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: e.p.tissink@vu.nl, 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

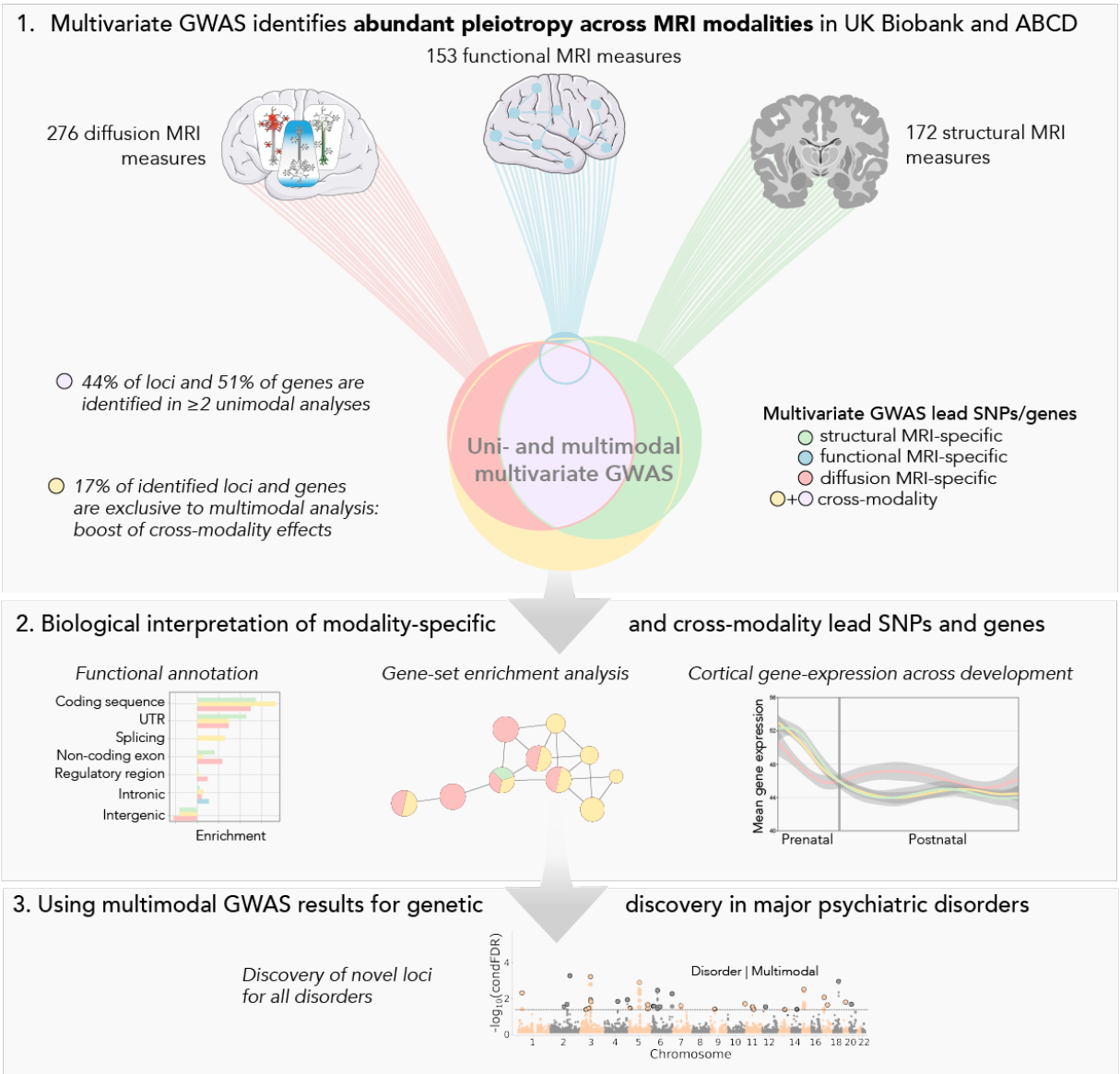
[12] Department of Child and Adolescent Psychology and Psychiatry, section Complex Trait Genetics, Amsterdam Neuroscience, VU University Medical Centre, Amsterdam, The Netherlands

[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

Graphical abstract



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) method, we boost the discovery of loci and genes associated with brain features 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 genetic discovery for 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 – features often concurrently altered within psychiatric disorders.

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].

Multivariate GWAS approaches gain their statistical power due to the distributed nature of genetic influences across phenotypes[12]. Such approaches are well-suited for the identification of pleiotropic variants and genes with effects across neuroimaging modalities. 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. Compared to conventional mass-univariate approaches these multivariate studies showed a boost in genetic discovery (whilst maintaining correct type-I errors[7] and ensuring replicability within and across samples[13]) due to extensive genetic overlap across spatially distributed modality-specific measures. The identified loci and genes inform us about the biological signal that is picked up by MRI. Previous work has linked 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 specificity of these links remains unclear. The extent to which these genetic effects overlap and show pleiotropy across neuroimaging modalities can be investigated by 1) overlapping the previously described multivariate effects for each modality and 2) combining all MRI-derived phenotypes into one

multimodal multivariate GWAS, obtaining an additional boost for discovery of cross-modality pleiotropic loci and genes. Identifying such loci, which are currently not discoverable in unimodal analyses, 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.

Here, we demonstrate evidence of extensive pleiotropy across neuroimaging modalities using 583 structural (sMRI), resting-state functional (fMRI) or diffusion (dMRI) MRI-derived phenotypes in the UK Biobank and the Adolescent Brain Cognitive Development (ABCD) Study®. We do so by performing unimodal and multimodal multivariate GWAS with the Multivariate Omnibus Statistical Test (MOSTest), which was designed to boost statistical power by capitalizing on the distributed nature of genetic influences across imaging-derived phenotypes[7]. We functionally annotate modality-specific (identified in only one unimodal analysis) and cross-modality (identified in ≥ 2 unimodal analyses or unique to multimodal analysis) loci and genes to describe the biological signal unique to and shared across MRI modalities. We improve locus discovery of the major psychiatric disorders by using the multimodal, multivariate, genetic signal of brain morphology, functional connectivity, and tissue composition. Thereby the current study provides insight into pleiotropic effects across neuroimaging modalities and their relevance for understanding the neurobiology of the human brain and mental health conditions.

