Title

Multivariate genome-wide association studies on the tissue compartments of human brain identify novel loci underpinning brain development and neuropsychiatric outcomes.

Authors

Chun Chieh Fan[1,2,3*], Robert Loughnan[4], Carolina Makowski[2,3], Diliana Pechva[2,3], Chi-Hua Chen[3], Donald Hagler[2,3], Wesley K. Thompson[1,3], Dennis van der Meer[5,6], Oleksandr Frei[5,7], Ole Andreassen[5], Anders M. Dale[2,3,4,8]

1. Population Neuroscience and Genetics Lab, University of California, San Diego, USA.

2. Center for Multimodal Imaging and Genetics, University of California, San Diego, USA

3. Department of Radiology, School of Medicine, University of California , San Diego, USA

4. Department of Cognitive Science, University of California , San Diego, USA

5. NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway

6. School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands

7. Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway

8. Department of Neuroscience, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92037, USA

* corresponding author:

Chun Chieh Fan, MD, PhD

c9fan@ucsd.edu

Abstract

It is important to understand the molecular determinants for microstructures of human brain. However, past genome-wide association studies (GWAS) on microstructures of human brain have had limited results due to methodological constraints. Here, we adopt advanced imaging processing methods and multivariate GWAS on two large scale imaging genetic datasets (UK Biobank and Adolescent Brain Cognitive Development study) to identify and validate key genetic association signals. We discovered 503 unique genetic loci that explained more than 50% of the average heritability across imaging features sensitive to tissue compartments. The genome-wide signals are strongly overlapped with neuropsychiatric diseases, cognitive functions, risk tolerance, and immune responses. Our results implicate the shared molecular mechanisms between tissue microstructures of brain and neuropsychiatric outcomes with astrocyte involvement in the early developmental stage.

Main

As a circuitry, human cortical and subcortical structures are coordinately developed and integrated for proper functioning. The underpinning molecular processes work as chemical gradients across brain regions through guiding the migration of neurons, promoting the cell differentiations, and forming of synapses. It is hard to imagine this fine-tuned developmental process would be independent from one anatomical structure to the other. However, current genome-wide association studies (GWAS) on brain imaging measurements often adopt univariate approach, associating one brain region at a time. This type of univariate GWAS is particularly problematic for imaging data, as it ignores the widely distributed effects across brain regions and suffers from the burden of multiple comparisons.

The cell compositions and the orientations of brain tissues can provide a clue on how the circuitries were formed and supported. The diffusion magnetic resonance imaging (dMRI) is designed to measure the properties for those so-called microstructures of human brain. Although GWAS on the dMRI have yield important insight into the molecular mechanisms in the integrities of white matter fiber tracts, the traditional diffusivity metrics those GWAS relied on failed to capture the properties of cell compositions and had lower sensitivities on the differences in cellular orientations. Because of these limitations, the GWAS on dMRI only focused on the diffusivity of the major fiber tracts in the cerebral white matters.

To overcome aforementioned constraints, we applied advanced imaging methods on the dMRI to derive features sensitive to tissue compartments and then performed GWAS with multivariate statistical inference, taking whole brain into consideration. We used UK Biobank (UKB) imaging data obtained before 2019 as our discovery set while the independent set of UKB (MRI performed after 2019) and data from Adolescent Brain Cognitive Development study (ABCD) were used as the validating set. While showcasing the benefits of performing multivariate GWAS on voxelwise whole brain imaging

data, we extended several methods to improve the interpretation of the multivariate results, providing insight into the biological processes of brain development and its relationships to neuropsychiatric outcomes.

Results

Multivariate genome-wide associations on features of tissue compartments across whole-brain

To extract the features of tissue compartments of human brain from MRI scans, we processed the diffusion imaging data from UKB and ABCD with the restricted spectrum imaging (RSI) [1-9], using ABCD consistent imaging pipeline [10,11]. The RSI deconvoluted the angular momentums of the multiple diffusion signals, deriving three main features of tissue compartments (Figure 1A). The normalized isotropic-restricted volume, N0, captures the relative amount of cell bodies within a voxel, such as the densities of neurons, astrocytes, and oligodendrocytes. The normalized directional-restricted volume, ND, captures the relative amount of tubular structures within a voxel, hence highlighting the regions rich in axons and dendrites. The free water component, NF, captures the relative amount of the unhindered free water outside of cell structures.

We then performed the multivariate GWAS on N0, ND, and NF, separately. In the discovery stage (UKB discovery set, imaging scanned before 2019, n = 23,543), we implemented the combined principal components GWAS (CPCs) to identify associated loci for each feature [12,13]. Given the UKB discovery set, cPCs first calculated the principal components (PCs) from the tissue feature across all voxels (Figure 1D). First 5000 PCs were used in the subsequent analyses, explaining more than 70% of total variance of the imaging data. Since each PC is orthogonal to each other, the statistical inference is based on combining the associations between genetic variants and each of the derived PCs. A global-local combination function was further used to optimize the power [12,14]. We searched four different

global-local cut-points (Supplemental Materials) and found setting first 100 PCs as global components and the remaining as local can maximize the discovery. To reflect these experiments, we set the p-value threshold for genome-wide significance as 5e-8/12 (four cut-points, three tissue features).

The CPCs on UKB discovery set found 432, 350, and 273 independent genetic loci associated with N0, ND, and NF, respectively (Figure 1E). The multivariate SNP-heritabilities [15] for N0, ND, and NF are 0.09 (95%CI 0.04 - 0.13), 0.06 (95%CI 0.02 - 0.10), and 0.05 (95%CI 0.01 - 0.09). The discovered loci explained 59%, 60%, and 58% of the SNP-heritabilities for N0, ND, and NF. Combined together, there are 503 unique loci across all three tissue features. Of those discovered loci, 40% are shared across all features whereas 34% are associated exclusively with one tissue feature (Figure 1F).

