Cerebral Cortex, June 2020;30: 3439–3450

doi: 10.1093/cercor/bhz319
Advance Access Publication Date: 9 February 2020
Original Article

ORIGINAL ARTICLE

Minimal Relationship between Local Gyrification and General Cognitive Ability in Humans

Samuel R. Mathias[1,2], Emma E. M. Knowles[1,2], Josephine Mollon[1,2], Amanda Rodrigue[1,2], Marinka M. C. Koenis[3], Aaron F. Alexander-Bloch[4], Anderson M. Winkler[5], Rene L. Olvera[6], Ravi Duggirala[7], Harald H. H. Göring[7], Joanne E. Curran[7], Peter T. Fox[8,9], Laura Almasy[10], John Blangero[8] and David C. Glahn[1,2,3]

[1]Department of Psychiatry, Boston Children's Hospital, Boston, MA 02115, USA, [2]Harvard Medical School, Boston, MA 02115, USA, [3]Olin Neuropsychiatric Research Center, Hartford Hospital, Harford, CT 06106, USA, [4]Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, and Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA, [5]Section on Development and Affective Neuroscience, National Institute of Mental Health, Bethesda, MD 20892, USA, [6]Department of Psychiatry, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA, [7]Department of Human Genetics and South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX 78520, USA, [8]Research Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA, [9]South Texas Veterans Health System, San Antonio, TX 78229, USA and [10]Department of Genetics, University of Pennsylvania and Department of Biomedical and Health Informatics at Children's Hospital of Philadelphia, Philadelphia, PA19104, USA

Address correspondence to Samuel R. Mathias, Harvard Medical School, Boston, MA 02215, USA. Email: samuel.mathias@childrens.harvard.edu.

Abstract

Previous studies suggest that gyrification is associated with superior cognitive abilities in humans, but the strength of this relationship remains unclear. Here, in two samples of related individuals (total $N = 2882$), we calculated an index of local gyrification (LGI) at thousands of cortical surface points using structural brain images and an index of general cognitive ability (*g*) using performance on cognitive tests. Replicating previous studies, we found that phenotypic and genetic LGI–*g* correlations were positive and statistically significant in many cortical regions. However, all LGI–*g* correlations in both samples were extremely weak, regardless of whether they were significant or nonsignificant. For example, the median phenotypic LGI–*g* correlation was 0.05 in one sample and 0.10 in the other. These correlations were even weaker after adjusting for confounding neuroanatomical variables (intracranial volume and local cortical surface area). Furthermore, when all LGIs were considered together, at least 89% of the phenotypic variance of *g* remained unaccounted for. We conclude that the association between LGI and *g* is too weak to have profound implications for our understanding of the neurobiology of intelligence. This study highlights potential issues when focusing heavily on statistical significance rather than effect sizes in large-scale observational neuroimaging studies.

Key words: cognition, general intelligence, gyrification

Introduction

Gyrification refers to the folding of the cerebral cortex. A popular method of quantifying gyrification is the "gyrification index" (GI), or the ratio of the total and superficially exposed cortical surface areas (Schaer et al. 2008; Zilles et al. 1988). Ontogenetically, gyrification is likely caused by processes such as neurogenesis and dendritic growth (Rakic 2009; Van Essen 1997; Welker 1990; Zilles et al. 2013). The phylogeny of gyrification is less clear. Human brains are highly gyrenphalic, suggesting that there is an association between gyrification and comparative intelligence across species. However, it could argue that this association is somewhat weak because other large and less intelligent mammals, such as elephants and whales, have higher GIs than humans (Zilles et al. 2013).

Based on the putative between-species correlation between gyrification and intelligence, several authors have explored the possibility that gyrification and cognitive abilities are correlated within samples of healthy humans (Chung et al. 2017; Docherty et al. 2015; Gautam et al. 2015; Green et al. 2018; Gregory et al. 2016). To date, the strongest evidence for this hypothesis comes from a study involving 1102 participants from two independent samples (Gregory et al. 2016). This study reported positive and statistically significant associations between an index of local gyrification (LGI; Schaer et al. 2008), calculated at thousands of points on the cortex, and an index of general cognitive ability (*g*; Spearman 1904), calculated from performance on cognitive tests. The authors considered LGI, rather than whole-brain GI, to examine potential regional specificity; indeed, LGI–*g* associations were more often significant in brain regions implicated by the parieto-frontal integration theory of intelligence (P-FIT; Jung and Haier 2007) than in the rest of the brain (cf., Green et al. 2018). Several other studies also provided evidence for an association between LGI and cognitive abilities (Chung et al. 2017; Gautam et al. 2015; Green et al. 2018), although their observed associations only reached significance within small subsets of brain regions or when certain cognitive measures were considered.

Previous studies certainly provide compelling evidence for the existence of an association between gyrification and cognitive abilities. However, it could be argued that because these studies sparingly reported correlation coefficients or other transparent measures of effect size, the strength of this association remains unclear. Is the association stronger or weaker than previously observed relationships between cognitive abilities and other anatomical traits, such as height (Keller et al. 2013), brain volume (Pietschnig et al. 2015), and cortical surface area (Vuoksimaa et al. 2015)? If stronger, gyrification per se may play a specific and important role in the neurobiology of human intelligence. Is the association strong enough such that cognitive abilities may be predicted from gyrification, or vice versa? If so, gyrification would hold considerable promise for future studies, especially those that consider the clinical utility of gyrification as a potential endophenotype of psychiatric disorders (Glahn et al. 2014).

What might explain an association between gyrification and cognitive abilities? Shared genetic factors (pleiotropy) is one intriguing possibility. However, while a previous study estimated genetic correlations between whole-brain GI and an index of general cognitive ability, it did not conclude that this correlation was statistically significant (Docherty et al. 2015). One could argue that this previous study may have missed potential regional specificity because it used a singular index of whole-brain GI, which reflects aggregate gyrification over the entire cortex, rather than many region-specific LGI values (cf., Gregory et al. 2016). Another possibility is that because neither GI nor LGI are independent of other neuroanatomical traits (Gautam et al. 2015; Hogstrom et al. 2013), the LGI–*g* association is an indirect consequence of another brain-behavior relationship. Previous studies considered this possibility by covarying for a single potential confounding variable, such as intracranial volume (ICV; Chung et al. 2017; Gautam et al. 2015; Green et al. 2018; Gregory et al. 2016) or overall cortical surface area (Docherty et al. 2015). However, it could be argued that these studies may have obtained different results if they had each considered a range of possible confounders.

Here, we calculated LGIs and other neuroanatomical traits using high-resolution structural magnetic resonance imaging (MRI) scans in two large samples of related individuals: 1) the Genetics of Brain Structure and Function Study (GOBS; McKay et al. 2014), a cohort of multigenerational extended pedigrees ($N = 1769$); and 2) the Human Connectome Project (HCP; Van Essen et al. 2013), a twin study ($N = 1113$). We also calculated *g* in the same participants and estimated phenotypic LGI–*g* correlations. However, in contrast to previous studies where all participants were unrelated (e.g., Chung et al. 2017; Gautam et al. 2015; Green et al. 2018; Gregory et al. 2016), we exploited the degrees of relatedness between individuals to estimate narrow-sense heritabilites of the traits, followed by genetic and environmental LGI–*g* correlations. In contrast to a previous twin study (Docherty et al. 2015), we considered LGI, rather than a single value of whole-brain GI per individual, so that our analyses were sensitive to potential regional specificity. Because this is the largest study on this topic to date (total $N = 2882$), we were best powered to detect statistically significant correlations. Crucially, however, we also provide summaries of all raw correlation coefficients, significant and nonsignificant, to make the strengths of these correlations abundantly clear. To consider the possible confounding roles of other traits, we estimated partial phenotypic, genetic, and environmental LGI–*g* correlations under models including height, ICV, cortical surface area, and cortical thickness as potential confounders. We also considered the extent to which LGIs considered together explained the phenotypic variance of *g* via ridge regression. Finally, we examined the consistency of regional specificity of LGI–*g* correlations across samples and compared these spatial patterns to the P-FIT (Jung and Haier 2007).

