S19 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and No-changers controlling for site using fixed-effect modelling (left column), ComBat batch harmonization (middle column), and random effect modelling (right column). Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

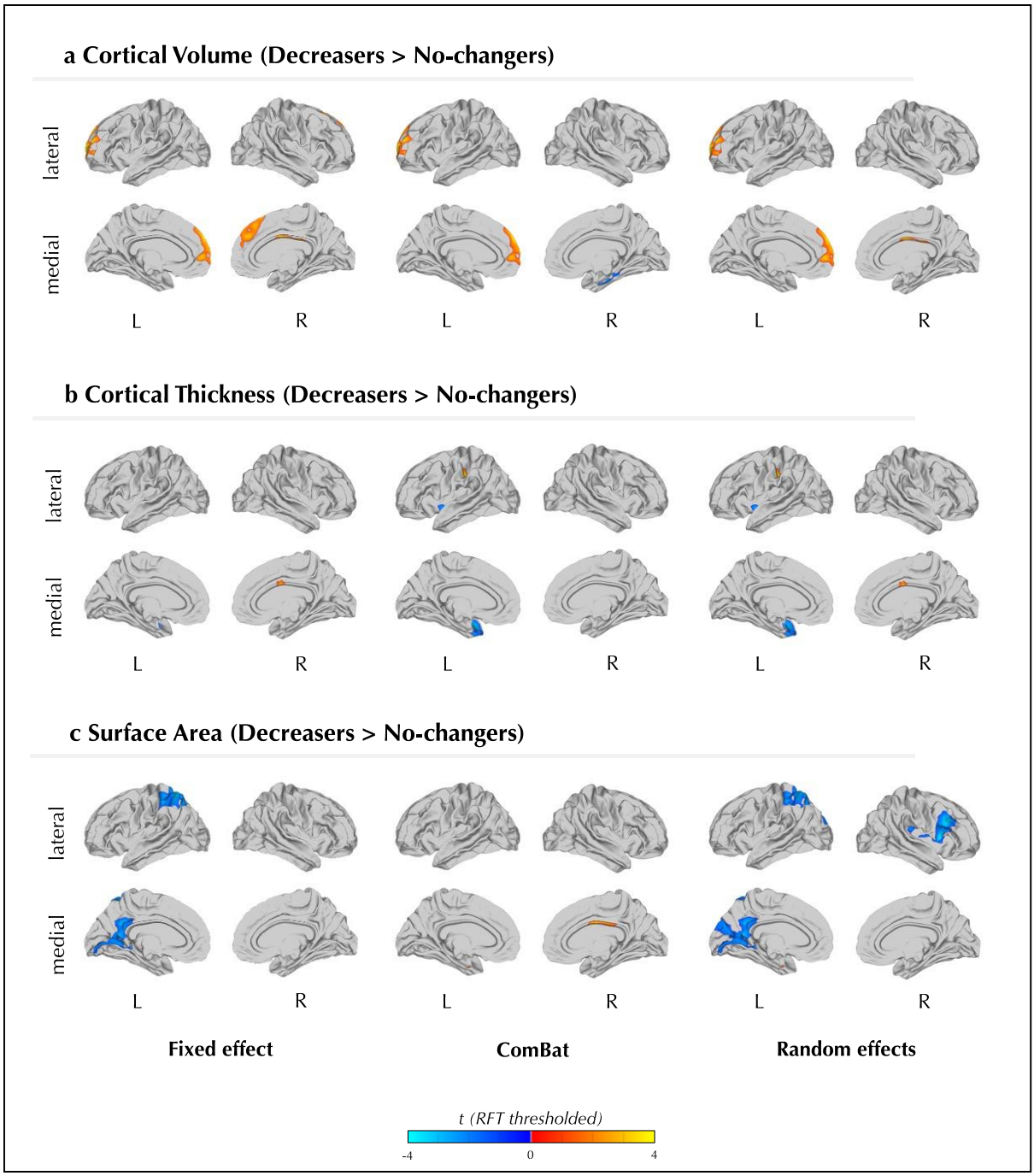

Figure S20 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Decreasers and No-changers controlling for site using fixed-effect modelling (left column), ComBat batch harmonization (middle column), and random effect modelling (right column). Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

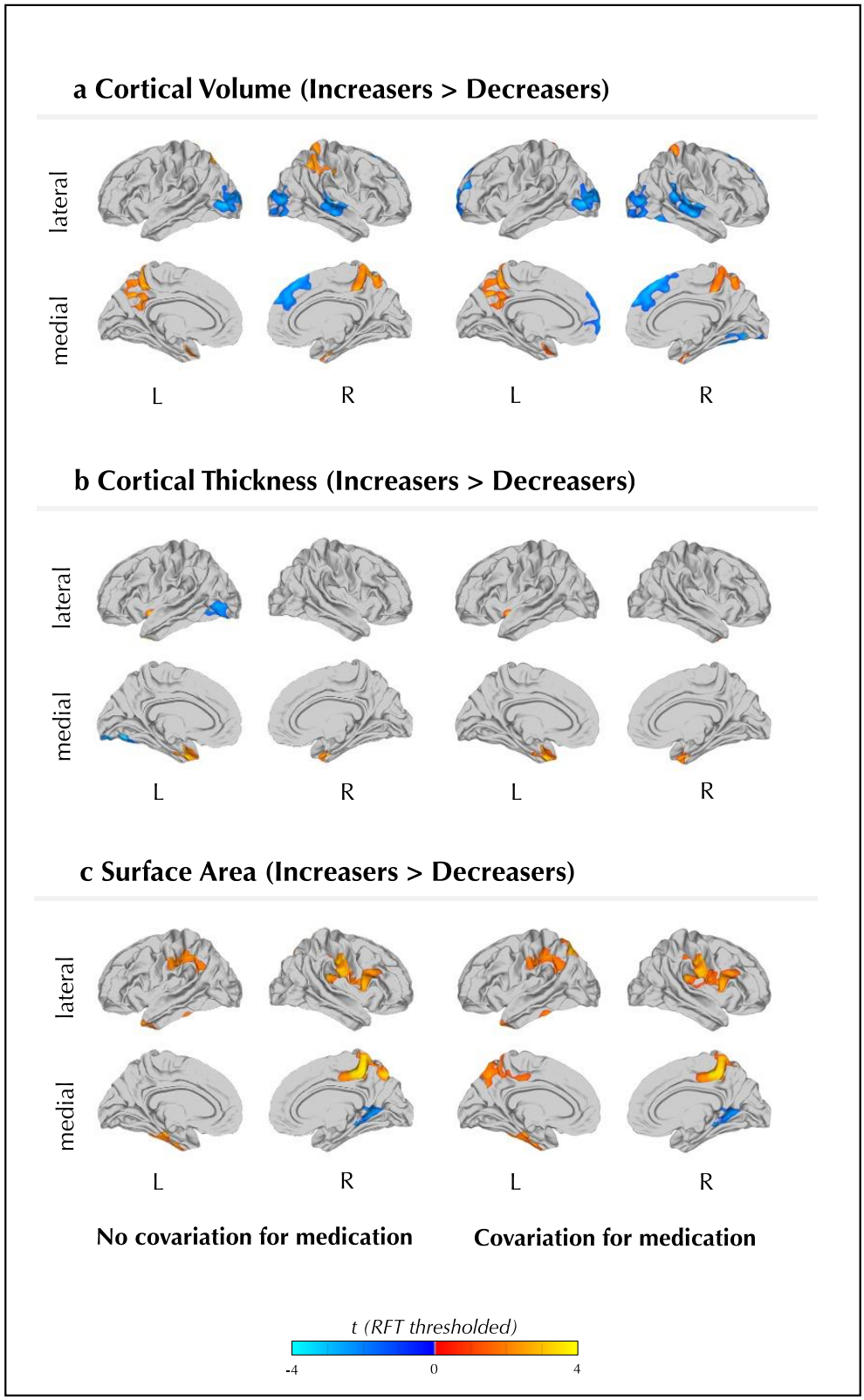

Figure S21 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers not covarying for (left column) and covarying for (right column) medication. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

