Whole-Brain Whole-Genome Framework for Imaging Genetics of Psychopathology and Cognition

By Alexander S. Hatoum

B.A., University of Texas at Austin, 2012

M.A., University of Colorado, 2017

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2019

This thesis entitled:

Whole-Brain Whole-Genome Framework for Imaging Genetics of Psychopathology and Cognition

By Alexander S. Hatoum

_____

Chaired By Naomi P. Friedman

_____

Matthew C. Keller

_____

Marissa A. Ehringer

_____

Soo Hyun Rhee

_____

Marie T. Banich

Date_____

The final copy of this thesis has been examined by the signatories, and we

Find that both the content and the form meet acceptable presentation standards

Of scholarly work in the above mentioned discipline

HRC protocol # ____11-0614_____

Abstract

Hatoum, Alexander S. (Ph.D, Psychiatric, Behavioral, and Statistical Genetics. Department of Psychology and Neuroscience)

Whole-Brain Whole-Genome Framework for Imaging Genetics of Psychopathology and Cognition

Dissertation Directed by Dr. Naomi P. Friedman

The conception of the Research Domain Criteria (RDoC) apotheosized integration across differing biostatistical domains in studying psychiatric dysfunction. For example, an RDoC aim is the integration of MRI imaging and genetics, which represent differing foundations for the biological etiology of psychiatric and cognitive states. Though seemingly distal, recent work in both fields suggests that psychological phenomena are not represented by simple *a proiri* hypothesized candidates; instead we require large-data statistical approaches to find pathways for detection and treatment. That is to say, as genetics moves towards more whole-genome approaches, imaging moves towards more whole-brain approaches, and these whole-system approaches will likely be useful for research and clinical practice in the future. Imaging genetics, the field of study at the intersection of these biostatical perspectives, is unequipped to integrate these distal whole-system practices at the current time, with the most popular current approaches in imaging genetics resorting back to the antiquated candidate gene methods that failed to find replicable results. The purpose of this dissertation is to offer a framework for (some) integration in these fields at the whole-brain whole-genome (WBWG) level. While not exhaustive, the research, procedures, scripts, and methods here will help navigate this translational space. **Study 1** demonstrates an approach to high-resolution mapping using the classical twin design to find patterns of association across the human cortex at high resolution for a dimensional measure of depression and a related psychiatric behavior. We then integrate these

mapping results with results from the whole-genome and whole-brain literature more broadly in a big data framework. **Study 2** demonstrates the effectiveness of whole-brain models as phenotypes for genetic studies of intelligence and discusses some practices for the application of whole-brain phenotypes in genetic studies. Finally, **Study 3** conducts a whole-genome association study of cEF and uses the patterns across the genome to implicate particular biological/neurological pathways for analysis. The final chapter returns to discuss each study and how it fits into the general Whole-Brain Whole-Genome framework we lay out in chapter 1.

Table of Contents

List of Tables and Figures

Tables

Figures

Chapter 1

Candidates vs. a Whole-brain Whole-Genome Perspective in Imaging Genetics

Imaging Genetics: MRI and Genetics as Theoretical Complements

Magnetic resonance imaging (MRI) and genetic analyses are perhaps the two most popular biological approaches for studying human psychological phenomena, and their effective integration should be key for advancing our knowledge of psychological traits. In the post-genome-wide association studies (GWAS) era, it is worth revisiting imaging, genetics, their intersection, and their relations to human health and wellness to evaluate and improve current imaging genetics perspectives. Specifically, in this thesis, we argue that current perspectives should shift to integrating more whole-genome and whole-brain models into studies of psychological traits.

This integrative perspective is currently being demonstrated by the National Institute of Mental Health's Research Domain Criteria (RDoC)[1]. The RDoC framework organizes findings from biological psychiatry along a matrix from genetic to tissue and behavior to translate across perspectives and is dominated at particular levels by MRI and genetic approaches. In the RDoC, MRI studies often focus on "tissue" and "circuit" level phenomena by directly measuring in vivo brain patterns and relating them to behavioral outcomes. Genetic studies parse inherited vulnerability by comparing family members or measuring DNA/DNA products directly, illuminating "molecular" and sometimes "cellular" mechanisms underlying psychological states.

MRI and genetics studies deserve their focus and have been useful in creating meaningful biological theories of behavior. Imaging, for example, helps distinguish functional components of the brain from one another, as MRI maps can be compared between traits to create biological

representations of psychological phenomena[2]. For example, De La Vega et al 2016[3] used a large-scale meta-analysis to divide the human medial frontal cortex into psychologically meaningful subdivisions that helped distinguish between motor, social and cognitive processes. However, a well-known limitation of this line of work is that that MRI patterns are correlational rather than causal.

In contrast, genetic studies focus almost entirely on etiology to discover the causes of disorders. The most popular designs in behavioral genetics are twin and genome-wide studies that explore whether traits are coinherited due to the same genotype or quasi-experimental designs that test causal claims in the general population[4]. However, after nearly half a century of behavioral genetics research[5], we know that post-genetic events (everything from transcription to personality) are likely influenced by some combination of genes and environments. The numerous possibilities for development of psychiatric disorders makes genetic association patterns difficult to contextualize without integration of other biological domains.

We argue that imaging genetics can act as a theoretical conduit between genetics and psychology, contextualizing the findings of genetic studies in a meaningful framework. Essentially, MRI provides a visual representation, a biostatical unit to represent and analyze psychological phenomena that is some mix of etiologies and outcomes, while genetics provides an underlying etiological pattern across the population. Thus, the two fields complement one another theoretically.

Do Clinically Relevant Findings in MRI and Genetics Match the RDoC Framework?

The combination of imaging and genetics may also extend beyond the academic, as both fields separately developed relevant clinical applications. MRI and functional MRI (fMRI)

studies have been used to develop biomarkers of treatment effectiveness/resistance[6,7], pain[8], sustained attention[9] and many other psychological traits and states. Genetics has also demonstrated clinical relevance, as family history remains one of the most robust predictors of psychiatric dysfunction[10] and treatment response, such as lithium response[11]. Now, in the era of molecular genetics, patterns across many implicated variants for psychiatric conditions support effectiveness for existing psychopharmaceuticals[12], and are beginning to suggest new potential treatments[13].

While these advances are certainly encouraging, these clinically relevant findings remain *specific to each field*. Further, the translational RDoC framework remains discrepant from these clinically relevant applications of genetics and MRI. The RDoC framework attempts to simplify psychological phenomena, with the current matrix showing a handful of mechanisms at any level of analysis, for example relating a gene product to one brain region. While this has been effective for other fields of medicine, clinically relevant applications of genetics and MRI research *to psychiatry* have embraced complexity. For example, the RDoC matrix originally used many candidate genes under the "genetics" category and recently reformed the matrix to remove candidate genes only after these perspectives failed to replicate (https://www.nimh.nih.gov/research-priorities/rdoc/update-on-genes-in-the-rdoc-matrix.shtml). However, other levels of the RDoC continue to use candidates (i.e. candidate transcripts or brain regions). Instead, translational researchers need to embrace the complexity of psychiatric phenomenon. Family history in psychiatry (genetics), pattern analysis (polygenic risk scores and MRI predictive models), and trait pattern recognition (fRMI multi-voxel pattern analysis) all rely on patterns across the entire system (i.e. the whole-brain or the whole-genome), and it is these perspectives that are offering a path to clinical application in the biological psychiatry literature.

Thus, we call the perspective that uses a priori hypothesized pieces of biology to relate to one another the "Candidate Biomarker" approach. In contrast, we call the emerging whole-brain or whole-genome approaches the "Whole-Brain Whole-Genome Biomarker" approach or WBWG biomarkers. The purpose of this dissertation is to demonstrate a framework for WBWG biomarkers for imaging genetics.

Candidate Approach: Prominence and Issues

There is some evidence for the popular *a priori* genes or brain regions approach in brain disorders. Huntington's disorder is a rare single gene disorder caused by a mutation in the huntingtin gene. The pattern of the expression of this single gene led to discoveries of neuronal breakdown in the basal ganglia, a key area in movement and an area where GABA genes are highly expressed, illuminating the etiology of Huntington's disease. Thus, in past work a single gene did provide insight into a disorder, so it was logical to continue this practice.

However, psychological phenomena are showing a different pattern. Candidate genes have failed to elucidate the etiology and mechanisms underlying psychological phenomena for multiple reasons. First, candidate genes fail to replicate across studies and the classic "file drawer" problem masked this failure[14,15]. This lack of replication likely occurred because candidate genes were typically examined in underpowered samples/designs[14]. After the failure of candidate gene and linkage studies[16] in behavior genetics, the success of many recent large-scale GWAS [17] shows that psychological phenomena are highly polygenic, with each variant attributing an extremely small proportion of the variance to psychological traits. (An exhaustive review of this debate is beyond this scope of this work; see Duncan & Keller 2011[14], Johnson et al. 2017[18], and Chabris et al. 2012[19].)

MRI studies are beginning to experience the same paradigm shift as behavioral genetics. While imaging phenotypes typically have (much) larger effect sizes than individual genetic variants, many recent imaging meta-analyses have determined that the effect sizes of the association between each brain region and an outcome are likely much smaller than previously thought[20]. Further, Redden and Wager 2018[21] used simulation to demonstrate the likelihood of false positives in small samples that only focus on clusters of analysis (rather than whole-brain patterns), arguing that small patterns greatly increases the false positive discovery rate over using a whole-brain pattern analysis method. Finally, Woo and Wager et al. 2015[22] argue that whole-brain patterns are more likely replicate across imaging sites and scanners[21], analogous to results in genetic studies.

While past results in each field are encouraging, imaging genetics is lagging behind both fields and has been dominated by candidate approaches. For example, one of the most popular imaging candidates is the insula[23], and a prominent genetic candidate is the serotonin transporter gene (5HTTLPR) in depression research[14]. Figure 1 shows the number of papers published each year for the past 8 years with the words "imaging serotonin transporter gene insula" in the text as of February 17th, 2019. Over 10,000 publications matched this search across all years and more than 500 in each year. However the high polygenicity of the insula[24] and behavioral phenotypes means these discoveries are likely multiplying the false positives in their respective fields. Even more damning, 5HTTLPR is likely not even influencing the insula (anatomically) at the genetic level[24].

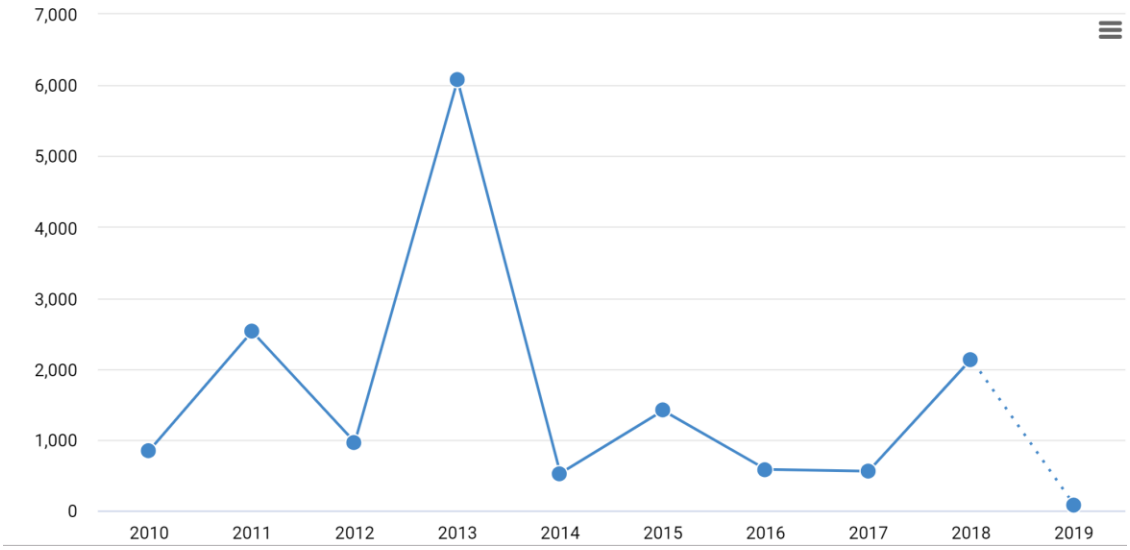


Figure 1. Number of papers published in the last 9 years using the words "imaging serotonin transporter gene insula" in a google scholar search by year. The search was done February 17, 2019.

Further, it is unlikely that high polygenicity of a psychiatric disorder is acting through a single brain region. For example, reduced hippocampal volume is one of the most robust findings in schizophrenia imaging literature[20], making hippocampal volume a key imaging candidate for large-scale schizophrenia research. However, the genetic associations between hippocampal (and all other subcortical brain) volumes with schizophrenia are small and non-significant[25]. While others have argued this lack of association could be due to hippocampal volume being a result of schizophrenia rather than a consequence, another explanation is that this lack of association is due to a candidate approach[25]. Notably, there is likely a genetic association (rather than a causal process), as hippocampal volume is smaller in unaffected relatives[26,27] and during the prodromal period of schizophrenia[28], meaning the development of schizophrenia is not necessary for the reduction in volume, but the risk for schizophrenia is. Further, the genetic correlations across limbic regions in studies of the genetic association between schizophrenia and

subcortical brain measures follow a similar pattern as the effect sizes for associations in the phenotypic literature[25]. Instead, it is likely a power issue that is driving the non-significant findings. We argue that using the whole-genome and whole-brain techniques as translational tools will be useful in predicting and elucidating mechanisms underlying psychological traits.

What About Knockout and Lesion Studies?

By WBWG approaches, we mean approaches that (1) consider the whole-system (i.e. whole-genome or whole-brain) and (2) utilize the patterns across the system to predict/theorize psychological outcomes. The main criticism/limitation of this line of thinking may come from past work in lesion and gene knockout studies that have demonstrated the impact of single brain regions or single genes. Here, I argue that these perspectives are not the antecedent of the WBWG approach. Instead, it is likely that contextualization of knockouts or lesions in WBWG models is still important.

For example, dorsal lateral prefrontal cortex (DLPFC) activity is key for executive functions, which was first demonstrated by lesion studies[29]. However in the general population, individual differences in executive functioning are also influenced by patterns in lower order brain areas and even the cerebellum[30,31]. In this example, the DLPFC is necessary but not sufficient for EF ability: We need the DLPFC, but better executive functioning may rely on other brain areas as well. Further, single areas and whole brain patterns can be modeled simultaneously to study mechanisms of neurological dysfunction. Even within local lesions of stroke patients, connectivity among other unaffected brain regions often determines degree of impairment and recovery[32].

In contrast to necessary but not sufficient biological units in imaging, geneticists have argued for redundancy in the genome. A single knockout or mutation (alone) is unlikely to produce *complex disease* or trait like psychiatric dysfunction (in most instances), as individuals carry several deleterious mutations throughout their lives[33]. For another example, it is thought that the polygenic background of Alzheimer's disease is important for the degree of impairment caused by APOE allele[34]. Thus, similar to MRI, in genetics there is nuance in the pattern of effects, but in the opposite way as in MRI research.

Therefore, WBWG biomarkers may be necessary but not sufficient models, like in imaging, or may mean effects working in unison, like genetics, or some combination. However, these nuanced complex mechanisms in each field are only known because candidate biomarkers *and* the rest of the biological system were studied jointly.

Current Whole-Brain Whole-Genome Approaches

There are a few key examples of WBWG approaches in the literature. First, the relationship between family similarity and anatomical and functional brain patterns has been studied extensively. These studies have found that genetic influences on brain tissue are likely highly anatomically complex and do not conform to a priori brain regions; in other words, heritability varies within and between popular regions of interest anatomically[35,36] and functionally[37]. Second, despite popular ROI maps being inaccurate representations of inheritance of neuroanatomical patterns, it is possible to estimate patterns of inheritance directly from brain maps using twin models. Twins are a popular whole-genome approach, as they allow tests of relationships among family members matched for age and common rearing environment. Twins are also substantially more powerful in discovery of genetic correlations (*r*G) and less computationally expensive than measured genotypes. Chen et al.[38] 2012 demonstrate that there

is high spatial autocorrelation in the inheritance of neuroanatomy; put simply, areas that are closer to one another (within large brain lobes) tend to share the same genetic influences. Thus, it is possible to estimate theoretically useful regions or "clusters" within the brain, but these patterns may differ from typical brain atlases.

Three previous applications of the knowledge of inherited anatomical patterns exist to form a foundation for the work of this thesis as an application to behavior. First, Couvy-Duchesne et al. 2018[39] demonstrated this perspective by genetically relating each vertex to one another to create a coarse genetically informed brain parcellation, and then relating the clusters from their parcellation to anxiety symptoms. They found that anxiety related (genetically) to the lingual gyrus.

Second, researchers have found that using genetic models at the voxel-wise/vertex-wise level addresses the lack of specificity in single region analyses. One study found that by using vertex-wise patterns across different developmental stages, one can decompose the genetic variance at each piece of the brain then relate these variance components to outcomes of interest, in this case cognitive development[40].

Finally, summary statistics of whole-genome effects are very useful in smaller imaging studies. In particular, using polygenic scores at the voxel-wise level has been a popular approach for whole-genome brain mapping. Notably, using polygenic scores from psychiatric disorders can predict activation during a cognitive tasks (e.g., word appropriateness) in healthy subjects[41], showing that these polygenic scores are tapping mechanisms that are more normally distributed throughout the population.

Whole-Genome Whole-Brain Approach to Psychological Studies: Review and Grounding

For this work, I focus on a few key methodological perspectives to ground the thesis and reduce the number of paths to explore for the overlap between imaging genetics. The methods for the WBWG biomarker approach in genetics are the classical twin design and GWAS. In imaging, I focus on the whole-brain mapping approaches (popular in packages like freesurfer and FSL) and connectivity pattern analysis. My work on the WBWG approach integrates these perspectives in three different studies.

Study 1 utilizes the classical twin design to generate high-resolution anatomical brain maps to estimate directly what neural patterns are co-inherited with pathological behaviors. Specifically, we imitate the classic parametric mapping approach but replace the general linear model with a bivariate genetic model to study symptoms of depression. We follow up on the depression brain map in two ways. First, we show how bivariate approaches in imaging and behavior genetics (twins) can be used in a complementary fashion to generate new discoveries by relating a scale of depression to a correlated scale of unfeeling/uncaring behavior. Then, we relate our results to results from the broader genetics and MRI literature.

Study 2 changes imaging modalities and focuses on patterns across whole-brain resting-state functional MRI connectivity as endophenotypes, in comparison to the modern approach of using single brain regions as endophenotypes. Here, we use intelligence as a practical example phenotype. We develop connectivity-based predictive models and probe their effectiveness as endophenotypes vs. the most associated connections alone.

Finally, **study 3** conducts a GWAS of common executive functioning (cEF) and uses whole-genome patterns to discover likely neurological pathways underlying cEF. While GWAS can implicate particular variants, it is likely that any individual variant is exerting a small effect for highly polygenic traits. Instead, it is likely that combined effects of many variants on a

biological or cellular mechanism (i.e. GABA or calcium channels) is contributing to the outcome of interest. To demonstrate this and discover biological pathways underlying cEF, we use regression-based bioinformatic techniques to test whether variants influencing all genes in a particular biological pathway (or transcriptional profile) are implicated by the cEF GWAS. We then use the whole-brain profile from one of these methods to conduct a drug relabeling study to see what psychopharmaceuticals may be targeting cEF in their improvement of psychiatric conditions.

Finally, in chapter 5 we will attempt to contextualize each of these three studies in line with the principles we have set out here in chapter 1. Specifically, chapter 5 will answer two questions (1) How did each study use a whole-system biological approach to make biological discoveries? (2) Did each study implicate genetic pathways of psychological states that can be contextualized in the brain?

Chapter 3: Whole-Cortex Mapping of Common Genetic Influences on Depression and A Social Deficits Dimension

Chapter Summary

Social processes are associated with depression, particularly understanding and responding to others, deficits in which can manifest as callousness/unemotionality (CU). Thus, CU may reflect some of the genetic risk to depression. Further, this vulnerability likely reflects the neurological substrates of depression, presenting biomarkers to capture genetic vulnerability of depression severity. However, heritability varies within brain regions, so a high-resolution genetic perspective is needed. In a sample of 258 same-sex twin pairs from the Colorado Longitudinal Twin Study (LTS), we developed a toolbox that maps genetic and environmental associations between brain and behavior at high resolution. We used this toolbox to estimate brain areas that are genetically associated with both depressive symptoms and CU. We then overlapped the two maps to generate coordinates that allow for tests of downstream effects of genes influencing our clusters. Genetic variance influencing cortical thickness in the right dorsal lateral prefrontal cortex (DLFPC) sulci and gyri, ventral posterior cingulate cortex (PCC), pre-somatic motor cortex (PreSMA), medial precuneus, left occipital-temporal junction (OTJ), parietal-temporal junction (PTJ), ventral somatosensory cortex (vSMA), and medial and lateral precuneus were genetically associated with both depression and CU. Split-half replication found support for both DLPFC clusters. Meta-analytic term search identified “theory of mind”, “inhibit”, and “pain” as likely functions. Gene and transcript mapping/enrichment analyses implicated calcium channels. CU reflects genetic vulnerability to depression that likely involves executive and social functioning in a distributed process across the cortex. This approach works to unify neuroimaging, neuroinformatics, and genetics to discover pathways to psychiatric vulnerability.

