bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title:

Multivariate genome-wide association study on tissue-sensitive diffusion metrics identifies key molecular pathways for axonal growth, synaptogenesis, and astrocyte-mediated neuroinflammation

Authors

Chun Chieh Fan[1,2,3*], Robert Loughnan[4], Carolina Makowski[2,3], Diliana Pecheva[2,3], Chi-Hua Chen[3], Donald Hagler[2,3], Wesley K. Thompson[1,3], Nadine Parker[5], Dennis van der Meer[5,6], Oleksandr Frei[5,7], Ole A. 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, 9500 Gilman Drive, La Jolla, CA 92037, USA

*Corresponding author:

Chun Chieh Fan, MD, PhD

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

c9fan@ucsd.edu

Abstract

The molecular determinants of tissue composition of the human brain remain largely unknown. Recent genome-wide association studies (GWAS) on this topic have had limited success due to methodological constraints. Here, we apply advanced whole-brain analyses on multi-shell diffusion imaging data and multivariate GWAS to two large scale imaging genetic datasets (UK Biobank and the Adolescent Brain Cognitive Development study) to identify and validate 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. We identified key molecular pathways involved in axonal growth, astrocyte-mediated neuroinflammation, and synaptogenesis during development. Our results provide critical implications for potential targets for pharmacological intervention on neuropsychiatric outcomes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Main

The human brain develops through complex yet carefully orchestrated neurobiological processes, whereby cortical and subcortical circuitries are integrated for proper functioning[1]. Neural migration, axonal guidance, and synapse formation are coordinated through spatially distributed molecular gradients spanning across several brain regions [2]. Differences in tissue composition are the end result of these developmental processes. We can gain substantial insight into how neural circuitries were formed and supported by investigating the genetic determinants of whole brain patterning with respect to tissue composition.

Recent advances in multi-shell diffusion magnetic resonance imaging (ms-dMRI) and diffusion signal modeling have created an opportunity to evaluate tissue composition *in vivo* [3-7]. Differences in diffusion signals between water molecules of intracellular, extracellular, and unhindered compartments are captured by ms-dMRI, allowing for the estimation of the relative proportions of cell bodies, axonal fibers, and interstitial fluids within a voxel [3-5,8-12]. This imaging modality has been used to detect compositional changes driven by neurodegeneration [8,11], development [3], obesity [4], and carcinogenesis [9,10]. However, there is currently no genome-wide association study (GWAS) on compositional features derived from ms-dMRI. This omission is critical, as traditional imaging measurements are insensitive to neurite density, short-range fibers, and cellular properties of cortical gray matter and subcortical nuclei[7].

Moreover, GWAS of brain imaging measurements usually adopt a univariate approach, performing associations with one brain region at a time [13-18]. Patterns encompassing the whole brain have been mostly ignored or controlled away as global effects, potentially biasing associations toward purely regional effects. This risks misattributing the nature of genetic effects on the brain, e.g., cortical surface area is driven by local cortical expansion when it may instead be due to underlying axonal growth. The univariate region-of-interest approach may also be underpowered to detect the full extent of

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genetic variants associated with canonical neurodevelopmental pathways, especially when effects are spatially distributed [19,20]. A multivariate GWAS, focused on detecting loci that have effects across multiple brain regions, has been shown to be highly efficient in discovering many loci [21-23]. Nevertheless, to gain biological insight, there is a need to extend these methods to enable meaningful interpretation of results.

Here, we performed a multivariate GWAS on the metrics derived from ms-dMRI to examine the genetic determinants of whole brain patterning of cellular compartments. Using two largest extant imaging genetic studies that have compatible ms-dMRI scans, the UK Biobank[24] (UKB) and the Adolescent Brain Cognitive Development Study (ABCD Study®) [25,26], we identified 503 unique loci for tissue sensitive diffusion metrics. The discovered loci were enriched for neurogenesis, neuron differentiation, and axonal development. Among the replicated loci, 152 have not been reported previously by GWAS of brain imaging phenotypes. By investigating the spatial distribution of the associated effects, we identified molecular pathways involved in neuroinflammation and axonal growth. Signal overlaps at both the locus level and genome-wide with neuropsychiatric outcomes indicate the functional relevance of our GWAS results, providing grounds for further understanding the biological underpinnings of neuropsychiatric disorders and potential pharmacological targets.

Results

Multivariate GWAS on features of tissue composition across the whole brain

We processed ms-dMRI data from UKB and ABCD with restriction spectrum imaging (RSI) to extract the tissue composition features of the human brain [3-5,8-12,27]. RSI decomposes the diffusion-weighted signals as emanating from three separable tissue compartments: intracellular, extracellular, and free water (Figure 1a). Each compartment is characterized by its intrinsic diffusion properties. In this

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

study we consider the intracellular compartment, which is defined by restricted diffusion bounded by cellular membranes, and the free water compartment characterized by the unimpeded diffusion of water molecules. RSI estimates the normalized isotropic restricted signal volume fraction, N0, which captures the relative amount of cell bodies within a voxel, such as the densities of neurons, astrocytes, and oligodendrocytes. The normalized directional restricted signal volume fraction, ND, captures the relative amount of tube-like structures within a voxel, such as axons and dendrites. The free water component, NF, captures the relative amount of free water outside of cell structures.