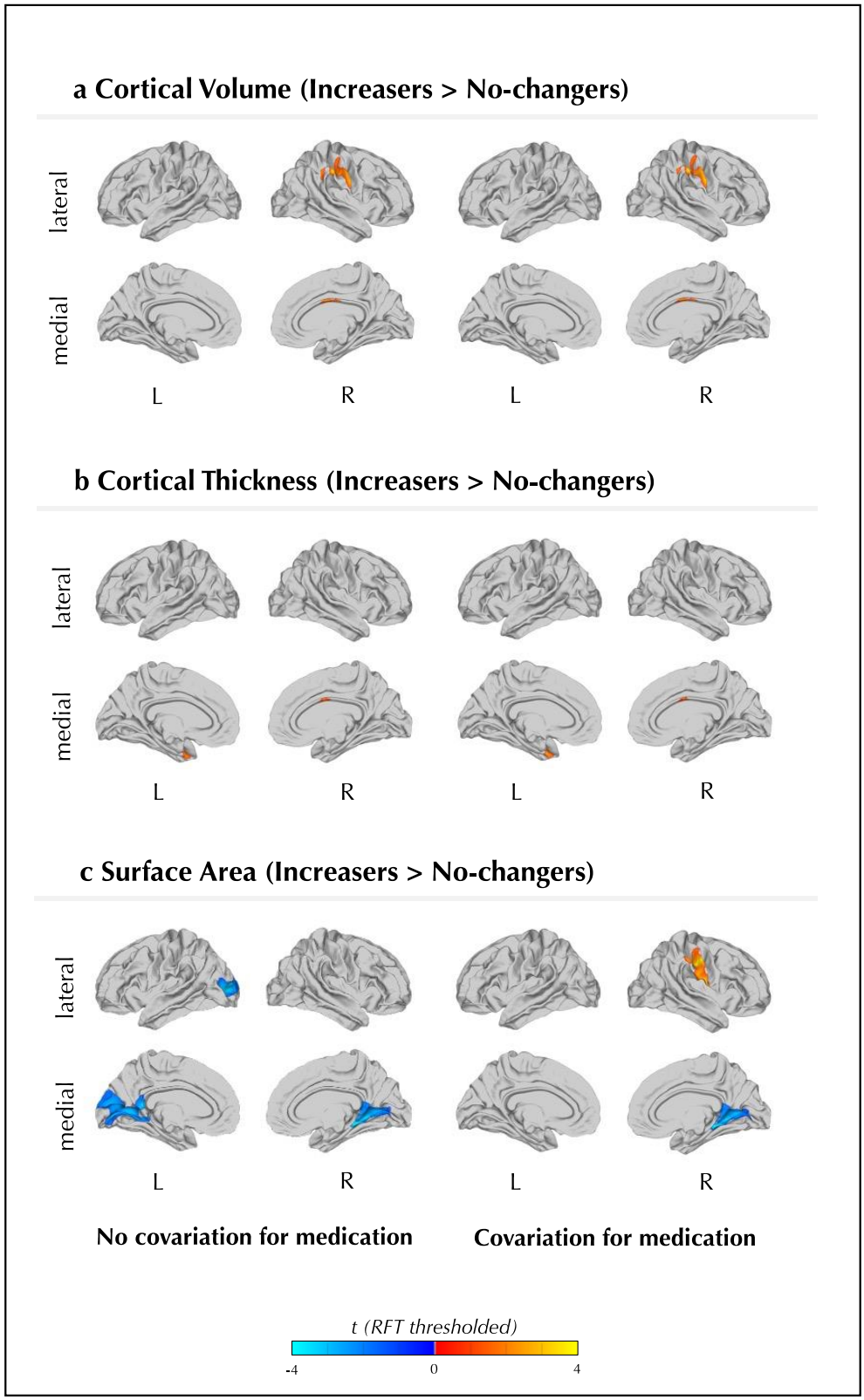

Figure S22 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and No-changers not covarying for (left column) and covarying for (right column) medication. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

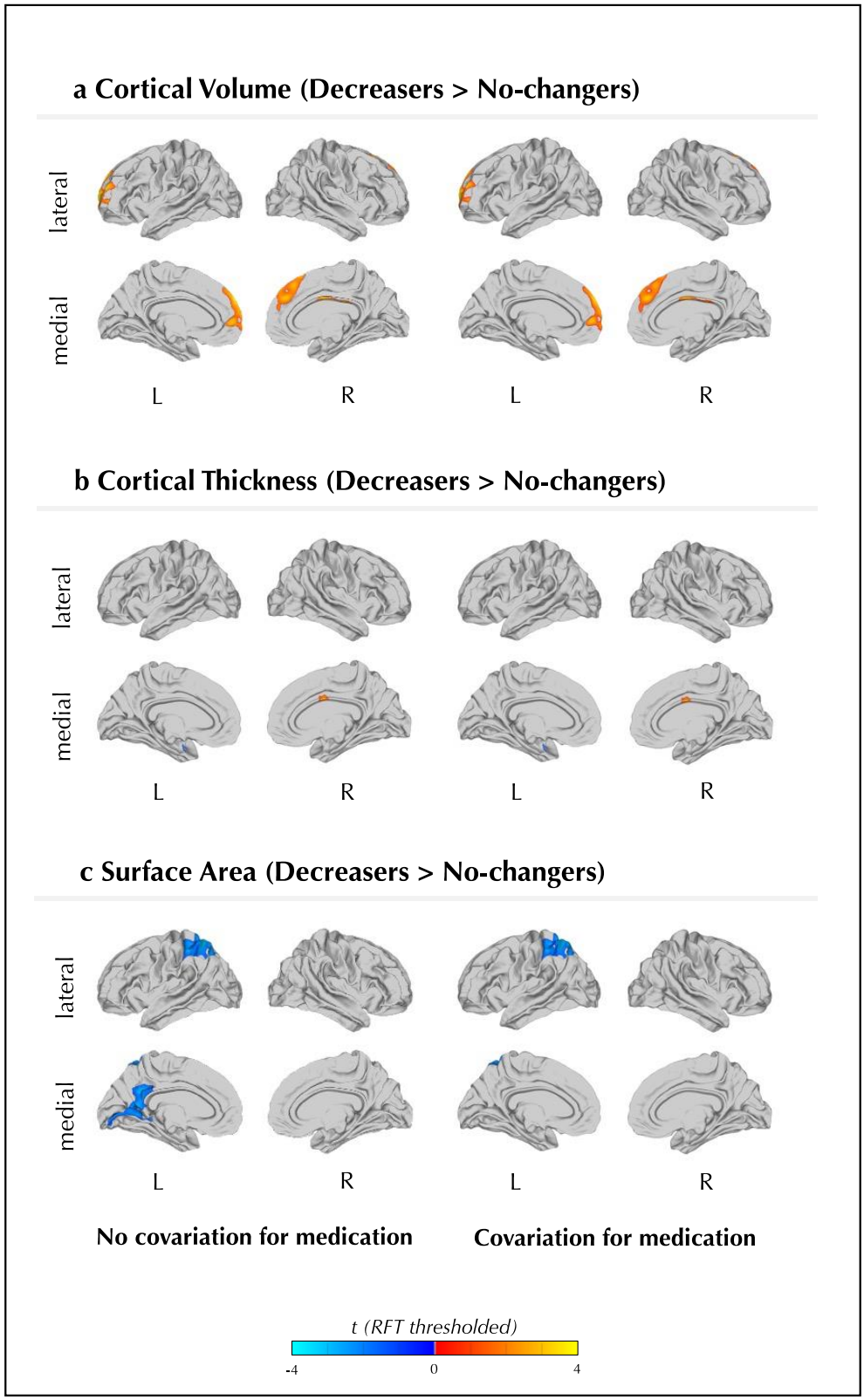

Figure S23 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Decreasers and No-changers not covarying for (left column) and covarying for (right column) medication. Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