Introduction

As depression follows a normal distribution of risk across the population[42], relating depression to psychological features will better define pathways for addressing disorder vulnerability [1]. Disruption in the ability to process social cues can lead to deficits in daily functioning and is often seen in depression[43]. "Social Dimensions" is one of the main dimensions in the U.S. National Institute of Mental Health Research Domain Criterion (RDoC). Depressed individuals' symptoms relate to specific facets of social behavior, namely, reasoning through others emotions[44–46], fitting under the subcategory of the "social dimensions" matrix, "understanding mental states." Unsurprisingly, social deficits in theory of mind, the ability to understand others' thoughts, are related to poor mentalizing/metacognition, or inability to understand the self. Further, theory of mind even predicts depression diagnosis above and beyond metacognition in behavioral studies[47].

An inability to understand and respond to others' emotions may manifest as callousness/unemotionality (CU, 8). Although typically examined in the context of externalizing disorders (10, 11), CU has also been consistently associated with depression [50,51]. This association may arise because, while CU may reflect a disregard for others, it may also reflect an inability to empathize with others, perhaps due to poor theory of mind and metacognition about one's own emotions. Consistent with this interpretation, CU has been related to mechanisms in social processing, like inability to understand others in development[48]. Thus, poor social processing/CU may be a mechanism that sustains depression[43] or index the severity of depression[52]. In either case, CU may reflect genetic influences on internalizing vulnerability[53], and brain mapping the overlap between depression and CU could help us determine whether cognitive or lower-order neurological systems are involved.

Multiple biological perspectives could advance our understanding of CU in depression. Family and genetic studies can estimate the relative importance of common genes and environments across two traits. Coinheritance between depression and CU is likely, as behavior genetics has established that depression is partially genetic in origin[54]. Further, a recent genome-wide association analysis implicated over 150 genes in depression etiology[42], any of which could relate more specifically to social processing. However, while genetics is an excellent basis for studying psychiatric symptoms in the population, genes/variants and their downstream mechanisms are difficult to scrutinize[55].

In contrast to this lack of contextualization in genetic research, brain mapping integrates nicely onto other areas of biology (like transcriptomics[56]), thanks to the specificity gained when using high-resolution scanning coordinates. Here, we implement an integrative framework in which we directly map areas of the brain that represent the genetic overlap of CU and depression. Specifically, the goal of the current study is test whether CU captures some of the genetic vulnerability to depression; and localize the brain areas contributing to this vulnerability. These genetically associated brain areas can then be used with bioinformatic tools for mapping across different levels of the RDoC, such as RNA expression and biological pathway analyses, to expand our understanding of the coinheritance of CU and depression and find likely mechanisms of this behavioral vulnerability.

Depression and CU in the Brain

Spatial brain mapping studies can localize where behavioral measures are associated with brain morphology. By overlaying neural correlates of depression with neural correlates from other measured behaviors, we gain specificity on areas associated with aspects of depression, like CU. While the neuroanatomical correlates of depression and CU have been studied

extensively, this will be the first study examining the anatomical overlap between CU and depression.

The largest meta-analysis of neuroanatomical differences in depression to date used region of interest (ROI) measures of cortical thickness. It found that major depressive disorder (MDD) was associated with cortical thinning in the insula, anterior and posterior cingulate, and temporal gyri[57]: areas key in salience[23], internal *mentation*[58], and switching between internal thought and executive control[59]. However, this ROI approach does not consider how subcomponents of large areas may differentially relate to more specific facets of psychological phenomena; for example, anterior cingulate cortex shows differential gene expression and differential task activation across the ROI [60]. One meta-analysis of voxel-based morphometry (VBM) found that MDD was associated with lower brain volume specifically in the rostral anterior cingulate cortex and the dorsolateral and dorsomedial frontal cortex[61].

For neuroanatomical correlates of CU, decreases in the volume of the rostral and dorsal cingulate cortex have been observed, overlapping spatially with the regions that have been identified for depression [62]. Additionally, the rostral and dorsal anterior cingulate cortex areas that overlap between CU and depression were also found to distinguish suicidal cases from controls in another VBM study[63], giving some evidence that CU could represent a social severity dimension of depression.

This Study

Using structural magnetic resonance imaging (MRI) data from 258 young adult twin pairs, we asked, where are the genes influencing the vulnerability to social deficits and depression influencing brain morphology? Do these morphological differences overlap? And, can we map a specific pattern and use this pattern to speculate further on mechanisms? To

answer these questions, we used the methodology pictured in Figure 1 (a tutorial for this approach can be found on our github: https://github.com/AlexHatoum/Wild-Card-Toolbox). In step 1, we estimated the genetic and environmental association between depression and CU to evaluate the relative importance of shared inherited vulnerability. In step 2, we developed a toolbox that creates environmental and genetic brain maps for each trait. Rather than map standard beta coefficients (i.e., clusters associated with phenotypic variability), our procedure maps effect sizes for genetic and environmental variances (i.e., clusters associated with our traits via a genetic or environmental etiology), creating brain maps of genetic association between cortical thickness and the two behavioral traits. We estimate areas that represent the genetic vulnerability to CU and depression by overlaying the clusters from the separate depression and CU genetic maps onto one map. Finally, by integrating these brain maps with neuroinformatic tools in step 3, we can begin to characterize likely functions and specific molecular mechanisms of the genes influencing CU and depression, which is impossible in a standard biometrical design. Specifically, in step 3, we used MNI coordinates to align our genetically associated clusters with a meta-analytic database of effects across multiple fMRI and transcriptomic studies. Thus, our main analysis is the generation of genetically influenced brain map for depression and CU, and our follow-up analyses explore likely effects of this genetic variance implicated by this map by using high-resolution brain coordinates.

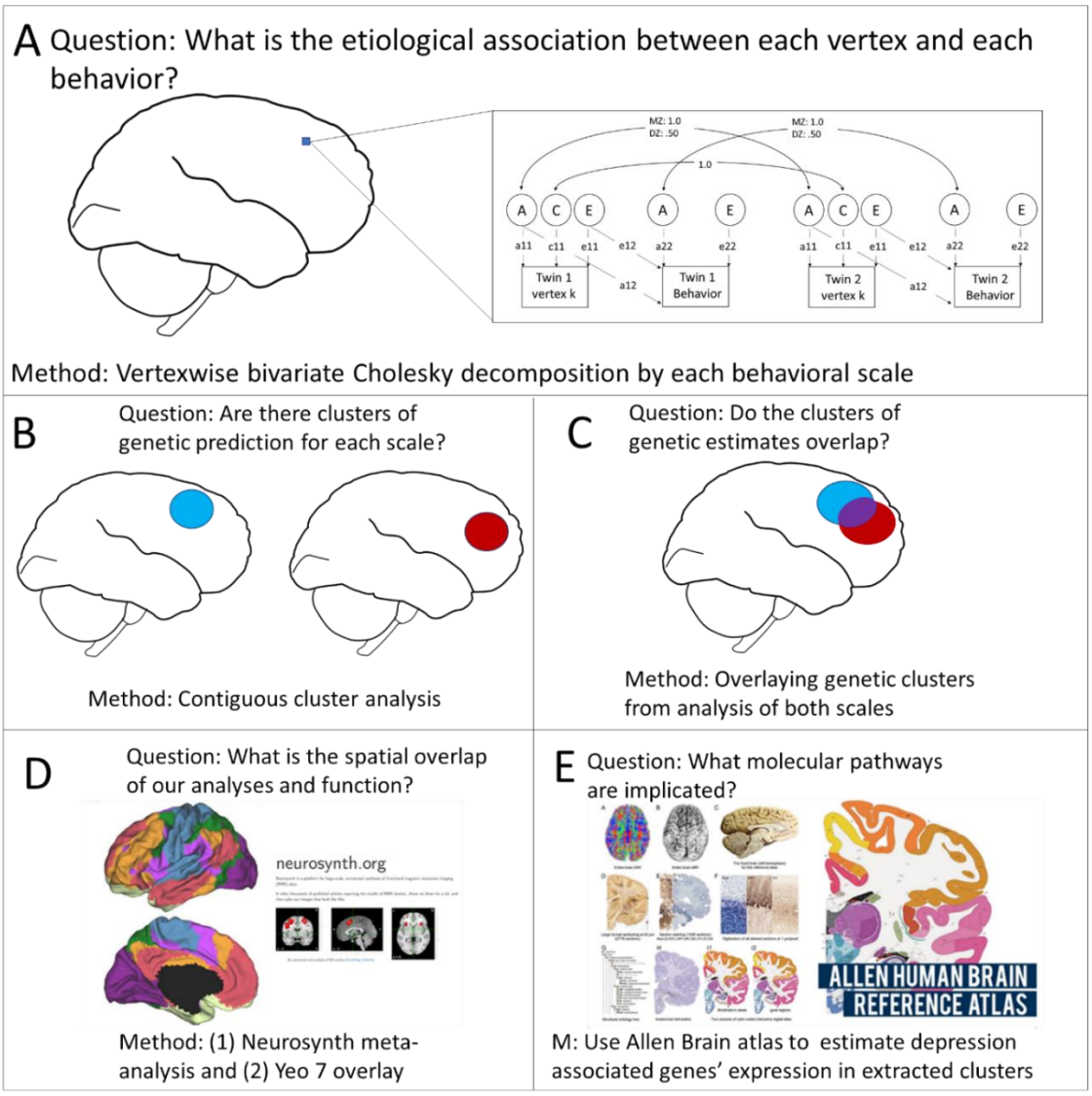


Figure 1. Five steps for whole cortex mapping by genetic association and follow up using informatic tools. Panel A: Additive genetic (A), Common environmental (C) and non-shared Environmental (E) Cholesky decomposition is used to find the etiological association of each vertex with each behavioral scale. Multiplication of standardized paths labeled 11 and 12 represents the phenotypic correlations predicted by additive genetic (Gr) and non-shared environmental (Er) influences, respectively. Panel B: Vertices whose associations with behavior are significant (p<.05) and are part of a contiguous cluster of larger than 20 mm (cluster-extent correction) are estimated across the cortex surface separately by each trait and separately for A and E components. This procedure recovered 4 categories of clusters: additive genetic clusters influencing CESD, additive genetic clusters influencing ICU, non-shared environmental clusters influencing CESD, and non-shared environmental clusters influencing ICU. Panel C: Areas that represent significant conjunction of genetic association is created by overlaying the genetic clusters from CESD and ICU after cluster-extent correction. Panel D: The coordinates for overlap were transformed in MNI space and were used to map onto the Yeo 7 functional connectivity patterns and conduct meta-analytic term searches of likely associated functions. Panel E: Genes associated with depression in a large genome-wide association study were

extracted from Neurosynth-gene/Allen Brain Atlas dataset to examine the expression of each of those genes in our clusters.

We conduct this analysis in a general population sample to include subsyndromal levels of depression and CU within a large enough sample to find patterns of coheritability between brain and phenotype. We chose high-resolution brain mapping because prior literature in neuroimaging genetics suggests vertex-wise approaches will more appropriately capture the individual differences patterns of genetic effects. In particular, common brain atlases used in anatomical research were derived agnostically to genes influencing individual differences and do not capture the specificity of the architecture of genetic effects on behavioral traits, as has been shown for language-related brain areas[64]. Further, past work has shown there are differences in the genetic variance structure within and between commonly utilized ROIs; thus, measuring genetic variability in ROIs vs. vertices leads to relative differences in genetic variance effects between regions being overestimated[35] and more fine-grained metrics, such as voxel or vertex measures, are preferable to ROI approaches for making comparisons across the cortex for individual differences genetics[35]. Notably for this study, it is these genetic individual differences patterns that are implicated in the mechanisms of psychopathology, requiring high-resolution coordinates to specify accurately. Finally, using high-resolution analysis and MNI coordinates allows for integration with functional and transcriptomics literature more broadly.

Methods and Materials

Sample

Participants were 258 same-sex twin pairs (225 complete, 120 monozygotic [MZ], and 115 dizygotic [DZ], 132 female pairs and 93 male pairs, singletons were used in calculating the mean and variance), aged 28 - 31 years (M = 28.7, SD = 0.6), recruited from the Colorado

Longitudinal Twin Study (LTS). Twin pairs who had completed an ongoing neuroimaging study of neural substrates of executive functions and psychopathology and whose imaging data passed quality control were included. More about the LTS can be found in the online methods.

Structural MRI Scan

Images were acquired on a Siemens Trio 3 Tesla MRI scanner with 32-channel parallel imaging located at the University of Colorado Boulder. The total scanning session lasted 1 hour 25 minutes; the current analyses focus on gray matter structure, obtained with a high-resolution T1-weighted Magnetization Prepared Gradient Echo sequence in 224 sagittal slices, with a repetition time (TR) = 2400 ms, echo time (TE) = 2.01 ms, flip angle = 8◦, field of view (FoV) = 256 mm, and voxel size of 0.8 mm^3.

Behavioral Assessment

On the day of the scan, participants completed the Center for Epidemiological Studies-Depression (CESD) scale, a 20-item Likert scale assessing the frequency of past-week depression symptoms[65]. We chose this measure because (1) tendencies toward an emotional vulnerability should manifest itself in higher frequency of depression, (2) we wanted to include subsyndromal levels of depression, and (3) this measure has shown reasonable stability across 10 years of longitudinal data[66].

Prior to the scanning session, participants completed an online questionnaire battery that included the Inventory of Callous and Unemotional traits (ICU)[51], a 24-item Likert questionnaire with three subscales: callousness (e.g., *The feelings of others are unimportant to me*), uncaring (e.g., *I do things to make others feel good*, reverse coded), and unemotional (e.g., *I do not show my emotions to others*). We used this scale as a measure of CU because it has been used to define clinical subtypes of conduct disorder in the past[67], the ICU total score relates to social and

emotional processing[50], and, though the factor structure changes in adulthood, the scale retains a high internal consistency and predicts social, emotional, and depressive behaviors in individuals similar in age to our sample[67]. We conducted all analyses with the ICU total scale, which is more reliable and normally distributed than the subscales, which were all highly intercorrelated (see Supplemental Table S1).

For the CESD and ICU, the dependent variable was the average item rating provided that at least 80% of the items were answered, multiplied by the number of items. To improve normality, both scales were then square-root transformed (see Supplemental Table 1).

Data Analysis

All twins' cortical thickness estimates were processed using a standard Freesurfer pipeline[68] (full description in online methods). Each vertex and psychopathology measure was residualized on brain mean thickness and sex prior to model estimation.

Behavioral genetic ACE models decompose phenotypic variance into three sources: Additive genetic (A; the sum of a large number of genetic variants), Common environmental (C; environmental influences that lead siblings to correlate), and non-shared Environmental (E; environmental influences that lead siblings to be uncorrelated). Because MZ twins share all their genes, their additive genetic influences correlated at 1.0; DZ twins share on average half their genes identical by descent, so their additive genetic influences correlate at 0.5. By definition, C effects correlate 1.0 and E effects correlate 0.0 for both types of twins.

To examine the genetic and environmental covariance between the psychopathology measures and brain measures, the standard ACE model for a single variable can be extended to multivariate analyses. To ensure that the estimated component covariance matrices are positive definite, they are expressed as the product of a lower triangular matrix and its transpose (Figure

1A). This is the Cholesky decomposition[69], which decomposes the phenotypic covariance between two measures into that explained by genes and environments. The genetic correlation (rG) of the two phenotypes equals $(a11*a12)/\sqrt{(a11^2*(a12^2 + a2^2))}$.

Depression and CU coinheritance. To examine the etiological overlap between Depression and CU, we started by estimating their phenotypic overlap through a partial correlation analyses (accounting for sex and mean cortical thickness). We used a series of structural models to show that our association is specific to our measure of depressive symptom frequency and CU, rather than a broad psychiatric vulnerability (Supplemental Figure S1). Finally, we used a standard bivariate Cholesky decomposition to estimate the relative contribution of genes and environment to the overlap between the measures.

Discovery procedure for brain maps. A full diagram of the analysis plan is available in Figure 1. For each vertex, we estimated a separate Cholesky decomposition with the first variable being the vertex and the second being the CESD or ICU scale. We noticed substantial C variance across some areas of the cortex (Supplemental Figure S2) so we specified our Cholesky decompositions with a freed C path loading on the vertex but set the C cross path and specific C loading on the psychopathology variable to be zero, as there were no C effects on the CESD or ICU measures. We then computed the parameter representing the bivariate heritability, the phenotypic correlation predicted by the overlap in genetic influences (standardized a11*a12), at each vertex and projected it to a surface map in Freeview[70] to create whole-cortex heat maps of genetic effects on the brain-behavior association. From the generated whole-cortex map, we estimated clusters above significance for CESD and ICU, respectively, and then overlaid the CESD and ICU clusters.

To determine significant clusters for each disorder, we (1) estimated a chi-square

difference test *p*-value for each Cholesky bivariate cross path, and (2) used vertex-wise cluster extent *p*-value correction of values below (.05) significance at a window of twice the original smoothing kernel (i.e. cluster extent threshold = 20 mm). We chose this procedure partially due to its practicality in integration with genetic estimates and to estimate clusters that were contiguous for follow-up analyses.

Split-half replication. To explore the replicability of our approach, we split our sample into halves by families (so that twin pairs would be kept together) by random draw (sample 1 *n* = 132 twin pairs, sample 2 *n* = 126 twin pairs) and ran the full analyses separately in each sample. In each half of the sample we used a conjunction minimum alpha of .05[71] and cluster extent correction of 20mm to define significant clusters. We then overlaid the clusters from the (1) the full sample analysis, (2) the analysis in sample 1, and (3) analysis in sample 2. Because the full sample was more conservative than either half, we wanted to use the criteria of significant overlap in all three analyses as our standard. i.e. a cluster must have been independently associated below the split half criteria in both half-samples and by a more conservative threshold with the full sample for us to have "high confidence" in its effect.

Transcripts, cell types, and functions associated with our genetic clusters. Using MNI coordinates, we examined the overlap of our clusters with other sources of data: (1) The Allen Brain atlas transcriptomic atlas and genome-wide association study (GWAS) results from the Psychiatric Genomics Consortium Depression mega-analyses[42], (2) Neurosynth meta-analytical database of functional activation across over 10,000 fMRI studies[72], and the (3) Yeo et al. 2011 7-network parcellation[73].

With the Allen Brain atlas, we took the list of associated genes from the psychiatric genomics consortium MDD GWAS gene-burden results[42] and used Allen brain atlas through

Neurosynth gene[74] by downloading each gene image, renormalizing them across the cortex with FSL[75] and visualizing their expression. We excluded genes from the major histocompatibility complex, as these associations may be spurious due to long-range linkage disequilibrium (LD), and any genes not obtained through RNA arrays in the Allen Brain Atlas, leaving us with 100 genes. We put the expression values by each region in one matrix with k-means clustering. We used the elbow method (Supplemental Figure S3) see how many genetic clusters were recovered from our analyses. We then put the gene list of each cluster through the Reactome[76] pathway analysis database, using FDR to account for multiple comparisons.

With Neuro-synth, we entered our clusters from the discovery sample into Neuro-synth decoder to obtain "terms" that were most associated with functional activation across studies, as determined by a meta-analytic naïve Bayes classifier across over 10,000 fMRI studies[72]. This analysis finds which of our coordinates most overlap with those found in the literature and which terms (fMRI patterns, tasks, or studied behaviors) are associated with those studies. We then identified which of these terms most commonly appeared across clusters (after filtering out non-specific brain terms). Finally, we overlaid the coordinate of our clusters on the 7 resting-state networks from the Yeo parcellation[73] to identify to which networks the clusters belonged.

Results

Is CU Genetically Correlated with Depressive Symptoms?

We began by estimating the phenotypic, genetic, and environmental overlap between the depressive symptom frequency, measured by the CESD, and CU, measured by the ICU. Figure 2 shows the AE Cholesky decomposition. Based on the best fitting models for each univariate trait, C paths were not estimated (see Supplemental Table S2 for full model comparisons for each trait). We derived the genetic correlation between the two measures as $r_G = .40$ ($p < .001$, see

Supplemental Table S3 for genetic correlations between CESD and the ICU callous, uncaring, and unemotional subscales). The environmental association was not significantly greater than zero (*r*E=.04, *p*=.50). We concluded that the correlation between CU and depressive symptoms was due almost entirely to genetic covariance.

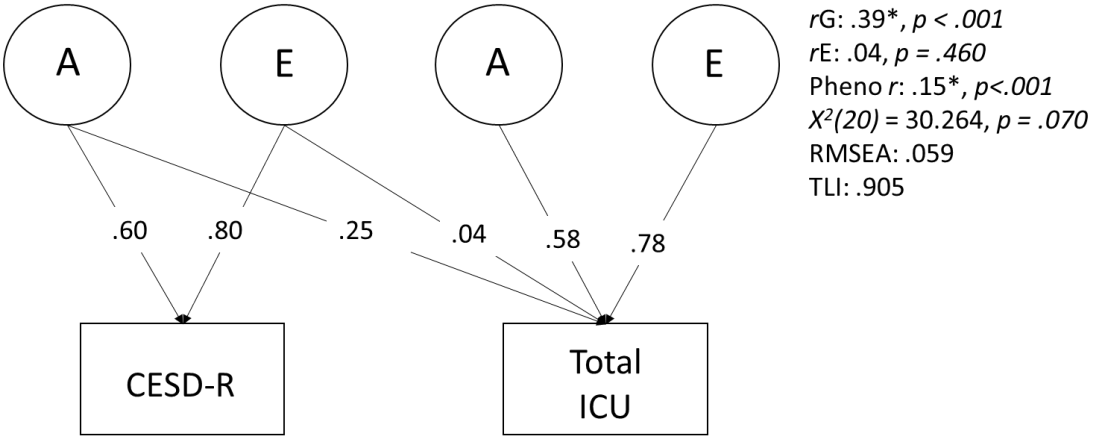


Figure 2. Additive genetic (A) and non-shared Environmental (E) Cholesky decomposition of the relationship between the Center for Epidemiological Studies Depression scale (CESD) and the Inventory of Callous and Unemotional traits (ICU). Numbers on arrows are standardized path estimates. Each task was residualized on sex and mean thickness prior to analysis. The derived genetic (*r*G), environmental (*r*E), and phenotypic (pheno *r*) associations are shown to the right of the path model. The model fit well by chi-square ($X^2(20) = 30.264$, $p = .070$ and RMSEA (.059). **p*<.05, determined by 1-df chi-square difference test. Dotted lines indicate *p*>.05.

