



Published in final edited form as:

*Nat Genet.* 2021 June ; 53(6): 830–839. doi:10.1038/s41588-021-00827-w.

## Shared heritability of human face and brain shape

**Sahin Naqvi**<sup>#1,2,†</sup>, **Yoei Sleyp**<sup>#3</sup>, **Hanne Hoskens**<sup>3,4</sup>, **Karlijne Indencleef**<sup>4,5</sup>, **Jeffrey P. Spence**<sup>2</sup>, **Rose Bruffaerts**<sup>6</sup>, **Ahmed Radwan**<sup>4,7</sup>, **Ryan J. Eller**<sup>8</sup>, **Stephen Richmond**<sup>9</sup>, **Mark D. Shriver**<sup>10</sup>, **John R. Shaffer**<sup>11,12</sup>, **Seth M. Weinberg**<sup>11,12,13</sup>, **Susan Walsh**<sup>8</sup>, **James Thompson**<sup>14</sup>, **Jonathan K. Pritchard**<sup>2</sup>, **Stefan Sunaert**<sup>4,7</sup>, **Hilde Peeters**<sup>3</sup>, **Joanna Wysocka**<sup>1,15,16,‡,†</sup>, **Peter Claes**<sup>3,4,5,17,‡,†</sup>

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California, United States of America

<sup>2</sup>Departments of Genetics and Biology, Stanford University School of Medicine, Stanford, California, United States of America

<sup>3</sup>Department of Human Genetics, KU Leuven, Leuven, Belgium

<sup>4</sup>Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium

<sup>5</sup>Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium

<sup>6</sup>Department of Neurosciences, KU Leuven, Leuven, Belgium; Neurology Dept, University Hospitals Leuven, Leuven, Belgium; Biomedical Research Institute, Hasselt University, Hasselt, Belgium

<sup>7</sup>Department of Imaging and Pathology, Translational MRI, KU Leuven, Leuven, Belgium

<sup>8</sup>Department of Biology, Indiana University Purdue University Indianapolis, Indianapolis, Indiana, United States of America

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>†</sup>Corresponding author (naqvi@stanford.edu, wysocka@stanford.edu, peter.claes@kuleuven.be).

<sup>‡</sup>These authors jointly supervised this work

Author contributions

Conceptualization, P.C., J.W., S.N.; Methodology, J.P.S., P.C., J.T.; Software, P.C., H.H., K.I., R.J.E., J.T., S.N., Y.S.; Formal analysis, J.P.S., P.C.; Investigation, S.N., Y.S., P.C., R.B., A.R.; Resources, S.R., M.D.S., J.R.S., S.M.W., S.W., P.C.; Data Curation, S.N., Y.S., P.C., H.H., K.I.; Writing – original draft preparation, S.N., Y.S., H.P., J.W., P.C., A.R., S.S., R.B.; Writing – review and editing, H.P., Y.S., H.H., K.I., J.T., J.K.P., J.P.S., S.W., S.M.W., J.R.S., M.D.S., R.J.E.; Visualization, P.C., Y.S., S.N., A.R.; Supervision, P.C., J.W., H.P., S.S., J.K.P., S.W.; Project Administration, P.C., J.W., H.P., S.W., S.R., M.D.S., J.R.S., S.M.W.; Funding acquisition, P.C., J.W., H.P., S.W., S.R., M.D.S., J.R.S., S.M.W.

Ethics statement

This study was conducted in compliance with the principles of the Declaration of Helsinki, the principles of GCP and in accordance with all applicable regulatory requirements. Local ethics review and approval for this study (S63179) was performed and obtained from the ethical committee for research of the University Hospital UZ Leuven and the University KU Leuven.

Competing interests statement

The authors declare no competing interests.

Code availability

Matlab implementations of the hierarchical spectral clustering to obtain phenotypic shape segmentations are available from a previous publication <https://doi.org/10.6084/m9.figshare.7649024.v1><sup>7</sup>. Updated implementations used in this work are provided as described in Supplementary Table 7. The statistical analyses in this work were based on functions of the statistical toolbox in Matlab as mentioned throughout the Methods. Other materials and software used are available online (see URL section). No other custom software packages were used.

<sup>9</sup>Applied Clinical Research and Public Health, School of Dentistry, Cardiff University, Cardiff, United Kingdom

<sup>10</sup>Department of Anthropology, Pennsylvania State University, State College, Pennsylvania, United States of America

<sup>11</sup>Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

<sup>12</sup>Department of Oral and Craniofacial Sciences, Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

<sup>13</sup>Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

<sup>14</sup>Department of Psychology George Mason University, Fairfax, Virginia, United States of America

<sup>15</sup>Department of Developmental Biology, Stanford University School of Medicine, Stanford, California, United States of America

<sup>16</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, United States of America

<sup>17</sup>Murdoch Children's Research Institute, Melbourne, Victoria, Australia

# These authors contributed equally to this work.

## Abstract

Evidence from model organisms and clinical genetics suggests coordination between the developing brain and face, but the role of this link in common genetic variation remains unknown. We performed a multivariate genome-wide association study (GWAS) of cortical surface morphology in 19,644 European-ancestry individuals, identifying 472 genomic loci influencing brain shape, of which 76 are also linked to facial shape. Shared loci include transcription factors involved in craniofacial development, as well as members of signaling pathways implicated in brain-face crosstalk. Brain shape heritability is equivalently enriched near regulatory regions active in either forebrain organoids or facial progenitors. However, we do not detect significant overlap between shared brain-face GWAS signals and variants affecting behavioral-cognitive traits. These results suggest that early in embryogenesis, the face and brain mutually shape each other through both structural effects and paracrine signaling, but this interplay may not impact later brain development associated with cognitive function.

---

The human cerebral cortex forms the outer layer of gray matter of the brain and underpins cognitive function. It is characterized by complex folding patterns varying between species and individuals<sup>1,2</sup>. Family- and twin-based studies indicate substantial heritability of brain shape<sup>3,4</sup>, and a recent genome-wide association study (GWAS) found that brain shape is highly polygenic with genetic correlations to a broad range of neuropsychiatric disorders and behavioral-cognitive phenotypes<sup>5</sup>. These studies focused on pre-defined, univariate measures of brain shape, such as total or regional surface area, extracted from structural magnetic resonance imaging (MRI) scans<sup>6</sup>, which cannot capture morphological complexities of the cortical surface. We recently developed a data-driven approach to

phenotyping complex, multidimensional traits<sup>7</sup>; this multivariate approach, when applied to facial surface images, revealed numerous loci with no previously known role in human face shape variation<sup>7,8</sup>. Here, we implemented this approach to discover associations between common genetic variants and brain shape, using MRI data from middle-aged participants in the UK Biobank (UKB) free of disease diagnosis.

In addition to sharing complex morphologies, the development of the brain and face is highly integrated due to shared developmental lineage, spatial proximity, and signaling crosstalk between both structures<sup>9</sup>. Early in embryonic development, the rostral end of the ectodermally-derived neural tube gives rise to the forebrain, which in turn gives rise to the cerebrum that encompasses the cerebral cortex<sup>10</sup>. Just before forebrain formation, a subset of neuroepithelial cells within the neural folds give rise to facial progenitor cells called cranial neural crest cells (CNCCs)<sup>11</sup>. Following specification, CNCCs undergo an epithelial-to-mesenchymal transition and migrate ventrally<sup>12</sup>, giving rise to most of the craniofacial skeleton and connective tissue<sup>13</sup>. Early brain growth rates can modulate both positioning and outgrowth of the facial prominences<sup>14,15</sup>, as well as induce flexion and bone deposition of CNCC-derived basicranial bones<sup>16,17</sup> and neurocranial sutures<sup>18,19</sup>, respectively. Finally, paracrine factors secreted by either the developing forebrain<sup>20–23</sup> or CNCCs<sup>24–26</sup> modulate the face or brain development, respectively.

These physical and molecular interactions have been detailed by studies in developing chick and mouse embryos, but are also supported by widespread co-occurrence of neurodevelopmental and craniofacial malformations in rare human syndromes<sup>27</sup>. This phenomenon was noticed by Demyer et al.<sup>28</sup> in 1964, when he coined the phrase “the face predicts the brain” to describe correlations between the severity of brain and face malformations in holoprosencephaly patients. While in some cases this co-occurrence may be caused by pleiotropic gene functions, a number of human syndromes have been mapped to genes functioning in brain-face crosstalk through paracrine signaling<sup>29–31</sup>. Nonetheless, close developmental links between face and brain are underappreciated; whether and how they extend to common human genetic variation influencing brain and face shape is unknown.

## RESULTS

### Multivariate GWAS of brain shape

We adapted our previously published data-driven phenotyping approach<sup>7</sup> to brain shape, as measured by MRI scans of 19,644 individuals in UKB. Participants included were of primarily European ancestry, such that results do not pertain to cross-population brain shape differences. We focused on the mid-cortical surface (midway between the white-grey matter interface and the pial surface with the cerebrospinal fluid, as extracted using FreeSurfer<sup>6</sup>), which we refer to as brain shape. Using mid-cortical surfaces represented by a mesh of three-dimensional vertices, the method segments brain shape in a global-to-local manner, yielding brain segments at different hierarchical levels of scale. Within each segment, principal component analysis (PCA) is used to describe effects in multivariate shape-space explaining between-individual variation, and canonical correlation analysis (CCA) is used to define, for each variant, the linear combination of PCs maximally associated with single

nucleotide polymorphism (SNP) dosage. Unsurprisingly, GWAS of left and right hemispheres from the same individuals showed highly concordant results (Supplementary Fig. 1); we therefore performed subsequent analyses using left-right hemisphere averaged data.

Applying this pipeline to UKB MRI data defined 285 hierarchical segments (Fig. 1, Supplementary Table 1), decomposing brain shape into different levels of detail, from larger brain segments with integrated variation, to more smaller brain segments with local effects. Each hierarchical level is a bipartition of its parent; the first level consisted of the entire brain, while the second and third levels segmented the whole brain into halves and quadrants, respectively, and the final, ninth level resulted in numerous smaller segments (Figure 1b, right). Many smaller segments from the seventh hierarchical level onwards were discarded due to small surface areas, resulting in fewer total segments than the 511 ( $2^9 - 1$ ) expected. Nevertheless, the ninth hierarchical level yielded a substantial number (74) of retained segments; a tenth level would contribute few additional segments (Supplementary Fig. 2). The segmentation broadly agreed with the commonly-used Desikan-Killiany<sup>32</sup>, Destrieux<sup>33</sup>, and Glasser<sup>34</sup> brain atlases (Supplementary Fig. 3). Prior to GWAS, we adjusted for covariates including total brain volume, height, BMI, sex, and population structure, as well as performing standard SNP filtering and quality control (Methods). Applying linkage disequilibrium score regression (LDSC)-based heritability estimation to each segment's GWAS (see online Methods, Supplementary Note for details on extension to multivariate traits) yielded intercept values close to 1 (range across segments 0.987–1.007, mean 1.001, Supplementary Table 1), indicating minimal confounding by population structure or cryptic relatedness. In total, we conducted 285 multivariate GWAS using CCA, each corresponding to one segment. 38,630 SNPs showed genome-wide significant ( $P < 5 \times 10^{-8}$ ) association with brain shape in at least one segment; of these, 23,413 reached study-wide significance ( $P < 2.07 \times 10^{-10}$  correcting for the number of effective GWAS, estimated by permutation, Methods) in at least one segment. Collapsing these SNPs into independent signals based on linkage disequilibrium and distance yielded 472 and 242 loci reaching genome- and study-wide significance, respectively (Supplementary Table 2). Most of the 472 loci showed effects on multiple segments (305/472, 65%), and many showed effects on multiple quadrants (158/472, 33%) (Figure 1, Supplementary Table 2), consistent with global-to-local effects at multiple levels of brain shape. Masking of associations from progressively higher hierarchical levels revealed that segments from higher levels contributed a substantial fraction of associations – for example, segments beyond the first three levels contributed 169 and 55 loci reaching genome-wide and study-wide significance, respectively (Extended Data Fig. 1). Associations between the 472 loci and brain shape were depleted from the frontal lobe segments (except for the most anterior orbitofrontal cortex) and enriched in the occipital and temporal lobe segments (Supplementary Fig. 4), mostly in agreement with point-wise heritability estimates (Extended Data Fig. 2).

We assessed the overlap between the 472 loci and previous GWAS of brain surface areas or subcortical volumes<sup>5,35–39</sup>. The 472 loci recapitulated 27–78% of the associations reported in previous studies; the highest overlap of 78% was with a recent study of univariate brain surface area<sup>5</sup>, the phenotype most comparable to the shape measures studied here (Table 1). Of the 472 loci, 121 overlapped with those reported in previous studies on brain surface area

or subcortical volume, while 351 represent previously undescribed associations with brain morphology. To assess the reproducibility of the 472 loci on the same shape measures, we analyzed MRI data from the Adolescent Brain Cognitive Development (ABCD) study<sup>40</sup>. Of the 472 loci, 466 were tested for replication (Methods). At 5% FDR, we replicated at least one associated segment for 305 of 466 (65.4%) loci, and 2,645 of 3,586 (73.8%) locus-segment combinations (Supplementary Table 3). We observed consistent rates when subdividing based on the hierarchical level of the segments being replicated, albeit with a slight decrease in replication rate at higher levels (Extended Data Fig. 3). These replication rates are notable given the substantial age difference of the ABCD cohort (9–10 years versus 40–70 years in UKB). The high reproducibility of GWAS results between the two cohorts suggests that despite the known continued growth and morphological changes of the brain throughout adolescence and into adulthood<sup>41</sup>, many of the observed associations with brain shape originate during development and are maintained throughout life.

We next used Functional Mapping and Annotation of GWAS (FUMA)<sup>42</sup> and the Genomic Regions Enrichment of Annotations Tool (GREAT)<sup>43</sup> to identify pathways enriched among genes near the 472 loci, as well as curated gene panels used to guide disease diagnoses<sup>44</sup> to identify disease associations (Methods). As expected, we found strong enrichment for brain-specific processes (neurogenesis, axonogenesis, neuron differentiation, nervous system development, neuron projection guidance), morphogenesis-related processes (anatomical structure morphogenesis, animal organ morphogenesis), and neurodevelopmental disorders (intellectual disability, malformations of cortical development, ciliopathies). We also observed a weak enrichment of terms related to formation and closure of the neural tube, suggesting that early developmental events impact adult brain shape. Surprisingly, we also observed strong enrichment of terms related specifically to CNCC development and migration, as well as weaker enrichment of broader terms encompassing skeletal system development, chondrogenesis, and osteogenesis (Supplementary Data 1). Furthermore, we found strong and weak enrichments for craniosynostosis (premature closure of the cranial bone sutures) and clefting gene panels, respectively. These enrichments suggest a link between variation in brain shape and craniofacial skeletal development.

### Loci affecting both brain and face shape

To more directly test for sharing of genetic effects between brain and face shape, we intersected the 472 loci described in this study with 203 loci previously associated with face shape in European-ancestry individuals through a similar, open-ended phenotyping approach<sup>8</sup>. Thirty-seven of the loci for brain shape are linked to ( $r^2 > 0.2$ ) at least one of the face shape loci, significantly above random expectation ( $P = 2.03 \times 10^{-22}$ , OR = 10.6) and greater than the overlap with other traits that have similar numbers of genome-wide significant associations in the NCBI-EBI GWAS Catalog<sup>45</sup> (Extended Data Fig. 4). Identifying signals showing genome-wide significant association with one of brain or face shape and suggestive ( $P < 5 \times 10^{-7}$ ) association with the other resulted in 76 brain-face shared loci (Figure 2a).