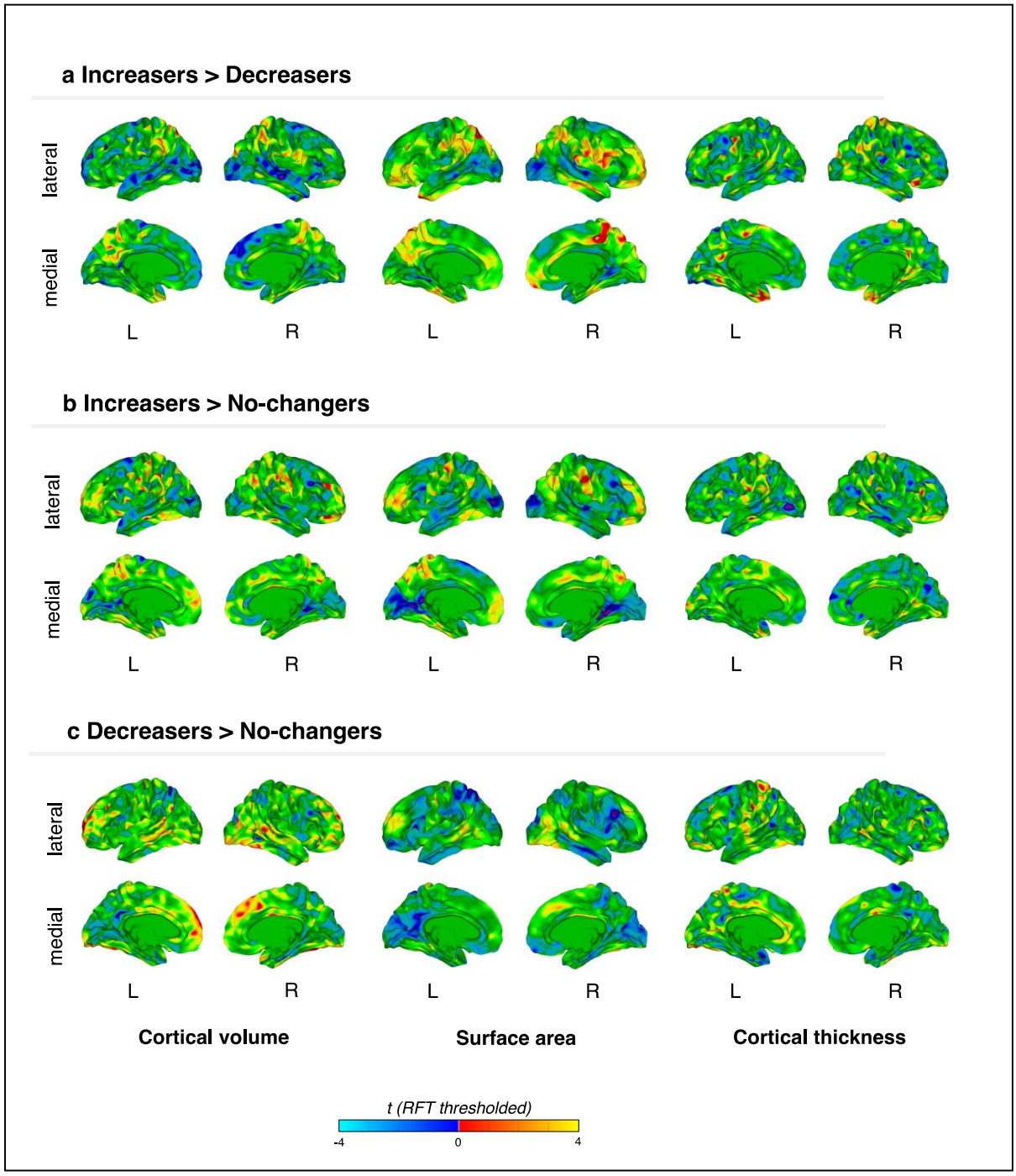

Figure S24 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers, No-changers, and Decreasers not covarying for intelligence quotient (IQ). Each row displays random field theory (RFT)-corrected t-values, as indicated through the colorbar. Abbreviations: A, anterior view; L, left; R, right.

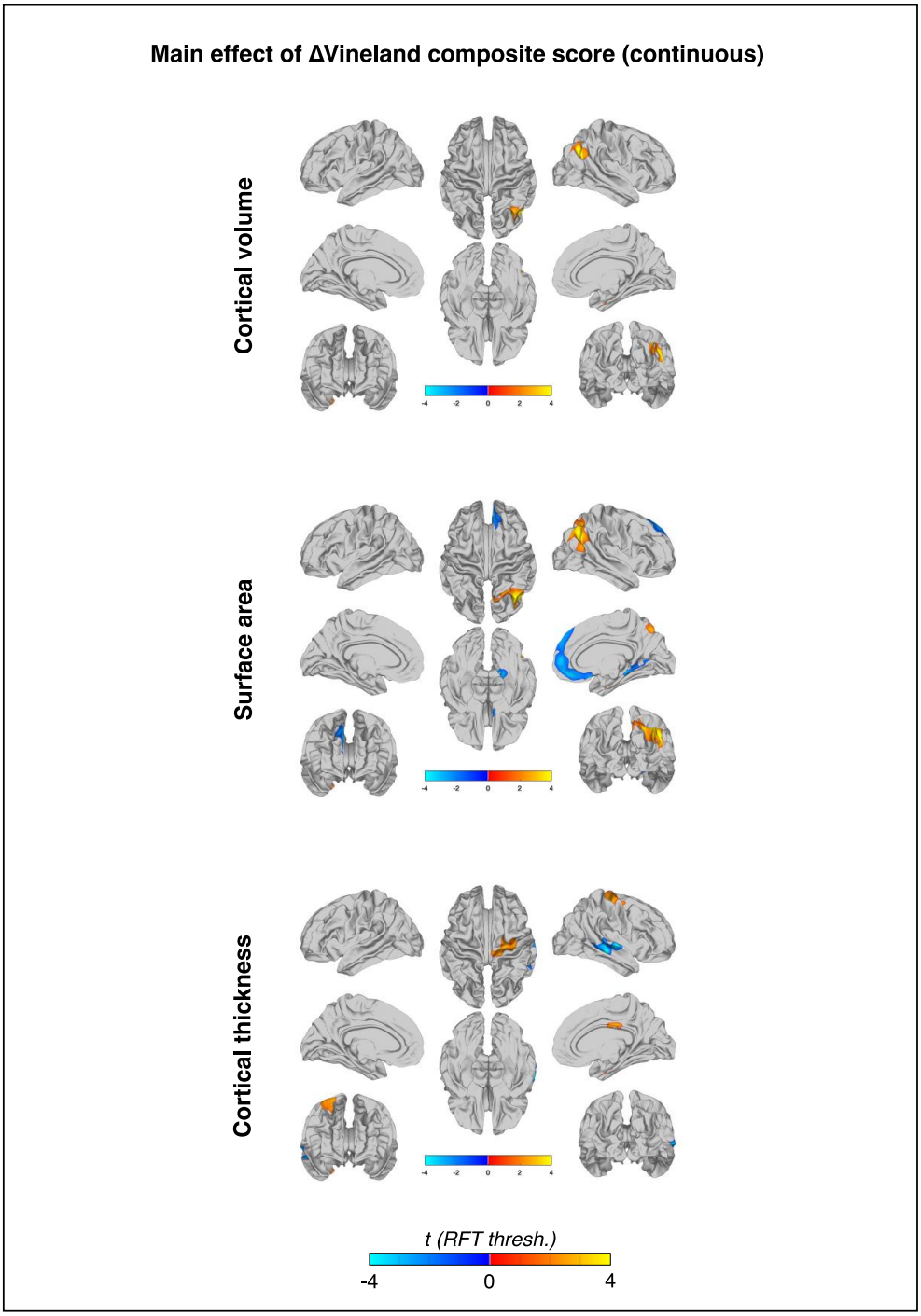

Figure S25 Main effect of change in adaptive behaviour on neuroanatomy. Cortical volume: effects observed in inferior parietal cortex. Surface area: effects in inferior parietal cortex, superior parietal cortex, precuneus cortex, superior frontal gyrus, medial orbital frontal cortex, lingual gyrus, and parahippocampal gyrus. Cortical thickness: effects in precentral gyrus, posterior cingulate cortex, superior temporal gyrus, middle temporal gyrus, and the banks of the superior temporal sulcus. T-values are random field theory (RFT)-corrected and indicated by colorbars.