After the images were harmonized and registered to a common atlas to ensure the alignment of each voxel across subjects (see Method for detailed imaging processing pipelines [25,26]), we then performed separate voxel-wise multivariate GWAS on N0, ND, and NF. In the discovery stage (UKB discovery set, imaging acquisition before 2019, n = 23,543), we implemented a combined principal components GWAS (CPCs) to identify associated loci for each feature [22,28] (Figure 1b). Using the UKB discovery set, CPCs calculated the principal components (PCs) from the tissue feature across all voxels. From the whole-brain images in 2 mm resolution per voxel, spanning across 100 by 100 by 130 voxels, the first 5000 PCs were extracted and used in the subsequent analyses, explaining more than 70% of total variance of the imaging data (Extended Data Figure 1). Since all PCs are orthogonal to each other, the statistical inference can be based on combining the associations between genetic variants and each of the derived PCs (Figure 1b). We tuned the hyper-parameters for the combination function to optimize the power for discovery [19,22] by searching through four possible combination sets (see Methods). To account for hyper-parameter tuning and the three tissue features, we set the p-value threshold for genome-wide significance as 5e-8 divided by 12 = 4.2e-9.

After Linkage-disequilibrium pruning (LD $R^2 > 0.1$) and positional clumping (distance < 250K bp), we found 432, 350, and 273 independent genetic loci associated with N0, ND, and NF, respectively

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 2a). Combined, there are 503 unique loci across all three tissue features. Of those significant loci, 40% are shared in the same genomic regions across all features whereas 34% are associated exclusively with one tissue feature (Figure 2b). The gene set analyses performed by Functional Mapping and Annotation (FUMA) [29] of the discovered loci shows each tissue feature has distinct pattern of Gene Ontology enrichment. While all tissue features were highly enriched for the Gene Ontology term of neurogenesis, N0 showed stronger enrichment in anatomical morphogenesis, while ND demonstrated more enriched in axon development, neuron projection guidance, and tangential neuronal migration (Figure 2c). This suggests that at the level of the genomic loci, modeling tissue compositions captured differential molecular effects associated with the human brain.

While the SNP heritability for each PC can be as high as 0.32 to 0.36 (Extended Data Figure 2), the multi-dimensional heritabilities [23] indicate the mean signal 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). Hence, the discovered loci explained 59%, 60%, and 58% of the average SNP-heritabilities for N0, ND, and NF.

Loci validated in adults and adolescents

To validate the discovered loci in independent studies, we first calculated imaging scores [30-33] based on eigenvectors and association weights from the discovery set, and then performed the association tests between genetic variants and the derived scores (see Methods). This procedure is similar to confirmatory canonical correlation analysis [23], except with only one variant involved in each regression. We repeated the same confirmatory analysis in the UKB replication set (n = 6,396, scanned after 2019) and ABCD samples (n = 8,189), except for including study-specific covariates and random effects controlling for family relatedness and diverse genetic background in ABCD (see Methods). Among the discovered loci, 335 (79%), 298 (85%), and 222 (81%) were replicated in the independent

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

UKB replication set for N0, ND, and NF, after Bonferroni correction for number of loci discovered. In ABCD, 106 (25%), 153 (43%), and 88 (32%) of the discovered loci were replicated for N0, ND, and NF, despite the large differences in age and other sample characteristics between UKB and ABCD (Figure 2d).

To examine the overlap 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", "subcortical", and "white matter". The final list of reported loci included GWAS on brain connectivity [15], cortical surface measures [13,21,34], derived imaging instruments across all modalities [35], subcortical volumes [14,21,36], brain volumes [16,34,37], white matter hyperintensities [38], and white matter microstructure [18]. We queried if any of our validated loci were in linkage-disequilibrium (LD) with or located in 250kb regions of previously reported neuroimaging loci. The results are summarized in Figure 2d. Among the replicated loci, 262 unique loci overlapped with previously published brain imaging GWAS, many of them found to be associated across several different imaging measurements (Figure 2d). This suggests widespread pleiotropy between the development of cortical surfaces and cerebral white matter, with molecular processes working across a spatial gradient rather than focalized in one single anatomical region.

Loci showing differential effects on tissue compositions

Closer inspection of the effect size distributions of the loci provides a unique angle into the molecular processes shaping the human brain. For instance, the 5q14.3 locus at the gene body of *VCAN*, tagged by a common SNP rs12653308, was found to be strongly associated with N0 (Figure 3a, Extended Data Table 1). It was reported to be associated with various diffusion metrics from white matter fiber tracts[18] and cortical surface measurements[21] (Figure 2a). Instead of fiber tracts or cortical