Genes near the 76 brain-face shared loci were strongly enriched for disease associations, including “skeletal disorders” and “hearing and ear disorders”, consistent with the

contribution of CNCCs to craniofacial skeleton and ear structures. We next manually scanned the 76 brain-face shared loci for genes with known roles in craniofacial or brain development from human syndromes and/or mouse knockouts (Supplementary Table 4). We observed that many of the shared brain-face loci include genes encoding transcription factors (TFs) involved in neural crest formation and/or craniofacial skeletal development. Some of those TFs (for example *DLX5/6*, *SOX9*, *ZEB2*, *ZIC2*, *ZIC3*, *TCF4*) have known functions in both neural crest and brain development, and this pleiotropy may account for the shared brain-face genetic signals. However, other shared brain-face signals are associated with TFs thought to function primarily during neural crest rather than brain development, and whose mutation causes specific craniofacial defects; those TFs include *ALX1* and *ALX4* (associated with frontonasal dysplasias<sup>46,47</sup>), *TWIST1* (associated with Saethre-Chotzen Syndrome<sup>48,49</sup>), *PAX3* (associated with Waardenburg syndrome<sup>50</sup>), and *TFAP2B* (associated with CHAR syndrome<sup>51</sup>). Consistent with the primary role of these TFs in facial development, transcriptome analysis showed high expression in in-vitro derived human CNCCs and their chondrocyte derivatives<sup>52</sup>, but low/no expression in either glia or neurons of human forebrain organoids spanning a range of developmental stages<sup>53</sup> (Figure 2b). These observations suggest that genetic variants affecting key craniofacial TFs have a greater than previously appreciated impact on brain shape.

Interactions between face and brain can be architectural, with the forebrain acting as a structural support for facial development, and facial skeletal structures flexing to accommodate early brain growth<sup>54</sup>. However, these interactions can also involve paracrine signaling, with fibroblast growth factor (FGF), Hedgehog, and bone morphogenetic protein (BMP) pathways known to mediate the signaling from the developing brain to the face<sup>20–22</sup>. Interestingly, genes encoding members of all three pathways, FGF (*FGF2*, *FGF13*, *FGF18*, *SPRY2*), Hedgehog (*PTCH1*), and BMP (*BMP2*, *BMP4*) are among the shared brain-face loci. For example, mutations in *PTCH1*, encoding the receptor for the sonic hedgehog ligand, cause holoprosencephaly<sup>55</sup>, a congenital, structural forebrain anomaly with associated craniofacial malformations. Conversely, CNCCs secrete anti-BMP signaling molecules that modulate forebrain development<sup>24,25</sup>; expression of these BMP antagonists is dependent on the SIX family TFs, whose perturbation in CNCCs leads to both craniofacial malformations and secondary pre-otic brain defects<sup>56</sup>. *SIX1/4* is also among the 76 brain-face shared loci (Figure 2a). Furthermore, genes linked to other signaling pathways, including Wnt (*DAAM1*, *DAAM2*, *TNKS*, *AH11*, *FBXW11*, *MCC*) and transforming growth factor beta (*LEMD3*, *PPP2R3A*), are among the shared brain-face loci. Not unexpectedly, and in contrast to craniofacial TFs, signaling pathway ligands, receptors and regulators are variably expressed between in-vitro derived CNCCs and brain organoids (Figure 2b).

Phenotypically, these highlighted loci largely affect brain shape in the frontal and temporal lobes, and face shape in the forehead and nose, as exemplified by *PAX3* and *ALX1* (Figure 2c), consistent with the physical proximity of the frontonasal prominence and the forebrain during development. Phenotypic effects distinct from this pattern include effects of variants near *BMP4* and *DLX6* on jaw and chin morphology, consistent with their known roles in mandibular development<sup>57,58</sup>, and effects of variants near *PTCH1* on occipital lobe morphology (Figure 2c). Together, these results suggest that both cell-intrinsic mechanisms



and paracrine signaling pathways contribute to the substantial number of loci with shared associations with brain and face shape.

### **Genome-wide sharing of signals with neuropsychiatric disorders and behavioral-cognitive traits**

We next asked whether the brain-face overlap among genome-wide significant loci held across the genome, also considering GWAS of neuropsychiatric disorders and behavioral-cognitive traits. LDSC can estimate genetic correlations between univariate traits using signed summary statistics<sup>59</sup>. However, this approach is not applicable to unsigned statistics yielded by CCA. We therefore applied an alternative method of assessing genome-wide sharing of signals between two GWAS, summarizing SNP p-values within approximately independent LD blocks and computing Spearman correlations between the two summarized profiles (Methods). When applied to pairs of univariate GWAS, the Spearman correlation method was largely concordant with, albeit generally smaller in magnitude than, unsigned estimates of LDSC-estimated genetic correlations (Extended Data Fig. 5), indicating that it is a conservative, robust measure for quantifying genome-wide sharing of GWAS signals.

We first assessed sharing of association signals between 63 face segments and 285 brain segments (Supplementary Table 5). All four main facial quadrants, representing shape variation within the forehead, nose, lower face (mandible and cheeks) and philtrum, respectively, showed the most sharing with frontal lobe segments, particularly the most anterior portions such as the rostral prefrontal cortex, and the least sharing with parietal lobe segments (Figure 3a). Furthermore, among the facial quadrants, the forehead and nose showed more sharing with frontal lobe segments than the philtrum and lower face. These genome-wide correlations are consistent with the phenotypic effects of top brain-face shared loci (Figure 2c, Supplementary Fig. 5).

We next assessed sharing of signals with other brain-related traits. We used publicly-available genome-wide summary statistics for a range of neuropsychiatric disorders, behavioral-cognitive traits, and subcortical brain volumes from studies other than UKB, since our Spearman correlation measure does not control for sample overlap (Supplementary Table 6). As approximate negative controls, we used four immune-related diseases shown to have minimal genetic correlation with schizophrenia and bipolar disorder<sup>60</sup>. Subcortical volumes showed the most sharing with brain shape in the corresponding regions, but the magnitude of these correlations was relatively low (on par with sharing between brain and face shape), indicating that our multivariate GWAS approach detects effects beyond those resulting from changes in relative subcortical volume (Figure 3b). We found that disorders with primarily developmental etiology showed substantial sharing with brain shape in regions previously linked to these disorders. For instance, schizophrenia and attention deficit hyperactivity disorder (ADHD) showed sharing with shape variation in the primary auditory<sup>61,62</sup> and prefrontal cortex regions<sup>63</sup>, respectively. In contrast, we did not observe this association for Alzheimer's disease, caused by plaque buildup and neurodegeneration much later in life. Behavioral-cognitive traits such as intelligence, neuroticism and worry showed broader patterns of sharing with brain shape, reflecting the involvement of distributed cortical regions in these traits<sup>64-66</sup> (Figure 3b). Sharing between brain shape and

the immune diseases was generally lower than with neuropsychiatric disorders, behavioral-cognitive traits, or subcortical volumes, but reached significance for type 1 diabetes and rheumatoid arthritis (Figure 3c). This overlap may be because these immune traits have genetic correlation with brain-related traits other than those tested previously (schizophrenia and bipolar disorder), as suggested by a significant genetic correlation between rheumatoid arthritis and intelligence (Extended Data Fig. 6).

Finally, we compared the degree to which face shape shares signals with neuropsychiatric disorders, behavioral-cognitive traits, and subcortical volumes. Brain shape shares significant (FDR 5%) signal with most neuropsychiatric traits, as well as all behavioral-cognitive and subcortical volume traits analyzed. In contrast, face shape does not show significant sharing with any of the neuropsychiatric disorders or behavioral-cognitive traits, and significant but weaker sharing with the subcortical volume measures (Figure 3c). To confirm these patterns using univariate approaches, we performed GWAS on the most heritable individual PCs of full brain or face shape and computed genetic correlations using LDSC. Although genetic correlation estimates were noisy due to low heritability of univariate shape GWAS, they agreed with our Spearman correlation measure, finding non-zero genetic correlations between both brain and face shape and subcortical volumes, and between brain shape and both autism spectrum disorder and bipolar disorder (Extended Data Fig. 7). Thus, the substantial sharing of signals between brain and face shape (Figure 3a) appears to be mostly independent of neuropsychiatric disorder risk and behavioral-cognitive traits, perhaps because mutual influences of face and brain shape on each other involve phenotypic effects on brain shape distinct from those influencing neuropsychiatric disorder risk and behavioral-cognitive traits.

### Cell-types influencing brain and face shape

Our results thus far suggest that a substantial fraction of brain shape variation is underpinned by face shape, but that these observed effects are largely independent of effects shared between brain shape and other cognitive traits. To test this idea further, we sought to identify the cell-types most enriched for heritability of brain shape, face shape, and other cognitive traits. Partitioning heritability into cell-type-specific functional annotations via stratified LD score regression (S-LDSC) can prioritize trait-relevant cell-types, but was developed for univariate traits<sup>67</sup>; we thus sought to extend the theoretical framework of S-LDSC to multivariate traits such as our brain and face shape GWAS. We proved that when applying unstratified LDSC<sup>59</sup> to  $\chi^2$  statistics obtained from multivariate traits with independent dimensions and further corrected for dimensionality, the LDSC-estimated heritability equals the average heritability of the component univariate traits (see Methods, Supplementary Note), a proof that we validated through heritability estimation of each PC making up the full face (Extended Data Fig. 8). By extension, heritability enrichments obtained by applying S-LDSC on multivariate, corrected  $\chi^2$  statistics partitioned by annotation represent the average heritability enrichment for each component univariate trait (see Methods, Supplementary Note).

We collected genome-wide data on open chromatin (inferred from Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)) and active regulatory regions (inferred



from chromatin immunoprecipitation followed by sequencing (ChIP-seq) of histone marks) from a variety of cell-types and tissues, including in-vitro derived CNCCs and their chondrocyte derivatives<sup>52,68</sup>, embryonic craniofacial tissue at different stages of development<sup>69</sup>, neuronal and glial cells from three-dimensional forebrain organoids at various differentiation stages<sup>53</sup>, and both fetal and adult brain tissue<sup>70</sup>. We quantified brain and face shape heritability enrichments for these cell-type-specific annotations (Supplementary Data 2). Face shape showed significant (5% FDR) heritability enrichment specific to regulatory regions in craniofacial cell-types (mean Z-score 4.58) (Figure 4a). Brain shape showed significant and comparable heritability enrichments for regulatory regions in craniofacial cell-types and tissues, brain organoids, and fetal brain tissue (mean Z-scores 4.23, 3.23, 3.33, respectively) (Figure 4b). Within brain organoids, the strongest enrichments were for early-stage glial cells and whole organoids (mean Z-score 4.11) (Extended Data Fig. 9), consistent with an important role for radial glial cells in corticogenesis and in agreement with enrichments of brain surface area heritability<sup>5</sup>. The strong enrichments for craniofacial cell-types, which were more significant than organoid enrichments in the orbitofrontal and medial temporal lobes (Supplementary Fig. 6), suggest that heritability shared between brain and face shape is mediated primarily by CNCCs and their derivatives early in embryonic development. Consistent with this idea, quantifying brain shape heritability enrichments after removing the 76 brain-face shared loci resulted in decreased enrichment for CNCCs (Z-score difference  $-0.68$ ) and slightly increased enrichment for early-stage glial cells (Z score difference 0.23) (Extended Data Fig. 10).

Finally, we quantified heritability enrichments for neuropsychiatric disorders, behavioral-cognitive traits, and subcortical volumes. Neuropsychiatric disorders and behavioral-cognitive traits showed enrichment patterns distinct from those of brain shape, with significant enrichment for both fetal and adult brain tissue (mean Z-scores 2.17 and 2.64, respectively), and broad enrichment across stages and cell-types of brain organoids (mean Z-score 2.46). In contrast to brain shape, these traits showed no enrichment for craniofacial cell-types or tissues (mean Z-score  $-0.92$ ) (Figure 4c). Subcortical volumes showed mixed enrichment patterns, with some regions (amygdala, caudate) similar to those of multivariate brain shape and others (putamen) closer to those of neuropsychiatric disorders and behavioral-cognitive traits. These results suggest that while much of the shared genetic variation between brain and face shape is mediated by regulatory regions in CNCCs and their craniofacial derivatives, variation in these regions does not appear to impact neuropsychiatric disorder risk or other behavioral-cognitive traits.

## DISCUSSION

Here, we applied multivariate phenotyping to discover numerous loci underlying common variation in brain shape. While these loci broadly implicate known pathways in brain development, the precise mechanisms by which they modulate brain shape are unknown, suggesting further avenues of investigation. As part of our study, we extended techniques for estimating genome-wide and partitioned heritability, originally developed for univariate traits, to multivariate traits. We anticipate that these and similar extensions will become increasingly useful with the greater availability of high-dimensional imaging or morphological data in large sample sizes.

We found a striking convergence of common genetic variation affecting brain and face shape, at least in part mediated by regulatory regions active in CNCCs and their derivatives. These observations suggest a larger than previously appreciated role of the face in shaping development of the brain and its morphological variation between individuals. However, these shared genetic effects do not appear to significantly impact neuropsychiatric disorder risk or cognitive functions. Our results are therefore consistent with a model whereby CNCCs and their derived cranial structures significantly influence brain shape through both physical interactions and paracrine signaling early in embryogenesis, but later shaping of cortical morphology, through processes such as the folding of the cortical surface<sup>71</sup>, has a greater impact on cognitive traits. Nevertheless, we cannot exclude the possibility that future GWAS of cognitive traits shows significant overlap with brain-face shared genetic effects, perhaps due to alternative trait definitions or to greater statistical power.

A number of developmental mechanisms could mediate shared brain-face genetics. One potential contribution comes from the common neuroepithelial origins of the two structures, with genes influencing growth, patterning and cell fate decisions within the neural plate ultimately affecting cell allocation within distinct parts of the brain and face; examples of such neural plate genes within brain-face shared loci include *ZIC2* and *ZIC3*<sup>72–74</sup>. Another potential mechanism entails common genetic variation modulating expression of genes with independent roles in both brain and face development. *SOX9*, encoding a TF with key functions in neural crest development and chondrogenesis, but which is also required for gliogenesis<sup>75</sup>, is an attractive candidate for this mechanism. Nonetheless, the primary impact of brain-face shared genetic effects on facial regions from the frontonasal prominence and anterior forebrain regions of the brain suggests additional, proximity-based mechanisms, which can be either structural, or mediated by paracrine signaling. While brain and face development must be tightly coordinated, the brain is thought to have greater structural effects on craniofacial development than the reverse, as the forebrain can serve as structural support for facial development<sup>54</sup> as well as induce flexion of the basicranium and bone deposition at coronal sutures through growth-dependent tensile forces<sup>17,18,54</sup>. However, we find multiple brain-face shared loci near TFs with known, cell-intrinsic roles in, and expression specific to, CNCCs and their derivatives. Furthermore, mutations in these TFs result in malformations of the frontal facial skeleton, such as coronal synostosis (*TWIST1*)<sup>48,49</sup> or frontonasal dysplasias (*ALX1* and *ALX4*)<sup>46,47</sup>. One explanation for these results is that these TFs control regulatory programs ultimately modulating the ability of the craniofacial skeleton to respond to and accommodate brain growth, causing subtle changes in brain shape. It is also possible, however, that these TFs exert some phenotypic effects on brain shape by regulating expression of signaling ligands secreted from the face. For example, CNCCs secrete BMP antagonists that modulate forebrain development by blocking BMP and FGF production in the anterior neural ridge<sup>24,25</sup>. BMP antagonist production in CNCCs is regulated by the SIX family TFs<sup>56</sup>, with *SIX1/4* lying near a shared brain-face GWAS signal (Figure 2a). In the reverse direction, studies in chick embryos have shown that Fgf, Shh, and BMP ligands are secreted by the forebrain and regulate the formation of the frontonasal ectodermal zone (FEZ), a signaling center that in turn patterns the frontonasal prominence of the developing face<sup>20–22,76</sup>. Notably, our study implicates all three of these signaling pathways, nominating specific ligands and receptors whose

modulation may be associated with the brain-face crosstalk. Furthermore, our study nominates other pathways, such as Wnt and TGF-beta, for roles in paracrine brain-face signaling. Altogether, we uncovered common genetic variants yielding numerous candidate molecular players whose diverse mechanistic roles in mediating brain-face interactions during development can be examined in future studies.