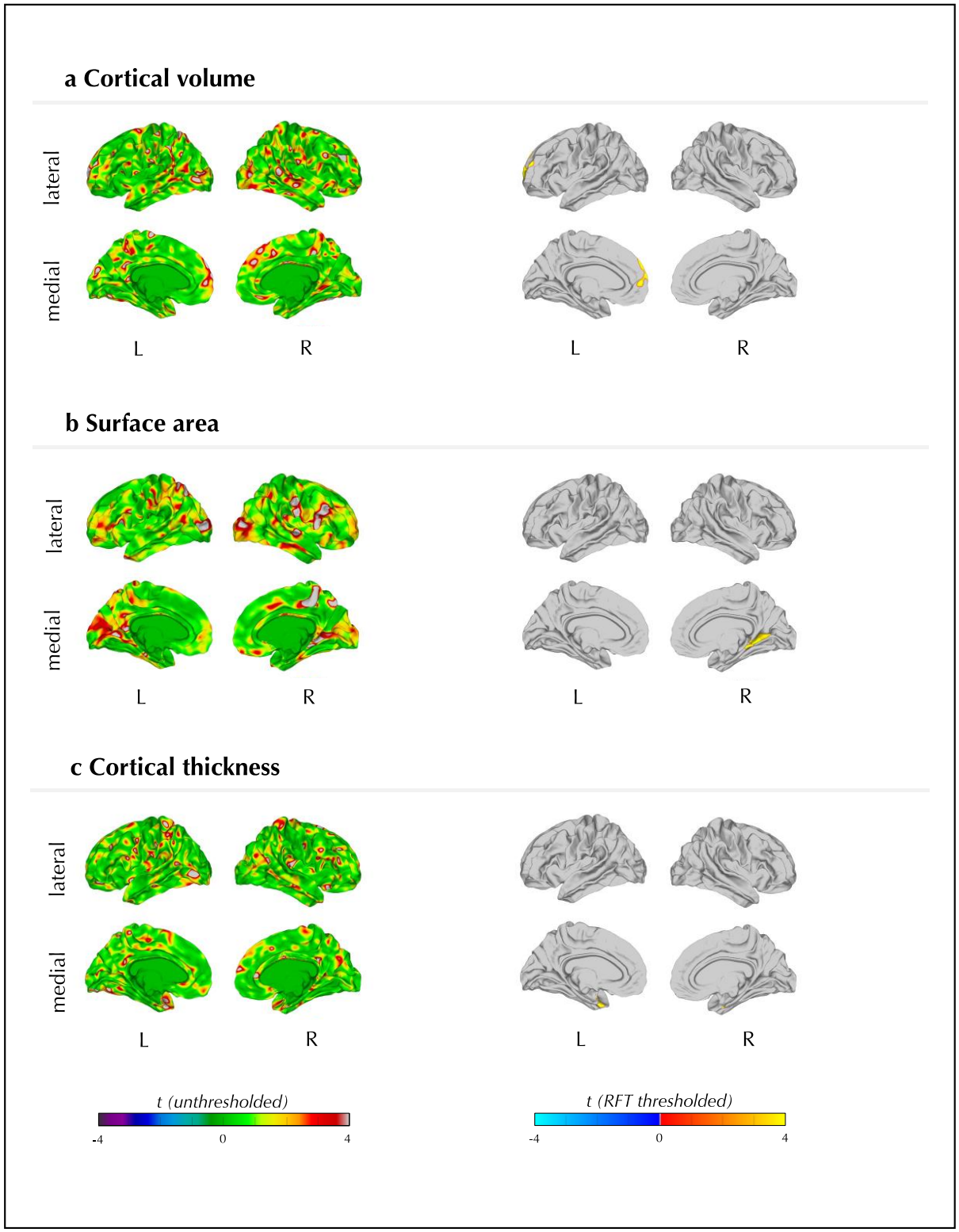

Figure S26 Main effect of outcome group on neuroanatomy. Cortical volume: effects observed in frontal cortex. Surface area: effects found in posterior temporal cortex. Cortical thickness: effects in anterior temporal cortex. Left column: T-values are random field theory (RFT)-corrected. Right column: uncorrected T-values. Abbreviations: L, left; R, right.

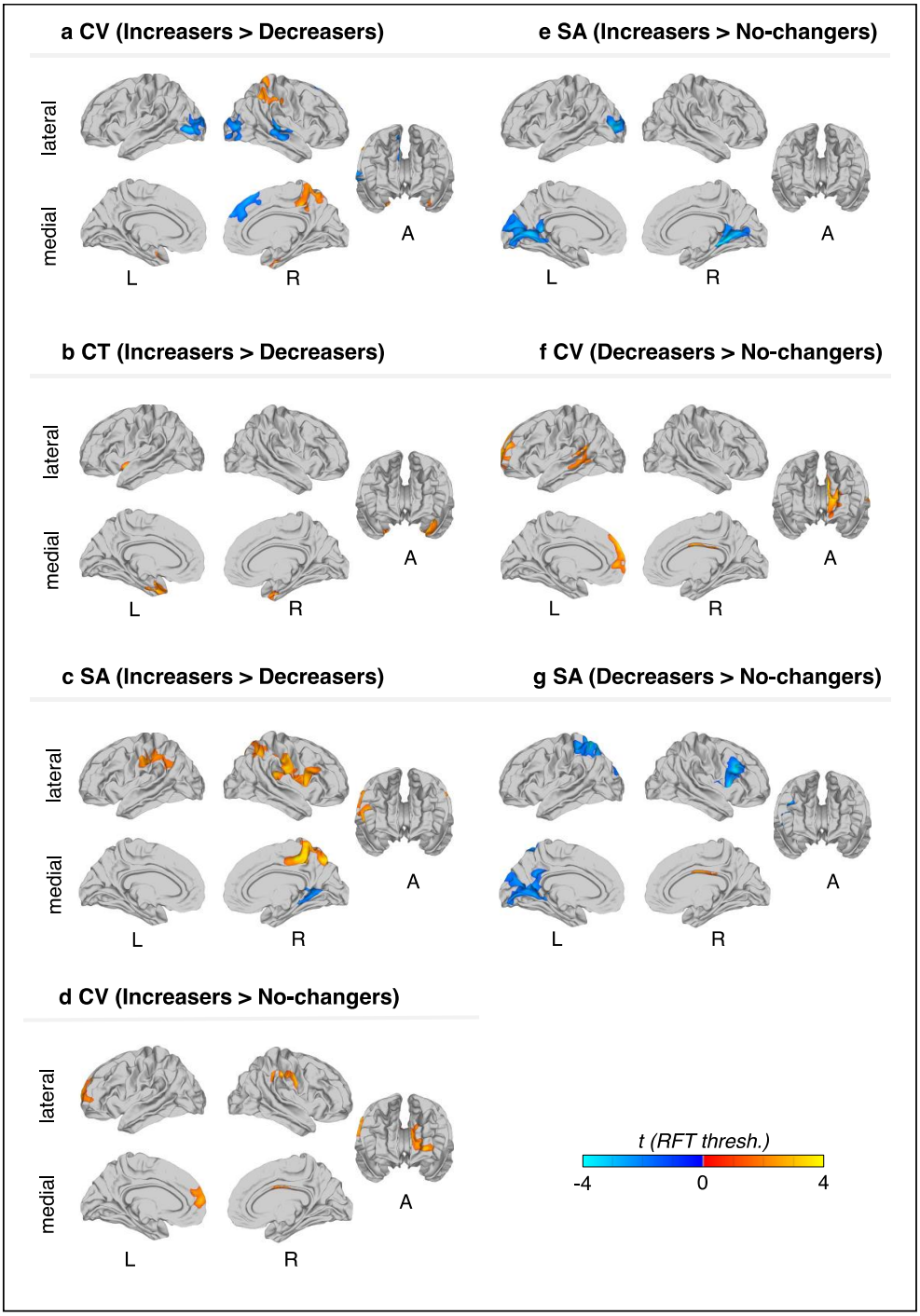

Figure S27 Neuroanatomical baseline differences between adaptive outcome groups additionally controlling for follow-up duration and its interaction with age. a-c: differences between Increasers and Decreasers in cortical volume (a), cortical thickness (b) and surface area (c). d-e: differences between Increasers No-changers in cortical volume (d) and surface area (e). f-g: differences between Decreasers and No-changers in cortical volume (f) and surface area (g). T-values are random field theory (RFT)-corrected and indicated by colorbars. Abbreviations: A, anterior view; CT, cortical thickness; CV, cortical volume; L, left; R, right; SA, surface area.