Results

Abundant pleiotropy of genome-wide significant loci and genes across neuroimaging modalities

We used data from three previous studies that applied MOSTest on single neuroimaging modality phenotypes: 172 sMRI-derived brain morphology measures (68 regional surface area and 68 regional thickness of the cerebral cortex, and 36 volumes of subcortical structures)[7], 153 fMRI-derived BOLD signal and connectivity measures (17 network variances and 136 network correlations)[8], and 276 dMRI-derived brain tissue composition principal components (65 restricted isotropic diffusion PCs, 124 restricted directional diffusion PCs, 87 normalized free water diffusion PCs)[9]. We investigated white British UK Biobank samples (as derived from both self-declared ethnic background and genetic

principal component analysis) with quality-controlled genotypes and neuroimaging available (N_{sMRI} = 34,029, N_{fMRI}=31,023, N_{dMRI} = 30,106). We also included the ABCD cohort which had identical sMRI- and dMRI-derived measures and similar fMRI-derived measures (see Methods) available in a European (N_{sMRI} = 4,794, N_{fMRI} = 4,132, N_{dMRI} = 4,418) and mixed ancestry (assigned based on genetic ancestry factor[15] as defined in Methods) quality-controlled samples (N_{sMRI} = 8,607, N_{fMRI} = 7,277, N_{dMRI} = 7,853). These ABCD samples were used to test the generalizability of our results across age and genetic ancestral compositions. Heritability estimates obtained in UK Biobank with linkage disequilibrium (LD) Score Regression[16] 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 Figure 1 and Supplementary Table 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 Figure 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 Figure 3). 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 640, 44 and 562 genome-wide significant loci associated with sMRI, fMRI and dMRI respectively ($p < 5\times10^{-8}$; Supplementary Table 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 mixed ancestries (42.12% sMRI, 15.38% fMRI, 35.39% dMRI; Supplementary Table 6). Applying MAGMA[17] gene-level analyses to unimodal multivariate summary statistics identified 1,809, 45 and 1,638 genome-wide significant genes ($p < 2.65\times10^{-6}$) for sMRI, fMRI

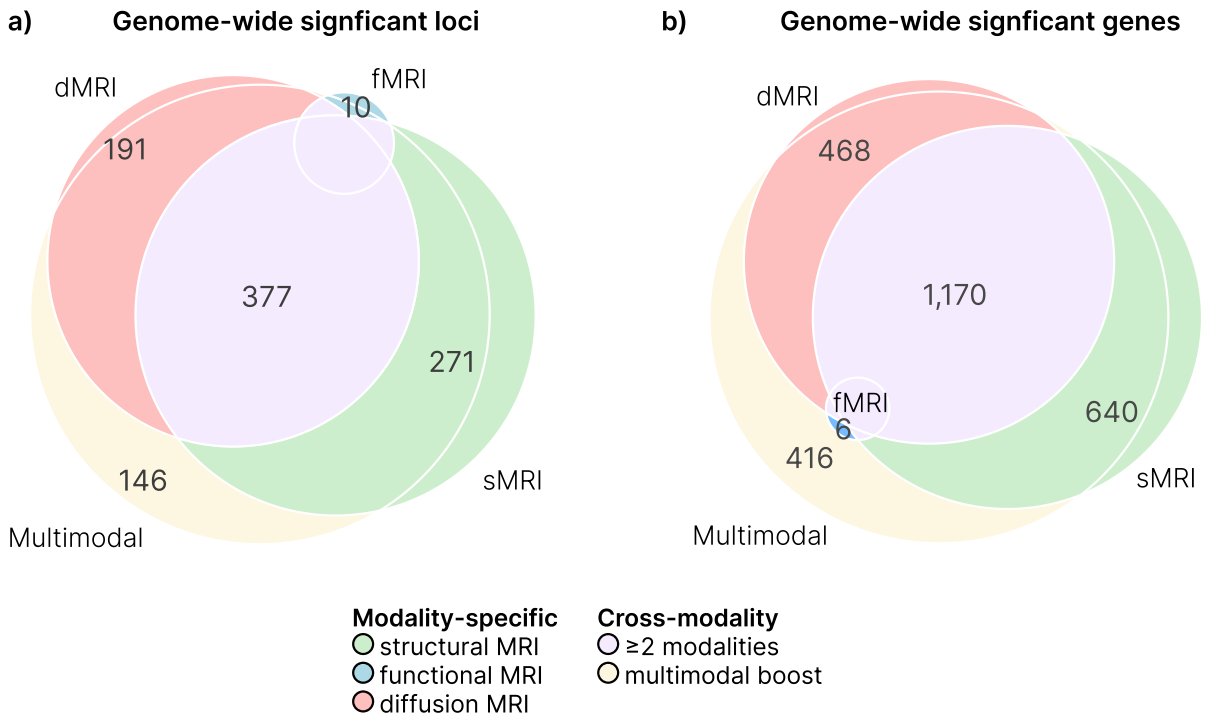


Figure 1. Overlap of genome-wide significant a) loci and b) genes observed across neuroimaging modalities in modality-specific and joint multimodal analyses ($p < 5\times10^{-8}$). When sMRI, fMRI and dMRI-derived phenotypes are jointly analysed in MOSTest (multimodal analysis), a boost in discovery of pleiotropic loci and genes is observed (yellow). This pattern partially replicates in the ABCD cohort (Supplementary Figure 4).

and dMRI respectively (Supplementary Table 3). When overlapping the identified loci ($p < 5\times10^{-8}$) and genes ($p < 2.65\times10^{-6}$) from each modality (Methods), we observed 377 loci (44.41% of total) and 1,170 genes (51.23% of total) associated with at least two out of three modalities (Figure 1). 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 (Supplementary Figure 4, Supplementary Table 4-5). 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 mixed ancestry ABCD sample.