Where are CU/Depressive Symptom Genetic Influences Related to Brain Morphology?

We created a map of areas where cortical thickness genetically correlated with CESD and ICU scores. We then overlaid the clusters from the two maps to discover regions that showed conjunction for genetic prediction.

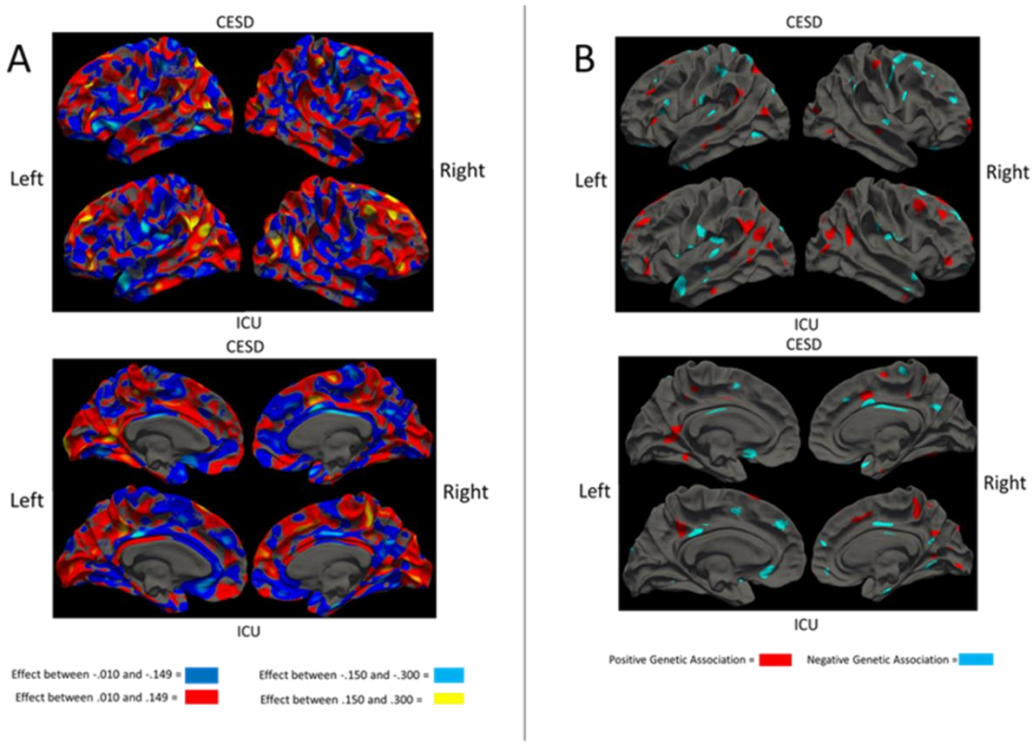
As shown in Figure 3 and Table 1, we found that genetic influences on thicker cortex in the right dorsal lateral prefrontal cortex (DLFPC) sulci, the right pre-somatic motor cortex (PreSMA), left medial and lateral precuneus, occipital-temporal junction (OTJ), and temporoparietal junction (TPJ) were associated with both traits (i.e., these areas showed positive

genetic associations above significance with both measures). We found genetic influences on thinner cortex in the right ventral posterior cingulate cortex (PCC), right medial precuneus, right DLPFC gyrus, and left ventral somatosensory cortex in the pathophysiology of both traits. Finally, split half-replication gave support for both right DLPFC areas in the same direction as discovered in the full sample (Supplemental Figure S4). Comparison to phenotypic maps (Supplemental Figure S6) showed that overlay regions discovered would have been qualitatively different without the genetic approach, as phenotypic areas did not overlap substantially with our genetic areas.

Table 1. **Cluster Coordinates for Overlay Clusters in mm Space**

Cluster	COG X	COG Y	COG Z	Number of Vertices
L-Lateral Precuneus	-14	-67	57	6
L-Medial Precuneus	-6.54	-42.3	43.6	38
L-Occipital Junction	-46.1	-72.8	14.7	138
L-Temporal Junction	-57.7	-49.1	29.9	191
L-ventral SMA	-60.8	-16.7	23.9	73
R-DLPFC*	23.5	32.6	35.2	61
R-Lateral Frontal*	23	16.4	57.2	21
R-PCC	4.87	-13	30.6	42
R-Posterior Precuneus	5.48	-59	31.1	99
R-PreSMA	11.1	12.6	43.6	28

Note. Cluster coordinates for each of the overlay clusters discovered in our analysis. Coordinates for the Center Of Gravity (COG) of the peak activation are given in mm space for X, Y and Z coordinates and size was determined based on the number of vertices in each cluster. The name of each area was determined by entering the coordinates into Neurosynth and using the top gyri/sulci name. R = right hemisphere and L = left hemisphere. DLPFC = Dorsal lateral prefrontal cortex, PCC = Posterior Cingulate Cortex, SMA = Somatamotor area.
*clusters that replicated in the sample split half replication.



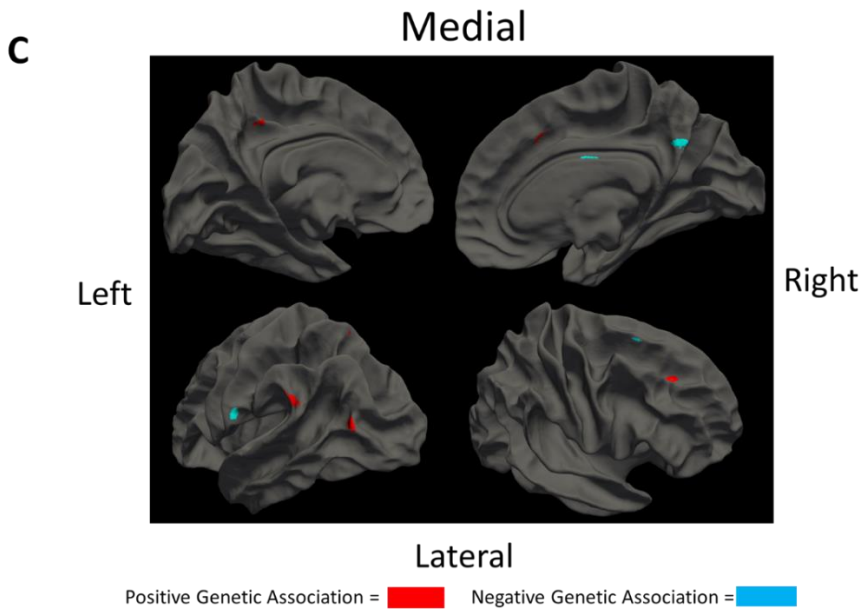


Figure 3. Neural associations with the Center for Epidemiological Studies Depression scale (CESD) and Inventory of Callous and Unemotional traits (ICU). Panel A depicts whole-cortex heat maps of the genetic association of each vertex with each behavioral measure as bivariate heritability. Panel B depicts *p*-values for genetic association between each vertex and each behavioral scale below correct significance ($p < .05$). Lateral views are on top and medial views below. These analyses correspond to those outlined by Figure 1 panel B. Panel C depicts overlap areas for our genetic clusters. These genetic clusters coordinates were used in all future analyses.

Our method also creates an environmental association map. If genetic and environmental association are in the same direction, it is consistent with an explanation of causality[77], though not sufficient to establish a causal relationship. Environmental associations were not consistently in the same direction of effect as the genetic clusters that were discovered (see Supplemental Figure S5). Thus, from the environmental map analysis and the bivariate Cholesky decomposition of ICU and CESD, we concluded these areas are likely biomarkers that reflect genetic vulnerability to CU and CESD and implicate a shared genetic liability.

What Genetic Pathways are Implicated?

We used follow-up analyses to gain insight into potential mechanisms involved in this genetic vulnerability. Results of clustering of Psychiatric Genomics Consortium depression-related genes are shown in Figure 4. We found three clusters: overexpressed, mixed expression, and under-expressed (genes listed in axis of Figure 4). The overexpressed cluster showed significant enrichment for genes in "Depolarization of the Presynaptic Terminal Triggers the Opening of Calcium Channels_Homo sapiens_R-HSA-112308" pathway (FDR corrected p=.03). No other pathways were significant across any of our clusters after FDR correction.

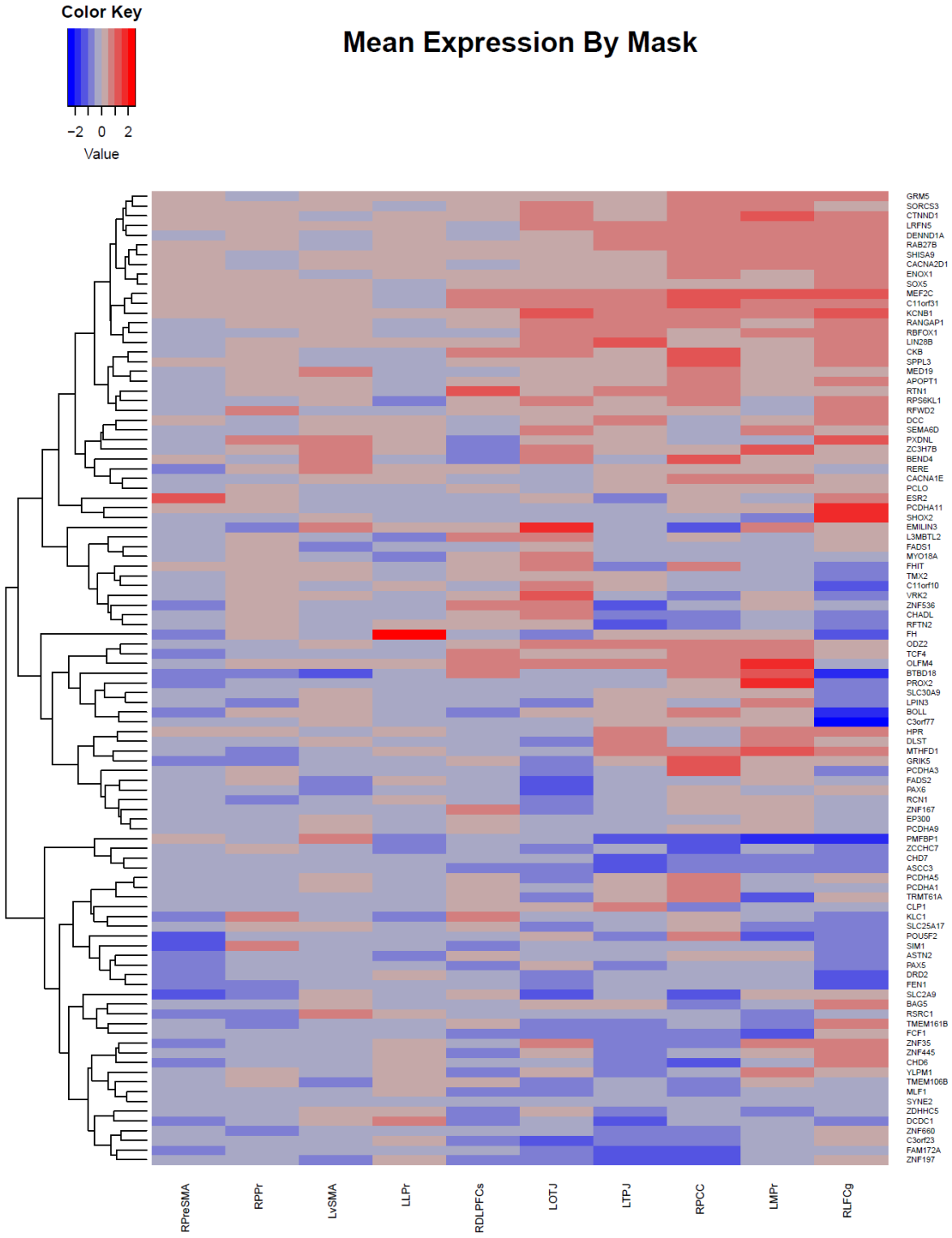


Figure 4. Hierarchical clustering of expression patterns of depression genes in derived clusters. Color scale is the *z*-score for the degree of expression of that gene in the derived area mask compared to the whole cortex. Depression genes were obtained from the Psychiatric Genomics Consortium GWAS gene-burden tests, excluding genes from the major histocompatibility complex region[42]. Gene expression values were recovered from Neurosynth-gene, which

processed data from the Allen Brain Atlas, Human Brain Atlas. R = right hemisphere clusters; L = left hemisphere clusters. RDLPFCs = right dorsal lateral prefrontal cortex sulci, RLFCg = Right Lateral frontal gyri, LLPr = Left Lateral Precuneus, LMPr = Left Medial Precuneus, LOTJ = Left Occipital Temporal Junction, RPreSMA = Right Pre-Somatosensory Area, RPCC = Right Posterior Cingulate Cortex, LvSMA = Left Ventral somatosensory, LTPJ = Left temporoparietal Junction, and RPPr= Right Posterior Precuneus.

What Likely Cognitive/Behavioral Pathways are Involved?

To identify likely cognitive/behavioral mechanisms reflecting this vulnerability, we conducted a meta-analytic term search using Neurosynth. Supplemental Tables S4 and S5 show the 25 most positively associated function terms from Neurosynth for each genetic overlap cluster from the full sample (in some cases, fewer than 25 terms were positively associated). The top repeated behavioral terms were "Theory of Mind", "inhibit", and "pain" across all regions (using a wildcard* for different forms of the same word and spelling out acronyms).

We projected our genetic derived clusters onto the Yeo 7-network parcellation, a popular, low-dimensionality parcellation derived from a clustering analysis of resting state data from 1000 participants[73]. Supplemental Table S6 reports the results of this analyses. The default network was the most common network (4 areas); all but one positively associated cluster from our genetic analysis fell in this network, in line with past research that implicated default network functions to depression [78]. All but two areas (8 of 10 positively and negatively associated areas) fell in networks with higher-level cognitive functions (i.e., default mode, ventral and dorsal attention, and frontal networks).

Discussion

By directly estimating brain areas genetically associated with depression and CU we found (1) the association between CU and depressive symptoms was entirely genetic in origin. (2) Genetic influences on thicker cortex in right DLFPC sulci, the right PreSMA, left medial and

lateral precuneus, OTJ, and TPJ were associated with both traits, and genetic influences on thinner cortex in the right ventral PCC, right medial precuneus, right DLPFC gyrus, and left somatosensory cortex were associated with both traits. (3) Likely molecular pathways are influencing calcium channel depolarization. (4) Likely associated behaviors are "theory of mind", "inhibit", and "pain". (5) Connectivity is associated with default-mode and higher-level cognitive systems. Figure 5 links our results across different methods to the RDoC social dimensions matrix. We discuss the implications of these findings below.

Domain: Social Processes	**Genetics**	**Molecule**	**Cells**	**Circuits (tissues)**	**Physiology**	**Behavior**	**Self-report**	**Paradigms**
Understanding Mental States	Coinheritance of uncaring traits and depression[1]	Ca+[3]	Influx of Ca+ into the neuron[3]	posterior cingulate cortex[1] DLPFC[1] Pre-SMA[1] OTJ[1] TPJ[1] Precuneus non-laterality[1]	Resting state connectivity in Default, frontal executive, and attention systems[2]	Pain[2]	Depressed mood[1] Poor response to others[1] somatic issues[1]	Inhibition[2] Theory of Mind[2]

Figure 5. We used the RDoC "Social Processing: Understanding Mental States" domain dimension matrix to organize our results across different levels of biology and literature. DLPFC=Dorsal Lateral Prefrontal Cortex, Pre-SMA=Pre-Somatosensory Area, OTJ=Occipital Temporal Junction, TPJ = temporoparietal Junction, Ca+=Calcium, positive charge.
[1]Results were estimated directly in this study.
[2]Results were found using MNI coordinates that overlap spatially with those found in the fMRI literature, including the Yeo 7 networks and Neurosynth meta-analytic database.
[3]Results use the Allen Brain Atlas to visualize expression of PGC MDD associated genes.

Advantages of Brain Mapping Approach

We are the first to directly estimate the cortical pattern that represents genetic vulnerability to a psychiatric disorder. Importantly, this approach is not limited to known associations (i.e., brain regions that are not phenotypically associated with depression can reflect genetic vulnerability due to sampling error and environmental associations), and can account for the architecture of genetic effects on brain structure[36]. Further, our approach allows for expansion of hypotheses in genetic association studies by integrating MRI atlas-based approaches to contextualize the genetic association patterns and implicate molecular pathways and brain functions.

In this case, we focused on the vulnerability for CU in depression, chosen due to its importance in depression severity[79] and integration with RDoC domains[1]. Reassuringly, this approach converges on several areas previously associated with depression and social processing literature[57,61], which means past neuroimaging studies of these behaviors may be driven by genetics. However, cortical thickness associations with depression in the temporoparietal and temporo-occipital junctions, key social processing areas, were novel. Finally, we identified likely mechanisms for follow-up analyses using Bayesian meta-analysis, such as theory of mind and inhibition, that are likely targets for behavioral intervention.

This vulnerability reflects an expanded cognitive network. We found clusters specific to the posterior ventral cingulate cortex and DLFPC, which show broad connectivity patterns (functional and anatomical) between limbic/emotional systems and the association cortex(45, 46, 47). Further, almost all clusters were in higher-order cognitive systems. At the molecular level, we implicated positively charged calcium channels. Further informatic analysis implicated

theory of mind, meaning talk therapy may be an effective intervention target, with adaptions of interpersonal mindfulness showing efficacy for depression[83].

Limitations

There are limitations to our approach. First, the sample is tightly matched on age (at around age 28). While this protects against both linear and non-linear confounding by age and developmental effects, results may not generalize to young or old age cohorts. Second, we did not have enough power to explore interactions with sex. Although sex interaction may be a factor in genetic depressive symptomology, there is still a genetic correlation between males and females[54]. Additionally, informatic analyses focused on overlap based on spatial coordinates. While inclusion of results from neuroinformatic tools is more expansive, we did not explicitly model the patterns of association between RNA transcripts, inferred behaviors, etc.

Conclusion

We directly mapped genetic vulnerability to CU and depressive symptoms on the brain. We found common genetic variance in CU and depressive symptoms was associated with higher-order cognitive areas and functions. As the genetic vulnerability to psychiatric disorders is discovered, the use of high-resolution cortical methods will be invaluable in contextualizing the patterns of genetic effects.

Chapter 3: Inferring the Genetic influences on Psychological Traits Using Whole-Brain Models: Application with Functional MRI Connectivity

Chapter Summary

Genetic correlations between brain and behavior phenotypes in major genetic consortiums have been weak and mostly non-significant. Integrating more complex fMRI procedures will improve our capability and utility to conduct gene-finding studies of neuropsychological outcomes, even expanding beyond measured phenotypes to predicted ones. In particular, "Connectivity-Based Predictive Modeling" is an approach for creating brain-based proxies of a neuropsychological variables. To this end, we compare different approaches for predicting intelligence (IQ) and test their effectiveness as endophenotypes by developing predictive models of IQ in a sample of 3,000 individuals and test their SNP genetic correlation with measured IQ in a sample of 13,092 individuals. We compare the additive connectivity-based model to the LASSO and Ridge models phenotypically and genetically in the full sample and develop learning curves to demonstrate their effectiveness in different sample sizes in the UK biobank parcellation. We also compare these approaches to single "candidate" brain areas. We find that predictive models capture about half of the genetic variance underlying IQ, which far exceeded the phenotypic overlap. LASSO and Ridge were both slightly more predictive at the phenotypic and genetic levels, but the additive model had the most power to detect polygenic signal. We assert that predicted behavior can be an effective GWAS phenotype, improve gene-finding efforts, and another key application of whole-brain models.

Introduction

Genetics and MRI imaging are perhaps the two most utilized biostatiscal approaches in studying psychological phenomena. In addition, both whole-genome analyses[84,85] and whole-brain analyses[86] are expanding the inferences and prediction we can make about psychological traits. In genetics, the standard discovery procedure is a Genome-Wide Association Analyses (GWAS) of large (10's-100's of thousands) samples, agnostic to a proiri hypothesized SNP associations. However, the power to detect variants in GWAS is incredibly small, which has led to a coarse phenotype approach to create larger sample sizes. We are now beginning to understand the consequences of this movement, as coarse phenotypes can often make results difficult to interpret and can lead to less specificity in understanding behavioral traits, like depression[87].

One early hypothesized approach to improve genome-wide discovery and genome-wide theoretical interpretation is to use intermediary traits, or endophenotypes, traits that are more closely related to the genetic expression of a distal behavioral outcome and likely to be more heritable[88]. For example, hippocampal volume is decreased in individuals with schizophrenia and non-affected related family members[89] and likely closer to genetic action than schizophrenia. Thus, it was thought that a GWAS of hippocampal volume should capture the genes influencing schizophrenia.

For years, the endophenotype approach presented an attractive alternative to the coarse approach. The most ubiquitous endophenotype for psychological and neurological phenomena is the brain, or imaging genetics studies. Figure 1 shows a chart of the terms "endophenotype" and "imaging genetics" that have appeared since 2010, showing thousands of articles published on the topic each year. Further, imaging is a useful vignette in the endophenotype literature as large

GWAS cohorts are being assembled for MRI imaging phenotypes and offers some insight into the relative success of endophenotypes to date.