Materials and Methods

Participants

GOBS is a cohort of randomly ascertained, multigenerational extended pedigrees of Mexican American ancestry from San Antonio, TX (McKay et al. 2014). MRI images, age, sex, and pedigree information were available for 1433 participants (mean age: 40.8 years; age range: 18–85 years; mean family size: 14.9; 819 females). Height was available for 1661 participants (mean age: 41.2 years; age range: 17–97 years; mean family size: 17.5; 982 females), and *g* was calculated for 1769 participants (1057 females; mean age: 42.3 years; age range: 18–97 years; mean family size: 18.6). The subset with complete data was 1216 participants (mean age: 39.6 years; age range: 18–81 years; mean family size: 13.1; 689 females). However, multivariate analyses (described later) made maximal use of unbalanced data and were not restricted to individuals with complete data.

HCP is a cohort of twins and siblings whose data are freely available for researchers (Van Essen et al. 2013). We used data from the "S1200" release, which included height, cognitive data, age, sex, and zygosity information for 1205 participants (mean age 28.8 years; age range 22–37 years; 655 females). MRI images were available for a subset of 1113 participants (mean age 28.8 years; age range 22–37 years; 605 females). Again, analyses made use of all data.

All participants provided written informed consent in accordance with institutional review boards at their respective institutions.

Neuroanatomical Traits

In GOBS, up to seven T_1-weighted images were acquired per participant on a Siemens 3 T TIM Trio scanner (for details, cf., McKay et al. 2014). Images were averaged and then processed using the FreeSurfer 5.3.0 pipeline (Dale et al. 1999; Fischl et al. 1999). Among other traits, this produced the estimates of ICV.

In HCP, T_1-weighted MRI images were acquired on a customized Siemens 3 T Skyra scanner (Van Essen et al. 2013; Glasser et al. 2013). We used "processed" versions of the images (e.g., skull-stripped, corrected for field-strength heterogeneity) produced by a modified version of the FreeSurfer pipeline. This pipeline produced estimates of ICV. To be consistent with GOBS, we also ran HCP images through standard the FreeSurfer pipeline.

FreeSurfer cortical surface models were registered to the "fsaverage5" template. Three vertexwise traits (LGI, area, and thickness) were calculated at each vertex using the "qdec" and "qcache" commands (cf., Dale et al. 1999; Fischl et al. 1999; Schaer et al. 2008) and smoothed with a Gaussian kernel of 10 mm of full width at half maximum, which was chosen because it is sufficient to approximate the areal interpolation method used by FreeSurfer to an ideal, mass-conservative (pycnophylactic) method (Winkler et al. 2018); for consistency, we also applied 10-mm smoothing to LGI and thickness. Vertices were labeled as belonging to regions of interest (ROIs) according to several atlases. Vertices labeled as "unknown" or the corpus callosum according to the Desikan–Killiany atlas (Desikan et al. 2006) were dropped, along with a small number of additional vertices which had values of 0 across all participants, leaving 18 408 vertices for analysis.

Calculation of *g*

GOBS participants completed a cognitive test battery covering a wide range of domains, such as episodic memory, fluid intelligence, processing speed, recognition memory, sustained attention, working memory, and verbal fluency. Details of these tests are provided elsewhere (e.g., Glahn et al. 2007; Glahn et al. 2010; Knowles et al. 2014). Measures used in the present study are summarized in Table 1.

HCP participants completed a different, but similarly comprehensive, cognitive test battery, described by Barch et al. (2013), also summarized in Table 1. We used the measures of raw performance (not age-adjusted).

All cognitive measures were standardized, and outliers greater or less than 3.29 were dropped (for a normally distributed variable, dropped values would lie beyond the 99.95th percentile). Participants with 50% or more missing cognitive data were dropped. To avoid entirely excluding remaining participants with incomplete cognitive data, missing values were imputed using the "mice" package (van Buuren and Groothuis-Oudshoorn 2011) in R (R Core Team 2013). The composite index of general cognitive ability, *g*, was the first principal component extracted using the FactoMineR package (Le et al. 2008).

Statistical Analysis

Univariate Models

All traits were rank-based inverse-normal transformed to ensure that they were normally distributed, and then analyzed using univariate variance components (polygenic) models in SOLAR (Almasy and Blangero 1998). One such model was fitted per trait (*g*, height, ICV, LGI at each vertex, area at each vertex, and thickness at each vertex), resulting in 110 456 univariate polygenic models (55 228 models per sample). These models exploit the expected proportion of alleles that are identical by descent among relatives to provide estimates of narrow-sense heritability, defined as the proportion of phenotypic variance that can be explained by additive genetic factors, denoted by h^2 (Visscher et al. 2008). An intercept, age, age squared, sex, and their interactions were included as fixed-effect covariates in all models. We performed a single one-tailed test per model to determine whether h^2 was significantly greater than 0. These tests involved comparing the likelihood of the polygenic model and that of a model with an identical design except that h^2 was constrained to 0. For the vertexwise traits, *P* values were corrected using a false-discovery rate (FDR) (Benjamini and Hochberg 1995) of 0.05. FDR correction was applied separately per sample and vertexwise trait.

Bivariate Models

A bivariate polygenic model is an extension of a univariate model and provides heritability estimates for two traits, as well as estimates of their genetic and environmental correlations. The genetic correlation, denoted by ρ_G, describes the correlation between the latent additive genetic factors influencing both traits. The environmental correlation, denoted by ρ_E, describes the correlation between the traits' nongenetic components. Both ρ_G and ρ_E were free parameters in these models. The phenotypic correlation, denoted by ρ_P, was not a free parameter and was estimated deterministically. We fitted 36 816 bivariate models (18 408 per sample) with LGI at one vertex as the first trait and *g* as the second trait. We fitted four additional bivariate models (two per sample) to examine the relationships between height and *g* and ICV and *g*.

For each bivariate model, we performed three two-tailed significance tests to determine whether the phenotypic, genotypic, and environmental LGI–*g* correlations differed from 0. Tests of genetic and environmental correlations involved comparing the likelihood of the full bivariate polygenic model to that of a model where the parameter of interest was constrained to 0. Since ρ_P was deterministic, testing it required the reparameterization of each bivariate model so that ρ_P was a free parameter that could be constrained it to 0. FDR correction was applied to separately per sample and type of correlation.