Relationships of facial shape with cognitive and personality traits fascinated humans since ancient times, from the Ancient Greeks, who introduced ‘physiognomy’ to describe a practice of assessing one’s personality from facial appearance<sup>77</sup>, through the Vedic traditions of Samudrika Shastra<sup>78</sup> and to the Chinese art of face reading<sup>79</sup>. The concept of physiognomy was revived in the 18<sup>th</sup> century by Johan Kaspar Lavater, and later led to a related pseudoscientific theory, phrenology, popularized by Franz Josef Gall. Both theories have a troubled history, as they have been used to justify racial discrimination as well as eugenic theories<sup>80,81</sup>. While the original formation of physiognomy has been debunked, modern studies have found correlations between facial width-to-height ratios and aggressive tendencies<sup>82</sup>, with regrettable renewed efforts in using machine learning approaches to detect such correlations raising serious ethical concerns<sup>83,84</sup>. Our results argue that while the ancient human intuition of a close relationship between the face and brain has genetic support at the morphological level, there does not appear to be genetic evidence for the supposed predictive value of face shape in behavioral-cognitive traits, which formed the core of physiognomy and related theories.

## METHODS

### UK Biobank data preprocessing

The UK Biobank project (UKB) encompasses ~500,000 British volunteers with informed consent containing genetics, non-imaging variables and brain imaging data acquired using a fixed protocol<sup>85</sup>. Hereby, brain T1-weighted magnetic resonance imaging (MRI) scans of the UKB, as well as genotyping and covariate information (e.g. sex, age, height, weight, among others), were used as the discovery dataset. We utilized release v1.5 (August 2018) which holds a cohort of 21,780 subjects. This cohort was composed of an adult population (40 to 70 years old, mean of 60 years old), with slightly more females than males (51.6% vs. 48.4% respectively), a predominantly self-reported white British ancestry (97.1%), and an average body mass index (BMI) of 26.6.

For 21,780 subjects, we processed raw MRI data for a surface-based analysis of the cortex using the following four-step procedure. Further details for each step are provided in Supplementary Note, section ‘UK Biobank data processing.’

First, the cortical surfaces were segmented and reconstructed from the MRI volumetric data using recon-all (FreeSurfer<sup>86</sup> v.6.0.0; URL section). In this step 20,409 images were processed successfully.

Second, to obtain a minimally preprocessed pipeline similar to the one of the Human Connectome Project (HCP – URL Section), the Connectivity Informatics Technology

Initiative file format) (CIFTIFY, URL section) was used to convert FreeSurfer's recon-all command output to a HCP-style file format and structure<sup>87</sup>.

Third, from the CIFTIFY output, we selected the mid-cortical surface of the left and right hemisphere, which is the surface that runs at the mid-distance between the white surface (at the interface between gray and white matter) and the pial surface (the external cortical surface)<sup>88</sup>. The mid-cortical surface does not over or under-represent gyri or sulci<sup>89</sup>, but is otherwise an arbitrary choice.

Fourth, as quality control for each hemisphere, we checked the resulting mid-cortical surfaces for mesh artifacts in a semi-automatic manner. All images passed this quality control, yielding 20,407 processed images.

For the list of 20,407 subjects with preprocessed images, we selected genomic data from the UK Biobank, which consisted of the version 3 (March 2018) imputed SNP genotypes, imputed to the Haplotype Reference Consortium and merged UK10K and 1000 Genomes (phase 3) panels. See Supplementary Note, section 'UK Biobank data processing' for more details on filtering of SNPs and individuals based on ancestry and relatedness. This resulted in 9,705,931 filtered SNPs for GWAS analysis on 19,670 unrelated subjects of European descent.

For the list of 19,670 subjects with preprocessed brain and genetic data, we collected the following list of covariates to control for during statistical testing: genetic sex, age, age-squared, height, weight, diastolic blood pressure, systolic blood pressure, and the first 20 genetic PCs. Furthermore, the following imaging specific parameters were also included, following Elliot et al.<sup>90</sup>: volumetric scaling from T1 head image to standard space, XYZ-position of brain mask in scanner co-ordinates, Z-position of table/coil in scanner co-ordinates, date of attending assessment center, and assessment center (coded as a dummy variable for each of the 21 centers). See Supplementary Note, section 'UK Biobank data processing' for more details on covariate-based filtering individuals. Next, to symmetrize brain shape, the right hemisphere was reflected to the side of the left hemisphere, by changing the sign of the x-coordinate for all of the 29,759 3D vertices on the surface of the right hemisphere. We performed a generalized Procrustes superimposition (GPA)<sup>91</sup>, thus eliminating differences in position, orientation, and scale (measured by centroid size) of all left and right hemispheres pooled together. We computed the symmetric brain component as the vertex-wise averaged brain surface of paired and superimposed left and right hemispheres. This resulted in a final discovery dataset of 19,644 subjects containing preprocessed MRI image data on the mid-cortical symmetrized surface, 9,705,931 imputed SNPs, and 54 covariates.

### **ABCD study data preprocessing**

The Adolescent Brain Cognitive Development Study (ABCD) (URL section) is a longitudinal study following brain development and health through adolescence<sup>40</sup>. A total of 11,411 MRI scans with additional information on sex and age were available from the data release of April 2019 and of those 11,393 images were processed successfully using the four-step imaging preprocessing described above.

In total 10,627 individuals from the ABCD dataset provided with genetic data on 517,724 SNP variants. These were imputed via the *Odyssey*<sup>92</sup> pipeline using the SHAPEIT4<sup>93</sup> and IMPUTE5<sup>94</sup> workflow to phase and impute respectively. The Haplotype Reference Consortium (HRC)<sup>95</sup> reference panel was used for imputation. Quality control prior to phasing and imputation includes using the McCarthy Group's Imputation preparation program (URL section) to check and fix strand, alleles, position, and reference/alternative problems as well as removing ambiguous A/T and G/C SNPs with minor allele frequencies greater than 0.4. See Supplementary Note, section 'ABCD study data preprocessing' for more details on phasing, imputation, and ancestry-based selection. These steps resulted in a final replication dataset of 4,470 individuals with preprocessed MRI image data, representing brain shape, 15.3M imputed SNPs and 7 covariates (sex, age and the first 5 genetic PCs). The minimum and maximum age of this final replication dataset, was 8.9 years and 11 years, respectively, with a mean age of 9.9 years. 46.5% are female and 53.5% are male.

### Auxiliary traits GWAS summary statistics

We collected publicly available genome-wide summary statistics for 26 auxiliary traits encompassing neuropsychiatric disorders<sup>96–101</sup>, behavioral-cognitive traits<sup>102–104</sup>, subcortical volume measures<sup>36–38</sup>, and immune-related disorders<sup>105–108</sup> with limited genetic correlation with schizophrenia or bipolar disorder<sup>60</sup>. In Supplementary Table 6, we provide links to relevant publications and URLs for these summary statistics.

### Point-wise SNP-heritability estimation of the mid-cortical surface

For each of the 29,759 vertices of the averaged mid-cortical 3D surfaces in UKB we computed a multivariate (X, Y and Z coordinate per vertex) narrow-sense heritability from common SNP variants using a linear mixed model (LMM). A genomic relationship matrix (GRM) modelled as random effects in the LMM was computed from LD-pruned SNPs (Plink 1.9, 50 variant window-size, 5 variant step size, 0.2  $r^2$ ). The first 10 genetic PCs and additional covariates (sex, age, height, weight, diastolic and systolic blood pressure) were modelled as fixed effects in the LMM. We used the open-source software SNPLib (URL Section)<sup>109</sup>, whose implementation is equivalent to GCTA<sup>110</sup> for a homogenous population.

### Global-to-local (G2L) segmentation of the mid-cortical surface

UKB served as discovery cohort using a data-driven global-to-local (G2L) segmentation of brain shape similar to previous work on face shape<sup>7,111</sup>. First, the superimposed and symmetrized mid-cortical surfaces were corrected using a partial least-squares regression (PLSR, function `plsregress` from Matlab 2019b) for all UK Biobank covariates listed above, augmented with centroid size to eliminate allometric effects of size on brain shape<sup>91</sup>. Second, pair-wise structural connections based on the RV-coefficient<sup>112</sup> between each pair of 3D surface vertices generated a squared similarity matrix (29,759 x 29,759). Third, a Laplacian transformation was applied to enhance similarities prior to eigendecomposition of this squared matrix. Finally, within the eigen spectral map, K-means++ clustering was used to group highly correlated vertices that segment the brain into separate modules. This was done in a bifurcating hierarchical manner using eight levels, resulting in a total of 511 hierarchically linked facial segments, with 1, 2, 4, 8, 16, 32, 64, 128, 256 non-overlapping

modules at levels 0, 1, 2, 3, 4, 5, 6, 7, 8. In contrast to our work on facial shape<sup>7,111</sup>, we removed segments with fewer vertices than 1% of the total vertex count. This resulted in 285 segments across eight levels as depicted in Figure 1. While the precise choice of number of hierarchical levels and vertex cutoff is arbitrary, the number of additional segments followed an “elbow” trajectory, with few segments being retained at hierarchical levels greater than the ninth (Extended Data Fig. 2). Segmentation depth and cutoff criteria were determined prior to performing GWAS. See Supplementary Note, section ‘Global-to-local (G2L) segmentation of the mid-cortical surface’ for details on the multivariate phenotyping approach within each brain segment.

### Overlap of brain atlases with G2L segmentation

We investigated the overlap of brain segments at each of the eight levels from our G2L segmentation with brain regions from three commonly used brain atlases (Desikan Killiany (34 distinct gyral based regions)<sup>32</sup>, Destrieux (74 distinct gyral and sulcal based regions)<sup>33</sup>, and the Glasser (180 distinct multi-modal based regions)<sup>34</sup>). See Supplementary Note, section ‘Overlap of brain atlases with G2L segmentation’ for details on computing overlap between our brain segments and brain atlases.

### G2L multivariate genome-wide discovery

The global-to-local phenotyping partitioned brain shape into overlapping (across different hierarchical levels), as well as non-overlapping (within a single hierarchical level) segments, each of which was represented by a different subset of mid-cortical surface vertices and spanned by multiple dimensions of variation (principal components). See Supplementary Note, section ‘G2L multivariate genome-wide discovery’ for details on the canonical correlation (CCA)-based approach used to discover SNP-phenotype associations.

A significance threshold of  $P = 5 \times 10^{-8}$  was used to declare “genome-wide significance”, which corresponds to a Bonferroni correction for 1 million independent tests in a European-ancestry cohort<sup>113</sup>. Due to 285 multivariate GWAS runs, the multiple comparisons burden was magnified. Therefore, we also determined a more stringent threshold for declaring “study-wide significance” which accounts for the effective number of independent tests. In a first instance, the number of eigenvalues larger than one of a pairwise multivariate correlation (RV-coefficient) matrix (285 x 285)<sup>114</sup>, determined a total of 210 independent tests. In a second instance, following the procedure in Kanai et al.<sup>115</sup>, we obtained an empirical estimate of the number of independent tests using the 472 lead SNPs representing the genome-wide significant independent loci, to keep the estimations computationally tractable. See Supplementary Note, section ‘G2L multivariate genome-wide discovery’ for details on empirical estimation of the number of independent tests.

### Peak detection, overlap and annotations

We observed 38,630 SNPs and 23,413 SNPs at the level of genome-wide and study-wide significance, respectively. These were clumped into 472 (genome-wide) and 243 (study-wide) independent loci in three steps, detailed in Supplementary Note, section ‘Peak detection, overlap and annotations’.



To study functional enrichment for genes near the 472 genome-wide lead SNPs, we performed gene ontology (GO) analysis using GREAT<sup>43</sup> (v4.0.4) and FUMA<sup>42</sup> (v1.3.6) using default settings. GO terms that were significant by both binomial and hypergeometric tests (False Discovery Rate (FDR)  $q$ -value  $< 0.05$ ) across three or two windows were reported as strongly and weakly enriched respectively.

In determining overlap between lead SNPs from different GWAS, we used a similar strategy: two lead SNPs tag the same genetic locus if they are within 10kb of each other or if they are within 1Mb of each other and with  $r^2 > 0.2$ . For quantifying overlap between the 472 brain shape loci and 430 other studies from the NCBI-EBI GWAS Catalog, we defined LD blocks of 0.2 around the 472 loci using Plink 1.9, and then calculated odds ratio and  $P$  for the overlap between these blocks and any given GWAS using bedtools v2.27.1, fisher function.

In determining brain-face shared loci, we first considered the 472 genome-wide lead SNPs from the brain GWAS and looked for any SNP within 10kb or within 1Mb and  $LD > 0.2$  of these lead SNPs with at least a genome-wide suggestive ( $P \leq 5 \times 10^{-7}$ ) association with face shape<sup>111</sup>. This resulted in 57 loci with evidence of association in brain and face shape. Then we took the 203 genome-wide lead SNPs reported in the face GWAS<sup>111</sup>, and clumped them if two lead SNPs were within 10kb or within 10Mb with  $r^2 > 0.01$ . For the resulting 197 independent genome-wide facial lead SNPs we selected any SNP within 10kb or within 1Mb and with  $r^2 > 0.2$  with at least suggestive ( $P \leq 5 \times 10^{-7}$ ) association with brain shape. This resulted in another 54 loci with evidence of association in brain and face shape and together with the previous 57 loci they were clumped (within 10kb or within 1Mb and  $r^2 > 0.2$ ) into a final set of 76 brain-face shared loci. We manually identified candidate genes in the vicinity of the 76 brain-face shared loci. For each locus, we first considered all genes within 500kb of the lead SNP. We primarily relied on evidence for these genes' involvement in a human craniofacial or neurodevelopmental syndrome, or for evidence of craniofacial or neurodevelopmental defects in knockouts of their orthologs in mice. We also considered associations with Gene Ontology (GO) terms related to craniofacial development, neurodevelopment, or skeletal system development. In some cases (i.e. *SOX9*, where enhancer-promoter interactions over 1Mb have been described<sup>52</sup>), we extended the window to within 750kb of the lead SNP.

### ABCD replication testing

The ABCD study data was used for replication, with UKB discovery cohort used as a phenotyping reference. First, the GPA superimposed and symmetrized mid-cortical shapes were corrected for sex, age, and the first five genetic PCs, augmented with centroid size to eliminate allometric effects of size on brain shape<sup>91</sup> using PLSR. Second, the PLSR residuals that were centered on average brain shape of the ABCD study were added to average brain shape of UKB. Third, the corrected and re-centered brain shapes were segmented using the G2L segmentation and projected onto the principal components of the UKB segments.

For each discovery lead SNP in a particular brain segment the replication panel was projected onto the latent shape trait of the lead SNP. This generated univariate projection scores as phenotypes<sup>116</sup> to test for in the replication panel that are equivalent to the latent

shape traits or phenotypes in the discovery panel. See Supplementary Note, section ‘ABCD replication testing’ for details on replication testing and FDR thresholds<sup>117</sup>.

### Clinical gene-panel overlap

Gene panels were downloaded from the Genomics England PanelApp website. Only panels used for clinical interpretation in the 100,000 Genomes project were selected (provided by PanelApp<sup>44</sup>). The clinical gene-panels were merged in disease (sub)categories according to the 100,000 Genomes Project criteria (e.g. the clinical gene panel “Intellectual Disability” belongs to the sub-category “Neurodevelopmental Disorders”, which is part of the “Neurology and Neurodevelopment” disease category). Only genes with high confidence for gene-disease association were included in the clinical gene panels. See Supplementary Note, section ‘Clinical gene-panel overlap’ for details on calculation of gene set overlaps and significance.

### Expression analyses of candidate genes at brain-face overlapping loci

Gene expression levels ( $\log_2(\text{TPM})$  values) for three-dimensional forebrain organoids and purifying neuronal or glial lineages were obtained from Trevino et al.<sup>53</sup> (GSE132403). Raw RNA-seq reads from CNCCs at passages 1-4, as well as day 9 chondrocytes derived from P4 CNCCs, were obtained from Long et al.<sup>52</sup> (GSE145327), and TPM values were quantified using kallisto v0.44.0<sup>118</sup> with sequence-biased bias correction.

