bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Determining the association between cortical morphology and cognition in 10,145 children from the Adolescent Brain and Cognitive Development (ABCD) study using the MOSTest

C E Palmer[1,3], W Zhao[2], R Loughnan[2], C C Fan[1,3], W Thompson[4], T L Jernigan[1,2,5,7] & A M Dale[5,6,7]

1. Center for Human Development, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92161, USA
2. Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
3. Center for Multimodal Imaging and Genetics, University of California, San Diego School of Medicine, 9444 Medical Center Dr, La Jolla, CA 92037, USA
4. Division of Biostatistics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA
5. Department of Radiology, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA
6. Department of Neuroscience, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA
7. Department of Psychiatry, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA

ABSTRACT

During development individuals undergo protracted changes in cortical morphology, which coincide with changes in cognition. A number of studies have measured the association between brain structure and cognition in children, however these frequently rely on mass univariate statistics or ROI-based comparisons and do not always measure cortical morphology relative to global brain measures. After controlling for global brain measures, we can see a residual regionalisation pattern indicating the size or thickness of different regions relative to the total cortical surface area or mean thickness. Individual variability in this regionalisation may be important for predicting between subject variability in cognitive performance. Here we sought to determine whether the relative configuration of cortical architecture across the whole cortex was associated with cognition using a novel multivariate omnibus statistical test (MOSTest). This method is optimally powered to detect associations that are widely distributed across the cortex. We then used the Polyvertex score (PVS) to quantify the magnitude of the association between vertex-wise cortical morphology and cognitive performance. We have determined that the relative pattern of cortical architecture, after removing the variance associated with global brain measures, predicted unique variance associated with cognition across different modalities.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

INTRODUCTION

During development there are protracted increases in cortical surface area (which peaks at 9 and 10 years of age) and independent monotonic decreases in cortical thickness (Brown & Jernigan, 2012; Jernigan & Tallal, 1990; Jernigan, Trauner, Hesselink, & Tallal, 1991; Lyall et al., 2015; Wierenga, Langen, Oranje, & Durston, 2014) that coincide with rapid cognitive development. However, the relationship between these distinct neuroanatomical trajectories and cognition is not well understood in late childhood. This is in part due to methodological differences, which conflate these independent morphometrics (Brown & Jernigan, 2012). Many studies have looked at regional associations between brain structure and cognition; however, these do not always partition out the variance associated with modality-specific global brain measures. However, regional differences in cortical morphology relative to total cortical surface area (CSA) or mean cortical thickness (CTH) also appear to be important for predicting behaviour (Curley et al., 2018; Fjell et al., 2012; Newman, Jernigan, et al., 2016; Newman, Thompson, et al., 2016; Vuoksimaa et al., 2016). This suggests that individual differences in the regionalisation of the cortex (relative to global brain measures) may be important for understanding individual variability in cognition. However, previous studies have been underpowered to detect significant effects of relative cortical configuration across the whole cortex particularly when using univariate statistics with stringent control for multiple comparisons (Vuoksimaa et al., 2016).

Brain-behaviour relationships are likely to represent weak diffuse effects across the brain, which are difficult to detect with small sample sizes. A recent study has highlighted how stringent thresholding procedures used in neuroimaging can produce variability among study replications and mask small, but true effects (Cremers, Wager, & Yarkoni, 2017). From genetics research we have discovered that complex behavioural traits are polygenic and the additive effect of many small effects across the genome is more predictive of behavioural phenotypes than the effect of stringently thresholded genome-wide significant single nucleotide polymorphisms (SNPs). The same appears to be true of the brain. We have recently demonstrated that we can predict more variance in behaviour by taking into account effect sizes across the whole cortex using novel analysis methods adapted from genetics research (Zhao et al, in prep). The poly*vertex* score (PVS), like the poly*genic* risk score, allows us to aggregate vertex-wise (or voxel-wise) effects across the cortex to predict behaviour. Based on this previous research and our understanding of how regionalisation occurs during development, we may assume that the association between cortical morphology and cognition is more likely to be widely distributed across the cortex.

In the current study, we hypothesised that the relative configuration of brain morphology (controlling for global measures) across the whole cortex could explain individual differences in cognitive task performance. To test this we are using a novel multivariate omnibus statistical test (MOSTest) for determining an association between vertex-wise imaging data and behaviour. The MOSTest aggregates effects across the cortex and is therefore optimally powered to detect distributed associations. This is in contrast to a traditional neuroimaging omnibus test based on the most significant vertex or voxel, which is alternatively optimally powered for sparse, localised effects of interest. The MOSTest has three main steps: 1) calculate a multivariate statistic by calculating the multivariate norm (Mahalanobis distance) across the cortex using the decorrelated mass univariate z statistics; 2) estimate the distribution of the test statistic under the null hypothesis using an appropriate permutation scheme and determine statistical significance controlling for a family-wise error rate at 5%; and 3) fit a gamma cumulative density function to the tail of the permuted distribution in order to extrapolate a MOSTest theoretical p-value that has greater precision and reduces the need for a high number of permutations. Van der Meer et al (in prep) have recently

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

validated this method in a different context and demonstrated the increased power of the MOSTest for detecting genome-wide associations with regional imaging phenotypes in UK Biobank. Here we demonstrate an application of the MOSTest for improving the discoverability of vertex-wise brain-behaviour associations in the Adolescent Brain and Cognitive Development (ABCD) study.

The large-scale ABCD study uses neuroimaging, genetics and a multi-dimensional battery of behavioural assessments to investigate the role of biological, environmental, and behavioural factors on brain, cognitive, and social/emotional development. The baseline data for this longitudinal study includes 11,875 nine and ten year old children, which allows us to generate a comprehensive cross-sectional understanding of the relationship between brain structure and cognition. Using the complete baseline data (release 2.0.1), we determined whether there was an association between performance on the thirteen cognitive tasks in the neurocognitive battery and the regionalisation of CSA (controlling for total CSA) and CTH (controlling for mean CTH) using the MOSTest. We then quantified the magnitude of these relationships using the PVS. We hypothesised that individual variability in the relative regionalisation of cortical morphology would explain cognitive task performance above and beyond that predicted by global brain measures.

METHOD

Sample

The ABCD study is a longitudinal study across 21 data acquisition sites following 11,875 children starting at 9 and 10 years old. This paper analysed the full baseline sample from release 2.0.1 (NDAR DOI: 10.15154/1504041). The study used specific recruitment strategies to create a population-based, demographically diverse sample, however it is not necessarily representative of the U.S. national population (Compton, Dowling, & Garavan, 2019; Garavan et al., 2018). Due to the inclusion of a wide range of individuals across different races, ethnicities and socioeconomic backgrounds, it is important to control for effects of self-declared race and ethnicity, household income and parental education when using this dataset due to potential confounding associations between these demographic variables and our effects of interest. Sex and age were also used as covariates in all analyses. Embedded within the sample is a large twin cohort and many siblings, therefore family relatedness was also controlled for as a random effect in all analyses for using a restricted exchangeability permutation scheme (see *Statistical Analysis 1)*.

