



Generalization of cortical MOSTest genome-wide associations within and across samples

Robert J. Loughnan^{a,b,i,*}, Alexey A. Shadrin^d, Oleksandr Frei^{c,d}, Dennis van der Meer^{c,e},
Weiqi Zhao^a, Clare E. Palmer^f, Wesley K. Thompson^{b,g,h}, Carolina Makowski^{i,j},
Terry L. Jernigan^{a,f,j,k}, Ole A. Andreassen^d, Chun Chieh Fan^{b,h,j,m}, Anders M. Dale^{a,i,j,l,*}

^a Department of Cognitive Science, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92093, USA

^b Population Neuroscience and Genetics, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92161, USA

^c Division of Mental Health and Addiction, NORMENT Centre, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway

^d Department of Informatics, Centre for Bioinformatics, University of Oslo, Oslo, Norway

^e Faculty of Health, Medicine and Life Sciences, School of Mental Health and Neuroscience, Maastricht University, the Netherlands

^f Center for Human Development, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92161, USA

^g Herbert Wertheim School of Public Health, University of California, La Jolla, San Diego, CA, USA

^h Center for Population Neuroscience and Genetics, Laureate Institute for Brain Research, Tulsa, OK, USA

ⁱ Center for Multimodal Imaging and Genetics, San Diego School of Medicine, University of California, 9444 Medical Center Dr, La Jolla, CA 92037, USA

^j Department of Radiology, San Diego School of Medicine, University of California, 9500 Gilman Drive, La Jolla, CA 92037, USA

^k Department of Psychiatry, San Diego School of Medicine, University of California, 9500 Gilman Drive, La Jolla, CA 92037, USA

^l Department of Neuroscience, San Diego School of Medicine, University of California, 9500 Gilman Drive, La Jolla, CA 92037, USA

^m Department of Radiology, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92037, USA

A B S T R A C T

Genome-Wide Association studies have typically been limited to univariate analysis in which a single outcome measure is tested against millions of variants. Recent work demonstrates that a Multivariate Omnibus Statistic Test (MOSTest) is well powered to discover genomic effects distributed across multiple phenotypes. Applied to cortical brain MRI morphology measures, MOSTest has resulted in a drastic improvement in power to discover loci when compared to established approaches (min-P). One question that arises is how well these discovered loci replicate in independent data. Here we perform 10 times cross validation within 34,973 individuals from UK Biobank for imaging measures of cortical area, thickness and sulcal depth (>1,000 dimensionality for each). By deploying a replication method that aggregates discovered effects distributed across multiple phenotypes, termed PolyVertex Score (MOSTest-PVS), we demonstrate a higher replication yield and comparable replication rate of discovered loci for MOSTest (# replicated loci: 242–496, replication rate: 96–97%) in independent data when compared with the established min-P approach (# replicated loci: 26–55, replication rate: 91–93%). An out-of-sample replication of discovered loci was conducted with a sample of 4,069 individuals from the Adolescent Brain Cognitive Development® (ABCD) study, who are on average 50 years younger than UK Biobank individuals. We observe a higher replication yield and comparable replication rate of MOSTest-PVS compared to min-P. This finding underscores the importance of using well-powered multivariate techniques for both discovery and replication of high dimensional phenotypes in Genome-Wide Association studies.

1. Introduction

Performing Genome Wide Association Studies (GWAS) on high dimensional phenotypes using methods that fail to model pleiotropic effects distributed across phenotypes can result in low power to detect associations. Vertex-wise measures of cortical morphology (area, thickness and sulcal depth) represent high dimensional phenotypes (>1000 dimensions) and, from twin studies, are known to have high heritabilities of up to 90% and 50% for total and regional area respectively, and 80% and 60% for mean and regional thickness respectively (Panizzon et al., 2009; Eyer et al., 2012). Our group has previously developed a novel Multivariate Omnibus Test (MOSTest) (van der Meer et al., 2020;

van der Meer et al., 2021; Shadrin et al., 2021), which aggregates the effect of a genomic variant across the cortex. This method significantly boosts discovery of genetic loci linked to cortical morphology, with an up to 10x increase in number of loci discovered – when compared to an established approach (min-P) deployed for the same phenotypes (Shadrin et al., 2021). Additionally, discovered loci show strong enrichment with pathways involved in neurogenesis and cell differentiation. The major statistical boost in power of MOSTest over established techniques, like min-P, is its ability to aggregate pleiotropic effects across the dimensionality of phenotypes into a single statistical test – whilst appropriately correcting for genome-wide multiple comparisons. Given such a dramatic increase in discovery of genomic loci, it is of interest

* Corresponding authors at: Center for Human Development, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92161, USA
E-mail addresses: rloughna@ucsd.edu (R.J. Loughnan), andersmdale@gmail.com (A.M. Dale).

<https://doi.org/10.1016/j.neuroimage.2022.119632>.

Received 20 May 2022; Received in revised form 25 August 2022; Accepted 13 September 2022

Available online 14 September 2022.