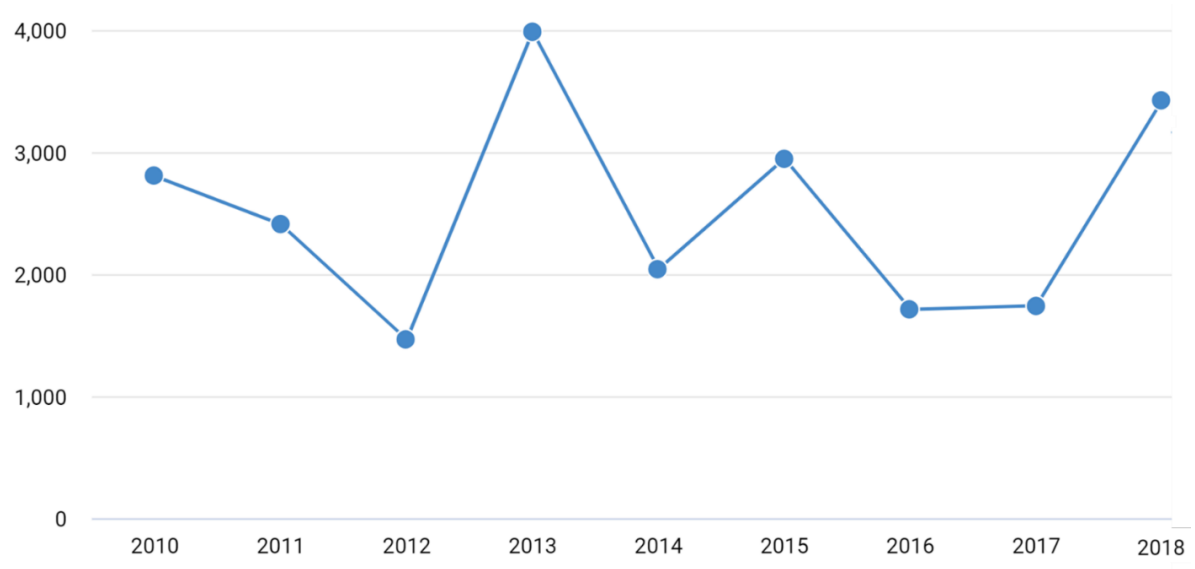


Figure 1. Rate of words "Imaging Genetics" and "Endophenotype" appearing together by google scholar search across 8 years. Years are organized on the x-axis and number of publications on the y-axis.

Unfortunately, the two goals of endophenotypes, prediction of new GWAS variants and explanation of mechanisms, have been largely unsuccessful. No new genetic variants have been discovered due to imaging endophenotypes (yet)[90]. Further, genetic associations between well-known brain endophenotypes and traits are low and non-significant in the largest and most powerful studies to date[25], weakening the claim that current endophenotypes are explanatory tools for scientific theory.

We argue that the issues with the imaging endophenotype approach can be (partially) alleviated by considering a more whole-brain approach to endophenotypes, i.e. it is more likely that a whole-brain model will capture (more of) the genetic variance underlying behavioral phenotypes than a single brain measure. Essentially, with the current approach, specific

associations are found in the literature and then established as endophenotypes based on univariate tests of their association with the outcome and in related family members, akin to a "candidate endophenotype" method. It is this approach that has failed to live up to the expectations of the endophenotype claims. Instead, whole-brain approaches should be leveraged for GWAS discovery.

Whole-Brain Models vs. Candidate approaches

While there are many neuroimaging brain-to-outcome associations found in the literature, many of these have failed to make impact on clinical practice. Woo, Chang, Lindquist, & Wager (2017)[91] offered a possible solution through the use of whole-brain predictive modeling. These approaches are preferred to candidate brain regions as they are more powerful omnibus summaries for the whole-brain, allow for flexible tests of reproducibility, incorporate more information and are more predictive than standard approaches.

Expanding, we argue that the prediction by these whole-brain models can be utilized as a GWAS phenotypes. In other words, we propose "whole-brain predictive endophenotypes" as new path forward in the gene discovery literature. While complete literature reviews of these whole-brain models are available elsewhere[86,92], models of outcomes such as pain responsivity[8] and sustained attention[9], have been developed and shown to replicate across large cohorts. Either of these models and numerous others could offer a theoretically relevant endophenotype closely tied to its underlying biology and useful for GWAS studies. Finally, while these phenotypes have statistical and theoretical advantages to standard approaches, they also give the added advantage of offering many new phenotypes for GWAS discovery than those measured, for example conducting a GWAS of sustained attention patterns to understand ADHD.

Connectivity-Based Predictive Modeling (CBPM)

Brain-imaging has many measurement approaches, or "modalities", that can be used to generate measures of the brain, i.e. aspects of anatomical size or flow of oxygenated blood to the brain can be used separately, or in combined analyses. Thus, a first step in developing a whole-brain model is the choice of brain data. For our demonstration we used Connectivity-Based Predictive Modeling (CBPM). Connectivity-based predictive models use the correlation in the time-course of oxygenated blood flow between two brain regions within each individual as a feature to train a predictive model[93]. For example, one feature could be the correlation in blood flow between the dorsal-lateral prefrontal cortex and the cingulate cortex, and each *jth* person gets a correlation between the regions as a "score" for their connectivity. These patterns are collected from low-frequency oscillations when the individual is at rest, i.e. not during a task. We chose CBPM because: many popular models across samples, such as the sustained attention and fingerprinting models[9,94], are replicable, the procedures can be relatively simple[93], resting-state data is being collected across many groups and resting-state data is easier to align than task activation. Finally, many approaches to resting-state data include a parcellation that was derived from the data, like Independent Components Analyses (ICA), making them powerful examples of feature engineering, a key step in predictive models that often improves predictive performance.

This Study

We demonstrate this approach using CPBMs of intelligence (IQ) in the UKBiobank, using the UKBiobank functional connectivity parcellation. We chose this parcellation because it was estimated in one of the largest sample sizes to date[95], and because this sample is publicly available and large enough to estimate predictive models for GWAS discovery, so other groups

could engage in the techniques herein. Further, the UKbiobank is also one of the largest samples of MRI, GWAS, and cognitive data, allowing for an effective sample size to demonstrate our claims.

For predictive models to be useful endophenotypes, they should be useful as GWAS discovery targets, associated with the trait of interest[88], and be theoretically relevant, interpretable, and generalizable[96]. We demonstrate these points in two parts. In part 1, we test whether CBPMs are effective endophenotypes for GWAS discovery. First, in a sample of 3,000 individuals from the UK biobank we develop three predictive models of IQ with three different machine learning algorithms. Next, to test the effectiveness of the prediction of these three models as endophenotypes, in a genetic validation sample of 13,092 unrelated individuals we estimate the phenotypic and genetic correlation between predicted and measured IQ and conduct a genome-wide association study of each model's prediction and interpretability. In this case, we use the larger sample as validation to have more power to estimate genetic correlations and more power for comparison of whole-brain measures and single brain measures.

In part 2, we compare the interpretability of these models for theory building across a range of sample sizes. To compare the different algorithms, we use the 13,902 individuals as a training set and the 3,000 individuals as a test set to estimate learning curves based on increasing sample size for each algorithm to compare their predictive ability in training and test sets.

Method

Participants.

UKBiobank. The UKBiobank is a large population cohort of over 500,000 individuals. We restricted our analysis to 16,092 individuals who had MRI data, IQ scores, and who were of European descent. We split the sample (randomly) into a sample of 3,000 (S3K) individuals and

13,092 individuals, for different aspects of the project. Though non-intuitive, we chose to use the 13,092 (S13K) individuals as a genetic validation set as this sample size had appropriate power to detect genetic correlations[97]. Further, the sample of 3,000 is orders of magnitude larger than most samples previously used in MRI research to develop predictive models[9,92]. Because inference about the performance of these techniques would benefit from larger sample sizes, to estimate learning curves at increasing sample sizes (phenotypically), we use S13K to estimate the model and S3K to test the three learning techniques phenotypically in a post-hoc analysis.

IQ in the UKBiobank was measured at four time points, three in person assessments and an online assessment. All tests were two-minute assessment of fluid intelligence via a handheld device. Participants had 2 minutes to answer as many questions as possible in a sequence of 13 questions flashed across the screen. We estimated a latent factor model across all four time points in the full sample of ~500,000 individuals and extracted the factor score as a measure of IQ to increase reliability due to the short assessments. Data from individuals who "abandoned" the later time points were treated as missing in factor analysis. Factor scores from the imaging sample are used in this analysis.

Brain data: Parcellation and Measurement

The UK Biobank project provides highly summarized rs-fMRI data in the form of 2 full connectivity matrices per participant –between 25 and 100 time courses derived from large independent components analyses (ICAs) conducted with FSL's MIGP-Melodic tool. These correlations within each individual represent the training features/variables of analysis, typically called "edges" in the literature. Because these 25 and 100 dimension parcellations include noise components, the actual number of signal components is lower than 25 and 100. We chose the 100 dimension parcellation which was reduced to 55 signal components based on an analysis by

the UK Biobank imaging group. The resulting data, per participant, was a full 55 x 55 correlation matrices, leaving us with 1485 functional connections/edges after excluding the diagonal in each. Each Edge was residualized on age and sex before training.

SNP data processing and Associations

We used UKBiobank participants who were imputed to the Haplotype Reference Consortium (McCarthy et al. 2016 Nature Genetics), 1000 Genomes and UK10K reference panels by the UK Biobank[98] and of European descent. Subjects were genotyped on a UK BiLEVE array or the UKBiobank axiom array. After removing individuals with mismatched self-reported and genetic sex, we filtered imputed SNPs using a Hardy-Weinberg equilibrium P value threshold of $>1\times10^{6}$, variant missingness > 0.05, imputation quality score (INFO) >0.95, and minor allele frequency (MAF) above 0.01, retaining 7,391,068 SNPs. More information is available in publication (Bycroft et al. 2018 Nature).

Training Procedures: Different scores generated and studied here

In this work, we generate three scores to test possible avenues to using whole-brain predictive models as endophenotypes. First, we adapt the Shen et al. 2017 procedure, which considers each edge of the connectome as an individual additive predictor. To do this, the procedure uses sum scores of positively and negatively associated edges (above an associated p-value threshold) in a linear model to predict the outcome of interest. We test the Shen et al. model at p-value cutoffs of .05 in initial analysis, as this is the typical value chosen in the literature[9]. In the post-hoc analysis, we use learning curves to demonstrate how adjusting the p-value changes the prediction.

The second and third models were based off the procedure from Tobyne et al. 2018. Specifically, cross-validated ridge regression (a.k.a. L2 regularization) estimates to train connectivity-based models. We train ridge regression and LASSO regression (Least Absolute Shrinkage and Selection Operator, a.k.a L1 regulations) with 10-fold validation, implemented in the R package glmnet[99]. We conducted a grid search of values from 0 to 10,000 for optimal regularization parameters (called "Rho" or "Lambda" in the literature) for each machine learning model. To interpret the predictive models, we extract the weights for each edge and from each model. We then plot the weights against the Yeo 7 resting state networks[73], a commonly used resting state parcellation.

To see how effective these models are at predicting IQ, in the training 3SK sample we trained a ridge regression, LASSO Regression, and Shen CBPM and then used the weights extracted from the 3SK sample to predict IQ in 13SK genetic validation sample. After estimation of the scores we used a Pearson's correlation to determine phenotypic association between predicted IQ and measured IQ, as we thought Pearson's r would be most comparable to the genetic correlation (rG).

Part 1: Test of scores as GWAS proxies for measured phenotypes

Univariate Heritabilities. In the 13SK sample, we used a mixed-effects model procedure through BOLT-LMM to estimate the univariate heritabilities of predicted IQ scores for the ridge regression, LASSO regression, and Shen et al. procedure. We also compared them to the heritability of measured IQ in the same sample.

Bivariate Heritabilities. To determine if the GWAS proxies were capturing the same genes as measured IQ, we estimated the genetic correlation between measured IQ and out-of

sample predicted IQ in the 13SK genetic validation data set. To do this, we used Bolt-LMM to estimate the genetic correlations between all three predicted IQ scores and measured IQ, separately.

To compare the genetic correlations between whole-brain models and individual brain regions, we ran a test of association (controlling for sex and age) in the 3SK sample of each edge and IQ. After Bonferroni correction, 6 individual edges were significantly associated with IQ (Table 2) in the 3SK. We then estimated the genetic correlation between each of these 6 edges and IQ in the 13SK. We compare these to the whole-brain correlations and correct for multiple testing in the genetic correlations (9 tests) using Bonferroni.

Genome-wide Association Analyses. To see if these scores find useful GWAS discoveries we then used BOLT-LMM (with leave on chromosome out) to conduct a mixed-effects genome-wide discovery for IQ and each predictive model in the 13SK sample. We chose BOLT-LMM to increase power and account for the polygenic background of our trait[100]. For each GWAS, we compare the genome-wide discovery via a Manhattan plot of individual SNP p-values and also explore the observed p-values deviation from expected p-values using QQ plots.

Genetic Correlations with IQ covariates Finally, to see if predicted IQ from functional connectivity showed similar patterns of genetic correlation as IQ, we tested genetic correlations between summary statistics of the GWAS for each IQ prediction (Ridge, Lasso, and Shen CBPM) and related traits through the LDhub[97,101]. We chose 7 traits that have been previously shown to be genetically correlated with IQ in a major meta-analysis[102] and were theoretically relevant to IQ. We considered IQ proxies (educational attainment[103]), or psychiatric covariates of IQ (autism and depressive symptoms[103]), anthropometric traits (infant head circumference[104], height[105]) or evolutionary linked traits (age of first birth[106]), as these are the typical types of

hypotheses tested using genetic correlations between IQ and a covariate of IQ. We did not consider schizophrenia and IQ genetic correlation because the sample was mixed ethnicity. We also did not run exhaustive LD score associations between predicted IQ and all trait summary statistics available online to reduce the number of tests.

Testing generalizability of Brain Predictive Models for IQ.

To test effective sample sizes for out of sample prediction of IQ, we estimated learning curves for each algorithm at increasing sample sizes. To preface results, we chose to do this because more predictive models were also more genetically associated with measured IQ. We plotted the canonical correlation between predicted IQ and measured IQ in both the training set and test set increasing sample sizes of 50 to 13092 individuals in increments of 50. We used the S13K as the training and S3K as the validation set for this analysis to increase the spectrum of sample sizes tested. We estimate learning curves for the (1) ridge, (2) LASSO, (3) Shen et al. at p-value threshold of .005, (4) p-value threshold of .01, (5) p-value threshold of .05, (6) p-value threshold of .10, and (7) no p-value threshold (summing up all edges). Importantly, this is the first work to compare the Shen et al. predictive models at varying p-value thresholds.

Results

Phenotypic prediction

We began by training our models in the S3K sample and testing their predictive accuracy in the S3K and out-of-sample S13K. Table 1 shows the results from the phenotypic prediction in both samples; measured by the correlation between measured and predicted IQ. All three models showed significant phenotypic prediction in the training and test set. Phenotypic prediction in the training set varies between r=.322-.476. In the test set, the prediction ranged

between r=.187-212. In both the train and test set the ridge regression was the most predictive model.

Table 1. Phenotypic Prediction of IQ From Each Algorithm in the Training and Test Set

Algorithm	Training *r*	Se Lower	Se Upper	P-value
Lasso Train	0.380	0.349	0.410	<2.2e-16
Ridge Train	0.476	0.447	0.503	<2.2e-16
CBPM Train	0.322	0.290	0.353	<2.2e-16
Lasso Test	0.188	0.172	0.205	<2.2e-16
Ridge Test	0.212	0.195	0.228	<2.2e-16
CBPM Test	0.187	0.171	0.204	<2.2e-16

Note. Phenotypic prediction of IQ by each algorithm, measured as the Pearson's r between predicted IQ and measured IQ. Training was done in the 3SK sample and testing the 13SK sample. The top three columns represent the algorithm predicting in the sample it was trained in, the bottom three represent the out of sample prediction going from the 3SK to the 13SK sample.

Predictive Features of CBPM

To demonstrate the utility of this approach for theory building, we plotted edges included in the Shen et al. model and weights for the ridge and LASSO regression against the Yeo 7 resting state-network parcellation. Figure 2 shows the results for each. Due to the large sample size and number of edges, both the Shen et al. procedure and the ridge lack interpretability. In contrast, the LASSO was highly interpretable, as only a handful of edges were needed to account for similar proportions of variance as explained by the Shen et. al. model and the ridge. The LASSO weights were largely negative weights on edges connecting the default-mode network

and other networks connections, i.e. more connectivity between the default-mode network and attention and frontal networks is predictive of lower IQ (by the LASSO output). The edges were anatomically largely default-mode edges in the posterior cingulate cortex, and frontal-parietal connections to the default-mode network.

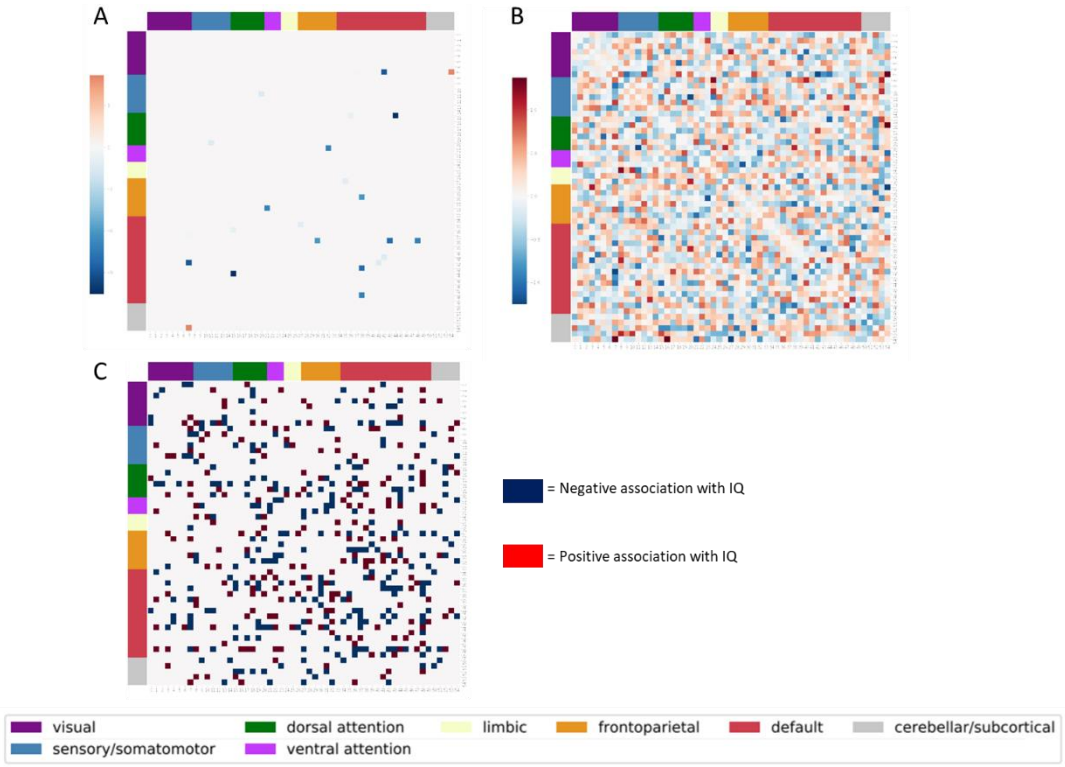


Figure 2. Weights implicated by each predictive modeling technique organized by the Yeo 7 resting state network to increase their interpretability. (A) Weights from the LASSO regression model. (B) Weights by the ridge regression model. (C) Weights that went into the positive (Red) and negative (navy blue) sum scores for the Shen et al. procedure.

Molecular Genetic Analyses

Univariate Heritability. First, to ensure the heritability of these scores with this data and check for the ridge regression heritability in out of sample, we used Bolt-REML to run a univariate heritability analysis of each model's prediction. All three model predictions were significantly heritable (Table 2), and had similar heritability at ~.15. We also tested the heritability of measured IQ in this sample, which was .28 and higher than any predicted model.

Genetic Correlations between Predicted and Measured IQ. Next, we estimated the genetic correlation between measured IQ and the prediction of IQ by each method. All models were significantly genetically correlated with measured IQ, with moderate genetic correlations ranging from .422-.576 (Table 2). The genetic correlations followed the same pattern as the phenotypic associations, and the ridge was most genetically correlated with measured IQ. Importantly, this means that *more than half of the genetic variance influencing IQ* can be captured by Ridge regression and LASSO regression predicted IQ scores. Considering that various measured IQ cohorts from Savage et al.[102] found an average rG of .67, our predicted IQ variable is approaching utility as a GWAS IQ cohort.

Next, we wanted to compare genetic correlations between single edges associated with IQ to genetic correlations from whole brain models. Table 2 shows the results and which UKbiobank edges were significantly phenotypically associated with IQ (notably, all of these edges were edges weighted above zero by the LASSO and shown in Figure 2), post Bonferroni correction. All six single edges were theoretically relevant, the top associated edge was a connection of the angular gyrus, known in the literature from smaller samples[107], and several edges were part of the posterior cingulate cortex and the task negative system, known to be negatively correlated with cognitive abilities[108].

