Abundant pleiotropy across neuroimaging modalities identified through a multivariate genome-wide association study

E.P. Tissink[1*], A.A. Shadrin[2], D. van der Meer[2,3], N. Parker[2], G. Hindley[2,4], D. Roelfs[2], O. Frei[2], C.C. Fan[5,6], M. Nagel[1], T. Nærland[7], M. Budisteanu[8,9], S. Djurovic[2,7,10], L. T. Westlye[2,7,11], M.P. van den Heuvel[1,12], D. Posthuma[1,12], T. Kaufmann[2,13], A.M. Dale[6,14,15] & O.A. Andreassen[2,7*]

[*] Corresponding authors: elleketissink@hotmail.com, ole.andreassen@medisin.uio.no

[1] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, 1081 HV Amsterdam, The Netherlands

[2] NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Building 48, Oslo, Norway

[3] School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands

[4] Psychosis Studies, Institute of Psychiatry, Psychology and Neurosciences, King's College London, 16 De Crespigny Park, London SE5 8AB, United Kingdom

[5] Laureate Institute for Brain Research, Tulsa, Oklahoma, USA.

[6] Department of Radiology, University of California San Diego, La Jolla, CA 92037, USA

[7] K.G. Jebsen Centre for Neurodevelopmental disorders, Division of Paediatric Medicine, Institute of Clinical Medicine, University of Oslo, Building 31, Oslo, Norway

[8] Prof. Dr. Alex Obregia Clinical Hospital of Psychiatry, Bucharest, Romania

[9] "Victor Babes" National Institute of Pathology, Bucharest, Romania

[10] Department of Medical Genetics, Oslo University Hospital, Oslo, Norway

[11] Department of Psychology, University of Oslo, Oslo, Norway

[12] Department of Child and Adolescent Psychology and Psychiatry, section Complex Trait Genetics, Amsterdam Neuroscience, VU University Medical Centre, Amsterdam, The Netherlands

[13] Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Germany

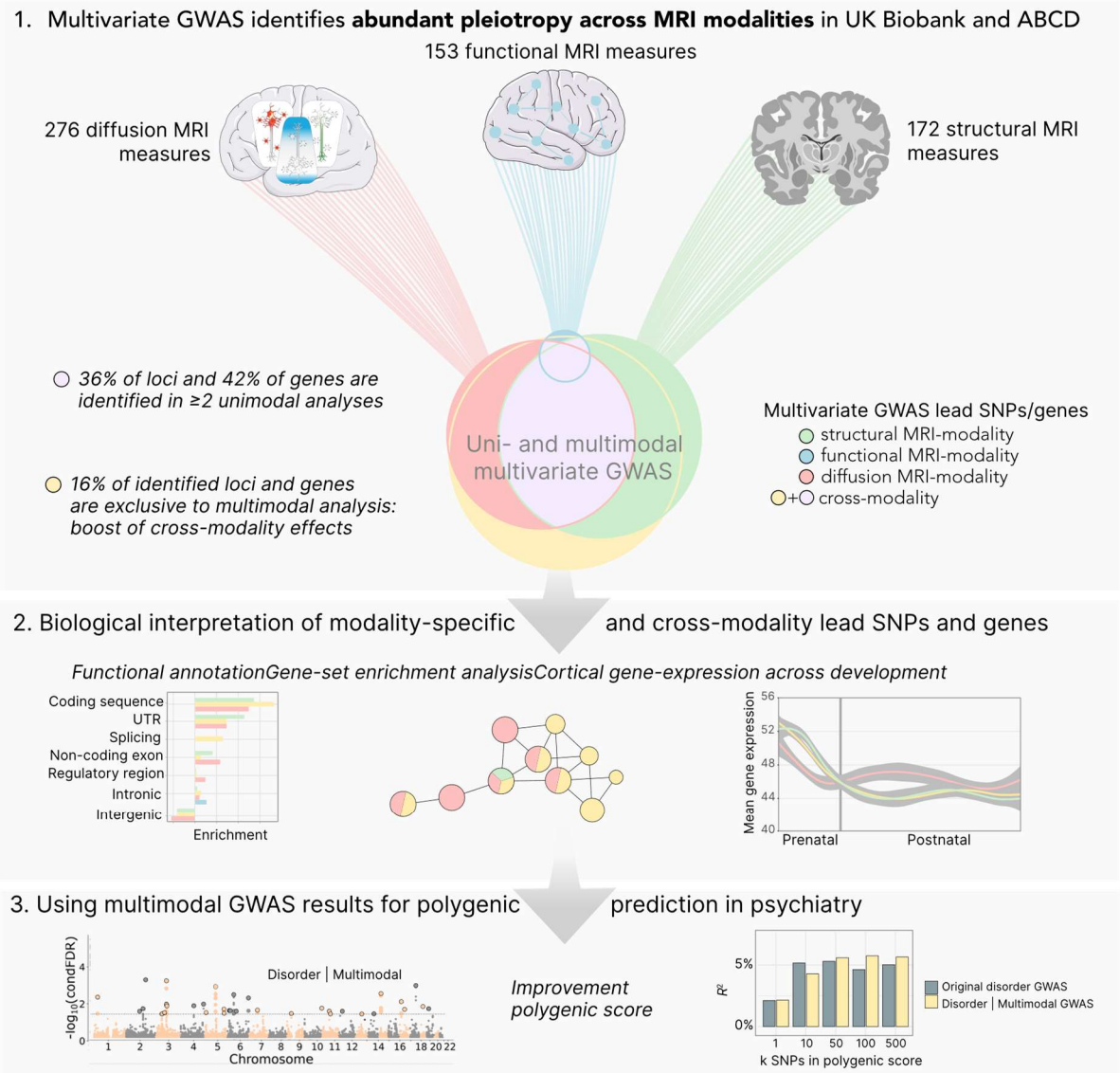
[14] Center for Multimodal Imaging and Genetics, University of California San Diego, La Jolla, CA 92037, USA

[15] Department of Neurosciences, University of California San Diego, La Jolla, CA 92037, USA

Abstract

Genetic pleiotropy is abundant across spatially distributed brain characteristics derived from one neuroimaging modality (e.g. structural, functional or diffusion MRI). A better understanding of pleiotropy across modalities could inform us on the integration of brain function, micro- and macrostructure. Here we show extensive genetic overlap across neuroimaging modalities at a locus and gene level in the UK Biobank (N=34,029) and ABCD Study (N=8,607). When jointly analysing phenotypes derived from structural, functional and diffusion MRI with the Multivariate Omnibus Statistical Test (MOSTest), we boost the discovery of loci and genes beyond previously identified effects for each modality individually. Cross-modality genes are involved in fundamental biological processes and predominantly expressed during prenatal brain development. We additionally boost prediction of psychiatric disorders by conditioning independent GWAS on our multimodal multivariate GWAS. These findings shed light on the shared genetic mechanisms underlying variation in brain morphology, functional connectivity, and tissue composition.

Graphical abstract



Introduction

The brain is our most complex organ, rapidly integrating information from many different sources[1], with strong genetic influences[2]. Studying genetic influences through genome-wide association studies (GWAS) has shown that most loci and genes show association with multiple traits[3], a phenomenon known as "statistical pleiotropy" (e.g. one gene influences multiple phenotypes directly, or indirectly via a causal pathway or common factor)[4]. Recently, numerous loci and genes with pleiotropic effects across brain characteristics derived from a single neuroimaging modality (e.g. structural, functional or diffusion MRI) have been discovered[5–9]. Yet the majority of pleiotropic loci act across rather than within phenotype domains[3], indicating that the genes associated within these loci may show pleiotropic effects across neuroimaging modalities, but the extent of this is underexplored. Investigating how genes influence a wide variety of brain imaging phenotypes may shed light on the mechanisms underlying alterations in brain morphology, activity, connectivity, and tissue composition that often co-occur in heritable psychiatric disorders[10,11].

Complex traits are affected by thousands of genetic variants scattered throughout the genome which makes pleiotropy between them inevitable[12]. In presence of such abundant pleiotropy, multivariate GWAS models have greater statistical power than univariate GWAS models[13]. This is because the cumulative evidence from different phenotypes leads to more sensitive detection of genetic associations and reduces the burden of multiple testing[7]. Previous studies have investigated either structural MRI-derived (sub)cortical volumes, surface area, and thickness[5–7], functional MRI-derived brain connectivity[8], or diffusion MRI-derived brain tissue composition[9] in a multivariate GWAS framework. The identified loci and genes inform us about the biological signal that is picked up by MRI, such as linking genetic effects of diffusion MRI to synaptic pruning, neuroinflammation, and axonal growth[9], structural MRI to neurogenesis and cell differentiation[5], and functional MRI to mental health[14] and psychiatric disorders[8]. Yet the modality specificity of these links remains unclear. Moreover, many loci are currently not discoverable in unimodal analyses: a multimodal multivariate GWAS could identify those loci that can improve our understanding of the interplay of biological processes contributing to brain structural organisation on multiple scales and reveal underpinnings of the structure-function relationship from a genetics viewpoint.

A large body of neuroimaging research established associations between psychiatric disorders and brain structure[15–20], function[21,22] and diffusion metrics[23,24]. However, previous studies were not able to detect non-null genetic correlations[25–27], possibly due to mixed effect directions[28] and differential polygenicity between neuroimaging phenotypes and psychiatric disorders[29]. Alternative methods that leverage the shared genetic signal between brain morphology and psychiatric disorders have been more fruitful[28] allowing improved prediction of disease liability[30]. Given that alterations of brain morphology, activity, connectivity, and tissue composition often co-occur in heritable psychiatric disorders[10,11], using multimodal GWAS associations may potentially further aid in leveraging the shared genetic signal between neuroimaging traits and psychiatric conditions.