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

surface regions, we found that the association strength is particularly strong in the hippocampus bilaterally (Figure 3b, 3c), based on the regional enrichment analysis with 50,000 bootstraps (see Method). *VCAN* encodes versican, which is a lectican-binding chondroitin sulfate proteoglycan (CSPG) and serves critical roles in astrocyte-mediated neuroinflammation[39] (Figure 3d). CSPGs were found to be associated with astrocyte-dependent synaptogenesis within the hippocampus [40]. When we examined the associations between genetic variants of genes encoding CSPGs (*BCAN*, *NCAN*, and *VCAN*) and tissue features, we found N0 showed stronger association signals than ND and NF (Figure 3e). Since the effects were replicated in ABCD, our results support the early effects of astrocytic mediated processes on the human hippocampus via CSPGs. Changes in the distribution of CSPGs in the hippocampal formation were observed among patients with Schizophrenia and patients with Bipolar disorders [41,42], linking our findings to neuropsychiatric outcomes. The Drug-Gene Interaction database (DGIdb) shows *VCAN* as tier one druggable target and was found to be interacting with cyclosporine [43,44], indicating a potential path for pharmacological interventions.

The locus located at 2p23.3, tagged by rs11126784, has strong signals associated with ND (Figure 3f). This locus resides within the gene body of *DPYSL5* and has been reported to be associated with cortical surface measures [21]. Instead of cortical surface, our whole-brain multivariate GWAS indicates the effect sizes were more diffusely distributed among white matter tracts, especially within cortico-striatal circuitry (Figure 3g, 3h). *DPYSL5* belongs to the collapsin response mediator protein (CRMP) family, including *DPYSL2*, *DPYSL3*, and *DPYSL4*, which are essential for axonal growth and neurite morphogenesis [45-47] (Figure 3i). Indeed, all tagged SNPs of the CRMP family proteins show stronger association signals with ND than with N0 and NF (Figure 3j). Our results are concordant with CRMP involvement in neurodevelopment and showing that their effects can be observable among major white matter fiber bundles early on. Our findings are also relevant to neuropsychiatric outcomes, as

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CRMP has been implicated in schizophrenia and mood disorders, and can be pharmacologically intervened [48].

The 152 novel loci we discovered and replicated in this study are relevant for neuropsychiatric phenotypes and warrant further investigation (Extended Data Table 1-3). An N0-specific novel locus at 5q14.3 is within the gene body of *MEF2C*, which can influence neural progenitor cell differentiation and regulation of synaptic densities [49,50]. This locus overlaps with GWAS findings of educational attainment and intelligence [51]. Another locus at 20p12.1, on the gene body of *MACROD2*, showed consistent signals among adults and adolescents (ND: UKB discovery p = 1e-29, UKB replication p = 1.7e-18, and ABCD replication p = 4.6e-8), and has previously been linked to autism [52] and general cognitive ability [53]. The gene *MACROD2* was also implicated in educational attainment [51] and risk-taking behaviors [54].

Cell-type enrichment analysis

Although N0, ND, and NF were designed to capture different properties of tissue compartments, the strong overlapping signals across the three features indicates that similar cell processes and populations may shape all three microstructural features. To investigate this, we analyzed the heritability enrichment given cell type annotations using stratified LD score regression (LDSC) [55]. A dimensionally-corrected multivariate statistic, such 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 [23]. Hence, we ran stratified LDSC with tissue-specific chromatin annotations [55] and cell type-specific annotations [56] to obtain cell type-specific enrichment patterns for our RSI phenotypes (Figure 4).

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 sites ($P_{bon} < 0.05$). All three features were

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enriched in the chromatin state of fetal brain and hippocampal tissues whereas ND also shows enrichment in the cingulate cortex and substantia nigra (Figure 4a). With respect to cell populations, cell type-specific chromatin state analysis indicates all three features have significant enrichment in embryonic dopaminergic interneurons and astrocytes ($P_{bon} < 0.05$). Moreover, ND shows stronger enrichment signals in oligodendrocytes, as expected for an imaging feature capturing the integrity of tubular structures such as the myelin sheath.

Genetic overlap with neuropsychiatric and immune-related phenotypes

We investigated the proportion of genome-wide signals of the three tissue features which overlap with neuropsychiatric phenotypes [51,54,57-63] and immune disorders [64]. Based on a method tailored for unsigned multivariate statistics [23], we evaluated the signal shared between each pair of traits using their summary statistics. The amount of shared signal was the Spearman correlation of the average SNP $-\log_{10}$ *p* values within each approximately independent LD block. All three tissue features consistently show significant overlap with immune disorders [64], schizophrenia [61], attention deficit hyperactivity disorder [65], bipolar disorder [59], cross-psychiatric-disorders [57], and Alzheimer's disease [63] (ρ: 0.10 ~ 0.29; all $P_{bon} < 0.05$). Educational attainment [51] and risk-related behaviors [54] are also significantly correlated, indicating their relevance to general cognitive features (Extended Data Figure 4). Nevertheless, the patterns of genome-wide signals shared with neuropsychiatric phenotypes were not evidently different across three tissue features, despite the distinct patterns we observed at the locus level and cell-type specific enrichments. While the limited resolutions by the LD block can contribute to this lack of differences, the evident similarities in the genome-wide level can mean that the pleiotropic effects on neurodevelopmental traits are highly polygenic, sharing multiple loci but with different functional outputs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Using imaging features of whole brain tissue compositions, a multivariate GWAS discovered and validated 503 loci, of which 152 had not been reported in previous GWAS of neuroimaging phenotypes. Through in-depth examination of effect size distributions, we demonstrated the specific impact of molecular pathways, including CSPGs and CRMP, on the tissue composition underlying the human brain *in vivo*. Our findings are highly relevant for neuropsychiatric outcomes, including cognitive functions and psychiatric disorders. By identifying the key protein families and highlighting the susceptible brain regions through enrichment analyses, these results indicate a path to further investigate a potential target for pharmacological interventions.