1053-8119/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

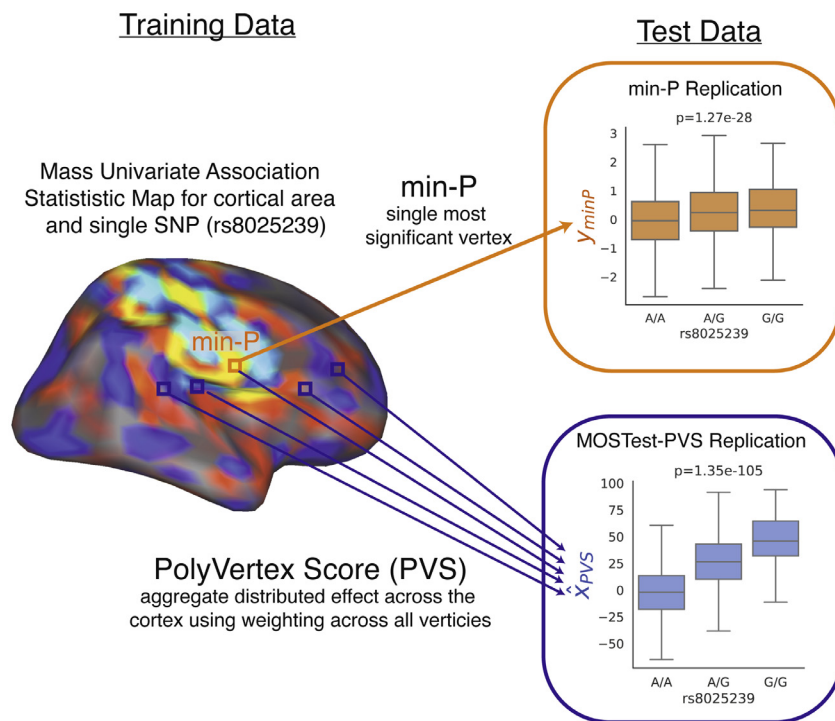


Fig. 1. Schematic of replication process for a single SNP. Variant rs8025239 is discovered in training fold and has mass univariate map of association statistics with cortical area. Min-P replication (indicated by orange box and arrow) takes most significant vertex and associates that vertex with variant rs8025239 in test data. MOSTest-PVS replication (indicated by blue box and arrow) computes a PolyVertex Score (PVS) in test data which aggregates all effects across the cortex by taking a weighted sum (using association statistics from training set) across all vertices – the PVS is then correlated with the variant rs8025239. This process is repeated for all discovered variants in training set with a separate PVS being generated for each MOSTest discovery. Replication of a variant is defined as $p < 0.05$ in one tailed t -test.

to understand how well these discoveries replicate in independent data.

Here we perform 10-times cross validation with simulated and real brain imaging data taken from the UK Biobank, and randomly split the sample into $\frac{2}{3}$ training and $\frac{1}{3}$ replication splits. For the training samples we perform discovery of vertex-wise measures of area, thickness and sulcal depth as in (van der Meer et al., 2021). Having discovered genomic loci in training folds, we perform replication of these loci in the test sets. To perform replication for each SNP we calculate a PolyVertex Score (PVS) (similar to (Zhao et al., 2019, 2021)) from imaging data in the test set for each MOSTest discovered locus. This PVS aggregates the distributed effects across the cortex by taking a weighted sum across all vertices using mass univariate z statistics as weights from the training set. This approach is similar to the widely used method of Polygenic Risk Scores (PRS) in genetics (Lewis and Vassos, 2020), where instead of predicting a phenotype we are predicting a single genomic variant and instead of using distributed effects across the genome as predictors we use the distributed effects across the cortex, estimated in the training set. For each discovered loci in the training set we generate a PVS for each individual, which represents a continuous prediction of the genotype in the test set. We then correlate each PVS with its corresponding measured genomic variant in the test set to test how well these discovered loci replicate (one tailed t -test, $p < 0.05$). We test this MOSTest discovery and PVS replication (termed MOSTest-PVS), against an established GWAS approach (min-P) (Grasby, 2020). Fig. 1 displays a schematic of how replication of how min-P and MOSTest-PVS differs for a single discovered variant. Firstly, we perform this process on simulated data under a range of different genetic architectures to demonstrate higher power to detect and replicate associations for MOSTest-PVS vs min-P while appropriately controlling type-I errors. Next, repeating this for real imaging data from the UK Biobank, we confirm a higher replication yield and comparable replication rate MOSTest-PVS versus min-P. Finally, we test the generalization of loci discovered in UK Biobank to a developmental cohort of 9–10 year old children from the Adolescent Brain Cognitive Development® (ABCD; <https://abcdstudy.org>) Study, where we see a higher yield of replicated loci for MOSTest-PVS versus min-P.

2. Results

2.1. Simulations

Across a range of simulated conditions MOSTest-PVS exhibited greater power than min-P to discover and replicate causal variants – see Fig. 2. At low polygenic ($nc=10$) phenotypes the two methods showed comparable power. Both min-P and MOSTest-PVS demonstrated a well-controlled type 1 error rate, with MOSTest-PVS having zero and min-P having one false positive replication across all simulations and folds. For min-P across 12 sets of simulated phenotypes, 10 folds and 7.2 million variants this single false positive represents an error rate of less than 10^{-8} . This low type 1 error rate for both methods is likely due to the stringent criteria required for variants to be both discovered ($p < 10^{-8}$) in training and replicated (0.05) on test data.

2.2. Real data – cross validation

Across training folds, the UK Biobank sample, we confirm that MOSTest confers up to a 10-fold increase in discovered loci over min-P – see solid bars Fig. 3. When replication of loci is defined at the nominal level ($p < 0.05$, see methods) we see a higher number of replicated loci for MOSTest-PVS (area: 292, thickness: 242, sulcal depth: 496) vs min-P (area: 34, thickness: 26, sulcal depth: 55), as well as comparable replication rate for MOSTest-PVS (area: 96%, thickness: 97%, sulcal depth: 96%) vs min-P (area: 91%, thickness: 93%, sulcal depth: 92%) – see hatched bars Fig. 3. Averaged across cross-validation folds, we found that the lead SNP of the top locus accounted for more variance in the replication set with MOSTest-PVS (R^2 (95%CI)= area: 0.029 (0.024–0.034), thickness: 0.058 (0.040–0.076), sulcal depth: 0.043 (0.037–0.049)) compared to min-P (R^2 (95%CI)= area: 0.0092 (0.0062–0.012), thickness: 0.012 (0.010–0.14), sulcal depth: 0.014 (0.0095–0.019)). If replication is defined more conservatively with significance corrected for the number of discovered loci ($p < 0.05/\#$ of discovered loci), we again find that MOSTest-PVS confers a comparable replication rate (area: 71%, thickness: 77%, sulcal depth: 70%) to min-P (area: 66%, thickness: 77%, sulcal depth: 67%).