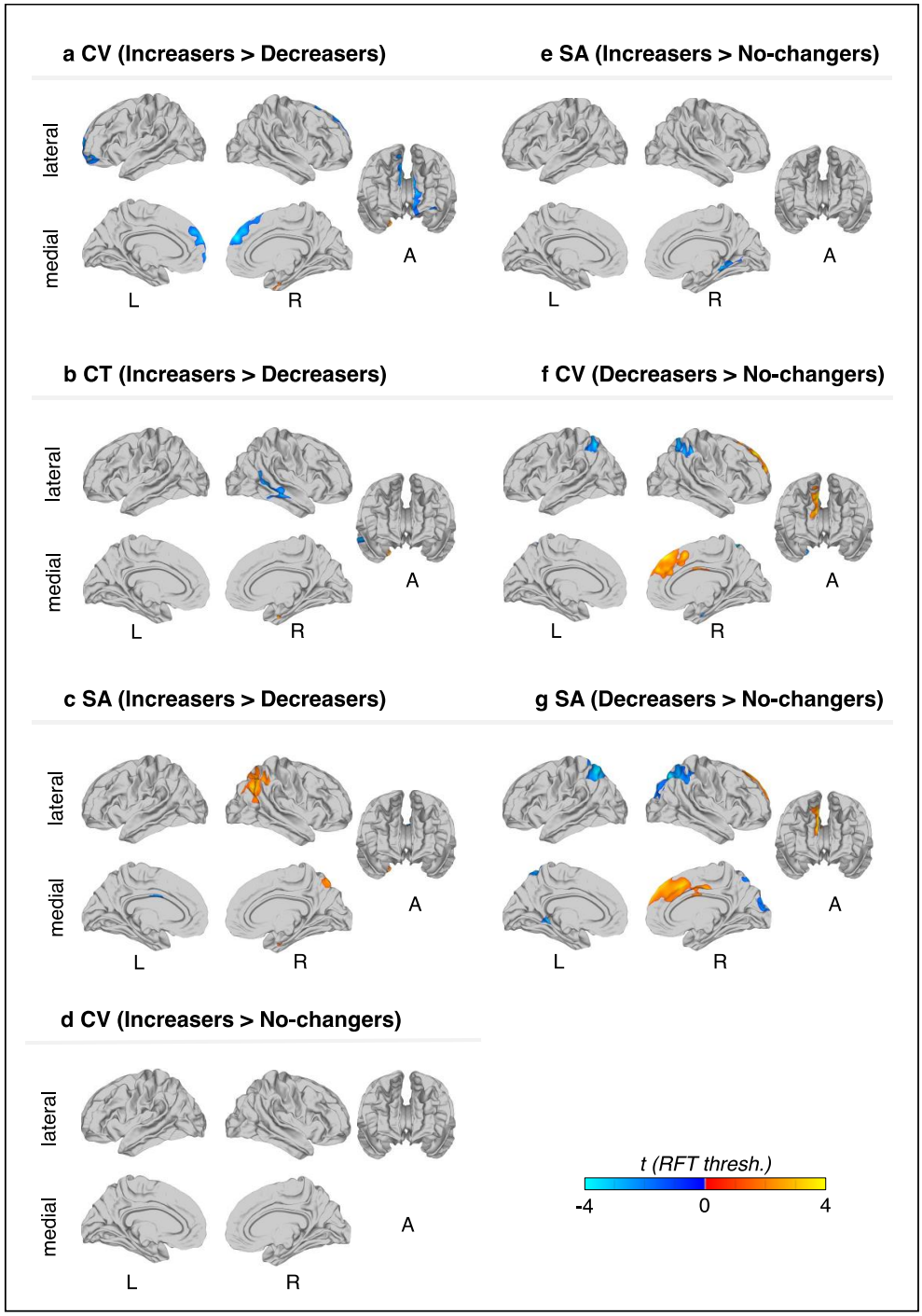

Figure S28 Neuroanatomical baseline differences between adaptive outcome groups in a subsample matched for age, sex, and IQ (n=83). a-c: differences between Increasers and Decreasers in cortical volume (a), cortical thickness (b) and surface area (c). d-e: differences between Increasers and No-changers in cortical volume (d) and surface area (e). f-g: differences between Decreasers and No-changers in cortical volume (f) and surface area (g). T-values are random field theory (RFT)-corrected and indicated by colorbars. Abbreviations: A, anterior view; CT, cortical thickness; CV, cortical volume; L, left; R, right; SA, surface area.

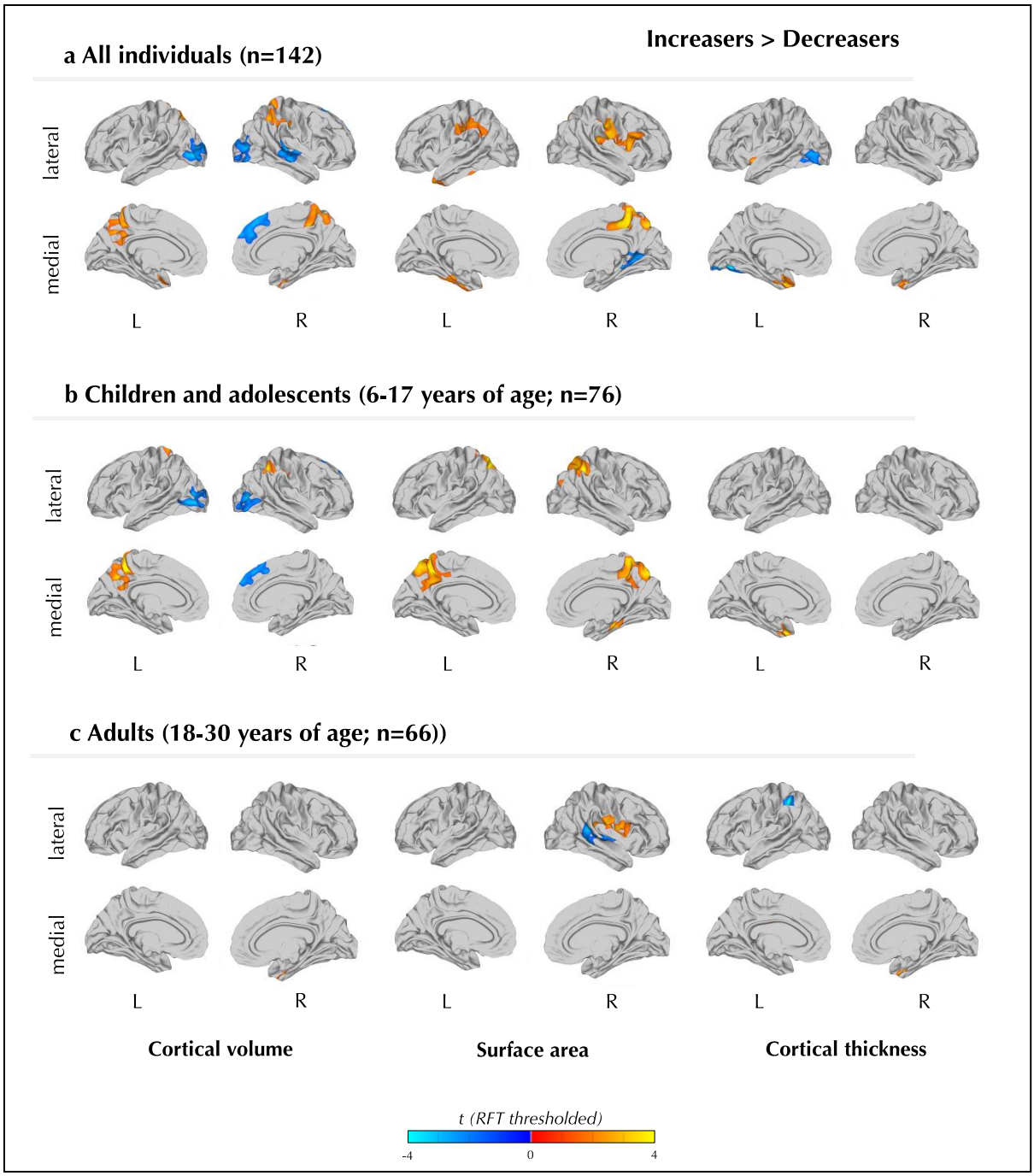

Figure S29 Neuroanatomical differences in cortical volume, cortical thickness, and surface area between Increasers and Decreasers. a-c: analyses within individuals across age (n=142). b: analyses within children and adolescents (6-17 years of age, n=76). c: analyses within adults (18-30 years of age, n=66). T-values are random field theory (RFT)-corrected and indicated by colorbars. Abbreviations: L, left; R, right.