Multimodal genome-wide association study boosts discovery of cross-modality loci and genes

We next investigated whether combining all sMRI, fMRI, and dMRI-derived measures in one multivariate analysis generated greater statistical power to identify novel pleiotropic loci and genes which show sub-threshold associations in each unimodal multivariate analysis (Manhattan plot in Supplementary Figure 3). We therefore applied MOSTest across neuroimaging modalities, combining 583 phenotypes, identifying 851 genetic loci (Supplementary Table 2; replication rates EUR 26.45%, 37.55% mixed ancestries; Supplementary Table 6). One-hundred-forty-six (17.16%) of these loci did not overlap with any of the loci identified in the unimodal multivariate analyses, suggesting that MOSTest leveraged the shared genetic signal across imaging modalities to boost the discovery of pleiotropic loci. Gene-based GWAS from MAGMA showed that of the 2,515 genome-wide significant multimodal genes, 416 (16.54%) were not discovered for unimodal gene-based GWAS (Supplementary Table 3). We used the ABCD cohort to investigate the generalizability of our findings and observed a similar boost in multimodal discovery on a genome-wide significant locus and gene level (Supplementary Figure 4, Supplementary Table 4-5). Additionally, 81.82% of multimodal loci identified in the European ABCD sample were overlapping with the loci identified in the mixed ancestry ABCD sample.

We next determined the extent to which multivariate loci that were uniquely identified in the multimodal analysis demonstrated pleiotropic effects compared to those that were uniquely identified for one modality. We examined the univariate associations underlying different parts of the Venn diagram (Figure 1) by extracting the minimum univariate *p*-value for every locus' lead SNP (Supplementary Figure 5). A relatively high minimum univariate *p*-value would indicate that the signal was highly distributed across other measures for the variant to become genome-wide significant in the multivariate analysis. Lead SNPs boosted by the multimodal analysis had relatively high minimum univariate *p*-values ($0.05 > p > 5\times10^{-5}$) more frequently (89%) than lead SNPs of loci identified in one unimodal analysis only (sMRI-specific 75%; fMRI-specific 22%; dMRI-specific 78%), indicating that the discovery of these lead SNPs was driven by pleiotropic signals across modalities.

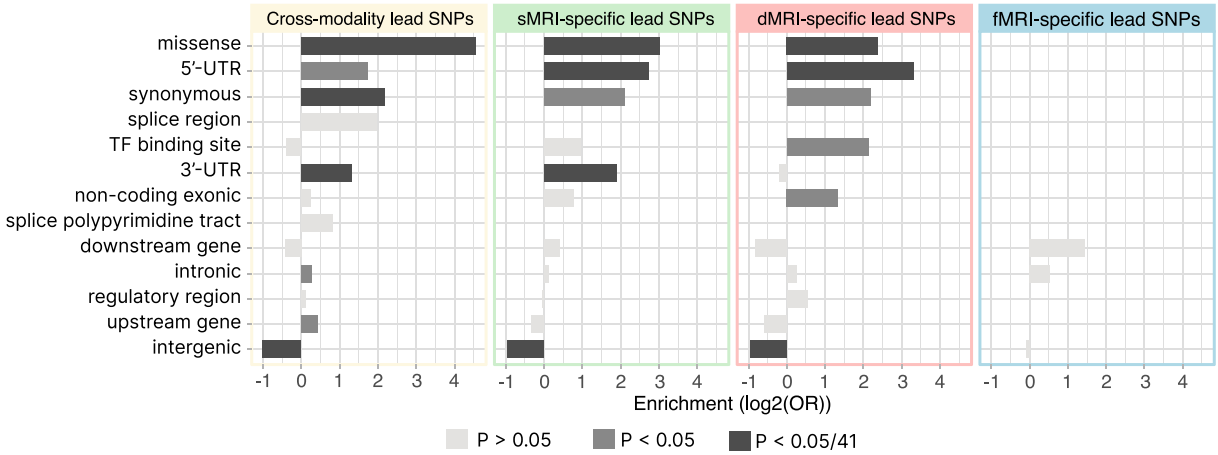


Figure 2. Functional consequences of modality-specific and cross-modality lead SNPs as annotated with VEP. The averaged counts of annotations within ten random sets of pruned SNPs were used as background in Fisher's Exact Test to obtain enrichment estimates and *p*-values (Supplementary Table 9).

Comparing the characteristics of modality-specific and cross-modality loci and genes

We investigated to what extent modality-specific (identified in only one unimodal analysis) and cross-modality (identified in ≥ 2 unimodal analyses or boosted by the multimodal analysis) loci and genes differ in their biological effects (Supplementary Table 7-8). To this end, we annotated 320 sMRI-, 10 fMRI-, 230 dMRI-specific, and 1,261 cross-modality lead SNPs using the Ensembl Variant Effect Predictor[18] (Supplementary Table 9). Figure 2 demonstrates that cross-modality and modality-specific lead SNPs were generally similarly enriched in coding regions and depleted in intergenic regions. dMRI-specific lead SNPs showed unique enrichment of transcription factor binding site (TFBS) variants (OR = 4.38, $p = 3.74\times10^{-2}$) and exonic variants in non-coding genes (OR = 2.53, $p = 1.89\times10^{-2}$), although these association were only nominally significant.

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 640 sMRI-, 6 fMRI-, 468 dMRI-specific, or 1,586 cross-modality genes (Figure 3a). We identified 40 gene-sets that consisted of both modality-specific and cross-modality subsets of genes and showed significant enrichment (Supplementary Table 10). These processes were mostly related to nervous system development and neuronal growth and differentiation. Overall, showing that extensive pleiotropy across neuroimaging modalities is also present on the level of gene-sets. Another 113 gene-

