

Exploring the molecular basis of the genetic correlation between body mass index and brain morphological traits

Daniela Fusco¹, Camilla Marinelli¹, Mathilde André¹, Lucia Troiani¹, Martina Noè¹, Fabrizio Pizzagalli¹, Davide Marnetto^{1*} and Paolo Provero^{1,2*}

¹ Dept. of Neurosciences “Rita Levi Montalcini”, University of Turin, Italy

² Center for Omics Sciences, IRCCS Ospedale San Raffaele, Milan, Italy

* Correspondence to davide.marnetto@unito.it, paolo.provero@unito.it

Abstract

Several studies have demonstrated significant phenotypic and genetic correlations between body mass index (BMI) and brain morphological traits derived from structural magnetic resonance imaging (sMRI). We use the sMRI, BMI, and genetic data collected by the UK Biobank to systematically compute the genetic correlations between area, volume, and thickness measurements of hundreds of brain structures on one hand, and BMI on the other. In agreement with previous literature, we find many such measurements to have negative genetic correlation with BMI. We then dissect the molecular mechanisms underlying such correlations using brain eQTL data and summary-based Mendelian randomization, thus producing an atlas of genes whose genetically regulated expression in brain tissues pleiotropically affects brain morphology and BMI. Fine-mapping followed by colocalization analysis allows, in several cases, the identification of credible sets of variants likely to be causal for both the macroscopic phenotypes and for gene expression. In particular, epigenetic fine mapping identifies variant rs7187776 in the 5' UTR of the *TUFM* gene as likely to be causal of increased BMI and decreased caudate volume, possibly through the creation, by the alternate allele, of an ETS binding site leading to increased chromatin accessibility, specifically in microglial cells.

Introduction

In 2022 one out of eight people were diagnosed with obesity worldwide, amounting to over 890 million adults and 160 million children and adolescents living with this chronic complex disease. In the majority of cases, obesity is a combination of environmental, psycho-social, and genetic factors. If the current increasing trend continues, 60% of the entire human population is estimated to be overweight or obese by 2030 [1].

The central nervous system has a role in susceptibility to obesity through the control of food intake [2]. This notion is strengthened by the fact that the heritability of body mass index (BMI), a parameter typically used by the World Health Organization to distinguish normal-weight ($18.5 \leq \text{BMI} \leq 25$) from obese people ($\text{BMI} \geq 30$), was found to be enriched in genes expressed in the brain and central nervous system [3,4]. In particular, a recent enrichment analysis of the genes located near 97 BMI-associated single-nucleotide polymorphisms (SNPs), aimed at identifying tissues and cell types with high expression of such genes, revealed that 27 out of 31 significantly enriched tissues were part of the central nervous system. However, these results are not enough to identify the specific brain regions involved [4].

Beyond the aforementioned results, several neuroimaging studies investigated the correlation between obesity and structural alterations in brain regions, especially differences in gray matter volumes. While the results of these studies have been somewhat conflicting, with both positive and negative correlations reported between gray matter volumes and obesity, most neuroimaging studies reported reduced gray matter volume in many brain regions, including several involved in executive control, to be associated with higher BMI [5]. In

particular, reduced volume of cerebellum [6,7], basal ganglia [8], putamen [7], prefrontal cortex, temporal lobes, and subcortical structures [9], as well as reduced cortical thickness, were found to be associated with higher BMI [10–13].

On the other hand, a positive correlation between gray matter volume and BMI was found in nucleus accumbens and hypothalamus [14], while a twin study identified positive correlation with ventromedial prefrontal cortex and the right cerebellum [9].

The relationship between BMI and white matter integrity is more difficult to characterize since alterations show a more complex pattern [15]. Studies involving diffusion tensor imaging (DTI) investigated the influence of obesity on white matter integrity by comparing BMI with DTI parameters in adults, revealing a negative correlation between several microstructure architectures of the white matter, such as in the corpus callosum, and BMI [15].

To date, most studies have focused on phenotypic correlation. Regarding genetic correlation, a study explored the effects of a collection of obesity-related SNPs on 164 regional brain volume traits from the UK Biobank, finding that 17 such SNPs were associated with 51 regional brain volumes (both positively and negatively) [16]. A bivariate linkage and quantitative analysis on Mexican American individuals looked for genetic factors associated with increased BMI and reduced cortical surface area and subcortical volume, localizing two genome-wide significant QTLs at 17p13.1 and 3q22.1. The former pleiotropically affects ventral diencephalon volume and BMI, suggesting the involvement of such region in obesity through leptin-induced signaling in the hypothalamus. The latter involves the surface area of the supramarginal gyrus and BMI and might be relevant to the food-related reward mechanism [17].

Finally, a study integrating single-cell-RNA-sequencing and genome-wide association studies (GWAS) has shown that susceptibility to obesity is enriched for some hypothalamic cell types such as VMH Sf1-expressing neurons, most of which are involved in the integration of sensory stimuli, learning and memory. Despite neuroanatomical differences, these brain cell types share transcriptional patterns related to obesity [18].

Here, we systematically investigated the genetic correlation between BMI and brain morphology using GWAS summary data from the UK Biobank. To investigate the molecular mechanisms behind these correlations, we used summary-based Mendelian randomization (SMR) to identify the genes whose genetically regulated expression (GReX) in brain tissues pleiotropically affects BMI and brain morphology. Finally, we used fine-mapping and colocalization to identify cases in which a single variant is associated with both BMI and brain morphology through gene expression, and epigenetic fine-mapping based on single-cell chromatin accessibility data to formulate mechanistic hypotheses on the relevant regulatory mechanisms and cell types.

Results

Genetic correlation

In order to assess the shared genetic basis between brain morphology and BMI, we relied on variant-trait associations for 435 cortical and subcortical measurements estimated on more than 30 thousand donors from the UK Biobank [19]. These measurements included volume, thickness, areas from white and pial surfaces, gray-white matter contrast for 54 cortical regions [20]; volumes and mean intensities for subcortical segments [21] (see Methods for details on the selection of traits).