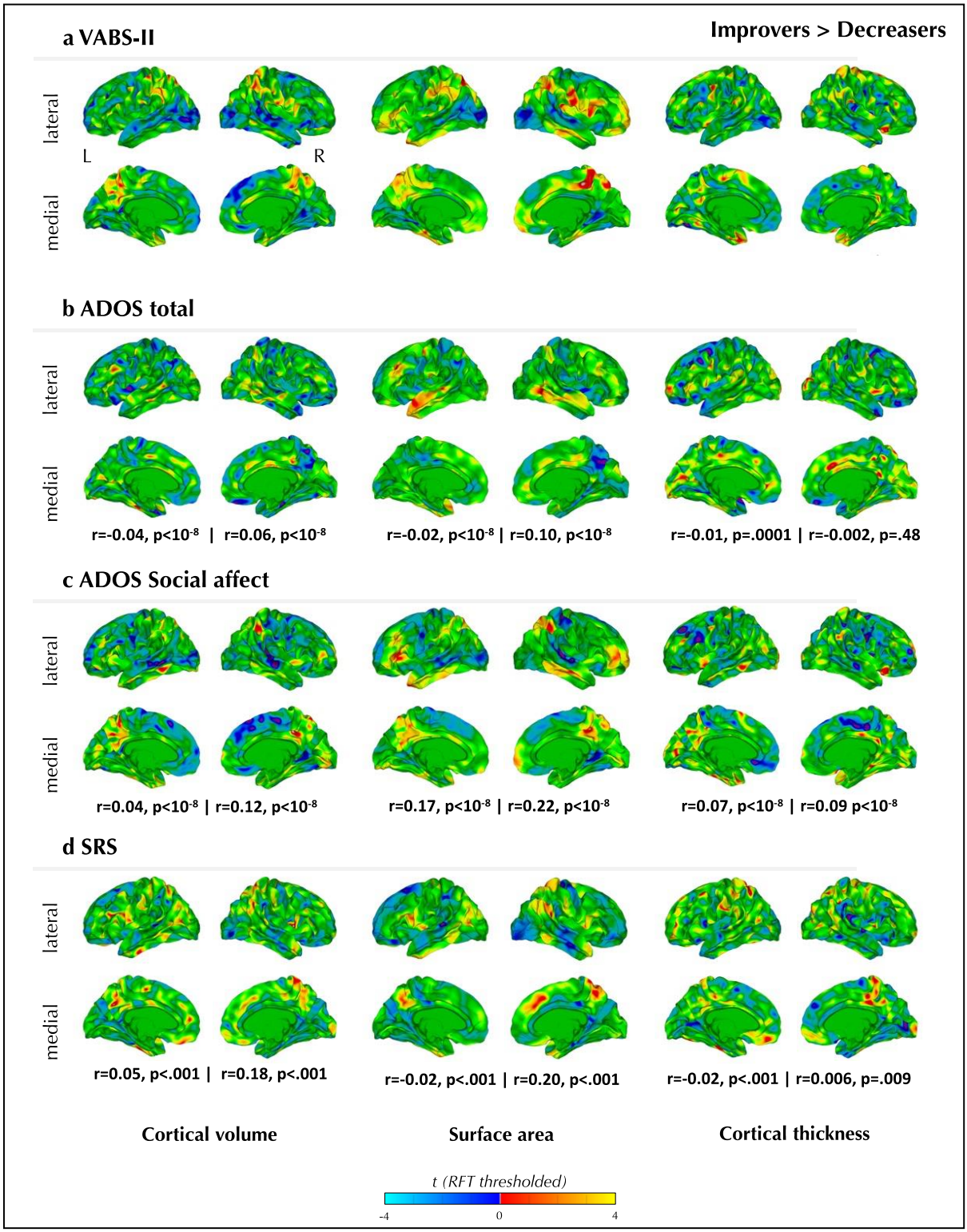

Figure S30 Neuroanatomical baseline differences in cortical volume, surface area, and cortical thickness between outcome groups based on different symptom/behaviour measures. a: differences between those individuals whose adaptive behavioural scores increased/improved (n=78) and decreased/deteriorated (n=64). b: differences between individuals whose ADOS total scores decreased/improved (n=36) vs increased/deteriorated (n=42). c: differences between individuals whose ADOS social affect domain scores decreased/improved (n=39) vs increased/deteriorated (n=37). d: differences between individuals whose SRS scores decreased/improved (n=30) vs increased/deteriorated (n=29). Numbers below images indicate Pearson correlation coefficients (r=correlation coefficient, p=p-value) of each map with the corresponding map in row A. T-values are unthresholded and indicated by colorbars. Abbreviations: L, left; R, right.

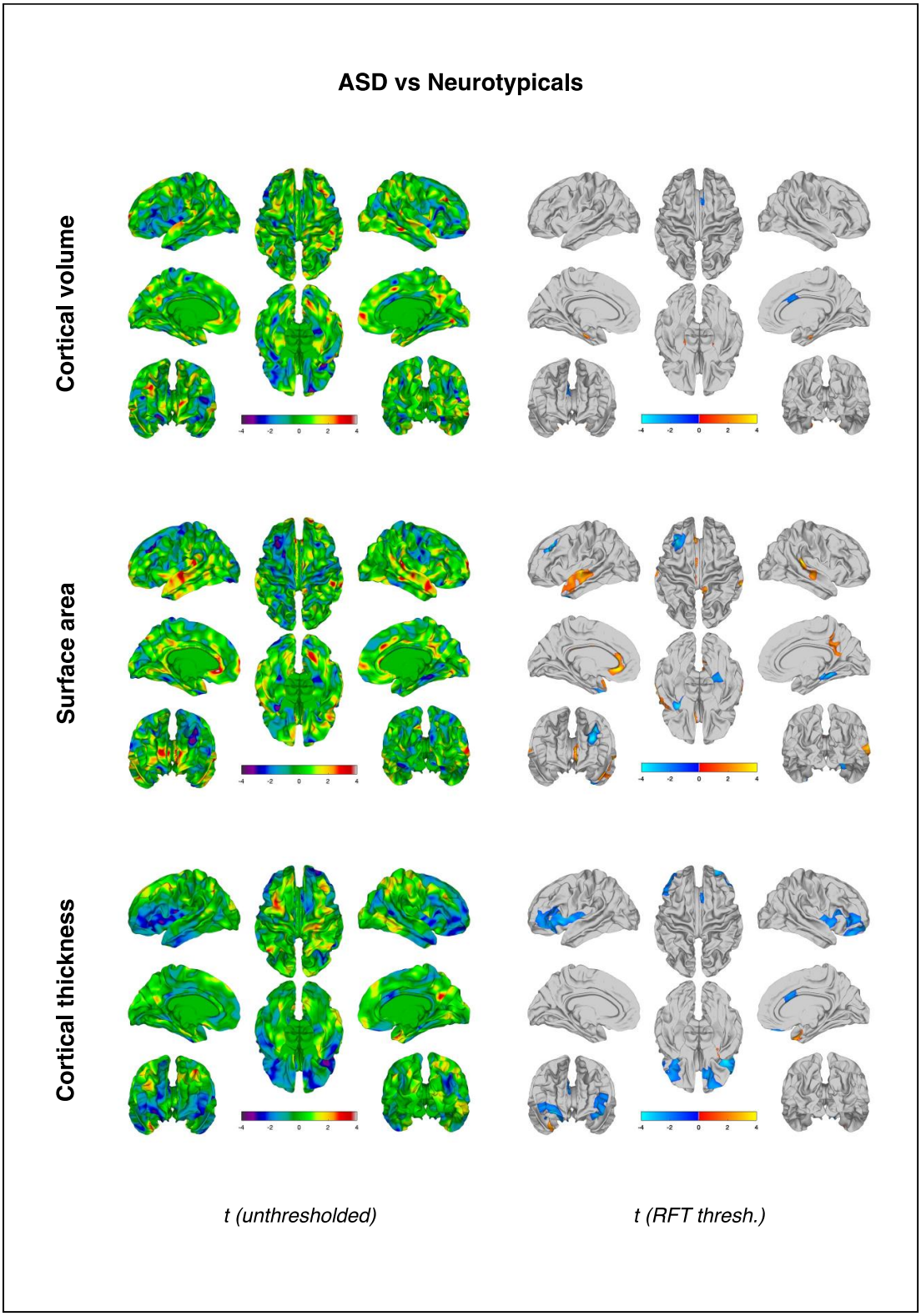

Figure S31 Neuroanatomical differences between all outcome groups combined (ASD) and the neurotypicals. Cortical volume: effects observed e.g., in anterior cingulate cortex. Surface area: effects seen in anterior cingulate cortex, insula, inferior/middle frontal gyrus, and orbital frontal cortex. Left column: T-values are random field theory (RFT)-corrected. Right column: uncorrected T-values.