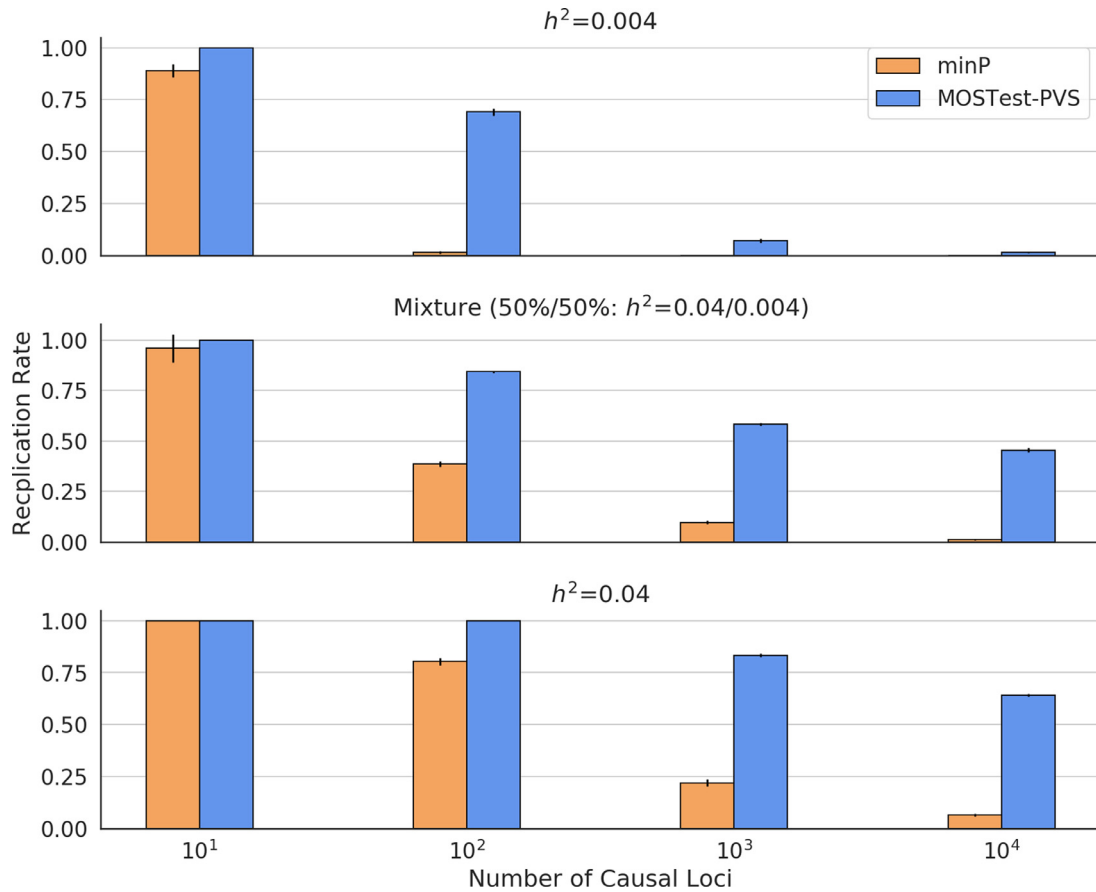


Fig. 2. Power to detect and replicate causal associations from simulated high dimensional phenotypes ($T = 1175$) for min-P and MOSTest-PVS under different genetic architectures. Error bars represent standard deviations across cross-validation folds.

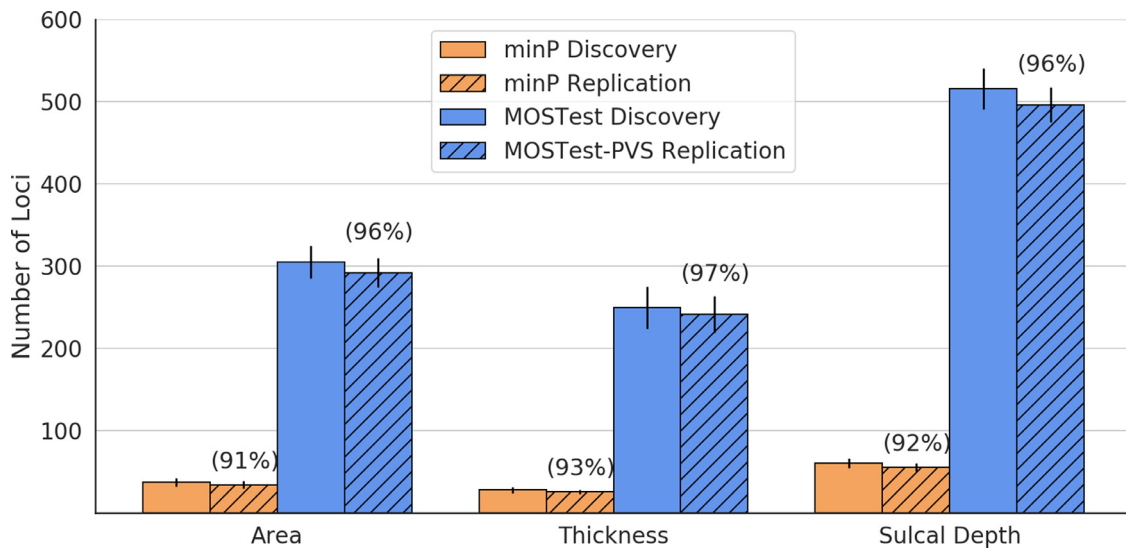


Fig. 3. Cross-validation discovery and replication yield within 10-times cross validation within UK Biobank for cortical morphometry measures. Solid bars represent the number of genome wide significant loci associated with each measure. Hashed bars represent the number of loci that replicate in test folds at a nominal significance level ($p < 0.05$). Error bars are standard deviations across 10 cross-validation repetitions. Numbers in parentheses represent replication rate (# of discovered loci / # replicated loci) for each method-phenotype pair.