### LD score regression SNP-heritability for multivariate traits

In the Supplementary Note, we provide a general proof showing that when applying LDSC to summary statistics of a multivariate GWAS, albeit with a small correction to the resulting  $\chi^2$  statistics, the heritability estimated by the LDSC slope is equal to  $\frac{1}{D}\text{trace}(\Sigma_G \Sigma_P^{-1})$ , which is a  $D$ -dimensional generalization of heritability for genetic and phenotypic covariance matrices  $\Sigma_G$  and  $\Sigma_P$ , respectively. When the dimensions of the multivariate trait are either genetically or phenotypically uncorrelated, this expression simplifies to the average SNP-heritability across dimensions. Similarly, when applying stratified LD Score regression (S-LDSC), one obtains enrichments for partitioned average heritability. We further show that  $\frac{1}{D}\text{trace}(\Sigma_G \Sigma_P^{-1})$  is an appropriate multivariate generalization of heritability since it satisfies the following four properties: 1) invariance to units of measurement, 2) coordinate-free, 3) linear in  $\Sigma_G$  and 4) maximized with a value of 1 when  $\Sigma_G = \Sigma_P$ .

Thus, for brain and face shape, we applied LDSC and S-LDSC using the published software (URL section) to corrected  $\chi^2$  statistics from GWAS of each brain or face segment. We used unmodified chi-squared values for the univariate traits analyzed (including indicated cases where we performed individual, univariate GWAS for each brain and face shape PC). While using unmodified chi-square values results in a small bias, we used unmodified statistics for consistency with previous studies. We limited S-LDSC analyses to traits with SNP-heritability Z-scores  $> 7$ , as in Finucaine et al.<sup>67</sup>

## Functional annotations for stratified LD score regression

We downloaded a range of publicly-available cell-type and sample-specific annotations representing open chromatin and/or active regulatory regions. Specifically, we obtained data on open chromatin (all ATAC-seq peaks) from brain organoids<sup>53</sup>, fetal brain tissue<sup>119</sup>, and CNCCs and derived chondrocytes<sup>52</sup>. ATAC-seq reads from Long et al. were mapped to hg19 with bowtie2<sup>120</sup> with default settings, and peaks were called using MACS2<sup>121</sup> with default settings. Annotations for active regulatory regions (based on a range of epigenomic marks) were obtained from CNCCs<sup>68</sup>, embryonic craniofacial tissues<sup>69</sup>, fetal and adult brain tissue<sup>70</sup>, and broad groupings of cell-types<sup>67</sup>. For CNCCs<sup>68</sup>, we combined all regions annotated as enhancers (weak, intermediate, strong) or promoters (weak and strong); For embryonic craniofacial tissues, we combined all regions with the following annotations from the 25-state chromHMM model: ‘Enh,’ ‘TxReg,’ ‘PromD1,’ ‘PromD2,’ ‘PromU,’ ‘TssA.’ For fetal and adult brain tissue, we combined all regions with the following annotations from the 15-state chromHMM model: ‘1\_TssA,’ ‘2\_TssAFlnk,’ ‘7\_Enh,’ ‘6\_EnhG.’ Each annotation was individually added to the baselineLD model from Finucaine et al. The resulting S-LDSC output (heritability fold-enrichment magnitude and significance, as well as coefficient Z-scores) are provided in Supplementary Data 2. When quantifying heritability enrichments with brain-face shared loci removed, we removed all SNPs within the same approximately independent LD block<sup>122</sup> as any of the 76 brain-face shared loci and re-computed LD scores.

## Quantifying sharing of signals between pairs of GWAS

To assess the extent to which genome-wide profiles of association were shared between a pair of GWAS, we computed a Spearman correlation between two vectors of LD-block organized association p-values. First, genome-wide SNPs were selected to overlap with the HapMap3 SNPs<sup>123</sup> and SNPs within the major histocompatibility complex region were removed. Second, we organized SNPs within 1,725 blocks in the human genome that can be treated as approximately independent in individuals of European ancestries<sup>122</sup>. For every LD block we computed the mean SNP  $-\log_{10}(\text{p-value})$ , and then computed a rank-based Spearman correlation using the averaged association value ( $n=1,725$ ) for each LD block. A standard error of the Spearman correlation was estimated using statistical resampling with 100 bootstrap cycles with replacement from the 1,725 LD blocks.

## URLs

UK biobank: <https://www.ukbiobank.ac.uk/about-biobank-uk/>

Human Connectome Project: <http://www.humanconnectomeproject.org/>

Adolescent Brain Cognitive Development (ABCD) study: <https://abcdstudy.org/about/>

SNPLIB: <https://github.com/jiarui-li/SNPLIB>

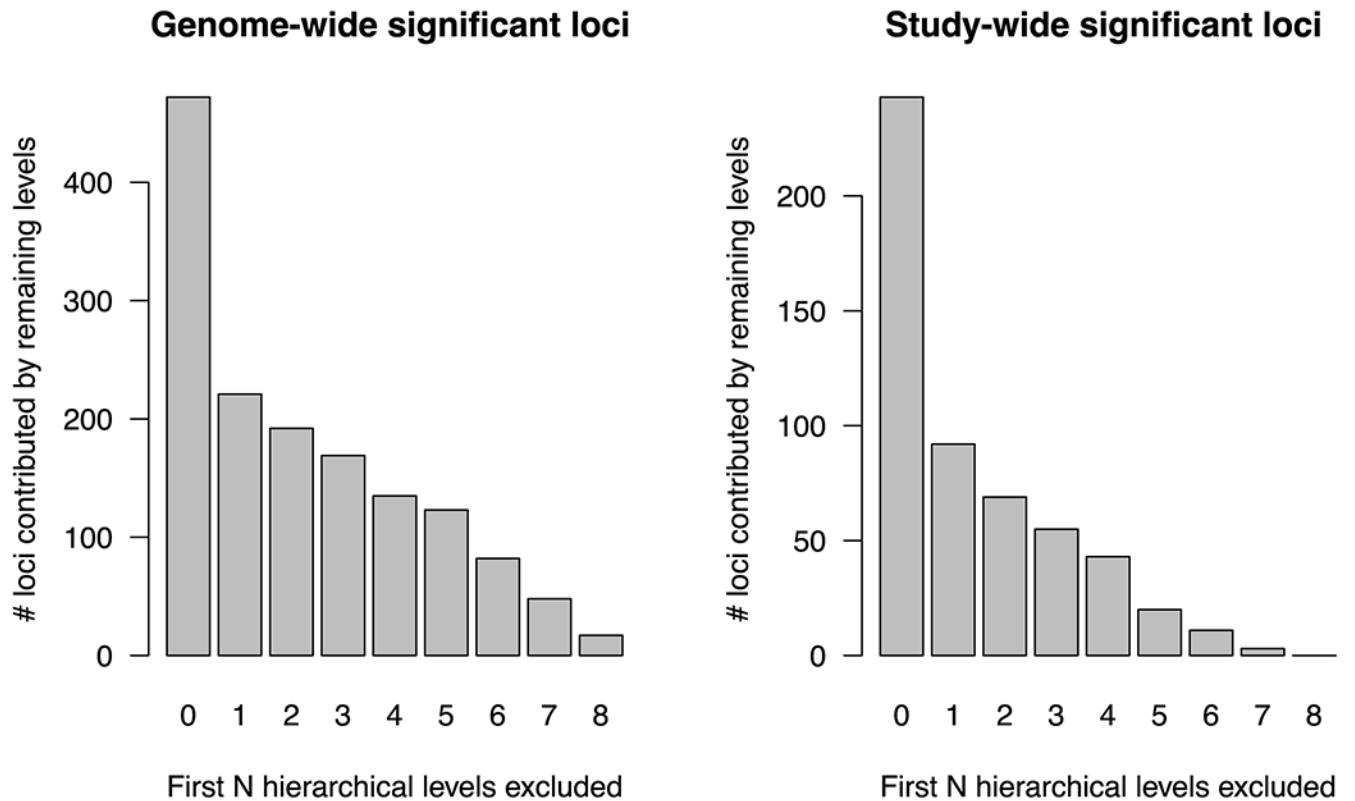
Freesurfer: <https://surfer.nmr.mgh.harvard.edu/>

Ciftify: <https://github.com/edickie/ciftify> and <https://www.nitrc.org/projects/cifti/Conte69>  
Atlas: [http://brainvis.wustl.edu/wiki/index.php//Caret:Atlases/Conte69\\_Atlas](http://brainvis.wustl.edu/wiki/index.php//Caret:Atlases/Conte69_Atlas)

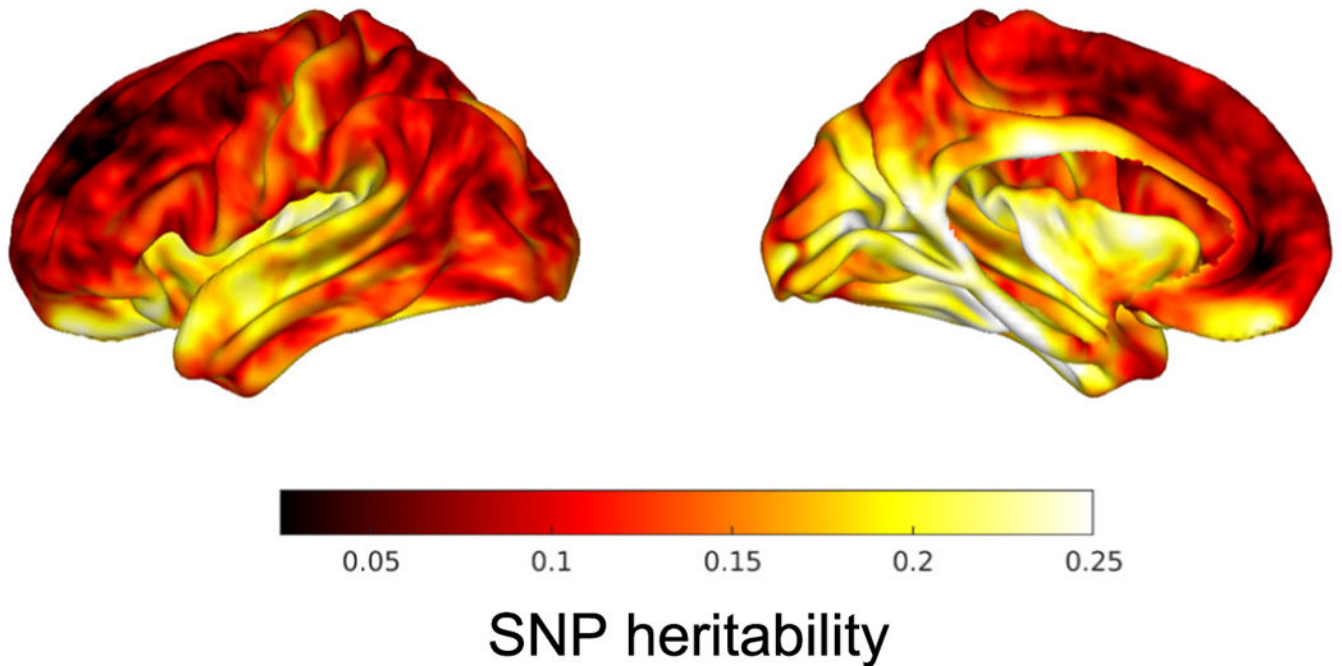
McCarthy Tools. <https://www.well.ox.ac.uk/~wrayner/tools/>

LDSC: <https://github.com/bulik/ldsc/wiki>

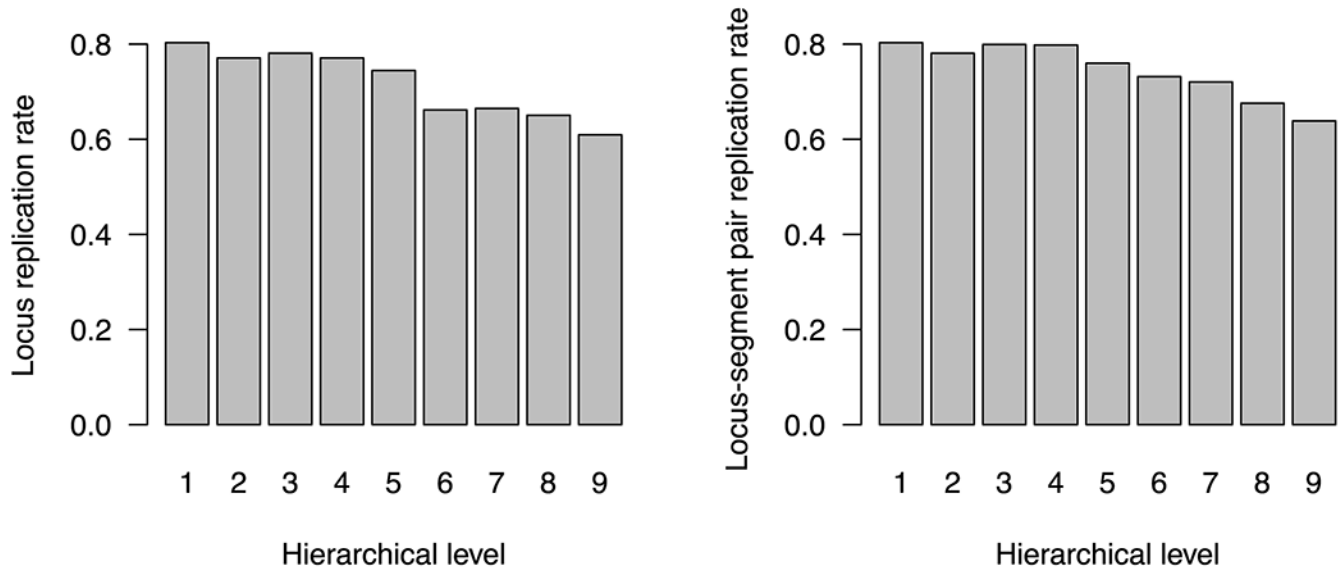
## Extended Data



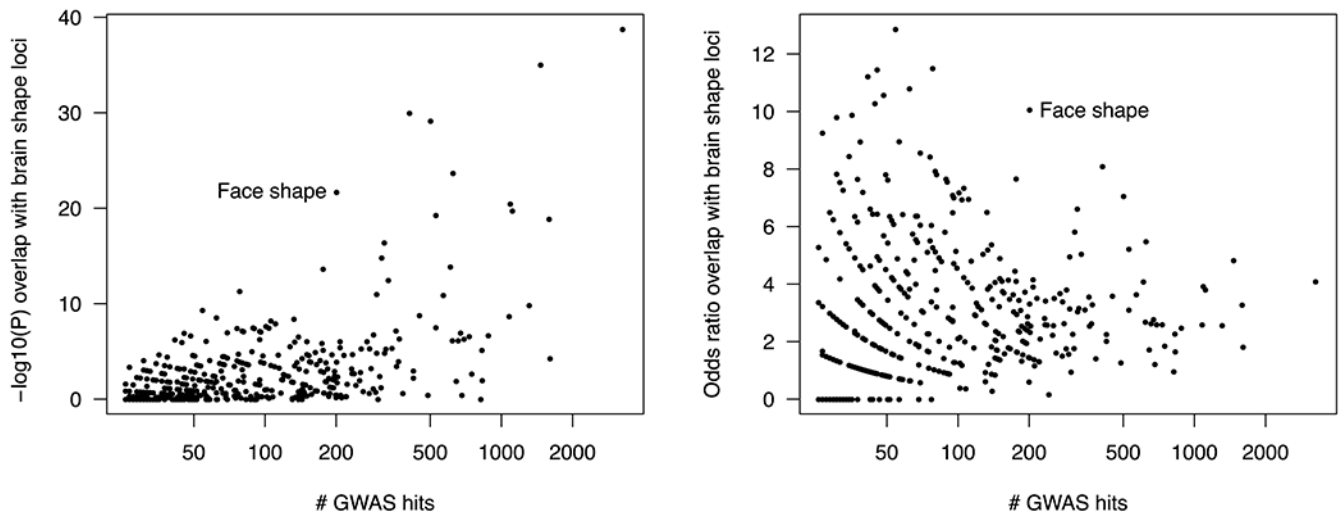
**Extended Data Fig. 1. Number of additional brain shape loci contributed by hierarchical levels.** For all genome-wide (left) or study-wide (right) significant associations, associations with all segments in hierarchical levels up to the indicated number were masked, and the number of remaining associations was assessed.