Adopting cross-trait LD Score regression [22], the state-of-the-art summary-based approach, we found 108 nominally significant ($P < 0.05$) genetic correlations (r_g) across all categories. Most of the significant correlations turned out to be negative (Fig. 1A), including in particular those between BMI and cortical areas (white and pial surfaces), cortical and subcortical volumes (all genetic correlation values are reported in Supplementary Data 1). The most significant correlations ($P < 0.001$ in at least one hemisphere) were all negative (Fig. 1B), suggesting that, overall, genetic variants associated with increased BMI are also related to reduced size of brain structures. This is evident from the global subcortical volume and the globus pallidus in particular, and the total areas of pial and white surfaces, with especially negative r_g values in cortical regions adjacent to the central sulcus, and in the temporal and cingulate lobes (Fig. 1B). As expected, results for the two hemispheres are largely consistent with each other (Fig. 1C).

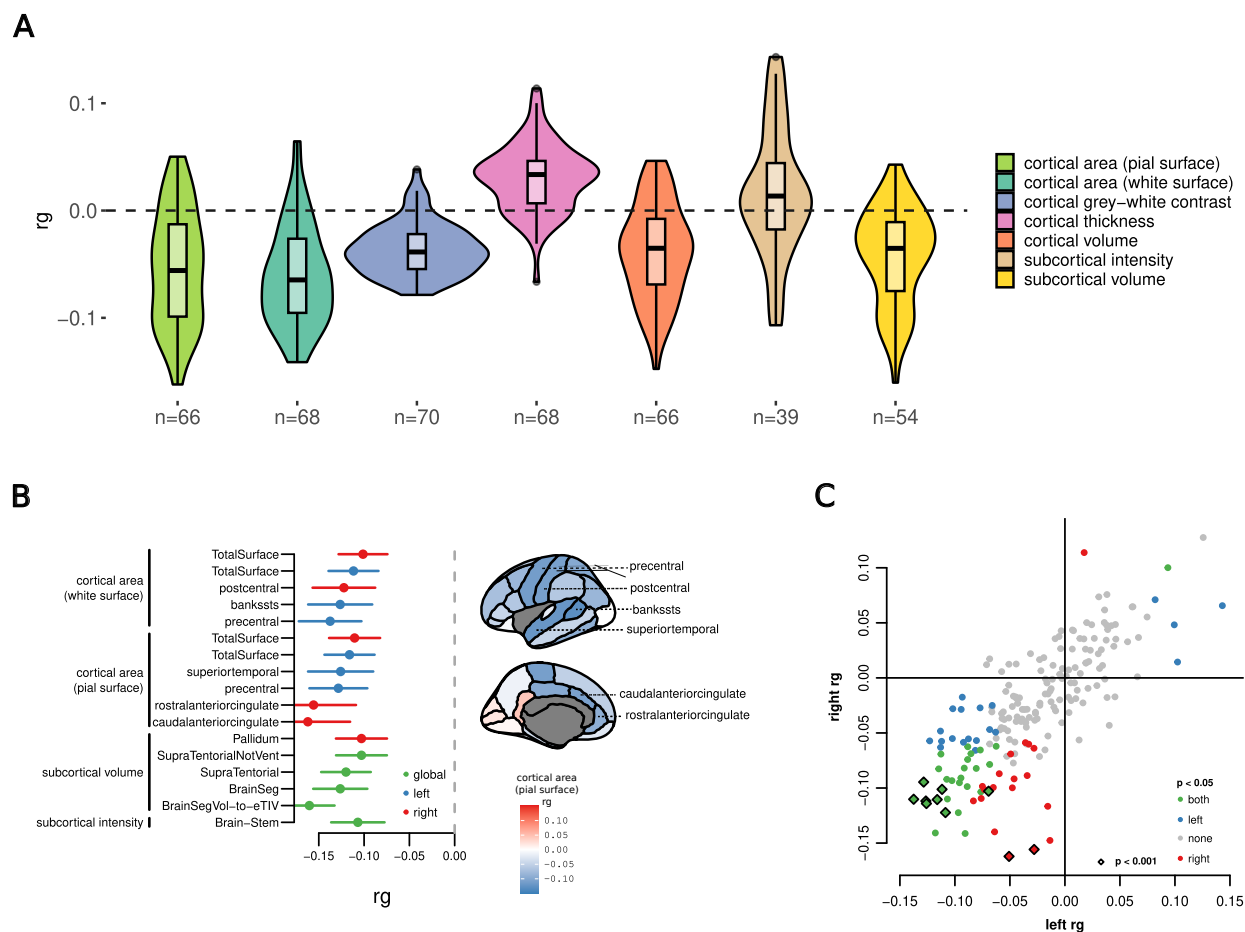


Figure 1: Genetic correlations between sMRI traits and BMI. (A) Genetic correlations (y axis) of all 435 sMRI traits with BMI separated by category. n , number of traits for each category. (B) Traits with the most significant ($P < 0.001$) genetic correlation (x axis) with BMI, grouped by category. A few subcortical volume traits were removed to avoid redundancy. Measurements refer to the left (blue) or right (red) hemisphere, or are global (green). Significant cortical regions are indicated on the brain surfaces on the right, which are colored according to their genetic correlation between pial surface area and BMI (left-right averages). (C) Comparison between genetic correlations in left (x axis) and right (y axis) hemispheres. Dots are colored based on their significance level. Diamonds represent correlations with $P < 0.001$ in either or both hemispheres.

Subcortical intensity traits, as provided by the UK Biobank (UKB), were analyzed, revealing a significant negative genetic correlation for brainstem intensity ($P < 0.001$). However, interpreting this association is not straightforward. Although intensity values partially reflect what voxel-based morphometry (VBM) measures, VBM leverages the optimized intensity contrast among gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissues from T1-weighted images to classify voxels and estimate tissue volumes. Voxel intensity itself, however, can be influenced by MRI scanner artifacts and other sources of noise, introducing biases which are difficult to control. To address this, studies employing this method typically perform statistical analyses after normalizing tissue intensities, thereby estimating the contrasts between different tissues [23].

Summary-based Mendelian randomization

We hypothesized the existence of molecular mechanisms at the level of gene expression underlying these genetic correlations. Mendelian randomization [24] can help elucidate the causal effect of an exposure on an outcome. Thanks to large databases of expression quantitative trait loci (eQTLs), gene expression can be used as an exposure, allowing the identification of putative genes whose regulation is causally (or at least pleiotropically) related to a complex trait. We used summary-based Mendelian randomization (SMR) [25] with gene expression in each of 13 brain tissues included in the GTEx database [26] as the exposure and sMRI traits and BMI as outcomes, reasoning that genes associated with both sMRI and BMI in such analysis could provide a basis for the investigation of the molecular mechanisms driving the genetic correlation between the two types of traits.