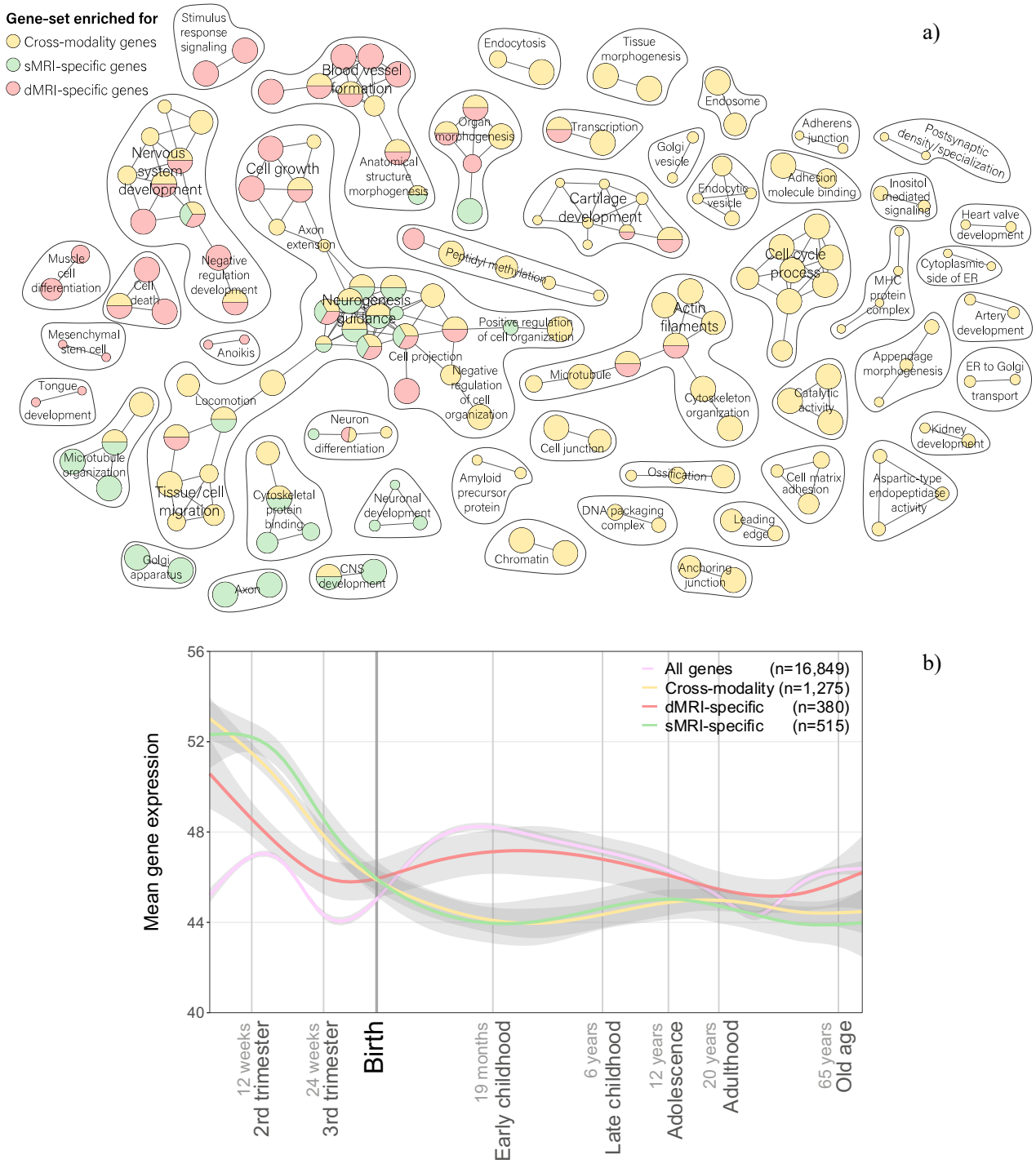


Figure 3. a) Gene-ontology biological processes, molecular functions and cellular components that were (FDR-corrected) significantly enriched for cross-modality, sMRI-specific and/or dMRI-specific genes (none of the GO terms tested showed enrichment for the 6 fMRI-specific genes). Node size reflects gene-set size, edges reflect pathway similarity scores (Methods). AutoAnnotate was used in Cytoscape to name clusters of gene-sets with similar semantics. Note that gene-sets without overlap with any other gene-set are not presented here, they are listed in Supplementary Table 10. b) Mean-normalized expression (y-axis) of cross-modality and modality-specific genes over developmental timepoints (x-axis; log10 scale). Gray shading indicates 95% confidence intervals. The mean-normalized expression of fMRI-specific genes is displayed in Supplementary Figure 6, since the number of genes (n=5) was low and therefore created an unreliable pattern.

sets were enriched for 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 modality-specific 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 (Figure 3b)[19]. The number of fMRI-specific genes in the data (n=5) was too low to generate a reliable expression pattern (Supplementary Figure 6). Prenatal gene expression was high for both sMRI-, dMRI-specific and cross-modality genes compared to all genes in the data. In (early) childhood until adolescence, a differential dMRI-specific expression pattern was not apparent (compared to all genes in the data), whereas sMRI-specific 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[20].

Leveraging shared genetic architecture with psychiatric disorders for genetic discovery

Alterations in brain morphology, connectivity, and tissue composition often co-occur in heritable psychiatric disorders[11], suggesting that our multimodal, multivariate genetic signal may have relevance for genetic discovery and prediction of psychiatric disorders. It is possible to boost locus discovery and polygenic prediction by re-ranking the test-statistics from a given GWAS based on a genetically related secondary GWAS[21]. We conditioned major psychiatric disorder GWAS summary statistics from the Psychiatric Genomics Consortium (Supplementary Table 11) on our multimodal MOSTest summary statistics using the conditional false discovery rate approach (cFDR)[22] to identify novel loci associated with schizophrenia[23], bipolar disorder[24], major depressive disorder[25], attention-deficit hyperactivity disorder[26], and autism spectrum disorder[27]. The rationale behind cFDR is that, in the presence of cross-trait enrichment, a variant with strong associations with both traits is more likely to represent a true association[28]. Compared to the number of genome-wide significant loci identified in the original GWAS, we observed a 6-19 fold increase in locus yield (Supplementary Table 12). We calculated the sign concordance of these newly discovered loci using independent disorder GWAS summary

234 statistics[29–31] (Supplementary Table 11) and observed higher and more significant sign concordance
235 across all disorders compared to the loci identified in the original GWASs (Supplementary Table 13).
236 This illustrates that genetic overlap may be used to identify novel genetic variants that may play a role
237 in psychiatric disorders with relevance for a wide range of neuroimaging traits. We followed-up by
238 testing whether our multimodal cFDR summary statistics could improve polygenic prediction of the
239 disorders in independent samples using a pleiotropy-informed polygenic scoring method[28]. This method
240 improved the prediction for bipolar disorder (R^2 = 5.75%, p = 2.28×10^{-15}) compared to the original
241 polygenic risk score (R^2 = 4.63%, p = 8.93×10^{-13}) (see Supplementary Results 2 and Supplementary
242 Table 14).
243

244 **Discussion**

245 The current findings demonstrated that many loci and genes show pleiotropic effects across brain
246 characteristics derived from three distinct MRI modalities. We find evidence of extensive pleiotropy
247 across structural, functional and diffusion MRI from genetic overlap (377 loci and 1,170 genes being
248 associated with at least two out of three modalities), and a boost in discovery of loci (n=146) and genes
249 (n=416) when all MRI-derived phenotypes are jointly analysed using MOSTest. The results in the
250 ABCD cohort show generalizability of structural and diffusion MRI pleiotropic loci from adulthood to
251 childhood, and from European ancestry to mixed ancestries. We show that each MRI modality captures
252 a genetic signal that includes modality-specific and cross-modality biological processes. Moreover, we
253 show how these results can be leveraged to improve locus discovery for major psychiatric disorders.