**Extended Data Fig. 2. Point-wise SNP heritability estimates across the mid-cortical surface.** Colors represent the total SNP heritability (computed by a linear mixed model approach, see Methods) at each point on the mid-cortical surface, represented by a set of three-dimensional coordinates in each individual.



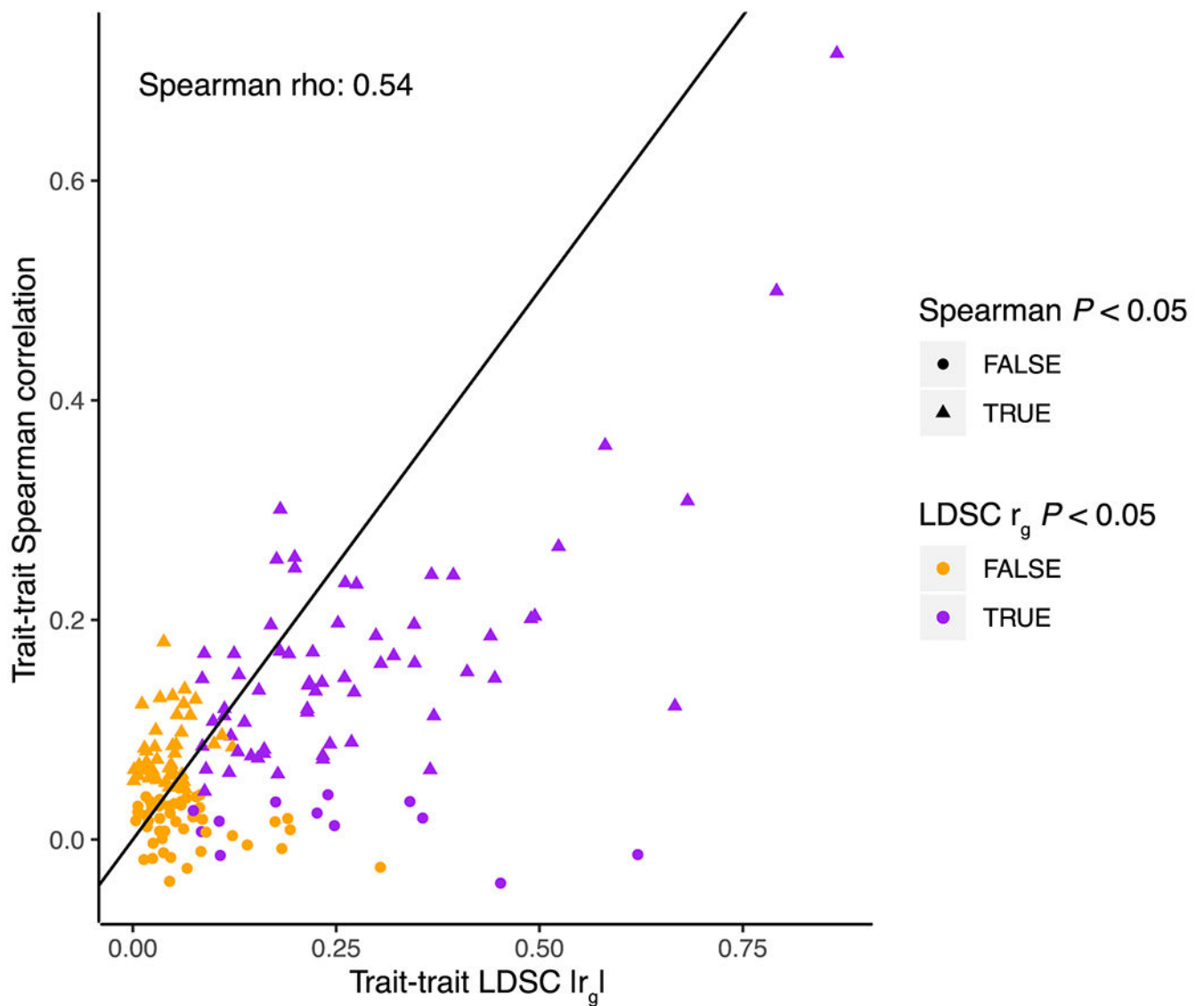
**Extended Data Fig. 3. Replication rates in the ABCD cohort by hierarchical level.** Only segments in the indicated hierarchical level were considered, and all loci (left) or locus-segment pairs (right) reaching genome-wide significance in those segments were tested for replication in the ABCD cohort at a 5% FDR.



**Extended Data Fig. 4. Overlap between genome-wide significant brain shape loci and genome-wide significant loci from 430 other studies.**

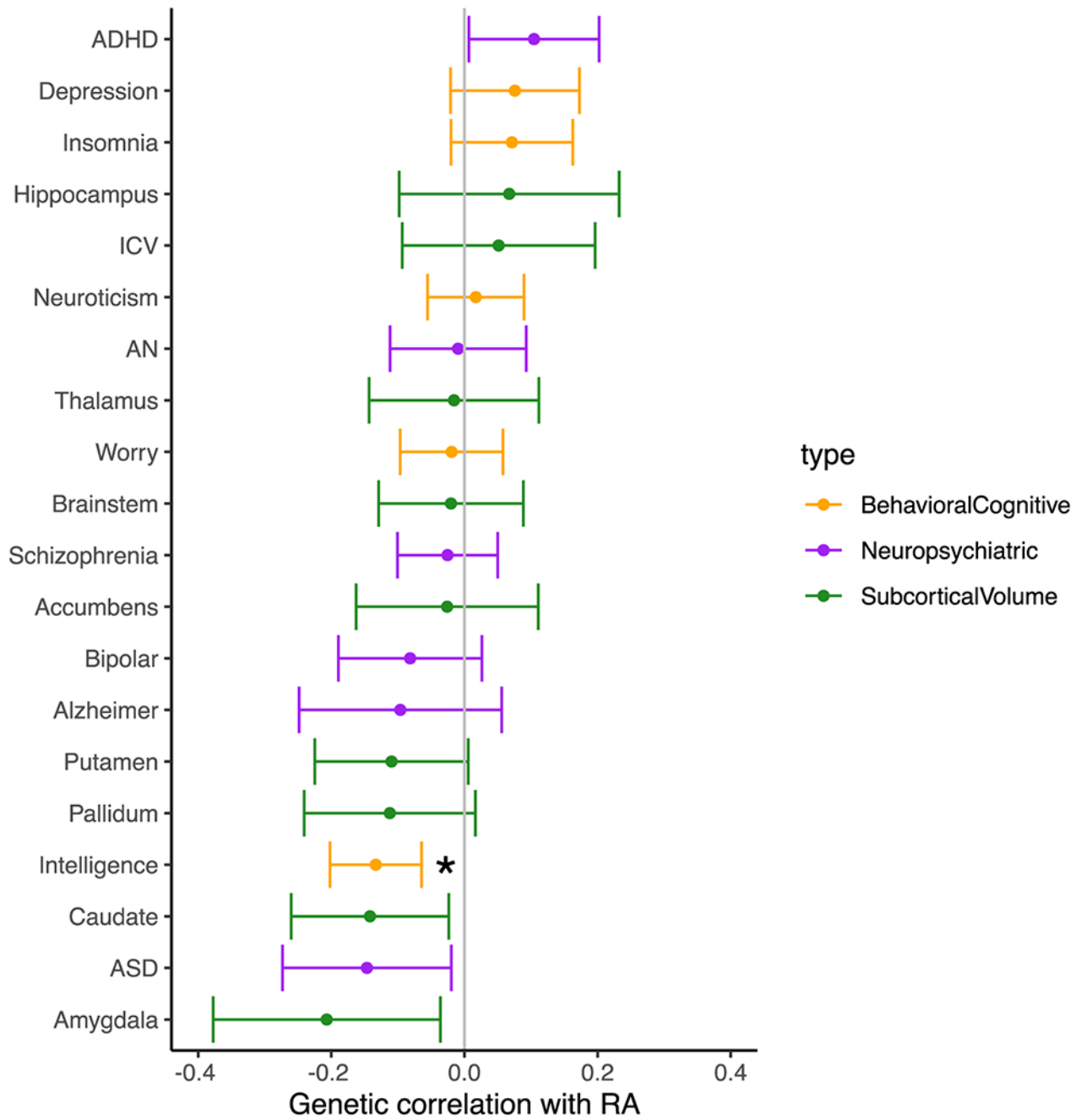
GWAS hits (number on x-axis) for other studies were obtained from the NCBI-EBI GWAS Catalog, and  $P$ -values (left, y-axis) and odds ratios (right, y-axis) for significance of overlap with regions in LD ( $> 0.2$ ) with brain shape loci were computed using bedtools' fisher function (see Methods). Note that relative to other traits with equivalent numbers of GWAS hits, face shape shows overlap with brain shape loci greater in both significance and magnitude.





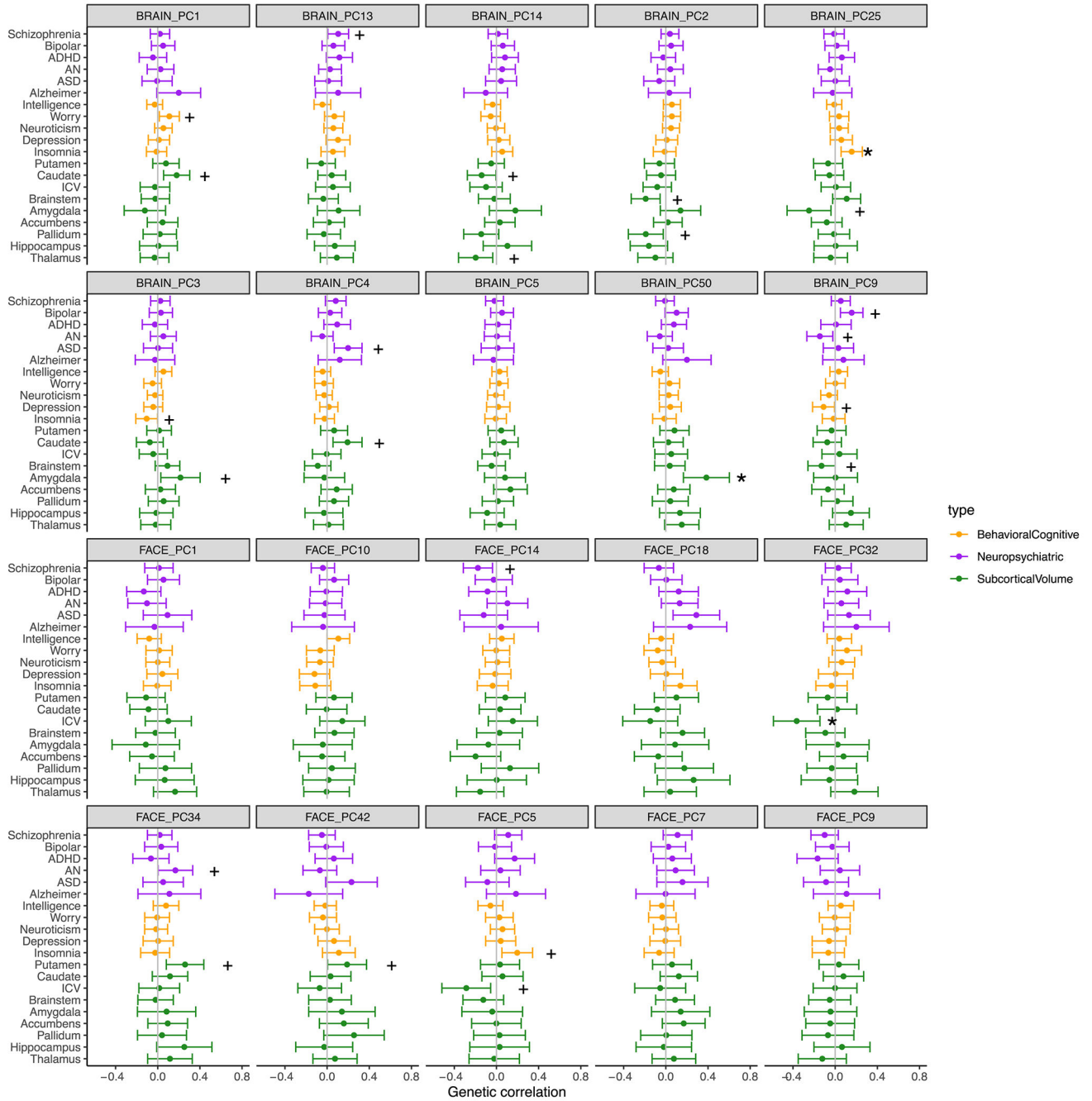
**Extended Data Fig. 5. Comparison of LDSC genetic correlations and Spearman correlation between pairs of univariate traits.**

Each point represents a pair of univariate traits (of all those considered in this study, see Methods), while the x- and y-axes indicate the absolute value of the LDSC-estimated genetic correlation and the estimated genome-wide sharing of effects by the Spearman correlation method. Point colors and shapes indicate significance ( $P < 0.05$ ) from LDSC or the Spearman correlation method, respectively. Exact p-values are provided in Supplementary Table 6.

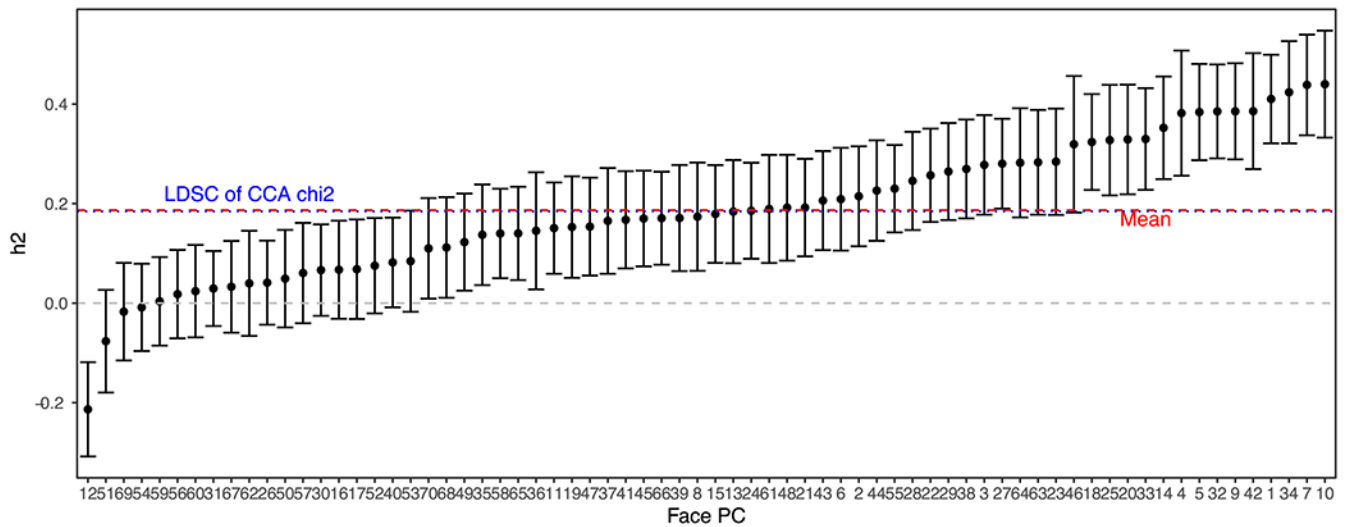


**Extended Data Fig. 6. Genetic correlations between RA (rheumatoid arthritis) and univariate brain-related traits.**

Points (center of error bars) represent estimated genetic correlations. Error bars represent 95% confidence intervals. \*, 5% FDR.