88 Here, we demonstrate evidence of extensive pleiotropy across neuroimaging modalities using
89 583 structural (sMRI), resting-state functional (fMRI) or diffusion (dMRI) MRI-derived phenotypes in
90 the UK Biobank and the Adolescent Brain Cognitive Development (ABCD) Study®. We do so by
91 performing unimodal and multimodal multivariate GWAS with the Multivariate Omnibus Statistical
92 Test (MOSTest), which was designed to boost statistical power by capitalizing on the distributed nature
93 of genetic influences across imaging-derived phenotypes[7]. Specifically, we 1) overlap the previously
94 described multivariate effects for each modality and 2) combine all MRI-derived phenotypes into one
95 multimodal multivariate GWAS, obtaining an additional boost for discovery of cross-modality
96 pleiotropic loci and genes. Next, we functionally annotate single-modality (identified in only one
97 unimodal analysis) and cross-modality (identified in ≥ 2 unimodal analyses or unique to multimodal
98 analysis) loci and genes to describe the biological signal across MRI modalities. Last, we improve
99 polygenic prediction of bipolar disorder and ADHD after conditioning the GWASs of five major
100 psychiatric disorders on our multimodal, multivariate, genetic signal of brain morphology, functional
101 connectivity, and tissue composition. Thereby the current study provides insight into pleiotropic effects
102 across neuroimaging modalities and their relevance for understanding the neurobiology of the human
103 brain and mental health conditions.
104

105 ## Results

106 **Abundant pleiotropy of loci and genes across neuroimaging modalities**

107 We used data from three previous studies that applied MOSTest on single neuroimaging modality
108 phenotypes: 172 sMRI-derived brain morphology measures (68 regional surface area and 68 regional
109 thickness of the cerebral cortex, and 36 volumes of subcortical structures)[7], 153 fMRI-derived BOLD
110 signal and connectivity measures (17 network variances and 136 network correlations)[8], and 276 dMRI-
111 derived brain tissue composition principal components (65 restricted isotropic diffusion PCs, 124
112 restricted directional diffusion PCs, 87 normalized free water diffusion PCs)[9]. We investigated white
113 British UK Biobank samples (as derived from both self-declared ethnic background and genetic
114 principal component analysis) with quality-controlled genotypes and neuroimaging available (N_{sMRI} =
115 34,029, N_{fMRI}=31,023, N_{dMRI} = 30,106). We also included the ABCD cohort which had identical sMRI-
116 and dMRI-derived measures and similar fMRI-derived measures (see Methods). The ABCD sample
117 has a heterogeneous and admixed ancestral background. We created a replication sample within ABCD
118 similar to our UKB discovery sample by selecting the subset of European individuals (N_{sMRI} = 4,794,
119 N_{fMRI} = 4,132, N_{dMRI} = 4,418) assigned based on genetic ancestry factor[31] as defined in Methods. All
120 quality controlled ABCD samples (N_{sMRI} = 8,607, N_{fMRI} = 7,277, N_{dMRI} = 7,853) served as an additional
121 (ad)mixed ancestral replication sample. Next to genetic ancestry, ABCD samples were used to test the
122 generalizability of our results across age.

Heritability estimates obtained in UK Biobank with linkage disequilibrium (LD) Score Regression[32] ranged from median = 5.80% (IQR = 3.09%) for fMRI-derived network variances to 28.00% (IQR = 6.92%) for sMRI-derived subcortical volumes (Supplementary Fig. 1 and Supplementary Data 1), with dMRI-derived measures ranking in-between (median = 13.70%, IQR = 9.63%). Genetic and phenotypic correlations were generally similar ($\rho(r_g,r_p) = 0.53$, $p = 2.2\times10^{-16}$), and stronger within modalities than between modalities (Supplementary Fig. 2).

First, we combined all heritable (nominal $p < 0.05$) phenotypes derived from the same modality in multivariate GWAS analyses using MOSTest (Manhattan plots in Supplementary Fig. 4). This heritability filter was applied (as previously by Roelfs *et al*[8]) because including non-heritable phenotypes into MOSTest analyses have been shown to reduce statistical power[7]. MOSTest estimates the correlation between measures from univariate GWAS (on randomly permuted genotype data) and sums the squared decorrelated *z*-values across univariate GWAS summary statistics (from the original genotype data) to integrate the effects across the measures into a multivariate test statistic[7]. These unimodal multivariate analyses identified 590, 42 and 512 genome-wide significant loci associated with sMRI, fMRI and dMRI respectively ($p < 5\times10^{-8}/3$; Supplementary Data 2). These results are in line with earlier research (Supplementary Results 1). The number of genome-wide significant lead SNPs from UK Biobank that replicated at nominal significance ($p < 0.05$) in ABCD-based MOSTest summary statistics differed across modalities (EUR: 24.46% sMRI, 8.70% fMRI, 23.63% dMRI), and were higher for the larger sample with (ad)mixed ancestries (42.12% sMRI, 15.38% fMRI, 35.39% dMRI). Replication rates after correcting for the number of lead SNPs tested are provided in Supplementary Data 3. Applying MAGMA[33] gene-level analyses to unimodal multivariate summary statistics identified 1,620, 39 and 1,453 genome-wide significant genes ($p < ((0.05/18{,}877)/3 =) 8.83\times10^{-7}$) for sMRI, fMRI and dMRI respectively (Supplementary Data 4).

When overlapping the identified loci ($p < 5\times10^{-8}/3$) and genes ($p < 8.83\times10^{-7}$) from each modality (Methods), we observed 326 loci (36.18% of total) and 1,021 genes (41.79% of total) associated with at least two out of three modalities (Fig. 1a). This indicates pleiotropy across neuroimaging modalities both at the genome-wide significant locus and gene level. We replicated this pattern of overlap between sMRI and dMRI genome-wide significant loci and genes from UK Biobank in the ABCD cohort, though fMRI genome-wide significant loci did not overlap and fMRI genome-wide significant genes were not identified in ABCD (Supplementary Fig. 5, Supplementary Data 5-6). Replication patterns were similar across ancestries, with all of the sMRI and dMRI loci identified in the European ABCD sample overlapping with the loci identified within the (ad)mixed ancestry ABCD sample. Locus definition parameters (LD cut-off and window size) did not affect on the observed locus overlap (Fig. 1a), as was apparent from sensitivity analyses presented in Supplementary Fig. 3.

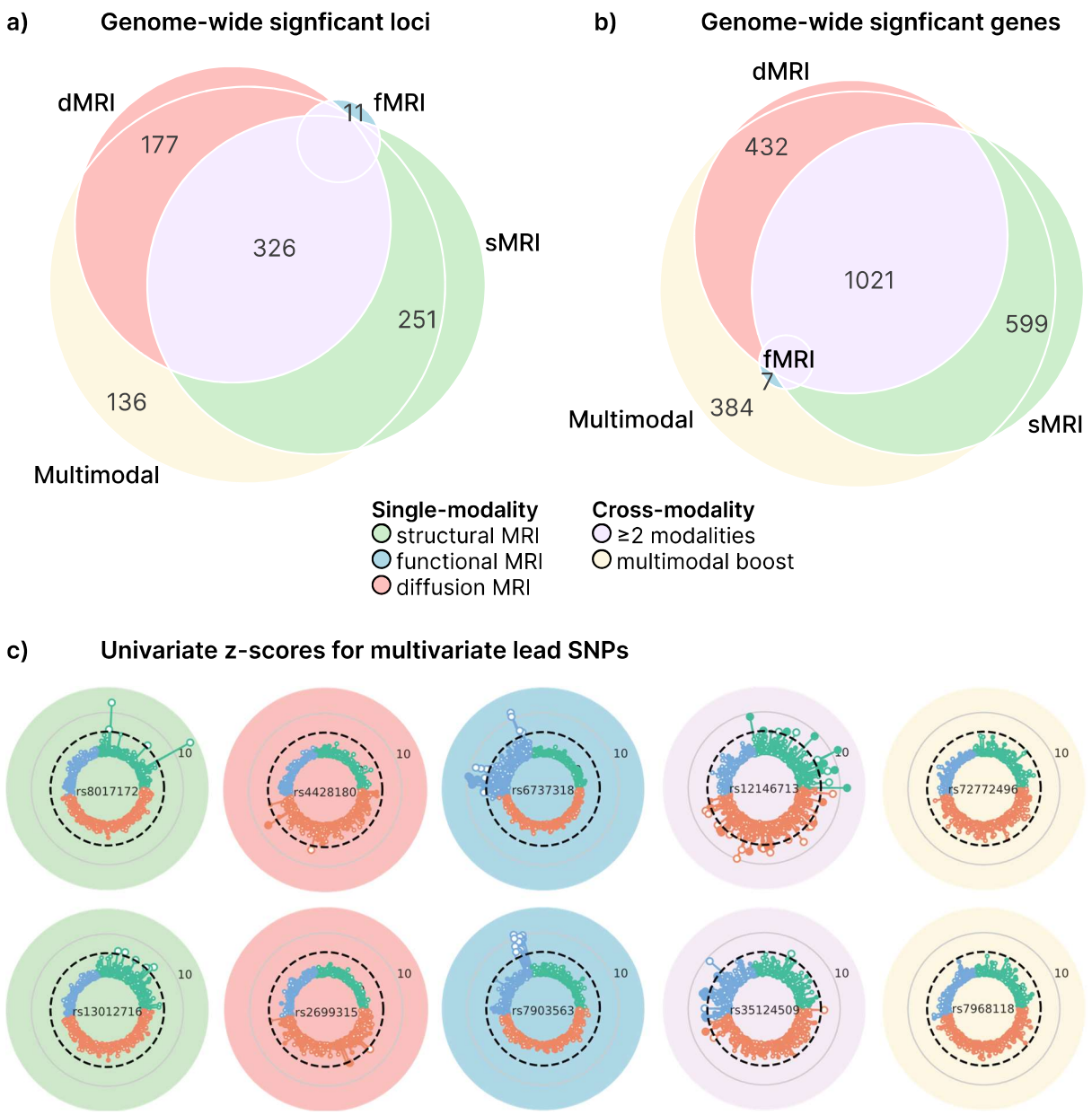


Figure 1. **Abundant pleiotropy of loci and genes across neuroimaging modalities.** Overlap of genome-wide significant a) loci and b) genes observed across neuroimaging modalities in single-modality and joint multimodal analyses ($p < 5\times10^{-8}/3$). When sMRI, fMRI and dMRI-derived phenotypes are jointly analysed in MOSTest (multimodal analysis), a boost in discovery of pleiotropic loci and genes is observed (yellow). This pattern partially replicates in the ABCD cohort (Supplementary Fig. 5). For 10 example lead SNPs from different parts of a) the univariate *z*-scores of all phenotypes used in this study (sMRI-derived phenotypes in green, dMRI-derived phenotypes in red, fMRI-derived phenotypes in blue) are plotted in c) as the distance from the centre of each circular plot. The dashed black line corresponds to $p = 5\times10^{-8}$ ($z = 5.45$) and *z*-scores that pass this threshold are depicted larger. Positive effect direction is shown as a filled circle and negative effect direction as a white circle.