SMR is unable, by itself, to distinguish between a causal relationship and a pleiotropic one, that is when gene expression and complex trait are influenced by the same genetic variants but not causally related. In addition, it can identify relationships which are actually mediated by different causal variants for exposure and outcome, when they are in linkage disequilibrium (LD) with each other. For this reason, SMR is often coupled with a test for heterogeneity in dependent instruments (HEIDI) [25], which can rule out a LD-mediated relationship.

A SMR + HEIDI analysis using the 435 sMRI traits and 13 GTEx brain tissues [26] found several genes mediating the shared genetic components of BMI and sMRI traits through their expression (Fig. 2A, respectively blue and green dots). A total of 21 genes (Fig. 2A, red dots; Supplementary Data 2) were associated with both BMI and one or more sMRI traits.

We confirmed the gene/trait associations found by SMR through an individual-level study of the association between GREX and traits, as in transcriptome-wide association studies (TWAS) [27]: the genetic component of gene expression for the 21 genes was predicted in the relevant brain tissues for the UKB subjects for which the relevant phenotype was available (BMI or sMRI), and its correlation with the trait was computed. For BMI the TWAS association was nominally significant in 23 out of 34 cases, and the sign of the correlation was concordant with the one found by SMR for all of them. For sMRI traits, nominal significance and sign concordance was achieved in all of the 687 tested associations.

We asked whether the overlap between BMI- and sMRI-associated genes was greater than expected by chance. Since the significant genes for all traits are clearly clustered into genomic loci (see e.g. the loci in chromosomes 8 and 16 in Fig. 2A, the latter including the well-known obesity associated *FTO* gene [28]), we devised a circular permutation procedure preserving the gene order (see Methods) which revealed that the overlap between BMI- and sMRI-associated genes is indeed greater than expected by chance (empirical $P < 10^{-4}$).

Furthermore, the cis- regions around the pleiotropic genes were enriched in genetic covariance between BMI and sMRI traits, although only nominally and for just 10 sMRI traits out of 97 tested (Monte-Carlo P -value from partial genetic covariance estimates). Nevertheless, as shown in Fig. 2B, when considering the genetic regions surrounding all genes mediating BMI, i.e. without restricting to the 21 found to also mediate sMRI traits, we could not observe any significant enrichment: their estimated log-fold changes were markedly smaller and never distinguishable from 0. This suggests that the portion of the genome surrounding the 21 identified pleiotropic genes, albeit small, indeed carries a larger-than-expected proportion of the shared genetic determinants of brain morphology and BMI.

As shown in Supplementary Fig. 1, there is a dense network of genetic correlations across the measurements

from different brain regions for which we found pleiotropic genes with BMI. This is in line with a common genetic basis across many sMRI traits: it is therefore non-trivial to quantitatively pinpoint which morphological traits are the most affected by these genes. However, all categories show at least one pleiotropic gene (Fig. 2C), and even though cortical gray-white contrast measurements seem very prominent, this can be readily explained by an especially strong genetic correlation among them (Supplementary Fig. 1). Similarly, all GTEx tissues are involved, although with a suggestive prominence of subcortical tissues.

SMR also provides an estimate of the direction and size of the effect of gene expression on each trait. We observed that in most cases the effects of gene expression on BMI and sMRI traits have opposite signs (Supplementary Fig. 2), in agreement with the prevalence of negative genetic correlations. Taken together, these results suggest that the genetic correlations between BMI and morphological brain measures are indeed driven in part by regulatory variants acting on brain gene expression, and provide us with a list of candidate genes for the more detailed variant-level investigations described in the following.

Fine-mapping and colocalization analysis

SMR reveals genes associated to both brain morphology and BMI through their expression in brain tissues, but does not provide direct information about the genetic variants involved. Although the heterogeneity test should rule out non-colocalizing associations between gene expression variation and complex traits (but not necessarily between BMI and sMRI), we used colocalization analysis to determine in which cases the variants likely to be causal of gene expression variation, BMI, and brain morphology coincided. Colocalization was investigated for all the BMI/sMRI trait/gene expression trios found through SMR analysis. Specifically, given such a trio, we used the Sum of Single Effects (SuSiE) regression framework combined with the COLOC algorithm (coloc-SuSiE) [29], which avoids the assumption of a single causal variant for each trait in each locus, to test colocalization for the three pairs of traits (gene expression - sMRI, gene expression - BMI, and BMI - sMRI). The results are shown in Fig. 3, as a network whose nodes are traits, and edges indicate colocalized causal variants.

Of particular interest are cliques involving BMI, an sMRI trait, and the expression of a gene. In these cases a single variant is likely to be causal for the three traits. In particular, this happens for the protein-coding genes *VPS11* and *TUFM*, and the antisense transcripts *RP11-624M8.1* and *RP11-552F3.9*. For other genes, colocalization is limited to gene expression and sMRI traits. Note that the HEIDI test we used after SMR is also a test of colocalization between gene expression on one hand and BMI or sMRI traits on the other, which was passed by all the trait pairs shown in the figure. The fact that not all of them pass the coloc-SuSiE test indicates that the results of HEIDI and coloc-SuSiE agree only partially, a fact previously noted in [30]. Thus we conservatively consider the cliques in the network as having strong evidence of three-way colocalization, and the other phenotypic trios as having weak evidence of colocalization.

The Manhattan plots (Fig. 4) show the SuSiE 95% credible sets for the three traits for the *TUFM* and *VPS11* loci (the loci associated to the two antisense transcripts are shown in Supplementary Fig. 3). In both cases, as expected, the three-way intersection of the SuSiE credible sets contains variants in high LD with each other ($R^2 > 0.5$). Notably, such intersections overlap the *TUFM* and *VPS11* gene loci. Coloc analysis of each pair of traits showed strong evidence of colocalization (PPH4>0.8).

Epigenetic fine-mapping

In order to further restrict the set of candidate causal variants for each locus, we investigated the effect of each of the variants included in the intersection of the credible sets of the three traits (gene expression, BMI, and sMRI) on chromatin accessibility in brain cells. This analysis was limited to the four loci with strong evidence of three-way colocalization, for which the intersections of the credible sets included a total of 128 variants, henceforth referred to as “candidate pleiotropic variants”.