In addition to these biological insights, our findings also showcase the need for novel analytic approaches in brain imaging genetics. Multivariate GWAS on whole brain phenotypes circumvents the potential "spotlight bias" that region-of-interest approaches are susceptible to. Diffuse effects across brain regions and neurobiological pathways are more easily detected with this approach, as the inference is based on the total sum of the effects. Detecting genetic loci associated with axonal growth, astrocyte-mediated neuroinflammation, and changes in the synaptic density require a more nuanced modeling approach of imaging signals. Moving beyond the metrics of structural volumes or fiber orientation enabled us to detect molecular effects on brain tissue properties, highlighting relevant biological pathways important for human brain development and neuropsychiatric outcomes.

References

1 Molnár, Z., Luhmann, H. J. & Kanold, P. O. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, doi:10.1126/science.abb2153 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2 Geschwind, D. H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633-647, doi:10.1016/j.neuron.2013.10.045 (2013).

3 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).

4 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).

5 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).

6 Pines, A. R. *et al.* Leveraging multi-shell diffusion for studies of brain development in youth and young adulthood. *Dev Cogn Neurosci* **43**, 100788, doi:10.1016/j.dcn.2020.100788 (2020).

7 Tournier, J. D., Mori, S. & Leemans, A. Diffusion tensor imaging and beyond. *Magn Reson Med* **65**, 1532-1556, doi:10.1002/mrm.22924 (2011).

8 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).

9 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).

10 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11 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).

12 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).

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

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

15 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).

16 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).

17 Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* **51**, 1637-1644, doi:10.1038/s41588-019-0516-6 (2019).

18 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).

19 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20 Dick, A. S. *et al.* Meaningful associations in the adolescent brain cognitive development study. *Neuroimage* **239**, 118262, doi:10.1016/j.neuroimage.2021.118262 (2021).

21 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).

22 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).

23 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).

24 Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523-1536, doi:10.1038/nn.4393 (2016).

25 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).

26 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).

27 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).

28 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).

29 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30 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).

31 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).

32 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).

33 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).

34 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).

35 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).

36 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).

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

38 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39 Stephenson, E. L. *et al.* Chondroitin sulfate proteoglycans as novel drivers of leucocyte infiltration in multiple sclerosis. *Brain* **141**, 1094-1110, doi:10.1093/brain/awy033 (2018).

40 Pyka, M. *et al.* Chondroitin sulfate proteoglycans regulate astrocyte-dependent synaptogenesis and modulate synaptic activity in primary embryonic hippocampal neurons. *Eur J Neurosci* **33**, 2187-2202, doi:10.1111/j.1460-9568.2011.07690.x (2011).

41 Pantazopoulos, H., Woo, T. U., Lim, M. P., Lange, N. & Berretta, S. Extracellular matrix-glial abnormalities in the amygdala and entorhinal cortex of subjects diagnosed with schizophrenia. *Arch Gen Psychiatry* **67**, 155-166, doi:10.1001/archgenpsychiatry.2009.196 (2010).

42 Shah, A. & Lodge, D. J. A loss of hippocampal perineuronal nets produces deficits in dopamine system function: relevance to the positive symptoms of schizophrenia. *Transl Psychiatry* **3**, e215, doi:10.1038/tp.2012.145 (2013).

43 Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aag1166 (2017).

44 Freshour, S. L. *et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* **49**, D1144-D1151, doi:10.1093/nar/gkaa1084 (2021).

45 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).

46 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).

47 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48 Quach, T. T., Honnorat, J., Kolattukudy, P. E., Khanna, R. & Duchemin, A. M. CRMPs: critical molecules for neurite morphogenesis and neuropsychiatric diseases. *Mol Psychiatry* **20**, 1037-1045, doi:10.1038/mp.2015.77 (2015).

49 Harrington, A. J. *et al.* MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife* **5**, doi:10.7554/eLife.20059 (2016).

50 Li, H. *et al.* Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proc Natl Acad Sci U S A* **105**, 9397-9402, doi:10.1073/pnas.0802876105 (2008).

51 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).

52 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).

53 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).

54 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).

55 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).

56 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

57 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).

58 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).

59 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).

60 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).

61 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).

62 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).

63 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).

64 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

65 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).

66 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).

67 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Overview of the study design. a. Illustration of tissue composition imaging features. The first row highlights which cellular compartments the metrics intend to capture. The second row is the formula used for calculating each metric, i.e. N0, ND, and NF (see Methods section). The third row shows the actual signal intensities for N0, ND, and NF, respectively. b. Illustration of the analytic sequence of multivariate GWAS. The discovery stage involves summarizing the whole-brain voxel-wise data into k principal components (PCs) and then performing the GWAS inference based on combined association signals across PCs. The validation stage involves replicating the findings by confirmatory associations with imaging scores.