Table 2. Heritability and Coheritability of Measured IQ and the Prediction of IQ by Each Modeling Procedure

Edge	Heritability	SE	rG	SE	rE	SE	rG CI Corrected
Angular Gyrus to PCC	0.104	0.031	-0.226	0.130	-0.037	0.028	±0.361

Frontal Parietal to PCC	0.070	0.030	-0.353	0.167	-0.008	0.027	±0.464
Frontal Parietal to unknown	0.087	0.030	-0.130	0.141	-0.064	0.028	±0.390
DLPFC To middle Temporal	0.102	0.030	-0.252	0.133	0.004	0.028	±0.368
Temporal junction to FP	0.118	0.030	0.105	0.122	-0.049	0.028	±0.339
Middle cingulate to PCC	0.038	0.029	-0.310	0.241	0.009	0.027	±0.669
CBPM*	0.181	0.031	0.422	0.092	0.123	0.029	±0.256
Lasso*	0.143	0.031	0.518	0.105	0.107	0.028	±0.290
Ridge*	0.155	0.031	0.576	0.100	0.116	0.028	±0.278

Note. The heritability of each edge that was significantly associated after Bonferroni correction and the prediction of IQ and coheritability between each edge and each predicted model of IQ with measured IQ. rG = genetic correlation between the IQ predictor and measured IQ, standard errors (SE) follow their respective statistic. We Bonferroni corrected the confidence interval to account for multiple testing. PCC= posterior cingulate cortex, FP = frontal pole, DLPFC = dorsal lateral prefrontal cortex. *P < .05 rG with predicted IQ post Bonferroni correction.

After correction for multiple tests, no single edge was significantly genetically associated with measured IQ (9 tests). Nominally, all genetic correlations between single edges and measured IQ were lower than genetic correlations between predicted and measured IQ. In line with our hypothesis, single edges showed non-significant and weaker genetic correlations, in comparison to models that combined across multiple edges via a predictive algorithm.

GWAS of Predicted IQ. To test if predicted IQ is a useful GWAS proxy, we ran a GWAS of the prediction by each machine learning algorithm as the phenotype and compared it to a GWAS of intelligence in the same individuals. No GWAS found significant SNPs above genome-wide p-value significance (Figure 3 for Manhattan plots of each of the 4 traits). This

suggests very large GWAS samples are going to be needed to detect significant variants of predictive models. We used the QQ Plots of estimated vs. expected p-values to probe the power from each of these models' predictions and measured IQ (Figure 4). Both the LASSO and the Ridge Regression have similar polygenic signals as IQ. However, the Shen et al. procedure led to the greatest deviation from expected p-values, arguably this method may offer more power to detect variants than a standard GWAS phenotype but, the increase would only be small.

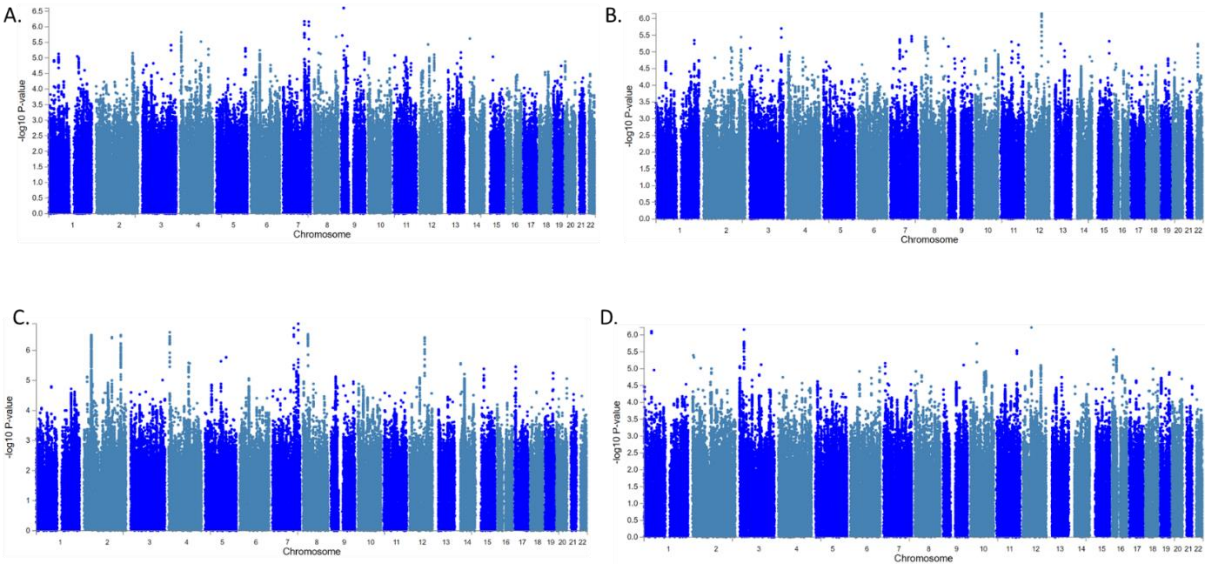


Figure 3. Manhattan plots for genome-wide association discovery across the 3 predictive models (A-C) and measured IQ (D) in the same individuals from the 13SK sample. P-values of each SNP are organized by chromosome on the x-axis and by the -log10 of the p-value on the y axis. No p-values were significant below genome-wide discovery correction (p = 5e-8). Panel A, IQ predicted by ridge regression. Panel B, IQ predicted by LASSO regression. Panel C, IQ predicted by Shen et al. procedure. Panel D. measured IQ in the UKbiobank.

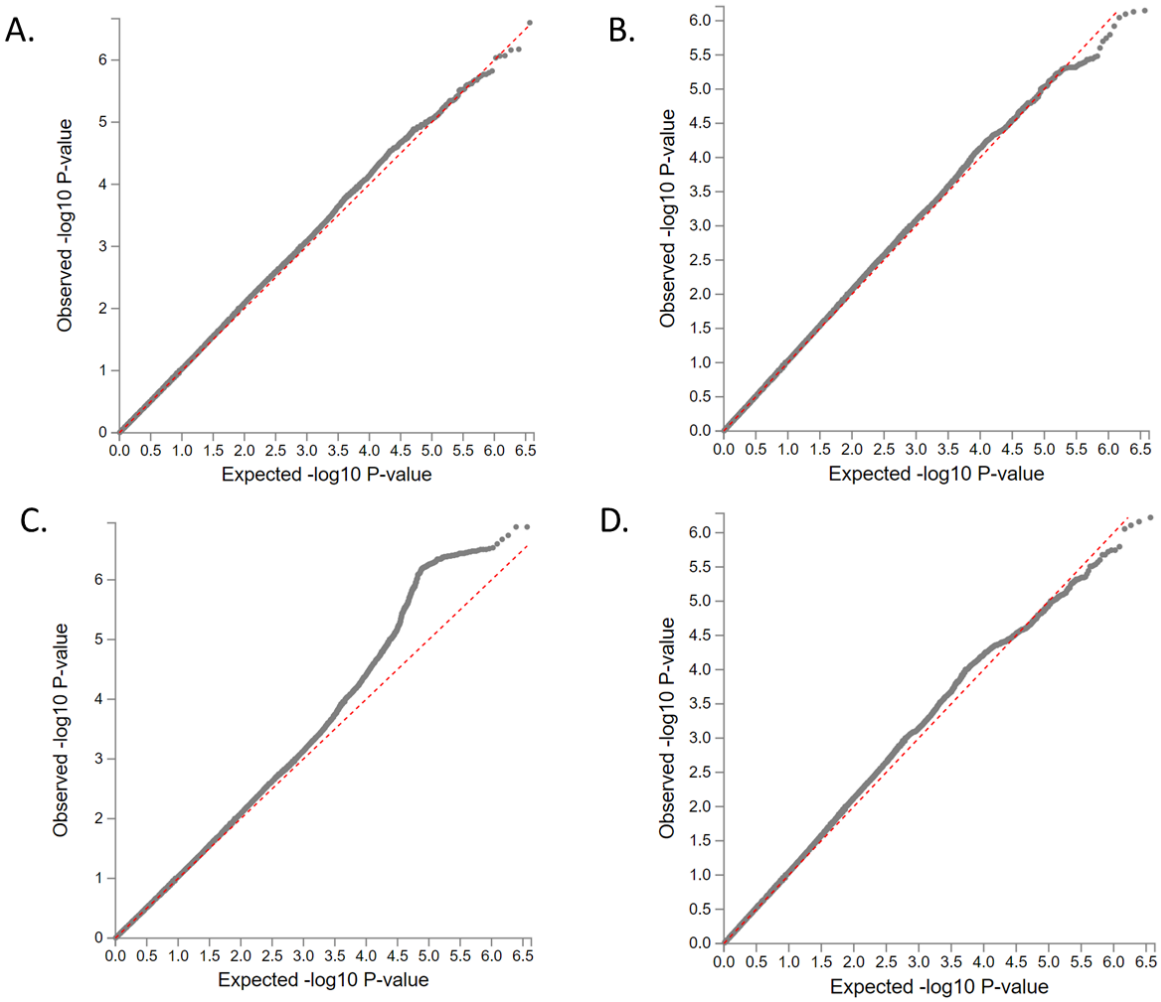


Figure 4. QQ-plots for the p-values from the genome-wide association discovery across the 3 predictive models (A-C) and measured IQ (D) in the same individuals from the 13SK sample. P-values are the dotted line plotted based on their expected -log 10 p-value on the x-axis and by their observed – log 10 p-value on the right access. The dashed line is the expected p-value distribution under the null model. Deviation above the line represents signal more significant than expected.

Genetic Correlations between Predicted IQ and Correlates of IQ. To see if predicted IQ endophenotypes also were genetically correlated with phenotypes genetically correlated with measured IQ, we conducted LD score regression genetic correlations of each predicted model with 7 key traits of interest: years of education, college completion, depressive symptoms, autism spectrum disorder, height, infant head circumference, and age at first birth. We compared these

genetic correlations between predicted IQ and traits of interest to genetic correlations of measured IQ (in the same individuals) and these same 7 key traits of interest. Table 3 shows the results. Broadly, the three predictive models showed a strikingly similar pattern of association with IQ correlates, probably due to their high genetic correlations with one another (Ridge-Lasso rG=.946, CI=±.033; Ridge-Shen rG=.902, CI=±.031; Lasso-Shen, rG = .843, CI=±.044). IQ was more strongly genetically correlated (nominally) with: Educational attainment (and significantly), height and age at first birth, to be expected from a proxy measure. However depressive symptoms, autism, and infant head circumference were more correlated with predicted IQ than measured IQ (nominally).

Table 3. Genetic Correlations Between Predicted IQ and Known Genetic Correlates of IQ

Ridge	**rG**	**Se**	**Z**	**p-value**
Years of schooling	0.353	0.082	4.311	<0.001
College completion	0.420	0.120	3.493	0.001
Depressive symptoms	-0.456	0.139	-3.285	0.001
Age of first birth	0.289	0.090	3.206	0.001
Childhood IQ	0.453	0.154	2.941	0.003
Infant head circumference	0.638	0.222	2.878	0.004
Height	0.023	0.062	0.370	0.711
Autism spectrum disorder	0.246	0.131	1.877	0.061
Lasso				
Years of schooling 2016	0.381	0.113	3.359	0.001
College completion	0.397	0.148	2.687	0.007
Depressive symptoms	-0.440	0.176	-2.495	0.013
Age of first birth	0.238	0.111	2.146	0.032
Childhood IQ	0.573	0.232	2.470	0.014
Infant head circumference	0.747	0.286	2.614	0.009
Height	0.011	0.074	0.144	0.886
Autism spectrum disorder	0.263	0.170	1.550	0.121
Shen Procedure				
Years of schooling	0.344	0.079	4.356	<0.001
College completion	0.391	0.121	3.230	0.001
Depressive symptoms	-0.428	0.148	-2.891	0.004

Age of first birth	0.248	0.095	2.613	0.009
Childhood IQ	0.298	0.159	1.871	0.061
Infant head circumference	0.676	0.242	2.793	0.005
Height	0.086	0.072	1.192	0.233
Autism spectrum disorder	0.195	0.127	1.529	0.126
Measured IQ				
Years of schooling 2016	0.618	0.062	9.936	<0.001
College completion	0.653	0.087	7.481	<0.001
Depressive symptoms	-0.239	0.087	-2.740	0.006
Age of first birth	0.446	0.072	6.172	<0.001
Childhood IQ	0.780	0.151	5.184	<0.001
Infant head circumference	0.438	0.151	2.895	0.004
Height_2010	0.182	0.051	3.567	<0.001
Autism spectrum disorder	0.136	0.095	1.436	0.151

Note. Genetic correlations (rG) between predicted IQ and genetic correlates of IQ, compared to genetic correlations with measured IQ. Genetic correlations were estimated using LD Score regression. Summary statistics were recovered through GWAS. Data on IQ genetic correlates are from summary statistics of major GWAS consortia.

Utility of Predictive Models Across Samples

Next, we conducted post-hoc analysis to see what sample sizes would be needed to use in development of CBPM in the UKBiobank parcellation. Here, we also allowed the p-value threshold of the Shen et al. procedure to vary. Figure 5 shows the results of learning curves training the S13K and predicting in the S3K sample. With this parcellation, more than 1000 individuals are needed to reduce overfitting of the machine learning procedures. In line with results from genetic predictive models[109], the Shen et al. count model seems to perform the best in smaller sample sizes (less overfitting) but becomes less useful as the sample sizes increase. The LASSO regression outperformed the Ridge regression at small samples and performed weaker but with similar prediction at large sample sizes. Thus, in this study LASSO regression offered interpretability and was a useful predictor across a range of sample sizes, meaning that LASSO is like a well-balanced procedure for future work. Finally, as sample sizes became large, the ridge prediction slightly outperformed the LASSO, but still showed more overfitting.

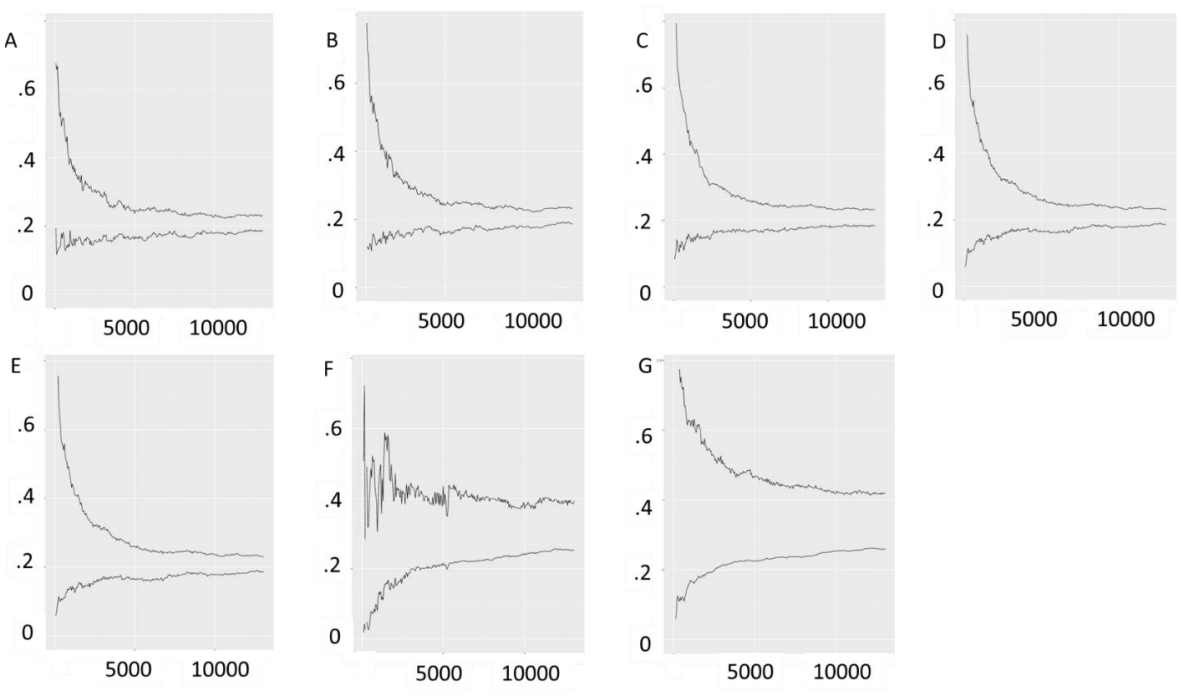


Figure 5. Learning curves for increasing sample sizes of predictive models of intelligence. The top line is the curve in the training set as sample sizes increases in the training set. The bottom line is prediction in test set of 3000 individuals as training sample size increasing. (A) Shen et al. p-value threshold =.005. (B) Shen et al. p-value threshold =.01. (C) Shen et al. p-value threshold =.05, value that is often chosen. (D) Shen et al. p-value threshold =.1. (E) Shen et al. procedure at p-value threshold =1. (F) LASSO regression training curve. (G) Ridge regression training curve.

Discussion

We demonstrated the application of whole-brain predictive models as GWAS proxies for discovery and theory building. We also presented novel innovations for estimation of the scores phenotypically with CBPM procedures outlined by Shen et al. 2017 and compared different approaches for developing these models in the UKBiobank resting state connectivity parcellation. We found that predictive modeling was an effective way to generate endophenotypes. LASSO regression developed endophenotypes that were interpretable and predicted half of the genetic variance underlying IQ in out of sample predictions. The Shen et al. procedure seemed to offer the most power to detect polygenic signal, based on deviation from expected p-values. Conclusions for this approach are presented below.

Whole-brain Endophenotypes and Unmeasured Variance

We established that whole-brain predictive models are useful endophenotypes by genetic correlations, and should be applied to work in the future. One application of this approach could be to expand the range of possible behavioral phenotypes to conduct GWAS of traits in these larger samples. For example, in the imaging literature there are examples were brain models of dichotomous traits with small samples are combined with brain models from continuous measured mechanisms to increase power, which has led to increased predictive accuracy for fibromyalgia using a pain sensitivity model[110]. This could expand the number of GWAS phenotypes and direct efforts to even predict particular mechanisms underlying these traits.

When endophenotypes were first purposed as useful biomarkers, it was believed they would play a key role in gene discovery[88]. While they largely were unsuccessful at this goal, they have transitioned into useful mechanisms for understanding underlying psychological variability[90]. Another way to expand gene-finding is to conduct GWAS of these predictive models that show theoretical relevance. As many of connective models reflect behavioral phenotypes close to brain and mechanisms in behavior (i.e. sustained attention in ADHD), GWAS of these models may offer a path forward in understanding the interplay of behavioral and genetic mechanisms in downstream health and wellness outcomes.

While it is often argued that predictive modeling approaches are difficult to interpret, plotting the weights can lead to insights into the brain systems underlying cognitive and behavioral phenotypes. In line with modern endophenotype approaches, we also demonstrate how prediction across the connectome is useful for theory building; and align our approach with the common Yeo 7 parcellation to establish functional implications of these associations. Specifically, using LASSO regression, we found that default-mode to other network connections

had high utility in predicting IQ, in line with the current literature[107,108], and also implicated to several nodes that were associated after multiple correction in univariate tests. Thus, our approach aligns with modern endophenotype research and brain discovery but importantly, our approach does not limit discovery to single brain regions; interestingly more edges were weighted in the LASSO then were associated via correction after the univariate tests of association and the full weighted model showed much stronger genetic associations than any single edge.

Further, CBPMs can be used to find genetic correlations predicted outcomes and covariates of the trait of interest. In this study, traits genetically correlated with IQ were genetically correlated with the predictions of IQ. Interestingly, infant head circumference and depressive symptoms were more genetically correlated with predicted IQ than measured IQ. This may mean that the association between IQ and connectivity is strongly mediated by these phenotypes or vice versa. For example, it appears that the main edges driving the prediction of IQ (according to the LASSO output) were in the default mode network, which is thought to play a key role in depressive symptoms[78], therefore it is likely that default mode connectivity genetics is influencing the overlap between intelligence and depressive symptoms.

Predictive Modeling Insights from This Study

Based on patterns of results, we recommend that CBPMs utilize large samples. After samples exceeded ~1,000 individuals, when comparing the different approaches, the standard CBPM (Shen et al. 2017) seemed to find solutions that worked effectively in smaller samples, with adjustments on the p-values only providing marginal change to the model performance. LASSO regression was useful across a range of sample sizes. Ridge regression seems to require very large sample sizes to avoid overfitting. The sum score approach outperforming in smaller

samples, but regularization in larger samples mirrors results from the statistical genetics literature[111]. The key difference being that much smaller sample sizes are needed for the same level of prediction in whole-brain models then polygenic risk scores.

Further, there are many neuroimaging modalities across the functional and anatomical literature and possible parcellations that can be used for whole-brain predictive modeling. This study establishes the utility of whole-brain modeling with the UKB functional connectivity parcellation and CBPM procedures. The procedures here predicted some, but not all, of the genetic variability underlying IQ. Though it would be outside the scope of any one manuscript to test all possible ways of predicting an outcome and their genetic associations, in the future, many other sources of data and brain parcellations may be used to further expand on this method. Woo et al.[22] offer a procedure for selecting the best model that involves first testing different models in the discovery sample, selecting the one that is predictive (and makes theoretical sense) and then holding out the validation just for that model in particular. Using procedures like this that have been established in imaging may improve use of endophenotypes in the future.

Limitations

There were a number of limitations to the study. First, our training and validation sets were closely aligned as they are both part of the UKBiobank. It is possible that these results may generalize poorly to other parcellations or subgroups (like younger cohorts) that were not used in generation of the UKBiobank parcellation. Consideration of brain parcellation and feature engineering is an important step in choosing predictive models, and this work does not argue against the importance of those considerations.

In our gene finding we focused on common variants (MAF > .01) and highly probable variants (INFO > .95). We chose these as we were running several GWAS studies and we wanted to conduct reasonable genome-wide screening, but also reduce the computational burden. This choice means, however, that we cannot speculate at how rare variants are likely influencing whole-brain models.