Only subjects who had complete data across all of the measures analysed were included in the neuroimaging analyses. There was a large number of subjects with missing income data (n=1018), therefore missing values were imputed by taking the median income value across participants from the same testing site. The sources of missing data were as follows: incomplete across all demographic variables (n=189), incomplete across all cognitive measures (n=944), unavailable T1-weighted MRI scan for reasons outlined in the ABCD release notes (e.g. did not get scanned, motion artefacts) (n=339) and imaging data that was made available but did not pass the free-surfer QC flag (n=462). These missing data values are not mutually exclusive. The permutation testing used in this study was dependent on having multiple families with the same number of children, therefore the single family with 5 children was excluded from these analyses. This resulted in a final sample of 10,145 subjects. Supplementary Table 1 outlines the names of each variable used in these analyses from data release 2.0.1 in accordance with the NDA data dictionary.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Neurocognitive assessment

NIH Toolbox

The NIH Toolbox Cognition Battery (NTCB) is a widely used battery of cognitive tests that measures a range of different cognitive domains. All of the tasks within the NTCB were administered using an iPad with support or scoring from a research assistant where needed. Below is a brief description of each task.

The ***Toolbox Oral Reading Recognition Task (TORRT)*** measured language decoding and reading. Children were asked to read aloud single letters or words presented in the center of an iPad screen. The research assistant marked pronunciations as correct or incorrect. Extensive training was given prior to administering the test battery. Item difficulty was modulated using computerised adaptive testing (CAT).

The ***Toolbox Picture Vocabulary Task (TPVT)***, a variant of the Peabody Picture Vocabulary Test (PPTV), measured language and vocabulary comprehension. Four pictures were presented on an iPad screen as a word was played through the iPad speaker. The child was instructed to point to the picture, which represented the concept, idea or object name heard. CAT was implemented to control for item difficulty and avoid floor or ceiling effects.

The ***Toolbox Pattern Comparison Processing Speed Test (TPCPST)*** measured processing speed. Children were shown two images and asked to determine if they were identical or different by touching the appropriate response button on the screen. This test score is the sum of the number of items completed correctly in the time given.

The ***Toolbox List Sorting Working Memory Test (TLSWMT)*** measured working memory. Children heard a list of words alongside pictures of each word and were instructed to repeat the list back in order of their actual size from smallest to largest. The list started with only 2 items and a single category (e.g. animals). The number of items increased with each correct answer to a maximum of seven. The child then progressed to the next stage in which two different categories were interleaved. At this stage children were required to report the items back in size order from the first category followed by the second category. Children were always given two opportunities to repeat the list correctly before the experimenter scored the trial as incorrect.

The ***Toolbox Picture Sequence Memory Test (TPSMT)*** measured episodic memory. On each trial, children were shown a series of fifteen pictures in a particular sequence. The pictures illustrated activities or events within a particular setting (e.g. going to the park). Participants were instructed to touch the pictures in the original sequence in which they were shown. The Rey-Auditory Verbal Learning Task (RAVLT) was also included in the ABCD neurocognition battery as a more comprehensive measure of episodic memory.

The ***Toolbox Flanker Task (TFT)*** measured executive function, attentional and inhibitory control. This adaptation of the Eriksen Flanker task (Eriksen & Eriksen, 1974) captures how readily a participant is influenced by the congruency of stimuli surrounding a target. On each trial a target arrow was presented in the center of the iPad screen facing to the left or right and was flanked by two additional arrows on both sides. The surrounding arrows were either facing in the same (congruent) or different (incongruent) direction to the central target arrow. The participant was instructed to push a response button to indicate the direction of the central target arrow. Accuracy and reaction time scores were combined to produce a total score of executive attention, such that

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

higher scores indicate a greater ability to attend to relevant information and inhibit incorrect responses.

The ***Toolbox Dimensional Change Card Sort Task (TDCCS)*** measured executive function and cognitive flexibility. On each trial, the participant was presented with two objects at the bottom of the iPad screen and a third object in the middle. The participant was asked to sort the third object by matching it to one of the bottom two objects based on either colour or shape. In the first block participants matched based on one dimension and in the second block they switched to the other dimension. In the final block, the sorting dimension alternated between trials pseudorandomly. The total score was calculated based on speed and accuracy.

In the current study, the uncorrected scores for each task were used for statistical analyses. Composite scores of crystallised intelligence (mean of TPVT and TORRT), fluid intelligence (mean of TPCPST, TLSWMT, TPSMT, TFT and TDCCS) and total score (mean of all tasks) were also analysed. These measures are highly correlated with 'gold standard' measures of intelligence in adults (Heaton et al., 2014) and children (Akshoomoff et al., 2013).

Rey-Auditory Verbal Learning Task (RAVLT)

This task measures auditory learning, recall and recognition. Participants listened to a list of 15 unrelated words and were asked to immediately recall these after each of five learning trials. A second unrelated list was then presented and participants were asked to recall as many words as possible from the second list and then recall words again from the initial list. Following a delay of 30 minutes (during which other non-verbal tasks from the cognitive battery are administered), longer-term retention was measured using recall and recognition. This task was administered via an iPad using the Q-interactive platform of Pearson assessments (Daniel, Wahlstrom, & Zhang, 2014). In the current study, the total number of items correctly recalled across the five learning trials was summed to produce a measure of auditory verbal learning.

Little Man Task (LMT)

This task measures visuospatial processing involving mental rotation with varying degrees of difficulty (Acker, 1982). A rudimentary male figure holding a briefcase in one hand was presented on an iPad screen. The figure could appear in one of four positions: right side up vs upside down and either facing the participant or with his back to the participant. The briefcase could be in either hand. Participants indicated which hand the briefcase was in using one of two buttons. Performance across the 32 trials was measured by the percentage of trials in which the child responded correctly. This was divided by the average reaction time to complete the task (in seconds) to produce a measure of efficiency of visuospatial processing. This was the dependent variable analysed in this study.

Matrix reasoning

Nonverbal reasoning was measured using an automated version of the Matrix Reasoning subtest from the Weschler Intelligence Test for Children-V (WISC-V; Weschler, 2014). On each trial the participant was presented with a series of visuospatial stimuli, which was incomplete. The participant was instructed to select the next stimulus in the sequence from four alternatives. There

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

were 32 possible trials and testing ended when the participant failed three consecutive trials. The total raw score, used in the current study, was the total number of trials completed correctly.

Imaging acquisition

The ABCD MRI data were collected across 21 research sites using Siemens Prisma, GE 750 and Philips 3T scanners. Scanning protocols were harmonised across sites. The T1w acquisition (1 mm isotropic) was a 3D T1w inversion prepared RF-spoiled gradient echo scan using prospective motion correction, when available (Tisdall et al., 2012; White et al., 2010) (echo time = 2.88 ms, repetition time = 2500 ms, inversion time = 1060 ms, flip angle = 8°, FOV = 256x256, FOV phase = 100%, slices = 176). Only the T1w scans were analysed in this paper. Full details of all the imaging acquisition protocols used in ABCD are outlined by Casey et al (Casey et al., 2018). Scanner ID was included in all analyses to control for any differences in image acquisition across sites and scanners.