254 The human brain is a highly complex and inter-connected structure for which "the whole is
255 more than the sum of its parts"[32]. This complexity emerges from tight interplay between different units
256 and processes, where disturbance of any part may change the state of the whole system. With this in
257 mind, pervasive effects of genetic variants on brain-related traits are inevitable, resulting in abundant
258 pleiotropy not only across phenotypically linked traits such as brain morphology measures[20], but also
259 in widespread genetic overlap between distinct aspects of brain functioning such as personality and
260 cognition[21]. Our results implicate that the multivariate genetic signals of structural, functional, or
261 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.

Genetic overlap was apparent 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 mixed ancestries. A lack of power in this smaller 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 pleiotropy.

Our enrichment analyses showed that the biological processes and molecular components implicated can be decomposed to those 1) enriched for either sMRI-specific (neuron development, differentiation, and migration, the synapse, axon) or dMRI-specific genes (cell death, cellular response to stimuli), 2) converging from modality-specific and cross-modality genes (regulation of neuron/cell projection development, organization and guidance, microtubule), and 3) only enriched for cross-modality genes (more general functions such as cell cycle processes, regulation of gene expression, cell junctions). That gene-sets implicated by cross-modality genes alone are involved in fundamental biological processes is consistent with previous findings that genes associated with multiple trait domains are more likely to be involved in general biological functions[3].

The functional enrichment of modality-specific and cross-modality lead SNPs hints towards the mechanisms through which pleiotropic effects could be exerted. Both within-modality and cross-modality pleiotropic lead SNPs were enriched for protein coding exonic and 3'-UTR variants and depleted for intergenic variants as found previously for pleiotropic SNPs[3,35]. There is previous evidence to suggest that pleiotropy emerges from the variants' effect on total expression of functional protein,

for example by the selective exclusion of missense exons from the gene transcript[36] or post-transcriptional influence on gene expression by variants in 3'-UTR regulatory elements[37]. However, identification of the (unmeasured) causal SNPs tagged by these multivariate GWAS lead SNPs is necessary in future studies to uncover the mechanisms through which variants exerts their pleiotropic effect.

Some limitations are worth noting when interpreting our results. First, despite our efforts to harmonize the three sets of phenotypes to the greatest degree, the differential spatial granularity and number of features across modalities can result in differential representation of certain brain regions or brain characteristics in the multivariate signal. Second, our definition of modality-specificity is inherent to the currently available data and dependends on 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].

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 modality-specific 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 Table 15.

Adolescent Brain Cognitive Development (ABCD) Study

Baseline data from ABCD participants from release 3.0 [NIMH Data Archive (NDA) DOI:10.151.54/ 1519007] were used for the replication efforts in this study. All children in this cohort assented before participation and their parents or guardians provided written informed consent. The procedures were approved by a central Institutional Review Board (IRB) at the University of California, San Diego, and, in some cases, by individual site Institutional Review Boards. We included participants with data for the structural or functional MRI-derived phenotypes of interest (N_{sMRI} = 11,760; N_{fMRI} = 11,801) or quality controlled diffusion MRI-derived phenotypes as previously described[9] (N_{dMRI} = 11,904). For all three subsamples, we excluded participants based on recommended criteria for either modality as provided by ABCD (e.g. imgincl_t1w_include), relatedness (first cousin), 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 a sample with mixed ancestries (Table 1). We additionally

343 excluded participants with a genetic ancestry factor[15] of European ancestry < 90% (as provided by
344 ABCD and applied previously in Loughnan *et al*[38]) to create a replication sample that matched the
345 ancestry characteristics of the discovery sample more specifically (Supplementary Table 15).
346
347 **Genotype data**
348 *UK Biobank*
349 UK Biobank samples were genotyped from whole blood either using the UK BiLEVE or the UK
350 Biobank axiom array and subsequently quality controlled and imputed by the UK Biobank Team[39].
351 Additional quality control was performed in-house and included SNP filters on minor allele frequency
352 (MAF > 0.1%), imputation information score (INFO > 0.5), Hardy-Weinberg equilibrium (HWE; $p <$
353 1×10^{-9}) and missingness (< 10%). This resulted in 9,061,587 SNPs used for association testing.
354 Ancestral principal components were computed within European samples by UK Biobank and used to
355 control for population stratification.
356
357 *ABCD*
358 Release 3.0 genotype data from ABCD participants was obtained through the Affymetric NIDA
359 SmokeScreen Array, using either saliva or whole blood based on higher successful calls, higher non-
360 missingness, matched genetic sex and less excessive identity by state. Initial quality control was
361 performed by ABCD based on calling signals and variant call rates and subsequently following pre-
362 imputation RICOPILI (Rapid Imputation and Computational Pipeline). We complemented ABCD
363 quality control after creating two subsamples (European and mixed ancestries as described above) by
364 further filtering pre-imputed variants on call rates (<5% missingness), MAF > 0.01, passing the HWE
365 test ($p < 1\times10^{-9}$) and heterozygosity rate (deviating >6SD from the mean value) in PLINK2[40]. A pruned
366 set of SNPs (r^2=0.1) was used to estimate 20 genetic principal components within each subsample to
367 use downstream as covariates in multivariate GWAS. Genetic data was phased and imputed using the
368 TOPMed imputation server and only SNPs with high imputation quality were retained (INFO > 0.9).
369
370