The functional mapping and annotation of GWAS (FUMA) [16] shows those 503 unique loci highly enriched for gene ontology in neurogenesis (bonferroni corrected $p < 5e\text{-}11$). The N0 loci shows more evident enrichment in the anatomical and tube development (bonferroni corrected $p < 5e\text{-}11$). The ND loci shows more enrichment in neuron projection guidance and axon development (bonferroni corrected $p < 5e\text{-}11$). The NF loci, in contrast with N0 and ND, has ranked higher the gene sets in positive regulation of gene expression (bonferroni corrected $p < 5e\text{-}10$).

Loci validated in adults and adolescents

To validate the discovered loci in independent studies, we first calculated imaging scores [17-20] based on eigenvectors and association weights from the discovery set, and then performed the association tests between genetic variants and the derived scores (Methods). This procedure is similar to the confirmatory canonical correlation analysis except only one variant involved in each regression. We repeated the same confirmatory analysis in UKB replication set (n = 6,396, MRI scanned after 2019) and ABCD samples (n = 8,189), except adding study specific covariates and random effects for controlling

family relatedness and diverse genetic background in ABCD (Methods and Supplemental Materials). Among discovered loci, 335 (79%), 298 (85%), and 222 (81%) were replicated in the independent UKB replication set for NF, N0, and ND, after bonferroni correction. In ABCD, 106 (25%), 153 (43%), and 88 (32%) of the discovered loci can be validated despite the evident differences in age between UKB and ABCD.

To examine the degrees of overlaps between our validated loci and previously reported loci in neuroimaging GWAS, we curated the reported loci lists from the NHGRI-EBI Catalog based on keywords in "brain", "imaging", "cortical", and "white matter". The final list of reported loci included GWAS on brain connectivity [21], cortical surface measures [22-24], derived imaging instruments across all modalities [25], subcortical volumes [24,26,27], brain volumes [23,28,29], white matter hyperintensities [30], and white matter microstructures [31]. We queried if any of our validated loci is in linkage-disequilibrium with or located in 250kb regions of previously reported neuroimaging loci. The results are summarized in Figure 2A. Although our regional enrichment analyses indicating that majority of our validated loci have stronger signals in cerebellum, subcortical, and major fiber tracts (Figure 2B), there are substantial overlapping with loci discovered by GWAS on cortical surface measurements [22-24] (39 to 45 percent of the validated loci). This suggested wide-spread pleiotropies on the development of cortical surfaces and brain white matters, which molecular processes are working in spatial gradient rather than focalized in one single anatomical region.

Among those genetic loci overlaps with GWAS on cortical surface measurements, a locus located at the gene body of *DPYSL5* highlights the need for whole brain analyses (Figure 2C). The tagged SNP, rs11126784, was strongly associated with patterns of ND ($p = 2e\text{-}54$) and less so with N0. Previously it was reported to be associated with cortical surface measurements. However, our analyses indicate a much stronger enrichment for the cortico-striatum circuitries (Figure 2D - E). *DPYSL5*

belongs to the collapsin response mediator protein (CRMP) family, and was implicated specifically with axonal growth by converging evidence from case studies [32], molecular mapping [33], and animal studies [34]. Interestingly, all CRMP family proteins shows association signals more evidently with ND (*Dpysl2*, p = 1.3e-14; *Dpysl3*, p = 9e-7; *Dpysl4*, p = 5e-8). Our results were more aligned with the potential biological mechanism, indicating the importance of performance whole brain multivariate GWAS.

Additionally, we found 86, 61, and 41 novel loci from N0, ND, and NF, respectively (Figure 2A), encompassing genes such as *AREL1*, *YLPM1*, *DLST*, *RPS6KL1*, *EIF2B2*, and *MLH3* that are associated with mood instability among adults [35] (enrichment p value = 6e-5). For those loci showing consistent effects among adults and adolescents, the mapped genes include *OTOR*, *PREX1*, *CSE1L*, STAU1, *DDX27*, *ZNFX1*, *ZFAS1*, *SNORD12C*, *SNORD12B*, and *SNORD12* that implicated in perceived loneliness (enrichment p value = 9e-9). Many have been associated with immune functions, hormonal metabolism, personality traits, and cognitive functions. For example, a novel loci (Figure 2F) that shows consistent signals among adults and adolescents (UKB discovery p = 1e-29, UKB replication p = 1.7e-18, and ABCD replication p = 4.6e-8), rs6079582 locating at 20p12.1 on the gene body of *MACROD2*, has overlapping with genetic loci for Autism [36] and general cognitive ability [37]. The gene *MACROD2* was also implicated in the education attainment [38] and adventurous personality [39]. The association signals for rs6079582 show strong regional enrichment in the cerebellum white matter (Figure 2G-H), implicating its relationship with motor functions. A recent knockout mouse experiment has shown that knocking out the *Macrod1* and *Macrod2* would lead to disruption in Adenosine diphosphate ribosylation-mono-hydrolases, resulting in motor-coordination defect and age-dependent hyperactivities [40], corroborate with what we found here.

<u>Genetic overlaps with neuropsychiatric and immune-related phenotypes</u>

We next investigated how much of the genome-wide signals of three tissue features overlap with other imaging measurements [22,26,27], neuropsychiatric phenotypes [38,39,41-47], and immune disorders [48]. Type 2 Diabetes [49] and puberty timing [50] were also included. Based on a method that tailored for unsigned multivariate statistics [15], we evaluated the signal shared between two traits using their summary statistics. The amount of shared signals was the Spearman correlation of the average SNP *p* values within each approximately independent LD blocks. The results were summarized in the Figure 3.