Imaging pre-processing

Pre-processing of all MRI data for ABCD was conducted using in-house software at the Center for Multimodal Imaging and Genetics (CMIG) at University of California San Diego (UCSD) as outlined in Hagler et al (Hagler et al., 2019). Manual quality control was performed prior to the full image pre-processing and structural scans with poor image quality as well as those that did not pass FreeSurfer QC were excluded from all analyses. Brain segmentation and cortical surface reconstruction were completed using FreeSurfer v5.3.0 (Dale, Fischl, & Sereno, 1999; Bruce Fischl, Sereno, & Dale, 1999).

T1-weighted structural images were corrected for distortions caused by gradient nonlinearities, coregistered, averaged, and rigidly resampled into alignment with an atlas brain. See previous publications for details of the surface based cortical reconstruction segmentation procedures (Dale et al., 1999; B Fischl & Dale, 2000; Bruce Fischl et al., 2004, 1999; Jovicich et al., 2006). In brief, a 3D model of the cortical surface was constructed for each subject. This included segmentation of the white matter (WM), tessellation of the gray matter (GM)/WM boundary, inflation of the folded, tessellated surface, and correction of topological defects.

Measures of cortical thickness at each vertex were calculated as the shortest distance between the reconstructed GM/WM and pial surfaces (B Fischl & Dale, 2000). To calculate cortical surface area, a standardised tessellation was mapped to the native space of each subject using a spherical atlas registration, which matched the cortical folding patterns across subjects. Surface area of each point in atlas space was calculated as the area of each triangle. This generated a continuous vertex-wise measure of relative areal expansion or contraction. Cortical maps were smoothed using a Gaussian kernel of 20 mm full-width half maximum (FWHM) and mapped into standardised spherical atlas space. Vertex-wise data for all subjects for each morphometric measurement were concatenated into matrices in MATLAB R2017a and entered into general linear models for statistical analysis using custom written code.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Statistical Analysis 1: determining the association between relative brain structure and cognition using the MOSTest

In order to determine whether there was a significant association between relative, vertex-wise cortical morphology and cognitive task performance, we employed the MOSTest. We applied a general linear model (GLM) associating a given behaviour y from a set of covariates W and the vertex-wise morphology data, X. W represents a standardized N by m matrix of covariates of no interest where N represents the number of subjects and m the number of covariates. X denotes the standardised N by V matrix of vertex-wise imaging data where V denotes the number of vertices. y represents an N by 1 vector of a behavioural phenotype. β and α are the V by 1 and m by 1 vectors of parameters of interest and no interest respectively (Eq 1).

(1) $y = \alpha_0 + \boldsymbol{W}\alpha + \boldsymbol{X}\beta + \varepsilon$

Using this linear model, we evaluated whether the vertex-wise data could predict the behaviour of interest, while controlling for a set of covariates, and obtained the least-squares estimate for each β_j from the GLM. This was used to compute the multivariate vector of association Wald statistics $\boldsymbol{z} = (z, \dots, z_V)'$ across the whole cortex. The multivariate statistic was calculated as the Mahalanobis norm, ${\chi^2}_{MOST}$ (Eq 2). This is an approximation of the multivariate normal cumulative distribution with known variance-covariance matrix, given by $\boldsymbol{R}$. The correlation matrix was estimated from the P by V matrix of permuted z statistics, $\boldsymbol{Z_{perm}}$, where p = 1,…,P denotes permutations. This was then regularised using a Tikhonov regularization to ensure the correlation matrix was invertible (Eq 3). The magnitude of regularisation dictated by the penalty factor λ was optimised based on the negative log likelihood of the fit of the gamma cumulative density function to the permuted distribution of ${\chi^2}_{MOST}$ (see supplementary figure 3).

(2) ${\chi^2}_{MOST} = z'\widehat{\boldsymbol{R}}^{-1}z$

(3) $\widehat{\boldsymbol{R}} = cov(\boldsymbol{Z_{perm}})(1-\lambda) + \lambda\boldsymbol{I}$

A permutation test was used to determine the distribution of ${\chi^2}_{MOST}$ under the null hypothesis H_0. We rejected H_0 if ${\chi^2}_{MOST}$ was greater than the permutation distribution more than 99.62% of the time (alpha level of 95% Bonferroni corrected for 13 tests). Theoretical p-values for ${\chi^2}_{MOST}$ were calculated by fitting a gamma cumulative density function to the permuted scores and estimating the likelihood of the observed omnibus test statistic given this permuted distribution. Supplementary figures 4A & 5A show the fit of the gamma distribution to the permuted data. All analyses were also conducted using the most significant vertex, ${z^2}_{MINP}$, as the omnibus statistic (Eq 4). Theoretical p-values for these permutation tests were generated by fitting a beta cumulative density function to the permuted data. This estimates the likelihood of these observations using a Sidak correction. The fits of these data are shown in supplementary figures 4B & 5B.

(4) $z_{MINP} = \max(z^2)$

To correctly estimate the null distribution from permutations, subject labels were shuffled according to exchangeability blocks (EBs) defined based on the family structure within ABCD (Winkler, Webster, Vidaurre, Nichols, & Smith, 2015). Families with the same number of siblings were allowed to be shuffled as a whole and siblings within a family were allowed to be shuffled with each other. This was necessary to account for the joint distribution of the observed data due to the relatedness of the ABCD sample. Permutations that adhered to this shuffling scheme were generated using the

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

PALM toolbox (Winkler, Ridgway, Webster, Smith, & Nichols, 2014). All statistical imaging analyses were conducted using custom code written in Matlab R2017a.

All imaging analyses included sex, self-declared race and ethnicity, household income, parental education and scanner as categorical fixed covariates and age as a standardised (z-scored) continuous fixed covariate. For vertex-wise cortical thickness analyses, standardised global mean thickness was included as a covariate, and for vertex-wise cortical surface area analyses, standardised total surface area was included.

Statistical Analysis 2: quantifying the magnitude of the association between brain structure and cognition using the PVS

The PVS was calculated for each cognitive task to determine the behavioural variance explained by the vertex-wise cortical morphology. All behavioural and imaging data were pre-residualised using the covariates of no interest prior to calculation of the PVS. For the imaging data, the global CSA and CTH measures specific to each modality were also included in this. This allowed us to determine the unique association between relative cortical morphology and cognition and compare this to the predictive power of brain structure without controlling for global measures. The method used here is outlined in detail in Zhao et al (in prep). The association between each imaging phenotype and each cognitive task was modelled using the mass univariate approach such that the behaviour of interest was modelled independently at each vertex using a general linear model (GLM). In the equations described below, $\boldsymbol{X}$ denotes a standardized N by V matrix of a given imaging phenotype where N and V denotes the number of subjects and vertices respectively. y represents an N by 1 vector of a behavioural phenotype, and β is the V by 1 vector of parameters of interest.

Parameter estimation

Across subjects, the imaging data and the behaviour were modelled as a linear combination of the imaging data at each vertex plus measurement error (Eq 5). This is used to generate mass univariate beta estimates, $\widehat{\beta_U}$, based on the assumption that each vertex is independent (Eq 6).