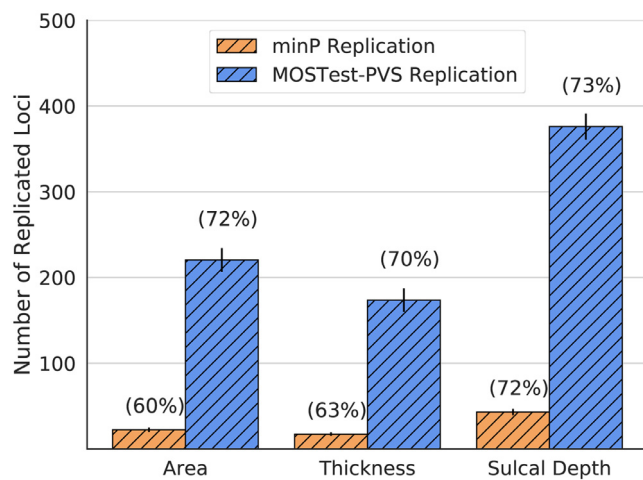


Fig. 4. Replication yield within the ABCD dataset across 10 training folds of UK Biobank for cortical morphometry measures. Bars represent the number of loci that replicate in ABCD at a nominal significance level ($p < 0.05$). Error bars are standard deviations across 10 training sets of UK Biobank. Numbers in parentheses represent replication rate (# of discovered loci / # replicated loci). ABCD: Adolescent Brain Cognitive Development Study.

Next, we tested the generalization performance of loci discovered in each training fold of UK Biobank to a developmental cohort of adolescents from the Adolescent Brain Cognitive Development study. Here we once again see a higher absolute number of replicated loci (nominal $p < 0.05$ level), as well as a comparable replication rate for MOSTest-PVS (area: 72%, thickness: 70%, sulcal depth: 73%) to min-P (area: 60%, thickness: 63%, sulcal depth: 72%) - see Fig. 4. Again, the variance explained by the lead SNP of the top locus (averaged across cross-validation folds) accounted for more variance for MOSTest-PVS (R^2 = area: 0.019 (0.016–0.022), thickness: 0.042 (0.025–0.059), sulcal depth: 0.033 (0.030–0.036)) than for min-P (R^2 = area: 0.0083 (0.0073–0.0093), thickness: 0.0066 (0.0046–0.0086), sulcal depth: 0.013 (NA-NA*)). * Confidence interval was not estimable due to zero variance across folds.

2.3. Real data - full sample training and testing

Finally, we performed MOSTest discovery on the full UK Biobank sample and validated these in the ABCD sample by i) MOSTest Re-discovery (performing MOSTest discovery again in ABCD) and ii) MOSTest-PVS Replication (PolyVertex Score replication as presented in cross-validation results above). For this analysis in the full UK Biobank sample we discovered 543 loci for area, 569 loci for thickness and 884 loci for sulcal depth. We found in ABCD, MOSTest-PVS replication yielded a larger number of replicated loci across measures (area: 321, thickness: 298 and sulcal depth: 492) than MOSTest Re-discovery (area: 144, thickness: 137 and sulcal depth: 243). Supplementary tables present the results of this analysis for all discovered loci. For area and sulcal depth we found rs59203590 displayed maximum discovery (both: $P_{UKB, MOSTest} < 10^{-300}$) and PVS replication (area: $P_{ABCD, PVS} = 2.42 \times 10^{-13}$, sulcal depth: $P_{ABCD, PVS} = 6.19 \times 10^{-22}$). For cortical thickness the lead variant was rs8032326 ($P_{UKB, MOSTest} < 10^{-300}$ and $P_{ABCD, PVS} = 1.27 \times 10^{-19}$). Both rs59203590 and rs8032326 are on chromosome 15, proximal and in large linkage disequilibrium with each other and as well as with rs1080066 (rs59203590: $r^2=0.18$ and rs8032326: $r^2=0.31$) - a variant previously reported to be strongly linked to cortical morphology (van der Meer et al., 2020; Grasby, 2020). Additionally, we find rs13107325, a variant in *SLC39A8* which codes for ZIP8, a transmembrane protein responsible for influx of zinc, magnesium, iron, and cadmium (Wang et al., 2011). This variant which has previously been shown to display strong diverse associa-

tions across cortical morphology, BMI, alcohol consumption and cholesterol (Elliott et al., 2018; Richardson et al., 2020; Pulit et al., 2019; Zhou et al., 2020), was associated with cortical thickness in UK Biobank ($P_{UKB, MOSTest} = 1.45 \times 10^{-155}$) and replicated in ABCD ($P_{ABCD, PVS} = 2.51 \times 10^{-6}$). Interestingly, this variant did not display an association with cortical area in ABCD using PVS replication ($P_{ABCD, PVS} = 0.76$), but did across all discovery and validation folds of UK Biobank (median across folds: $P_{UKB, MOSTest} = 1.07 \times 10^{-17}$, $P_{UKB, PVS} = 4.61 \times 10^{-9}$) - possibly indicating an age dependent effect of this SNP on cortical area.

3. Discussion

In simulations we demonstrated the increased power of MOSTest vs min-P to replicate through the deployment of a PVS method, whilst exhibiting appropriate control of type-I errors. On real imaging data from UK Biobank we show a higher replication yield and comparable replication rate of MOSTest-PVS compared to a conventional min-P approach. The comparable replication rate for MOSTest-PVS loci (96–97% vs 91–93% for min-P) indicates that the difference in absolute number of replicated loci for MOSTest-PVS vs min-P is not merely a result of MOSTest discovering a higher number of loci. Furthermore, we still see a comparable replication rate when we penalize the replication significance threshold by the number of loci discovered by each method (i.e. $p < 0.05 / \#$ of discovered loci). This underscores the distributed effects of the genome across the cortex, which multivariate methods are better powered to capture and in turn, will display stronger generalization to independent data. Of note although other methods exist for performing GWAS discovery across multiple phenotypes (Galesloot et al., 2014; Zhang et al., 2014), we found none of these were able to feasibly run on the large dimensionality of phenotypes in this analysis - see supplementary note.