To this purpose, we first trained a gapped k-mer support vector machine (gkm-SVM [31]) on single-nuclei ATAC-seq data of 107 brain cell types produced in [32]. For 105 out of 107 cell types we had sufficient data to train a SVM, which we then used to predict the cell type-specific effect of each candidate pleiotropic variant

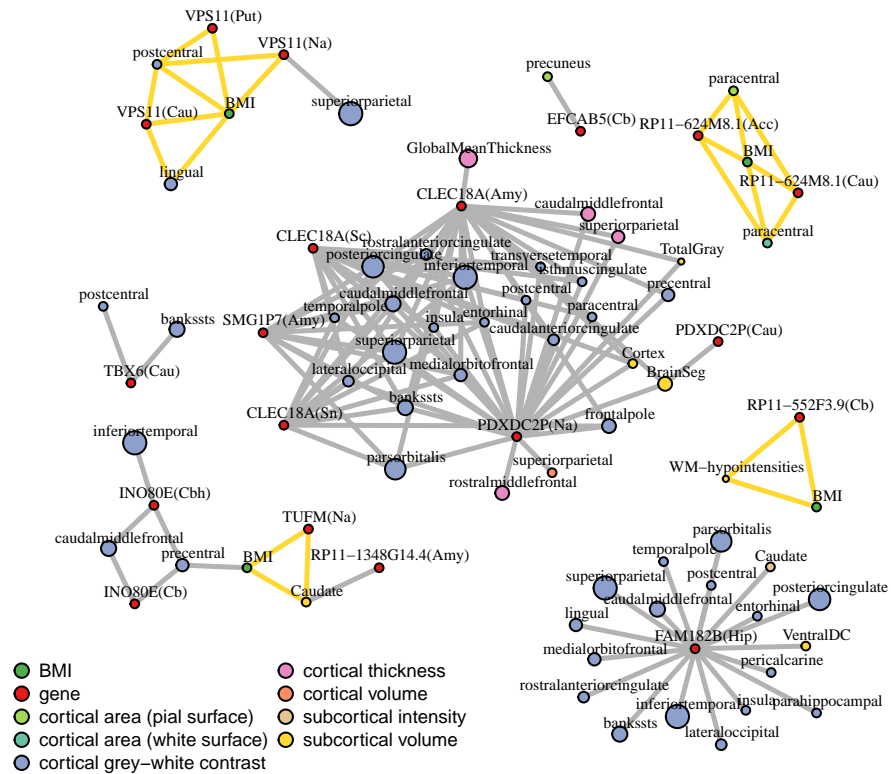


Figure 3: Network of colocalized causal variants. Each edge represents an instance of colocalization, that is a pair of causal variants (or rather genetic components tagged by such variants) for which H4 is the most likely hypothesis according to COLOC; edges highlighted in yellow form a colocalization clique between eQTL, BMI and sMRI causal variants. Each node represents a trait: red nodes stand for gene expression in a GTEx tissue, green nodes for BMI, while sMRI nodes are colored according to their trait category. Traits can be duplicated if multiple independent causal variants are found for them. sMRI traits belonging to the same category with genetic correlation higher than 0.9 were pruned to leave only a representative node, sized proportionally to the trait count that it represents. For all left-right hemisphere pairs this pruning left a single representative node.