169 **Multimodal GWAS boosts discovery of cross-modality loci and genes**

170 We next investigated whether combining all sMRI, fMRI, and dMRI-derived measures in one
171 multivariate analysis generated greater statistical power to identify significant pleiotropic loci and genes
172 which show sub-threshold associations in each unimodal multivariate analysis (Manhattan plot in
173 Supplementary Fig. 4). We therefore applied MOSTest across neuroimaging modalities, combining 583
174 phenotypes, identifying 794 genetic loci (Supplementary Data 2). The LD Score Regression intercept
175 (1.04, SE=0.04) indicated that the inflation in these multimodal MOSTest *p*-values (lambdaGC = 2.83)
176 was driven by polygenicity and not population stratification or cryptic relatedness. Replication rates at
177 nominal significance were comparable to previous MOSTest studies: EUR 26.45%, 37.55% (ad)mixed
178 ancestries (see Supplementary Data 3 for rates after correction for the number of lead SNPs). One-
179 hundred-thirty-six (15.09%) of these loci did not overlap with any of the loci identified in the unimodal
180 multivariate analyses, suggesting that MOSTest leveraged the shared genetic signal across imaging
181 modalities to boost the discovery of pleiotropic loci. Gene-based GWAS from MAGMA showed that
182 of the 2,242 genome-wide significant multimodal genes, 384 (15.72%) were not discovered for
183 unimodal gene-based GWAS (Supplementary Data 4). We used the ABCD cohort to investigate the
184 generalizability of our findings and observed a similar boost in multimodal discovery on a genome-
185 wide significant locus and gene level (Supplementary Fig. 5, Supplementary Data 5-6). Additionally,
186 81.82% of multimodal loci identified in the European ABCD sample were overlapping with the loci
187 identified in the (ad)mixed ancestry ABCD sample.

188 We examined the univariate associations underlying multivariate associations from different
189 parts of the Venn diagram (Fig. 1a). Fig. 1c shows the univariate *z*-scores across sMRI-, fMRI-, and
190 dMRI-derived phenotypes for ten lead SNPs representing different parts of the Venn diagram (Fig. 1a).
191 Cluster map of univariate *z*-scores for all multimodal MOSTest lead SNPs is displayed in
192 Supplementary Fig. 6. Alternatively, for every lead SNP identified in the MOSTest analyses presented
193 in the Venn diagram (Fig. 1a) we extracted the minimum univariate *p*-value across all analysed
194 phenotypes. A lead SNP with a relatively high minimum univariate *p*-value would indicate that the
195 signal should have been highly distributed across other measures for the variant to become genome-
196 wide significant in the multivariate analysis. Supplementary Fig. 7 shows that lead SNPs boosted by
197 the multimodal analysis had relatively high minimum univariate *p*-values ($0.05 > p > 5\times10^{-5}$) more
198 frequently (87%) than lead SNPs of loci identified in one unimodal analysis only (sMRI-modality 72%;
199 fMRI-modality 25%; dMRI-modality 78%), indicating that the discovery of these lead SNPs is driven
200 by pleiotropic signals across modalities.

201

202 **Comparing characteristics of single- and cross-modality loci and genes**

203 We investigated to what extent single-modality (identified in only one unimodal analysis) and cross-
204 modality (identified in ≥ 2 unimodal analyses or boosted by the multimodal analysis) loci and genes
205 (Supplementary Data 7-8) differ in their biological effects compared to other complex traits. To this

end, we annotated 251 sMRI-, 11 fMRI-, 177 dMRI-modality, and 462 cross-modality lead SNPs using ANNOVAR[34] (Supplementary Data 9). We used the annotations of 43,492 (unique) lead SNPs derived from 558 traits (with reasonable power, i.e. N > 50,000) across 24 trait domains from Watanabe et al.[3] as a reference. Fig. 3b demonstrates that cross-modality and single-modality lead SNPs were very similarly located in genomic regions as other complex traits[3,35]. Notably, cross-modality lead SNPs showed unique enrichment of exonic variants (6.58%, OR = 2.17, 95% CI = 1.43-3.17, $p = 2.91\times10^{-4}$). Watanabe et al. already demonstrated that the contribution of exonic SNPs increases from ~1% (SNPs associated to 1 trait domain) to ~5% (associated to ≥10 trait domain) with increasing pleiotropy[3]. The enrichment we observed therefore emphasizes the strong pleiotropic nature of the cross-modality lead SNPs.

Next, we tested whether any difference in results from gene-set enrichment analyses with Gene Ontology (GO) biological processes, cellular components and molecular functions could be observed using the 599 sMRI-, 7 fMRI-, 432 dMRI-modality, or 1,405 cross-modality genes (Fig. 3a). We identified four networks of gene-sets that consisted of both single-modality and cross-modality subsets of genes and showed significant enrichment after Bonferroni correction (Supplementary Data 10). These processes were mostly related to nervous system development and neuronal growth and differentiation. However, most gene-sets were significantly enriched of cross-modality genes only and highlight the genes' involvement in fundamental biological processes, such as cell cycle processes, cellular structure (chromatin, cytoskeleton, cell junction), and vesicle transport.

To investigate potential differential temporal patterns of cortical gene expression between single-modality and cross-modality genes, we investigated transcriptome data of post-mortem brain tissue (N=56) representing males and females of multiple ethnicities across the life span (Fig. 3c)[36]. The number of fMRI-modality genes in the data (n=5) was too low to generate a reliable expression pattern (Supplementary Fig. 8). Prenatal gene expression was high for both sMRI-, dMRI-modality and cross-modality genes compared to all genes in the data. In (early) childhood until adolescence, a differential dMRI-modality expression pattern was not apparent (compared to all genes in the data), whereas sMRI-modality and cross-modality genes were generally lower expressed. This matches a large body of research showing that pre- and postnatal cortical transcriptomes differ largely and pronounced prenatal expression matches the course of cortical development[37].

Given the predominant prenatal gene-expression pattern, we investigated whether single-modality and cross-modality genes were enriched for any cell-type identified by Bhaduri et al.[38] using single-cell RNA sequencing analyses of the fetal brain (Fig. 2d). We observed that cross-modality genes were enriched across all cell-types, confirming their importance in fetal brain tissue and suggesting their expression pattern is not cell-specific and more general in nature (Supplementary Data 11).