Additionally, we have shown that genetic-cortical morphology associations learned within an adult population (mean age 64 years) of individuals from the UK generalize out of sample to adolescents aged 9–10 years old in the United States of America taken from the ABCD study. There are marked differences between the training sample of UK Biobank and validation sample of ABCD including: large age differences, different scanners used, imaging protocols and the number of individuals in validation sets. In spite of these differences we observe a high replication rate in ABCD of discoveries found within UK Biobank via MOSTest-PVS. We see higher replication for cortical area and sulcal depth in ABCD than for cortical thickness. Cortical thickness changes more dynamically over the lifespan (Walhovd et al., 2017), therefore, given the large age disparity between the two samples, perhaps it is not a surprise to see that cortical thickness is the measure that exhibits the largest reduction in replication rates in ABCD when compared across cross-validation folds of UK Biobank for MOSTest-PVS (70% vs 97%). We may expect that the replication rate of discovered cortical thickness loci to increase as the children develop, a hypothesis that can be tested as more longitudinal ABCD data is collected. Despite differences across these datasets we observe greater replication of UK Biobank discovered loci in ABCD when taking into account the multivariate nature of associations across the cortex (i.e. MOSTest and PVS).

Furthermore, we demonstrated that lead MOSTest discoveries explained a notable amount of variance out of sample, by GWAS standards: 3–6% in UK Biobank and 2–4% in ABCD. Methods, such as MOSTest and PVS, that result in high replication yield and out of sample variance explained may support precision medicine efforts (Denny and Collins, 2021). In particular if these methods are deployed on disorders of the brain they may provide complimentary predictive power to well established models such as Polygenic Risk Scores. Here we have shown the utility of PVS in quantifying the association between cortical morphology and genotype, however this method has been used for quantifying brain-behavior relationships (Zhao et al., 2021) and could be extended to clinical neuroimaging problems unrelated to genetics e.g. predicting language recovery from MRI scans following stroke.

The training data used here to detect loci and train PVS projections weights were taken from individuals of European ancestry from the UK Biobank. We may expect that the genetic architecture of cortical morphology to differ between ancestry groups (Fan et al., 2015). Additionally, although our simulations were built using a variety of genetic architectures they were not sparse across dimensions of generated phenotypes—we may expect min-P and MOSTest-PVS performance to be more comparable for sparse signals. We also acknowledge that our use of PVS to predict genotypes out of sample is just one possible projection weighting scheme, which may not provide optimal out of sample prediction. Here we have demonstrated the high generalization performance of cortical morphology discoveries using MOSTest-PVS to independent data. This was shown both within study (UK Biobank) and across studies (UK Biobank to ABCD) despite substantial age differences of participants. This work underscores the importance of deploying well powered multivariate methods when performing GWAS on high dimensional phenotypes, both for discovery and replication.

4. Methods

The UK Biobank sample used and methods described for min-P and MOSTest discovery overlap with previous work (van der Meer et al., 2021; Shadrin et al., 2021).

4.1. UK biobank sample

Genotypes, MRI scans, demographic and clinical data were obtained from the UK Biobank under accession number 27,412, excluding participants who withdrew their consent. For this study we selected white British individuals (derived from both self-declared ethnicity and principal component analysis) who were unrelated (no relation greater than 3rd degree relatives using PLINK (Chang et al., 2015) –king-cutoff 0.0625) and who had undergone the neuroimaging protocol. This was done to obtain a homogenous group of unrelated individuals of a single ancestry as is considered best practice to control for confounds in genetic analyses (Marees et al., 2018). The resulting sample contained 34,973 individuals with a mean age of 64.4 years (standard deviation 7.5 years), 18,031 female. T₁-weighted MRI scans were collected from three scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receive head coils. We used UK Biobank v3 imputed genotype data (Bycroft et al., 2018).

4.2. Adolescent brain cognitive development® (ABCD) sample

The ABCD study is a longitudinal study across 21 data acquisition site following 11,878 children starting at 9 and 10 years old. This paper analyzed the full baseline sample from data release 3.0 (NDA DOI:10.151.54/1,519,007). The ABCD study used school-based recruitment strategies to create a population-based, demographically diverse sample with heterogeneous ancestry. T₁-weighted MRI scans were collected using Siemens Prisma, GE 750 and Phillips 3T scanners. Scanning protocols were harmonized across 21 acquisition sites. Genetic ancestry factors were estimated using fastStructure (Raj et al., 2014) with four ancestry groups. Genotype data was imputed at the Michigan Imputation Server (Das et al., 2016), using the HRC reference panel as described in (Loughnan et al., 2021; Loughnan et al., 2022). We selected individuals who had passed neuroimaging and genetic quality control checks. Additionally, we restricted to individuals who were unrelated (no relation greater than 3rd degree relatives) and of European ancestry (>90% estimated European from fastStructure) as training and testing in the same ancestry groups is considered best practice for genetic analyses (Duncan et al., 2019). This resulted in 4,069 individuals with a mean age of 9.9 years (standard deviation 0.62 years), 1889 female.

4.3. Data processing

T1-weighted structural MRI scans were processed with the FreeSurfer v5.3 standard “recon-all” processing pipeline (Dale et al., 2004) to generate 1284 non-smoothed vertex-wise measures (ico3 downsampling with the medial wall removed) summarizing cortical surface area, thickness and sulcal depth. Vertices with no variation across subjects were removed. All measures were pre-residualized for age, sex, scanner site, Euler number and the first ten genetic principal components. In contrast to other MOSTest work (van der Meer et al., 2020; Shadrin et al., 2021) we did not pre-residualize for global measures specific to each set of variables (total cortical surface area or mean cortical thickness) as there is no clear analogous global measure for sulcal depth. Subsequently, a rank-based inverse normal transformation was applied to the residualized measures. For genomic data we carried out standard quality-checks as described previously (van der Meer et al., 2020), setting a minor allele frequency threshold of 0.5% and finding the intersecting variants between UK Biobank and ABCD, leaving 8230,886 variants. Variants were tested for association with cortical surface area, cortical thickness and sulcal depth at each vertex using the standard univariate GWAS procedure. Resulting univariate p-values and effect sizes were further combined in the MOSTest and min-P analyses to identify area, thickness and sulcal depth associated loci.

4.4. Cross validation

We performed 10 times cross validation within UK Biobank with random $\frac{2}{3}$ training and $\frac{1}{3}$ testing splits. Validation in ABCD was performed across each of these $\frac{2}{3}$ UKB Training samples.