(5) $\mathrm{y} = \boldsymbol{X}\beta + \varepsilon$

(6) $\widehat{\beta_U} = I\boldsymbol{X}'y$

However, the assumption that the signal at each vertex is independent is not correct. We therefore used a Bayesian parameter estimation procedure adapted from the LDpred framework used to improve accuracy for polygenic risk scores(Vilhjálmsson et al., 2015). Full details of this procedure are outlined in Zhao et al (in prep) where the authors applied this method to imaging data and show improvements in prediction performance. This Bayesian parameter estimation procedure calculates the posterior mean effect size at each vertex using the mass univariate beta estimates, the observed correlation structure of the brain structure, D, the estimated signal-to-noise ratio (SNR) of the imaging phenotype, S , (computed using the GWASH framework(Schwartzman, Schork, Zablocki, & Thompson, 2017)), the number of vertices, V, and sample size, N (Eq 7).

(7) $E(\beta_{Bayesian} \mid \hat{\beta}_U, \boldsymbol{D}) = (\frac{V}{NS} + \boldsymbol{D})^{-1}\hat{\beta}_U$

The Bayesian parameter estimation procedure takes into account the correlation structure of the brain providing a more valid estimate of the association between the brain and behaviour. Furthermore, by incorporating an estimate of the SNR of the imaging phenotype we can reduce the

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

variance in prediction error, which leads to better predictive performance. We note that this method explicitly assumes a prior that all vertices have true associations with behaviour.

Behavioural prediction

The estimated beta parameters were then used to predict the behavioural phenotype and thus produce the PVS for each subject. The PVS was calculated by aggregating the Bayesian parameter estimates across all of the vertices (Eq 8). This measure harnesses the explanatory power of all of the vertices on behaviour. The PVS can then be compared with the observed behaviour in order to provide an estimate for how much variance in the observed behaviour can be predicted using the vertex-wise imaging phenotype.

$$(8)\ \hat{y}_{PVS-B} = \sum_j^V X_j \hat{\beta}_{Bayesian,j}$$

In order to generate an unbiased PVS for every subject we used a leave-one-out 10 fold cross-validation procedure. The Bayesian parameters were estimated in the training set (90% of full sample) and multiplied with the imaging phenotype of participants in the test set (10% of full sample). This was repeated 10 times for each fold until a PVS was calculated for every participant in the full sample. The subjects in each fold were randomly selected based on unique family IDs, such that subjects within the same family were always within the same fold. The association between the imaging phenotype and behaviour across the whole sample was calculated as the squared correlation (R^2) between the observed behaviour and the predicted behaviour (the PVS). This process was repeated for four imaging phenotypes: CSA, relative CSA (controlling for total CSA), CTH. relative CTH (controlling for mean CTH).

RESULTS

Determining the association between relative cortical morphology and cognition

In order to test the association between the regionalisation of cortical morphology across the whole cortical surface and cognition in this study, we employed the MOSTest. The multivariate statistic was calculated and used within a permutation test for each cognitive task and cortical morphology measure independently.

Relative CSA

For each cognitive task, our multivariate statistic, ${\chi^2}_{MOST}$, was calculated for each permutation (n=10,000) for statistical inference. In figure 1, we have plotted the distribution of ${\chi^2}_{MOST}$ across permutations and superimposed the observed (non-permuted) statistic (green) to demonstrate the likelihood of observing the non-permuted statistic by chance. The Bonferroni corrected alpha threshold (0.05/13=0.0038) is overlaid (red) on each plot in order to demonstrate the significance of the permutation test. The MOSTest showed a significant association between relative cortical configuration of CSA (controlling for total CSA) and cognition for all of the cognitive performance measures. We also conducted the same permutation test using the most significant vertex (${z^2}_{MINP}$) as the test statistic for comparison. The results are shown in supplementary figure 1. Here we can see that using the same threshold for significance the flanker task, pattern speed, picture sequencing, LMT and RAVLT did not show significant associations.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Relative CTH

In figure 2, we show the permuted and non-permuted statistics for the association between cognitive performance and the relative cortical configuration of CTH (controlling for mean CTH). The relative cortical configuration of CTH was significantly associated with all of the cognitive measures at the Bonferroni corrected alpha level (p<0.0038) according to the MOSTest. The same permutation test was carried out using the most significant vertex, z^2_{MINP} , and the results are shown in supplementary figure 2. According to this analysis, for comparison, the following measures would not have been considered to show a significant association with relative CSA: picture vocabulary, flanker, pattern speed, LMT and RAVLT.

Across both relative CSA and CTH, the p-values for the MOSTest associations were magnitudes smaller than for the min-p associations (supplementary table 2). The increased power of the MOSTest is likely due to the distributed nature of the associations between cognition and relative cortical morphology. This can clearly be seen in the surface maps of the unthresholded mass univariate effect sizes for relative CSA and CTH predicting the total composite score from the NIH Toolbox (Figure 3).

Quantifying the association between brain structure and cognition

Having established there was a significant association between relative cortical configuration and cognition across tasks, we aimed to quantify the unique proportion of individual variability in each behaviour predicted by the relative cortical configuration. To do this we used the PVS in order to provide a conservative, out-of-sample estimate of the variance in each behaviour predicted by the relative vertex-wise imaging data. Each PVS was orthogonal to the modality specific global imaging measures (total CSA and mean CTH respectively) as all imaging data were pre-residualised using these measures (and the additional covariates of no interest) prior to calculating the PVS. We then used a linear model to determine the unique contribution of each PVS (relative CSA and relative CTH) to cognitive task performance and compared to this to the independent contribution of the global imaging measures. Figure 4 shows the variance (%R^2) in each cognitive task performance explained by total CSA (dark blue), relative CSA (light blue), mean CTH (dark pink) and relative CTH (light pink).

Total CSA predicted more variance in cognitive performance compared to relative CSA for the crystallised measures (reading recognition, picture vocabulary, crystallised composite score and the total composite score); however, this relationship was not always true for the other cognitive measures. Indeed, for the fluid composite score, total and relative CSA predicted a similar proportion of variance. This suggests that brain size is more predictive of crystallised compared to fluid cognitive measures. Mean CTH was not predictive of cognitive performance across any domains, however, relative CTH was predictive across cognitive domains. In general, the variance in behaviour explained by these morphology measures was very low with the largest association at 1.45%R^2; however, in this analysis we have residualised for a wide range of confounding demographic variables (race and ethnicity, household income and parental education), which removes a large portion of shared variance amongst these variables, brain structure and cognition. These estimates of R^2 therefore provide a conservative lower bound of the contribution of cortical morphology to cognition that is not attributable to confounding variables.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

DISCUSSION

In this study, we have used a novel multivariate omnibus statistical test, the MOSTest, to demonstrate that individual differences in the regionalisation of cortical morphology can predict cognitive performance across multiple domains in a large sample of 9 and 10 year old children. This method aggregates many small effects across the cortical surface to improve our power for discovering distributed brain-behaviour associations. The MOSTest has allowed us to uncover patterns of association across the cortical surface between the regionalisation of cortical morphology and cognition that have not previously been detected. We then used the PVS to quantify the effect size of these discovered associations. The regionalisation of CSA and CTH predicted unique variance in cognition beyond that predicted by global brain measures, however the relative contribution of arealisation and total CSA to predicting behaviour differed depending on the cognitive modality being measured. These results highlight the importance of studying the configuration of brain structure relative to global brain measures. Individual differences in these regionalisation patterns may uncover important mechanisms underlying the development of cognitive processes.