Trivariate Models

A trivariate polygenic model provides heritability estimates for three traits, along with genetic and environmental correlations between all pairings of traits. Phenotypic correlations were estimated deterministically. Trivariate models also provide partial genetic, environmental, and phenotypic correlations,

Table 1 Cognitive measures used to calculate *g*

Cohort	Test	Cognitive domain	Measure
HCP	Picture sequence memory	Episodic memory	Unadjusted scale score
	Dimensional change card sort	Executive function	Unadjusted scale score
	Flanker	Executive function	Unadjusted scale score
	Progressive matrices	Fluid intelligence	Number correct
	Oral reading recognition	Language	Unadjusted scale score
	Picture vocabulary	Language	Unadjusted scale score
	Pattern comparison	Processing speed	Unadjusted scale score
	Line orientation	Spatial orientation	Number correct
	Continuous performance	Sustained attention	Number of hits
	Word Memory	Episodic memory	Number correct
	List sorting	Working memory	Unadjusted scale score
	Emotion recognition	Emotion recognition	Number correct
GOBS	Verbal learning	Episodic memory	Number recalled
	Digit-symbol coding	Processing speed	Number correct
	Conditional exclusion	Executive function	Number correct
	Conditional exclusion	Executive function	Number of learned rules
	Spatial working memory	Working memory	Number correct
	Verbal learning (delayed)	Episodic memory	Number recalled
	Facial memory	Episodic memory	Number correct
	Facial memory (delayed)	Episodic memory	Number correct
	Trail making (A)	Processing speed	Time taken
	Trail making (B)	Executive function	Time taken
	Continuous performance	Sustained attention	Number of hits
	Digit span (forward)	Working memory	Span
	Digit span (backward)	Working memory	Span
	Letter–number sequencing	Working memory	Span
	Balloon analog	Risk taking	Number correct
	Oral word association (FAS)	Language	Number valid
	Oral word association (semantic)	Language	Number valid
	Emotion recognition	Emotion recognition	Number correct

which describe the correlation between two traits adjusted for the third trait. This is advantageous to a bivariate model including two traits of interest residualized for the third trait because the former approach provides a clearer picture of the genetic and environmental relationships between all three traits. Partial correlations were estimated deterministically. We fitted 147 264 trivariate models (73 632 per sample) in total. Each model contained LGI at one vertex as the first trait, *g* as the second trait, and either height, ICV, area (at the same vertex), or thickness (at the same vertex) as the third trait.

Trivariate models took a long time to converge. We did not perform significance tests of parameters from the trivariate models because it was unfeasible to fit additional reparameterized and/or constrained versions of these models given the constraints of our computational resources.

Ridge Regression

One could argue that LGI measured at a single vertex is likely to have greater associated measurement error than, for instance, whole-body or whole-brain traits, such has height or ICV. If so, it would not be surprising if individual LGI–*g* correlations were weak. By extension, it would not be surprising for LGI at a typical vertex to explain only a small portion of the variance of *g*. It would therefore make sense to quantify the variance of *g* explained by all LGIs together. However, this could not be achieved with ordinary least-squares regression because the model would be overidentified (more predictors than observations). An alternative strategy is to calculate the variance of *g* explained by whole-brain GI. However, as argued by Gregory et al. (2016) and others, this strategy may miss regional specificity. We opted for ridge regression (Hoerl and Kennard 1970), which mitigates the problem of overidentification by regularizing regression coefficients. We performed ridge regression with *g* as the outcome and all LGIs as predictors using the sklearn package (Pedregosa et al. 2011) in Python (Python Software Foundation 2018). All traits were residualized, using the covariates described previously, prior to this analysis. Since LGI is not independent of ICV, the ensemble predictive ability of LGI would likely be inflated by the well-established association between *g* and ICV; therefore, *g* was additionally residualized for ICV. The regularization factor for ridge regression was selected by fitting with different regularization factors and choosing the model with the lowest leave-one-out cross validation estimator. Variance explained was calculated in the usual way, by squaring the correlation between the observed and predicted outcomes.

Across-Sample Correlations

Because vertexwise traits were registered to the same template, it was possible to examine across-sample consistency in regional specificity. For example, suppose genetic LGI–*g* correlations exhibited a strong positive across-sample correlation; this would suggest that if LGI and *g* were strongly genetically correlated at a given brain region in GOBS, they were strongly genetically correlated at the same brain region in HCP. We computed such across-sample correlations and subjected them to

two-tailed significance tests with FDR correction applied separately per sample.

P-FIT Tests

Previous studies suggest that the spatial pattern of LGI–*g* correlations may be consistent with the P-FIT (Jung and Haier 2007). The P-FIT implicates, by Brodmann areas: the dorsolateral prefrontal cortex (6, 9, 10, 45, 46, 47); the inferior (39, 40) and superior (7) parietal lobule; the anterior cingulate (32); and regions within the temporal (21, 37) and occipital (18, 19) lobes. We followed Green et al. (2018) by considering a given vertex to be implicated by the P-FIT if it belonged to one of a subset of ROIs from the Destrieux et al. atlas (Destrieux et al. 2010) selected by Green et al. to closely match the original Brodmann areas. We tested whether parameter estimates from polygenic models related to vertexwise measures were stronger at P-FIT than non-P-FIT vertices by computing Bayes factors (BFs) using JASP (JASP Team 2018). We placed a half-Cauchy prior on effect sizes to reflect the direction of hypothesized effects (P-FIT > non-P-FIT) with "default" scale 0.707 (Rouder et al. 2009). The advantage of BFs over *P* values is that they state evidence for or against the null hypothesis. Here, the null hypothesis was that correlations were not stronger in P-FIT than non-P-FIT regions.

Results

Heritability

Figure 1A shows the estimates of narrow-sense heritability for *g*, height, and ICV. In both samples, height was the most heritable of these traits (GOBS: h^2 (Height) = 0.85; HCP: h^2 (Height) = 0.89), followed by ICV (GOBS: h^2 (ICV) = 0.79; HCP: h^2 (ICV) = 0.88), followed by *g* (GOBS: $h^2(g) = 0.69$; HCP: $h^2(g) = 0.82$). These were all significant (FDR-corrected $P < 0.001$) and consistent with previous estimates (Baare et al. 2001; Deary et al. 2006; Silventoinen et al. 2003).

Figure 1A also shows the distributions of heritability estimates for vertexwise traits. LGI was significantly heritable (FDR-corrected $P < 0.05$) at all vertices in both samples (GOBS: median h^2 (LGI) = 0.44, interquartile range = 0.37, 0.51; HCP: median h^2 (LGI) = 0.55, interquartile range = 0.47, 0.62). To our knowledge, heritability estimates of LGI have not been reported previously. Area was significantly heritable (FDR-corrected $P < 0.05$) at 18 316 vertices (99.5%) in GOBS (median h^2 (Area) = 0.36, interquartile range = 0.28, 0.44) and 18 400 vertices (99.96%) in HCP (median h^2 (Area) = 0.39, interquartile range = 0.31, 0.48). Thickness was significantly heritable at 16 924 vertices (92%) in GOBS (median h^2 (Thickness) = 0.20, interquartile range = 0.14, 0.26) and 18 375 vertices (99.8%) in HCP (median h^2 (Thickness) = 0.34, interquartile range = 0.28, 0.40). Heritability estimates of vertexwise area and thickness were consistent with those reported previously (Eyler et al. 2012).

LGI–*g* Correlations

Figure 1B shows the distributions of LGI–*g* correlation estimates from the bivariate polygenic models. In GOBS, these were mostly positive yet extremely weak (median ρ_P (LGI, *g*) = 0.05; interquartile range = 0.03, 0.07). Nevertheless, 4570 of them (25%) were significant (FDR-corrected $P < 0.05$). The corresponding correlations in HCP were somewhat stronger but also weak in absolute terms (median ρ_P (LGI, *g*) = 0.10; interquartile range = 0.06, 0.14); 13 651 of them (74%) were significant (FDR-corrected $P < 0.05$). Thus, at a typical vertex, LGI explained between 0.25% and 1% of the total phenotypic variance of *g*.

Genetic LGI–*g* correlation estimates (Fig. 1B) were mostly positive and stronger than phenotypic LGI–*g* correlations, particularly in HCP. However, again, they were all weak in absolute terms (GOBS: median ρ_G (LGI, *g*) = 0.07; interquartile range = 0.02, 0.011; HCP: median ρ_G (LGI, *g*) = 0.18; interquartile range = 0.12, 0.24). None of these correlations reached the threshold of statistical significance in GOBS. In HCP, 14 022 (76%) of these correlations were significant (FDR-corrected $P < 0.05$).

Environmental LGI–*g* correlation estimates (Fig. 1B) were all extremely weak in both samples (GOBS: median ρ_E (LGI, *g*) = 0.03; interquartile range = −0.01, 0.08; HCP: median ρ_E (LGI, *g*) = −0.06; interquartile range = −0.10, 0.02). None was significant in either cohort.