Figure 2. Results of multivariate GWAS on whole brain imaging features. a. Ideogram of the discovered loci, colored according to the imaging features. b. Offset plot shows the unique and overlapping loci of each neuroimaging feature. c. The top-ranking enriched Gene Ontologies, derived from gene set analyses on significant loci from each imaging feature. +: positive regulation. d. Offset plot shows the unique and overlaping replicated loci with previous neuroimaging GWAS, including brain connectivity [15], cortical surface measures [13,21,34], derived imaging instruments across all modalities [35], subcortical volumes [14,16,21,34,36,37], white matter hyperintensities [38], and white matter microstructures [18].

Figure 3. Illustrations of the selected loci. a. Locus plot of the association signals with N0 in 5q14.3. b. Regional enrichment of the association signals on the human brain. The top five enriched regions are shown. c. 3D visualization of the top two enriched regions (hippocampus). d. Visual representations of the functions of CSPG. e. Association magnitudes across CSPGs, including *BCAN*, *NCAN*, and *VCAN*. f. Locus plot of the association signals with N0 in 2p23.3. g. Regional enrichment of the association

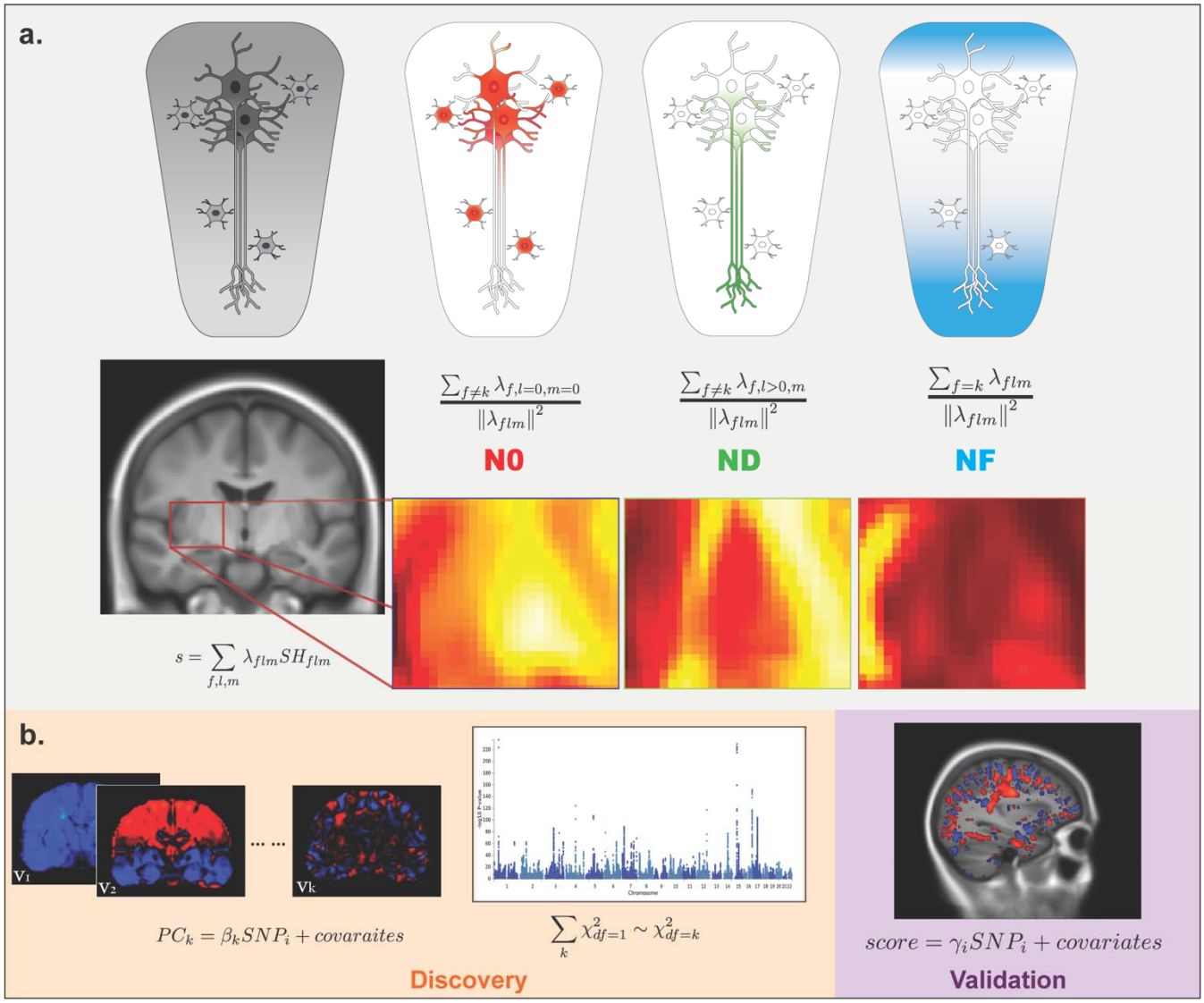
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

signals on the human brain. The top five enriched regions are shown. h. 3D visualization of the top two enriched regions (cortical striatum). i. Visual representations of the functions of CRMP. e. Association magnitudes across CRMP, including *DPYSL2*, *DPYSL3*, *DPYSL4*, and *DPYSL5*.