Conclusions

Whole-brain models will be useful endophenotypes for neuropsychological outcomes in the future. Connectivity-based models performed well for predicting IQ in this study, predicting about half the genetic variance underlying IQ. Future expansions should consider other neuroimaging modalities as ways to further improve these scores.

Chapter 4: Executive Functioning GWAS of over 427,000 Individuals Establishes Molecular Pathways of Neurocognitive Processing

Chapter Summary

Executive functions (EFs) are top-down cognitive control mechanisms that enable goal-directed behavior. While EF is likely neurological, and deficits are broadly associated with brain disorders, little is known about the molecular underpinnings of EF individual differences. Furthermore, genome-wide studies of EFs use individual tasks, which are impure measures of higher-order processes. Multiple tasks can be used to tap general EF and probe its genetic associations with health and behavior. We conducted a GWAS of Common EF (cEF), measured with multiple tasks (n=93,024~427,037 across tasks) in the UK Biobank. 10,122 significant SNPs were discovered across the full sample. 394 SNPs were independently associated in different sub-samples and were on chromosomes 1, 3, 6, and 8. Gene-based analysis found neuronal, potassium channel and GABA pathways associated with cEF. We found genetic correlations between cEF and almost all psychiatric traits, behavioral traits and health outcomes. This work presents a molecular profile of cEF.

Introduction

A hallmark of dysfunction across neurological and behavioral disorders is impairment in neurocognitive executive functioning (EF), or the ability to control and influence one's thoughts and actions[112]. Though EF abilities can be measured across the general population[113], EF is in an important dimension of clinical neuroscience and is associated with several brain disorders, including neurological disorders like Alzheimer's[114], vascular dementia[115], and lateral sclerosis[116] and almost all psychiatric disorders, including schizophrenia[117], depression[66], ADHD[118], antisocial personality disorder[119], sleeping dysfunction[120], Suicidal ideation[121] and with common variability across psychiatric symptoms[112,122,123]. Further, in studies of schizophrenia, lower scores on EF tasks relate directly to patients' daily functioning[124], rate of hospitalization, and symptom severity[125], and in studies of Alzheimer's disease, better EF predicts better daily functioning[126]. Thus, EF distinguishes cases from controls and (for some disorders) relates to degree of disorder impairment.

Past twin and family studies have established that EF is heritable in childhood[127], early adulthood[128] and middle age[129], and the genetic variance influencing EF is stable across multiple time points[130]. Further, twin studies have shown that EF relates genetically to several different psychiatric disorders[66,131] and behavioral dimensions of health, like sleep[120]. However, little is known about the molecular underpinnings of EF in humans. Genome-wide association studies (GWAS) are an excellent way to characterize specific single nucleotide polymorphisms (SNPs) influencing EF and speculate on molecular mechanisms. Further, recent gene-based[132] and whole-genome[97] approaches have expanded the possible inferences we can draw from SNP GWAS results to include pathways, or biological systems, on which these SNPs act and to probe

tissue specific questions, like in which neurological tissues are genes influencing EF likely expressed.

There are two major issues that remain to characterize the molecular underpinnings of EF. First, we need larger samples that are well-powered to discover the molecular pathways of EF. To date, the largest GWAS of neurocognitive tasks included 1311 to 32,070 individuals (across tasks), and found only one hit, for a processing speed task[133], in the largest sample studied.

Second, we need purer measures of EF[113]. Response inhibition, working memory maintenance, updating, mental set shifting, and other EFs are measured using a single task in past GWAS. However, because EFs are control processes, each task includes a mixture of the target EF and the lower-level cognitive processes on which that EF operates[134]. These lower-level processes can contribute to individual differences in performance, leading to the "task impurity problem."[134] For example, the classic Stroop task requires visual processing of color information, word reading, and control of interference from highly dominant conflicting information (i.e., word reading). It is only the last aspect of the Stroop task that qualifies as EF. Past work has demonstrated that across multiple cognitive tasks, a Common EF (cEF) factor can be used to remove task-specific processes to solve this task impurity problem[113]. Extraction of cEF isolates the process of interest, increasing effect sizes,[134], and interpretability. Further, past research indicates that psychiatric disorders are associated with a broad array of EF tasks[123] , suggesting that cEF is the aspect of EF that relates to psychiatric outcomes (i.e., vs. more specific EF components that isolate variance unique to working memory updating or task shifting)[112,122]. Finally, past twin and family studies have shown that the cEF factor is correlated with, but distinguishable from, a general intelligence factor at the phenotypic and genetic levels, and

predicts behavior over and above intelligence[128]. Thus, it is likely that there are unique biological systems acting on cEF, apart from those shared with intelligence. Thus, biological studies may benefit from a deeper phenotypic perspective of EF.

This study is the first to examine a GWAS of cEF with a factor based on multiple cognitive tasks, and is the largest GWAS sample for any cognitive ability to date. We generate a factor score of cEF in a the UK Biobank sample of over 427,000 individuals of European ancestry based on tasks that have cognitive control components: the trail-making task, a commonly used measure of EF that has been applied in past family studies of heritability as an indicator of cEF[127,129]; the digit span task; the symbol-digit substitution task; and two tasks requiring updating and overriding memory, a pairs-matching task and a prospective memory task. We generate a score of cEF and also conduct our study separately in (two) subsamples of the UKbiobank to look for consistent effects across different densities of cognitive measurement. More about the tasks and the subsamples can be found in Table 1 and in the online methods.

We ask: (1) what specific SNPs are associated with cEF? (2) What genetic and molecular pathways are implicated by the whole-genome pattern underlying cEF? And finally, (3) Is the association between cEF and psychiatric health and wellness accounted for (in part) by shared genes?

Results

cEF Phenotypic Model Results

Factor Scores. Table 1 presents the demographic information for each task. Figure 1A presents the zero-order correlations among the cognitive measures used in the cEF factor models,

as well as the cEF factor scores. The confirmatory factor analyses used to obtain the cEF factor score is shown in Figure 1B. All tasks loaded on the cEF factor, and orthogonal task-specific factors were used to account for repeated measurement of some tasks (prospective memory, pairs memory, and digit span). We did not analyze factor scores for the specific factors, which can be considered to reflect a combination of method variance as well as variance due to processes specific to that paradigm (i.e., uncorrelated with the other tasks in the model). The fit of this model was good, $\chi^2(44)=1786.53$, $p<.001$, CFI=.980, RMSEA=.009. With this sample size (total n=490,588), a large chi-square statistic for model fit is expected, but the model fit well by other fit criteria, particularly a CFI>.95 and RMSEA<.06[135]. When additional factors were added to capture time-specific effects, they had non-significant loadings, so they were not included.

Table 1. Descriptive Statistics for Cognitive Measures Used to Obtain Factor Scores

Measure	*N*	Mean	SD	Min	Max	Skewness	Kurtosis
Trail making							
Online[a]	104,050	0.00	0.11	-0.44	0.44	0.48	0.73
Numeric[b]	104,052	1.57	0.14	1.14	2.87	0.65	0.67
Alphanumeric[b]	104,050	1.80	0.15	1.31	2.87	0.49	0.46
Symbol-digit substitution							
Online	117,785	19.76	5.11	0	40	-0.40	0.54
Prospective Memory[c]							
Initial visit	171,309	0.24	--	--	--	--	--
Repeat visit	20,314	0.15	--	--	--	--	--
Imaging visit	15,880	0.12	--	--	--	--	--
Pairs Matching[d]							
Initial visit	484,340	0.76	0.37	0.00	2.22	0.39	0.56

Repeat visit	20,085	0.70	0.34	0.00	2.06	0.33	0.55
Imaging visit	15,472	0.66	0.33	0.00	2.00	0.35	0.61
Online	114,828	0.83	0.37	0.00	2.31	0.39	0.26
Digit Span							
Initial visit	50,116	6.69	1.34	2	12	-0.32	0.84
Imaging visit	4,237	6.80	1.24	2	11	-0.20	0.68
Online	111,086	6.92	1.49	2	11	-0.38	1.09

Note. Descriptive statistics and sample information for each task loading on the common executive functioning (cEF) factor from the UKBiobank sample.
[a]Unstandardized residual of the log10-transformed alphanumeric path time after regressing out the log10-transformed numeric path time; only this score was used in the model.
[b]Log10-transformed total times in seconds to complete the numeric and alphanumeric paths; these variables were not used in the confirmatory factor analysis model but were used to obtain the residualized trails measure used in the model.
[c]Categorical variable coded as 1 for correct and 0 for incorrect on first try. The mean described proportion correct. Dashes indicate that other descriptive statistics were not calculated.
[d]Sum of the log10-transformed number of incorrect matches +1 in the 6- and 12-card rounds.

Table 2. Genetic Correlation between common EF indicators and common EF samples

	Symbol Digit	Pairs Memory	Digit Span	Prospect. Memory	Trail Making	Trails+ cEF	Trails- cEF	Full cEF
Symbol Digit	**0.1245 (0.0079)**							
Pairs Memory	0.6603 (0.0271)	**0.0713 (0.003)**						
Digit Span	0.3226 (0.0345)	0.442 (0.0263)	**0.1337 (0.0069)**					
Prospective Memory	0.4479 (0.0414)	0.5982 (0.0348)	0.4539 (0.0355)	**0.0527 (0.0039)**				

Trail Making	0.7126 (0.0322)	0.7085 (0.0317)	0.653 (0.0293)	0.5927 (0.0463)	**0.1136 (0.0084)**			
Trails+ sample cEF	0.8428 (0.0138)	0.858 (0.0207)	0.6653 (0.0214)	0.6416 (0.0365)	0.9274 (0.0133)	**0.1894 (0.0105)**		
Trails- sample cEF	0.7031 (0.0307)	0.9831 (0.0074)	0.558 (0.0259)	0.7052 (0.0308)	0.7771 (0.0381)	0.923 (0.0286)	**0.0696 (0.0038)**	
Full sample cEF	0.7683 (0.0178)	0.9527 (0.0047)	0.6164 (0.0178)	0.7046 (0.0255)	0.8452 (0.0215)	0.9629 (0.0106)	0.9892 (0.0073)	**0.0906 (0.0039)**

Note. Lower diagonal matrix representing the genetic correlation and standard error of each indicator and common executive functioning (cEF) factor scores in theTrails+, Trails-, and full samples. as estimated by LD score regression. The heritability of each measure is shown on the diagonal.

A Phenotypic Correlations

	Trails	SymDig	PM1	PM2	PM3	Pairs1	Pairs2	Pairs3	PairsO	Digit1	Digit3
Trails											
SymDig	0.34										
PM1	0.20	0.24									
PM2	0.22	0.25	0.71								
PM3	0.22	0.24	0.52	0.69							
Pairs1	0.19	0.25	0.28	0.22	0.19						
Pairs2	0.18	0.26	0.24	0.23	0.22	0.23					
Pairs3	0.19	0.28	0.19	0.21	0.19	0.23	0.22				
PairsO	0.22	0.27	0.18	0.22	0.20	0.23	0.24	0.24			
Digit1	0.25	0.19	0.29	0.31	0.21	0.19	0.16	0.05	0.14		
Digit3	0.27	0.25	0.12	0.21	0.18	0.18	0.15	0.16	0.16	0.57	
DigitO	0.24	0.23	0.16	0.21	0.19	0.13	0.13	0.13	0.15	0.43	0.50

B

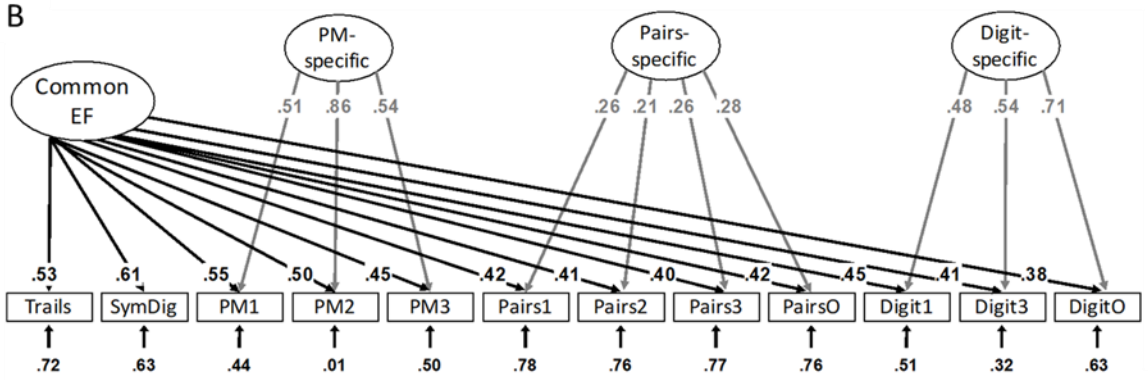


Figure 1. Justification for a common executive functioning (cEF) factor across cognitive tasks in the UK Biobank: (A) Correlations taken from Mplus; (B) Confirmatory factor analysis model used to extract factor scores. Ellipses indicate latent variables; rectangles indicate observed variables. Numbers on arrows are standardized factor loadings, and numbers at the end of arrows are residual variances. All parameters were statistically significant (p<.05). Trails= trail making (online); SymDig= symbol-digit substitution (online); PM= prospective memory; Pairs= pairs memory; Digit= digit span; IQ= intelligence; RT= reaction time. Task names with 1=first assessment; with 2=repeat assessment; with 3=imaging visit assessment; with O=online follow-up. Directionality was reversed for some variables so that for all variables, higher scores indicate better performance.

SNP associations and Annotations in Full Sample

From the confirmatory factor analysis, we obtained a factor score on cEF in the full UKbiobank sample of 427,037 individuals. We used this score to conduct a genome-wide association study in the full sample as our main analysis. To ensure consistency across different

measurement in subsamples of the UKB we also conducted GWAS in two subsamples of the UKB. First, we conducted a GWAS of the cEF factor score in the more densely measured sample of 93,024 individuals who had trails-making task and completed online battery; we chose this sample based on the trail-making task because trail-making has been used as an indicator of cEF in past genetic studies of cEF and this sample had more dense measurement, we call this the "Trails+ sample". Our second UKB sample were individuals who completed at least one cognitive task requiring direction of cognitive abilities and were part of the UKB but did not complete the trail-making task and were unrelated to people measured on the online cognitive battery (n=256,135), we call this the "Trails- sample". All genome-wide results and their annotations for this study can be accessed via http://fuma.ctglab.nl/browse.

We found 10,122 significant SNPs associated with cEF in the full sample analysis. Manhattan plots for the full sample and both subsamples are shown in Figure 2. The top SNP was a protein-coding SNP on EXOC4 that is an eQTL in cerebellar tissue. QQ plots (supplemental Figure S1) show departure from expected *p*-values under the null hypothesis for all three samples. Further, the LD-score regression intercepts were low (Full = 1.0381, Trails+ = 1.0128, Trails- = 1.0238), which implicates a highly polygenic signal. Table 3 shows the top significant lead independent SNPs from the full sample. All significant SNPs are shown in supplemental Table S1 (independent significant SNPs in Table S2, lead SNPs in Table S3, all possible candidates annotated in Table S4, and genomic loci in Table S5). Supplemental Table S6 and Figure S2 show the Circos plots for enhancer promotor associations for independent significant SNPs. Circos plots showed a long-range regulatory connection between SNPs on C17 (cytokine gene) and LRRC37A2 and SNPs in the peak of association on chromosome 17.

Supplemental Table S7 shows full annotations for SNPs discovered. We estimated the SNP-heritability of cEF score in the full sample to be .104 (se=0.002) using BOLT-LMM.

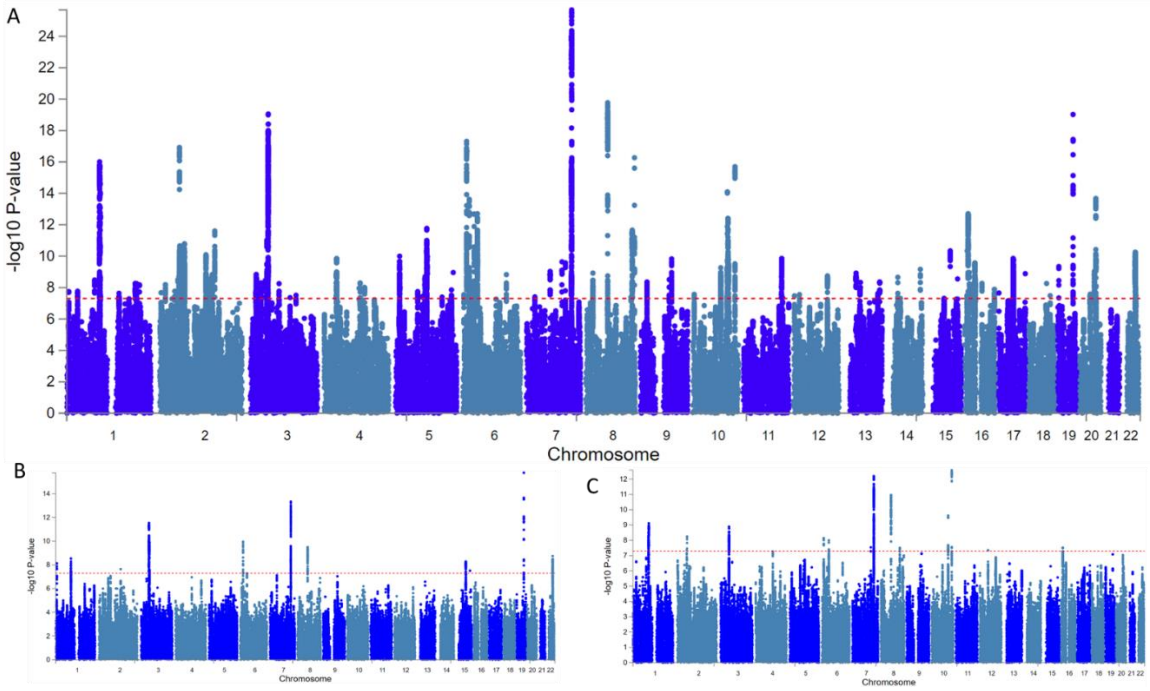


Figure 2. Manhattan plots for GWAS of common executive functioning (cEF) in the full sample (Panel A), the Trails+ sample (Panel B), and the Trails- sample (Panel C). SNPs reaching GWAS Bonferroni significance ($p<5x10^{-8}$) independently in both the Trails- and Trails+ samples were considered GWAS-significant. All models were run using Bolt-LMM to account for polygenicity and family structure.

Table 3. Top 10 Lead Independent SNPs

Alleles	Rsid	Chr	Beta	*p*-value
A:G	rs12707117	7	-0.01205	2.10E-26
C:T	rs812603	8	0.01149	1.70E-20
C:T	rs2581789	3	0.01090	8.90E-20
C:T	rs429358	19	0.01415	9.50E-20

C:T	rs36120363	6	0.01000	7.20E-18
C:G	rs7582485	2	0.00989	1.20E-17
A:G	rs13262595	8	-0.00950	5.40E-17
C:T	rs9659182	1	0.00963	9.80E-17
G:T	rs2280141	10	-0.00928	2.00E-16
C:G	rs978346	1	0.00901	2.30E-15

Note. Of the identified significant SNPs, those independent at $r^2 < 0.1$ were defined as lead SNPs. The top 10 lead SNPs by *p*-value of association are shown with their alleles (minor:major), registered SNP ids (rsid), chromosomes (Chr), betas, and *p*-values.

Comparison of Trails+ and Trails- Sample

The LD score heritability for the Trails+ sample was higher (h^2=.19, se=0.0136) than for the Trails- sample ($h^{2=}$.07, se=0.0039) and the Trails+ sample showed more significant SNPs (Trails+= 1120, Trails-= 713), despite a smaller sample size. This difference suggests that adequate measurement of phenotypes should remain important in SNP discovery. However, both samples showed less enrichment for polygenic signal than the full sample.

Despite these differences, both samples are likely measuring the same construct. Trails+ and Trails- assessments showed a high genetic correlation: *r*G=0.9181 (se=0.0288). We found 394 SNPs that were consistently independently associated at genome wide significance ($p<5 \times 10^{-8}$) in both the Trails+ and Trails- sample (shown in supplemental Table S1). There were consistent peaks of association on chromosomes 1, 3, 6, and 8. Supplemental Tables S8-S13 show discovery and significant findings in the Trails+ sample, and supplemental Tables S14-S19 show findings in the Trails- sample.

Gene-Based Analysis

Gene-Wise Analysis. To speculate individual genes that are associated with cEF, we ran a gene-wise test of association by using combinations of all the SNPs in each gene to see if they related to cEF, using the FUMA/MAGMA[132] pipeline[136]. The gene-based analysis in the full cEF sample found 319 genes significantly associated with cEF, after multiple corrections (Bonferroni $p = 0.05/18597 = 2.689e\text{-}6$). 22 genes were consistent across both subsamples, with the strongest association being EXOC4 (supplemental Tables 20-22 for genes in each sample, and Manhattan plots in supplemental Figure S3). QQ plots for the *p*-values in the gene-based test (supplemental Figure S2) show that *p*-values differed from expected.

Gene-Set Analyses. To discover molecular pathways we ran a gene-set analysis of pathways from Msigdb v5.2[137] for "Curated gene sets" and "GO terms" using FUMA/MAGMA. Essentially, the annotation category for combined SNPs in a biological category is tested for its association with cEF. Gene-based regression analysis showed significant association between larger SNP effects on cEF and potassium channel activity, neuronal pathways; and GABA-A receptor activity. Table 4 shows the significantly associated gene sets after Bonferroni correction.