Partial LGI–*g* Correlations

Figure 1B also shows the distributions of all partial LGI–*g* correlation estimates. Partial phenotypic LGI–*g* correlations were weakest with ICV as the third trait (GOBS: median ρ_P (LGI, *g*|ICV) = 0.03, interquartile range = 0.001, 0.05; HCP: median ρ_P (LGI, *g*|ICV) = 0.06, interquartile range = 0.03, 0.09), followed by area (GOBS: median ρ_P (LGI, *g*|Area) = 0.04, interquartile range = 0.01, 0.06; HCP: median ρ_P (LGI, *g*|Area) = 0.08, interquartile range = 0.04, 0.12), followed by height (GOBS: median ρ_P (LGI, *g*|Height) = 0.05, interquartile range = 0.03, 0.07; HCP: median ρ_P (LGI, *g*|Height) = 0.10, interquartile range = 0.06, 0.14), followed by thickness (GOBS: median ρ_P (LGI, *g*|Thickness) = 0.06, interquartile range = 0.03, 0.07; HCP: median ρ_P (LGI, *g*|Thickness) = 0.10, interquartile range = 0.07, 0.14). Thus, partialing out the influence of other anatomical traits, particularly those related to brain size (ICV and area), further reduced already weak phenotypic LGI–*g* correlations.

Estimates of partial genetic LGI–*g* correlations were also weakest with ICV as the third trait (GOBS: median ρ_G (LGI, *g*|ICV) = 0.02, interquartile range = −.02, 0.07; HCP: median ρ_G (LGI, *g*|ICV) = 0.12, interquartile range = 0.07, 0.17), followed by area (GOBS: median ρ_G (LGI, *g*|Area) = 0.02, interquartile range = −.03, 0.08; HCP: median ρ_G (LGI, *g*|Area) = 0.10, interquartile range = 0.03, 0.17), followed by height (GOBS: median ρ_G (LGI, *g*|Height) = 0.07, interquartile range = 0.01, 0.12; HCP: median ρ_G (LGI, *g*|Height) = 0.18, interquartile range = 0.11, 0.24), followed by thickness (GOBS: median ρ_G (LGI, *g*|Thickness) = 0.06, interquartile range = 0.03, 0.07; HCP: median ρ_G (LGI, *g*|Thickness) = 0.18, interquartile range = 0.12, 0.25).

Other Correlations

Height–*g* and ICV–*g* correlations were estimated with additional bivariate polygenic models. In GOBS, phenotypic and environmental height–*g* correlations were significant (FDR-corrected $P < 0.001$), but not the genetic height–*g* correlation (ρ_P (Height, *g*) = 0.16; ρ_G (Height, *g*) = 0.08; ρ_E (Height, *g*) = 0.22); and only the phenotypic ICV–*g* correlation was significant (FDR-corrected $P < 0.001$) (ρ_P (ICV, *g*) = 0.12; ρ_G (ICV, *g*) = 0.14; ρ_E (ICV, *g*) = 0.11). In HCP, none of the height–*g* correlations was significant (ρ_P (Height, *g*) = 0.00; ρ_G (Height, *g*) = 0.15; ρ_E (Height, *g*) = −0.12); but the phenotypic and genetic ICV–*g* correlations were significant (FDR-corrected $P < 0.001$) (ρ_P (ICV, *g*) = 0.27; ρ_G (ICV, *g*) = 0.30; ρ_E (ICV, *g*) = 0.02). Notably, in both samples, ICV–*g* correlations were stronger than most of the LGI–*g* correlations.

We did not explicitly test associations between vertexwise area/thickness and *g* because these have been explored thoroughly in previous studies (e.g., Vuoksimaa et al. 2015). However, our trivariate models provided estimates area–*g* and thickness–*g* correlations (although they were not tested for statistical significance), as well as corresponding partial correlations adjusted for LGI. Figure 1B shows the distributions of these correlations. Consistent with previous studies, phenotypic area–*g* correlations (GOBS: median $\rho_P\,(\text{Area}, g) = 0.07$, interquartile range = 0.05, 0.09; HCP: median $\rho_P\,(\text{Area}, g) = 0.11$, interquartile range = 0.08, 0.14) and genetic area–*g* correlations (GOBS: median $\rho_G\,(\text{Area}, g) = 0.12$, interquartile range = 0.06, 0.18; HCP: median $\rho_G\,(\text{Area}, g) = 0.21$, interquartile range = 0.15, 0.27) were mostly positive. Notably, these correlations were similar in strength to those between LGI and *g*. After partialing out LGI, phenotypic and genetic area–*g* correlations were slightly weaker. Also consistent with previous studies (e.g., Vuoksimaa et al., 2015), all thickness–*g* correlations were very weak (GOBS: median $\rho_P\,(\text{Thickness}, g) = 0.03$, interquartile range = 0.01, 0.05; HCP: median $\rho_P\,(\text{Thickness}, g) = 0.00$, interquartile range = −0.03, 0.04) and genetic thickness–*g* correlations (GOBS: median $\rho_G\,(\text{thickness}, g) = 0.00$, interquartile range = −0.08, 0.08; HCP: median $\rho_G\,(\text{Thickness}, g) = 0.01$, interquartile range = −0.07, 0.07).

Ridge Regressions on *g*

In both GOBS and HCP, ridge regression estimated the phenotypic variance explained by ensemble variation in LGI. In GOBS, this analysis yielded $R^2 = 0.11$ (regularization factor = 10^5). In HCP, the same analysis yielded $R^2 = 0.05$ (regularization factor = 10^6). Thus, all LGIs considered together failed to account for at least 89% of the phenotypic variance of *g*. The regularization factors of the ridge regressions were high because LGI at one vertex tended to correlated very strongly with all other LGIs.

Neuroanatomical Specificity

Figure 2B shows the neuroanatomical patterns of heritability estimates for LGI, area, and thickness. In both samples, regions of strong heritability included the perimeter of the central sulcus, insula, anterior cingulate, and junctions of the isthmus, precuneus, and cuneus. Regions of relatively weak heritability included middle frontal, superior parietal, and inferior temporal lobes. The across-sample correlation in LGI heritability was strong (Fig. 2A; $r = 0.59$; $R^2 = 0.35$; FDR-corrected $P < 0.001$). In other words, if LGI was strongly heritable at a given vertex in GOBS, it tended to be strongly heritable at the same vertex in HCP. Thus, the neuroanatomical specificity of LGI heritability was consistent across samples. The same was true of area ($r = 0.73$; $R^2 = 0.54$; FDR-corrected $P < 0.001$) and, to a lesser extent, thickness ($r = 0.44$; $R^2 = 0.19$; FDR-corrected $P < 0.001$).

Figure 3B shows the neuroanatomical patterns of LGI–*g* correlations. There was a strong across-sample correlation in phenotypic LGI–*g* correlation (Fig. 3A; $r = 0.63$; $R^2 = 0.40$; FDR-corrected $P < 0.001$), a moderate across-sample correlation in genetic LGI–*g* correlation ($r = 0.44$; $R^2 = 0.19$; FDR-corrected $P < 0.001$), but no correlation in environmental LGI–*g* correlation ($r = 0.00$; $R^2 = 0.00$).

Comparisons to the P-FIT

Figure 4 shows the distributions of LGI heritability estimates and LGI–*g* correlations as a function of whether vertices were implicated by the P-FIT. In all cases, the median heritability or correlation was actually weaker in P-FIT than non-P-FIT vertices, contrary to the hypothesized direction. Thus, unsurprisingly, all one-sided Bayesian *t* tests provided decisive evidence for the null hypothesis, namely that parameter estimates were not higher for vertices in P-FIT regions, according to the nomenclature proposed by Jeffreys (1961) (largest BF = 0.005).