4.5. Simulations

Within the UK Biobank sample we generated simulated phenotypes which exhibited effects distributed across a large number of phenotypes following a similar process to previously published work (van der Meer et al., 2020). For this we included the full sample of $N = 34,973$ individuals, across $M = 8230,886$ variants. Using an additive genetic model, we drew genetic effects β from a gaussian distribution, and then calculated quantitative phenotypes y_i of the i th sample as $y_i = \sum_{j=1}^M x_{ij}\beta_j + \epsilon_i$, where $\mathbf{X} = (x_{ij})$ is an N by M genotype matrix, containing the number of reference alleles for the i th sample and j th variant. β_j is the causal effect size, and ϵ is a normally distributed residual that yields a pre-defined level of heritability, $h^2 = \text{Var}(\mathbf{X}\beta)/\text{Var}(y)$, for our simulations $h^2 = 0.004, 0.04$ or a 50%/50% mixture of variants with $h^2 = 0.004/0.04$. To simulate a realistic multivariate scenario of genetically and phenotypically correlated phenotypes observed for cortical morphology, we introduced a covariance structure in both β and ϵ across a total of $T = 1175$ phenotypes set to the correlation structure across vertex wise measures of sulcal depth in UK Biobank. Each set of $T = 1175$ simulated phenotypes had a pre-defined number of causal variants, nc , where $nc=10, 100, 10,000$ and $100,000$. These causal variants were restricted to 102,484 markers of chromosome 21 to facilitate the calculation of type-I error rate, as null markers (i.e. those not on chromosome 21) would not be in LD (linkage disequilibrium) with causal markers. Type-I errors were calculated as the number of null variants discovered (genome-wide significance level $p < 5 \times 10^{-8}$) and replicated ($p < 0.05$) across training and testing folds, by min-P and MOSTest-PVS methods respectively – described below (without pruning described in “Locus Definition” section). Replication power was analogously calculated as the proportion of causal variants discovered and replicated by each method across training and testing folds. In total we generated 12 sets of simulated phenotypes (each of dimensionality $T = 1175$) under different genetic architectures at 3 h^2 values and 4 nc values. Assuming the same distribution of genetic effects across other chromosomes the simulated phenotype of $nc=100$ and $h^2=0.004$, represents a realistic magnitude of the genetic effects, with genome-wide

$h^2 = 28.3\%$ exhibiting the empirical phenotype correlation structure of sulcal depth. These simulated phenotypes were then each discovered and replicated using min-P and MOSTest-PVS for each of the 10 training and testing folds. Code for generating these simulations can be found at: <https://github.com/precimed/mostest/tree/master/simu>.

4.6. Full sample training and testing

To quantify full replicability of MOSTest discoveries in UK Biobank to ABCD, we additionally performed MOSTest discovery in the full UK Biobank (34,973 individuals). We then evaluated the replication of these discoveries in the ABCD dataset (4,069 individuals) using a) MOSTest re-discovery and b) MOSTest-PVS replication. Here “MOSTest re-discovery” refers to performing MOSTest discovery separately in ABCD to see which hits replicate and “MOSTest-PVS” refers to using the vertex weights learned from the full UK Biobank dataset to generate PolyVertex Scores in ABCD to test replication of discoveries – as described below. MOSTest discoveries were using genome-wide significance $p < 5 \times 10^{-8}$ (with locus definition as described below) and replicated hits were defined with a threshold of $p < 0.05$.

4.7. MOSTest discovery

Consider M variants and P (pre-residualized) phenotypes. Let $z_{j,k}$ be a z-score from the univariate association test between j th variant and k th (residualized) phenotype, and \mathbf{z}_j be the vector of z-scores of the j th variant across phenotypes. Let \mathbf{Y} be a matrix of (pre-residualized) phenotypes with N (individuals) rows and P (phenotypes) columns, and \mathbf{R} be its correlation matrix. \mathbf{R} can be decomposed using singular valued decomposition as $\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{U}^T$ (\mathbf{U} – unitary matrix, \mathbf{S} – diagonal matrix with singular values on its diagonal). Consider the regularized version of the correlation matrix $\mathbf{R}_r = \mathbf{U}\mathbf{S}_r\mathbf{U}^T$, where \mathbf{S}_r is obtained from \mathbf{S} by keeping r largest singular values and replacing the remaining with r^{th} largest. The MOSTest statistics for j th variant (scalar) is then estimated as $\chi_j = \mathbf{z}_j \mathbf{R}_r^{-1} \mathbf{z}_j^T$, where regularization parameter is selected separately for cortical area, thickness and sulcal depth to maximize the yield of genome-wide significant loci. As established in previous work (van der Meer et al., 2020, 2021; Shadrin et al., 2021) the largest yield for cortical surface area is obtained with $r = 10$; the optimal choice for cortical thickness and sulcal depth was $r = 20$. For simulations we set $r = 20$ as simulated phenotypes were based on correlation structure of sulcal depth. As performed in previous MOSTest work (van der Meer et al., 2021; Shadrin et al., 2021) the distribution of the test statistics under null (CD_{null}^{most}) is approximated from the observed distribution of the test statistics with permuted genotypes, using the empirical distribution in the 99.99 percentile and Gamma distribution in the upper tail, where shape and scale parameters of Gamma distribution are fit to the observed data. The p-value of the MOSTest test statistic for the j th variant is then obtained as $p_{most} = CD_{null}^{most}(\chi_j)$. To clarify our notation, the gamma(a, b) distribution is parametrized by shape (a) and scale (b), so that $p_{gamma}(x|a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}$, where $\Gamma(\cdot)$ is the gamma function (Evans et al., 1993).