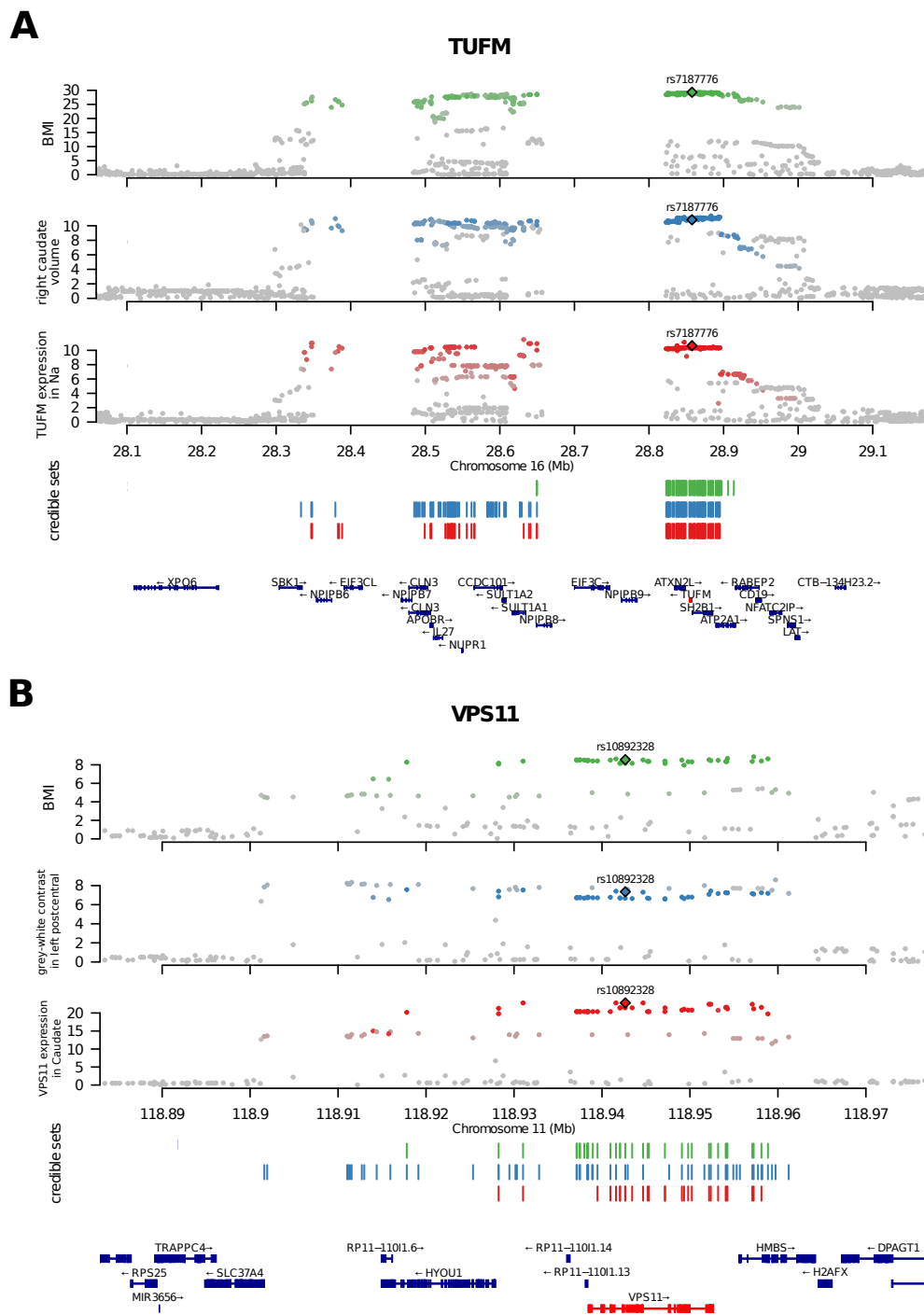


Figure 4: Credible sets and evidence of colocalization. Manhattan plots showing colocalization in the TUFM (A) and VPS11 (B) loci among the trait trios. Variants are plotted with their respective GWAS $-\log_{10}(P)$ (see y axes). BMI (green), sMRI (blue), and gene expression (red) credible sets are shown by vertical bars. For each gene, the three-way intersection of the SuSie credible sets is tagged by a genome-wide significant variant represented by a diamond. Other variants in each credible set are represented by dots whose color represents the level of LD with the tagging variant (gray: $R^2 < 0.5$; colored: $R^2 > 0.5$, with color intensity proportional to R^2). Na= Nucleus accumbens.

on chromatin accessibility. Nine of the 128 candidate pleiotropic variants were predicted ($|\text{deltaSVM}| > 2$) to affect chromatin state in at least one brain cell type (Supplementray Data 3).

The strongest epigenetic fine mapping prediction concerned variant rs7187776, located in the 5' UTR of the *TUFM* gene (Fig. 5A), and previously associated by GWAS to various complex traits including hip circumference adjusted for BMI [33]. The alternate allele was predicted by gkm-SVM to be associated with increased chromatin accessibility specifically in microglial cells ($\text{deltaSVM} = 3.17$). Indeed the same variant was found to be positively associated ($P = 1.1 \cdot 10^{-4}$) with chromatin accessibility in glia, but not neurons, in a recent study of the genetic determinants of chromatin accessibility in the human brain [34].

We thus decided to investigate in greater depth the possible mechanisms by which the alternate allele of rs7187776 could promote the opening of the chromatin in microglia. We used motifbreakR [35] to identify transcription factors (TFs) whose motif was altered by rs7187776. The analysis revealed that the alternate allele introduces a strong binding site for TFs of the ETS family, by creating the core ETS motif "GGAA" in lieu of "AGAA" found in the reference sequence. Specifically, motifbreakR identified several TFs whose affinity is predicted to be increased by the alternate allele, all belonging to the ETS family (Fig. 5B,C).

Thus we hypothesized that rs7187776 opens the chromatin near the *TUFM* transcription start site (TSS) by creating a binding site for an ETS TF. To gain insight into the precise identity of such TF we used single-cell RNA-seq data curated by the Human Protein Atlas [36] to identify which of these TFs are expressed in microglia (Fig. 5D). ETV6 and FLI1 showed the most robust expression in microglial cells, with a remarkable level of cell type-specificity especially for the latter.

Discussion

We have devised a strategy to dissect the molecular underpinnings of the genetic correlation between a macroscopic complex trait (BMI) and a set of endophenotypes (brain morphology parameters assessed by sMRI). Since the effects of genetic variants on complex traits is thought to be mediated mostly by gene regulation, we first used SMR to generate a catalog of 21 genes whose genetically regulated expression pleiotropically affects both the endophenotypes and BMI. The enrichment of both the number of genes found and of the contribution of the respective genomic loci to the genetic correlation suggests that this approach can indeed explain at the molecular level a sizable portion of the observed correlations. In four of the identified loci, colocalization analysis confirmed the three-way pleiotropic effect of the fine-mapped genetic variants on gene expression, BMI, and brain morphology. For these loci we used epigenetic fine mapping to identify putative causal variants and the related mechanisms, highlighting in particular variant rs7187776, located in the 5' UTR of the *TUFM* gene and predicted to alter chromatin accessibility in microglia, possibly by creating a strong ETS binding site.

SMR identified genetically regulated *TUFM* expression to be positively associated with BMI and negatively associated with the volume of caudate in the right hemisphere. *TUFM* encodes a ubiquitously expressed nuclear-encoded mitochondrial protein that has been associated to several biological processes, including autophagy, and human phenotypes, including obesity (see [37] for a recent review). A study on subcortical volumes across lifespan has shown that *TUFM* expression in several brain regions in older and young adults is associated with caudate nucleus volume. In particular, increased *TUFM* expression was associated with smaller caudate nucleus volume, with evidence for colocalization in several tissues [38]. Epigenetic fine mapping and motif analysis suggest a mechanism by which a variant in the 5' UTR of *TUFM* creates a strong ETS binding site resulting in microglia-specific chromatin opening. Of note, chronic microglia activation has been associated with decreased hippocampus and parahippocampus volume in neurodegenerative diseases [39].

Other three-way colocalization signals involved BMI, *VPS11* expression in various subcortical structures, and gray-white contrast phenotypes from sMRI. *VPS11* has an essential role in brain white matter development and neuron survival in zebrafish and its reduced expression leads to an impaired autophagic activity in human cells [40]. To the best of our knowledge it has not been investigated in the context of BMI and obesity, although GWASs found *VPS11*-associated variants associated to BMI [41] and HDL cholesterol measurements [42].

The remaining high-confidence colocalization signals involved two non-coding RNAs, namely *RP11-552F3.9* and *RP11-624M8.1* (also known as *HEY2-AS1*), antisense transcripts of *TRIM47* and *HEY2*, respectively. Variants mapped to these genes have been associated by GWASs to both brain morphology traits [19,43] and BMI [44,45]. Indeed *TRIM47* was among the genes found in [16] to harbor variants associated with both BMI and regional brain volumes.

Some limitations of this work should be noted. It is possible that some of the pleiotropy signals were lost by summarizing the genetic basis of our complex traits of interest into effect-sizes and applying summary-based analyses. While the high number of sMRI traits analyzed initially warranted this approach, individual-based techniques capable of dealing with large sample sizes and multiple traits simultaneously are now being developed, opening the possibility of an individual-based analysis. Notably, the fact that the estimated GReX was predictive of BMI and sMRI traits for the vast majority of the genes identified by SMR, is instead reassuring against the risk of false positives.

Another limitation is that even if we set out to evaluate genetic correlation and pleiotropy for fine-grained brain morphology measurements, the strong genetic correlation across all the sMRI traits analyzed and the lack of an explicitly differential strategy prevented us from pinpointing specific brain regions consistently involved throughout all steps of our analysis.