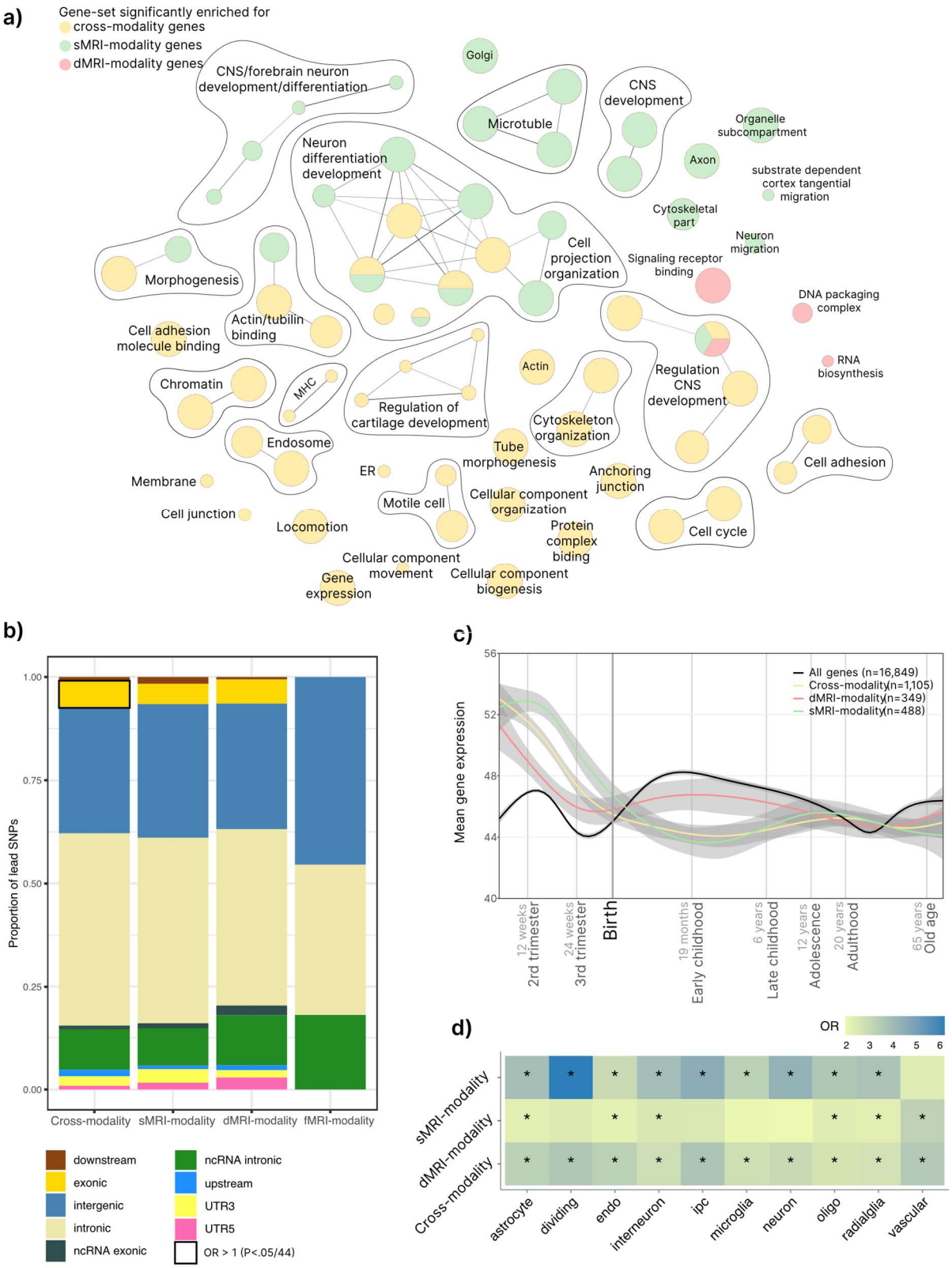


Figure 2. **Comparing properties of single-modality and cross-modality lead SNPs and genes.** a) Gene-ontology biological processes, molecular functions and cellular components that were (Bonferroni-corrected) significantly enriched for cross-modality, sMRI-modality and/or dMRI-modality genes (none of the GO terms tested showed enrichment for the 6 fMRI-modality genes). Node size reflects gene-set size, edges reflect pathway similarity scores (Methods). b) Functional consequences of single-modality and cross-modality lead SNPs as

245 annotated with ANNOVAR[34]. When the null hypothesis (OR=1) could be rejected after Bonferroni correction (p
246 $< 1.14\times10^{-3}$), the solid line indicates significant enrichment of the annotation. Annotations of 43,492 (unique)
247 lead SNPs derived from 558 traits across 24 trait domains from Watanabe et al. were used as reference for Fisher
248 Exact Test (Supplementary Data 9). ncRNA = non-coding RNA. c) Mean-normalized expression (y-axis) of
249 cross-modality and single-modality genes over developmental timepoints (x-axis; log10 scale). Gray shading
250 indicates 95% confidence intervals. The mean-normalized expression of fMRI-modality genes is displayed in
251 Supplementary Fig. 8, since the number of genes (n=5) was low and therefore created an unreliable pattern. d)
252 Cell-type enrichment analysis for fetal brain tissue from Bhaduri et al.[38]. Bonferroni corrected significant results
253 (p < (0.05 / 30 =) 1.67×10^{-3}) are indicated by an asterisk (*). ipc = intermediate progenitor cells, oligo =
254 oligodendrocytes.
255
256 **Leveraging shared genetics with psychiatric disorders for polygenic prediction**
257 Alterations in brain morphology, connectivity, and tissue composition often co-occur in heritable
258 psychiatric disorders[11], suggesting that our multimodal, multivariate genetic signal may have relevance
259 for the genetics of psychiatric disorders. It is possible to boost polygenic prediction by re-ranking the
260 test-statistics from a given GWAS based on a genetically related secondary GWAS[39]. We therefore
261 conditioned five major psychiatric disorder GWAS summary statistics on our multimodal MOSTest
262 summary statistics using the conditional false discovery rate approach (condFDR; Supplementary
263 Results)[40]. We included schizophrenia (SCZ)[41], bipolar disorder (BD)[42], major depressive disorder
264 (MDD)[43,44], attention-deficit hyperactivity disorder (ADHD)[45], and autism spectrum disorder (ASD)[46]
265 (Suppelementary Data 12). The rationale behind condFDR is that, in the presence of cross-trait
266 enrichment, a variant with strong associations with both traits is more likely to represent a true
267 association[30]. These disorder-multimodal condFDR summary statistics were used to construct
268 polygenic scores for the disorders in independent samples using a pleiotropy-informed polygenic
269 scoring method[30] (pleioPGS; see Methods). For that purpose, we constructed polygenic scores with
270 PRSice2 in two independent clinical datasets (TOP and MoBa, see Supplementary Data 12) based on
271 the top 10-150,000 original disorder GWAS ordered SNPs and condFDR-based ordered SNPs (Fig. 3).
272 For each disorder a likelihood ratio test was then applied to test if a model with both pleioPGS and
273 original PGS provides significantly better prediction compared to the model with only one PGS
274 included.
275 We observed a 1.24-fold increase in variance explained (R^2 = 6.22%, $p = 2.28\times10^{-15}$) for BD in
276 the TOP sample (N_{case} = 463; $N_{control}$ = 1,073) using the pleioPGS with 100,000 SNPs, compared with
277 the original GWAS-based PGS (R^2 = 5.02%, $p = 2.28\times10^{-15}$). The likelihood ratio test that compared the
278 model fit before and after adding BD pleioPGS to the original BD PGS showed that pleioPGS brings a
279 significant improvement in model fit (χ^2 (df = 2) = 13.35, $p = 1.26\times10^{-3}$). Also for ADHD, the pleioPGS
280 (R^2 = 1.47%, $p = 4.57\times10^{-38}$) performed better than the original PGS (R^2 = 1.20%, $p = 2.62\times10^{-31}$),
281 contributing significantly to explaining the variation in ADHD liability (χ^2 (df = 2) = 33.54, $p = 5.21\times10^{-}$

[8]) in MoBa cases (N = 2,216) and controls (N = 206,644). In MDD cases (N_{case} = 135) and controls ($N_{control}$ = 1,073), the pleioPGS (R^2 = 3.98%, $p = 2.16\times10^{-6}$) explained more variance than the original GWAS-based PGS (R^2 = 3.22%, $p = 1.92\times10^{-5}$) as well. Adding MDD pleioPGS to the original MDD PGS did not improve the model fit (χ^2 (df = 2) = 5.30, p = 0.07), but pleioPGS remained significant in the combined model while the original MDD PGS became insignificant (original PGS t = 0.90, p = 0.37; pleioPGS t = 2.29, p = 0.02). A similar observation was made in the case of ASD: although the higher variance explained by pleioPGS (R^2 = 1.47%, $p = 4.57\times10^{-38}$) compared to the original PGS (R^2 = 1.47%, $p = 4.57\times10^{-38}$) in BUPGEN (N_{case} = 331, $N_{control}$ = 1073) did not cause a significantly increased model fit (χ^2 (df = 2) = 5.93, $p = 5.17\times10^{-2}$), pleioPGS remained significant in the combined model while the original PGS became insignificant (original PGS t = -0.70, p = 0.48; pleioPGS t = 2.42, $p = 15.37\times10^{-3}$). When predicting SCZ in TOP (N_{case} = 735; $N_{control}$ = 1,073), the addition of the original GWAS-based PGS provided significant improvement (χ^2 (df = 2) = 23.17, $p = 9.30\times10^{-6}$).

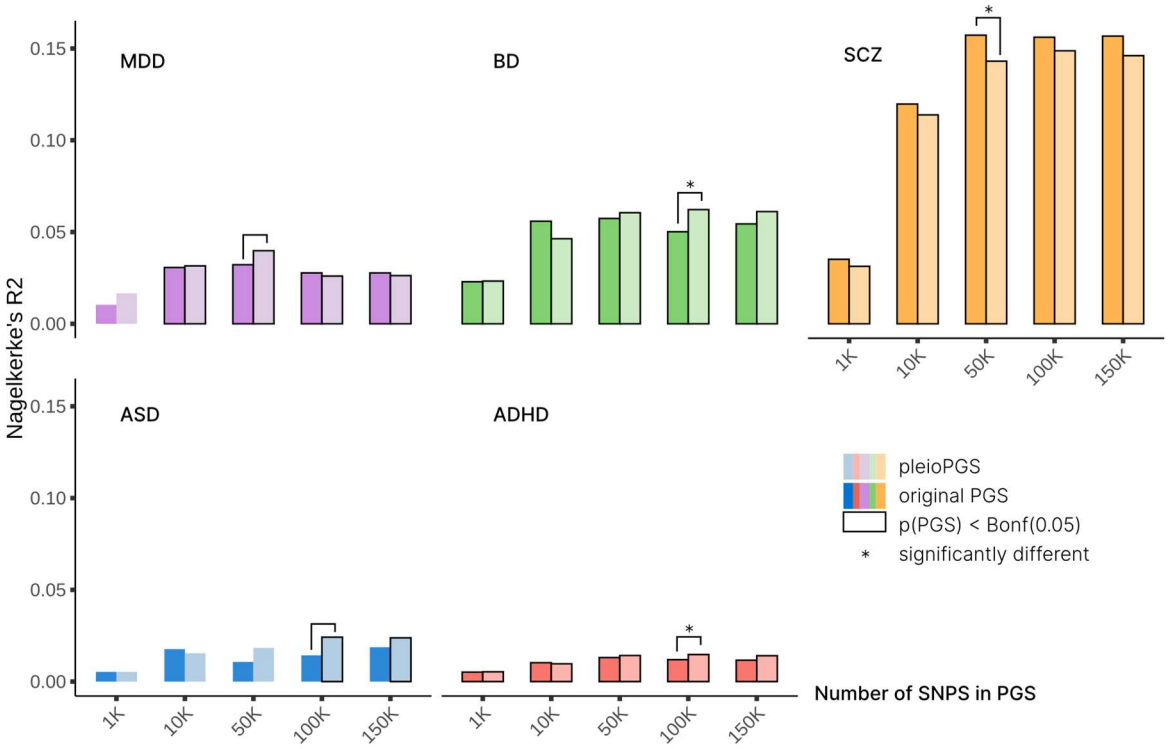


Figure 3. **Phenotypic variance explained for five psychiatric disorders explained by polygenic scores based on original disorder GWAS summary statistics (original PGS) and conditional disorder-multimodal GWAS summary statistics (pleioPGS).** Note that PGS using top 10 and top 100 SNPs had too low R^2 values to be visible and are therefore not plotted, but can be found in Supplementary Data 17. Independent target samples were used for prediction (see Methods). A likelihood ratio test was applied to compare the model which includes only the best PGS (highest R^2) with the model including both PGS for each disorder.