Discussion

In the present study, we analyzed data from two samples of related individuals to examine the association between gyrification and general cognitive ability. We used a popular automatic method to calculate LGI across the cortex from MRI images (Schaer et al. 2008), and calculated *g* from performance on batteries of cognitive tests. We estimated the heritability of height, ICV, and *g*, as well as the heritability LGI, area, and thickness at all vertices. We estimated phenotypic, genetic, and environmental LGI–*g* correlations, as well as partial LGI–*g* correlations with height, ICV, area (at the same vertex), and thickness (at the same vertex) as potential confounding variables. We estimated the amount of phenotypic variance of *g* explained by all LGIs together via ridge regression, and examined the across-sample consistency of neuroanatomical specificity in heritability of LGI, area, and thickness, as well as LGI–*g* correlations. Finally, we tested whether heritability estimates and LGI–*g* correlations were stronger in regions implicated by the P-FIT, a model of the neurological basis of human intelligence (Jung and Haier 2007).

A novel finding of the present study was that LGI was heritable across the cortex, extending a previous study that established the heritability of whole-brain GI (Docherty et al. 2015). This finding was not particularly surprising because many features of brain morphology are heritable. Nevertheless, it was necessary to establish the heritability of LGI before calculating genetic LGI–*g* correlations, which are only meaningful if both LGI and *g* are heritable traits. The previous study estimated the heritability of GI to be 0.71, which is much greater than most of the heritability estimates for LGI observed in GOBS or HCP. This result is also not surprising, because GI is likely to be contaminated by less measurement error than LGI. Heritabilities of all other traits were consistent with those published in previous studies.

The present study represents a replication of previous work and provides several important extensions to our understanding of the relationship between gyrification and cognition. First, we replicated previous work by finding positive and significant phenotypic LGI–*g* correlations (e.g., Gregory et al. 2016). Furthermore, we found that genetic LGI–*g* correlations were positive and significant (but only in HCP), suggesting that the relationship between gyrification and intelligence may be driven by pleiotropy. Since environmental LGI–*g* correlations were not significant, their net sign differed across GOBS and HCP, and their spatial patterns showed no consistency across samples, it is reasonable to conclude that they mostly reflected measurement error rather than meaningful shared environmental contributions to LGI and *g*.

In our view, the most important finding from the present study is that all LGI–*g* correlations, even the significant ones, were weak. Phenotypically, LGI at a typical vertex poorly

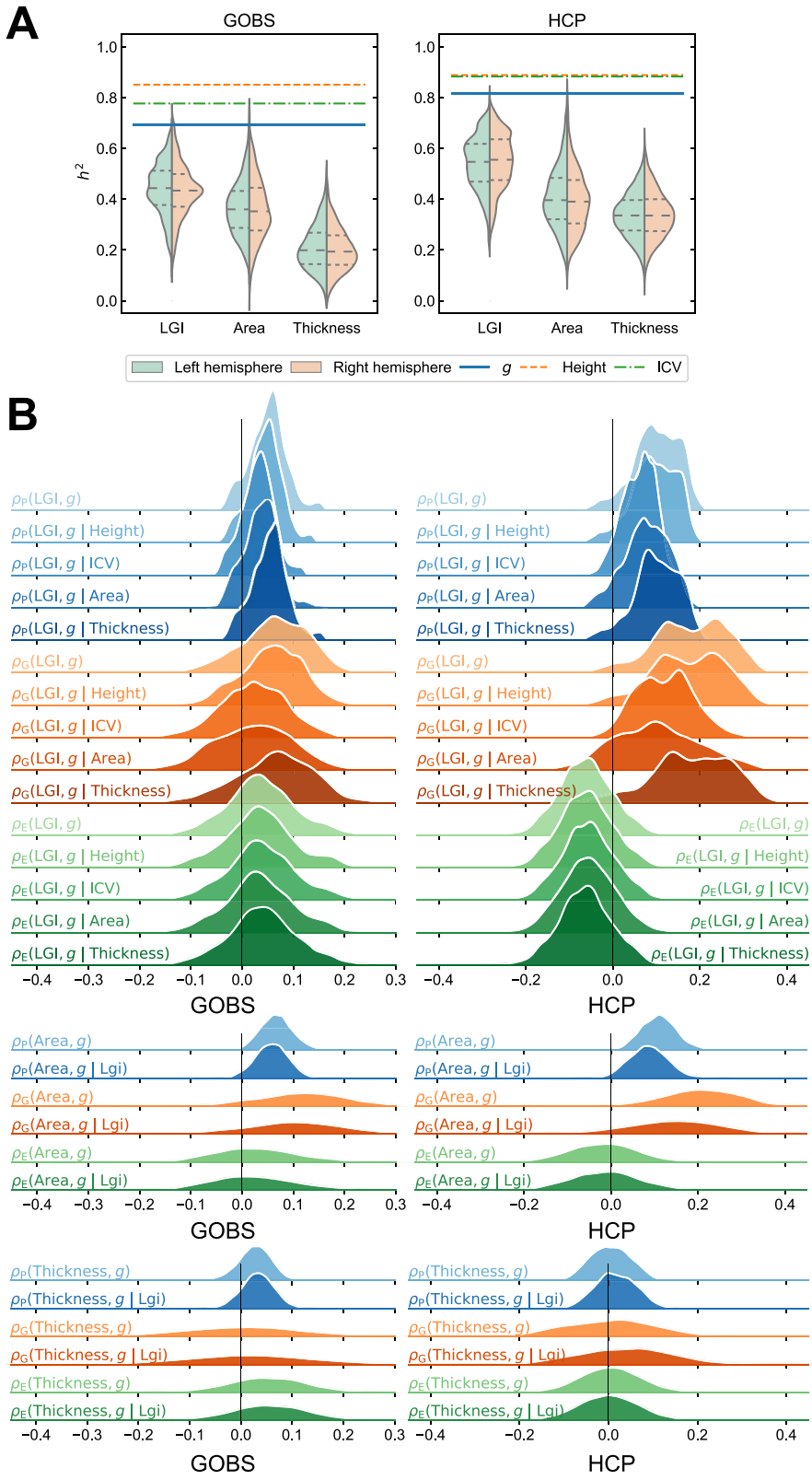

Figure 1. (A) Heritability estimates of all quantitative traits used in the present study. Violins are kernel density estimates (KDEs) of the distributions of heritability estimates of all vertexwise traits, split by hemisphere and with first, second, and third quartiles included as horizontal lines. (B) KDEs of full and partial LGI–*g*, area-*g*, and thickness-*g* correlation estimates.