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 Table 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[41,42] 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[43] and mapped onto 17 large-scale brain networks defined by Yeo & Krienen *et al*[44]. 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 Table 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[45]. We averaged the parcel variances belonging to

399 the same network to achieve comparability to our discovery phenotypes. Subsequently, we rank-based
400 inverse-normal transformed the 12 network variances and 66 network connectivity phenotypes.
401
402 *Diffusion MRI-derived phenotypes*
403 We used diffusion MRI-derived phenotypes from UK Biobank and ABCD based on a voxel-wise
404 restriction spectrum imaging (RSI) model (as in Fan *et al*[9]). In short, RSI estimates the signal volume
405 fractions of separable pools of water in the human brain (i.e. intracellular, extracellular, and unhindered
406 free water) and their corresponding spherical harmonic coefficients[46,47]. Three RSI features were used:
407 1) restricted isotropic diffusion (N0) is most sensitive to isotropically diffusing water in the restricted
408 compartment (within cell bodies), restricted directional diffusion (ND) is sensitive to anisotropically
409 diffusing water in the restricted compartment (within oriented structures such as axons and dendrites),
410 and 3) normalized free water diffusion (NF) is sensitive to cerebrospinal fluid or intravascular spaces[48].
411 Fan *et al*[9] calculated the principal components (PCs) across all voxels and extracted the first 5,000 PCs
412 explaining more than 70% of the total variance of each feature. Here, due to dimensionality constraints,
413 we reduced the number of PCs for each feature by estimating the "elbow" of each scree plot of
414 eigenvalues using the nScree function of the nFactors R package (Supplementary Figure 8). This
415 resulted in the first 124 ND-PCs, 87 NF-PCs and 65 N0-PCs used in our multivariate GWAS.
416
417 **Statistical analyses**
418 *SNP-based GWAS*
419 We performed discovery and replication SNP-based GWAS in PLINK2[40] for every MRI-derived
420 phenotype separately while controlling for sex, age, age^2, genotype array (UKB only), scanner, 20
421 genetic principal components, and modality specific covariates. The latter included Euler number, and
422 total surface area, mean thickness or intracranial volume (sMRI), signal to noise ratio and motion
423 (fMRI), and intracranial volume (dMRI). A linear regression model with additive allelic effects was
424 fitted for each SNP. Subsequently, SNP-based heritability (h^2_{SNP}) was estimated for each phenotype
425 using Linkage Disequilibrium Score Regression (LDSC)[16]. Univariate GWAS summary statistics from

non-heritable phenotypes (nominal significance threshold $\frac{h^2_{SNP}}{h^2_{SNP_{SE}}} > 1.96$ as used by Roelfs *et al*[8]) were dropped from further multivariate analyses, since including them may reduce statistical power[7]. This led to the exclusion of 3 dMRI-derived NF-PCs and 14 fMRI-derived connectivity phenotypes GWAS summary statistics (Supplementary Table 1). Then, variant z-scores from univariate GWAS were combined in the MOSTest framework to construct multivariate *p*-values as described by Shadrin *et al*[5]. This approach selects a regularization parameter optimized to the maximum yield of genetic loci (Supplementary Table 16). The alpha level for SNPs reaching genome-wide significance in the multivariate GWAS was $\alpha = 5\times10^{-8}$.

Locus definition

We defined genome-wide significant loci from MOSTest and conditional FDR summary statistics following a protocol as implemented in FUMA. First, independent genome-wide significant SNPs were obtained by clumping ($r^2 < 0.6$) and SNPs in linkage disequilibrium (LD) with them ($r^2 \geq 0.6$) were defined as candidate SNPs. LD was estimated using reference genotypes, using 5,000 random participants from the UK Biobank sample for UKB-based summary statistics and 1000 Genomes Phase 3 EUR for the European as well as mixed ancestry ABCD-based summary statistics. Second, independent significant SNPs with $r^2 < 0.1$ were defined as lead SNPs and the minimum and maximum positional coordinates of the corresponding candidate variants defined the locus start and end position. Loci in <250kb proximity were merged into a single locus. We excluded loci with a single SNP as these are more likely to be false positives.

Multivariate gene-based GWAS

We explored the overlap between modalities and multimodal MOSTest on a gene-level by applying a SNP-wise mean model for 18,877 genes with MAGMA (Multi-marker Analysis of GenoMic Annotation) v1.08[17] in FUMA. The SNP-based MOSTest summary statistics from sMRI, dMRI, fMRI and multimodal served as input with default settings and the UKB European population was used as

reference. The alpha level for genes reaching genome-wide significance was adjusted from $\alpha = 0.05$ to $\alpha = (0.05/18{,}877 =) 2.65\times10^{-6}$ according to Bonferroni correction for multiple testing.

Definition of modality-specific and cross-modality loci and genes

Locus overlap between the three MOSTest summary statistics (sMRI, fMRI, dMRI) was defined as physically overlapping genome-wide significant loci after clumping (see above). We used the GenomicRanges R-package[49] to compare the chromosome and start and end base pair positions of all loci between any pair of summary statistics. A locus was considered modality-specific when it did not overlap with any of the loci identified for other modalities. All loci that were found to overlap between two or more modalities also overlapped with the multimodal loci, hence we decided to represent these cross-modality loci with the association statistics of the multimodal locus' lead SNP(s) in downstream analyses. The sMRI, fMRI, dMRI and multimodal MOSTest genes that were found to be genome-wide significant in MAGMA were compared to provide a similar overview. The overlapping patterns were then plotted with the eulerr R-package. This procedure was repeated for both discovery (UKB) and replication (ABCD) MOSTest summary statistics, to investigate whether a similar overlapping pattern could be observed.

Comparison of modality-specific and cross-modality lead SNP properties

We were interested in potential differences between modality-specific and cross-modality loci. Therefore we selected the lead SNPs within the respective loci (see *Locus definition*) and annotated them with the Ensembl Variant Effect Predictor (VEP)[18]. As reference, we averaged the counts of annotations within ten randomly pruned sets of SNPs. Enrichment of the modality-specific and cross-modality lead SNPs in positional annotation categories was then tested using Fisher's Exact test. The alpha level for significant enrichment was Bonferroni corrected ($\alpha = 0.05 / 41 = 1.22\times10^{-3}$).