4.8. min-P discovery

Similar to the MOSTest analysis, consider M variants P and pre-residualized phenotypes. Let $z_{i,j}$ be a z-score from the univariate association test between j th variant and k th (residualized) phenotype and \mathbf{z}_j be the vector of z-scores of the j th variant across phenotypes. The min-P statistics for the j th variant is then estimated as $v_j = 2\Phi(-\max_{k=1\dots P}(|z_{j,k}|))$, where Φ is a cumulative distribution function of the standard normal distribution. The distribution of the min-P test statistics under null (CD_{null}^{minP}) is approximated from the observed distribution of the test statistics with permuted genotypes, using the empirical distribution in the 99.99th percentile and Beta distribution in the upper tail, where

shape parameters of Beta distribution (α and β) are fit to the observed data (Evans et al., 1993). The p-value of the min-P test statistic for the j th variant is then obtained as $p_{minP} = CD_{null}^{minP}(v_j)$.

4.9. Comparison with other multivariate GWAS methods

Although other methods exist for performing multivariate GWAS (Galesloot et al., 2014; Zhang et al., 2014), we found these would not feasibly run on the high dimensional data in this analysis – see supplementary note for details.

4.10. Locus definition

Independent significant SNPs and genomic loci were identified in accordance with the PGC locus definition, as also used in FUMA SNP2GENE (Watanabe et al., 2017). First, we select a subset of SNPs that pass genome-wide significance threshold 5×10^{-8} , and use PLINK (Chang et al., 2015) to perform a clumping procedure at LD $r^2=0.6$, to identify the list of independent significant SNPs. Second, we clump the list of independent significant SNPs at LD $r^2=0.1$ threshold to identify lead SNPs. Third, we query the reference panel for all candidate SNPs in LD r^2 of 0.1 or higher with any lead SNPs. Further, for each lead SNP, it's corresponding genomic loci is defined as a contiguous region of the lead SNPs' chromosome, containing all candidate SNPs in $r^2=0.1$ or higher LD with the lead SNP. Finally, adjacent genomic loci are merged if they are separated by less than 250 KB. Allele LD correlations are computed from EUR population of the 1000 genomes Phase 3 data. Obtained clumps of variants were considered as independent genome-wide significant genetic loci. This process was only performed for min-P and MOSTest discovery of loci for real data, and not on simulated data.

4.11. Replication of discovered variants

A schematic displaying the difference between min-P and MOSTest-PVS replication is displayed in Fig. 1. For genome-wide significant loci defined in the training folds, we performed replication in test folds of UK Biobank, as well as the whole sample of ABCD. Let \mathbf{X}^{test} represent the genotype matrix of individuals in the test set of N individuals and M variants and \mathbf{Y}^{test} represent the phenotype matrix of N individuals and P (pre-residualized) phenotypes. Replication was performed in one of two ways, depending on whether the genetic variant was discovered using min-P or MOSTest. Firstly, for a min-P discovery, implicated by the association statistic $z_{j,k}$, the j th variant, \mathbf{x}_j^{test} , is associated with the k th (residualized) phenotype \mathbf{y}_k^{test} , in the test set. Secondly, for MOSTest-PVS validation the j th discovered loci corresponds to a vector of mass univariate association statistics across all vertices \mathbf{z}_j - these are used to generate projection weights to create a PolyVertex Score (PVS) (Zhao et al., 2021), $\mathbf{x}_{PVS,j}^{test}$. This approach largely mirrors the use of polygenic scores used in genetics, where here we are aggregating effects of vertices across the cortex. For polygenic scores, it is well known that the correlation structure (i.e. linkage disequilibrium) across the genome can result in suboptimal out of sample performance. This has motivated techniques like LD-Pred (Vilhjalmsson et al., 2015) and PRSice (Choi et al., 2020) to first account for this genomic correlation before generating scores. Similarly, we decorrelate the association statistics, \mathbf{z}_j , as $\mathbf{w}_j = \tilde{\mathbf{R}}_j \mathbf{z}_j$ using the regularized correlation matrix $\tilde{\mathbf{R}}$, that was learned in the training fold. We then generate the polyvertex score for the j th genomic variant as the dot product of \mathbf{w}_j with the (pre-residualized) phenotype matrix, \mathbf{Y}^{test} , in the test set: $\hat{\mathbf{x}}_{PVS,j}^{test} = \mathbf{w}_j \mathbf{Y}^{test}$. This PVS, $\hat{\mathbf{x}}_{PVS,j}^{test}$, was then associated with its corresponding genetic variant, x_j , in the test set.

As the phenotype matrix, \mathbf{Y}^{test} , was pre-residualized for covariates before taking the most significant vertex (min-P) or computing the PVS (MOSTest-PVS) we did not need to further control for covariates. For both min-P and MOSTest-PVS validation, we calculated one-tailed p values from computed replication t statistics as we assume the effect to be

in the same direction for training folds and test sets. To define replicated loci we use a nominal p value threshold of 0.05 for associations. Due to the higher number of discovered loci for MOSTest vs min-P, we additionally report the number of loci validated at a Bonferroni corrected threshold, where this number of independent tests is taken to be the number of discovered loci in the training set. This corrected threshold penalizes MOSTest-PVS to a greater extent than min-P for discovering a larger number of loci. We calculate the variance explained by the single lead j th variant in the replication sample from t statistics of x_j and degrees of freedom (df) as: $R^2 = \frac{t^2}{(t^2 + df)}$. We report the mean of R^2 across cross validation folds, as well as an estimate of its 95% confidence interval as $1.96 \times \sigma$, where σ is the standard deviation of R^2 across cross validation folds. Code for discovery and replication is available at (https://github.com/robloughnan/MOSTest_generalization).

Funding

This work was supported by Kavli Innovative Research Grant under award number 2019-1624, and grant R01MH122688 and RF1MH120025 funded by the National Institute for Mental Health (NIMH).

Data and code availability

The data incorporated in the primary analysis were gathered from the public UK Biobank resource and will be made publicly available together with the code used to generate the data through the UK Biobank Returns Catalogue (<https://biobank.ndph.ox.ac.uk/showcase/docs.cgi?id=1>). ABCD study data release 3.0 is available for approved researchers in NIMH Data Archive (NDA DOI:10.151.54/1,519,007).

Code for conducting discovery and replication is available at https://github.com/robloughnan/MOSTest_generalization. Code for simulations is available at <https://github.com/precimed/mostest/tree/master/simu>.