A novel multivariate vertex-wise statistical inference tool

In order to make a statistical inference about the association between cortical configuration and cognition in this study we developed the MOSTest. This test uses a statistic that aggregates effects of interest across the cortex; therefore, it is optimally powered to detect effects that are distributed across the cortical surface. Traditionally, neuroimaging analysis methods assume neuroimaging effects are sparse and localised, therefore adopt the most significant vertex or voxel as an omnibus test statistic. However, this can lead to reduced statistical power and an increase in Type II errors (Cremers et al., 2017) particularly if the association between the brain and behaviour is continuous and distributed across the cortex. The continuous distribution of brain-behaviour associations has recently been highlighted across modalities using large samples such as the ABCD study. Indeed, a recent paper introducing the PVS, demonstrated that aggregating information across the cortex could produce greater out-of-sample prediction performance of task fMRI-behaviour associations than thresholding based on significance level (Zhao et al, in prep). This mirrors findings from genetics, which demonstrate that many complex behavioural phenotypes are polygenic: genome-wide significant single nucleotide polymorphisms (SNPs) only predict a very small proportion of the variance in complex behavioural phenotypes, however collating effects across the entire genome increases the variance explained. Zhao and colleagues have shown that the same is true when integrating effects across the cortex.

In this study, the MOSTest detected a greater number of significant associations compared to using the most significant vertex as an omnibus test, which demonstrates the increased power of this test for the discovery of vertex-wise associations with behaviour. Van der Meer and colleagues have applied the MOSTest to detect significant genome-wide associations with regional neuroimaging phenotypes using UK Biobank data and demonstrated a 3-fold increase in statistical power using this method compared to standard univariate analyses (Meer et al., 2019). This method improves power by capitalising on the shared variance across vertices and by negating the need to use stringent vertex-wise correction for multiple comparisons across the cortical surface; but importantly it does not inflate the type-1 error rate due to the use of a permutation test for statistical inference. Additionally, the MOSTest test statistic takes into account the correlation structure of the data across vertices rather than using the smoothness of the imaging measure to estimate an effective degrees of freedom for multiple comparison correction of the univariate test

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

statistics. The multivariate test statistic, the Mahalanobis norm, represents the sum of squared deviations of the vector of decorrelated mass univariate z statistics. This is a generalisation of measuring the Euclidean distance across a decorrelated multidimensional space. Using this multivariate method we can readily test for associations between the brain and behaviour using information across the entire cortex with increased statistical power. Although it is clear that not all behaviours will be associated with widespread cortical activity or brain structure, these data suggest that we should not assume sparseness of neuroimaging associations.

Relative cortical configuration is important for cognition

Global brain measures have been previously associated with cognition; however, there are many examples where there are differences in global brain measures, such as brain size between men and women, which do not correspond with differences in cognition. When measuring regional associations with behaviour it is therefore essential to control for these global measures in order to uncover specific relationships between regionalisation and behaviour. Indeed, a number of studies have highlighted associations between regional differences in cortical morphology and cognition; however, they have lacked the power and multivariate methods to measure associations specifically based on the entire configuration of cortical morphology. Importantly, the regionalisation of CSA and CTH has been associated with distinct, continuous gradients of genetic influences across the cortex (C.-H. Chen et al., 2013; C. Chen et al., 2013; Chi-Hua Chen et al., 2011) that coincide with the gene expression patterns that dictate specialisation of the neocortex during embryonic development. Individual differences in this molecular signalling could therefore lead to subtle alterations in this cortical configuration, which could lead to variability in behaviour. It is therefore important to use a statistical test that captures the graded and distributed nature of the biology of regionalisation, such as the MOSTest.

Here we have shown that performance across multiple cognitive tasks was associated with the regionalisation of cortical morphology across the whole cortex. The presence of negative and positive associations across the cortical surface suggests that it is the patterning of cortical morphology that may be important for supporting cognitive function. Indeed, the regionalisation of CSA predicted additional unique variance in cognition not explained by total CSA. This allows us to partition the variance associated with cortical morphology in order determine the relative importance of these components of CSA for cognition. This appeared to differ depending on the cognitive task, with more crystallised measures having a greater association with total CSA compared to the regionalisation of CSA. Comparatively the proportion of variance in behaviour explained by regional CTH was relatively similar for fluid and crystallised measures. Mean CTH did not predict cognitive performance across any measures; however this may be due to the developmental stage of this sample. These results support the independence of CSA and CTH and suggest they may have differential influences across different cognitive domains.

It is important to note that the effect sizes reported between brain structure and cognition in this study are relatively small (~1% R^2). This is important because we have used a very conservative, out-of-sample effect size estimate and have residualised for a number of confounding demographic variables, which are highly collinear with these measures. We are therefore providing the lower bound for the estimated effect size between regionalisation and cognition, which is not influenced by any confounding measures. Nevertheless, cognitive processing is highly complex, therefore we are unlikely to explain large proportions of individual variability using single neuroimaging modalities. Future work combining imaging measures within multi-modal predictive models will likely uncover larger associations between brain phenotypes and cognition.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Conclusions

With the availability of large multi-dimensional datasets such as the ABCD study, we can now begin developing more advanced predictive models of behaviour, which will allow for more model complexity whilst still remaining interpretable. It is essential that we generate models that can be used to understand how stable individual differences in complex behaviours relate to individual differences in brain phenotypes and their development. Future work will aim to understand how these structural associations change over time with longitudinal assessments in ABCD. Moreover, we aim to further elucidate the heterogeneity in the pattern of associations between cortical morphology and cognition across subjects and across cognitive tasks.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ACKNOWLEDGEMENTS

The authors wish to thank the youth and families participating in the Adolescent Brain Cognitive Development (ABCD) Study and all ABCD staff. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/Consortium_Members.pdf. ABCD consortium investigators designed and implemented the study and/or provided data but did not all necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The data was downloaded from the NIMH Data Archive ABCD Collection Release 2.0.1 (DOI: 10.15154/1504041).

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

REFERENCES

Acker, W. (1982). *A computerized approach to psychological screening—The Bexley-Maudsley Automated Psychological Screening and The Bexley-Maudsley Category Sorting Test*. *International Journal of Human-computer Studies / International Journal of Man-machine Studies - IJMMS* (Vol. 17). https://doi.org/10.1016/S0020-7373(82)80037-0

Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., … Heaton, R. K. (2013). VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development*, *78*(4), 119–132. https://doi.org/10.1111/mono.12038

Brown, T. T., & Jernigan, T. L. (2012). Brain Development During the Preschool Years. *Neuropsychology Review*, *22*(4), 313–333. https://doi.org/10.1007/s11065-012-9214-1

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/J.DCN.2018.03.001

Chen, C.-H., Fiecas, M., Gutierrez, E. D., Panizzon, M. S., Eyler, L. T., Vuoksimaa, E., … Kremen, W. S. (2013). Genetic topography of brain morphology. *Proceedings of the National Academy of Sciences*, *110*(42), 17089–17094. https://doi.org/10.1073/pnas.1308091110