Finally, the use of BMI as an obesity-related quantitative trait should be complemented by other obesity-related measures, such as waist-to-hip ratio, body fat percentage, fat-free mass, etc. The analysis and comparison of the genetic correlations of these traits with sMRI measurements is likely to allow a more nuanced view of the relationship between obesity and brain morphology.

Conclusion

We have shown how the integration of data from GWASs, population-based transcriptomic studies, and single-cell chromatin accessibility assays allows the molecular dissection of the genetic correlation between complex traits and the formulation of data-driven hypotheses on the relevant regulatory mechanisms. Specifically, we have confirmed and expanded previous observations on the genetic correlation between brain morphological features and BMI, and generated an atlas of 21 genes whose genetically regulated expression mediates a significant part of such correlation.

Methods

Summary statistics and individual data

We used the GWAS summary statistics of brain imaging-derived phenotypes (IDPs) processed with a sample size of about 33 thousands individuals (sample sizes vary across traits) from the UKB release 2020 [19]. In particular, we selected 435 structural IDPs defined according to the Desikan-Killiany cortical atlas [20] and subcortical volumetric segmentation (ASEG) [21], see Supplementary Data 1 for a full list. GWAS summary statistics for BMI were derived from the UK Biobank GWAS version 3, and made available by the Neale Lab (<http://www.nealelab.is/uk-biobank>). All summary statistics were filtered for imputation INFO score ≥ 0.95 and minor allele frequency (MAF) ≥ 0.01 .

For methods that required individual-level information (LD matrices, stratified LD scores, GReX), we used 38,406 samples of European ancestry from the UKB. Subjects were selected among the donors with brain MRI-derived phenotypes as of October 2023 (ASEG whole brain volume, ID 26514, used as representative trait for filtering) identified as British with West-European ancestry (field 22006, code 1), and not considered outliers for missingness and heterozygosity (field 22027). Finally, we pruned relatives with kinship coefficient ≥ 0.04419 (third or lower degree relatives), reaching a subset with the sample size mentioned above. A total of 9,252,961 variants with INFO score > 0.8 and minor allele frequency (MAF) > 0.01 were included in this dataset.

Genetic correlation

We estimated the genetic correlation among each of the 435 sMRI-derived traits and BMI using GWAS summary statistics with the cross-trait LD Score regression (LDSC) software obtained from <https://github.com/bulik/ldsc>. We used genome-wide LD scores computed by the Pan-UK consortium on UKB samples of European descent (<https://pan.ukbb.broadinstitute.org/>): scores were available for 1,094,844 SNPs which met their exporting criteria, most notably after filtering for HapMap3 SNPs, $MAF > 0.01$ and removing the human major histocompatibility complex (MHC) region.

Summary-based Mendelian randomization

We conducted summary-based Mendelian randomization with the Omics Pleiotropic Association (OPERA) software tool [46] to identify pleiotropic associations. Summary-level expression quantitative trait loci (eQTL) data of European individuals ($n = 698$) from the GTEx [26] dataset, including 13 brain tissues, were used for the analysis as the exposure. We estimated the cis-eQTLs with the R package Matrix eQTL [47], testing all variants with $MAF \geq 0.01$ within 2 megabases (Mbs) of each gene's TSS.

BMI and sMRI-derived traits were considered one by one as outcomes. We used 503 European samples from the 1000 Genomes Project [48] as the LD reference data required for the HEIDI test. We set the significance threshold of posterior probability of association (PPA) to 0.5, then performed the HEIDI test for the genes passing this threshold and discarded those for which the null hypothesis of non heterogeneity could be rejected with $P < 0.05$.

The circular permutation procedure to test for BMI/sMRI gene overlap was devised as follows: for 10,000 times, a list of all genes tested, ordered for TSS position and tagged for association status with BMI, was split into K genomic chunks and a random spin was applied to each chunk; chunks were then reassembled in a random order. This resulted in 10,000 random sets with the same positional clustering of the genes identified for BMI, which were used to compute an empirical P -value for the size of the intersection with genes associated with at least one sMRI trait. The procedure was repeated with $K = 1, 5, 10$, obtaining the same result.

Individual-level GREx-trait association

An in-house implementation of prediXcan [27] was used to train an elastic net regression for the 21 putatively pleiotropic genes on the 13 GTEx brain tissues mentioned above. We used for training 80% of the European samples used for eQTL estimation, using 1 Mb as cis-distance from each gene's TSS and including all encompassed variants in the training. Gene expression was first transformed into residuals using the approach described in [30], then regressed on the genotype matrices. We filtered out models achieving $R < 0.1$ or with zero variance in the 20% holdout testing set. The resulting models were applied on individual UKB genetic data described above, for which the GREx was imputed and used as predictor for the BMI and sMRI phenotypes identified by SMR+HEIDI analysis.

Partitioned genetic covariance

We aimed to test enrichment in genetic covariance in the genomic loci around two sets of genes: a) those found associated with both BMI and at least one sMRI trait by SMR analysis, and b) all genes associated with BMI by SMR. These regions were identified by considering a 4 Mb window surrounding the TSS of each gene in each set.

A cross-trait stratified LD Score Regression (sLDSC) [3] was conducted to estimate the partitioned genetic covariance between BMI and sMRI phenotypes at each of the two aforementioned annotations, using GWAS summary statistics as described previously. We limited this test to the 97 sMRI phenotypes previously found to be genetically correlated with BMI at nominal significance.

We used the individual-level UKB genetic data described above to compute partitioned LD scores, setting a 2 Mb LD window and keeping SNPs in HapMap3 and with $MAF > 0.01$ for the regression. Enrichments and the corresponding P -values were computed starting from the genetic covariance estimates and relative standard

errors reported in the LDSC logs, by generating 50,000 Monte Carlo randomizations and testing how often the genetic covariance fraction was less or equal than the SNP fraction, thus obtaining one-sided P -values. The enrichment estimates were extracted from the medians of such distributions and then log₂-transformed.

Fine-mapping and colocalization

In order to assess three-way colocalization assuming multiple causal variants per locus, we applied coloc to the decomposed signals found by SuSiE (coloc-SuSiE [29]) for each locus whose expression in at least one brain tissue was found by SMR+HEIDI analysis to be in association with BMI and a sMRI phenotype. The analysis was performed on each pair of phenotypes resulting from the combination of BMI, sMRI traits and gene expression, then the colocalization results were compared across the three variables. Only SNPs with effect sizes available for all three phenotypes were considered. The *coloc_susie* function from the coloc 5.2.3 R package was run with default parameters. We computed the LD matrix for effects on gene expression with the GTEx European samples [26] from each tissue. The LD matrix for the effects on the complex trait from UKB was computed on the individual UKB genetic data described above. We considered that colocalization occurred when the colocalization posterior probability (PPH4) was greater than the probability of association with different causal variants (PPH3). Manhattan plots were produced with locuszoom [49].