Figure 4. Cell-type specific enrichment results. Results from stratified LDSC analysis with dimensionally corrected effect sizes. a. tissue-specific histone markers. b. cell types.

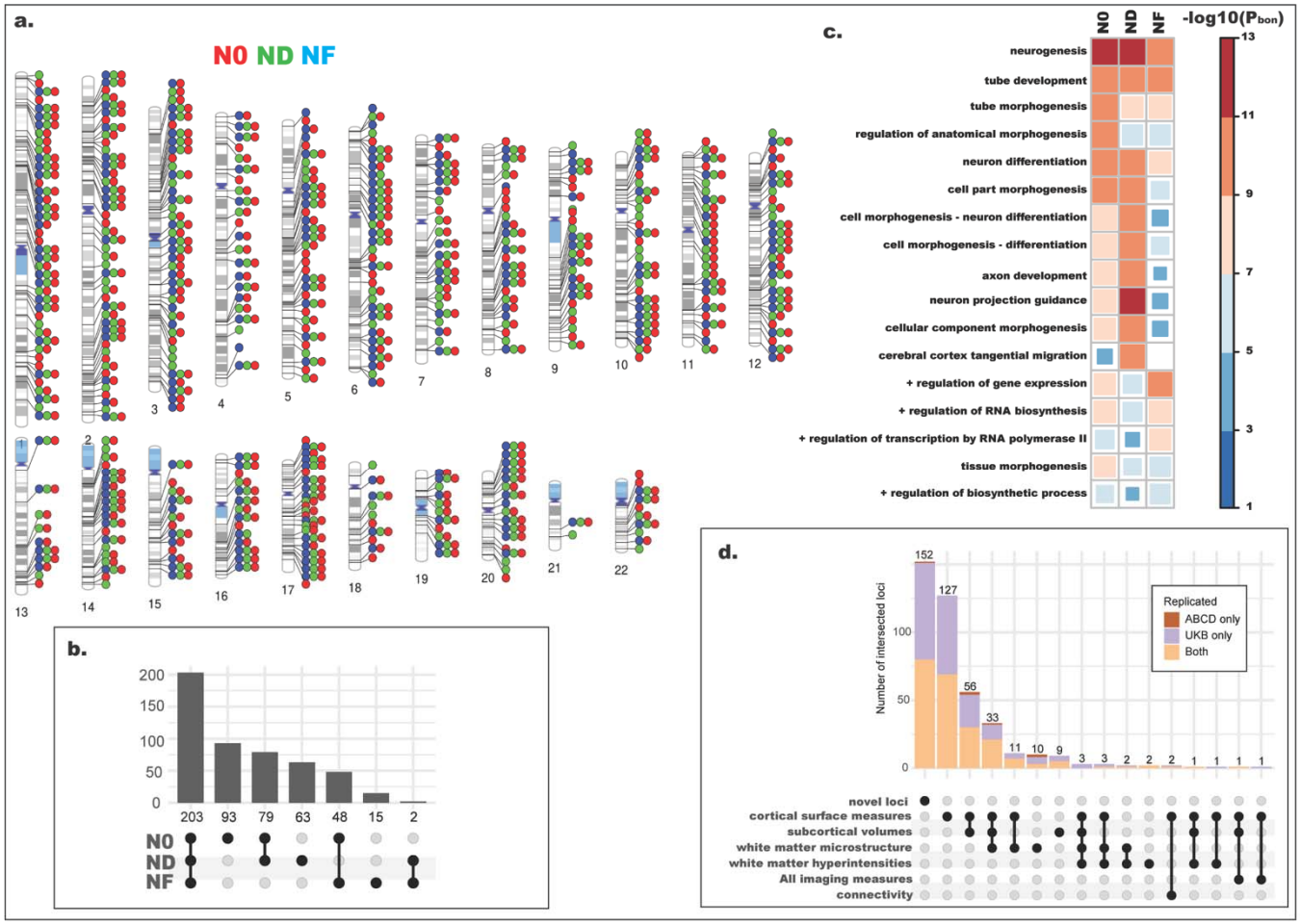
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1.