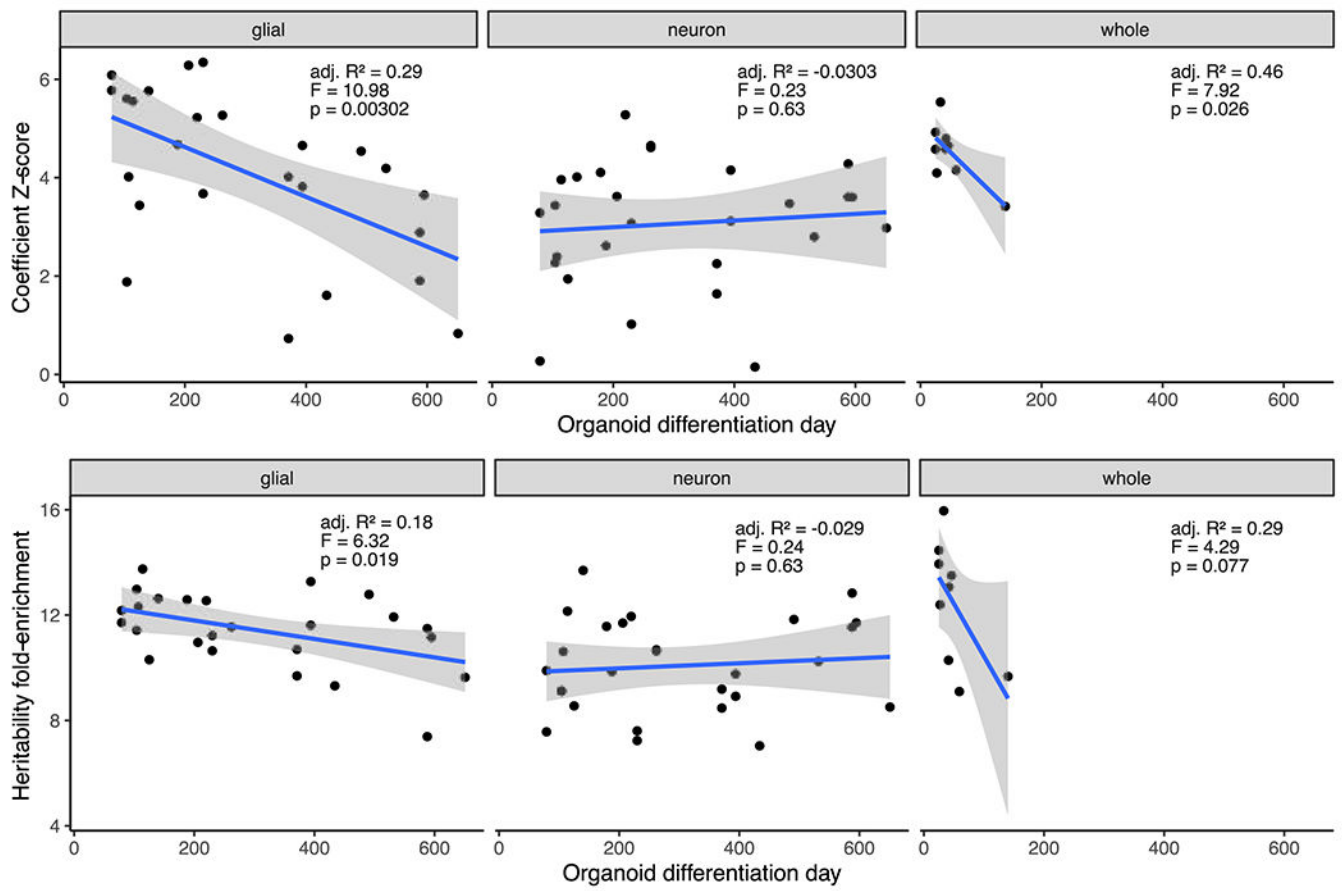


**Extended Data Fig. 7. Genetic correlations between the most heritable brain (top two rows) or face (bottom two rows) shape PCs and other traits.** Points (center of error bars) represent estimated genetic correlations ( $r_g$ ) between the top ten shape PCs (for segment 1, the full brain or face) with heritability z-score > 3 and each of the indicated univariate traits using LD score regression. Error bars represent 95% confidence intervals. \*, 5% FDR for indicated PC; +, 10% FDR.



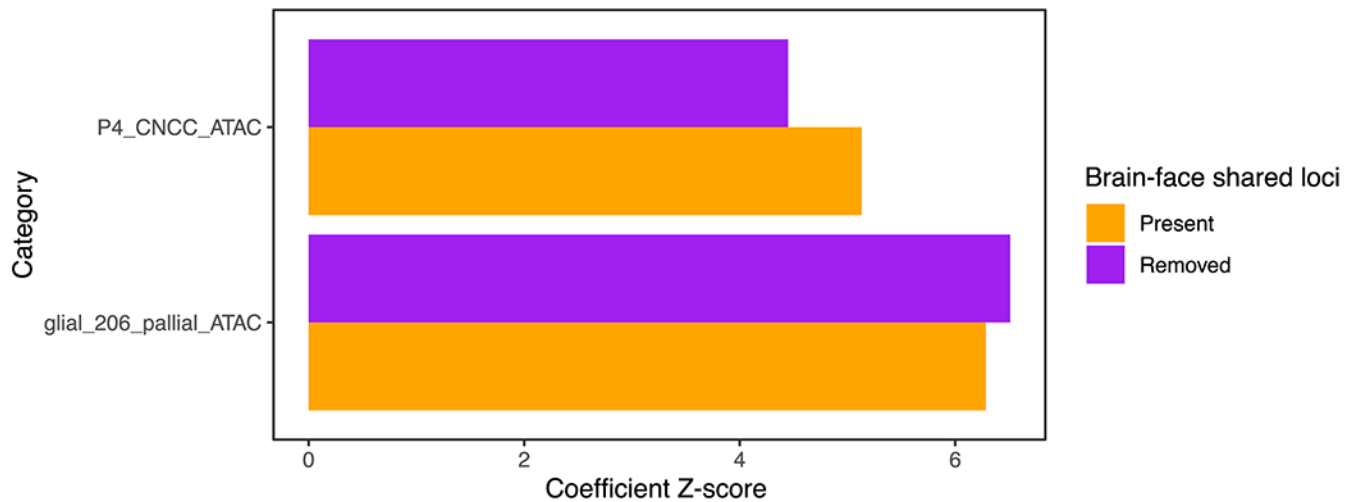
**Extended Data Fig. 8. SNP heritability of individual face shape PCs and multivariate face shape estimated by LDSC.**

Points (center of error bars) represent estimated SNP heritability of each PC. Error bars represent 95% confidence intervals. The red line represents the mean heritability of all 70 PCs, and the blue line indicates the heritability obtained by applying LDSC to corrected  $\chi^2$  statistics from the multivariate CCA GWAS using all 70 PCs.



**Extended Data Fig. 9. Partitioned heritability enrichments for brain shape with respect to stage- and cell-type-specific brain organoid open chromatin.**

S-LDSC coefficient Z-scores and heritability fold-enrichment for annotations corresponding to the indicated cell-type and differentiation day were computed as described in Methods. Regression lines represent the linear best fit with intercept and organoid differentiation day as dependent variable, and grey areas represent 95% confidence intervals. P-values are from a two-tailed F-test.



**Extended Data Fig. 10. Partitioned heritability enrichments for brain shape with respect to open chromatin in CNCCs or early glial organoid cells, with or without 76 brain-face shared loci.** S-LDSC Z-scores were calculated using full brain shape as the trait and the most enriched craniofacial (top) or brain organoid (bottom) ATAC-seq dataset as annotations. Z-scores were re-estimated (blue) after removing all SNPs in the same approximately independent LD block as one of the 76 brain-face shared loci (see Methods for details).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

JW was supported by the Howard Hughes Medical Institute, a Lorry Lokey endowed professorship and a Stinehart Reed award, SN was supported by a Helen Hay Whitney Fellowship. The KU Leuven research team and analyses were supported by the National Institutes of Health (1-R01-DE027023, 2-R01-DE027023), The Research Fund KU Leuven (BOF-C1, C14/15/081 & C14/20/081) and The Research Program of the Research Foundation - Flanders (Belgium) (FWO, G078518N). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI. JPS was supported by a National Institutes of Health training grant (5T32HG000044-23). JKP was supported by the National Institutes of Health (HG008140, HG009431). Pittsburgh personnel, data collection, and analyses were supported by the National Institute of Dental and Craniofacial Research (U01-DE020078, R01-DE016148, and R01-DE027023). Funding for genotyping by the National Human Genome Research Institute (X01-HG007821 and X01-HG007485) and funding for initial genomic data cleaning by the University of Washington provided by contract HHSN268201200008I from the National Institute for Dental and Craniofacial Research awarded to the Center for Inherited Disease Research (<https://www.cidr.jhmi.edu/>). JCT was supported by the National Institutes of Health (5R01-DA033431-07) and the National Science Foundation (1922598).

This research has been conducted [in part] using the UK Biobank Resource under application number 43193 (understanding the genetic architecture of human brain shape from MRI using global-to-local shape segmentations), and we are grateful for all the participants in that resource. This manuscript reflects the views of the authors and may not reflect the opinions or views of the UK Biobank funders and investigators.

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at <https://abcdstudy.org/federal->



[partners.html](#). A listing of participating sites and a complete listing of the study investigators can be found at [https://abcdstudy.org/Consortium\\_Members.pdf](https://abcdstudy.org/Consortium_Members.pdf). ABCD consortium investigators 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 consortium investigators.

## Data availability

All the data and detailed information for the UK Biobank, including genetic markers, covariates and MRI images are available to bona fide researchers via the UK Biobank data access process (see <http://www.ukbiobank.ac.uk/register-apply/>).

All the data and detailed information for the ABCD study, including genetic markers, covariates and MRI images are also available to bona fide researchers through the ABCD data depository (<https://nda.nih.gov/abcd/request-access>, controlled access due to highly identifiable facial scans and brain MRIs linked to genotype data)

Relevant data and materials from the facial GWAS study are available online (<https://doi.org/10.6084/m9.figshare.c.4667261>)<sup>124</sup>. The full facial GWAS summary statistics are available on the NHGRI-EBI GWAS catalog ([study accession GCST90007181](https://www.ebi.ac.uk/gwas/study-accession/GCST90007181)). Furthermore, relevant files generated from the face and brain GWAS summary statistics as input to (S-)LDSC regression and spearman correlations are available on FigShare, see Supplementary Table 7. The full brain GWAS summary statistics are available on the GWAS catalog, prepublished/unpublished studies section ([study accessions GCST90012880 - GCST90013164](https://www.ebi.ac.uk/gwas/study-accessions/GCST90012880-GCST90013164), one accession number per brain segment).

All relevant additional data related to this work are provided in the FigShare repository for this work (<https://doi.org/10.6084/m9.figshare.c.5089841.v1>). This includes additional figures, input files and updated implementations, listed in Supplementary Table 7.

## MAIN REFERENCES

1. Lui JH, Hansen DV & Kriegstein AR Development and evolution of the human neocortex. *Cell* 146, 18–36 (2011). [PubMed: 21729779]
2. Gu J & Kanai R What contributes to individual differences in brain structure? *Front. Hum. Neurosci* 8, 1–6 (2014). [PubMed: 24474914]
3. Strike L et al. Genetic complexity of cortical structure: Differences in genetic and environmental factors influencing cortical surface area and thickness. *Cereb. Cortex* 29, 952–962 (2019). [PubMed: 29377989]
4. Wen W et al. Distinct Genetic Influences on Cortical and Subcortical Brain Structures. *Sci. Rep* 6, 1–11 (2016). [PubMed: 28442746]
5. Grasby KL et al. The genetic architecture of the human cerebral cortex. *Science* (80-.). 367, (2020).
6. Fischl B *FreeSurfer*. *Neuroimage* 62, 774–781 (2012). [PubMed: 22248573]
7. Claes P et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet* 50, 414–423 (2018). [PubMed: 29459680]
8. White JD et al. Insights into the genetic architecture of the human face. *Nat. Genet* 53, 45–53 (2021). [PubMed: 33288918]
9. Marcucio R, Hallgrimsson B & Young NM Facial Morphogenesis. *Physical and Molecular Interactions Between the Brain and the Face*. *Current Topics in Developmental Biology* 115, (Elsevier Inc., 2015).
10. Stiles J & Jernigan TL The basics of brain development. *Neuropsychol. Rev* 20, 327–348 (2010). [PubMed: 21042938]

11. Simões-Costa M & Bronner ME Establishing neural crest identity: a gene regulatory recipe. *Development* 142, 242 LP–257 (2015). [PubMed: 25564621]
12. Theveneau E & Mayor R Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Dev. Biol* 366, 34–54 (2012). [PubMed: 22261150]
13. Santagati F & Rijli FM Cranial neural crest and the building of the vertebrate head. *Nat. Rev. Neurosci* 4, 806–818 (2003). [PubMed: 14523380]
14. Diewert VM A morphometric analysis of craniofacial growth and changes in spatial relations during secondary palatal development in human embryos and fetuses. *Am. J. Anat* 167, 495–522 (1983). [PubMed: 6624691]
15. Boughner JC et al. Short-faced mice and developmental interactions between the brain and the face. *J. Anat* 213, 646–662 (2008). [PubMed: 19094181]
16. Lieberman DE, Hallgrímsson B, Liu W, Parsons TE & Jamniczky HA Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. *J. Anat* 212, 720–735 (2008). [PubMed: 18510502]
17. Hallgrímsson B, Lieberman DE, Liu W, Ford-Hutchinson AF & Jirik FR Epigenetic interactions and the structure of phenotypic variation in the cranium. *Evol. Dev* 9, 76–91 (2007). [PubMed: 17227368]
18. Herring SW & Teng S Strain in the braincase and its sutures during function. *Am. J. Phys. Anthropol. Off. Publ. Am. Assoc. Phys. Anthropol* 112, 575–593 (2000).
19. Carver EA, Oram KF & Gridley T Craniosynostosis in Twist heterozygous mice: A model for Saethre-Chotzen syndrome. *Anat. Rec* 268, 90–92 (2002). [PubMed: 12221714]
20. Foppiano S, Hu D & Marcucio RS Signaling by bone morphogenetic proteins directs formation of an ectodermal signaling center that regulates craniofacial development. *Dev. Biol* 312, 103–114 (2007). [PubMed: 18028903]
21. Marcucio RS, Cordero DR, Hu D & Helms JA Molecular interactions coordinating the development of the forebrain and face. *Dev. Biol* 284, 48–61 (2005). [PubMed: 15979605]
22. Chong HJ et al. Signaling by SHH rescues facial defects following blockade in the brain. *Dev. Dyn* 241, 247–256 (2012). [PubMed: 22275045]
23. Kaucka M et al. Signals from the brain and olfactory epithelium control shaping of the mammalian nasal capsule cartilage. *Elife* 7, e34465 (2018). [PubMed: 29897331]
24. Le Douarin NM, Brito JM & Creuzet S Role of the neural crest in face and brain development. *Brain Res. Rev* 55, 237–247 (2007). [PubMed: 17765317]
25. Le Douarin NM, Couly G & Creuzet SE The neural crest is a powerful regulator of pre-otic brain development. *Dev. Biol* 366, 74–82 (2012). [PubMed: 22269168]
26. Aguiar DP, Sghari S & Creuzet S The facial neural crest controls fore- and midbrain patterning by regulating Foxg1 expression through Smad1 activity. *Development* 141, 2494 LP–2505 (2014). [PubMed: 24917504]
27. Plummer JT, Gordon AJ & Levitt P The genetic intersection of neurodevelopmental disorders and shared medical comorbidities - relations that translate from bench to bedside. *Front. Psychiatry* 7, 1–8 (2016). [PubMed: 26903886]
28. DeMyer W, Zeman W & Palmer CG THE FACE PREDICTS THE BRAIN: DIAGNOSTIC SIGNIFICANCE OF MEDIAN FACIAL ANOMALIES FOR HOLOPROSENCEPHALY (ARHINENCEPHALY). *Pediatrics* 34, 256 LP–263 (1964). [PubMed: 14211086]
29. Muenke M & Cohen MMJ Genetic approaches to understanding brain development: holoprosencephaly as a model. *Ment. Retard. Dev. Disabil. Res. Rev* 6, 15–21 (2000). [PubMed: 10899793]
30. Muenke M & Beachy PA Genetics of ventral forebrain development and holoprosencephaly. *Curr. Opin. Genet. Dev* 10, 262–269 (2000). [PubMed: 10826992]
31. Balk K & Biesecker LG The clinical atlas of Greig cephalopolysyndactyly syndrome. *Am. J. Med. Genet. A* 146A, 548–557 (2008). [PubMed: 18241058]
32. Desikan RS et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980 (2006). [PubMed: 16530430]