Epigenetic fine mapping and motif analysis

The open chromatin regions of the brain cell types identified in [32] were used to train gkm-SVM [31] models which were then used to predict the effect of each variant on the chromatin accessibility of each cell type. The analysis was performed on the 105 cell types (out of a total of 107 identified in [32]) for which at least 5,000 open chromatin regions were available. Each model was trained on 5,000 open chromatin regions drawn at random from the complete set. This analysis was applied to the 128 variants included in the three-way intersection of credible sets (BMI, sMRI, and gene expression) for the four loci for which we had strong evidence of pleiotropy (supported by both HEIDI and COLOC).

Candidate transcription factor binding sites altered by the variants of interest were identified with motifbreakR [35], using the whole MotifDb collection of positional frequency matrices and log probability scores. We considered as significantly motif-altering the variants whose effect was evaluated as “strong” by motifbreakR and such that the logarithmic score was positive for one allele and negative for the other one.

Acknowledgements

This work was supported in part by a Fondazione CRT grant (grant 2021.1787 to P.P.). Daniela Fusco is a PhD student enrolled in the National PhD in Artificial Intelligence (XXXVIII cycle) course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. UKB data were accessed under application 86275. GTEx data were accessed from dbGAP under project #18670.

Supplementary material

Supplementary Data

- Supplementary Data 1: Genetic correlation between BMI and 435 sMRI traits. The columns contain:
 - Pheno, UKB.ID : UKB phenotype codes
 - ROI: region of interest
 - Category.name: phenotype category
 - rg: estimated genetic correlation
 - se: standard error
 - P-value: P-value of the null $rg = 0$
- Supplementary Data 2: Genes associated by SMR to both BMI and at least one sMRI trait. The columns contain:

- Gene_ID: GENCODE gene id
 - Tissue: GTEX tissue
 - Pheno: UKB code of sMRI trait
 - p_HEIDI_sMRI: HEIDI P-value for the sMRI trait
 - b_SMR_sMRI: effect size of GReX on sMRI
 - PPA_sMRI: posterior probability of association between GReX and sMRI trait
 - p_HEIDI_BMI: HEIDI P-value for BMI
 - b_SMR_BMI: effect size of GReX on BMI
 - PPA_BMI: posterior probability of association between GReX and BMI
- Supplementary Data 3: Variants in the three-way intersections of credible sets predicted to alter chromatin accessibility in brain cell types. The columns contain:
 - cell_type: as in [50]
 - coordinates: SNP coordinates, hg38
 - rsid: SNP id
 - ref: reference allele
 - alt: alternate allele
 - delta_svm: deltaSVM value
 - ENSG00000267801, ENSG00000178952, ENSG00000237742, ENSG00000160695: whether the SNP is included in the three-way intersection of credible sets for each gene

Supplementary Figures:

- Supplementary Fig. 1: Genetic correlation between sMRI traits. Only the correlations between traits in the same category were computed, plus those between cortical area (pial surface) and cortical area (white surface). *, nominally significant ($P < 0.05$)
- Supplementary Fig. 2: Effect size of GReX on BMI and sMRI traits for the 21 common genes
- Supplementary Fig. 3: Manhattan plots as in Fig.4 for the non-coding RNAs *RP11-552F3.9* and *RP11-624M8.1*, respectively antisense transcripts of *TRIM47* and *HEY2*

References

1. Kelly T, Yang W, Chen C-S, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes.* 2008;32: 1431–1437. doi:10.1038/ijo.2008.102
2. Schwartz MW, Woods SC, Porte D, Seeley RJ, Baskin DG. Central nervous system control of food intake. *Nature.* 2000;404: 661–671. doi:10.1038/35007534
3. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47: 1228–1235. doi:10.1038/ng.3404
4. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518: 197–206. doi:10.1038/nature14177
5. García-García I, Michaud A, Dadar M, Zeighami Y, Neseliler S, Collins DL, et al. Neuroanatomical differences in obesity: Meta-analytic findings and their validation in an independent dataset. *Int J Obes.* 2019;43: 943–951. doi:10.1038/s41366-018-0164-4
6. Taki Y, Kinomura S, Sato K, Inoue K, Goto R, Okada K, et al. Relationship Between Body Mass Index and Gray Matter Volume in 1,428 Healthy Individuals. *Obesity.* 2008;16: 119–124. doi:10.1038/oby.2007.4
7. Pannacciulli N, Del Parigi A, Chen K, Le DSNT, Reiman EM, Tataranni PA. Brain abnormalities in human obesity: A voxel-based morphometric study. *NeuroImage.* 2006;31: 1419–1425. doi:10.1016/j.neuroimage.2006.01.047