Consistent with the number of discovered loci shared across three tissue features, the genome-wide signals are highly correlated among N0, ND, and NF (rho = 0.88 ~ 0.90). They are highly correlated with the genome-wide signals of cortical surface measurements, but less so for subcortical volumes. All three tissue features consistently show significant overlaps with immune disorders [48], schizophrenia [45], attention deficits hyperactivity disorder [51], bipolar disorders [43], cross-psychiatric-disorders [41], and Alzheimer's disease [47] ($p < 0.05$ after Bonferroni correction). Education attainment [38] and risk-related behaviors [39] are also significantly correlated, indicating the relevancy to general cognitive capacities.

Cell-type enrichment analysis

Although N0, ND, and NF were designed to capture the properties of tissue compartments, the strong overlapping signals across three features makes us wonder whether the same cell populations are driving the signals. To investigate this, we analyzed the heritability enrichment given cell-type annotations using the stratified LD score regression (LDSC) [52]. The dimensionally corrected multivariate statistic, as the scaled χ^2, can be used in the context of LDSC for deriving the relative enrichment in the average heritability of the high-dimensional phenotypes [15]. Hence, we run the stratified LDSC with tissue

specific chromatin annotations [52] and cell-type specific annotations [53]. Figure 4 demonstrates the results from the cell-type enrichment analyses.

While the overall patterns of the enrichment are similar across three tissue features, ND has the strongest enrichment signals across all activating histone markers (H3K27ac, H3K36me3, H3K4me1, H3K4me3, and H3K9ac) and DNase hypersensitivity site (Bonferroni corrected $p < 0.05$). All three features were evidently enriched in the chromatin state of fetal brain tissues and hippocampus tissues whereas ND also shows enrichment in Cingulate cortex and Substantia Nigra (Figure 4A). In terms of cell populations, the cell-type specific chromatin states indicating all three features have significant enrichment in the Embryonic dopaminergic interneurons and astrocytes (Bonferroni corrected $p < 0.05$). Meanwhile, the ND shows stronger enrichment signals in the oligodendrocytes, as expected for an imaging feature capturing the integrities of tubular structures as the myeline sheath.

Discussion

With the voxelwise diffusion data, we performed the multivariate GWAS on three main features of the tissue compartments across whole brain. In addition to numerous loci reported in previous neuroimaging GWAS, we discovered and validated many novel loci that are functionally related to neuropsychiatric outcomes, risk tolerance, and immune responses. Regardless the distinct spatial distribution and the sensitivity in capturing cell compositions across brain regions, all three tissue features show strikingly overlapped signals in terms of discovered loci, shared genome-wide signals, and cell-type specific enrichment. This indicates common molecular mechanisms for the microstructures of human brain, potentially driven by the developmental processes involving astrocytes.

The identification of the novel loci at *MACROD2* is a showcase for the power and flexibility of multivariate GWAS. Despite its improvement of statistical power over traditional univariate GWAS, the

multivariate GWAS is not as widely used, probably because the difficulty in the interpretation. Our extended methods, such as imaging scores for replication and regional enrichment analyses for effect localization, were aiming to improve the interpretability of the multivariate results. The converging evidence for the loci at 20p12.1 support the validity of our approach. The shared signals with GWAS on cognitive function and risk-taking tendencies, the effect localization on the cerebellum white matter, consistent associations across lifespan, and feature enrichment in the embryonic/fetal tissues are all corroborating with recent observations on the *Macrod1* and *Macrod2* knockout mice.

Our results implicate the involvement of astrocytes in the neuropsychiatric outcomes during developmental stages. All three tissue features have evident enrichment in the neurogenesis and cell differentiation, have enriched heritability in the compositions related to astrocytes, and have shared genome-wide signals with neuropsychiatric outcomes. It is remained to be seen how our results can be compatible with previous findings in supporting the relationships between the neuronal cells and psychiatric diseases [45,52] . As more and more single cell sequencing studies indicating the complexity of the type and function of astrocyte across brain regions, we envision a fuller picture on the relationships between astrocytes and neuropsychiatric disorders will be emerged.

Methods

UK Biobank samples

The inclusion criteria for the UKB samples were as following: individuals who had valid consent when the time we performed analyses (Dec 2020), were genetically inferred as European ancestry, and completed the neuroimaging protocols. Among individuals who were included in the analyses, we further divided samples into two groups based on when the neuroimaging was performed (before or after year 2019). We decided to use this naturally occurred cut-point instead of randomized allotment of the groups because of the best practice recommendations , avoiding potential systematic biases driven by temporally related imaging confounds . Individuals who had valid imaging data before 2019 were assigned as the discovery set (n = 23,543) and those who had valid imaging data, not before, but after 2019 were assigned as one of the validation sets (n = 6,396). The demographic information of the final selected UKB samples can be found in the Supplemental Table 1. Data from UKB is obtained under accession number 27412.

Adolescent Brain Cognitive Development study (ABCD) samples

For the purpose of replication tests, we selected the full baseline data of the ABCD samples from the public data release 3.0 (NDA DOI:10.151.54/1519007). Since ABCD was designed to recruit as many individuals with diverse ancestral background as the actual race/ethnic compositions in United States , we decided not to exclude individuals based on their genetic ancestries. Instead, we retained as much as the ABCD samples and used linear mixed effects models to control for the family relatedness and heterogeneous ancestral background . We only excluded those who did not have valid imaging and genetic data from the release 3.0, resulting in 8,189 individuals in the analyses. The demographic characteristics of the ABCD samples can be found in the Supplemental Table 1.