We also annotated gene sets based on cell-type specific RNA datasets in the human cortex, hippocampus and frontal cortex. Supplemental Figure S4 shows the results of this analysis. Post Bonferroni correction, GABA2 cells were significant in the hippocampus, GABAergic neurons in the prefrontal cortex (though this was specific to 26 weeks of gestation) and hybrid and neurons cells in the whole human cortex (across age).

Table 4. Significantly Associated GO Categories from MAGMA Gene-Set Analysis

Gene-set	*N* Genes	Beta	SE	*P* Corrected
GO: Synaptic membrane	243	0.351	0.071	0.004
GO: Synapse part	576	0.216	0.044	0.006
CGS: Gaba a receptor activation	10	1.940	0.401	0.007
CGS: neuronal system	262	0.325	0.068	0.008
GO: Voltage gated potassium channel activity	85	0.604	0.126	0.008
CSG: Potassium Channels	96	0.560	0.118	0.012
GO: regulation of synapse structure or activity	221	0.319	0.069	0.023
GO: Postsynapse	354	0.264	0.058	0.024
GO: Synapse	713	0.185	0.040	0.026
GO: Gaba receptor complex	14	1.530	0.340	0.037
CGS: Voltage gated potassium channels	42	0.844	0.190	0.047
GO: regulation of synaptic plasticity	135	0.385	0.087	0.048

Note. Signal GO term enrichment for SNPs influencing common executive functioning (cEF). GO terms were found through MAGMA, using a gene-level regression analysis accounting for gene-size and population structure. We present the number of genes in each term category, the beta and standard error from the gene level regression, and the Bonferroni corrected *p*-value for each category. GO = Recovered from MBSig Gene Ontology, CGS = recovered from MBsig Curated Gene Sets.

Gene-Property Analysis. To ascertain in which tissues our SNP results were likely influencing gene expression, we used MAGMA to conduct gene-property analysis by tissues by

30 broad human tissues and 53 specific human tissues in the GTeX sample[138] (supplemental Figure S4). After Bonferroni correction, broad tissues implicated were the brain and pituitary. All specific brain tissues were associated except the substantia nigra and spinal chord-c1, post Bonferroni correction.

Transcriptional Profiling. To examine the transcriptional profile across the implicated brain tissues (all GTeX tissues excluding the substantia nigra and the spinal chord c-1), we used PrediXcan[139] to predict transcription patterns from SNP data and tissue specific eQTL associations from the GTeX sample[138]. We found 441 brain tissue specific transcripts associated with cEF, post Bonferroni correction (supplemental Table S25). We then entered this transcriptional profile in the connectivity Map (cMAP)[140]. After filtering for transcripts found in multiple tissues, 78 were also associated with transcriptional changes after exposure to perturbagens in the cMAP. The top 15 substances shown to mimic the predicated transcriptional profile of cEF are shown in supplemental Table S23. Of note, three have previous psychiatric applications. Nicergoline, an anti-dementia drug that is shown to be effective in a broad array of behavioral and cognitive disorders in old age, nortriptyline, an earlier type of tricylic anti-depressant, and chlorpromazine, a typical anti-psychotic that is often prescribed to treat severe cases of schizophrenia, bipolar, OCD, and depression.

Genetic Correlations. We used LD Score regression to estimate the genetic correlation between the summary statistics from cEF and summary statistics of other major behavioral and brain GWAS studies, as many of these traits have been associated with EF phenotypically and/or genetically. We attempted to replicate findings that cEF genetically related to common psychiatric dysfunction like ADHD and depression; evaluated whether cEF related genetically to more severe psychiatric dysfunction, like schizophrenia and bipolar disorder; probed the degree

of association with behavioral and personality traits; and estimated the genetic overlap between cEF and life events, like age at first birth and educational attainment.

We first examined the genetic correlation of cEF with intelligence, given prior literature suggesting a close relationship. The correlation based on a meta-analysis of over 78,000 individuals by Sniekers et al. 2017[141] was estimated at .71 (se=.0215, p <.001). This correlation is about half-way between the genetic correlation estimated in prior twin studies (report values from Friedman et al., 2008; TX study, and Vetsa studies), and the confidence interval confirms that cEF, though related to intelligence, is genetically separable. Moreover, we found additional evidence for separability, as cEF was more associated with bipolar disorder and schizophrenia (beyond the 95% confidence interval) than intelligence in past studies[102]. As we have forthcoming work that examines the associations between intelligence and cEF in more depth, we defer further presentation of comparisons to that study.

All nominally significant genetic correlations and their effect sizes and confidence intervals are shown in Figure 2 (all correlations run shown in Supplemental Table S24). Post Bonferroni correction, cEF was significantly associated with cognitive phenotypes, including years of schooling (r=.34, se=.020, p< .001), college completion (r=.31, se=.0322, p< .001 and childhood IQ (r=.51, se=.060, p< .001). Almost all psychiatric disorders, including schizophrenia (r=-.35, se=.022, p< .001), PGC cross-disorder diagnosis (r=-.36, se=.029, p< .001), Alcohol dependence (Claire is rerunning) and alcohol use frequency (r=-.11, se=.024, p< .001), bipolar disorder (r=-.32, se=.034, p< .001), depressive symptoms (r=-.24, se=.035, p< .001), Major Depressive Disorder (r=-.27, se=.046, p< .001), Alzheimer's (r=-.34, se=.088, p< .001) and ADHD (r=-.40, se=.010, p< .001) (associations with Anorexia Nervosa were nominally significant). The only personality facet cEF was associated with was neuroticism (r=-

.26, se=.061, p< .001). Finally, cEF related to age of first birth (r=.1801, se=.029, p< .001), and parents' age at death (r=.22, se=.061, p< .001). Comparisons of effect sizes and standard errors suggests that cEF related more strongly to dimensions of impairment, like, psychiatric disorders broadly vs. neuroticism and alcohol dependence vs. alcohol use.

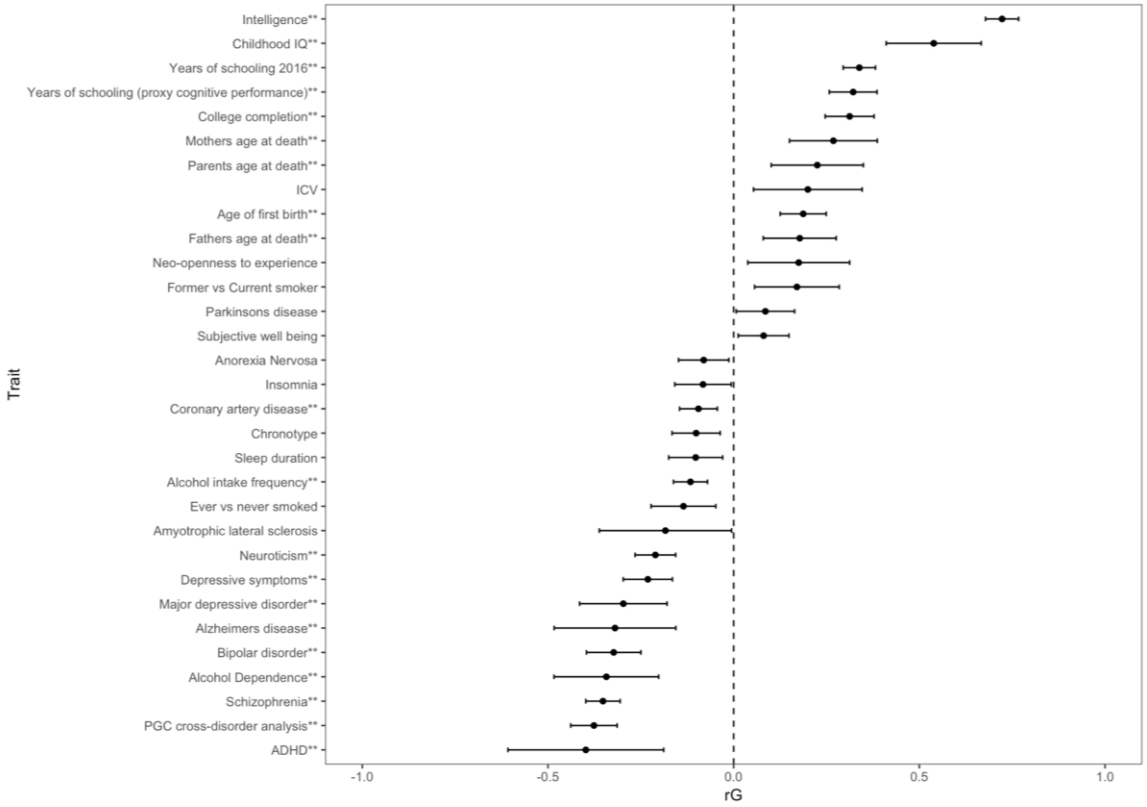


Figure 3. Genetic correlations between common executive functioning (cEF) and psychiatric, behavioral and health traits using LD score regression. Bars indicate 95% confidence intervals. All GWAS summary statistics were filtered for SNPs with imputation quality above .90 and minor allele frequency above .01. All results significant at nominal significance $p < .05$. ** represent significance that passed Bonferroni correction.

Discussion

This study uncovered likely molecular mechanisms influencing cEF. We found genome-wide significant, replicable signals on chromosomes 1, 3, 6, and 8, with the strongest effect for a

protein-coding variant on the EXOC4 gene that is an eQTL in the cerebellum. We found over 300 genes and transcripts associated with cEF from various informatic follow-ups, demonstrating an extremely high polygenic signal. Gene-set and gene-property analyses converged on neuronal cells, potassium channels, and GABA A receptor activity as likely biological pathways. Molecular-cellular analysis implicated GABA neuronal signaling in the cortex broadly, DLPFC, and hippocampus, and gene-property analysis tested across the whole-body indicated enriched patterns of expression in almost all brain tissues. We replicated the significant heritability of cEF[128], though this estimate was smaller than twin and family studies. We also replicated the coheritability of cEF with common psychiatric disorders and personality, and found novel genetic associations with some life outcomes and more severe psychiatric disorders. Interestingly, we found that cEF was nominally (and sometimes beyond the 95% CI) more genetically related to more impairing psychiatric conditions. Finally, expanding on these results, we found that Nicergoline, Nortriptyline, and chlorpromazine induce cellular changes similar to transcriptional profile for better cEF. We discuss the biological implications of these findings below.

One of the main strengths of this study has been the deeper phenotypic perspective gained by using a dense measurement of EF and generating a harmonized score across subsamples. Two perspectives currently exist in the genome-wide literature in relation to power for behavioral phenotypes. One advocates increasing sample size by combining across overlapping measures; the other advocates refining phenotyping to discover more meaningful signal[142]. This study (through the UKB) offered a path forward in the genome-wide literature that succeeded in incorporated both perspectives. In this study, the largest cohort, the full sample, did yield the most genetic associations. However, in comparison to the more sparsely

measured Trails- sample, there was much higher heritability (more than 2X) and more genome-wide significant variants in the more densely measured Trails+ sample. Because the genetic correlation is > .9 between Trails+ and Trails- samples, it is likely that the dense measurement sample simply had more power due to better phenotypic measurement. Finally, the full sample leveraged both sample size from sparse measurement sample and phenotyping from the smaller sample by estimating one model across the full UKBiobank.

We replicated genetic association of cEF with depression[143] and ADHD[131]. We also found novel genetic associations with schizophrenia, bipolar disorder, alcohol dependence, Alzheimer's disease, educational attainment, age of first birth, and parents age of death. In line with past literature that suggests that cEF is a broad risk factor for psychopathology[112,122], cEF related to cross-disorder variance and each psychiatric disorder examined (besides autism and anorexia nervosa) with a similar effect size. Further, we found evidence that cEF related to later parents age of death and earlier age of first birth. These associations with cEF and psychiatric, life outcomes, and later dementia follow a similar pattern as a fast life-strategy[144], a perspective in evolutionary theory. Individuals may express phenotypes, like impulsive behaviors, because earlier in life they lead to better fecundity (i.e. having more children, obtaining and using resources more quickly), but these early advantageous dimensions may convey less survivability in later life. In the population, there may be a balance of fast and slow life-strategies that are reflected in individual's behavior and life events, and, in context of this study, grounded in cognitive abilities.

While there is substantial and significant overlap between our cEF factor and studies of intelligence, there is some separability based on the genetic correlation and this separability is reflected in some differential correlations with outcomes of interest. cEF genetically correlates

more strongly (beyond the 95% CI) with schizophrenia and bipolar disorder than intelligence[102,141]. Also, past genome-wide association studies of alcohol consumption behaviors have suggested positive or no genetic association between intelligence and alcohol use and abuse[145], while cEF is significantly genetically negatively associated with alcohol use and abuse. Further, cEF's genetic associations to psychiatric and behavioral traits seem to follow a spectrum of impairment, and intelligence does not follow the same pattern. Additionally, intelligence is more genetically correlated with educational attainment, head size, autism, openness to experience, and smoking behaviors then cEF[102,141]. Finally, part of the genetic correlation between cEF and intelligence could be artificially inflated, as previous meta-analyses of intelligence (where these genetic correlations were estimated and summary statistics were recovered) include cohorts that are almost entirely EF tasks[102,141]. More deep phenotyping work is needed to establish the relationship between IQ and cEF in large population studies.

The molecular pathways of cEF are almost entirely neurological and notably spread across the brain. In addition to the frontal-parietal brain regions classically associated with EF, this work is consistent with imaging studies that have suggested associations between cEF and lower order brain areas, in particular, the cerebellum[31]. In this study, the strongest signal came from a protein coding variant on EXOC4, which influences exocytic vesicles docking to the plasma membrane. This variant is an eQTL in the cerebellum, and we found significant eQTL enrichment (across the genome) for the cerebellum and a number of other non-cortical brain regions. Additionally, cell-type specific analysis implicated GABA activity in both the cortex and hippocampus. Taken together, these results suggest that expression across the brain may be important for cEF individual differences, and inhibitory neurotransmitters in particular.

Finally, Nicergoline, Nortriptyline, and chlorpromazine were drugs that induce a transcriptional profile similar to higher cEF and are known to cross the blood-brain barrier. Interestingly, all three drugs have been used to treat broad array of psychiatric conditions, are older classes of psycho-pharmaceuticals with less specific drug targets[146], and treat disorders related to EF. As cEF is highly polygenic with an extremely complex molecular profile, it is possible that less specific drug targets are needed to improve cognitive performance. More work should be done to see how these drugs may influence cognitive abilities, particularly in clinical psychiatric populations, and how to reduce harmful side-effects of older classes of psych-pharmaceuticals.

Conclusion

Genetics of cEF represents a vulnerability to neurological dysfunction broadly. cEF is heritable and highly polygenic, but neuronal and GABAergic pathways were detectable from a whole genome screening. We establish here a molecular profile of neurocognitive ability that may serve to understand the neuro-molecular underpinnings of individual differences in cognitive control across the population and impairment in brain disorders.

Chapter 5: Conclusions and Discussion

This work set out to generate a framework for whole-genome whole-brain (WGWB) analysis in the post-genome-wide association study (GWAS) era. We conducted three studies to demonstrate possible avenues: (Study 1) a brain mapping study of genetic effects on depression and a correlated dimension that implicated neurological areas that share genetic variance with traits of interest; (Study 2) a demonstration of power gained from whole-brain predictive endophenotypes over individual brain measures; and (Study 3) a GWAS of common executive functioning (cEF) with bioinformatic techniques that leveraged patterns across the genome to discover neurological patterns. Below we discuss the results and philosophical grounding within the WBWG approach. Specifically, we ask (1) How did each study use a whole-system biological approach to make biological discoveries of psychological states? (2) Did each study implicate genetic pathways of psychological states that can be contextualized in the brain?

Study 1: A Demonstration of Whole-Brain Mapping with Whole-Genome Effects

In study 1, we estimated classic parametric brain maps across the human cortex at the genetic and environmental level for depression and a related dimension and looked at their overlap. We concluded by using the overlap maps to determine likely areas of future study.

While this work clearly expands brain imaging to create maps based on genetic and environmental effects, it also expands the types of inference we make from twin studies. The bivariate genetic model is one of the classic designs in behavior genetics, and this model which explores overlapping genetic variance between two traits. In study 1, we attempted to contextualize the heritability and coheritability of depression and a related dimension by using these overlay brain maps. To expand on this analysis, we also used other neuroinformatic tools

in imaging research. For example, we used a naïve Bayes estimator to see if prior fMRI studies have found effects in those areas and what phenotypes those effects were related to. Finally, we looked at overlap in spatial transcription patterns within our implicated brain areas.

We found that depression and unemotional behavior mapped most reliably on the dorsal lateral prefrontal cortex, but found several other regions were also nominally associated across different brain systems. In terms of likely functions, the spatial coordinates overlapped with executive and default-mode processes most closely. Finally, the only consistent pattern across brain transcription associations implicated calcium channels, a key neural firing process. As the effects overlapped with fMRI results in a theoretically meaningful way and implicated calcium channels, it is likely that cross-modality effect may be responsible; i.e. anatomy and function may be important for this dimension and could be a key area of future study.

By combining the whole-brain map with the classic twin design (a whole-genome approach), this paper is in line with the purpose of this thesis and qualifies as a WBWG biomarker approach, as it both combines a popular whole-brain approach (cortical mapping) with a popular whole-genome approach (twins) and contextualizes the findings using MRI and fMRI tools. WBWG biomarkers were also demonstrated in the follow-up, as we focused on patterns across transcription to find pathways rather than interpreting effects of single transcripts in these regions. Finally, we relate these comparisons to the modern RDoC matrix to qualify the translational nature of the approach.

Study 2: Whole-Brain Pattern Analysis as a Phenotype for Whole-Genome Studies

The purpose of study 2 was to expand the types of endophenotypes we use in gene-finding studies. While single brain regions have been the norm for endophenotypic research thus

far, we showed here how whole-brain predictive models also work as endophenotypes. We demonstrate this principle across three popular machine learning procedures and draw conclusions about genetic and phenotypic association from these models.

Specifically, we found that whole-brain models of intelligence demonstrated higher heritability than single brain areas, and much larger genetic correlations with measured IQ. Further, the LASSO regression model implicated neurological pathways that were known to be related to intelligence in the literature and discovered some additional brain areas that have not been associated previously.

In line with the philosophy of our approach, these endophenotypes are more powerful measures of genetic overlap (estimated by all measured variants) than single brain areas. Further, these approaches have nice properties for brain-behavior discovery as the whole-brain LASSO was highly interpretable and offered a more powerful omnibus test for discovery of associated brain areas.

Study 3: A GWAS of Common Executive Functioning

It is often stated that an ultimate goal of a GWAS is to find new candidate genes for follow-up; however, single variants are not the only way forward with this data source. Other groups have developed several regression-based methods that take effect of SNPs across the genome. While we did not develop the methods, these approaches fit with the context of the WBWG biomarkers philosophy we set out, and thus we demonstrate their utilization in a genome-wide association study of common executive functioning (cEF). Study 3 uses whole-genome (mostly exome and intergenic) patterns to search for biological pathways underlying common EF. Specifically we utilized regression based methods in MAGMA and PrediXcan to

that uses the combined effects of many SNPs to implicate transcriptional, molecular and cellular pathways. In both these methods, SNPs are weighted not based off their p-value significance, but their effect size in the GWAS and their annotation for belonging to a biological system (or associating to a particular transcript). Thus, for these techniques, it is whether the combined SNP effects across a system show a stronger pattern of association (than would be expected by random SNPs) that determines whether a significant discovery is made.

In using these techniques to estimate whole-genome patterns, we created a biological profile of cEF that implicates GABAergic, neuronal and potassium channels and shows (some) possible treatment pathways within known psychopharmaceuticals. As an example to put this approach in context, cEF is so highly polygenic that any one GABA single nucleotide polymorphism is going to show small effects that are similar patterns to several non-coding SNPs. However, the combination of GABA SNPs is more likely to influence cEF, compared to almost any other biological pathway. This increased likelihood is easily detected when we look at the patterns of all SNP effects, not only the significant ones. Finally, the regression-based approaches implicated transcription patterns across the whole-brain in cEF, mirroring some recent results from imaging studies. Thus, using the whole genome we can also discovery meaningful biological pathways, and the findings mirror results from past whole-brain imaging studies.

Future Directions

The methods here are not exhaustive and there are several directions I would like to take the WBWG biomarker approach in the future. First, to follow up on the ideas of study 1, I would like to compare polygenic score maps to twin maps. These are both ways of conducting whole-brain maps of genetic effects and their overlap and differences would be interesting for

determining the degree of overlap between twins and statistical genetics data. For example, do polygenic scores of depression find the same regions as the twin map? If there are differences, what behaviors are we detecting better with each method (and which worse)? Essentially, can the brain tell us whether twins or polygenic scores capture particular aspects of depression better?

To follow up on study 2, I would like to see if training procedures can be improved to increase the heritability and coheritabiliy of whole-brain endophenotypes. I have played with training machine learning models across the twins, essentially using twin 1's brain measures as features to predict twin 2's IQ (cross-twin training procedures). Another approach is to simply to use only highly heritable components in training the machine learning model. Initial results show increased heritability for whole-brain models made from either of these approaches, and this increased heritability could increase the utility of whole-brain endophenotypes.

To follow up on study 3, I would like to estimate polygenic scores of cEF and map them across different brain modalities. We also have several EF tasks in the scanner and genotypes on most of those individuals, so I could simply add the polygenic score for cEF into those brain maps through a covariate analysis and see where they EF predict task activation, this may be a good way at getting at what aspect of each task is capturing common EF genetic influences on that task. Additionally, we are comparing the cEF GWAS to a GWAS of intelligence to demonstrate that their uniqueness from one another (i.e., their $r_G < 1$) is meaningful for psychological traits of interest.