Chen, C., Gutierrez, E. D., Thompson, W., Panizzon, M. S., Terry, L., Eyler, L. T., … Dale, A. M. (2013). Hierarchical Genetic Organisation of Human Cortical Surface Area, *335*(6076), 1634–1636. https://doi.org/10.1126/science.1215330.Hierarchical

Chen, Chi-Hua, Panizzon, M. S., Eyler, L. T., Jernigan, T. L., Thompson, W., Fennema-Notestine, C., … Dale, A. M. (2011). Genetic Influences on Cortical Regionalization in the Human Brain. *Neuron*, *72*(4), 537–544. https://doi.org/10.1016/j.neuron.2011.08.021

Compton, W. M., Dowling, G. J., & Garavan, H. (2019). Ensuring the Best Use of Data. *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2019.2081

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLOS ONE*, *12*(11), e0184923. https://doi.org/10.1371/journal.pone.0184923

Curley, L. B., Newman, E., Thompson, W. K., Brown, T. T., Hagler, D. J., Akshoomoff, N., … Jernigan, T. L. (2018). Cortical morphology of the pars opercularis and its relationship to motor-inhibitory performance in a longitudinal, developing cohort. *Brain Structure & Function*, *223*(1), 211–220. https://doi.org/10.1007/s00429-017-1480-5

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Daniel, M. H., Wahlstrom, D., & Zhang, O. (2014). *Equivalence of Q-interactive™ and Paper Administrations of Cognitive Tasks: WISC ®-V Q-interactive Technical Report 8*. Retrieved from http://www.helloq.com.au/userfiles/830351450329565.pdf

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fischl, B, & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055. https://doi.org/10.1073/pnas.200033797

Fischl, Bruce, Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, *23*, S69--S84. https://doi.org/10.1016/j.neuroimage.2004.07.016

Fischl, Bruce, Sereno, M. I., & Dale, A. M. (1999). *Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System*. Retrieved from http://www.idealibrary.com

Fjell, A. M., Walhovd, K. B., Brown, T. T., Kuperman, J. M., Chung, Y., Hagler, D. J., … Gruen, J. (2012). Multimodal imaging of the self-regulating developing brain. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19620–19625. https://doi.org/10.1073/pnas.1208243109

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., … Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., … Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*, 116091. https://doi.org/10.1016/J.NEUROIMAGE.2019.116091

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., … Gershon, R. (2014). Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society : JINS*, *20*(6), 588–598. https://doi.org/10.1017/S1355617714000241

Jernigan, T. L., & Tallal, P. (1990). Late childhood changes in brain morphology observable with MRI. *Developmental Medicine and Child Neurology*, *32*(5), 379–385. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2354751

Jernigan, T. L., Trauner, D. A., Hesselink, J. R., & Tallal, P. A. (1991). Maturation of human cerebrum observed in vivo during adolescence. *Brain : A Journal of Neurology*, *114 (Pt 5*, 2037–2049. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1933232

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., … Dale, A. (2006). Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, *30*(2), 436–443. https://doi.org/10.1016/J.NEUROIMAGE.2005.09.046

Lyall, A. E., Shi, F., Geng, X., Woolson, S., Li, G., Wang, L., … Gilmore, J. H. (2015). Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cerebral Cortex*, *25*(8), 2204–2212. https://doi.org/10.1093/cercor/bhu027

Meer, D. van der, Frei, O., Kaufmann, T., Shadrin, A. A., Devor, A., Smeland, O. B., … Dale, A. M. (2019). Making the MOSTest of imaging genetics. *BioRxiv*, 767905. https://doi.org/10.1101/767905

Newman, E., Jernigan, T. L., Lisdahl, K. M., Tamm, L., Tapert, S. F., Potkin, S. G., … Group, M. N. (2016). Go/No Go task performance predicts cortical thickness in the caudal inferior frontal gyrus in young adults with and without ADHD. *Brain Imaging and Behavior*, *10*(3), 880–892. https://doi.org/10.1007/s11682-015-9453-x

Newman, E., Thompson, W. K., Bartsch, H., Hagler, D. J., Chen, C.-H., Brown, T. T., … Jernigan, T. L. (2016). Anxiety is related to indices of cortical maturation in typically developing children and

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

adolescents. *Brain Structure and Function*, *221*(6), 3013–3025. https://doi.org/10.1007/s00429-015-1085-9

Schwartzman, A., Schork, A. J., Zablocki, R., & Thompson, W. K. (2017). A simple, consistent estimator of heritability for genome-wide association studies. *BioRxiv*, 204446. https://doi.org/10.1101/204446

Tisdall, M. D., Hess, A. T., Reuter, M., Meintjes, E. M., Fischl, B., & van der Kouwe, A. J. W. (2012). Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magnetic Resonance in Medicine*, *68*(2), 389–399. https://doi.org/10.1002/mrm.23228

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., … Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics*, *97*(4), 576–592. https://doi.org/10.1016/J.AJHG.2015.09.001

Vuoksimaa, E., Panizzon, M. S., Chen, C.-H., Fiecas, M., Eyler, L. T., Fennema-Notestine, C., … Kremen, W. S. (2016). Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *NeuroImage*, *129*, 356–366. https://doi.org/10.1016/j.neuroimage.2016.01.049

Weschler, D. (2014). *Weschler Intelligence Scale for Children, 5th ed*. Pearson, Bloomington, MN.

White, N., Roddey, C., Shankaranarayanan, A., Han, E., Rettmann, D., Santos, J., … Dale, A. (2010). PROMO: Real-time prospective motion correction in MRI using image-based tracking. *Magnetic Resonance in Medicine*, *63*(1), 91–105. https://doi.org/10.1002/mrm.22176

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, *87*, 120–126. https://doi.org/10.1016/J.NEUROIMAGE.2013.11.010

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. https://doi.org/10.1016/J.NEUROIMAGE.2014.01.060

Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E., & Smith, S. M. (2015). Multi-level block permutation. *NeuroImage*, *123*, 253–268. https://doi.org/10.1016/J.NEUROIMAGE.2015.05.092