Imaging data processing

Both UKB and ABCD have diffusion imaging protocols that were compatible for applying RSI models. The MRI scans of UKB were performed at three scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receive head coils. The MRI scans of ABCD were collected by 21 study sites throughout the United States, with scanners from Siemens Prisma, GE 750 and Phillips 3T scanners. To harmonize those imaging data, we processed the dMRI data from UKB and ABCD using ABCD consistent imaging process pipeline implemented by ABCD Data Analysis, Informatics, and Resource Center (ABCD DAIRC). The detailed processing procedures have been published elsewhere. In short, the raw imaging data from multiple b shell dMRI were first gone through forward-reverse gradient warping, eddy current correction, and motion correction to reduce the spatial distortion and signal heterogeneities driven by the scanner differences. The corrected images were then undergone rigid body registration to a common atlas. The fiber orientation densities (FOD) were then calculated for each voxel and then the tensor information were fed into multi-channel none-linear smoothing spline registration, resulting in positional and orientational aligned voxelwise diffusion data in 2mm resolution.

The restriction spectrum imaging (RSI) modeled the diffusion signals as mixtures of spherical harmonic basis functions [2,3]. Based on the intrinsic diffusion characteristics of separable pools of water in human brain, i.e. intracellular, extracellular, and unhindered free water, RSI estimated the volume fractions of each components and their corresponding spherical harmonic coefficients. The measure of restricted isotropic diffusion (N0) is the coefficient of the zeroth order spherical harmonic coefficients, normalized by the Euclidian norm of all spherical harmonics. This feature is most sensitive to the restricted water compartment within a cell body, hence informative for the cellular compositions of human brain. The measure of restricted directional diffusion (ND) is the sum of second and fourth order spherical harmonic coefficients, normalized by the norm of all spherical harmonics. This feature is

sensitive to the restricted water compartment within tubular structures, hence informative for the axonal formation and the integrity of myeline in fiber tracts. The final feature is the normalized free water diffusion (NF). It is the zeroth order spherical harmonic coefficients for the unhindered water compartment. NF is also normalized with all other spherical harmonics, making those features unitless in 0 to 1. The tissue sensitive features, N0, ND, and NF, are particularly useful for understanding the variations of cellular compositions of human brain and highly informative for human brain development [1,4,5,9].

Genotype data processing

For UKB, we used the released v3 imputed genotype data. For ABCD, we used the public release 3.0 imputed genotype data. Both data were imputed with HRC reference panel [54]. We performed post-imputation quality controls to only allow for GWAS on common bi-allelic SNPs. We filtered SNPs have minor allele frequencies less than 0.5 percent, Hardy-Weinberg disequilibrium ($p < 1e-10$), and missingness greater than 5 percent. Genetic principal components and ancestral factors were derived using well called independent SNPs for both datasets, which were used for controlling population stratification in our analyses.

Combined principal component GWAS (CPC)

We implemented CPC for our multivariate GWAS on our UKB discovery set. CPC has been shown to be a robust multivariate GWAS methods that are well powered in detecting loci across different scenarios [12-14]. The procedures were as following. First, the principal components and their corresponding eigenvectors were derived given the voxelwise imaging data. Each SNPs was regressed on each of the derived principal component score, controlled for age, sex, 20 genetic principal components, and genotyping batches. For a given SNP, the Wald statistics for each principal component were combined as a simple linear sum (Figure 1B). Because PCs are orthonormal, the sum of the Wald

statistics follows the χ^2 distribution with k degree of freedom for k PCs combined. Although several different combination function can also be used [14], we found that the global-local combination with Fisher's method proposed in the original CPC paper has greatest power in detecting genetic loci [12]. Therefore, we experimented four different global-local cut points (50, 100, 500, and 1000 principal components) to see which combinations yielded most discoveries. To reflect this experiment, we raised the significance threshold to p value less than 4.2e-9 (corrected for 12 multiple comparisons, as 4 thresholds and 3 features were used in current study).

Replications with confirmatory imaging scoring

To perform replication test for the discovered loci, we decided to use the confirmatory imaging scoring instead of repeating the GWAS on the independent cohorts. The eigenvectors (v_k) and the regression coefficients (β_k) obtained from the discovery set were used to calculate the imaging scoring for everyone in the validating sets.

$$score = \sum \beta_k v_k x'$$

x stands for the raw imaging data. Because each PC is independent with each other, it can be easily shown the SNP regression on the imaging score is equivalent to the comparison of the consistencies of regression coefficients between discovery set and validating set (Supplemental Materials).

Regional enrichment for investigating spatial distribution across voxels

To provide more interpretability for the multivariate GWAS results, we developed a regional enrichment analyses to show which brain regions have relatively stronger signals. Most of previous imaging studies relied on re-doing the voxelwise association tests to show the effect distributions of the discovered loci. Given the distributed nature of the effect sizes among imaging measurements , we thought the voxelwise associations were not an ideal way for localizing the effects. Instead, we examined the overlaps between

association patterns and the region of interests in the co-registered anatomical atlas. The enrichment score is the probability weighted regression coefficients from CPC:

$$score = \frac{\sum_i P_i \widehat{\beta}_i}{\sum_i P_i}$$

The variance of the enrichment score was estimated by bootstrapping the association patterns from SNPs did not surpass significant threshold. We then calculate the corresponding enrichment z score and the corresponding p values. In current study, we have obtained 130 probability maps of brain regions defined in the common atlas. We applied the regional enrichment analyses on the loci that shows robust signals across adult and adolescent data.

Loci annotations, overlaps, and gene-set enrichment analyses

To annotate the identified genetic loci, we used FUMA [16] and GRanges functions in R. The SNPs have LD in 0.1 and within 250kb distance were considered as one single loci. The MAGMA was used for calculating the gene set enrichment.

Calculation of high dimensional heritability

Previous studies on the heritability of high-dimensional phenotypes indicated the average heritability is a valid way to estimate the genetic architecture of human traits [15,55]. It is equivalent of the weighted average of heritabilities across each principal components. We applied the LD score regression for each principal components, and then weighed according to the eigenvalues, deriving the average heritabilities across RSI features.

Calculation of shared genome-wide signals between two phenotypes

As proposed in other multivariate GWAS efforts[15], we calculated the average for each of the approximately independent LD blocks [56]. Spearman correlations were performed for each pair of the GWAS results.