Discussion

Our findings demonstrated that many loci and genes show pleiotropic effects across brain characteristics derived from three distinct MRI modalities. We found evidence of extensive pleiotropy across structural, functional and diffusion MRI from genetic overlap (loci and genes being associated with at least two out of three modalities), and a boost in discovery of loci (n=136) and genes (n=384) when all MRI-derived phenotypes are jointly analysed using MOSTest. The results in the ABCD cohort show generalizability of structural and diffusion MRI pleiotropic loci from adulthood to childhood, and from European ancestry to (ad)mixed ancestries. Genes with cross-modality pleiotropic effects were highly prenatally expressed, though not cell-type specific and mostly enriched in gene-sets of general biological functions. Moreover, we showed how these results can be leveraged to improve polygenic prediction of bipolar disorder and ADHD.

The human brain is a highly complex and inter-connected structure for which "the whole is more than the sum of its parts"[50]. This complexity emerges from tight interplay between different units and processes, where disturbance of any part may change the state of the whole system. With this in mind, pervasive effects of genetic variants on brain-related traits are inevitable, resulting in abundant pleiotropy not only across phenotypically linked traits such as brain morphology measures[37], but also in widespread genetic overlap between distinct aspects of brain functioning such as personality and cognition[39]. Our results implicate that the multivariate genetic signals of structural, functional, or diffusion MRI are not only composed of pleiotropic effects within modalities as previously shown[5–9], but also of a component that is shared across brain traits measured with different MRI modalities. This provides a new conceptual insight into the integration of human brain functional connectivity (fMRI), microstructure (dMRI) and macrostructure (sMRI) and highlights the importance of characterizing patterns of specificity and pleiotropy to improve our understanding of molecular neurobiological mechanisms. It is expected from theoretical analysis that rare variants would follow this pattern of pleiotropy under assumption of the infinitesimal model[51], which would be an interesting direction of future research given the recent availability of multivariate tools for gene-based rare-variant association studies[52,53].

We observed genetic overlap across all three modalities in our discovery sample from UK Biobank. Even though the heritability and number of loci and genes for fMRI was lower than in dMRI and sMRI, the proportions of genome-wide significant loci and genes that did and did not overlap with other modalities was generally equal across modalities. The results in the ABCD cohort showed generalizability of structural-diffusion MRI pleiotropy from old age to late childhood, and from European ancestry to (ad)mixed ancestries. Note that more sophisticated methods for analyses with admixed individuals are becoming available[54], that allow to model individuals' local ancestry. A lack of power in the smaller ABCD cohort most likely limited the robust estimation of genetic associations for the relatively low heritable functional MRI-derived phenotypes, which complicated examining the generalizability of structural-functional-diffusion MRI pleiotropy beyond structural-diffusion MRI

338 pleiotropy. The necessary future samples size to uncover the heritability of neuroimaging-derived
339 phenotypes depends on genetic architecture and varies across modalities or the brain trait of interest[6].
340 For example, researchers expect that genome-wide significant variants in the prospective UK Biobank
341 neuroimaging sample of 100,000 individuals will explain 32% and 24% of the SNP-based heritability
342 of cortical surface area and thickness respectively[5]. This highlights that both larger neuroimaging-
343 genetic datasets and novel statistical approaches are required to delineate the genetic background of
344 neuroimaging phenotypes.

345 Our enrichment analyses showed that the biological processes and molecular components only
346 enriched for cross-modality genes include more general functions such as cell cycle processes,
347 regulation of gene expression, cell junctions. That gene-sets implicated by cross-modality genes alone
348 are involved in fundamental biological processes is consistent with previous findings that genes
349 associated with multiple trait domains are more likely to be involved in general biological functions[3].
350 This is in line with our finding that, although cross-modality genes were highly expressed in fetal brain
351 tissue, those genes were not necessarily cell-specific – suggesting they may serve in cellular functions
352 taking place in all cell-types. Future research may explore this putative relationship between the degree
353 of pleiotropy and the specificity of biological functions. The functional enrichment of cross-modality
354 lead SNPs investigated here can hint towards the mechanisms through which pleiotropic effects could
355 be exerted. Cross-modality pleiotropic lead SNPs were enriched for exonic variants as found previously
356 for pleiotropic SNPs[3,12]. There is previous evidence to suggest that pleiotropy emerges from the
357 variants' effect on total expression of functional protein, for example by the selective exclusion of
358 missense exons from the gene transcript[55]. However, identification of the (unmeasured) causal SNPs
359 tagged by these multivariate GWAS lead SNPs is necessary in future studies to uncover the mechanisms
360 through which variants exerts their pleiotropic effect.

361 Conditioning GWAS summary statistics for five major psychiatric disorders on our multimodal
362 multivariate analysis prioritized variants that enhanced polygenic prediction for BD and ADHD. This
363 highlights that future diagnostic strategies targeting psychiatric disorders may benefit from aligning the
364 PGSs to a relevant endophenotype. This rationale is similar to a previously published study that showed
365 improved prediction of SCZ after filtering on genetic variants expressed in the placenta[56]. Exploring
366 which disorders benefit from either a more specific or broader range of neuroimaging phenotypes could
367 also be an intriguing area for future research.

368 Some limitations are worth noting when interpreting our results. First, despite our efforts to
369 harmonize the three sets of phenotypes to the greatest degree, the differential spatial granularity and
370 number of features across modalities can result in differential representation of certain brain regions or
371 brain characteristics in the multivariate signal. Second, our definition of single-modality loci and genes
372 is inherent to two factors: 1) the grouping of phenotypes in unimodal and multimodal analyses –
373 MOSTest may miss a small number of hits when genetic signal is sparse across the jointly analyzed
374 phenotypes[7], and 2) the currently available data limits our statistical power – loci that are now associated

with one modality could become genome-wide significant in another modality once sample sizes increase. Third, one should keep in mind that the presence of statistical pleiotropy as indicated in this study can include instances of pleiotropy where multiple traits are affected by one gene but different causal SNPs[4] or one locus with distinct gene effects that are in linkage disequilibrium[3]. Fourth, MOSTest does not provide effect directions due to its multivariate nature and requires the use of individual level data. This restricted access to certain post GWAS analyses that require direction of effect, but these limitations were outweighed by MOSTest's ability to boost identification of variants with shared effects across phenotypes and handle hundreds of phenotypes with sample size differences in a computationally efficient manner[7]. Fifth, the temporal gene-expression patterns of single-modality and cross-modality genes are based on a *postmortem* dataset with a limited number of donors per timepoint[36]. A study that was recently released compared gene-expression between *postmortem* and living brain samples and found a significant difference in 80% of the genes[57]. Although living postnatal gene-expression samples may become available in the future for validation of our findings, prenatal gene-expression will remain to rely on *postmortem* brain tissue.

In conclusion, we identified extensive cross-modality pleiotropy and demonstrated that combining different neuroimaging modalities in multivariate analysis substantially increases genetic variant and gene discovery compared to multivariate analyses within single modalities. The results presented improve our understanding of the biology implicated by single-modality and cross-modality genetic effects, and provide insights into the mechanistic pathways linking common genetic variation, brain structure and function, and psychiatric disorders.

Methods

Samples

UK Biobank

The primary analyses of this study were conducted using data from UK Biobank participants who provided written informed consent. This population-based resource obtained ethical approval from the National Research Ethics Service Committee North West–Haydock (reference 11/NW/0382) and the current study was conducted under application number 27412. We included participants that passed quality control for functional[8] or diffusion[9] MRI-derived phenotypes as described in previous publications (N_{fMRI} = 39,951; N_{dMRI} = 31,306). We increased our sample size for participants with structural MRI-derived phenotypes compared to the original MOSTest publication[7], since new data had been released (N_{sMRI} = 42,068). For all three subsamples, we excluded participants based on relatedness (kinship coefficient > 0.05 as estimated in PLINK), non-European ancestry (UKB field 22006), a genotype missing rate > 10%, and bad scan quality as indicated by an (age- and sex-adjusted) Euler number > 3 SDs lower than the scanner site mean. This resulted in the sample characteristics described in Supplementary Data 18.

412 *Adolescent Brain Cognitive Development (ABCD) Study*

413 Baseline data from ABCD participants from release 3.0 [NIMH Data Archive (NDA) DOI:10.151.54/
414 1519007] were used for the replication efforts in this study. All children in this cohort assented before
415 participation and their parents or guardians provided written informed consent. The procedures were
416 approved by a central Institutional Review Board (IRB) at the University of California, San Diego, and,
417 in some cases, by individual site Institutional Review Boards. We included participants with data for
418 the structural or functional MRI-derived phenotypes of interest (N_{sMRI} = 11,760; N_{fMRI} = 11,801) or
419 quality controlled diffusion MRI-derived phenotypes as previously described[9] (N_{dMRI} = 11,904). For all
420 three subsamples, we excluded participants based on recommended criteria for either modality as
421 provided by ABCD (e.g. imgincl_t1w_include), relatedness (first cousin), a genotype missing rate >
422 10%, and bad scan quality as indicated by an (age- and sex-adjusted) Euler number < 3 SDs lower than
423 the scanner site mean. This resulted in a sample with (ad)mixed ancestries (Table 1), as described by
424 Fan et al. recently[58]. We additionally excluded participants with a genetic ancestry factor[31] of European
425 ancestry < 90% (as provided by ABCD and applied previously in Loughnan *et al*[59]) to create a
426 replication sample that matched the ancestry characteristics of the discovery sample more specifically
427 (Supplementary Data 18).

428

429 **Genotype data**

430 *UK Biobank*

431 UK Biobank samples were genotyped from whole blood either using the UK BiLEVE or the UK
432 Biobank axiom array and subsequently quality controlled and imputed by the UK Biobank Team[60].
433 Additional quality control was performed in-house and included SNP filters on minor allele frequency
434 (MAF > 0.1%), imputation information score (INFO > 0.5), Hardy-Weinberg equilibrium (HWE; $p <$
435 1×10^{-9}) and missingness (< 10%). This resulted in 9,061,587 SNPs used for association testing.
436 Ancestral principal components were computed within European samples by UK Biobank and used to
437 control for population stratification.