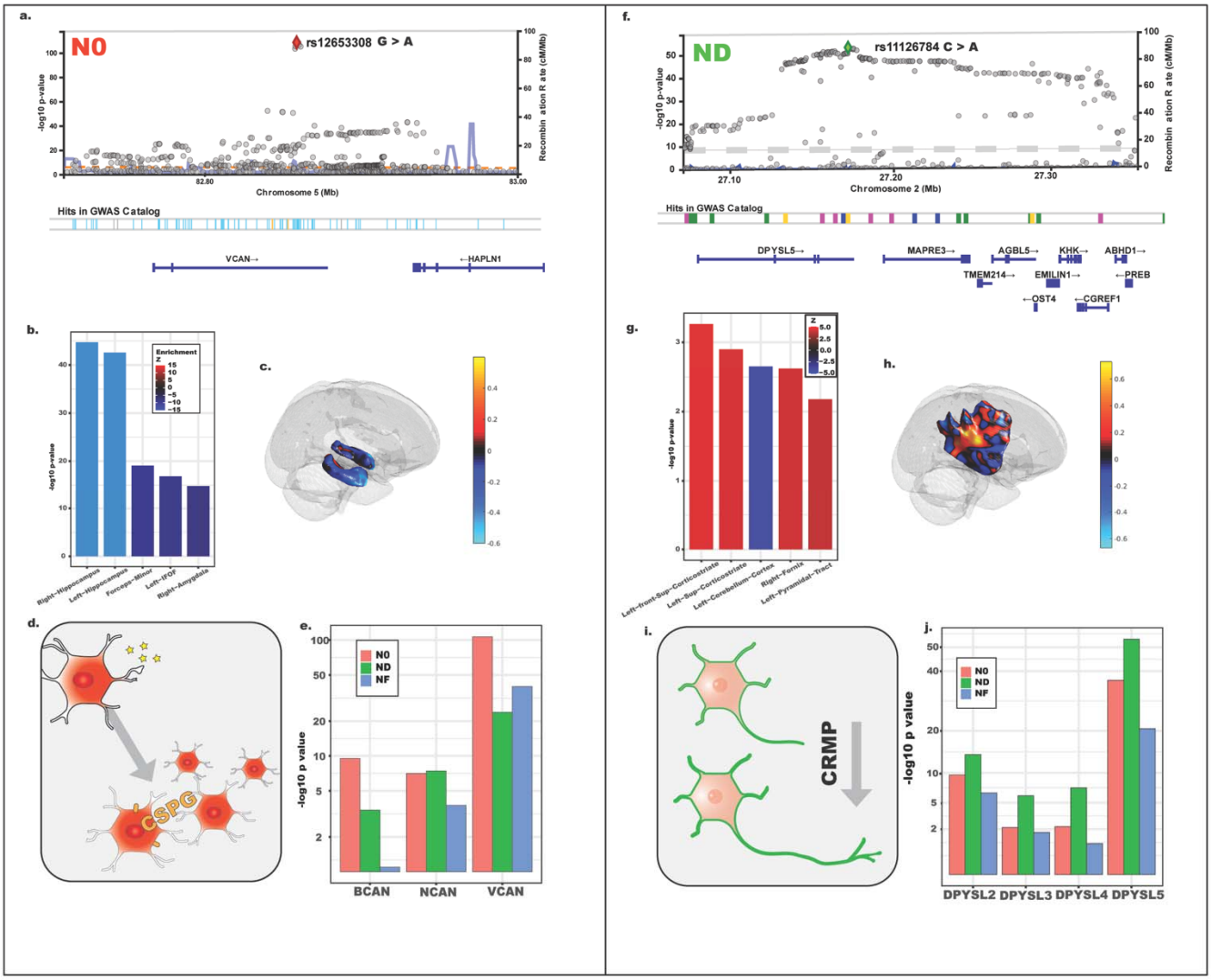
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2.



bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3.



bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

UK Biobank samples

The inclusion criteria for the UKB sample were as follows: individuals who had valid consent at the time the analyses were performed (Dec 2020), were genetically inferred as having 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 2019). We decided to use this naturally occurring cut-point instead of randomized allotment of the groups because of best practice considerations, 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 to the validation set (n = 6,396). The demographic information of the final selected UKB samples can be found in the Extended Data Table 4. Data from UKB is obtained under accession number 27412.

Adolescent Brain Cognitive Development study (ABCD) samples

For replication of results, we selected the full baseline data of the ABCD Study from public data release 3.0 (NDA DOI:10.151.54/1519007). Since ABCD was designed to recruit individuals with diverse ancestral background which reflect the racial/ethnic composition of the United States, we did not exclude individuals based on their genetic ancestries, using 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 release 3.0, resulting in 8,189 individuals in the analyses. The demographic characteristics of the ABCD samples can be found in the Extended Data Table 4.

Imaging data processing

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 in 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 the imaging data across the two studies, we processed the dMRI data from UKB and ABCD using the ABCD-consistent imaging processing pipeline implemented by the ABCD Data Analysis, Informatics, and Resource Center (ABCD DAIRC). The detailed processing procedures have been published elsewhere[25]. In short, multi-shell dMRI data acquired with seven b=0 s/mm2 frames and 96 non-collinear gradient directions, with 6 directions at b=500 s/mm2, 15 directions at b=1000 s/mm2, 15 directions at b=2000 s/mm2, and 60 directions at b=3000 s/mm2. Ms-dMRI data were first processed through forward-reverse gradient warping, eddy current correction, and motion correction to reduce the spatial distortion and signal heterogeneities driven by scanner differences. The corrected images were then aligned to a common atlas using rigid body registration [25]. The fiber orientation density (FOD) functions were then calculated for each voxel and then the tensor information was fed into multi-channel nonlinear smoothing spline registration, resulting in positional and orientational aligned voxel-wise diffusion data in 2mm resolution.

Restriction spectrum imaging (RSI) models the diffusion signals as mixtures of spherical harmonic basis functions [5,12]. Based on the intrinsic diffusion characteristics of separable pools of water in the human brain (i.e. intracellular, extracellular, and unhindered free water), RSI estimates the signal volume fractions of each compartment and their corresponding spherical harmonic coefficients. The measure of restricted isotropic diffusion (N0) is the coefficient of the zeroth order spherical harmonic coefficient, normalized by the Euclidian norm of all spherical harmonics. This feature is most sensitive to isotropically diffusing water in the restricted compartment, within cell bodies. The measure of

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 anisotropically diffusing water in the restricted compartment, within oriented structures such as axons and dendrites. The normalized free water diffusion (NF) measure is calculated as the zeroth order spherical harmonic coefficients for the unhindered water compartment. NF is also normalized by the Euclidean norm of all SH coefficients. This normalization makes the RSI features unitless and in the range of 0 to 1. N0, ND, and NF provide greater tissue specificity than the widely-used diffusion tensor metrics and are particularly useful for understanding the variations of cellular organization within the human brain and highly informative for human brain development [3,10,11,27].

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 datasets were imputed with the HRC reference panel [66]. We performed post-imputation quality control to only allow for GWAS on common bi-allelic SNPs. We filtered SNPs which 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 and 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 method that is well powered to detect loci across different scenarios [19,22,28]. The procedures were as follows. First, the principal components and their corresponding eigenvectors were derived given the voxel-wise imaging data. Each SNP was regressed on each of the derived principal component scores, controlled for age, sex, 20 genetic principal components, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genotyping batches. For a given SNP, the Wald statistics for each principal component were combined as a simple linear sum (Figure 1B). Given that PCs are orthonormal, the sum of the Wald statistics follows the χ^2 distribution with k degrees of freedom for k PCs combined. Although several different combination functions can be used [19], we found that the global-local combination with Fisher's method proposed in the original CPC paper has greatest power in detecting genetic loci [22]. Therefore, we experimented with four different global-local cut points (50, 100, 500, and 1000 principal components) to see which combinations yield the most discoveries. To reflect this experiment, we lowered the significance threshold to p <4.2e-9 (corrected for 12 multiple comparisons, as 4 thresholds and 3 features were used in the current study).

Replication with confirmatory imaging scoring

To perform the replication test for the discovered loci, we used 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 all subjects in the validation sets.

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

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

Regional enrichment for spatial distribution across voxels

To provide more interpretability for the multivariate GWAS results, we developed a regional enrichment analysis to show which brain regions have relatively stronger signals. Most previous imaging studies relied on re-doing the voxelwise association tests to show the effect distributions of the discovered loci [14,16-18,36]. Given the distributed nature of the effect sizes among imaging measurements, the voxel-wise

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associations were not an ideal way of localizing effects [20]. Instead, we examined the overlap between association patterns and regions of interest 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 that did not surpass the significance threshold. We then calculate the corresponding enrichment z-score and the corresponding p-values. In the current study, we obtained 130 probability maps of brain regions defined in the common atlas. We applied the regional enrichment analyses on the loci that showed robust signals across adult and adolescent data.

Loci annotations, overlaps, and gene-set enrichment analyses

To annotate the identified genetic loci, we used FUMA [29] and the GRanges function in R. SNPs with LD of r2<0.1 and within 250kb distance were considered as one single locus. MAGMA [67] 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 of estimating the genetic architecture of human traits [23,68]. It is equivalent to the weighted average of heritabilities across each of the principal components. We applied LD score regression for each principal component and then weighted these according to their eigenvalues, deriving the average heritabilities across RSI features.

Stratified LD score regression for cell-type specific heritability enrichment

We rescaled the χ^2 according to extended methods [23] and then used stratified LDSC to examine the relative enrichment of heritability for cell type-specific annotations.

Calculation of shared genome-wide signals between two phenotypes

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As proposed in other multivariate GWAS efforts[23], for a given summary statistics of a phenotype, we first calculated the average magnitudes of associations in each of the approximately independent LD blocks [69], deriving the unsigned polygenic signal profiles of a given trait. Spearman correlations were performed for each pair of the GWAS results, evaluating the level of overlapping in the genome-wide signals.

Acknowledgement

This work was supported by grant R01MH122688 and RF1MH120025 funded by the National Institute for Mental Health (NIMH). Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development Study (ABCD Study®) (https://abcdstudy.org), held in the NIMH Data Archive (NDA). The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers: U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147.

Author Contributions

CCF, AMD, and OA conceptualized and designed the study. CCF and RL performed the analyses. DJH and OF processed the data. CCF interpreted the results and wrote the manuscript. All other co-authors provide critical inputs for the revision of manuscripts.

Conflict of Interest

Dr. Andreassen has received speaker's honorarium from Lundbeck, and is a consultant to HealthLytix. Dr. Dale is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its Scientific Advisory

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.24.449723; this version posted July 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. and receives funding through research agreements with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies. The other authors declare no competing interests.