Stratified LD score regression for cell-type specific heritability enrichment

We rescaled the χ^2 according the extended methods [15] and then used the stratified LDSC to examine the relative enrichment in the cell-type specific annotations.

References

1 McDonald, C. R. *et al.* Recovery of white matter tracts in regions of peritumoral FLAIR hyperintensity with use of restriction spectrum imaging. *AJNR Am J Neuroradiol* **34**, 1157-1163, doi:10.3174/ajnr.A3372 (2013).

2 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, doi:10.1002/hbm.21454 (2013).

3 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. *AJNR Am J Neuroradiol* **34**, 958-964, S951, doi:10.3174/ajnr.A3327 (2013).

4 McDonald, C. R. *et al.* Restriction spectrum imaging predicts response to bevacizumab in patients with high-grade glioma. *Neuro Oncol* **18**, 1579-1590, doi:10.1093/neuonc/now063 (2016).

5 Reas, E. T. *et al.* Sensitivity of restriction spectrum imaging to memory and neuropathology in Alzheimer's disease. *Alzheimers Res Ther* **9**, 55, doi:10.1186/s13195-017-0281-7 (2017).

6 Khan, U. A. *et al.* Diagnostic utility of restriction spectrum imaging (RSI) in glioblastoma patients after concurrent radiation-temozolomide treatment: A pilot study. *J Clin Neurosci* **58**, 136-141, doi:10.1016/j.jocn.2018.09.008 (2018).

7 Hope, T. R. *et al.* Diffusion tensor and restriction spectrum imaging reflect different aspects of neurodegeneration in Parkinson's disease. *PLoS One* **14**, e0217922, doi:10.1371/journal.pone.0217922 (2019).

8 Rapuano, K. M. *et al.* Nucleus accumbens cytoarchitecture predicts weight gain in children. *Proc Natl Acad Sci U S A* **117**, 26977-26984, doi:10.1073/pnas.2007918117 (2020).

9 Beck, D. *et al.* White matter microstructure across the adult lifespan: A mixed longitudinal and cross-sectional study using advanced diffusion models and brain-age prediction. *Neuroimage* **224**, 117441, doi:10.1016/j.neuroimage.2020.117441 (2021).

10 Casey, B. J. *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci* **32**, 43-54, doi:10.1016/j.dcn.2018.03.001 (2018).

11 Hagler, D. J. *et al.* Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* **202**, 116091, doi:10.1016/j.neuroimage.2019.116091 (2019).

12 Aschard, H. *et al.* Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am J Hum Genet* **94**, 662-676, doi:10.1016/j.ajhg.2014.03.016 (2014).

13 Porter, H. F. & O'Reilly, P. F. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci Rep* **7**, 38837, doi:10.1038/srep38837 (2017).

14 Liu, Z. & Lin, X. A Geometric Perspective on the Power of Principal Component Association Tests in Multiple Phenotype Studies. *J Am Stat Assoc* **114**, 975-990, doi:10.1080/01621459.2018.1513363 (2019).

15 Naqvi, S. *et al.* Shared heritability of human face and brain shape. *Nat Genet* **53**, 830-839, doi:10.1038/s41588-021-00827-w (2021).

16 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).

17 Loughnan, R. J. *et al.* Generalization of Cortical Multivariate Genome-Wide Associations Within and Across Samples. *bioRxiv*, 2021.2004.2023.441215, doi:10.1101/2021.04.23.441215 (2021).

18 Zhao, W. *et al.* Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. *Cereb Cortex* **31**, 1478-1488, doi:10.1093/cercor/bhaa290 (2021).

19 Fan, C. C. *et al.* Williams syndrome-specific neuroanatomical profile and its associations with behavioral features. *Neuroimage Clin* **15**, 343-347, doi:10.1016/j.nicl.2017.05.011 (2017).

20 Fan, C. C. *et al.* Williams Syndrome neuroanatomical score associates with GTF2IRD1 in large-scale magnetic resonance imaging cohorts: a proof of concept for multivariate endophenotypes. *Transl Psychiatry* **8**, 114, doi:10.1038/s41398-018-0166-y (2018).

21 Jahanshad, N. *et al.* Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc Natl Acad Sci U S A* **110**, 4768-4773, doi:10.1073/pnas.1216206110 (2013).

22 Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *Science* **367**, doi:10.1126/science.aay6690 (2020).

23 Hofer, E. *et al.* Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nat Commun* **11**, 4796, doi:10.1038/s41467-020-18367-y (2020).

24 van der Meer, D. *et al.* Understanding the genetic determinants of the brain with MOSTest. *Nat Commun* **11**, 3512, doi:10.1038/s41467-020-17368-1 (2020).

25 Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216, doi:10.1038/s41586-018-0571-7 (2018).

26 Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229, doi:10.1038/nature14101 (2015).

27 Satizabal, C. L. *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet* **51**, 1624-1636, doi:10.1038/s41588-019-0511-y (2019).

28 Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* **44**, 552-561, doi:10.1038/ng.2250 (2012).

29 Zhao, Q. *et al.* Adolescent alcohol use disrupts functional neurodevelopment in sensation seeking girls. *Addict Biol*, e12914, doi:10.1111/adb.12914 (2020).

30 Persyn, E. *et al.* Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun* **11**, 2175, doi:10.1038/s41467-020-15932-3 (2020).

31 Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry*, doi:10.1038/s41380-019-0569-z (2019).

32 Jeanne, M. *et al.* Missense variants in DPYSL5 cause a neurodevelopmental disorder with corpus callosum agenesis and cerebellar abnormalities. *Am J Hum Genet* **108**, 951-961, doi:10.1016/j.ajhg.2021.04.004 (2021).

33 Hamdan, H. *et al.* Mapping axon initial segment structure and function by multiplexed proximity biotinylation. *Nat Commun* **11**, 100, doi:10.1038/s41467-019-13658-5 (2020).