Comparison of modality-specific and cross-modality gene properties

In order to interpret the biological processes, cellular components or molecular functions our modality-specific and cross-modality genes are involved in, Gene Ontology (GO)[50,51] gene-sets were tested for

enrichment of genes in these four lists using hypergeometric testing as implemented in FUMA[52]. Protein coding genes were used as background genes and correction for multiple comparisons was performed using the Benjamini-Hochberg method. The significant modality-specific and cross-modality GO terms were visualized as a graph using Cytoscape, EnrichmentMap[53] and AutoAnnotate[54] following the Nature Protocol by Reimand & Isserlin *et al*[55]. Stringent pathway similarity scores (Jaccard and overlap combined coefficient = 0.6 as used in Paczkowska, Barenboim *et al*[56]) were used as edges.

To visualize the temporal gene expression pattern of the sets of modality-specific and cross-modality genes, we made use of gene expression data derived from brain tissue from 56 donors[19]. This dataset ranges from 5 weeks post conception to 82 years of age and we used the data as pre-processed in Kang *et al*[19]. We selected the probe with the highest differential stability for each gene (n = 16,660). A number of modality-specific ($n_{sMRI\text{-}specific}$ = 125, $n_{dMRI\text{-}specific}$ = 88, $n_{fMRI\text{-}specific}$ = 1) and cross-modality genes (n=311) were not available in the data. Given the relatively high homogeneity of expression patterns across cortical brain samples[57], we subsequently averaged over 13 cortical regions, within donor, and normalized the expression values, within probe, across donors, to a range between 0 and 100. Mean expression over time per set of genes was plotted with ggplot2 in R v4.0.3., with geom_smooth(method="gam") using default settings.

Conditioning the genetic signal of major psychiatric disorders on multimodal MOSTest

We explored whether the multimodal multivariate summary statistics could be leveraged to improve locus discovery for major psychiatric disorders (schizophrenia[23], bipolar disorder[24], major depressive disorder[25], attention-deficit hyperactivity disorder[26], and autism spectrum disorder[27]). For that purpose, we applied Conditional False Discovery Rate (cFDR)[22] on Psychiatric Genomics Consortium GWAS summary statistics (listed in Supplementary Table 11) by conditioning on multimodal MOSTest summary statistics. We obtained disorder summary statistics excluding UK Biobank to prevent sample overlap. In cFDR analyses, original *p*-values are replaced by FDR values that reflect the posterior probability that a SNP is null for the disorder given that the *p*-values for both phenotypes are as small or smaller as the observed *p*-values:

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

with F = the conditional empirical cumulative distribution function and $\pi_0(p_{multimodal})$ = the conditional proportion of null SNPs for the disorder given that *p*-values for the multimodal phenotype are as small or smaller. We subsequently defined genome-wide significant loci (see *Locus definition*) for the original ($p < 5\times10^{-8}$) and conditioned (FDR < 0.05) summary statistics and compared the results.

We tested how replicable the identified loci from original and conditioned summary statistics were by performing sign concordance tests in independent disorder summary statistics. All summary statistics that were used to look up lead SNPs and test for replication are listed in Supplementary Table 11. An exact binomial was used test to test the null hypothesis that sign concordance was randomly distributed (p=0.5), given the total number of variants and the number of variants with concordant effects.

PleioPGS

We then compared the prediction power of the original and conditioned summary statistics by constructing polygenic scores (PGS) in independent case-control samples for the five major psychiatric disorders described above in independent samples. The TOP, BUPGEN and MoBa[58] samples are described in the Supplementary Methods (and Supplementary Table 11). We applied and compared two different set-ups, both based on the C+T (clumping + thresholding) approach[59] using different strategies for ranking SNPs: 1) original GWAS *p*-value-based ranking and original GWAS effect sizes (standard PGS); 2) cFDR-based ranking (described above) and original GWAS effect sizes (pleioPGS as introduced by Van der Meer *et al*[28]; https://github.com/precimed/pleiofdr). For these two setups PGS were calculated across five sets of LD-independent SNPs (N = 1,000, 10,000, 50,000, 100,000, 150,000) using PRSice-2 (v2.3.3)[60] with no additional clumping (--no-clump option). Sets of LD-independent SNPs were obtained using plink v1.90b6.1[40] based on the setup-defined SNP ranking with --clump-kb 250, --clump-r2 0.1 parameters and in-sample LD estimates. In both setups the phenotypic variance explained by the PGS (R^2) was estimated using linear regression model controlling for age, sex and first 10 genetic PCs.

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

561 The Netherlands Organization for Scientific Research (NWO Gravitation: BRAINSCAPES: A
562 Roadmap from Neurogenetics to Neurobiology - Grant No. 024.004.012, and VICI 452-16-015), The
563 European Research Council (Advanced Grant No ERC-2018-AdG GWAS2FUNC 834057,
564 Consolidator Grant No 101001062), American National Institutes of Health (NS057198, EB000790,
565 1R01MH124839, R01MH120025, R01MH122688), The Research Council of Norway (RCN) (229129,
566 213837, 324252, 300309, 273291, 223273, 248980, 276082, 323961), The South-East Norway
567 Regional Health Authority (2019-108, 2022-073), KG Jebsen Stiftelsen (SKGJ-MED-021), EEA
568 (#EEA2018-0573) collaborative grant in ASD: "Improving quality of life for Autism Spectrum
569 Disorders patients by promoting strategies for early diagnosis and preventive measures" and EEA-RO-
570 NO-Grant 2014-2021 (contract No 6/2019), and the European Union's Horizon 2020 research and
571 innovation programme (No 847776 and 964874 and 801133, Marie Sklodowska-Curie grant
572 agreement). We want to acknowledge the participants and investigators of the UK Biobank, ABCD
573 Study, TOP, BUPGEN, PGC and FinnGen studies. The Norwegian Mother, Father and Child Cohort
574 Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of
575 Education and Research. We are grateful to all the participating families in Norway who take part in
576 this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating
577 high-quality genomic data. This research is part of the HARVEST collaboration, supported by the
578 Research Council of Norway (#229624). The graphical abstract includes images obtained from Servier
579 Medical Art (smart.servier.com). Data used in the preparation of this article were obtained from the
580 Adolescent Brain Cognitive Development SM Study (ABCD Study®) (https://abcdstudy.org), held in
581 the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than
582 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study is
583 supported by the National Institutes of Health and additional federal partners under award numbers:
584 U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117,
585 U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123,
586 and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A
587 listing of participating sites and a complete listing of the study investigators can be found at
588 https://abcdstudy.org/principal-investigators.html. ABCD Study consortium investigators designed and