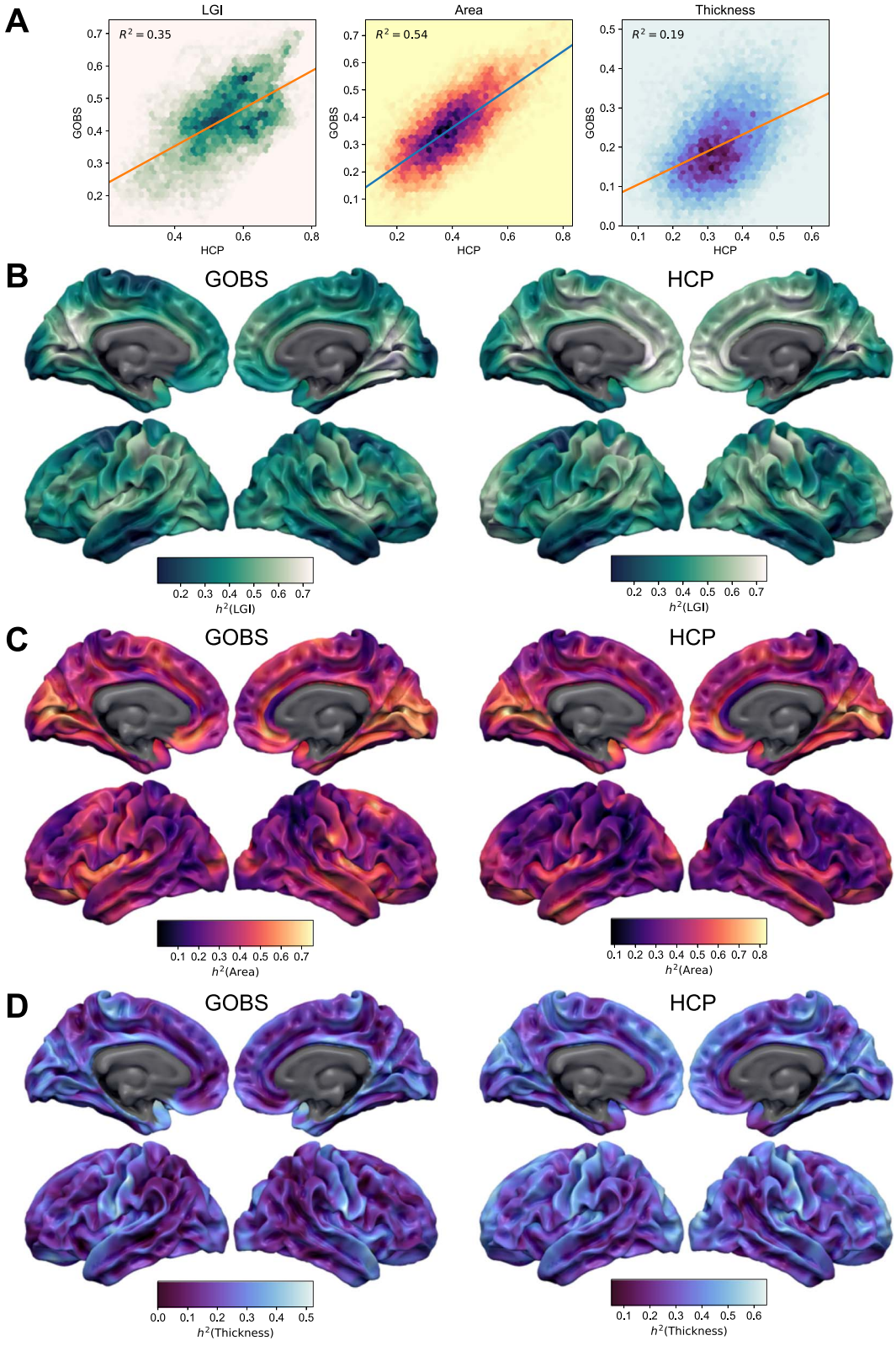

Figure 2. (A) Across-sample correlations in h2 parameter estimates. Darker hexagons indicate higher density and the oblique line denotes the linear trend. Such was the precision of each linear trend that 99% bootstrapped confidence intervals are not visible. (B–D) Neuroanatomical patterns of heritability estimates projected onto the fsaverage5 white-matter surface. While LGI, area, and thickness actually pertain to gray matter, it is convenient to use the white-matter surface instead because sulci are more readily viewable.

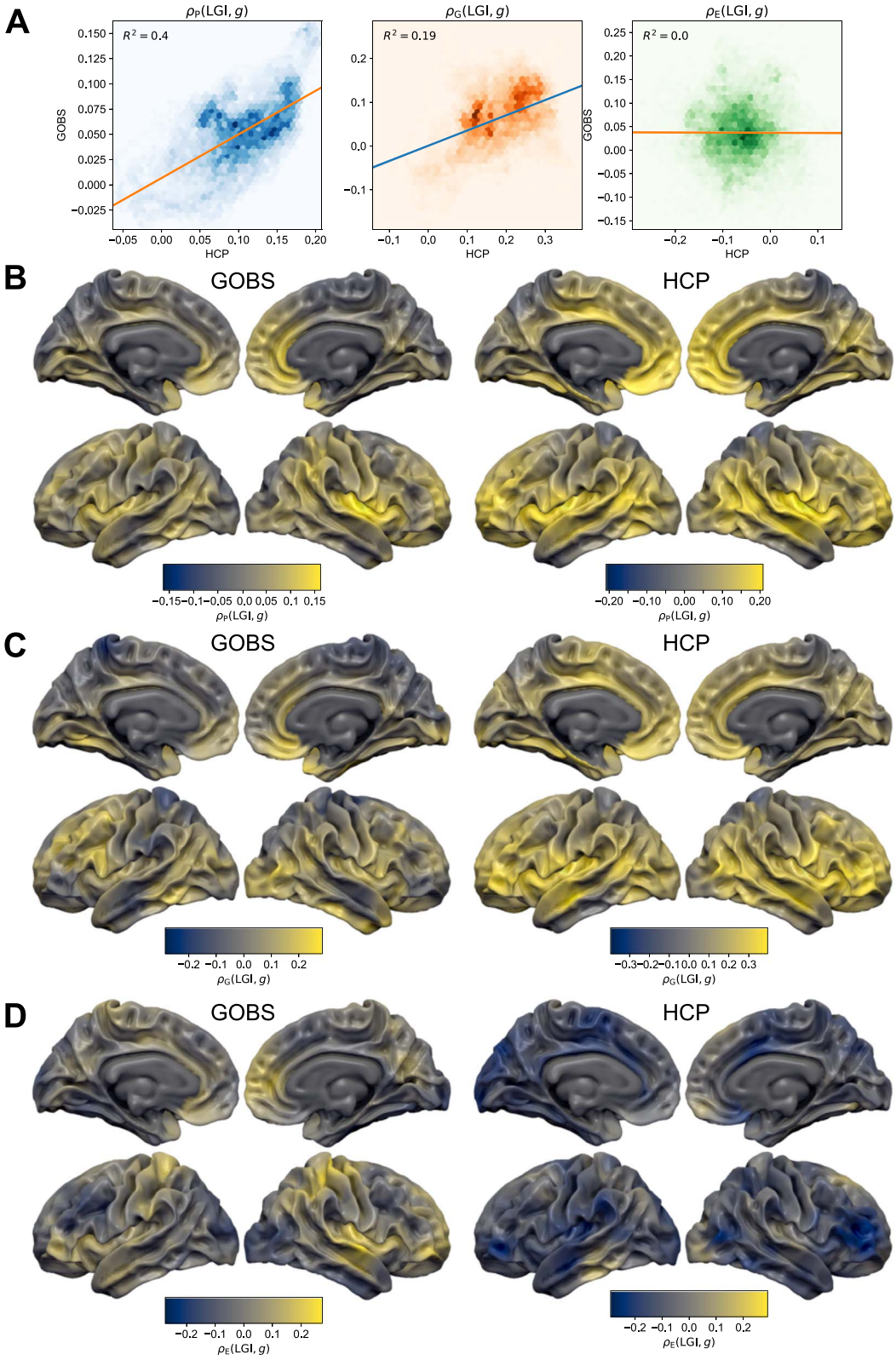

Figure 3. Same as Fig. 2, but for LGI–*g* correlations.