33. Destrieux C, Fischl B, Dale A & Halgren E Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15 (2010). [PubMed: 20547229]
34. Glasser MF et al. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178 (2016). [PubMed: 27437579]
35. Hibar DP et al. Common genetic variants influence human subcortical brain structures. *Nature* 520, 224–229 (2015). [PubMed: 25607358]
36. Adams HHH et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci* 19, 1569–1582 (2016). [PubMed: 27694991]
37. Hibar DP et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun* 8, 13624 (2017). [PubMed: 28098162]
38. Satizabal CL et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet* 51, 1624–1636 (2019). [PubMed: 31636452]
39. Zhao B et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet* 51, 1637–1644 (2019). [PubMed: 31676860]
40. Casey BJ et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci* 32, 43–54 (2018). [PubMed: 29567376]
41. Giedd JN et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci* 2, 861–863 (1999). [PubMed: 10491603]
42. Watanabe K, Taskesen E, van Bochoven A & Posthuma D Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun* 8, 1826 (2017). [PubMed: 29184056]
43. McLean CY et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol* 28, 495–501 (2010). [PubMed: 20436461]
44. Martin AR et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet* 51, 1560–1565 (2019). [PubMed: 31676867]
45. MacArthur J et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901 (2017). [PubMed: 27899670]
46. Uz E et al. Disruption of ALX1 Causes Extreme Microphthalmia and Severe Facial Clefting: Expanding the Spectrum of Autosomal-Recessive ALX-Related Frontonasal Dysplasia. *Am. J. Hum. Genet* 86, 789–796 (2010). [PubMed: 20451171]
47. Kayserili H et al. ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum. Mol. Genet* 18, 4357–4366 (2009). [PubMed: 19692347]
48. Howard TD et al. Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat. Genet* 15, 36–41 (1997). [PubMed: 8988166]
49. Ghouzzi V El et al. Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat. Genet* 15, 42–46 (2000).
50. Tassabehji M et al. Mutations in the PAX3 gene causing Waardenburg syndrome type 1 and type 2. *Nat. Genet* 3, 26–30 (1993). [PubMed: 8490648]
51. Zhao F et al. Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. *Am. J. Hum. Genet* 69, 695–703 (2001). [PubMed: 11505339]
52. Long HK et al. Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell* 27, 765–783.e14 (2020). [PubMed: 32991838]
53. Trevino AE et al. Chromatin accessibility dynamics in a model of human forebrain development. *Science* (80-.). 367, (2020).
54. Marcucio RS, Young NM, Hu D & Hallgrímsson B Mechanisms that underlie co-variation of the brain and face. *Genesis* 49, 177–189 (2011). [PubMed: 21381182]
55. Ming JE et al. Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Hum. Genet* 110, 297–301 (2002). [PubMed: 11941477]
56. Garcez RC, Le Douarin NM & Cruzet SE Combinatorial activity of Six1-2-4 genes in cephalic neural crest cells controls craniofacial and brain development. *Cell. Mol. Life Sci* 71, 2149–2164 (2014). [PubMed: 24061537]
57. Liu W et al. Threshold-specific requirements for Bmp4 in mandibular development. *Dev. Biol* 283, 282–293 (2005). [PubMed: 15936012]

58. Shimizu M et al. Probing the origin of matching functional jaws: roles of *Dlx5/6* in cranial neural crest cells. *Sci. Rep* 8, 14975 (2018). [PubMed: 30297736]
59. Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet* 47, 291–295 (2015). [PubMed: 25642630]
60. Pouget JG et al. Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk. *Hum. Mol. Genet* 28, 3498–3513 (2019). [PubMed: 31211845]
61. Kompus K et al. The role of the primary auditory cortex in the neural mechanism of auditory verbal hallucinations. *Frontiers in Human Neuroscience* 7, 144 (2013). [PubMed: 23630479]
62. Mørch-Johnsen L et al. Auditory Cortex Characteristics in Schizophrenia: Associations With Auditory Hallucinations. *Schizophr. Bull* 43, 75–83 (2016). [PubMed: 27605526]
63. Hoogman M et al. Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. *Am. J. Psychiatry* 176, 531–542 (2019). [PubMed: 31014101]
64. Assem M, Blank IA, Mineroff Z, Ademoglu A & Fedorenko E Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex* (2020). doi:10.1016/j.cortex.2020.06.013
65. Cai H, Zhu J & Yu Y Robust prediction of individual personality from brain functional connectome. *Soc. Cogn. Affect. Neurosci* 15, 359–369 (2020). [PubMed: 32248238]
66. Saviola F et al. Trait and state anxiety are mapped differently in the human brain. *Sci. Rep* 10, 11112 (2020). [PubMed: 32632158]
67. Finucane HK et al. Partitioning heritability by functional annotation using genomewide association summary statistics. *Nat. Genet* 47, 1228–1235 (2015). [PubMed: 26414678]
68. Prescott SL et al. Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimpanzee Neural Crest. *Cell* 163, 68–83 (2015). [PubMed: 26365491]
69. Wilderman A, VanOudenhove J, Kron J, Noonan JP & Cotney J High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep.* 23, 1581–1597 (2018). [PubMed: 29719267]
70. Kundaje A et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015). [PubMed: 25693563]
71. Gregory MD et al. Regional Variations in Brain Gyrfication Are Associated with General Cognitive Ability in Humans. *Curr. Biol* 26, 1301–1305 (2016). [PubMed: 27133866]
72. Nagai T et al. *Zic2* regulates the kinetics of neurulation. *Proc. Natl. Acad. Sci* 97, 1618 LP–1623 (2000). [PubMed: 10677508]
73. Elms P, Siggers P, Napper D, Greenfield A & Arkell R *Zic2* is required for neural crest formation and hindbrain patterning during mouse development. *Dev. Biol* 264, 391–406 (2003). [PubMed: 14651926]
74. Inoue T, Ota M, Mikoshiba K & Aruga J *Zic2* and *Zic3* synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. *Dev. Biol* 306, 669–684 (2007). [PubMed: 17490632]
75. Stolt CC et al. The *Sox9* transcription factor determines glial fate choice in the developing spinal cord. *Genes Dev.* 17, 1677–1689 (2003). [PubMed: 12842915]
76. Hu D & Marcucio RS Neural crest cells pattern the surface cephalic ectoderm during FEZ formation. *Dev. Dyn* 241, 732–740 (2012). [PubMed: 22411554]
77. Boys-Stones G, Elsner J, Ghersetti A, Hoyland R & Repath I Seeing the Face, Seeing the Soul: Polemon's Physiognomy from classical antiquity to medieval Islam. (OUP Oxford, 2007).
78. Zysk KG *Conjugal Love in India: Ratiâsâstra and Ratiramaðna: Text, Translation, and Notes.* 92, (Brill, 2002).
79. Mar TT *Face reading: The Chinese art of physiognomy.* (Dodd, Mead, 1974).
80. Gray RT *About face: German physiognomic thought from Lavater to Auschwitz.* (Wayne State University Press, 2004).
81. Collins AF *The enduring appeal of physiognomy: Physical appearance as a sign of temperament, character, and intelligence.* *Hist. Psychol* 2, 251–276 (1999).

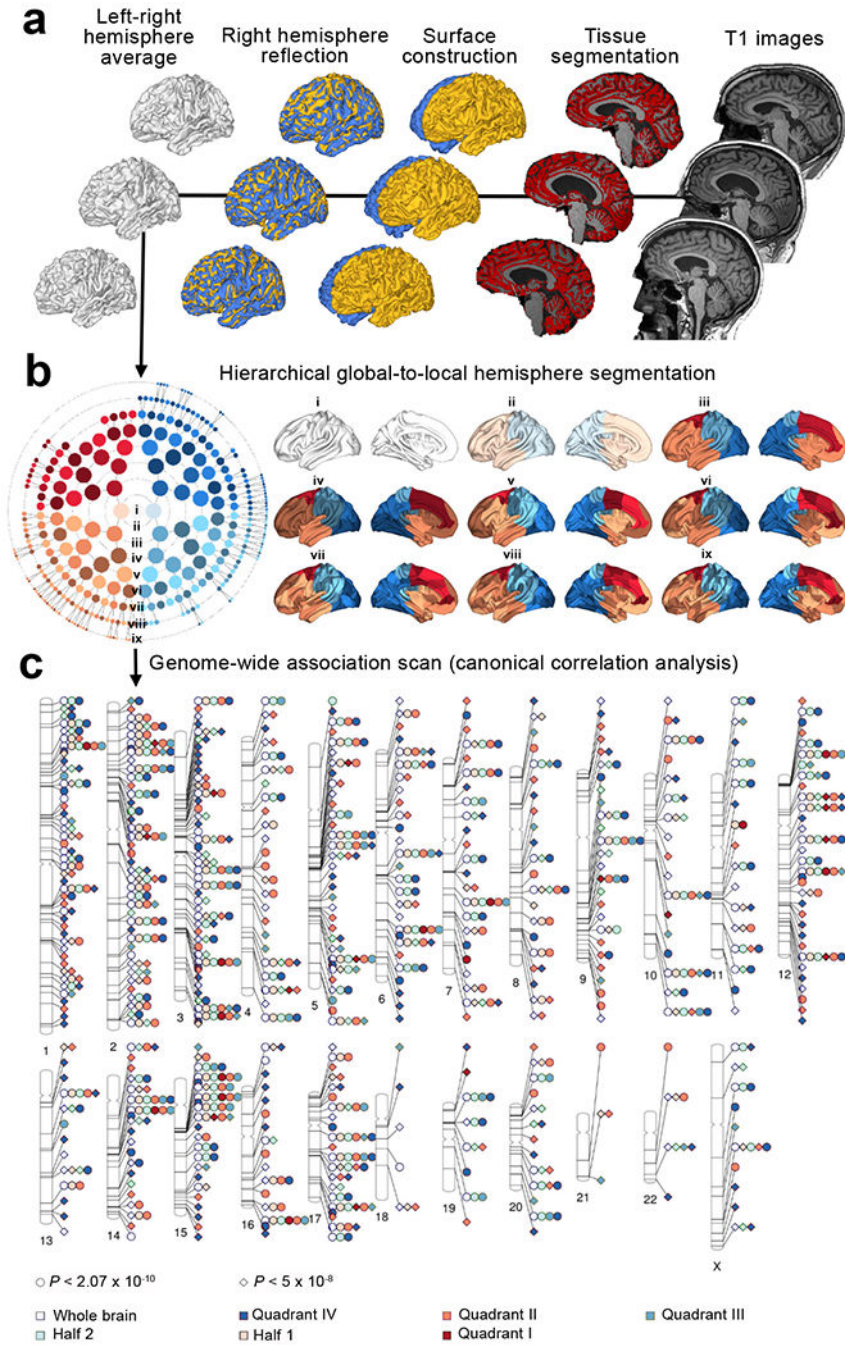
82. Haselhuhn MP, Ormiston ME & Wong EM Men's facial width-to-height ratio predicts aggression: A meta-analysis. *PLoS One* 10, 1–10 (2015).
83. Wu X & Zhang X Automated inference on criminality using face images. *arXiv Prepr. arXiv1611.04135* 4038–4052 (2016).
84. Wu X & Zhang X Responses to critiques on machine learning of criminality perceptions (Addendum of arXiv: 1611.04135). *arXiv Prepr. arXiv 1611.04135* (2016).

## METHODS REFERENCES

85. Miller KL et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci* 19, 1523–1536 (2016). [PubMed: 27643430]
86. Dale AM, Fischl B & Sereno MI Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194 (1999). [PubMed: 9931268]
87. Dickie EW et al. Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *Neuroimage* 197, 818–826 (2019). [PubMed: 31091476]
88. Winkler AM et al. Measuring and comparing brain cortical surface area and other areal quantities. *Neuroimage* 61, 1428–1443 (2012). [PubMed: 22446492]
89. Van Essen DC A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28, 635–662 (2005). [PubMed: 16172003]
90. Elliott LT et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210–216 (2018). [PubMed: 30305740]
91. Zelditch ML, Swiderski DL & Sheets HD Geometric morphometrics for biologists: a primer. (academic press, 2012).
92. Eller RJ, Janga SC & Walsh S Odyssey: a semi-automated pipeline for phasing, imputation, and analysis of genome-wide genetic data. *BMC Bioinformatics* 20, 364 (2019). [PubMed: 31253090]
93. Delaneau O, Zagury J-F, Robinson MR, Marchini JL & Dermitzakis ET Accurate, scalable and integrative haplotype estimation. *Nat. Commun* 10, 5436 (2019). [PubMed: 31780650]
94. Rubinacci S, Delaneau O & Marchini J Genotype imputation using the Positional Burrows Wheeler Transform. *bioRxiv* 797944 (2020). doi:10.1101/797944
95. McCarthy S et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet* 48, 1279–1283 (2016). [PubMed: 27548312]
96. Lam M et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet* 51, 1670–1678 (2019). [PubMed: 31740837]
97. Watson HJ et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet* 51, 1207–1214 (2019). [PubMed: 31308545]
98. Stahl EA et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet* 51, 793–803 (2019). [PubMed: 31043756]
99. Grove J et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet* 51, 431–444 (2019). [PubMed: 30804558]
100. Demontis D et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet* 51, 63–75 (2019). [PubMed: 30478444]
101. Jansen IE et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet* 51, 404–413 (2019). [PubMed: 30617256]
102. Jansen PR et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet* 51, 394–403 (2019). [PubMed: 30804565]
103. Nagel M et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet* 50, 920–927 (2018). [PubMed: 29942085]
104. Savage JE et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet* 50, 912–919 (2018). [PubMed: 29942086]
105. Bradfield JP et al. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLOS Genet.* 7, e1002293 (2011). [PubMed: 21980299]

106. Stahl EA et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet* 42, 508–514 (2010). [PubMed: 20453842]
107. López-Isac E et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun* 10, 4955 (2019). [PubMed: 31672989]
108. Jin Y et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet* 48, 1418–1424 (2016). [PubMed: 27723757]
109. Li J et al. Robust Genome-Wide Ancestry Inference for Heterogeneous Datasets and Ancestry Facial Imaging based on the 1000 Genomes Project. *bioRxiv* 549881 (2019). doi:10.1101/549881
110. Yang J, Lee SH, Goddard ME & Visscher PM GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet* 88, 76–82 (2011). [PubMed: 21167468]
111. White JD et al. Insights into the genetic architecture of the human face. *bioRxiv* 2020.05.12.090555 (2020). doi:10.1101/2020.05.12.090555
112. Robert P & Escoufier Y A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient. *J. R. Stat. Soc. Ser. C (Applied Stat)* 25, 257–265 (1976).
113. Pe'er I, Yelensky R, Altshuler D & Daly MJ Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol* 32, 381–385 (2008). [PubMed: 18348202]
114. Li J & Ji L Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*. 95, 221–227 (2005). [PubMed: 16077740]
115. Kanai M, Tanaka T & Okada Y Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet* 61, 861–866 (2016). [PubMed: 27305981]
116. Claes P et al. Modeling 3D Facial Shape from DNA. *PLoS Genet*. 10, (2014).
117. Benjamini Y & Yekutieli D The control of the false discovery rate in multiple testing under dependency. *Ann. Stat* 29, 1165–1188 (2001).
118. Bray NL, Pimentel H, Melsted P & Pachter L Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol* 34, 525–7 (2016). [PubMed: 27043002]
119. de la Torre-Ubieta L et al. The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* 172, 289–304.e18 (2018). [PubMed: 29307494]
120. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009). [PubMed: 19261174]
121. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008). [PubMed: 18798982]
122. Berisa T & Pickrell JK Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285 (2016). [PubMed: 26395773]
123. Altshuler DM et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010). [PubMed: 20811451]
124. Naqvi S Shared heritability of face and brain shape. *Figshare*. 10.6084/m9.figshare.c.5089841.v1 (2021).