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

FIGURES

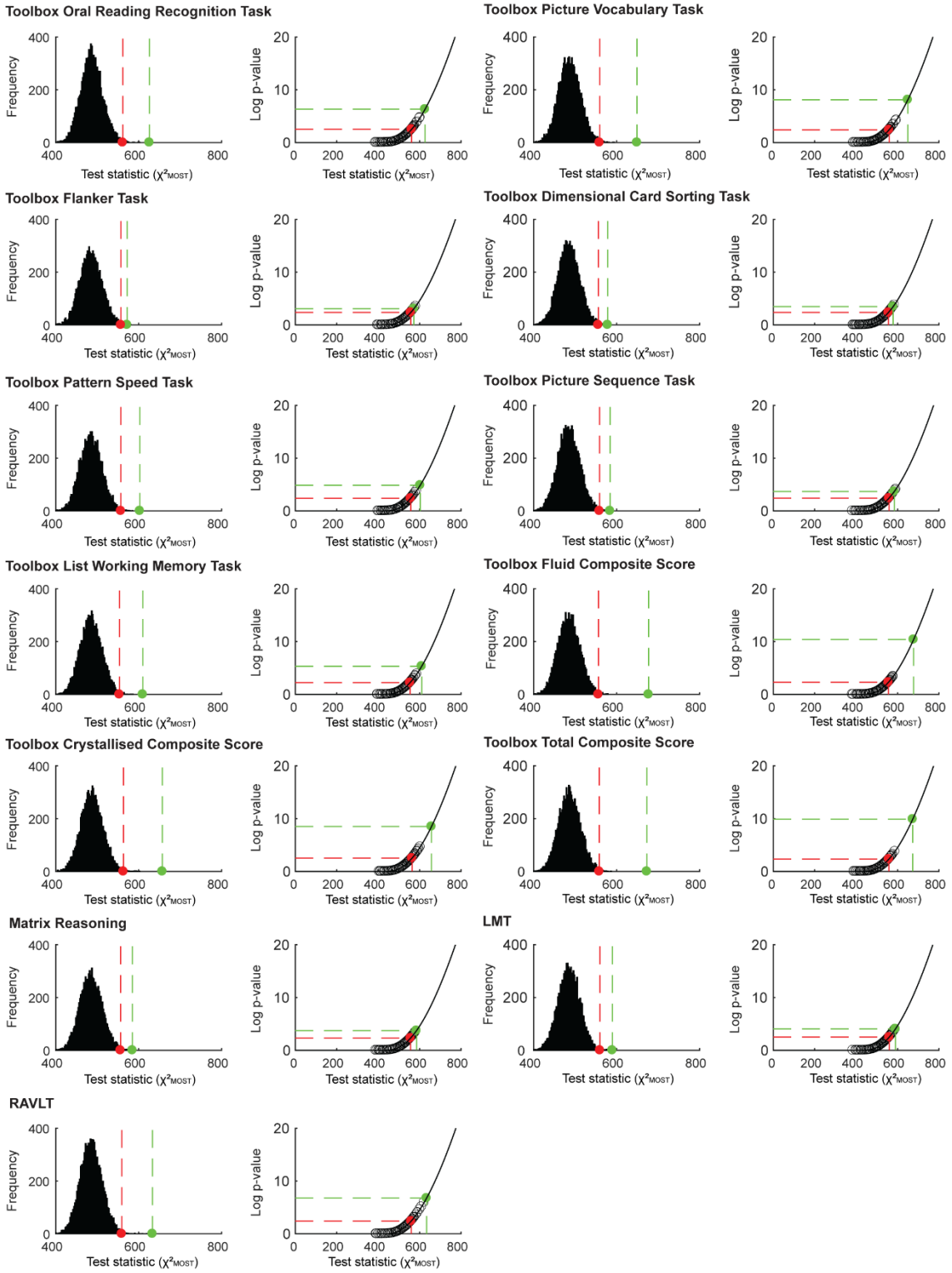

Figure 1. Multivariate omnibus permutation test for each cognitive task and relative CSA. The multivariate test statistic ($\chi^2{}_{MOST}$) was calculated as the Mahalanobis norm of the decorrelated mass univariate z statistics for the association between relative CSA and each of the cognitive tasks. All of the left side plots show the permuted null distribution of this statistic for each cognitive task. The observed, non-permuted test statistic is shown in green and the Bonferroni corrected alpha level in red. All of the right side plots show the likelihood values for the permuted observations after fitting a gamma cumulative density function to the tail of the null distribution. This can be used to approximate a theoretical p-value for the observed statistic.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

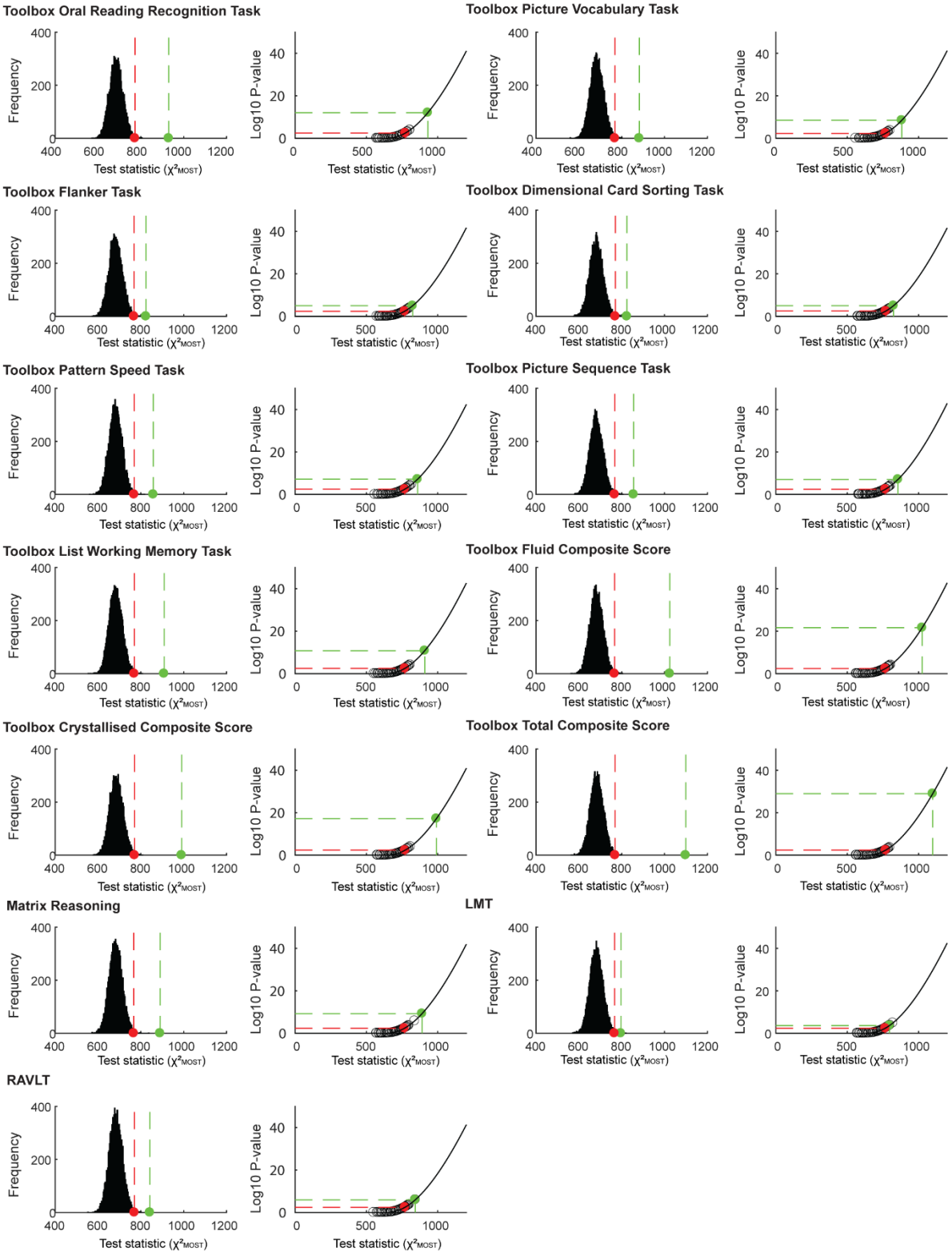

Figure 2. Multivariate omnibus permutation test for each cognitive task and relative CTH. The multivariate test statistic ($\chi^2{}_{MOST}$) was calculated as the Mahalanobis norm of the decorrelated mass univariate z statistics for the association between relative CTH and each of the cognitive tasks. All of the left side plots show the permuted null distribution of this statistic for each cognitive task. The observed, non-permuted test statistic is shown in green and the Bonferroni corrected alpha level in red. All of the right side plots show the likelihood values for the permuted observations after fitting a gamma cumulative density function to the tail of the null distribution. This can be used to approximate a theoretical p-value for the observed statistic.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