34 Brot, S. *et al.* CRMP5 interacts with tubulin to inhibit neurite outgrowth, thereby modulating the function of CRMP2. *J Neurosci* **30**, 10639-10654, doi:10.1523/JNEUROSCI.0059-10.2010 (2010).

35 Ward, J. *et al.* The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. *Mol Psychiatry* **25**, 3091-3099, doi:10.1038/s41380-019-0439-8 (2020).

36 Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* **19**, 4072-4082, doi:10.1093/hmg/ddq307 (2010).

37 Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* **9**, 2098, doi:10.1038/s41467-018-04362-x (2018).

38 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121, doi:10.1038/s41588-018-0147-3 (2018).

39 Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet* **51**, 245-257, doi:10.1038/s41588-018-0309-3 (2019).

40 Crawford, K., Oliver, P. L., Agnew, T., Hunn, B. H. M. & Ahel, I. Behavioural Characterisation of. *Cells* **10**, doi:10.3390/cells10020368 (2021).

41 plee0@mgh.harvard.edu, C.-D. G. o. t. P. G. C. E. a. & Consortium, C.-D. G. o. t. P. G. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e1411, doi:10.1016/j.cell.2019.11.020 (2019).

42 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).

43 Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).

44 Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920-927, doi:10.1038/s41588-018-0151-7 (2018).

45 Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).

46 Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681, doi:10.1038/s41588-018-0090-3 (2018).

47 Wightman, D. P. *et al.* Largest GWAS (N=1,126,563) of Alzheimer's Disease Implicates Microglia and Immune Cells. *medRxiv*, 2020.2011.2020.20235275, doi:10.1101/2020.11.20.20235275 (2020).

48 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261, doi:10.1038/ng.3760 (2017).

49 Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).

50 Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834-841, doi:10.1038/ng.3841 (2017).

51 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).

52 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).

53 Hook, P. W. & McCallion, A. S. Leveraging mouse chromatin data for heritability enrichment informs common disease architecture and reveals cortical layer contributions to schizophrenia. *Genome Res* **30**, 528-539, doi:10.1101/gr.256578.119 (2020).

54 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).

55 Ge, T. *et al.* Multidimensional heritability analysis of neuroanatomical shape. *Nat Commun* **7**, 13291, doi:10.1038/ncomms13291 (2016).

56 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283-285, doi:10.1093/bioinformatics/btv546 (2016).

Figure 1.

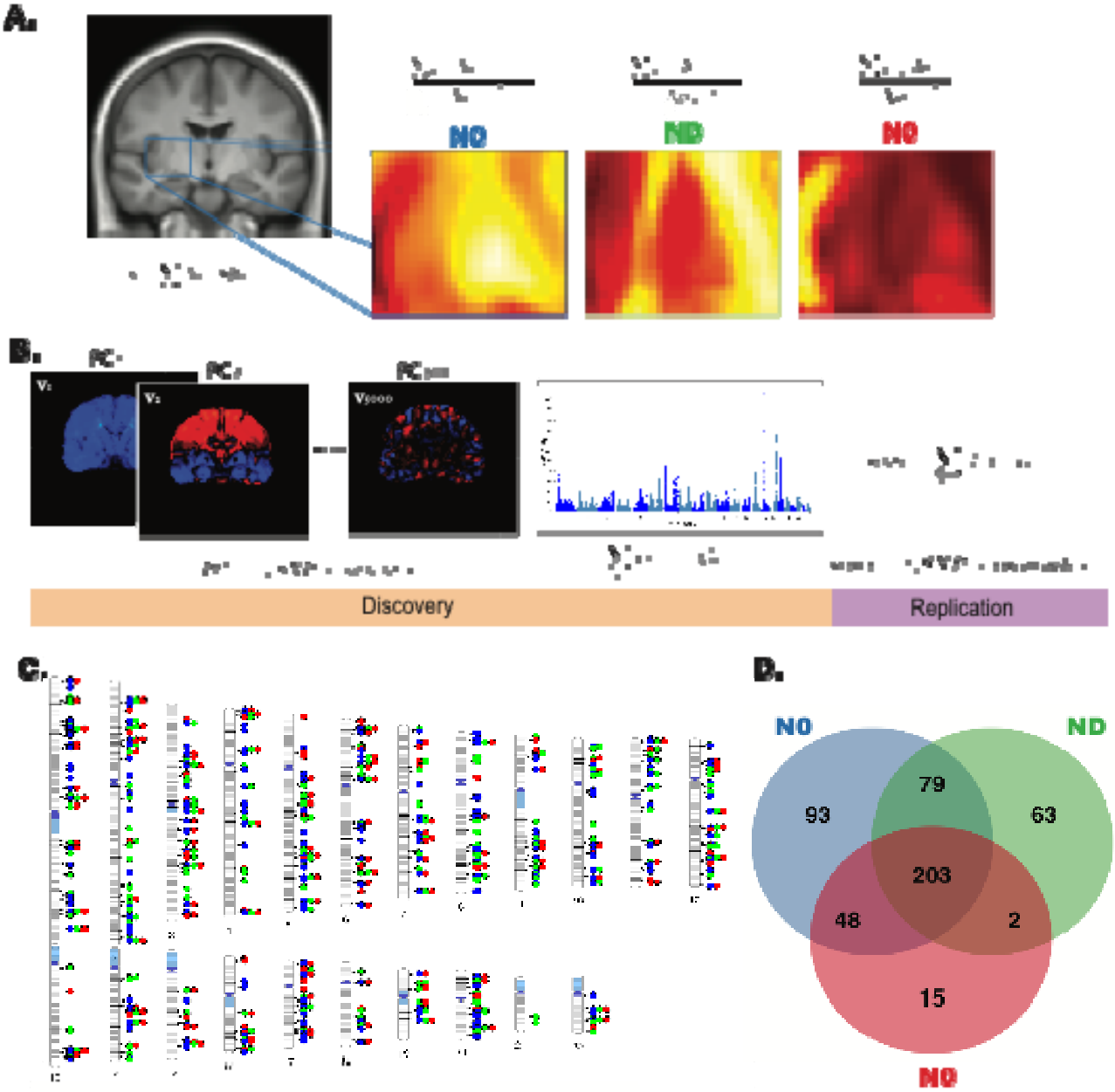
A.
NO
ND
NO
B.
Discovery
Replication
C.
D.
NO
ND
NO
93
79
63
203
48
2
15

Figure 2.