438

439 *ABCD*

440 Release 3.0 genotype data from ABCD participants was obtained through the Affymetric NIDA
441 SmokeScreen Array, using either saliva or whole blood based on higher successful calls, higher non-
442 missingness, matched genetic sex and less excessive identity by state. Initial quality control was
443 performed by ABCD based on calling signals and variant call rates and subsequently following pre-
444 imputation RICOPILI (Rapid Imputation and Computational Pipeline). We complemented ABCD
445 quality control after creating two subsamples (European and (ad)mixed ancestries as described above)
446 by further filtering pre-imputed variants on call rates (<5% missingness), MAF > 0.01, passing the
447 HWE test ($p < 1\times10^{-9}$) and heterozygosity rate (deviating >6SD from the mean value) in PLINK2[61]. A
448 pruned set of SNPs (r^2=0.1) was used to estimate 20 genetic principal components within each

subsample to use downstream as covariates in multivariate GWAS. Genetic data was phased and imputed using the TOPMed imputation server and only SNPs with high imputation quality were retained (INFO > 0.9).

Neuroimaging data

Structural MRI-derived phenotypes

Three previous publications have used MOSTest on sMRI-derived phenotypes. These included either region-of-interest (ROI)-based cortical thickness, surface area, and subcortical volume[7] or vertex-based cortical thickness, surface area[5], and sulcal depth[6]. Given that the aim of this study was to combine phenotypes derived from three modalities and MOSTest can currently analyse a few thousand of phenotypes simultaneously, we opted to use the ROI-based cortical thickness, surface area, and subcortical volume[7] phenotypes given their relative low dimensionality. Supplementary Data 1 contains all the regional morphology measures included in the current study and indicates which measures were analysed for the left and right hemisphere separately. The respective publication by Van Der Meer & Frei *et al*[7] describes how the sets of 36 regional subcortical volumes, 68 cortical thickness and 68 surface area, as well as estimated intracranial volume (for covariate use downstream), were extracted from T1-weighted MRI using FreeSurfer v5.3[62,63] in UK Biobank samples. A similar procedure was performed by the ABCD Data Acquisition and Integration Core and were readily available. As the importance of normally distributed phenotypes for MOSTest was demonstrated in the original publication[7], we applied rank-based inverse-normal transformation to each measure.

Functional MRI-derived phenotypes

We used functional MRI (fMRI)-derived phenotypes as previously described by Roelfs *et al*[8]. The UK Biobank resting-state fMRI scans were processed into 1,000 Schaefer parcels[64] and mapped onto 17 large-scale brain networks defined by Yeo & Krienen *et al*[65]. The averaged time series within each Yeo-defined network were Pearson correlated and represented 136 brain connectivity measures next to the 17 network variances (Supplementary Data 1). Rank-based inverse-normal transformation was applied to each measure.

The ABCD Data Acquisition and Integration Core provided similar, but not identical, resting-state fMRI-derived phenotypes for replication purposes. Instead of the 17 Yeo & Krienen networks based on 1,000 parcels, temporal variance in 333 Gordon-defined parcels and 66 average correlations between 12 Gordon-defined networks were available[66]. We averaged the parcel variances belonging to the same network to achieve comparability to our discovery phenotypes. Subsequently, we rank-based inverse-normal transformed the 12 network variances and 66 network connectivity phenotypes.

483 *Diffusion MRI-derived phenotypes*

484 We used diffusion MRI-derived phenotypes from UK Biobank and ABCD based on a voxel-wise
485 restriction spectrum imaging (RSI) model (as in Fan *et al*[9]). In short, RSI estimates the signal volume
486 fractions of separable pools of water in the human brain (i.e. intracellular, extracellular, and unhindered
487 free water) and their corresponding spherical harmonic coefficients[67,68]. Three RSI features were used:
488 1) restricted isotropic diffusion (N0) is most sensitive to isotropically diffusing water in the restricted
489 compartment (within cell bodies), restricted directional diffusion (ND) is sensitive to anisotropically
490 diffusing water in the restricted compartment (within oriented structures such as axons and dendrites),
491 and 3) normalized free water diffusion (NF) is sensitive to cerebrospinal fluid or intravascular spaces[69].
492 Fan *et al*[9] calculated the principal components (PCs) across all voxels and extracted the first 5,000 PCs
493 explaining more than 70% of the total variance of each feature. Here, due to dimensionality constraints,
494 we reduced the number of PCs for each feature by estimating the "elbow" of each scree plot of
495 eigenvalues using the nScree function of the nFactors R package (Supplementary Fig. 10). This resulted
496 in the first 124 ND-PCs, 87 NF-PCs and 65 N0-PCs used in our multivariate GWAS.

497

498 **Statistical analyses**

499 *SNP-based GWAS*

500 We performed discovery and replication SNP-based GWAS in PLINK2[61] for every MRI-derived
501 phenotype separately while controlling for sex, age, age^2, genotype array (UKB only), scanner, 20
502 genetic principal components, and modality specific covariates. The latter included Euler number, and
503 total surface area, mean thickness or intracranial volume (sMRI), signal to noise ratio and motion
504 (fMRI), and intracranial volume (dMRI). A linear regression model with additive allelic effects was
505 fitted for each SNP. Subsequently, SNP-based heritability (h^2_{SNP}) was estimated for each phenotype
506 using Linkage Disequilibrium Score Regression (LDSC)[32]. Univariate GWAS summary statistics from
507 non-heritable phenotypes (nominal significance threshold $\frac{h^2_{SNP}}{h^2_{SNP_{SE}}} > 1.96$ as used by Roelfs *et al*[8]) were
508 dropped from further multivariate analyses, since including them may reduce statistical power[7]. This
509 led to the exclusion of 3 dMRI-derived NF-PCs and 14 fMRI-derived connectivity phenotypes GWAS
510 summary statistics (Supplementary Data 1). Then, variant *z*-scores from univariate GWAS were
511 combined in the MOSTest framework to construct multivariate *p*-values as described by Shadrin *et al*[5].
512 This approach selects a regularization parameter optimized to the maximum yield of genetic loci
513 (Supplementary Data 19). The alpha level for SNPs reaching genome-wide significance in the
514 multivariate GWAS was $\alpha = 5\times10^{-8}$.

515

516 *Locus definition*

517 We defined genome-wide significant loci from MOSTest and conditional FDR summary statistics
518 following a protocol as implemented in FUMA. First, independent genome-wide significant SNPs ($p <$

519 $5\times10^{-8}/3$) were obtained by clumping ($r^2 < 0.6$) and SNPs in linkage disequilibrium (LD) with them (r^2
520 ≥ 0.6) were defined as candidate SNPs. LD was estimated using reference genotypes, using 5,000
521 random participants from the UK Biobank sample for UKB-based summary statistics and 1000
522 Genomes Phase 3 EUR for the European as well as (ad)mixed ancestry ABCD-based summary
523 statistics. Second, independent significant SNPs with $r^2 < 0.1$ were defined as lead SNPs and the
524 minimum and maximum positional coordinates of the corresponding candidate variants defined the
525 locus start and end position. Loci in <250kb proximity were merged into a single locus. We excluded
526 loci with a single SNP as these are more likely to be false positives. To test the robustness of our findings
527 we performed sensitivity analyses for the locus definition parameters LD cut-off and merging proximity
528 window, leading to twelve scenarios with different locus definition settings (Supplementary Fig. 3).
529
530 *Multivariate gene-based GWAS*
531 We explored the overlap between modalities and multimodal MOSTest on a gene-level by applying a
532 SNP-wise mean model for 18,877 genes with MAGMA (Multi-marker Analysis of GenoMic
533 Annotation) v1.08[33] in FUMA. The SNP-based MOSTest summary statistics from sMRI, dMRI, fMRI
534 and multimodal served as input with default settings and the UKB European population was used as
535 reference. The alpha level for genes reaching genome-wide significance was adjusted from $\alpha = 0.05$ to
536 $\alpha = (0.05/18{,}877)/3 = 8.83\times10^{-7}$ according to Bonferroni correction for multiple testing.
537
538 *Definition of single-modality and cross-modality loci and genes*
539 Locus overlap between the three MOSTest summary statistics (sMRI, fMRI, dMRI) was defined as
540 physically overlapping genome-wide significant loci after clumping (see above). We used the
541 GenomicRanges R-package[70] to compare the chromosome and start and end base pair positions of all
542 loci between any pair of summary statistics. A locus was considered single-modality when it did not
543 overlap with any of the loci identified for other modalities. All loci that were found to overlap between
544 two or more modalities also overlapped with the multimodal loci, hence we decided to represent these
545 cross-modality loci with the association statistics of the multimodal locus' lead SNP(s) in downstream
546 analyses. All univariate statistics that laid the basis of these single-modality and cross-modality lead
547 SNPs were plotted in a heatmap with the hierarchical clustering algorithm as implemented in
548 seaborn.clustermap (https://seaborn.pydata.org/generated/seaborn.clustermap.html) applied to the
549 phenotypes using (Pearson's) correlation as a distance metric and method='average' clustering. The
550 sMRI, fMRI, dMRI and multimodal MOSTest genes that were found to be genome-wide significant in
551 MAGMA were compared to provide a similar overview. The overlapping patterns were then plotted
552 with the eulerr R-package. This procedure was repeated for both discovery (UKB) and replication
553 (ABCD) MOSTest summary statistics, to investigate whether a similar overlapping pattern could be
554 observed.
555

556 *Comparison of single-modality and cross-modality lead SNPs*

557 We were interested in potential differences between single-modality and cross-modality loci compared
558 to other complex traits. Therefore we selected the lead SNPs within the respective loci (see *Locus*
559 *definition*) and annotated them with ANNOVAR[34]. As reference, we used the annotations of 43,492
560 (unique) lead SNPs derived from 558 traits (with reasonable power, i.e. N > 50,000) across 24 trait
561 domains (Table 2 from Watanabe et al.[3]). Enrichment of the single-modality and cross-modality lead
562 SNPs in positional annotation categories was then tested using Fisher's Exact test. The alpha level for
563 significant enrichment was Bonferroni corrected ($\alpha = 0.05 / 44 = 1.14\times10^{-3}$).