8. Raji CA, Ho AJ, Parikshak NN, Becker JT, Lopez OL, Kuller LH, et al. Brain structure and obesity. *Human Brain Mapping*. 2010;31: 353–364. doi:10.1002/hbm.20870
9. Weise CM, Piaggi P, Reinhardt M, Chen K, Savage CR, Krakoff J, et al. The obese brain as a heritable phenotype: A combined morphometry and twin study. *Int J Obes*. 2017;41: 458–466. doi:10.1038/ijo.2016.222
10. Caunca MR, Gardener H, Simonetto M, Cheung YK, Alperin N, Yoshita M, et al. Measures of obesity are associated with MRI markers of brain aging: The Northern Manhattan Study. *Neurology*. 2019;93. doi:10.1212/WNL.0000000000007966
11. Opel N, Thalamuthu A, Milaneschi Y, Grotegerd D, Flint C, Leenings R, et al. Brain structural abnormalities in obesity: Relation to age, genetic risk, and common psychiatric disorders: Evidence through univariate and multivariate mega-analysis including 6420 participants from the ENIGMA MDD working group. *Mol Psychiatry*. 2021;26: 4839–4852. doi:10.1038/s41380-020-0774-9
12. Shaw ME, Sachdev PS, Abhayaratna W, Anstey KJ, Cherbuin N. Body mass index is associated with cortical thinning with different patterns in mid- and late-life. *Int J Obes*. 2018;42: 455–461. doi:10.1038/ijo.2017.254
13. Chen L, Zhao S, Wang Y, Niu X, Zhang B, Li X, et al. Genetic Insights into Obesity and Brain: Combine Mendelian Randomization Study and Gene Expression Analysis. *Brain Sciences*. 2023;13: 892. doi:10.3390/brainsci13060892
14. Horstmann A, Busse FP, Mathar D, Müller K, Lepsien J, Schlögl H, et al. Obesity-Related Differences between Women and Men in Brain Structure and Goal-Directed Behavior. *Front Hum Neurosci*. 2011;5. doi:10.3389/fnhum.2011.00058
15. Kullmann S, Schweizer F, Veit R, Fritsche A, Preissl H. Compromised white matter integrity in obesity. *Obes Rev*. 2015;16: 273–281. doi:10.1111/obr.12248
16. Pan X, Zhang M, Tian A, Chen L, Sun Z, Wang L, et al. Exploring the genetic correlation between obesity-related traits and regional brain volumes: Evidence from UK Biobank cohort. *Neuroimage Clin*. 2022;33: 102870. doi:10.1016/j.nicl.2021.102870
17. Curran JE, McKay DR, Winkler AM, Olvera RL, Carless MA, Dyer TD, et al. Identification of Pleiotropic Genetic Effects on Obesity and Brain Anatomy. *Human Heredity*. 2013;75: 136–143. doi:10.1159/000353953
18. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types for obesity. *eLife*. 2020;9: e55851. doi:10.7554/eLife.55851
19. Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci*. 2021;24: 737–745. doi:10.1038/s41593-021-00826-4
20. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006;31: 968–980. doi:10.1016/j.neuroimage.2006.01.021
21. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole Brain Segmentation. *Neuron*. 2002;33: 341–355. doi:10.1016/S0896-6273(02)00569-X
22. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47: 1236–1241. doi:10.1038/ng.3406
23. Jefferson AL, Gifford KA, Damon S, Chapman GW, Liu D, Sparling J, et al. Gray & white matter tissue contrast differentiates Mild Cognitive Impairment converters from non-converters. *Brain Imaging Behav*. 2015;9: 141–148. doi:10.1007/s11682-014-9291-2
24. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. *Nat Rev Methods Primers*. 2022;2: 6. doi:10.1038/s43586-021-00092-5

25. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48: 481–487. doi:10.1038/ng.3538
26. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369: 1318–1330. doi:10.1126/science.aaz1776
27. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47: 1091–1098. doi:10.1038/ng.3367
28. Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. *Trends Genet.* 2010;26: 266–274. doi:10.1016/j.tig.2010.02.006
29. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* 2021;17: e1009440. doi:10.1371/journal.pgen.1009440
30. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun.* 2018;9: 1825. doi:10.1038/s41467-018-03621-1
31. Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, et al. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet.* 2015;47: 955–961. doi:10.1038/ng.3331
32. Li YE, Preissl S, Miller M, Johnson ND, Wang Z, Jiao H, et al. A comparative atlas of single-cell chromatin accessibility in the human brain. *Science.* 2023;382: eadf7044. doi:10.1126/science.adf7044
33. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015;518: 187–196. doi:10.1038/nature14132
34. Zeng B, Bendl J, Deng C, Lee D, Misir R, Reach SM, et al. Genetic regulation of cell type-specific chromatin accessibility shapes brain disease etiology. *Science.* 2024;384: eadh4265. doi:10.1126/science.adh4265
35. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics.* 2015;31: 3847–3849. doi:10.1093/bioinformatics/btv470
36. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv.* 2021;7: eabh2169. doi:10.1126/sciadv.abh2169
37. Liu N, Pang B, Kang L, Li D, Jiang X, Zhou C-M. TUFM in health and disease: Exploring its multifaceted roles. *Front Immunol.* 2024;15: 1424385. doi:10.3389/fimmu.2024.1424385
38. Le Grand Q, Satizabal CL, Sargurupremraj M, Mishra A, Soumaré A, Laurent A, et al. Genomic Studies Across the Lifespan Point to Early Mechanisms Determining Subcortical Volumes. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2022;7: 616–628. doi:10.1016/j.bpsc.2021.10.011
39. Femminella GD, Ninan S, Atkinson R, Fan Z, Brooks DJ, Edison P. Does Microglial Activation Influence Hippocampal Volume and Neuronal Function in Alzheimer’s Disease and Parkinson’s Disease Dementia? *J Alzheimers Dis.* 2016;51: 1275–1289. doi:10.3233/JAD-150827
40. Zhang J, Lachance V, Schaffner A, Li X, Fedick A, Kaye LE, et al. A Founder Mutation in VPS11 Causes an Autosomal Recessive Leukoencephalopathy Linked to Autophagic Defects. *PLoS Genet.* 2016;12: e1005848. doi:10.1371/journal.pgen.1005848
41. Huang J, Huffman JE, Huang Y, Do Valle Í, Assimes TL, Raghavan S, et al. Genomics and phenomics of body mass index reveals a complex disease network. *Nat Commun.* 2022;13: 7973. doi:10.1038/s41467-022-35553-2
42. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med.* 2020;17: e1003062. doi:10.1371/journal.pmed.1003062

43. Meer D van der, Frei O, Kaufmann T, Shadrin AA, Devor A, Smeland OB, et al. Understanding the genetic determinants of the brain with MOSTest. *Nat Commun.* 2020;11: 3512. doi:10.1038/s41467-020-17368-1
44. Koskeridis F, Evangelou E, Said S, Boyle JJ, Elliott P, Dehghan A, et al. Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat Commun.* 2022;13: 6939. doi:10.1038/s41467-022-34688-6
45. Zhu Z, Guo Y, Shi H, Liu C-L, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol.* 2020;145: 537–549. doi:10.1016/j.jaci.2019.09.035
46. Wu Y, Qi T, Wray NR, Visscher PM, Zeng J, Yang J. Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes. *Cell Genomics.* 2023;0. doi:10.1016/j.xgen.2023.100344
47. Shabalín AA. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28: 1353–1358. doi:10.1093/bioinformatics/bts163
48. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526: 68–74. doi:10.1038/nature15393
49. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26: 2336–2337. doi:10.1093/bioinformatics/btq419
50. Li X, Lappalainen T, Bussemaker HJ. Identifying genetic regulatory variants that affect transcription factor activity. *Cell Genomics.* 2023;0. doi:10.1016/j.xgen.2023.100382