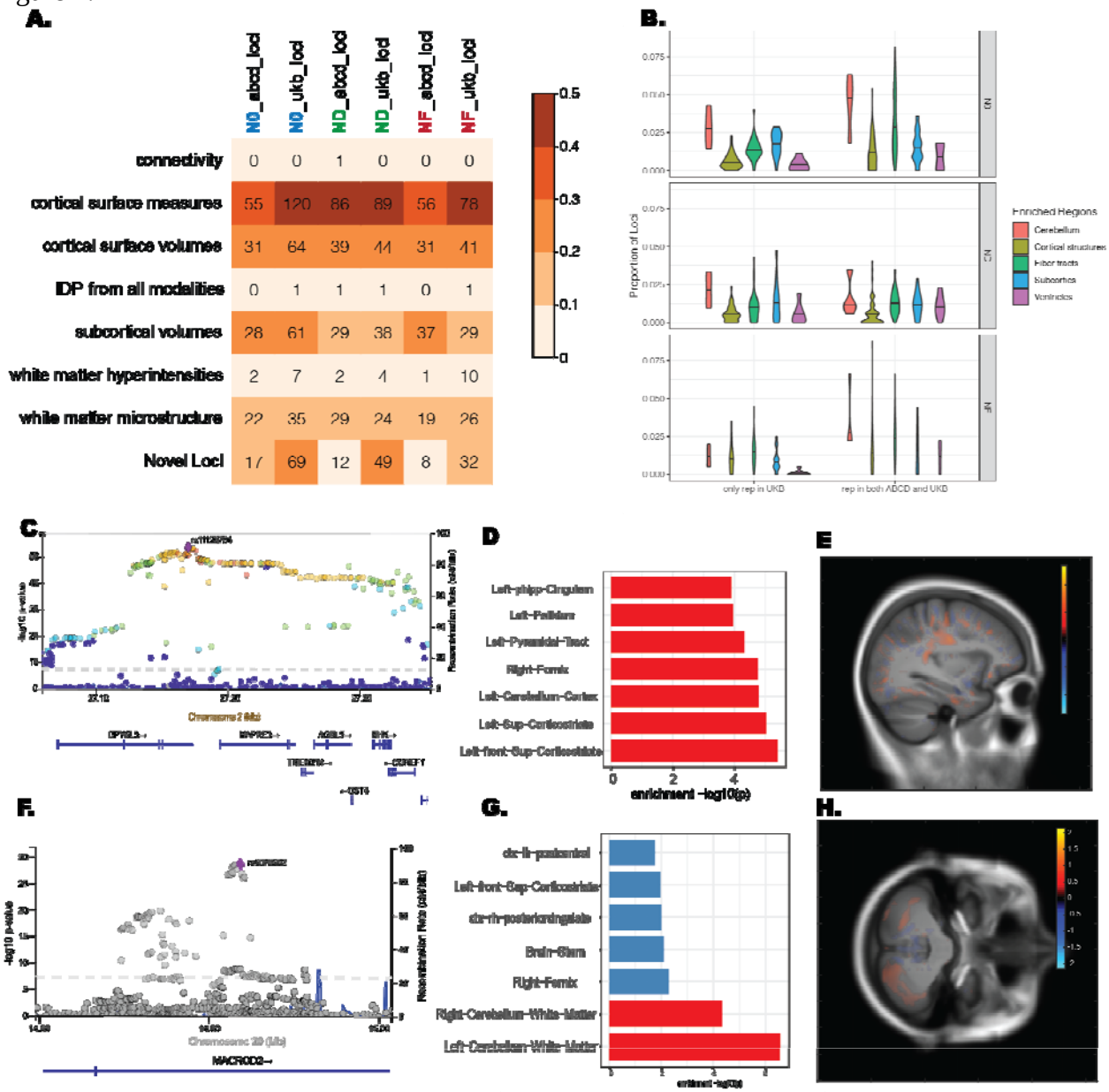
A.
N0_abcd_loci
N0_ukb_loci
ND_abcd_loci
ND_ukb_loci
NF_abcd_loci
NF_ukb_loci
connectivity 0 0 1 0 0 0
cortical surface measures 55 120 86 89 56 78
cortical surface volumes 31 64 39 44 31 41
IDP from all modalities 0 1 1 1 0 1
subcortical volumes 28 61 29 38 37 29
white matter hyperintensities 2 7 2 4 1 10
white matter microstructure 22 35 29 24 19 26
Novel Loci 17 69 12 49 8 32
0.5
0.4
0.3
0.2
0.1
0
B.
Proportion of Loci
Enriched Regions
Cerebellum
Cortical structures
Fiber tracts
Subcortics
Ventricles
only rep in UKB
rep in both ABCD and UKB
C.
D
Left-phipp-Cingulum
Left-Pallidum
Left-Pyramidal-Tract
Right-Fornix
Left-Cerebellum-Cortex
Left-Sup-Corticostriate
Left-front-Sup-Corticostriate
enrichment -log10(p)
E
F.
MACROD2
G.
dx-lh-postcentral
Left-front-Sup-Corticostriate
dx-rh-posteriorcingulate
Brain-Stem
Right-Fornix
Right-Cerebellum-White-Matter
Left-Cerebellum-White-Matter
H.

Figure 3.