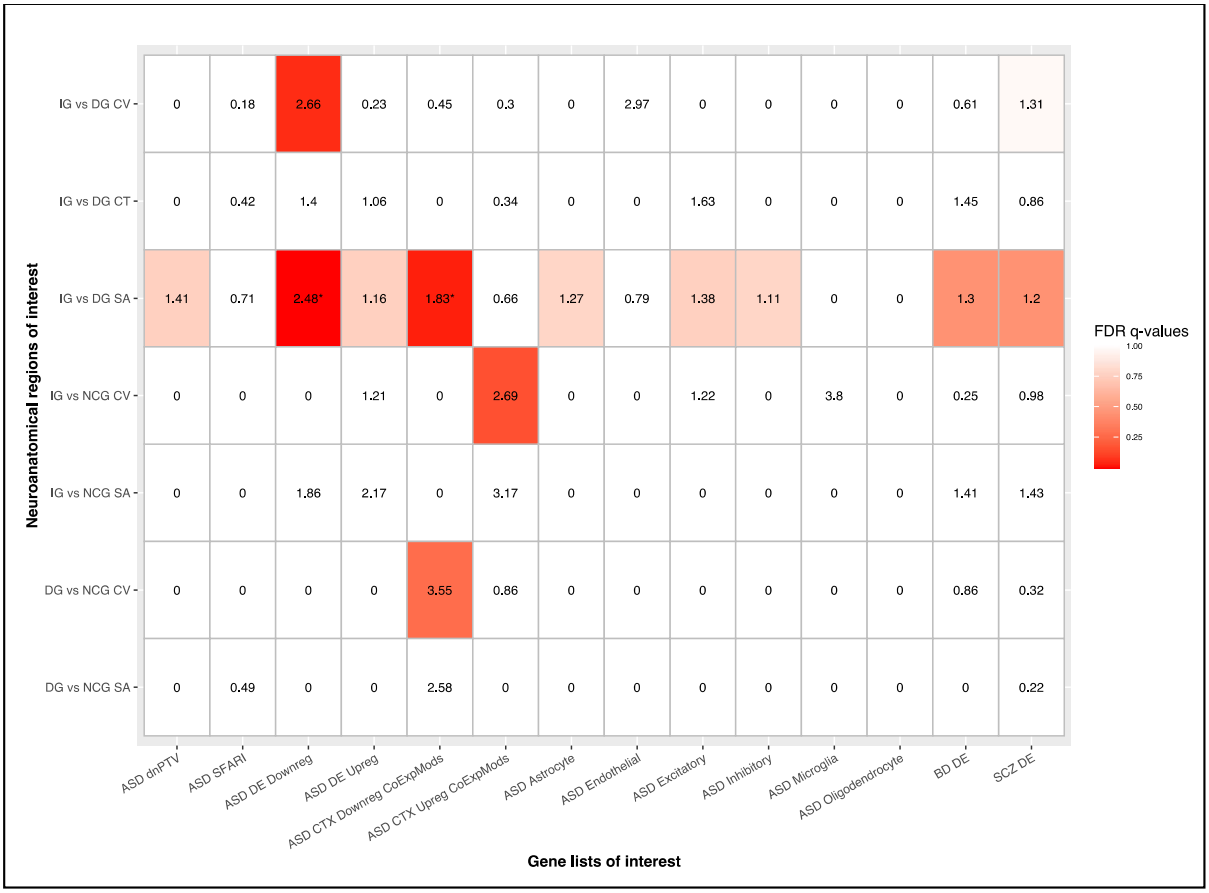

Figure S32 Genetic correlates of neuroanatomical variability (effect sizes). a: Enrichment analyses for cortical phenotypes (y-axis, rows) by ASD-associated gene lists (x-axis, columns). Tile colours indicate FDR q-values. Tile labels indicate enrichment effect sizes (Cohen's d), and significant values are marked with an asterisk. Negative values indicate un-enrichment/lessening of enrichment. Abbreviations: CT, cortical thickness; CV, cortical volume; DG, decrease group; IG, increase group; NCG, no change group; SA, surface area.

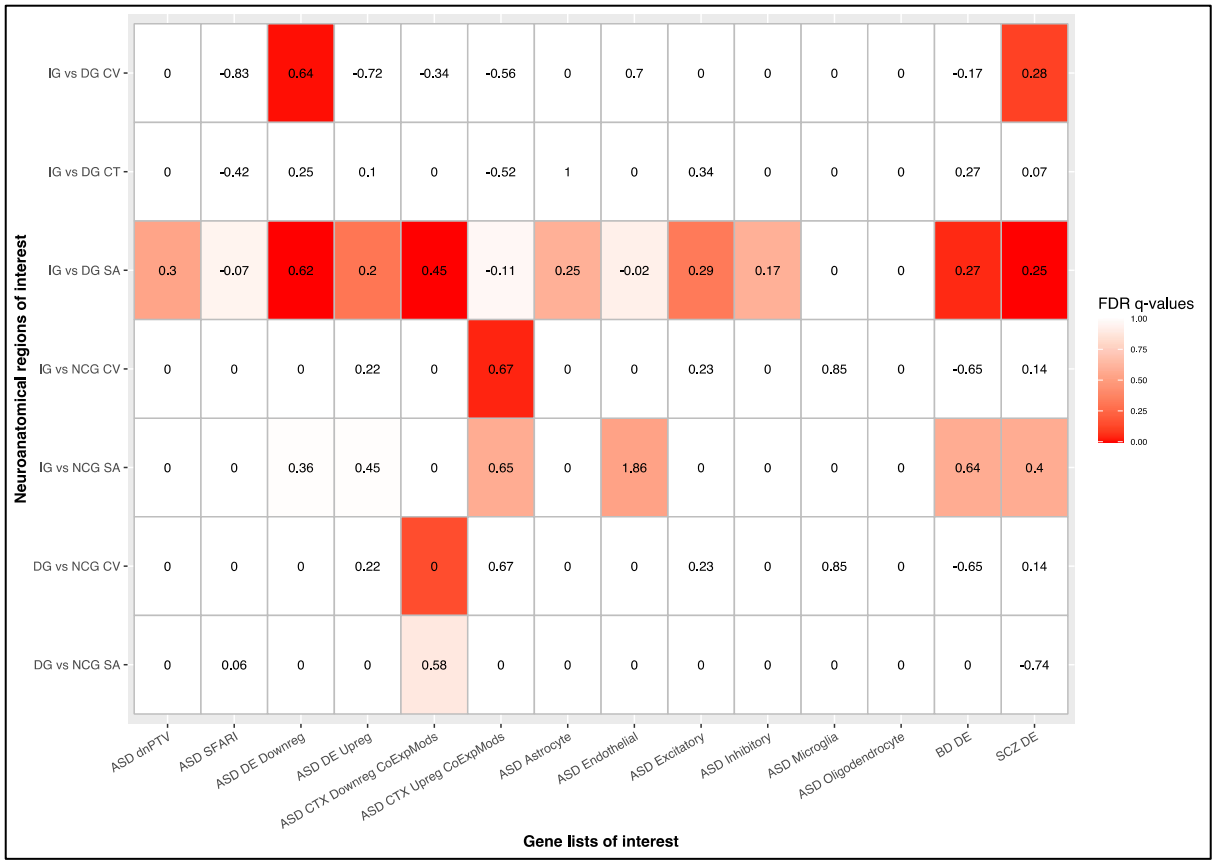

Figure S33 Enrichment analyses for cortical phenotypes (y-axis, rows) by ASD-associated gene lists (x-axis, columns) using a more restricted gene background list of 16,906 genes expressed in cortical tissue (vs 20,787 genes in the original analysis). Tile colours indicate FDR q-values. Tile labels indicate enrichment odd ratios, and significant values are marked with an asterisk. Abbreviations: CT, cortical thickness; CV, cortical volume; DG, decrease group; IG, increase group; NCG, no change group; SA, surface area.