Declaration of Competing Interest

Dr. Andreassen has received speaker's honorarium from Lundbeck, and is a consultant to HealthLytix. Dr. Dale is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. and receives funding through research agreements with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies. The other authors declare no competing interests.

Credit authorship contribution statement

Robert J. Loughnan: Conceptualization, Methodology, Validation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Alexey A. Shadrin:** Software, Methodology, Data curation, Writing – review & editing. **Oleksandr Frei:** Software, Methodology, Data curation, Writing – review & editing. **Dennis van der Meer:** Software, Methodology, Data curation, Writing – review & editing. **WeiQi Zhao:** Conceptualization. **Clare E. Palmer:** Visualization. **Wesley K. Thompson:** Writing – review & editing. **Carolina Makowski:** Writing – review & editing. **Terry L. Jernigan:** Supervision, Writing – review & editing. **Ole A. Andreassen:** Conceptualization, Writing – review & editing. **Chun Chieh Fan:** Supervision, Funding acquisition, Methodology, Conceptualization, Writing – review & editing. **Anders M. Dale:** Conceptualization, Supervision, Methodology, Software, Writing – review & editing.

ABCD Acknowledgement

Data used in the preparation of this article were obtained from the **Adolescent Brain Cognitive DevelopmentSM Study (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, and U24DA041147

A full list of supporters is available at <https://abcdstudy.org/federal-partners/>. A listing of participating sites and a complete listing of the study investigators can be found at <https://abcdstudy.org/principal-investigators.html>. ABCD Study consortium investigators designed and implemented the study and/or provided data but did not 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 Study consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this came from [NIMH Data Archive Digital Object Identifier (10.151.54/1519007)].

UK Biobank Acknowledgment

UK Biobank's research resource is a major contributor in the advancement of modern medicine and treatment, enabling better understanding of the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases and stroke.

UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. The organization has over 150 dedicated members of staff, based in multiple locations across the UK. UK Biobank data was accessed under accession number 27412.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2022.119632](https://doi.org/10.1016/j.neuroimage.2022.119632).

References

- Panizzon, M.S., et al., 2009. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19, 2728–2735.
- Eyer, L.T., et al., 2012. A comparison of heritability maps of cortical surface area and thickness and the influence of adjustment for whole brain measures: a magnetic resonance imaging twin study. *Twin Res. Hum. Genet.* 15, 304–314.
- van der Meer, D., et al., 2020. Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* 11, 3512.
- van der Meer, D. et al. The genetic architecture of human cortical folding. *bioRxiv* 2021.01.13.426555 (2021). doi:10.1101/2021.01.13.426555
- Shadrin, A.A. et al. Multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *bioRxiv* 2020.10.22.350298 (2021). doi:10.1101/2020.10.22.350298
- Zhao, W. et al. The Bayesian polyvertex score (PVS-B): a whole-brain phenotypic prediction framework for neuroimaging studies. *bioRxiv* 813915 (2019). doi:10.1101/813915
- Zhao, W., et al., 2021. Individual differences in cognitive performance are better predicted by global rather than localized BOLD activity patterns across the cortex. *Cereb. Cortex* 31, 1478–1488.
- Grasby, K.L., et al., 2020. The genetic architecture of the human cerebral cortex. *Science* 367.
- Lewis, C.M., Vassos, E., 2020. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12, 44.
- Wang, B., He, L., Dong, H., Dalton, T.P., Nebert, D.W., 2011. Generation of a Slc39a8 hypomorph mouse: markedly decreased ZIP8 Zn²⁺/(HCO₃⁻)₂ transporter expression. *Biochem. Biophys. Res. Commun.* 410, 289–294.

- Elliott, L.T., et al., 2018. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210–216.
- Richardson, T.G., et al., 2020. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med.* 17, 1–22.
- Pulit, S.L., et al., 2019. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* 28, 166–174.
- Zhou, H., et al., 2020. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* 23, 809–818.
- Galesloot, T.E., Van Steen, K., Kiemeny, L.A.L.M., Janss, L.L., Vermeulen, S.H., 2014. A comparison of multivariate genome-wide association methods. *PLoS ONE* 9, 1–8.
- Zhang, Y., Xu, Z., Shen, X., Pan, W., 2014. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *Neuroimage* 96, 309–325.
- Walhovd, K.B., Fjell, A.M., Giedd, J., Dale, A.M., Brown, T.T., 2017. Through thick and thin: a need to reconcile contradictory results on trajectories in human cortical development. *Cereb. Cortex* 27, 1472–1481.
- Denny, J.C., Collins, F.S., 2021. Precision medicine in 2030—Seven ways to transform healthcare. *Cell* 184, 1415–1419.
- Fan, C.C., et al., 2015. Modeling the 3D geometry of the cortical surface with genetic ancestry. *Curr. Biol.* 25, 1988–1992.
- Chang, C.C., et al., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- Loughnan, R.J., Palmer, C.E., Makowski, C., Thompson, W.K., Barch, D.M., Jernigan, T.L., Dale, A.M., Fan, C.C., 2022. Unique prediction of developmental psychopathology from genetic and familial risk. *J Child Psychol Psychiatr.* <https://doi.org/10.1111/jcpp.13649>.
- Marees, A.T., et al., 2018. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27, 1–10.
- Bycroft, C., et al., 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.
- Raj, A., Stephens, M., Pritchard, J.K., 2014. FastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589.
- Das, S., et al., 2016. Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284.
- Loughnan, R.J. et al. Gene-experience correlation during cognitive development: evidence from the Adolescent Brain Cognitive Development (ABCD) study. *bioRxiv* 637512 (2021). doi:10.1101/637512
- Duncan, L., et al., 2019. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10, 3328. doi:10.1038/s41467-019-11112-0.
- Dale, A.M., et al., 2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.
- Evans, M., Hastings, N., Peacock, J.B., 1993. *Statistical Distributions*. J. Wiley.
- Watanabe, K., Taskesen, E., van Bochoven, A., Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826.
- Vilhjálmsón, B.J., et al., 2015. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97, 576–592.
- Choi, S.W., Mak, T.S.H., O'Reilly, P.F., 2020. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15, 2759–2772.