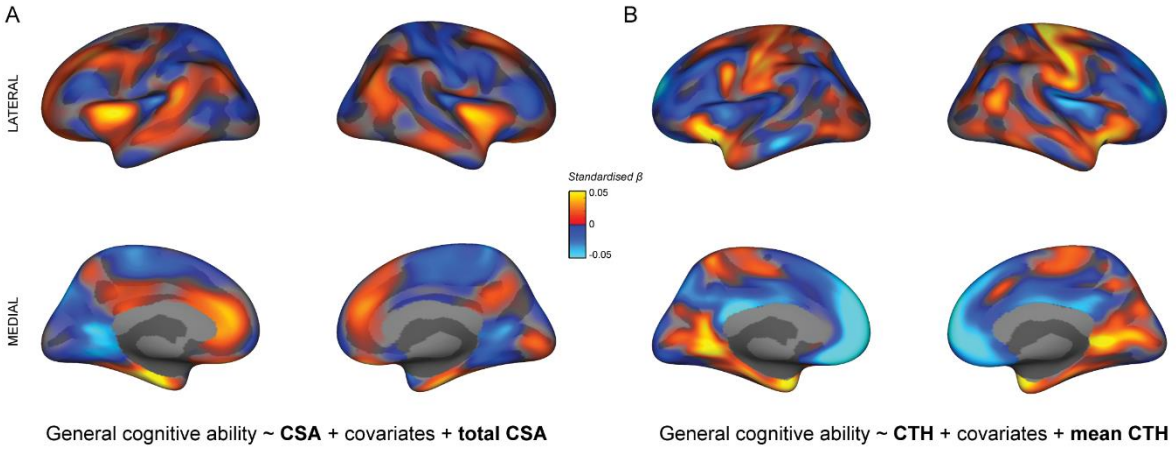

Figure 3. Surface maps showing the mass univariate standardised beta coefficients for the association between the NIH Toolbox Total Composite score and (A) relative CSA and (B) relative CTH. These unthresholded effect size maps show the continuously distributed associations across the cortical surface. These distributed patterns of association support the need to use a multivariate statistic that incorporates this entire pattern to determine the relationship between relative cortical morphology and cognition.

bioRxiv preprint doi: https://doi.org/10.1101/816025; this version posted October 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

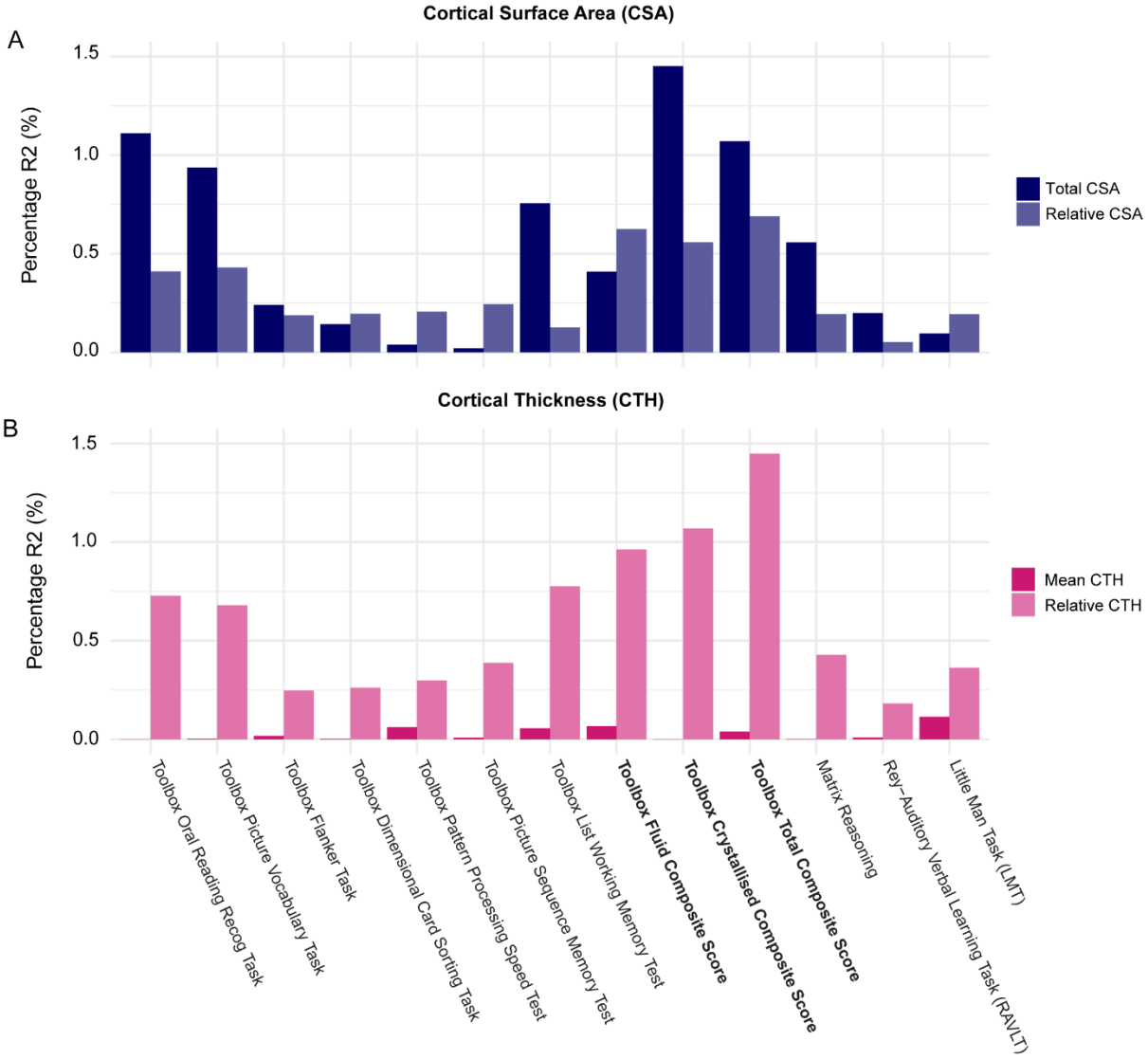

Figure 4. Percent behavioural variance explained across cognitive tasks by CSA and CTH. Variance explained was calculated by correlating each PVS with the observed behaviour for each cognitive task to generate a %R^2 value. Each PVS was generated by aggregating the Bayesian estimated effect sizes across the whole cortex from the association between the imaging data and behaviour. The composite scores, which represent averages of single task performance, are highlighted in bold. The percent variance explained for each cognitive measure was determined for the following imaging phenotypes: CSA (dark blue), relative CSA (controlling for total CSA; light blue), CTH (dark pink), relative CTH (controlling for mean CTH; light pink).