Supplementary Tables

Table S1 Summary of image acquisition parameters used at each study site*. Please note that all scanners operated at 3 Tesla. Abbreviations: FA, flip angle; FOV, field of view; TE, echo time; TR, repetition time.*

Site	Manufacturer	Model	Software version	Acquisition sequence	Slices	TR [s]	TE [ms]	FA [˚]	Coverage	Thickness [mm]	Resolution [mm^3]	FOV
Cambridge	Siemens	Verio	Syngo MR B17	Tfl3d1_ns	176	2.3	2.95	9	256*256	1.2	1.1*1.1*1.2	270
KCL	GE Medical systems	Discovery mr750	LX MR DV23.1_V02_1317.c	SAG ADNI GO ACC SPGR	196	7.31	3.02	11				
Mannheim	Siemens	TimTrio	Syngo MR B17	MPRAGE ADNI	176	2.3	2.93	9				
Nijmegen	Siemens	Skyra	Syngo MR D13	Tfl3d1_16ns	176	2.3	2.93	9				
Rome	GE Medical systems	Signa HDxt	24/LX/MR HD16.0_V02_1131.a	SAG ADNI GO ACC SPGR	172	5.96	1.76	11				
Utrecht	Philips Medical Systems	Achieva/ Ingenia CX	3.2.3/3.2.3.1/ 5.1.9/5.1.9.1	ADNI GO 2	170	6.76	3.1	9				

Table S2 Predicting clinical outcome (V_{T2}-V_{T1}) using the Atypicality Index (AI). *Significantly contributing neuroanatomical factors are highlighted in bold. Abbreviations: b, unstandardized beta; β, standardized beta; CT, cortical thickness; CV, cortical volume; DG, Decrease group; FSIQ, full-scale IQ; IG, Increase group; NCG, No-change group; SA, surface area; SE, standard error; V_{T1}, Vineland composite score at timepoint 1 (baseline).*

V_{T2}-V_{T1}Vineland composite score (N = 204)						
Predicted measure	Predictors	b	SE	β	t-value	Probability
IG vs DG R=.422, R^2=.178, R^2_{adj}=.144, F(8,195)=5.266, p=.000	(constant)	2.018	5.97		0.34	0.736
	Age (yrs)	-0.07	0.13	-0.04	-0.59	0.556
	FSIQ	0.13	0.04	0.25	3.27	0.001
	Sex	2.00	1.46	0.09	1.38	0.171
	Site	1.28	0.49	0.20	2.63	0.009
	V_{T1}	-0.28	0.06	-0.34	-4.33	0.000
	AI CV	**-5.35**	**2.01**	**-0.30**	**-2.66**	**0.008**
	AI SA	**7.16**	**1.76**	**0.44**	**4.06**	**0.000**
	AI CT	2.91	1.49	0.15	1.95	0.053
IG vs NCG R=.361, R^2=.130, R^2_{adj}=.099, F(7,196)=4.184, p=.000	(constant)	1.40	6.09		0.23	0.818
	Age (yrs)	-0.12	0.13	-0.07	-0.95	0.344
	FSIQ	0.15	0.04	0.29	3.79	0.000
	Sex	1.05	1.49	0.05	0.71	0.481
	Site	1.19	0.48	0.18	2.51	0.013
	V_{T1}	-0.26	0.07	-0.33	-4.04	0.000
	AI CV	**1.75**	**0.72**	**0.17**	**2.44**	**0.016**
	AI SA	-0.61	0.87	-0.05	-0.71	0.482
DG vs NCG R=.378, R^2=.143, R^2_{adj}=.112, F(7,196)=4.672, p=.000	(constant)	1.39	6.05		0.23	0.818
	Age (yrs)	-0.07	0.13	-0.04	-0.55	0.581
	FSIQ	0.14	0.04	0.26	3.50	0.001
	Sex	1.34	1.47	0.06	0.92	0.361
	Site	1.05	0.47	0.16	2.24	0.026
	V_{T1}	-0.26	0.07	-0.32	-4.02	0.000
	AI CV	-1.57	0.88	-0.14	-1.78	0.076
	AI SA	**3.76**	**1.26**	**0.22**	**2.97**	**0.003**

Table S3 Predicting clinical outcome (V_{T2}-V_{T1}) using the Atypicality Index (AI) using Bootstrapping (4000 iterations). *Significantly contributing neuroanatomical factors are highlighted in bold. Abbreviations: b, unstandardized beta; c, cluster; CT, cortical thickness; CV, cortical volume; DG, Decrease group; FSIQ, full-scale IQ; IG, Increase group; NCG, No-change group; SA, surface area; SE, standard error; V_{T1}, Vineland composite score at timepoint 1 (baseline).*

	V_{T2}-V_{T1}Vineland composite score (N = 204)						
Predicted measure	Predictors	b	Bias	SE	Bootstrap p-value	95% Confidence Interval	
						lower	upper
IG vs DG R=.422, R^2=.178, R^2_{adj}=.144, F(8,195)=5.266, p=.000, standard error of estimate: 9.361	(constant)	2.018	-.222	5.97	.719	-9.052	12.494
	Age (yrs)	-0.07	.005	0.13	.576	-.327	.191
	FSIQ	0.13	.000	0.04	.001	.061	.199
	Sex	2.00	-.015	1.46	.190	-.906	4.873
	Site	1.28	.002	0.49	.013	.310	2.294
	V_{T1}	-0.28	.003	0.06	.001	-.391	-.147
	AI CV	**-5.35**	**-.020**	**2.01**	.018	-9.966	-1.085
	AI SA	**7.16**	**-0.17**	**1.76**	**.001**	**3.791**	**10.810**
	AI CT	2.91	-0.14	1.49	**.046**	**.162**	**5.824**
IG vs NCG R=.361, R^2=.130, R^2_{adj}=.099, F(7,196)=4.184, p=.000, standard error of estimate: 9.557	(constant)	1.40	-.177	6.09	.793	-9.445	11.510
	Age (yrs)	-0.12	.003	0.13	.350	-.377	.148
	FSIQ	0.15	.000	0.04	.000	.076	.220
	Sex	1.05	-.017	1.49	.509	-2.045	4.074
	Site	1.19	.011	0.48	.014	.304	2.176
	V_{T1}	-0.26	.002	0.07	.000	-.379	-.132
	AI CV	**1.75**	**-.022**	**0.72**	**.008**	**.433**	**2.982**
	AI SA	-0.61	-.017	0.87	.419	-2.102	.785
DG vs NCG R=.378, R^2=.143, R^2_{adj}=.112, F(7,196)=4.672, p=.000, standard error of estimate: 9.486	(constant)	1.39	-.068	6.05	.820	-10.303	12.584
	Age (yrs)	-0.07	.005	0.13	.576	-.322	.190
	FSIQ	0.14	-.002	0.04	.001	.062	.205
	Sex	1.34	.037	1.47	.383	-1.533	4.379
	Site	1.05	-.010	0.47	.024	.145	1.927
	V_{T1}	-0.26	.003	0.07	.000	-.379	-.136
	AI CV	-1.57	.013	0.88	.156	-3.721	.529
	AI SA	**3.76**	**-.074**	**1.26**	**.003**	**1.263**	**6.085**

Table S4 Medication information for all participants included in this study.

Diagnostic group	Unknown % (n)	No % (n)	Yes % (n)	Medication and Categories
ASD (n = 204)	29% (60)	31% (63)	40% (81) One: n=81 Two: n=25 Three: n=3	Antidepressant (n=22) • 19 Selective serotonin reuptake inhibitor (SSRI) • 2 Tetracyclic antidepressant (TeCA) • 1 Tricyclic Antidepressant (TCA) Antiepileptics (n=6; no additional information) Antimigraine preparations (n=4; no additional information) Antipsychotics (n=14) • 4 Aripiprazole • 1 Clozapine • 1 Pipamperone • 8 Risperidone Anxiolytics (n=2; no additional information) Hypnotics and sedatives (n=28) • 28 Melatonin Other analgesics and antipyretics (n=4) • 4 Other analgesics and antipyretics Psychostimulants and other drugs used to treat ADHD (n=29) • 3 Atomoxtine • 1 Dexamfetamine • 25 Methylphenidate hydrochloride
Neurotypicals (n = 279)	56% (155)	37% (103)	8% (21) One: n=21 Two: n=4 Three: n=1	Antidepressant (n=4) • 3 Selective serotonin reuptake inhibitor (SSRI) • 1 Tetracyclic antidepressant (TeCA) Antiepileptics (n=2; no additional information) Anxiolytics (n=1; no additional information) Drugs used in addictive disorder (n=1; no additional information) Hypnotics and sedatives (n=3) • 3 Melatonin Other analgesics and antipyretics (n=5) • 1 Opioids • 4 Other analgesic and antipyretics Psychostimulants and other drugs used to treat ADHD (n=10) • 2 Atomoxetine • 8 Methylphenidate hydrochloride