589 implemented the study and/or provided data but did not necessarily participate in analysis or writing of
590 this report. This manuscript reflects the views of the authors and may not reflect the opinions or views
591 of the NIH or ABCD Study consortium investigators. The ABCD data repository grows and changes
592 over time. The ABCD data used in this came from [NIMH Data Archive Digital Object Identifier
593 (10.151.54/1519007)].
594

595 **Conflicts of interest**

596 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.,
597 D.P. and T.K. declare no conflicts of interest. O.A.A. is a consultant to HealthLytix and received
598 speakers honorarium from Lundbeck. A.M.D. is a Founder of and holds equity in CorTechs Labs, Inc,
599 and serves on its Scientific Advisory Board. The terms of this arrangement have been reviewed and
600 approved by UCSD in accordance with its conflict of interest policies.
601

602 **Author contributions**

603 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
604 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
605 E.P.T. Genotype data were processed by E.P.T. and D.v.d.M. All authors contributed with conceptual
606 input on methods and/or interpretation of results. E.P.T. wrote the manuscript and all authors
607 contributed to the final manuscript.
608

609 **References**

610 1. Tononi, G., Edelman, G. M. & Sporns, O. Complexity and coherency: Integrating information
611 in the brain. *Trends Cogn. Sci.* **2**, 474–484 (1998).
612 2. Jansen, A. G., Mous, S. E., White, T., Posthuma, D. & Polderman, T. J. C. What Twin Studies
613 Tell Us About the Heritability of Brain Development, Morphology, and Function: A Review.
614 *Neuropsychol. Rev.* **25**, 27–46 (2015).
615 3. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits.

Nat. Genet. (2019). doi:10.1101/500090

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* 1–24 (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**, (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. Vroom, C.-R., Leeuw, C. de, Posthuma, D., Dolan, C. V. & Sluis, S. van der. The more the merrier? Multivariate approaches to genome-wide association analysis. *BioRxiv* 1–19 (2019).

13. Loughnan, R. J. *et al.* Generalization of cortical MOSTest genome-wide associations within and across samples. *Neuroimage* **263**, 119632 (2022).

14. Roelfs, D. *et al.* Shared genetic architecture between mental health and the brain functional connectome in the UK Biobank. *medRxiv* (2022).

15. Raj, A., Stephens, M. & Pritchard, J. K. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).

16. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

17. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set

644 Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).

645 18. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).

646 19. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489

647 (2011).

648 20. van der Meer, D. & Kaufmann, T. Mapping the genetic architecture of cortical morphology

649 through neuroimaging: progress and perspectives. *Transl. Psychiatry* (2022).

650 doi:10.1038/s41398-022-02193-5

651 21. Hindley, G. *et al.* Multivariate genetic analysis of personality and cognitive traits reveals

652 abundant pleiotropy and improves prediction. *bioRxiv* (2022).

653 22. Andreassen, O. A. *et al.* Improved Detection of Common Variants Associated with

654 Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery

655 Rate. *PLoS Genet.* **9**, (2013).

656 23. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in

657 schizophrenia. *Nature* **604**, 502–508 (2022).

658 24. Mullins, N. *et al.* Genome-wide association study of over 40,000 bipolar disorder cases

659 provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

660 25. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the

661 genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

662 26. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention

663 deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

664 27. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder.

665 *Nat. Genet.* **51**, 431–444 (2019).

666 28. van der Meer, D. *et al.* Boosting Schizophrenia Genetics by Utilizing Genetic Overlap With

667 Brain Morphology. *Biol. Psychiatry* 1–8 (2022). doi:10.1016/j.biopsych.2021.12.007

668 29. Middeldorp, C. M. *et al.* A Genome-Wide Association Meta-Analysis of Attention-

669 Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *J. Am. Acad.*

670 *Child Adolesc. Psychiatry* **55**, 896–905 (2016).

671 30. Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P. & Kristiansson, K. FinnGen: Unique

genetic insights from combining isolated population and national health register data. *medRxiv* 1–55 (2022).

31. Lam, M. *et al.* Comparative Genetic Architectures of Schizophrenia in East Asian and European Populations. *Nat. Genet.* **29**, (2019).

32. Sporns, O. The complex brain: connectivity, dynamics, information. *Trends Cogn. Sci.* (2022). doi:10.1016/j.tics.2022.08.002

33. De, S., Lopez-Bigas, N. & Teichmann, S. A. Patterns of evolutionary constraints on genes in humans. *BMC Evol. Biol.* **8**, 275 (2008).

34. Orr, H. A. Adaptation and the cost of complexity. *Evolution (N. Y).* **54**, 13–20 (2000).

35. Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* **89**, 607–618 (2011).

36. Drivas, T. G., Wojno, A. P., Tucker, B. A., Stone, E. M. & Bennett, J. Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. *Sci. Transl. Med.* **7**, 1–8 (2015).

37. Halbeisen, R. E., Galgano, A., Scherrer, T. & Gerber, A. P. Post-transcriptional gene regulation: From genome-wide studies to principles. *Cell. Mol. Life Sci.* **65**, 798–813 (2008).

38. Loughnan, R. J. *et al.* Gene-experience correlation during cognitive development: Evidence from the Adolescent Brain Cognitive Development (ABCD) Study. *bioRxiv* (2021).

39. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

40. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

41. Desikan, R. S. *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).

42. Fischl, B. *et al.* Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* **33**, 341–355 (2002).

43. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

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

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

46. 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).

47. 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).

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

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

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

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

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

53. 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).

54. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations [version 1; referees: 2 approved]. *F1000Research* **5**, 1–14 (2016).

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

56. Paczkowska, M. *et al.* Integrative pathway enrichment analysis of multivariate omics data.

Nat. Commun. **11**, 1–16 (2020).

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

58. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **45**, 382–388 (2016).

59. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* (2020). doi:10.1038/s41596-020-0353-1

60. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, 1–6 (2019).