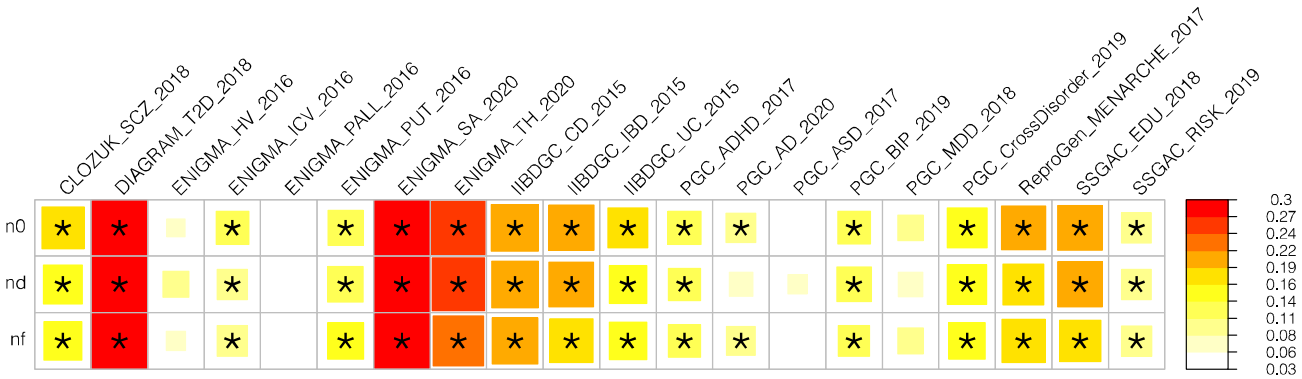
CLOZUK_SCZ_2018
DIAGRAM_T2D_2018
ENIGMA_HV_2016
ENIGMA_ICV_2016
ENIGMA_PALL_2016
ENIGMA_PUT_2016
ENIGMA_SA_2020
ENIGMA_TH_2020
IIBDGC_CD_2015
IIBDGC_IBD_2015
IIBDGC_UC_2015
PGC_ADHD_2017
PGC_AD_2020
PGC_ASD_2017
PGC_BIP_2019
PGC_MDD_2018
PGC_CrossDisorder_2019
ReproGen_MENARCHE_2017
SSGAC_EDU_2018
SSGAC_RISK_2019
n0
nd
nf
0.3
0.27
0.24
0.22
0.19
0.16
0.14
0.11
0.08
0.06
0.03

Figure 4.

A.
NO
ND
NF
DNase_Fetal_Brain_Female
DNase_Fetal_Brain_Male
H3K27ac_Brain_Angular_Gyrus
H3K27ac_Brain_Anterior_Caudate
H3K27ac_Brain_Cingulate_Gyrus
H3K27ac_Brain_Dorsolateral_Prefrontal_Cortex
H3K27ac_Brain_Hippocampus_Middle
H3K27ac_Brain_Inferior_Temporal_Lobe
H3K27ac_Brain_Substantia_Nigra
H3K36me3_Brain_Angular_Gyrus
H3K36me3_Brain_Anterior_Caudate
H3K36me3_Brain_Cingulate_Gyrus
H3K36me3_Brain_Dorsolateral_Prefrontal_Cortex
H3K36me3_Brain_Germinal_Matrix
H3K36me3_Brain_Hippocampus_Middle
H3K36me3_Brain_Inferior_Temporal_Lobe
H3K36me3_Brain_Substantia_Nigra
H3K36me3_Fetal_Brain_Female
H3K36me3_Fetal_Brain_Male
H3K4me1_Brain_Angular_Gyrus
H3K4me1_Brain_Anterior_Caudate
H3K4me1_Brain_Cingulate_Gyrus
H3K4me1_Brain_Dorsolateral_Prefrontal_Cortex
H3K4me1_Brain_Germinal_Matrix
H3K4me1_Brain_Hippocampus_Middle
H3K4me1_Brain_Inferior_Temporal_Lobe
H3K4me1_Brain_Substantia_Nigra
H3K4me1_Fetal_Brain_Female
H3K4me1_Fetal_Brain_Male
H3K4me3_Brain_Angular_Gyrus
H3K4me3_Brain_Anterior_Caudate
H3K4me3_Brain_Cingulate_Gyrus
H3K4me3_Brain_Dorsolateral_Prefrontal_Cortex
H3K4me3_Brain_Germinal_Matrix
H3K4me3_Brain_Hippocampus_Middle
H3K4me3_Brain_Inferior_Temporal_Lobe
H3K4me3_Brain_Substantia_Nigra
H3K4me3_Fetal_Brain_Female
H3K4me3_Fetal_Brain_Male
H3K9ac_Brain_Angular_Gyrus
H3K9ac_Brain_Anterior_Caudate
H3K9ac_Brain_Cingulate_Gyrus
H3K9ac_Brain_Dorsolateral_Prefrontal_Cortex
H3K9ac_Brain_Inferior_Temporal_Lobe
H3K9ac_Brain_Substantia_Nigra
B.
NO
ND
NF
CD4_T-cells
CD4_T-cells_human
CD8_T-cells
CD8_T-cells_human
Cones_(blue)
Cones_(green)
Rods
Embryonic_DA_forebrain
Embryonic_DA_midbrain
Excitatory_Camk2a
Excitatory_DG*
Excitatory_Layer_IV
Excitatory_Layer_V
Excitatory_Layer_VI
Excitatory_Layer_VI*
Excitatory_Layers_II-III
Excitatory_Layers_II-V*
Inhibitory*
Inhibitory_Gad2
Inhibitory_MSN*
Inhibitory_PV
Inhibitory_VIP
Microglia
Microglia*
NeuN_negative
Oligodendrocytes*
Astrocytes*
5.5
3.71
1.93
0.14
-1.64
-3.43
-5.21
-7