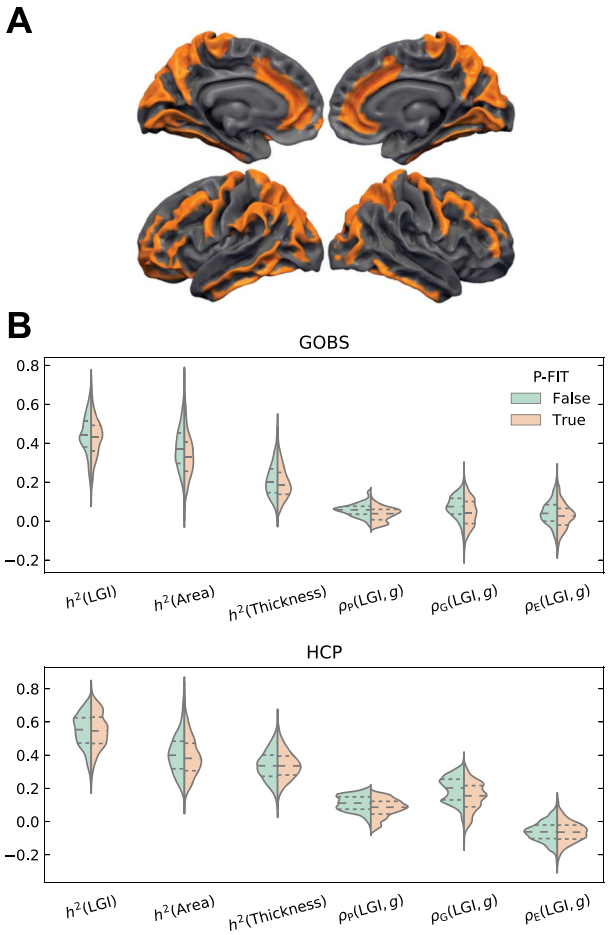

Figure 4. (A) Regions from the Destrieux et al. (2010) atlas considered to be implicated by the P-FIT by the present study, following Green et al. (2018). (B) KDEs of heritability LGI–*g* correlation estimates partitioned by whether each vertex was implicated by the P-FIT, with first, second, and third quartiles included as horizontal lines.

predicted *g*. Even when the predictive ability of all LGIs was considered together via ridge regression, at least 89% of the variance of *g* remained unaccounted for. Phenotypic and genetic LGI–*g* correlations were weaker than ICV–*g* correlations in the same participants, and about the same as area–*g* correlations. Partialing out ICV or area further reduced LGI–*g* correlations.

The volume of cortical mantle is often computed as the product of its area and thickness, but at the resolution of meshes typically used to represent the cortex, the variability of area is higher than the variability of thickness such that surface area is the primary contributor to the variability of cortical volume (Winkler et al. 2010), and therefore of its relationship to other measurements; the same holds, more strongly even, for parcellations of the cortex in large anatomical or functional regions. This means that the association between overall brain volume and cognitive abilities reported by previous studies (e.g., Pietschnig et al. 2015) is probably primarily driven by area–*g* correlations (Vuoksimaa et al. 2015). LGI is strongly correlated with area (Gautam et al. 2015; Hogstrom et al. 2013), which explains why partialing out either ICV or area reduced phenotypic and genetic LGI–*g* correlations in the present study. Thus, we conclude, based on our results, that the association between gyrification and cognitive abilities to a large extent reflects the already well-established relationship between surface area and cognitive abilities, and that the particular association between the unique portion of gyrification and cognitive abilities is extremely small.

The above conclusion is consistent with that of a previous twin study (Docherty et al. 2015), which examined genetic associations between overall cortical surface area, whole-brain GI, and cognitive abilities. The authors concluded that the genetic GI–*g* correlation could be more or less fully explained by the area–*g* correlation. It has been argued previously that focusing on whole-brain GI may miss important neuroanatomical specificity; however, our findings suggest that Docherty et al.'s conclusion holds for both local and global gyrification.

The P-FIT is a popular hypothesis concerning which brain regions matter most for human cognition (Jung and Haier 2007). The P-FIT was initially proposed to explain activation patterns observed during functional MRI experiments, but has been extended to aspects of brain structure. Previous studies have suggested that the association between gyrification and cognitive abilities may be stronger in P-FIT regions than the rest of the brain (Green et al. 2018; Gregory et al. 2016). However, when we tested this hypothesis, we actually found evidence to the contrary. Since neuroanatomical patterns of phenotypic and genetic LGI–*g* correlations were consistent across GOBS and HCP, this unexpected finding was unlikely to have been caused by a lack of specificity, such as if LGI–*g* correlations were distributed randomly over the cortex. Instead, while LGI–*g* correlations exhibited a characteristic neuroanatomical pattern, this pattern did not match the P-FIT. A potential limitation of the present study in this regard is that there is no widely accepted method of matching Brodmann areas (used to define P-FIT regions) to surface-based ROIs (used to group vertices). Therefore, one could argue that our selection of P-FIT regions was incorrect. While our selection was based on that of a previous study (Green et al. 2018), we nevertheless reperformed our analysis several times with different selections of P-FIT regions, and the results remained the same. Importantly, although we argue that the P-FIT is not a good model for the association between gyrification—a purely structural aspect of cortical organization—and cognitive abilities, our results should not be used to criticize the P-FIT as a hypothesis of the brain's functional organization, because function does not necessarily follow structure.

Most of our results were consistent across samples. However, estimates of heritability and genetic correlations were generally weaker in GOBS than HCP. Notably, some genetic LGI–*g* correlations were strong enough to surpass the FDR-corrected threshold for significance in HCP, but not GOBS. Such differences could be related to study design. One limitation of all family studies is that polygenic effects are susceptible to inflation due to shared environmental factors, which would cause overestimation of both heritability and genetic correlations. It could be argued that extended-pedigree studies, such as GOBS, are less susceptible to this kind of inflation than twin studies, such as HCP, because there are usually fewer shared environmental factors between distantly related individuals than twins (Almasy and Blangero 2010); this reduction in inflation comes at the expense of a reduction in power to detect polygenic effects, which could also explain the lack of significant genetic correlations in GOBS. It is unlikely that differences in results between samples were caused by differences in scanner or scanning protocol (Han et al. 2006). Furthermore, while GOBS and HCP participants completed

different cognitive batteries, both were comprehensive in terms of measured cognitive abilities, ensuring that *g* indexed a similar construct in both samples.

With the recent emergence of large, open-access data sets and international consortia, neuroimaging and genetics studies have entered a new era characterized by samples comprising many thousands of participants. In such large studies, trivial effects may be labeled as statistically significant. This observation is not new (Berkson 1938) and numerous solutions have been proposed, such as adopting more stringent significance criteria (Benjamin et al. 2018), scaling criteria by sample size (Mudge et al. 2012), testing interval-null rather than point-null hypotheses (Morey and Rouder 2011), and, most radically, abandoning the notion of statistical significance altogether (McShane et al. 2019). One could argue that these solutions suffer from their own drawbacks and are unlikely to be adopted by the scientific mainstream in near future. Therefore, in the meantime, we believe that it is imperative to judge, at least qualitatively, whether the sizes of statistically significant effects are large enough to justify one's conclusions, particularly when these conclusions may have broad, overarching implications. This idea is not new either (Kelley and Preacher 2012) but deserves to be restated. Based on the results of the present study, we are inclined to believe that gyrification minimally explains variation in cognitive abilities and therefore has somewhat limited implications for our understanding of the neurobiology of human intelligence.

Funding

National Institute of Mental Health (Grants MH078143 to principal investigator D.C.G., MH078111 to principal investigator J.B., and MH083824 to principal investigator D.C.G.); SOLAR (Sequential Oligogenic Linkage Analysis Routines) was supported by National Institute of Mental Health Grant (MH059490 to J.B.).

Notes

We thank the participants in HCP and GOBS, and researchers involved in the HCP for providing their data to all.
Conflict of Interest: None declared.

References

Almasy L, Blangero J. 2010. Variance component methods for analysis of complex phenotypes. *Cold Spring Harb Protoc*. 5:pdb.top77.

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 62(5):1198–1211.

Baare WF, Hulshoff Pol HE, Boomsma DI, Posthuma D, de Geus EJ, Schnack HG, van Haren NE, van Oel CJ, Kahn RS. 2001. Quantitative genetic modeling of variation in human brain morphology. *Cereb Cortex*. 11(9):816–824.

Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C *et al*. 2013. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*. 80:169–189.

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, Bollen KA, Brembs B, Brown L, Camerer C *et al*. 2018. Redefine statistical significance. *Nat Hum Behav*. 2(1):6–10.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 57(1):289–300.

Berkson J. 1938. Some difficulties of interpretation encountered in the application of the chi-square test. *J Am Stat Assoc*. 33:526–536.

Chung YS, Hyatt CJ, Stevens MC. 2017. Adolescent maturation of the relationship between cortical gyrification and cognitive ability. *NeuroImage*. 158:319–331.

Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*. 9(2):179–194.

Deary IJ, Spinath FM, Bates TC. 2006. Genetics of intelligence. *Eur J Hum Genet*. 14(6):690–700.

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT *et al*. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 31(3):968–980.

Destrieux C, Fischl B, Dale A, Halgren E. 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*. 53(1):1–15.

Docherty AR, Hagler DJ, Panizzon MS Jr, Neale MC, Eyler LT, Fennema-Notestine C, Franz CE, Jak A, Lyons MJ, Rinker DA *et al*. 2015. Does degree of gyrification underlie the phenotypic and genetic associations between cortical surface area and cognitive ability? *NeuroImage*. 106:154–160.

Eyler LT, Chen CH, Panizzon MS, Fennema-Notestine C, Neale MC, Jak A, Jernigan TL, Fischl B, Franz CE, Lyons MJ *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(3):304–314.

Fischl B, Sereno MI, Dale AM. 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 9(2):195–207.

Gautam P, Anstey KJ, Wen W, Sachdev PS, Cherbuin N. 2015. Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. *Behav Brain Res*. 287:331–339.

Glahn DC, Knowles EE, McKay DR, Sprooten E, Raventos H, Blangero J, Gottesman II, Almasy L. 2014. Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. *Am J Med Genet B Neuropsychiatr Genet*. 165B(2):122–130.

Glahn DC, Almasy L, Blangero J, Burk GM, Estrada J, Peralta JM, Meyenberg N, Castro MP, Barrett J, Nicolini H *et al*. 2007. Adjudicating neurocognitive endophenotypes for schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 144B(2):242–249.

Glahn DC, Almasy L, Barguil M, Hare E, Peralta JM, Kent JW, Jr DA, Contreras J, Pacheco A, Lanzagorta N *et al*. 2010. Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. *Arch Gen Psychiatry*. 67(2):168–177.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR *et al*. 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*. 80:105–124.

Green S, Blackmon K, Thesen T, DuBois J, Wang X, Halgren E, Devinsky O. 2018. Parieto-frontal gyrification and working memory in healthy adults. *Brain Imaging Behav*. 12(2):303–308.

Gregory MD, Kippenhan JS, Dickinson D, Carrasco J, Mattay VS, Weinberger DR, Berman KF. 2016. Regional variations in brain

gyrification are associated with general cognitive ability in humans. *Curr Biol*. 26(10):1301–1305.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R *et al*. 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *NeuroImage*. 32(1):180–194.

Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1(12):55–67.

Hogstrom LJ, Westlye LT, Walhovd KB, Fjell AM. 2013. The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. *Cereb Cortex*. 23(11):2521–2530.

JASP Team. 2018. JASP (Version 0.9) [computer program].

Jeffreys H. 1961. *Theory of probability*. 3rd ed. Oxford: Clarendon Press.

Jung RE, Haier RJ. 2007. The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav Brain Sci*. 30(2):135–54;discussion 154–187.

Keller MC, Garver-Apgar CE, Wright MJ, Martin NG, Corley RP, Stallings MC, Hewitt JK, Zietsch BP. 2013. The genetic correlation between height and IQ: shared genes or assortative mating? *PLoS Genet*. 9(4):e1003451.

Kelley K, Preacher KJ. 2012. On effect size. *Psychol Methods*. 17(2):137–152.

Knowles EE, Carless MA, de Almeida MA, Curran JE, McKay DR, Sprooten E, Dyer TD, Goring HH, Olvera R, Fox P *et al*. 2014. Genome-wide significant localization for working and spatial memory: identifying genes for psychosis using models of cognition. *Am J Med Genet B Neuropsychiatr Genet*. 165B(1):84–95.

Le S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 1(25):1–18.

McKay DR, Knowles EE, Winkler AA, Sprooten E, Kochunov P, Olvera RL, Curran JE, Kent JW, Jr CMA, Goring HH *et al*. 2014. Influence of age, sex and genetic factors on the human brain. *Brain Imaging Behav*. 8(2):143–152.

McShane BB, Gal D, Gelman A, Robert C, Tackett JL. 2019. Abandon statistical significance. *Am Stat*. 73:235–245.

Morey RD, Rouder JN. 2011. Bayes factor approaches for testing interval null hypotheses. *Psychol Methods*. 16(4):406–419.

Mudge JF, Baker LF, Edge CB, Houlahan JE. 2012. Setting an optimal alpha that minimizes errors in null hypothesis significance tests. *PLoS One*. 7(2):e32734.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al*. 2011. Scikit-learn: machine learning in python. *J Mach Learn Res*. (12):2825–2830.

Pietschnig J, Penke L, Wicherts JM, Zeiler M, Voracek M. 2015. Meta-analysis of associations between human brain volume and intelligence differences: how strong are they and what do they mean? *Neurosci Biobehav Rev*. 57:411–432.

Python Software Foundation. 2018. Python Language Reference, version 3.6.[computer program].

R Core Team. 2013. *R: a language and environment for statistical computing [computer program]*. Austria: Vienna.

Rakic P. 2009. Evolution of the neocortex: perspective from developmental biology. *Nat Rev Neurosci*. 10(10):724–735.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev*. 16(2):225–237.

Schaer M, Cuadra MB, Tamarit L, Lazeyras F, Eliez S, Thiran JP. 2008. A surface-based approach to quantify local cortical gyrification. *IEEE Trans Med Imaging*. 27(2):161–170.

Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmborg JV *et al*. 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res*. 6(5):399–408.

Spearman C. 1904. General intelligence, objectively determined and measured. *Am J Psychol*. 15(2):201–292.

van Buuren S, Groothuis-Oudshoorn K. 2011. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 3(45):1–67.

Van Essen DC. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*. 385(6614):313–318.

Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. 2013. The WU-minn human connectome project: an overview. *NeuroImage*. 80: 62–79.

Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era–concepts and misconceptions. *Nat Rev Genet*. 9(4):255–266.

Vuoksimaa E, Panizzon MS, Chen CH, Fiecas M, Eyler LT, Fennema-Notestine C, Hagler DJ, Fischl B, Franz CE, Jak A *et al*. 2015. The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cereb Cortex*. 25(8):2127–2137.

Welker W. 1990. Why does the cerebral cortex fissure and fold? In: Jones EG, Peters A, editors. *Cerebral cortex*. Boston, MA, USA: Springer, pp. 3–136.

Winkler AM, Greve DN, Bjuland KJ, Nichols TE, Sabuncu MR, Haberg AK, Skranes J, Rimol LM. 2018. Joint analysis of cortical area and thickness as a replacement for the analysis of the volume of the cerebral cortex. *Cereb Cortex*. 28(2):738–749.

Winkler AM, Kochunov P, Blangero J, Almasy L, Zilles K, Fox PT, Duggirala R, Glahn DC. 2010. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*. 53(3):1135–1146.

Zilles K, Palomero-Gallagher N, Amunts K. 2013. Development of cortical folding during evolution and ontogeny. *Trends Neurosci*. 36(5):275–284.

Zilles K, Armstrong E, Schleicher A, Kretschmann HJ. 1988. The human pattern of gyrification in the cerebral cortex. *Anat Embryol (Berl)*. 179(2):173–179.