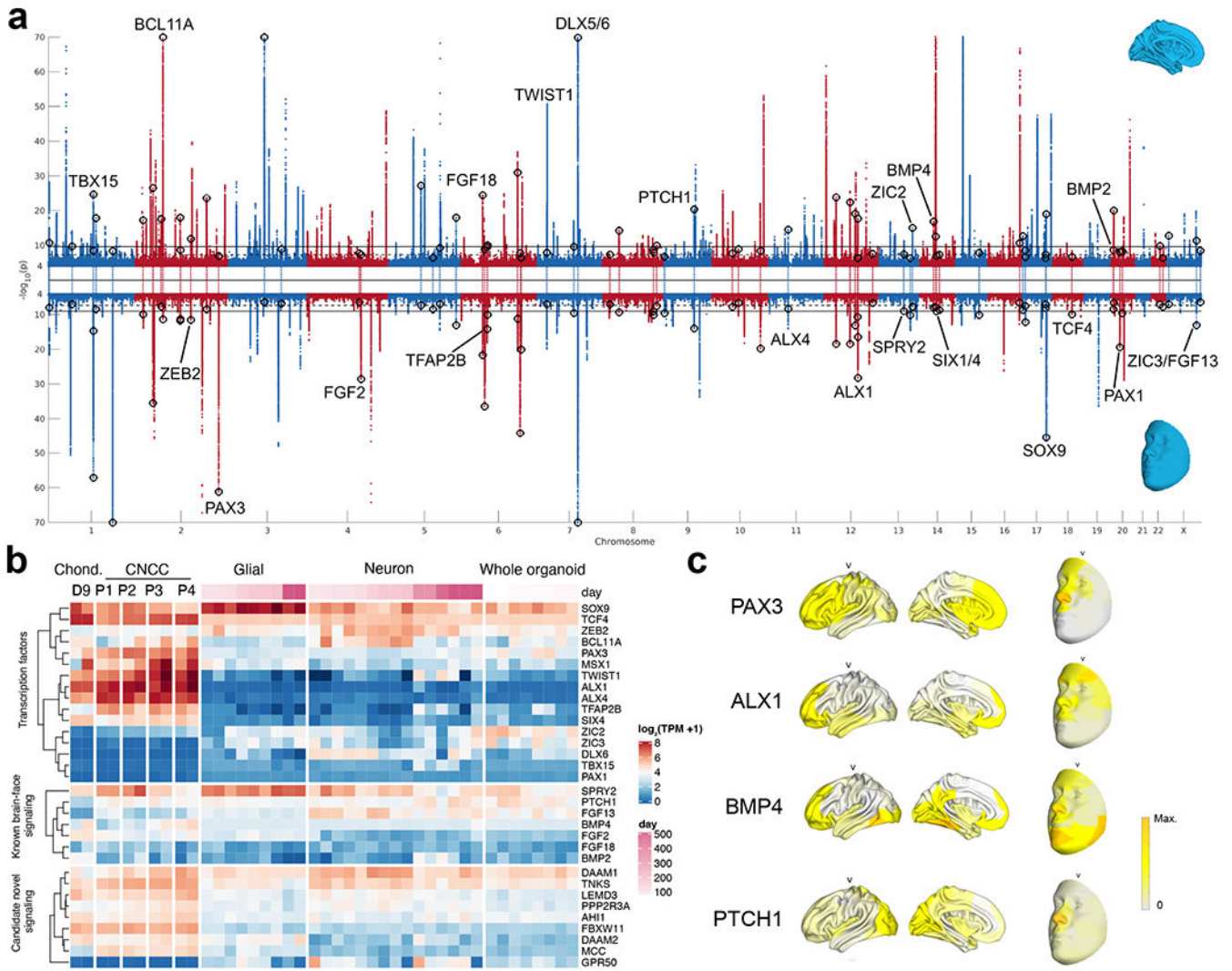




**Fig. 1: Multivariate genome-wide association study (GWAS) of brain shape.**

**a**, Upstream processing of UK Biobank (UKB) magnetic resonance imaging (MRI) images. **b**, in the polar dendrogram on the left, each concentric ring of filled circles corresponds to a hierarchical level (labeled i-ix) shown on the right, and the filled circle colors correspond to the respective segments in the same hierarchical level. **c**, Ideogram showing genomic locations and regional effects of 472 genome-wide significant loci for brain shape. Circles and diamonds represent associations passing the study-wide or genome-wide significance thresholds, colors represent broad regions of the brain with the indicated effects.





**Fig. 2: Loci affecting both brain and face shape.**  
**a**, Miami plot of GWAS for brain (top) and face (bottom) shape. For each SNP, p-values aggregated across all brain or face segments are plotted. All 76 loci reaching genome-wide ( $P < 5 \times 10^{-8}$ ) significance in one study and genome-wide suggestive ( $P < 5 \times 10^{-7}$ ) significance in the other are highlighted by unfilled circles. Right-tailed, one-sided P-values were computed based on canonical correlation analysis (CCA) chi-squared statistics; exact p-values are available in Supplementary Table 4. Loci near candidate genes highlighted in the text and in *b* and *c* are labeled, generally on the side where they show greater significance of association. *b*, Expression (in transcripts per million, TPM) of candidate genes near brain-face shared loci in cranial neural crest cells (CNCCs) of different passages, representing different stages of maturation, from early (P1) to late (P4) and their chondrocyte (Chond. D9) derivatives<sup>52</sup> (left), and three dimensional forebrain organoids at various stages of differentiation<sup>53</sup> (right), further sorted into glial or neuronal lineages or profiled as whole organoids. *c*, Regional phenotypic effects of four candidate loci, showing effects of linked single nucleotide polymorphisms (SNPs) on brain (left) or face (right) shape. Segments shown are of hierarchical level v,  $-\log_{10}(p\text{-values})$  are normalized to the

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

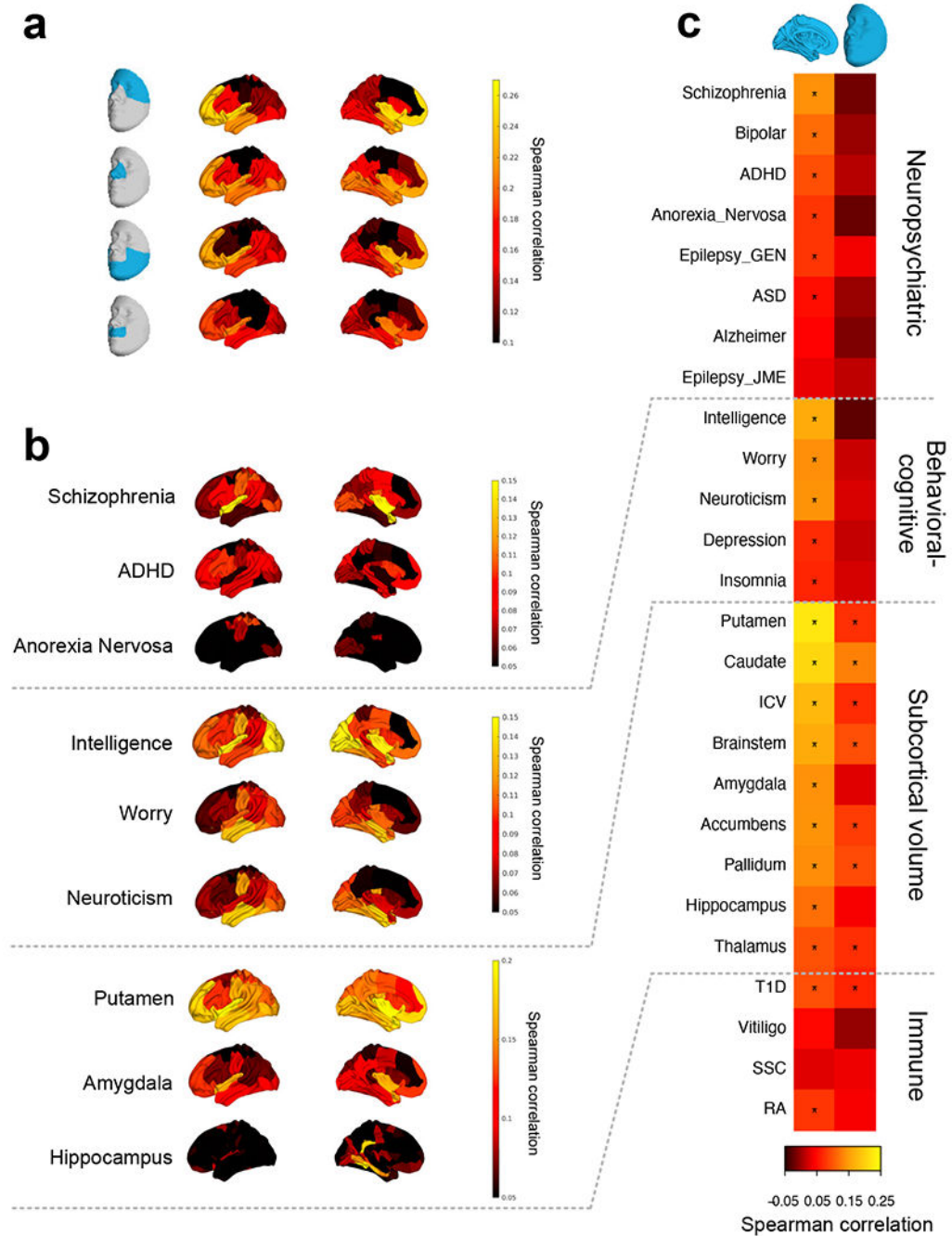
maximum at each locus. Full face and brain images from all 76 brain-face shared loci corresponding to all hierarchical levels can be found online (see Data Availability)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 3: Genome-wide sharing of signals with neuropsychiatric disorders and behavioral-cognitive traits.**

Genome-wide sharing of signals between any two given GWAS was assessed by Spearman correlation of linkage disequilibrium block-average SNP  $-\log_{10}(p\text{-values})$  (Methods). **a**, Spearman correlations between GWAS of indicated facial quadrants and brain segments. **b**, Spearman correlations between GWAS of selected neuropsychiatric disorders, behavioral-cognitive traits, subcortical volume measures and brain segments, or select immune traits. All brain segments in **a,b** are from hierarchical level v segmentation, with the exception of

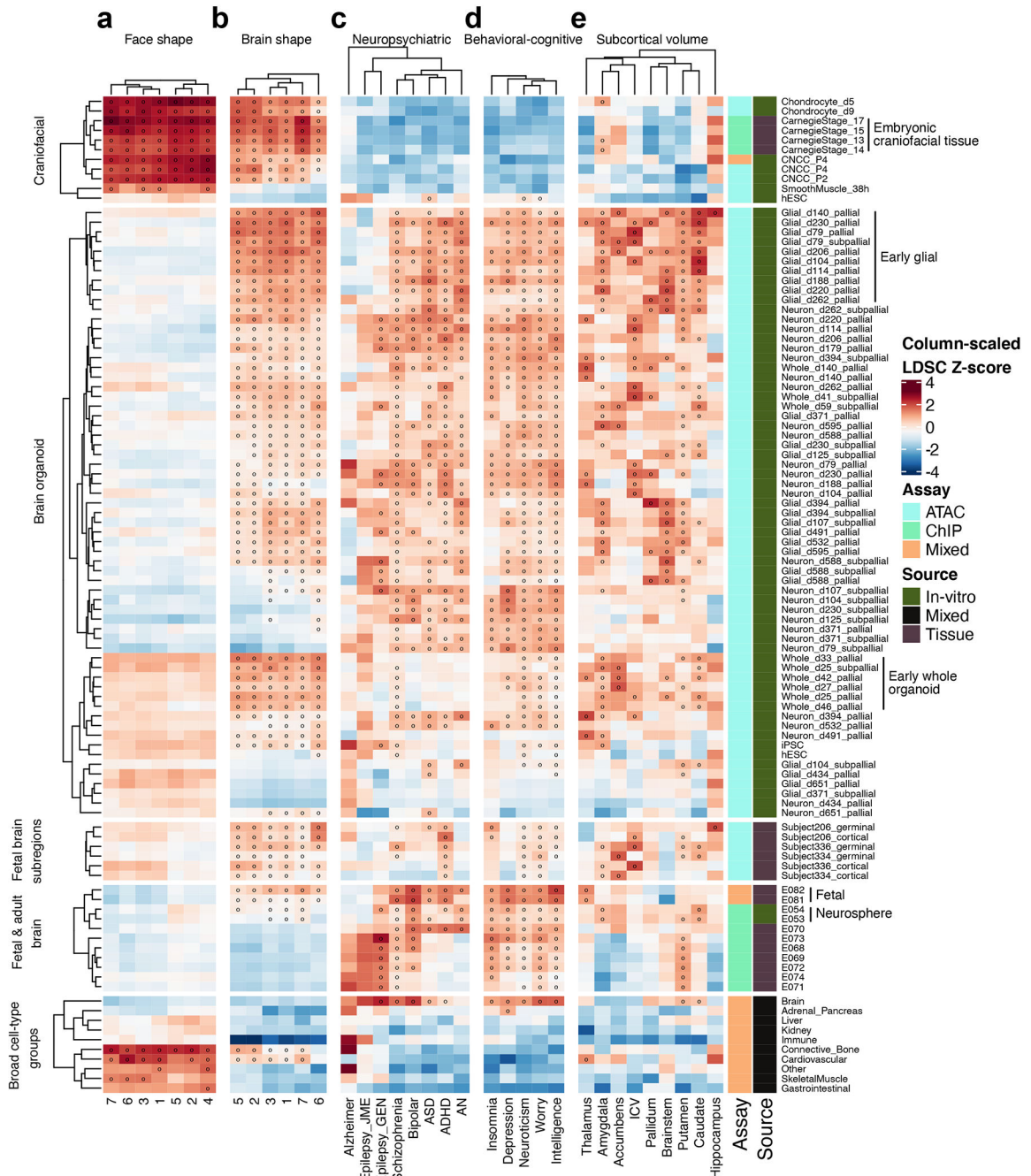
Hippocampus, where hierarchical level vi segmentation shows a strong correlation of shape of the hippocampal region with volume. **c**, Spearman correlations between shape effects on the full brain (left) or face (right) with the indicated traits. \* 5% false discovery rate (FDR) based on bootstrapped p-value (Methods). Images of brain-trait correlations at all six hierarchical levels can be found online (see Data Availability). Abbreviations: ADHD, Attention Deficit Hyperactivity Disorder ;GEN, generalized epilepsy; JME, juvenile myoclonic epilepsy; ICV, intracranial volume; T1D, type 1 diabetes; SSC, systemic sclerosis; RA, rheumatoid arthritis.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 4: Partitioned heritability enrichments based on cell-type-specific regulatory annotations.** Heritability enrichment Z-scores, as estimated by stratified linkage disequilibrium score regression (S-LDSC), of **a)** multivariate shape for the first 7 face segments, **b)** multivariate shape for the first 7 brain segments, excluding segment 4 which had low heritability, **c)** neuropsychiatric disorders, **d)** behavioral cognitive traits, and **e)** subcortical volume measures. Heritability enrichments were estimated for annotations based on open chromatin (based on Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)), regulatory regions (based on chromatin immune precipitation followed by sequencing

(ChIP-seq) of multiple histone modifications), or a combination of the two. Annotations for the indicated samples, representing in-vitro-derived cell-types, primary tissues, or a combination of both (see Methods for source papers), were added to the S-LDSC baseline model, and the resulting Z-score was scaled by column to visualize relative enrichments between traits. \* 5% FDR based on unscaled Z-scores. Trait abbreviations as in Figure 3, with AN representing anorexia nervosa.

**Table 1.**  
**Overlap between previous GWAS of brain surface areas or subcortical volumes with brain shape GWAS in this study.**

‘Subcortical combined’ refers to a combined set of loci from four studies of subcortical volume measures<sup>35–38</sup>.

Study	# loci tested	# lead SNP $P < 5 \times 10^{-8}$	# proxy SNP $P < 5 \times 10^{-8}$	% overlap
Subcortical combined <sup>31–34</sup>	65	15	18	27.6
Grasby et al <sup>5</sup>	301	195	236	78.4
Zhao et al <sup>35</sup>	494	212	273	55

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript