

Shared sets of correlated polygenic risk scores and voxel-wise grey matter across multiple traits identified via bi-clustering

^{1,2}Md Abdur Rahaman, ³Amanda Rodrigue, ³David Glahn, ^{1,2}Jessica Turner, ^{1,2}Vince Calhoun

¹Georgia Institute of Technology, Atlanta, GA, USA

²Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

³Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

Abstract—Neuropsychiatric disorders involve complex polygenic determinants as well as brain alterations. The combination of genetic inheritance and neuroimaging approaches could advance our understanding of psychiatric disorders. However, cross-disorder overlap is a current issue since psychiatric conditions share some neurogenetic correlates, symptoms, and brain effects. Exploring the impact of genetic risk on the brain across disorders could help understand commonalities across multiple psychopathologies. To do this, we first compute the linear relationship between PRS and voxel-wise grey matter volume to generate brain maps for five psychiatric and three control traits. Next, we use the biclustering approach to identify regions of the brain associated with polygenic risk scores in one or more traits. Our results demonstrate a significant overlap in brain regions connected to polygenic risk across psychiatric traits. Moreover, such brain domains are highly allied with the polygenic risk for non-psychiatric control traits. This multi-trait overlap characterizes the nonspecific relationship between neural anatomy and inherited risk factors in psychiatric conditions, and in some cases, the overlap in neural features linked to genetic risk for non-psychiatric attributes.

Clinical Relevance— This study presents biclusters of multiple psychiatric and control traits. The analysis reported various brain regions, including cerebellum, cuneus, precuneus, fusiform, supplementary motor area, that show significant correlation with polygenic risk scores across diverse groups of psychiatric conditions and non-psychiatric control traits.

I. INTRODUCTION

Idiopathic psychiatric disorders are polygenic – many genetic loci of minimal effects influence risk and, in some cases, have been linked to neuroimaging data which are thought to play a significant role in the biological foundation of brain disorders [1-3]. A considerable amount of research has been carried out to explore the genetic liability [4-7] and neural associations [8, 9] of psychiatric disorders. Researchers have recently started probing the combination of polygenic and neuroimaging modalities to explain the underlying dynamics of these illnesses [7, 10, 11]. Given the lack of common high penetrance loci, researchers have begun examining genetic background in aggregate with a polygenic risk score (PRS) - a measure of the overall genetic risk an individual carries for a disorder [10, 12, 13]. PRS association studies have already reported pivotal linkage with various diseases [14-17]. Additionally, phenome-wide PRS association studies evaluate

the alliance between genetic liability for a given trait and hundreds of diverse health outcomes [14, 18]. The evidence for neuro-genetic alliance in psychiatric conditions provokes the joint analysis of these modalities to understand the disease better. Studying the insightful shared space between neural and genetic information can potentially facilitate the research. Psychiatric abnormalities have significant overlap in both genomic coding and neural functioning. Thus, studying multiple traits can ease our investigation scheme and help researchers alleviate the trait-specific shortcomings, e.g., lack of test subjects, hardship in data collection, etc. Furthermore, inspecting multiple traits together helps to formulate more generalized and robust biomarkers for the disease. With that spirit, here, we aim to examine the association between the polygenic risk score (PRS) and the grey matter volume (GMV) of the brain. We compute the linear regression slope (beta map) between PRS and GMV in various physiological conditions. Exploring neural associations both across and between different traits due to polygenic risk could help guide future research regarding the development of more targeted treatments. This is especially needed given that current cross-disorder treatment studies [19-21] either a) largely focus on comorbidity within a single subject who expresses multiple psychiatric traits rather than the neurogenetic homogeneity across psychopathologies or b) mostly confined to a single modality (genetics or imaging).

Here, we generate beta maps for five psychiatric [schizophrenia (SZ), bipolar disorder (BPD), major depressive disorder (MDD), autism spectrum disorder (ASD), attention-deficit hyperactivity disorder (ADHD)], and three control traits [height (H), type 2 diabetes (T2D), and inflammatory bowel disease (IBD)]. The beta maps represent the voxel-wise linear slope between the PRS and grey matter volume across various subjects for each trait. Next, we employ a biclustering framework to study the similarity of beta values across a subset of traits and brain regions. Biclustering is a two-dimensional data mining technique that allows simultaneous clustering of rows (traits) and columns (voxels with beta values) of a two-dimensional data matrix [22]. Selecting biclustering is intuitive because we aim to analyze the homogeneity across two dimensions, traits and the brain regions. The biclustering procedure extracts similar beta values across the attributes by grouping them together. The objective is to approximate a concise multi-modal subspace among the physiologies. The method considers highly positive and negative beta values

separately and returns both positive and negative GMV/PRS associations. Our experimental results produced eight biclusters divided into three categories – only control traits, only disorder traits and all traits. Domains exhibiting significant GMV-PRS association include the cerebellum, cuneus, calcarine, precuneus, and supplementary motor area. Future research can use these regions to evaluate combined hypotheses on various traits. Moreover, these overlaps may tap into common mechanisms that would be relevant for more than one disorder.

II. DATA COLLECTION AND PREPROCESSING

A. Participants

The participants in this study were part of the UK Biobank (www.ukbiobank.ac.uk), a population-based prospective cohort of ~500,00 individuals, ages 40-69 years, and recruited between 2006 and 2010 throughout Great Britain. Recruitment procedures for the UK Biobank are described in this study [23]. Analyses were performed on a subsample of biobank participants [N=31,616, mean age (standard deviation) = 63.5 (7.4), 46% male] that passed our genetic and imaging quality control procedures described below.

Table 1. Discovery Sample Genome-wide Association Summary Statistics

| Disorder | Citation ¹ | Sample Size ² | Number of Overlapping Variants with UK Biobank ³ |
|----------|----------------------------|--------------------------|---|
| SZ | Ripke, S. et al. [24] | 43,456/ 33,640 | 803,029 |
| BPD | Stahl, E.A. et al. [16] | 31,358/ 20,352 | 7993,96 |
| MDD | Wray, N.R. et al. [17] | 95,680/ 43,204 | 792,206 |
| ASD | Grove, J. et al. [25] | 27,969/ 18,382 | 792,722 |
| ADHD | Demontis, D. et al. [26] | 34,194/ 19,099 | 779,137 |
| Height | Wood AR, et al. [27] | 252,149 | 699,512 |
| T2D | Bonás-Guarch S et al. [28] | 57,196/ 12,931 | 798,466 |
| IBD | Liu JZ et al. [15] | 21,770/ 12,882 | 794,364 |

¹ All GWA results were restricted to European cohorts and excluded participants from the UK Biobank.

² Sample size is noted for controls/cases where applicable.

³ Column indicates the number of variants included in the PRS calculation as weights were only assigned to variants overlapping between discovery and target samples.

B. Polygenic Risk Score (PRS) calculation

We computed the polygenic risk score (PRS) for the five psychiatric conditions and three control traits mentioned earlier with PRS-CS [29]. PRS-CS is a python-based command-line tool that implements Bayesian regression to place continuous shrinkage priors on single nucleotide polymorphism (SNP) effect sizes using genome-wide association (GWA) summary statistics and an external linkage disequilibrium (LD) reference panel (1000 Genomes Project European samples: N=503). Such a framework allows for adaptive shrinkage based on GWA signals from a discovery sample while modeling local LD patterns; LD blocks are updated jointly in a multivariate fashion, in contrast to

updating the effect sizes separately and sequentially for each marker. Thus, PRS-CS can accommodate diverse genetic architectures and avoids the need for pruning and GWA threshold selection, which can often discard valuable LD information and limit prediction accuracy in subsequent target samples [30]. Details related to GWA summary statistics used to calculate the PRS for each trait are noted in Table 1. We excluded variants with INFO < 0.8 (where possible) and filtered GWA results for ambiguous and duplicated SNPs. Details regarding blood sample ascertainment, processing, and handling, and genotyping, and imputation in the UK Biobank are described in Peakmann et al. [31] and Bycroft et al. [32]. We also removed variants if the MAF < 0.01, INFO < 0.8, HWE p-value < 1×10^{-6} , or missingness was > 0.01). We then took the intersection between the remaining subjects and those passing imaging quality control (QC). Plink v1.9 (<http://pngu.mgh.harvard.edu/purcell/plink/>) [33] was used for IBD estimation and further pruning related individuals from the sample (PI-HAT > 0.2), resulting in 31,616 individuals available for analysis. Although we limited our participants to those in the white British ancestry cohort, there is nontrivial genetic variance due to ancestry even within an ethnic group [34]. As such, we performed principal component analysis (PCA) with the LD-pruned imputed data and obtained ten ancestry PCs to adjust PRS scores for these effects. After preprocessing, PRS-CS was used to generate weights for overlapping variants between discovery GWA samples and the UK Biobank for each trait. The resulting weights were applied to the imputed genotypes in the UK Biobank to calculate the final trait-wise PRS via the PLINK score command. Last, PRS was standardized and residualized for the previously described 10 ancestry PCs.

B. Neuroimaging data acquisition and analysis

The UK Biobank Imaging Working Group designed imaging protocols (www.ukbiobank.ac.uk/expert-working-groups); the additional details are described in Alfaró-Almagro, et al.[35]. Each imaging center was equipped with a 32-channel head coil and used the following 3D MPRAGE protocol to quantify brain structure: T1-weighted, TI/TR=880/2000 ms, sagittal acquisition, resolution= $1.0 \times 1.0 \times 1.0$ mm, FOV= 208 × 256 × 256, in-plane acceleration factor=2, 4 min 54 sec scan duration. Voxel-based morphometry (VBM) analyses were performed with FMRIB Software Library v5.0.10 (FSL) [36] and included all imaging data available from the UK Biobank as of 7/13/20 (N=39,676). The image preprocessing was identical to that in Rodrigue et al. [37] and included generating a study-specific template. For quality assurance, we correlated the registered T1 image for each participant with the study-specific template and eliminated those individuals with a correlation less than 0.78 (N=1,829). A total of 31,616 participants remained after this QC procedure and after applying the exclusionary genetic filters. Voxel-wise output was used as the dependent variable in a general linear model (GLM) implemented with FSL's randomize v6.0.3 [38] using a polygenic risk score as a predictor covarying for age, sex, and scanning site.

III. METHODOLOGY

Our biclustering approach has two basic modules. First, we collect beta values for each trait and sort them using standard statistics. Then we apply biclustering on these trait-wise beta

maps. The biclustering framework adapts an N-way exploration of all possible subgroups of incoming sets. The input to the N-way search is a set of elements where each element could be a vector or matrix with an arbitrary size. All the required scripts for the algorithmic implementation and post hoc analysis are developed in MATLAB. We run the analysis on high-performance computing nodes in a distributed setting.

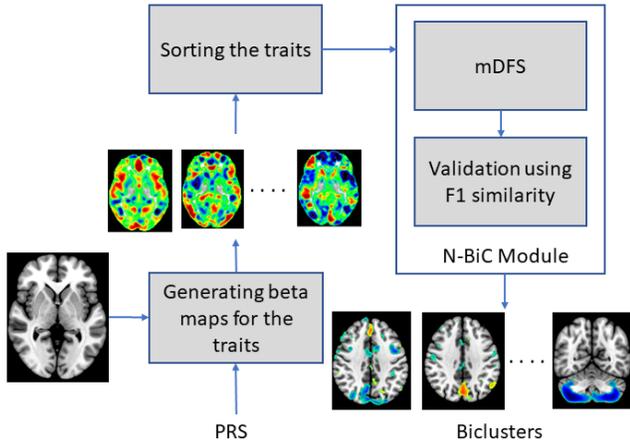


Figure 1. The architecture for biclustering the traits. It takes polygenic risk score (PRS) and gray matter volume (GMV) to create the beta maps for five psychiatric disorders and three control traits. The beta maps are sorted and thresholded for apprehending more associated brain domains. The sorted traits are sent through an N-way biclustering algorithm (N-BiC). It uses the modified depth-first search (mDFS) for exploring all possible combinations of input instances and an F1 similarity index for validation by controlling the overlaps in both dimensions.

A. Sorting the beta values

We generated a three-dimensional brain map for each trait where the voxel values denote the correlation between GMV and PRS. To make the inputs consistent for the biclustering algorithm and to reduce noise from weakly associated voxels, the submodule thresholds the beta values. For each beta map, it computes a positive mean and selects voxels with beta values \geq positive mean. Similarly, it takes the mean across the negative beta values and selects voxels with beta values \leq negative mean. This step generates a subset of voxels more strongly related to PRS, both highly correlated and anti-correlated. We also mask the beta maps to avoid ventricles and other brain regions that often show artifactual associations. We created a mask using the identical dimensions as the beta maps and assign ventricle voxels to zero. After thresholding, each trait becomes a vector of voxel indices, and each trait can contain a different number of voxels. We check multiple approach for sorting the beta-values and an exhaustive exploration that requires more time to run the algorithm.

B. N-way biclustering the traits

The main idea of biclustering voxels is to extract meaningful relationships/patterns across various subsets of traits via a depth-first-search algorithm [39]. We adopt a modified depth-first search (mDFS) based exploration technique called 'N-BiC' for checking all possible combinations of traits [40]. For each combination, the voxels are intersected across the traits to generate a common subspace. Since mDFS aims to explore all possible combinations of the input instances, the time

complexity grows exponentially with a brute-force approach. As such, the algorithm tweaks the searching technique by treating it as a graph traversal problem. It considers the inputs (list of vectors) as a set of vertices in an undirected graph. Then, mDFS seeks to create edges among those nodes by satisfying some user-defined constraints. It integrates an early abandoning technique dependent on user-defined parameters on the resulting biclusters. It stops extending the branches where it could not acquire enough homogeneous voxels after the intersection. The parameters are,

- S: List of sorted input instances (i.e., traits)
- N: Minimum number of voxels in a bicluster
- M: Minimum number of traits in a bicluster
- O: Overlap between biclusters

For every subset of input instance, mDFS evaluates an intersection between the traits and tries to augment the set for more traits depending on the validation - constraints on input parameters N, M, and O. Following a depth-first paradigm; it uses backtracking for traversing different branches of the search tree. The searching scheme depends on the feedback from the validator to abandon a branch for exploration. If the earlier iteration results in an inadequate bicluster, the algorithm stops exploring the path and backtracks to an earlier point. The validator computes the overlapping ratio with the biclusters that have already been listed. Here, we use the F1 similarity index to investigate the overlap between any earlier reported biclusters [41, 42]. The F1 similarity index is defined as follows for any two arbitrary biclusters A and B,

$$F1(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

$|A \cap B|$ = Size after intersection ; $V_{A \cap B} \times T_{A \cap B}$

$|A|$ = Size of A; i.e., the number of voxels \times traits

$|B|$ = Size of B; i.e., the number of voxels \times traits

After traversing the search space for a given set of the input stream, mDFS continues the iteration for different permutations of the input sequence. That ensures the robustness of the algorithm for any sequence of input data. We collect biclusters that show stability across the permutations. Following the thresholds determined in previous studies [40, 43], the minimum number of voxels required in a bicluster was one-fourth of the size of a beta map ($91 \times 109 \times 91$), and the minimum number of traits was two. The biclusters represent the overlap of associations between grey matter volume and polygenic risk score across a subset of traits.

IV. RESULTS

The biclusters show diverse associations between PRS and GMV patterns in the brain. Figure 2 shows eight biclusters estimated by our analysis. These clusters demonstrate cross-trait similarity in GMV-PRS association in distinct regions of the brain. The first bicluster includes two control traits, height H) and inflammatory bowel disease (IBD); we observe a positive association in the right and left cerebellum crus, right fusiform gyrus, and a strong negative association in the right cerebellum_4_5. In the second bicluster, three traits (SZ, IBD, and T2D) showed overlap in the left precuneus and calcarine

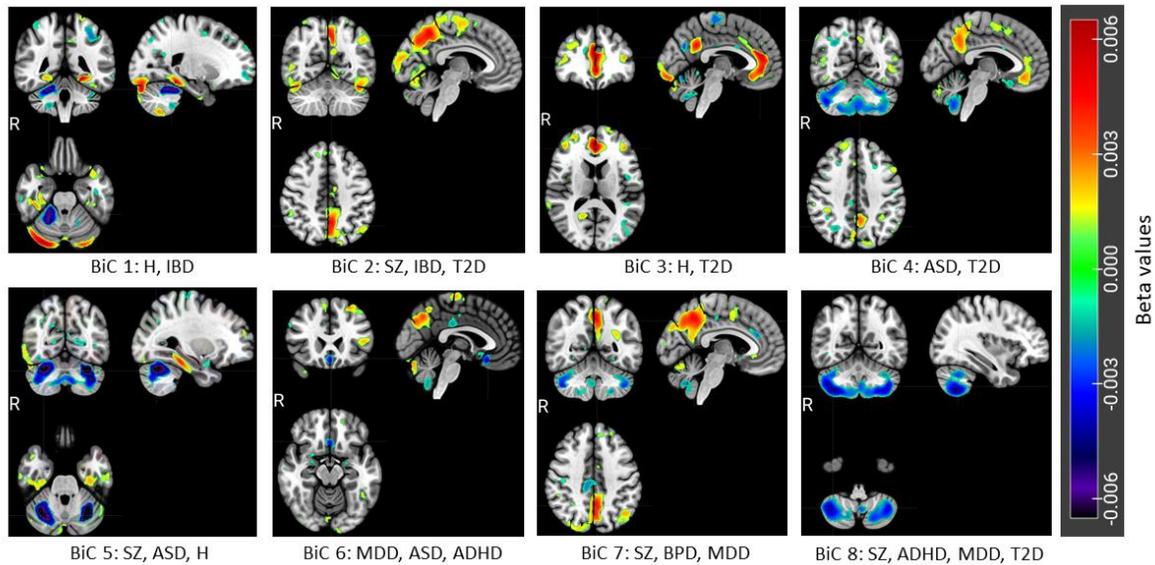


Figure 2. Results of the Bicluster Analysis. The positive and negative biclusters are extracted separately and then merged for common traits. The traits are Schizophrenia (SZ), Bipolar disorder (BPD), Major depressive disorder (MDD), Autism spectrum disorder (ASD), Attention deficit hyperactivity and impulsive disorder (ADHD), and three control traits Height (H), Type 2 diabetes (T2D), Inflammatory bowel disease (IBD). We found eight biclusters, six of which show overlap in positive and negative beta values, and two (6th and 8th) which show overlap in negative voxels only. The voxel values in the brain map of each bicluster indicate the mean betas across the traits included in the bicluster. We selected the slice with maximum beta value and overlaid it on standard template ‘spm152’ using MRICroGL for visualization. The color bar is consistent for all the maps.

sulcus (positive). The third cluster showed similarity in both left anterior and posterior cingulum in Height and T2D. Between ASD and T2D (cluster 4), the left precuneus is positively correlated with polygenic risk score (PRS), and the cerebellum crus is anti-correlated. SZ, ASD, and H overlap in cerebellum crus and vermis regions - negatively associated with the PRS (cluster 5). Bicluster 6 comprises three psychiatric traits MDD, ASD, and ADHD. These traits have similar beta values in the left cuneus, precuneus, and paracentral lobule (all positive) and left olfactory negative). Bicluster 7 also combines three psychiatric traits SZ, BPD, and MDD, including strong positive relationships in the precuneus and superior occipital regions and a negative relationship in the cerebellum crus. The last bicluster includes three psychiatric traits SZ, ADHD, MDD, and one control trait, T2D. There was a substantial overlap in the right and left cerebellum_8 (negative association). The boxplots in figure 3 show the overall distribution of beta values: GMV-PRS relationships of biclusters throughout the brain. Bicluster 2 and 8 demonstrate significantly higher positive and negative betas, respectively. Clusters 1 and 2 also showed positively skewed betas. Bicluster 4, 5, and 6 show weakly skewed negative beta values in overlapping brain domains. We further examine the domain-wise beta values of the biclusters. We compile a list of regions from all the biclusters (figure 2), showing significant GMV-PRS associations. We extracted domain domains with top 5 percent beta values from the max-slice we visualize in figure 2 using a python script in NiBabel (<https://nipy.org/nibabel/>). In figure 4, bicluster 1 and 2 show a positive correlation in cerebellum crus, where the rest of the biclusters exhibit a negative association. Cerebellum crus has been linked to divergent thinking [44], often disrupted by psychiatric conditions. The left cingulum (anterior and posterior) regions remain positively associated with almost all

biclusters except 8. The cerebellum crus and cerebellum 6 &

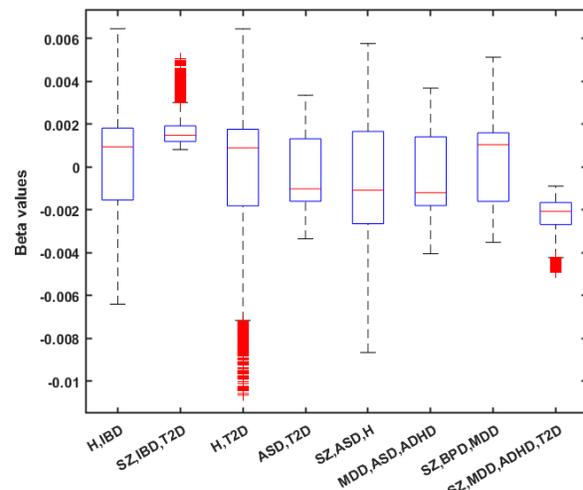


Figure 3. Boxplots for bicluster-wise beta values. The red line inside the box represents the population's mean, and the top and bottom lines demonstrate the highest and lowest values, respectively. The red data points outside the whisker are the outliers. We run a one-sample t-test with a null hypothesis of 'the mean is zero' to check whether the mean is significantly nonzero. The t-test rejects the null hypothesis for all the biclusters at a significance level of 0.001 (1%).

8 are negatively associated with the traits included in bicluster 3, 4, 5, 6, 7, and 8. The high negative correlation indicates a reverse association between grey matter volume and PRS. The cerebellum is a well-studied region of the brain in major psychiatric disorders [45, 46]. The distinctive beta relationship patterns in cerebellar subdomains explains the dynamics better in different psychiatric traits. With the prevalence of psychiatric traits in the biclusters, we observe the patterns in the remaining domains. In bicluster 5, SZ and

ASD reveal negative PRS-GMV relationships in the left calcarine sulcus, cuneus, and vermis; positive in the precuneus and right fusiform. A closer view of bicluster 6 and 7 reveal distinct beta patterns for MDD, ASD,

brain linked to one or more physical conditions. These results from biclustering are qualitatively more insightful than one-dimensional clustering. The subgroups are tightly coupled and highlight the benefits of focusing on homogeneous subsets

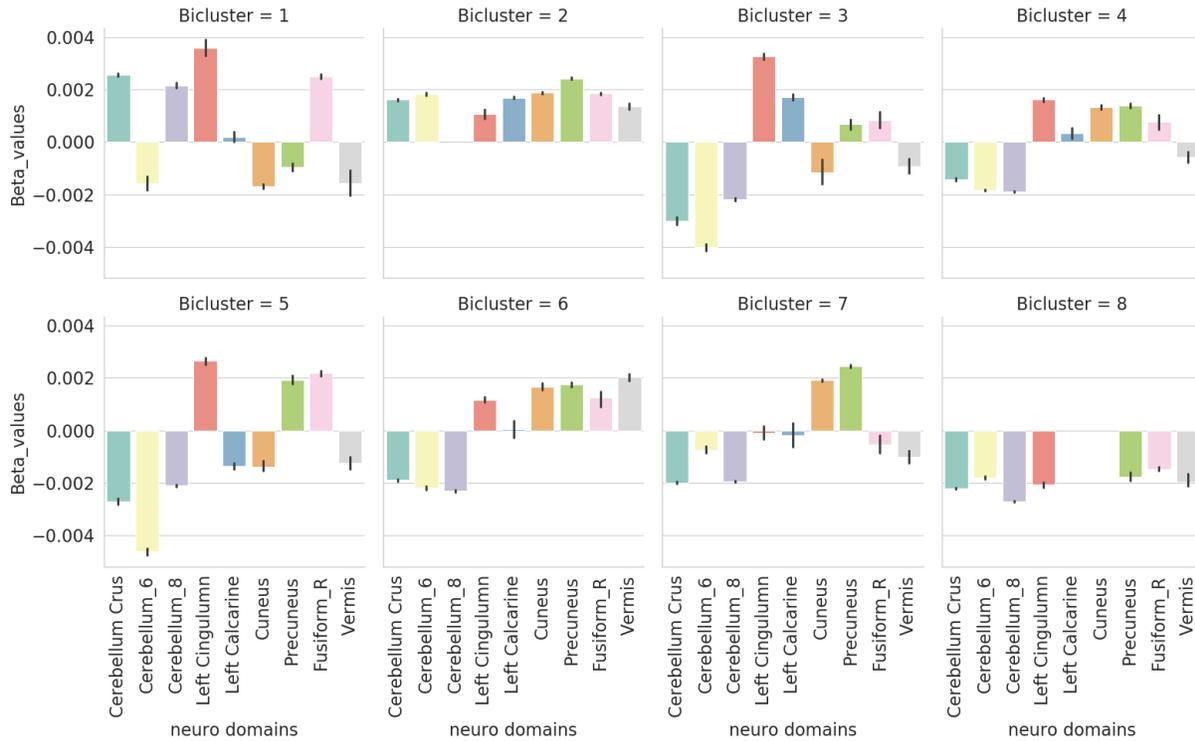


Figure 3. The bar plots for domain-wise beta values of all the biclusters. The X-axis shows nine different domains with distinct colors, and the y-axis stands for the beta values in these regions. Error bars are standard deviations. The domains are determined from the top 5 beta homogeneity across the biclusters. We summarize all the subdomains into nine broader domains namely Cerebellum crus (subdomains: crus_1_L, crus_1_R, crus_2_L, crus_2_R), Cerebellum_6 (left & right), Cerebellum_8 (left & right), left Cingulum (anterior & posterior), left Calcarine, Cuneus (left & right), Precuneus (left and right), right Fusiform, Vermis (6, 7, 8, 9). The name and domain index consistent with the ‘aal’ atlas in MRICroGL (<https://www.nitrc.org/projects/mricrogl/>). Used NiBabel (<https://nipy.org/nibabel/>) neuroimaging python package to resample the atlas to our image dimension and extract domain-wise indices.

and ADHD – all positive except cerebellum, left calcarine is zero. For SZ, BPD, and MDD, the GMV-PRS correlation pattern is organized as almost no association in the cingulum and calcarine; positive in cuneus and precuneus; negative in right fusiform and vermis. The vermis, precuneus, fusiform gyrus, and cingulum brain regions are affiliated with a variety of complex functions, including cognition, memory processing, facial recognition, etc. [46-50]. Therefore, the changes in those regions result in different psychiatric attributes of a person. The last subgroup's overlaps are mostly anti-correlated with PRS. The strength of this overlap is also consistent across domains (bicluster 8).

V. CONCLUSION

We implement a biclustering framework for analyzing the relation between the polygenic risk score and the grey matter volume of the brain in multiple psychiatric and control conditions. The contributions of this study are two-folded. First, we propose betas for capturing the neurogenetic association for a set of psychiatric and control traits, and secondly, bicluster them to identify similar betas among those traits throughout the brain. The beta value represents how likely a voxel is connected to the risk score of the disorder. So, the biclusters characterize the most relevant regions of the

of data across multiple disorders and brain regions. The results also reveal the level of association between the grey matter and PRS in a distinct subgroup of psychiatric and control traits. Future studies are needed to evaluate the implications of these results and their links with other behavior and clinical measures. A promising future direction of this project is to improve the runtime by reducing the search space. There are also benefits to making the approach permutation invariant by modifying the traversal process in the backtracking step. Future work will focus on analysis of the PRS values directly.

VI. COMPLIANCE WITH ETHICAL STANDARDS

Informed consent was obtained from each participant prior to scanning in accordance with the Internal Review Boards of corresponding institutions.

REFERENCES

- [1] Durston, S. and K. Konrad, *Integrating genetic, psychopharmacological and neuroimaging studies: A converging methods approach to understanding the neurobiology of ADHD*. Developmental Review, 2007. 27(3): p. 374-395.
- [2] Glahn, D.C., P.M. Thompson, and J. Blangero, *Neuroimaging endophenotypes: strategies for finding genes influencing brain*

- structure and function. *Human brain mapping*, 2007. **28**(6): p. 488-501.
- [3] Heck, A., et al., *Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity*. *Neuron*, 2014. **81**(5): p. 1203-1213.
- [4] Aguirre, M., et al., *Polygenic risk modeling with latent trait-related genetic components*. *European Journal of Human Genetics*, 2021: p. 1-11.
- [5] Cullen, H., et al., *Polygenic risk for neuropsychiatric disease and vulnerability to abnormal deep grey matter development*. *Scientific reports*, 2019. **9**(1): p. 1-8.
- [6] Fullerton, J.M. and J.I. Nurnberger, *Polygenic risk scores in psychiatry: Will they be useful for clinicians?* *F1000Research*, 2019. **8**.
- [7] Passchier, R.V., et al., *Schizophrenia Polygenic Risk and Brain Structural Changes in Methamphetamine-Associated Psychosis in a South African Population*. *Frontiers in Genetics*, 2020. **11**.
- [8] Linden, D.E., *The challenges and promise of neuroimaging in psychiatry*. *Neuron*, 2012. **73**(1): p. 8-22.
- [9] Masdeu, J.C., *Neuroimaging in psychiatric disorders*. *Neurotherapeutics*, 2011. **8**(1): p. 93-102.
- [10] Narlund, S., et al., *Associations between polygenic risk scores for four psychiatric illnesses and brain structure using multivariate pattern recognition*. *NeuroImage: Clinical*, 2018. **20**: p. 1026-1036.
- [11] Richards, A.L., et al., *The relationship between polygenic risk scores and cognition in schizophrenia*. *Schizophrenia bulletin*, 2020. **46**(2): p. 336-344.
- [12] Chatterjee, N., J. Shi, and M. García-Closas, *Developing and evaluating polygenic risk prediction models for stratified disease prevention*. *Nature Reviews Genetics*, 2016. **17**(7): p. 392.
- [13] Dima, D. and G. Breen, *Polygenic risk scores in imaging genetics: usefulness and applications*. *Journal of Psychopharmacology*, 2015. **29**(8): p. 867-871.
- [14] Fritsche, L.G., et al., *Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from The Michigan Genomics Initiative*. *The American Journal of Human Genetics*, 2018. **102**(6): p. 1048-1061.
- [15] Liu, J.Z., et al., *Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations*. *Nature genetics*, 2015. **47**(9): p. 979-986.
- [16] Stahl, E.A., et al., *Genome-wide association study identifies 30 loci associated with bipolar disorder*. *Nature genetics*, 2019. **51**(5): p. 793-803.
- [17] Wray, N.R., et al., *Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression*. *Nature genetics*, 2018. **50**(5): p. 668-681.
- [18] Denny, J.C., et al., *Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data*. *Nature biotechnology*, 2013. **31**(12): p. 1102-1111.
- [19] Docherty, A.R., A.A. Moscati, and A.H. Fanous, *Cross-disorder psychiatric genomics*. *Current behavioral neuroscience reports*, 2016. **3**(3): p. 256-263.
- [20] Leppert, B., et al., *A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank*. *PLoS genetics*, 2020. **16**(5): p. e1008185.
- [21] Opel, N., et al., *Cross-disorder analysis of brain structural abnormalities in six major psychiatric disorders: a secondary analysis of mega-and meta-analytical findings from the ENIGMA consortium*. *Biological Psychiatry*, 2020. **88**(9): p. 678-686.
- [22] Cheng, Y. and G.M. Church. *Biclustering of expression data*. in *Ismb*. 2000.
- [23] Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age*. *PLoS medicine*, 2015. **12**(3): p. e1001779.
- [24] Consortium, S.W.G.o.t.P.G., *Biological insights from 108 schizophrenia-associated genetic loci*. *Nature*, 2014. **511**(7510): p. 421-427.
- [25] Grove, J., et al., *Identification of common genetic risk variants for autism spectrum disorder*. *Nature genetics*, 2019. **51**(3): p. 431-444.
- [26] Demontis, D., et al., *Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder*. *Nature genetics*, 2019. **51**(1): p. 63-75.
- [27] Wood, A.R., et al., *Defining the role of common variation in the genomic and biological architecture of adult human height*. *Nature genetics*, 2014. **46**(11): p. 1173-1186.
- [28] Bonàs-Guarch, S., et al., *Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes*. *Nature communications*, 2018. **9**(1): p. 1-14.
- [29] Ge, T., et al., *Polygenic prediction via Bayesian regression and continuous shrinkage priors*. *Nature Communications*, 2019. **10**(1): p. 1776.
- [30] Vilhjálmsson, B.J., et al., *Modeling linkage disequilibrium increases accuracy of polygenic risk scores*. *The American journal of human genetics*, 2015. **97**(4): p. 576-592.
- [31] Peakman, T.C. and P. Elliott, *The UK Biobank sample handling and storage validation studies*. *International Journal of Epidemiology*, 2008. **37**(suppl_1): p. i2-i6.
- [32] Bycroft, C., et al., *Genome-wide genetic data on ~500,000 UK Biobank participants*. *bioRxiv*, 2017: p. 166298.
- [33] Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. *The American journal of human genetics*, 2007. **81**(3): p. 559-575.
- [34] Kerminen, S., et al., *Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland*. *The American Journal of Human Genetics*, 2019. **104**(6): p. 1169-1181.
- [35] Alfaro-Almagro, F., et al., *Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank*. *Neuroimage*, 2018. **166**: p. 400-424.
- [36] Jenkinson, M., et al., *Smith SM*. *FSL neuroimage*, 2012. **62**: p. 782-90.
- [37] Rodrigue, A.L., et al., *Genetic Contributions to Multivariate Data-Driven Brain Networks Constructed via Source-Based Morphometry*. *Cerebral Cortex*, 2020.
- [38] Winkler, A.M., et al., *Permutation inference for the general linear model*. *NeuroImage*, 2014. **92**: p. 381-397.
- [39] Kozen, D.C., *The design and analysis of algorithms*. 1992: Springer Science & Business Media.
- [40] Rahaman, M.A., et al., *N-BiC: A method for multi-component and symptom biclustering of structural MRI data: Application to schizophrenia*. *IEEE Transactions on Biomedical Engineering*, 2019. **67**(1): p. 110-121.
- [41] Hripcsak, G. and A.S. Rothschild, *Agreement, the f-measure, and reliability in information retrieval*. *Journal of the American Medical Informatics Association : JAMIA*, 2005. **12**(3): p. 296-298.
- [42] Santamaria, R., L. Quintales, and R. Therón, *Methods to Bicluster Validation and Comparison in Microarray Data*. Vol. 4881. 2007. 780-789.
- [43] Rahaman, M.A., et al., *A novel method for tri-clustering dynamic functional network connectivity (dFNC) identifies significant schizophrenia effects across multiple states in distinct subgroups of individuals*. *bioRxiv*, 2020.
- [44] Gao, Z., et al., *The indispensable role of the cerebellum in visual divergent thinking*. *Scientific reports*, 2020. **10**(1): p. 1-12.
- [45] Konarski, J.Z., et al., *Is the cerebellum relevant in the circuitry of neuropsychiatric disorders?* *Journal of Psychiatry and Neuroscience*, 2005. **30**(3): p. 178.
- [46] Phillips, J.R., et al., *The cerebellum and psychiatric disorders*. *Frontiers in public health*, 2015. **3**: p. 66.
- [47] Noordermeer, S.D., et al., *Structural brain abnormalities of attention-deficit/hyperactivity disorder with oppositional defiant disorder*. *Biological Psychiatry*, 2017. **82**(9): p. 642-650.
- [48] Pandya, M., et al., *Where in the brain is depression?* *Current psychiatry reports*, 2012. **14**(6): p. 634-642.
- [49] Wu, H., et al., *Brain Structural Abnormalities of Major Depressive Disorder*, in *Major Depressive Disorder*. 2020, Elsevier. p. 39-49.
- [50] Leech, R. and D.J. Sharp, *The role of the posterior cingulate cortex in cognition and disease*. *Brain*, 2014. **137**(1): p. 12-32.