564

565 *Comparison of single-modality and cross-modality gene properties*

566 In order to interpret the biological processes, cellular components or molecular functions our single-
567 modality and cross-modality genes are involved in, Gene Ontology (GO)[71,72] gene-sets were tested for
568 enrichment of genes in these four lists using hypergeometric testing as implemented in FUMA[73]. Protein
569 coding genes were used as background genes and Bonferroni correction was applied to adjust for the
570 multiple comparisons. The significant single-modality and cross-modality GO terms were visualized as
571 a graph using Cytoscape, EnrichmentMap[74] and AutoAnnotate[75] following the Nature Protocol
572 by Reimand & Isserlin *et al*[76]. Stringent pathway similarity scores (Jaccard and overlap combined
573 coefficient = 0.6 as used in Paczkowska, Barenboim *et al*[77]) were used as edges.

574 To visualize the temporal gene expression pattern of the sets of single-modality and cross-
575 modality genes, we made use of gene expression data derived from brain tissue from 56 donors[36]. This
576 dataset ranges from 5 weeks post conception to 82 years of age and we used the data as pre-processed
577 in Kang *et al*[36]. We selected the probe with the highest differential stability for each gene (n = 16,849).
578 A number of single-modality ($n_{sMRI\text{-}modality} = 111$, $n_{dMRI\text{-}modality} = 83$, $n_{fMRI\text{-}modality} = 2$) and cross-modality
579 genes (n=300) were not available in the data. Given the relatively high homogeneity of expression
580 patterns across cortical brain samples[78], we subsequently averaged over 13 cortical regions, within
581 donor, and normalized the expression values, within probe, across donors, to a range between 0 and
582 100. Mean expression over time per set of genes was plotted with ggplot2 in R v4.0.3., with
583 geom_smooth(method="gam") using default settings.

584 Lastly, we explored cell-type enrichment of single-modality and cross-modality genes by
585 performing Fisher Exact Tests for each cell-type identified by Bhaduri et al.[38] using single-cell RNA
586 sequencing analyses of the fetal brain. The minimum number of genes overlapping was set to >2.
587 Bonferroni correction was applied for the number of tests performed ($\alpha = 0.05 / 30 = 1.67\times10^{-3}$).

588

589 *Conditional FDR*

590 We explored whether the multimodal multivariate summary statistics could be leveraged to improve
591 polygenic predcition for major psychiatric disorders (schizophrenia[41], bipolar disorder[42], major
592 depressive disorder[43], attention-deficit hyperactivity disorder[45], and autism spectrum disorder[46]). For

593 that purpose, we applied Conditional False Discovery Rate (cFDR)[40] on Psychiatric Genomics
594 Consortium GWAS summary statistics (listed in Supplementary Data 12) by conditioning on
595 multimodal MOSTest summary statistics. We obtained disorder summary statistics excluding UK
596 Biobank to prevent sample overlap. In cFDR analyses, original *p*-values are replaced by FDR values
597 that reflect the posterior probability that a SNP is null for the disorder given that the *p*-values for both
598 phenotypes are as small or smaller as the observed *p*-values:

599 (1) $$FDR(p_{disorder}|p_{multimodal}) = \frac{\pi_0(p_{multimodal})p_{disorder}}{F(p_{disorder}|p_{multimodal})}$$

600 with F = the conditional empirical cumulative distribution function and $\pi_0(p_{multimodal})$ = the
601 conditional proportion of null SNPs for the disorder given that *p*-values for the multimodal phenotype
602 are as small or smaller.
603
604 *PleioPGS*
605 We then computed the prediction power of the original and conditioned summary statistics by
606 constructing polygenic scores (PGS) in independent case-control samples for the five major psychiatric
607 disorders described above in independent samples. The TOP, BUPGEN and MoBa[82] samples are
608 described in the Supplementary Methods (and Supplementary Data 12). We applied two different set-
609 ups, both based on the C+T (clumping + thresholding) approach[83] using different strategies for ranking
610 SNPs: 1) original GWAS *p*-value-based ranking and original GWAS effect sizes (standard PGS); 2)
611 cFDR-based ranking (described above) and original GWAS effect sizes (pleioPGS as introduced by
612 Van der Meer *et al*[30]; https://github.com/precimed/pleiofdr). For these two setups PGS were calculated
613 across five sets of LD-independent SNPs (n_{SNP} = 1,000, 10,000, 50,000, 100,000, 150,000) using
614 PRSice-2 (v2.3.3)[84] with no additional clumping (--no-clump option). Sets of LD-independent SNPs
615 were obtained using plink v1.90b6.1[61] based on the setup-defined SNP ranking with --clump-kb 250, -
616 -clump-r2 0.1 parameters and in-sample LD estimates. In both setups the phenotypic variance explained
617 by the PGS (R^2) was estimated as the difference between the R^2 of the full regression model and the R^2
618 of the null model only including the covariates (age, sex and the first 10 genetic PCs), also known as
619 Nagelkerke's pseudo-R^2. Bonferroni correction was applied for the number of tests ($p < (0.05 / (7\ n_{SNP}$
620 cut-offs × 2 PRS models × 5 disorders) =) 7.14×10^{-4}.
621 Per disorder we tested if the highest R^2 (MDD pleioPGS/SCZ original PGS with n_{SNP} = 50,000;
622 ADHD/ASD/BD pleioPGS with n_{SNP} = 100,000) was significantly higher than the R^2 of the alternative
623 PGS (MDD original PGS/SCZ pleioPGS with n_{SNP} = 50,000; ADHD/ASD/BD original PGS with n_{SNP}
624 = 100,000). For each disorder a likelihood ratio test was applied to test if a model with both pleioPGS
625 and original PGS provides significantly better prediction compared to the model with only one PGS
626 included.

Data availability

Genome-wide summary statistics will be made publicly available via https://ctg.cncr.nl/software/summary_statistics/ and GWAS Catalog upon acceptance. The individual-level data that support the discovery findings of this study are available from UK Biobank but restrictions apply to the availability of these data, which were used under license no. 27412 for the current study. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development^{SM} (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). ABCD data used for replication in this study is registered under the NDA study register at https://doi.org/10.15154/1527969. Data from the Norwegian Mother, Father and Child Cohort Study and the Medical Birth Registry of Norway used in this study are managed by the national health register holders in Norway (Norwegian Institute of public health) and can be made available to researchers, provided approval from the Regional Committees for Medical and Health Research Ethics (REC), compliance with the EU General Data Protection Regulation (GDPR) and approval from the data owners.

Code availability

Code to obtain the multimodal results presented in this manuscript are available via https://github.com/EPTissink/MOSTest-multimodal. Code from previously published studies based on different modalities from which we used phenotypic data are available at https://github.com/cmig-research-group/RSIGWAS (dMRI), https://github.com/precimed/mostest (sMRI) and https://www.github.com/norment/open-science (fMRI).

Acknowledgements

E.P.T. has been supported by the Foundation "De Drie Lichten" and The Simons Foundation Fund in The Netherlands. This work was partly performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT Department (USIT) (tsd-drift@usit.uio.no). We gratefully acknowledge support from the The Netherlands Organization for Scientific Research (NWO Gravitation: BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology - Grant No. 024.004.012, and VICI 452-16-015), The European Research Council (Advanced Grant No ERC-2018-AdG GWAS2FUNC 834057, Consolidator Grant No 101001062), American National Institutes of Health (NS057198, EB000790, 1R01MH124839, R01MH120025, R01MH122688), The Research Council of Norway (RCN) (229129, 213837, 324252, 300309, 273291, 223273, 248980, 276082, 323961), The South-East Norway Regional Health Authority (2019-108, 2022-073), KG Jebsen Stiftelsen (SKGJ-MED-021), EEA (#EEA2018-0573) collaborative grant in ASD: "Improving quality of life for Autism Spectrum Disorders patients by promoting strategies for early diagnosis and preventive measures" and EEA-RO-

NO-Grant 2014-2021 (contract No 6/2019), and the European Union's Horizon 2020 research and innovation programme (No 847776 and 964874 and 801133, Marie Sklodowska-Curie grant agreement). We want to acknowledge the participants and investigators of the UK Biobank, ABCD Study, TOP, BUPGEN, PGC and FinnGen studies. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). The graphical abstract includes images obtained from Servier Medical Art (smart.servier.com). Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development SM Study (ABCD Study®) (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers: U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators.html. ABCD Study consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD Study consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this came from [NIMH Data Archive Digital Object Identifier (10.151.54/1519007)].

Conflicts of interest

E.P.T., A.A.S., D.v.d.M, N.P., G.H., D.R., O.F., C.C.F., M.N., T.N., M.B., S.D., L.T.W., M.P.v.d.H., D.P. and T.K. declare no conflicts of interest. Dr. Andreassen has received speaker's honorarium from Lundbeck and Janssen, and is a consultant to coretechs.ai. Dr. Dale is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. and receives funding through research agreements with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies.

Author contributions

E.P.T., A.A.S., O.F., D.v.d.M. and O.A.A. conceived of the study. E.P.T. performed the analyses with A.A.S. and D.v.d.M. Neuroimaging data were processed by D.R., T.K., C.C.F., A.M.D, D.v.d.M and E.P.T. Genotype data were processed by E.P.T. and D.v.d.M. All authors contributed with conceptual input on methods and/or interpretation of results. E.P.T. wrote the manuscript and all authors contributed to the final manuscript.

References

1. Tononi, G., Edelman, G. M. & Sporns, O. Complexity and coherency: Integrating information in the brain. *Trends Cogn. Sci.* **2**, 474–484 (1998).
2. Jansen, A. G., Mous, S. E., White, T., Posthuma, D. & Polderman, T. J. C. What Twin Studies Tell Us About the Heritability of Brain Development, Morphology, and Function: A Review. *Neuropsychol. Rev.* **25**, 27–46 (2015).
3. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
4. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: Challenges and strategies. *Nat. Rev. Genet.* **14**, 483–495 (2013).
5. Shadrin, A. A. *et al.* Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, (2021).
6. van der Meer, D. *et al.* The genetic architecture of human cortical folding. *Sci. Adv.* **7**, (2021).
7. van der Meer, D. *et al.* Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, (2020).
8. Roelfs, D. *et al.* Genetic overlap between multivariate measures of human functional brain connectivity and psychiatric disorders. *medRxiv* (2022).
9. Fan, C. C. *et al.* Multivariate genome-wide association study on tissue-sensitive diffusion metrics highlights pathways that shape the human brain. *Nat. Commun.* **13**, 2423 (2022).
10. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482 (2019).
11. Liu, S. *et al.* Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain Informatics* **2**, 167–180 (2015).
12. Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* **89**, 607–618 (2011).
13. Zhu, W. & Zhang, H. Why do we test multiple traits in genetic association studies? *J. Korean Stat. Soc.* **38**, 1–10 (2009).
14. Roelfs, D. *et al.* Shared genetic architecture between mental health and the brain functional

732 connectome in the UK Biobank. *BMC Psychiatry* **23**, (2023).
733 15. Boedhoe, P. S. W. *et al.* Subcortical brain volume, regional cortical thickness, and cortical
734 surface area across disorders: Findings from the ENIGMA ADHD, ASD, and OCD working
735 groups. *Am. J. Psychiatry* **177**, 834–843 (2020).
736 16. Hibar, D. P. *et al.* Cortical abnormalities in bipolar disorder: An MRI analysis of 6503
737 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol. Psychiatry* **23**, 932–942
738 (2018).
739 17. van Erp, T. G. M. *et al.* Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia
740 and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis
741 (ENIGMA) Consortium. *Biol. Psychiatry* **84**, 644–654 (2018).
742 18. van Erp, T. G. M. *et al.* Subcortical brain volume abnormalities in 2028 individuals with
743 schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* **21**, 547–
744 553 (2016).
745 19. Schmaal, L. *et al.* Cortical abnormalities in adults and adolescents with major depression based
746 on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working
747 Group. *Mol. Psychiatry* **22**, 900–909 (2017).
748 20. Schmaal, L. *et al.* Subcortical brain alterations in major depressive disorder: findings from the
749 ENIGMA Major Depressive Disorder working group. *Mol. Psychiatry* **21**, 806–812 (2016).
750 21. van den Heuvel, M. P. & Sporns, O. A cross-disorder connectome landscape of brain
751 dysconnectivity. *Nat. Rev. Neurosci.* (2019). doi:10.1038/s41583-019-0177-6
752 22. Rudie, J. D. *et al.* Altered functional and structural brain network organization in autism.
753 *NeuroImage Clin.* **2**, 79–94 (2013).
754 23. Daianu, M. *et al.* Disrupted rich club network in behavioral variant frontotemporal dementia and
755 early-onset Alzheimer's disease. *Hum. Brain Mapp.* **37**, 868–883 (2016).
756 24. Kelly, S. *et al.* Widespread white matter microstructural differences in schizophrenia across
757 4322 individuals: Results from the ENIGMA Schizophrenia DTI Working Group. *Mol.*
758 *Psychiatry* **23**, 1261–1269 (2018).
759 25. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: Large-scale
760 proof of concept. *Nat. Neurosci.* **19**, 420–431 (2016).
761 26. Tissink, E. *et al.* Genome-wide association study of cerebellar volume provides insights into
762 heritable mechanisms underlying brain development and mental health. *Commun. Biol.* **5**,
763 (2022).
764 27. Ohi, K. *et al.* Genetic correlations between subcortical brain volumes and psychiatric disorders.
765 *Br. J. Psychiatry* **216**, 280–283 (2020).
766 28. Cheng, W. *et al.* Shared genetic architecture between schizophrenia and subcortical brain
767 volumes implicates early neurodevelopmental processes and brain development in childhood.
768 *Mol. Psychiatry* (2022). doi:10.1038/s41380-022-01751-z

29. Cheng, W. *et al.* Genetic Association Between Schizophrenia and Cortical Brain Surface Area and Thickness. *JAMA Psychiatry* 1–11 (2021). doi:10.1001/jamapsychiatry.2021.1435
30. van der Meer, D. *et al.* Boosting Schizophrenia Genetics by Utilizing Genetic Overlap With Brain Morphology. *Biol. Psychiatry* **92**, 291–298 (2022).
31. Raj, A., Stephens, M. & Pritchard, J. K. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).
32. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
33. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
34. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, 1–7 (2010).
35. Schork, A. J. *et al.* All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* **9**, (2013).
36. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
37. van der Meer, D. & Kaufmann, T. Mapping the genetic architecture of cortical morphology through neuroimaging: progress and perspectives. *Transl. Psychiatry* **12**, 447 (2022).
38. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
39. Hindley, G. *et al.* Multivariate genetic analysis of personality and cognitive traits reveals abundant pleiotropy. *Nat. Hum. Behav.* (2023).
40. Andreassen, O. A. *et al.* Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. *PLoS Genet.* **9**, (2013).
41. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
42. Mullins, N. *et al.* Genome-wide association study of over 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
43. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
44. Levey, D. F. *et al.* Reproducible Genetic Risk Loci for Anxiety: Results From ∼200,000 Participants in the Million Veteran Program. *Am. J. Psychiatry* **177**, 223–232 (2020).
45. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
46. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

806 47. Middeldorp, C. M. *et al.* A Genome-Wide Association Meta-Analysis of Attention-
807 Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *J. Am. Acad.*
808 *Child Adolesc. Psychiatry* **55**, 896–905 (2016).
809 48. Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P. & Kristiansson, K. FinnGen: Unique genetic
810 insights from combining isolated population and national health register data. *medRxiv* (2022).
811 49. Lam, M. *et al.* Comparative Genetic Architectures of Schizophrenia in East Asian and European
812 Populations. *Nat. Genet.* **29**, (2019).
813 50. Sporns, O. The complex brain: connectivity, dynamics, information. *Trends Cogn. Sci.* **26**,
814 1066–1067 (2022).
815 51. Turelli, M. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the
816 abdominal bristle. *Theor. Popul. Biol.* **25**, 138–193 (1984).
817 52. Luo, L. *et al.* Multi-trait analysis of rare-variant association summary statistics using MTAR.
818 *Nat. Commun.* **11**, 1–11 (2020).
819 53. Dutta, D., Scott, L., Boehnke, M. & Lee, S. Multi-SKAT: General framework to test for rare-
820 variant association with multiple phenotypes. *Genet. Epidemiol.* **43**, 4–23 (2019).
821 54. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals
822 in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
823 55. Drivas, T. G., Wojno, A. P., Tucker, B. A., Stone, E. M. & Bennett, J. Basal exon skipping and
824 genetic pleiotropy: A predictive model of disease pathogenesis. *Sci. Transl. Med.* **7**, 1–8 (2015).
825 56. Ursini, G. *et al.* Convergence of placenta biology and genetic risk for schizophrenia. *Nat. Med.*
826 **24**, 792–801 (2018).
827 57. Liharska, L. E. *et al.* A study of gene expression in the living human brain Living Brain Project
828 cohort. *bioRxiv* (2023).
829 58. Fan, C. C. *et al.* Genotype Data and Derived Genetic Instruments of Adolescent Brain Cognitive
830 Development Study® for Better Understanding of Human Brain Development. *Behav. Genet.*
831 159–168 (2023). doi:10.1007/s10519-023-10143-0
832 59. Loughnan, R. J. *et al.* Gene-experience correlation during cognitive development: Evidence
833 from the Adolescent Brain Cognitive Development (ABCD) Study. *bioRxiv* (2021).
834 60. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
835 **562**, 203–209 (2018).
836 61. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
837 Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
838 62. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex
839 on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
840 63. Fischl, B. *et al.* Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures
841 in the Human Brain. *Neuron* **33**, 341–355 (2002).
842 64. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic

Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

65. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–45 (2011).

66. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex* **26**, 288–303 (2016).

67. White, N. S., Leergaard, T. B., D'Arceuil, H., Bjaalie, J. G. & Dale, A. M. Probing tissue microstructure with restriction spectrum imaging: Histological and theoretical validation. *Hum. Brain Mapp.* **34**, 327–346 (2013).

68. White, N. S. *et al.* Improved conspicuity and delineation of high-grade primary and metastatic brain tumors using 'restriction spectrum imaging': Quantitative comparison with high B-value DWI and ADC. *Am. J. Neuroradiol.* **34**, 958–964 (2013).

69. Palmer, C. E. *et al.* Microstructural development from 9 to 14 years: Evidence from the ABCD Study. *Dev. Cogn. Neurosci.* **53**, 101044 (2022).

70. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, 1–10 (2013).

71. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

72. The Gene Ontology Consortium. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).

73. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

74. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, (2010).

75. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **5**, 1–14 (2016).

76. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, (2019).

77. Paczkowska, M. *et al.* Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.* **11**, 1–16 (2020).

78. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).

79. Bahrami, S. *et al.* Dissecting the shared genetic basis of migraine and mental disorders using novel statistical tools. *Brain* **145**, 142–153 (2022).

80. Torgersen, K. *et al.* Shared genetic loci between depression and cardiometabolic traits. *PLoS Genet.* **18**, 1–25 (2022).

81. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, 1318–1330 (2020).

880 82. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study
881 (MoBa). *Int. J. Epidemiol.* **45**, 382–388 (2016).
882 83. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score
883 analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
884 84. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data.
885 *Gigascience* **8**, 1–6 (2019).
886