Conclusion

By utilizing patterns across the whole system, we have powerful tools for biological discovery in human studies. These whole-system patterns may be the key to expanding inference and clinical application of biologically meaningful data. Through the work here, I both developed innovations for WBWG biomarkers and utilized existing tools to demonstrate this ability for discovery. Hopefully, the WBWG biomarker paints a path forward for discoveries critical in our understanding of our biology and ourselves.

Reference Cited

1. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *Am. J. Psychiatry* **167,** 748–751 (2010).

2. Poldrack, R. A. The role of fMRI in Cognitive Neuroscience: where do we stand? *Curr. Opin. Neurobiol.* **18,** 223–227 (2008).

3. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *J. Neurosci.* **36,** 6553–6562 (2016).

4. Testing Hypotheses About Direction of Causation Using Cross-Sectional Family Data. Available at: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/44109/10519_2005_Article_BF01067552.pdf?sequence=1&isAllowed=y. (Accessed: 10th August 2015)

5. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).

6. Reggente, N. *et al.* Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive-compulsive disorder. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 2222–2227 (2018).

7. Elliott, L. *et al.* The genetic basis of human brain structure and function: 1,262 genome-wide associations found from 3,144 GWAS of multimodal brain imaging phenotypes from 9,707 UK Biobank participants. *bioRxiv* 178806 (2017). doi:10.1101/178806

8. Wager, T. D. *et al.* An fMRI-Based Neurologic Signature of Physical Pain. *N. Engl. J. Med.* **368,** 1388–1397 (2013).

9. Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19,** 165–71 (2016).

10. McMahon, F. J. Prediction of treatment outcomes in psychiatry--where do we stand ? *Dialogues Clin. Neurosci.* **16,** 455–64 (2014).

11. MENDLEWICZ, J., FIEVE, R. R. & STALLONE, F. Relationship Between the Effectiveness of Lithium Therapy and Family History. *Am. J. Psychiatry* **130,** 1011–1013 (1973).

12. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47,** 856–860 (2015).

13. So, H.-C. *et al.* Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat. Neurosci.* **20,** 1342–1349 (2017).

14. Duncan, L. E. & Keller, M. C. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *Am. J. Psychiatry* **168,** 1041–1049 (2011).

15. Culverhouse, R. C. *et al.* Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of

depression. *Mol. Psychiatry* **23,** 133–142 (2018).

16. Hamer, D. Rethinking Behavior Genetics. *Science (80-.).* **298,** 71–72 (2002).

17. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

18. Johnson, E. C. *et al.* No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. *Biol. Psychiatry* **82,** 702–708 (2017).

19. Chabris, C. F. *et al.* Most Reported Genetic Associations With General Intelligence Are Probably False Positives. *Psychol. Sci.* **23,** 1314–1323 (2012).

20. van Erp, T. G. M. *et al.* Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* **21,** 547–553 (2016).

21. Reddan, M. C., Lindquist, M. A. & Wager, T. D. Effect Size Estimation in Neuroimaging. *JAMA Psychiatry* **74,** 207 (2017).

22. Woo, C.-W. & Wager, T. D. Neuroimaging-based biomarker discovery and validation. *Pain* **156,** 1379–81 (2015).

23. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.* **16,** 55–61 (2015).

24. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *bioRxiv* 399402 (2018). doi:10.1101/399402

25. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat. Neurosci.* **19,** 420–431 (2016).

26. Seidman, L. J. *et al.* Left Hippocampal Volume as a Vulnerability Indicator for Schizophrenia. *Arch. Gen. Psychiatry* **59,** 839 (2002).

27. Tepest, R., Wang, L., Miller, M. I., Falkai, P. & Csernansky, J. G. Hippocampal deformities in the unaffected siblings of schizophrenia subjects. *Biol. Psychiatry* **54,** 1234–1240 (2003).

28. Witthaus, H. *et al.* Hippocampal subdivision and amygdalar volumes in patients in an at-risk mental state for schizophrenia. *J. Psychiatry Neurosci.* **35,** 33–40 (2010).

29. Godefroy, O. Frontal syndrome and disorders of executive functions. *J. Neurol.* **250,** 1–6 (2003).

30. Smolker, H. R., Friedman, N. P., Hewitt, J. K. & Banich, M. T. Neuroanatomical Correlates of the Unity and Diversity Model of Executive Function in Young Adults. *Front. Hum. Neurosci.* **12,** 283 (2018).

31. Reineberg, A. E. & Banich, M. T. Functional connectivity at rest is sensitive to individual differences in executive function: A network analysis. *Hum. Brain Mapp.* **37,** 2959–75 (2016).

32. Grefkes, C. & Fink, G. R. Connectivity-based approaches in stroke and recovery of function. *Lancet Neurol.* **13,** 206–216 (2014).

33. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1405–9 (2002).

34. Bird, T. D. Genetic aspects of Alzheimer disease. *Genet. Med.* **10,** 231–9 (2008).

35. Eyler, L. T. *et al.* A comparison of heritability maps of cortical surface area and thickness and the influence of adjustment for whole brain measures: a magnetic resonance imaging twin study. *Twin Res. Hum. Genet.* **15,** 304–14 (2012).

36. Rimol, L. M. *et al.* Cortical Thickness Is Influenced by Regionally Specific Genetic Factors. *Biol. Psychiatry* **67,** 493–499 (2010).

37. Coskunpinar, A., Dir, A. L. & Cyders, M. A. Multidimensionality in impulsivity and alcohol use: a meta-analysis using the UPPS model of impulsivity. *Alcohol. Clin. Exp. Res.* **37,** 1441–50 (2013).

38. Chen, C.-H. *et al.* Hierarchical Genetic Organization of Human Cortical Surface Area. *Science (80-.).* **335,** 1634–1636 (2012).

39. Couvy-Duchesne, B. *et al.* Lingual Gyrus Surface Area Is Associated with Anxiety-Depression Severity in Young Adults: A Genetic Clustering Approach. *eNeuro* **5,** (2018).

40. Schmitt, J. E. *et al.* The Dynamic Associations Between Cortical Thickness and General Intelligence are Genetically Mediated. *Cereb. Cortex* (2019). doi:10.1093/cercor/bhz007

41. Whalley, H. C. *et al.* The influence of polygenic risk for bipolar disorder on neural activation assessed using fMRI. *Transl. Psychiatry* **2,** e130–e130 (2012).

42. 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 (2018).

43. Fekadu, A. *et al.* What happens to patients with treatment-resistant depression? A systematic review of medium to long term outcome studies. *J. Affect. Disord.* **116,** 4–11 (2009).

44. Schreiter, S., Pijnenborg, G. H. M. & aan het Rot, M. Empathy in adults with clinical or subclinical depressive symptoms. *J. Affect. Disord.* **150,** 1–16 (2013).

45. Wolkenstein, L., Schönenberg, M., Schirm, E. & Hautzinger, M. I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *J. Affect. Disord.* **132,** 104–111 (2011).

46. Cusi, A. M., MacQueen, G. M., Spreng, R. N. & McKinnon, M. C. Altered empathic responding in major depressive disorder: Relation to symptom severity, illness burden, and psychosocial outcome. *Psychiatry Res.* **188,** 231–236 (2011).

47. Ladegaard, N., Larsen, E. R., Videbech, P. & Lysaker, P. H. Higher-order social cognition in first-episode major depression. *Psychiatry Res.* **216,** 37–43 (2014).

48. Centifanti, L. C. M., Meins, E. & Fernyhough, C. Callous-unemotional traits and

impulsivity: distinct longitudinal relations with mind-mindedness and understanding of others. *J. Child Psychol. Psychiatry* **57,** 84–92 (2016).

49. Frick, P. J. & Ellis, M. Callous-Unemotional Traits and Subtypes of Conduct Disorder. *Clin. Child Fam. Psychol. Rev.* **2,** 149–168 (1999).

50. Kimonis, E. R. *et al.* Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. *Int. J. Law Psychiatry* **31,** 241–252 (2008).

51. Byrd, A. L., Kahn, R. E. & Pardini, D. A. A Validation of the Inventory of Callous-Unemotional Traits in a Community Sample of Young Adult Males. *J. Psychopathol. Behav. Assess.* **35,** (2013).

52. Lamers, F. *et al.* Identifying Depressive Subtypes in a Large Cohort Study. *J. Clin. Psychiatry* **71,** 1582–1589 (2010).

53. Carney, D. M. *et al.* The Twin Study of Negative Valence Emotional Constructs. *Twin Res. Hum. Genet.* **19,** 456–64 (2016).

54. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *Am. J. Psychiatry* **157,** 1552–1562 (2000).

55. Levinson, D. F. *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biol. Psychiatry* **76,** 510–2 (2014).

56. Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* **41,** D996–D1008 (2013).

57. Schmaal, L. *et al.* Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22,** 900–909 (2017).

58. Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R. & Buckner, R. L. Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron* **65,** 550–562 (2010).

59. The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *Neuroimage* **99,** 180–190 (2014).

60. Fischl, B., Sereno, M. I., Tootell, R. B. H. & Dale, A. M. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8,** 272–284 (1999).

61. Bora, E., Fornito, A., Pantelis, C. & Yücel, M. Gray matter abnormalities in Major Depressive Disorder: a meta-analysis of voxel based morphometry studies. *J. Affect. Disord.* **138,** 9–18 (2012).

62. De Brito, S. A. *et al.* Size matters: Increased grey matter in boys with conduct problems and callous–unemotional traits. *Brain* **132,** 843–852 (2009).

63. Wagner, G. *et al.* Structural brain alterations in patients with major depressive disorder and high risk for suicide: Evidence for a distinct neurobiological entity? *Neuroimage* **54,**

1607–1614 (2011).

64. Amunts, K. *et al.* Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* **412,** 319–41 (1999).

65. Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. *Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R).* (Lawrence Erlbaum Associates, 2004).

66. Friedman, N. P., du Pont, A., Corley, R. P. & Hewitt, J. K. Longitudinal Relations Between Depressive Symptoms and Executive Functions From Adolescence to Early Adulthood: A Twin Study. *Clin. Psychol. Sci.* **6,** 543–560 (2018).

67. Byrd, A. L., Kahn, R. E. & Pardini, D. A. A Validation of the Inventory of Callous-Unemotional Traits in a Community Sample of Young Adult Males. *J. Psychopathol. Behav. Assess.* **35,** (2013).

68. Hagler, D. J., Saygin, A. P., Sereno, M. I. & Sereno, M. I. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* **33,** 1093–103 (2006).

69. Neale, M. C. & Cardon, L. R. *Methodology for Genetic Studies of Twins and Families*. (Springer Netherlands, 1992). doi:10.1007/978-94-015-8018-2

70. Fischl, B. FreeSurfer. *Neuroimage* **62,** 774–781 (2012).

71. Friston, K. J., Penny, W. D. & Glaser, D. E. Conjunction revisited. *Neuroimage* **25,** 661–667 (2005).

72. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8,** 665–670 (2011).

73. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106,** 1125–65 (2011).

74. Fox, A. S., Chang, L. J., Gorgolewski, K. J. & Yarkoni, T. Bridging psychology and genetics using large-scale spatial analysis of neuroimaging and neurogenetic data. *doi.org* 012310 (2014). doi:10.1101/012310

75. Crum, W. R. Magnetic Resonance Brain Image Processing and Arithmetic with FSL. in 109–126 (Humana Press, 2011). doi:10.1007/978-1-61737-992-5_5

76. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33,** D428–D432 (2004).

77. Heath, A. C. *et al.* Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet.* **23,** 29–50 (1993).

78. Kaiser, R. H. *et al.* Large-Scale Network Dysfunction in Major Depressive Disorder. *JAMA Psychiatry* **72,** 603 (2015).

79. Szanto, K. *et al.* Social Emotion Recognition, Social Functioning, and Attempted Suicide

in Late-Life Depression. *Am. J. Geriatr. Psychiatry* **20,** 257–265 (2012).

80. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain* **137,** 12–32 (2014).

81. Pizzagalli, D. A. Frontocingulate Dysfunction in Depression: Toward Biomarkers of Treatment Response. *Neuropsychopharmacol. Publ. online 22 Sept. 2010; | doi10.1038/npp.2010.166* **36,** 183 (2010).

82. Goldapple, K. *et al.* Modulation of Cortical-Limbic Pathways in Major Depression. *Arch. Gen. Psychiatry* **61,** 34 (2004).

83. Harley, R., Sprich, S., Safren, S., Jacobo, M. & Fava, M. Adaptation of Dialectical Behavior Therapy Skills Training Group for Treatment-Resistant Depression. *J. Nerv. Ment. Dis.* **196,** 136–143 (2008).

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

85. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

86. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20,** 365–377 (2017).

87. Cai, N., Kendler, K. & Flint, J. Minimal phenotyping yields GWAS hits of low specificity for major depression. *bioRxiv* 440735 (2018). doi:10.1101/440735

88. Gottesman, I. I. & Gould, T. D. The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *Am. J. Psychiatry* **160,** 636–645 (2003).

89. Turetsky, B. I. *et al.* Neurophysiological Endophenotypes of Schizophrenia: The Viability of Selected Candidate Measures. *Schizophr. Bull.* **33,** 69–94 (2006).

90. Hall, M.-H. & Smoller, J. W. A New Role for Endophenotypes in the GWAS Era: Functional Characterization of Risk Variants. *Harv. Rev. Psychiatry* **18,** 67–74 (2010).

91. Woo, C.-W., Krishnan, A. & Wager, T. D. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *Neuroimage* **91,** 412–9 (2014).

92. Arbabshirani, M. R., Plis, S., Sui, J. & Calhoun, V. D. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* **145,** 137–165 (2017).

93. Shen, X. *et al.* Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* **12,** 506–518 (2017).

94. Liu, J., Liao, X., Xia, M. & He, Y. Chronnectome fingerprinting: Identifying individuals and predicting higher cognitive functions using dynamic brain connectivity patterns. *Hum. Brain Mapp.* **39,** 902–915 (2018).

95. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19,** 1523–1536 (2016).

96. Hall, M.-H. & Smoller, J. W. A new role for endophenotypes in the GWAS era: functional

characterization of risk variants. *Harv. Rev. Psychiatry* **18,** 67–74 (2010).

97. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).

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

99. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33,** 1–22 (2010).

100. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).

101. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33,** 272–279 (2017).

102. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50,** 912–919 (2018).

103. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).

104. Taal, H. R. *et al.* Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat. Genet.* **44,** 532–538 (2012).

105. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).

106. Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48,** 1462–1472 (2016).

107. Song, M. *et al.* Brain spontaneous functional connectivity and intelligence. *Neuroimage* **41,** 1168–1176 (2008).

108. Basten, U., Stelzel, C. & Fiebach, C. J. Intelligence is differentially related to neural effort in the task-positive and the task-negative brain network. *Intelligence* **41,** 517–528 (2013).

109. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet* **9,** e1003348 (2013).

110. López-Solà, M. *et al.* Towards a neurophysiological signature for fibromyalgia. *Pain* **158,** 34–47 (2017).

111. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet.* **9,** e1003348 (2013).

112. Hatoum, A. S., Rhee, S. H., Corley, R. P., Hewitt, J. K. & Friedman, N. P. Do executive functions explain the covariance between internalizing and externalizing behaviors? *Dev. Psychopathol.* 1–17 (2017). doi:10.1017/S0954579417001602

113. Miyake, A. *et al.* The unity and diversity of executive functions and their contributions to complex 'Frontal Lobe' tasks: a latent variable analysis. *Cogn. Psychol.* **41,** 49–100 (2000).

114. Lafleche, G. & Albert, M. S. Executive function deficits in mild Alzheimer's disease. *Neuropsychology* **9,** 313–320 (1995).

115. McGuinness, B., Barrett, S. L., Craig, D., Lawson, J. & Passmore, A. P. Executive functioning in Alzheimer's disease and vascular dementia. *Int. J. Geriatr. Psychiatry* **25,** n/a-n/a (2009).

116. Abe, K. *et al.* Cognitive function in amyotrophic lateral sclerosis. *J. Neurol. Sci.* **148,** 95–100 (1997).

117. Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S. & Glahn, D. C. Meta-analysis of 41 Functional Neuroimaging Studies of Executive Function in Schizophrenia. *Arch. Gen. Psychiatry* **66,** 811 (2009).

118. MARIJE BOONSTRA, A., OOSTERLAAN, J., SERGEANT, J. A. & BUITELAAR, J. K. Executive functioning in adult ADHD: a meta-analytic review. *Psychol. Med.* **35,** 1097–1108 (2005).

119. Friedman, N. P., Rhee, S. H., Ross, J. M., Corley, R. P. & Hewitt, J. K. Genetic and environmental relations of executive functions to antisocial personality disorder symptoms and psychopathy. *Int. J. Psychophysiol.* (2018). doi:10.1016/J.IJPSYCHO.2018.12.007

120. Friedman, N. P., Corley, R. P., Hewitt, J. K. & Wright, K. P. Individual Differences in Childhood Sleep Problems Predict Later Cognitive Executive Control. *Sleep* **32,** 323–333 (2009).

121. Brick, L. A. *et al.* Overlapping genetic effects between suicidal ideation and neurocognitive functioning. *J. Affect. Disord.* **249,** 104–111 (2019).

122. Martel, M. M. *et al.* A General Psychopathology Factor (P Factor) in Children: Structural Model Analysis and External Validation Through Familial Risk and Child Global Executive Function. *J. Abnorm. Psychol.* (2016). doi:10.1037/abn0000205

123. Snyder, H. R., Miyake, A. & Hankin, B. L. Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Front. Psychol.* **6,** 328 (2015).

124. Bowie, C. R. *et al.* Predicting Schizophrenia Patients' Real-World Behavior with Specific Neuropsychological and Functional Capacity Measures. *Biol. Psychiatry* **63,** 505–511 (2008).

125. Johnson-Selfridge, M. & Zalewski, C. Moderator Variables of Executive Functioning in Schizophrenia: Meta-Analytic Findings. *Schizophr. Bull.* **27,** 305–316 (2001).

126. Boyle, P. A. *et al.* Executive Dysfunction and Apathy Predict Functional Impairment in Alzheimer Disease. *Am. J. Geriatr. Psychiatry* **11,** 214–221 (2003).

127. Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P. & Tucker-Drob, E. M. Genes

Unite Executive Functions in Childhood. *Psychol. Sci.* **26,** 1151–1163 (2015).

128. Friedman, N. P. *et al.* Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol. Gen.* **137,** 201–25 (2008).

129. Gustavson, D. E. *et al.* Genetic and Environmental Influences on Verbal Fluency in Middle Age: A Longitudinal Twin Study. *Behav. Genet.* **48,** 361–373 (2018).

130. Friedman, N. P. *et al.* Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Dev. Psychol.* **52,** 326–40 (2016).

131. Young, S. E. *et al.* Behavioral Disinhibition: Liability for Externalizing Spectrum Disorders and Its Genetic and Environmental Relation to Response Inhibition Across Adolescence. *J. Abnorm. Psychol.* **118,** 117–130 (2009).

132. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11,** e1004219 (2015).

133. Ibrahim-Verbaas, C. A. *et al.* GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol. Psychiatry* **21,** 189–197 (2016).

134. Miyake, A. & Friedman, N. P. The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Curr. Dir. Psychol. Sci.* **21,** 8–14 (2012).

135. Hu, L. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* **6,** 1–55 (1999).

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

137. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).

138. GTEx Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–5 (2013).

139. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47,** 1091–1098 (2015).

140. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–35 (2006).

141. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49,** 1107–1112 (2017).

142. Bergen, S. E. & Petryshen, T. L. Genome-wide association studies of schizophrenia: does bigger lead to better results? *Curr. Opin. Psychiatry* **25,** 76–82 (2012).

143. Friedman, N. P., du Pont, A., Corley, R. P. & Hewitt, J. K. Longitudinal Relations Between Depressive Symptoms and Executive Functions From Adolescence to Early

Adulthood: A Twin Study. *Clin. Psychol. Sci.* **6,** 543–560 (2018).

144. Richardson, G. B. & Hardesty, P. Immediate Survival Focus: Synthesizing Life History Theory and Dual Process Models to Explain Substance Use. *Evol. Psychol.* **10,** 147470491201000 (2012).

145. Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am. J. Psychiatry* **176,** 107–118 (2019).

146. Stahl, S. M. *PsycNET Record Display - PsycNET*. (Cambridge University Press, 2014).

Chapter 2: Supplement

Index

Online Methods

A. Sample and Acquisition

The Longitudinal Twin Study (LTS) is composed of adult individuals that were ascertained through the Twin Infant Project in 1984, and the MacArthur Longitudinal Twin Study in 1986. The Colorado Department of Health solicited participation in the registry using birth records of all families in which (1) both twins survived, (2) were healthy and (3) were within ~3 hours driving time of CU-Boulder. The sample is relatively representative of twin births for the time considering these criteria (Rhea et al., 2006). We tested all eligible individuals (those who consented and could safely enter the scanning environment), regardless of handedness, medication status, head injury history, or substance use history. We asked subjects to contribute a urine and saliva sample to assess current levels of various substances. About 30% of the sample no longer lives locally. These subjects traveled to Colorado for the imaging sessions.

B. MRI and Questionnaire Procedures

After completing informed consent, participants completed screening measures and the Center for Epidemiological Studies Depression scale (CESD), and were informed about scanning procedures. The Inventory of Callous and Unemotional traits (ICU) was collected online prior to the scanning session.

Cortical thickness was calculated with the Freesurfer analysis suite (http://surfer.nmr/mgh.harvard.edu)(1). T1-wieghted images were extracted using watershed/surface deformation procedure(2), surface deformation along intensity gradients to optimally differentiate gray matter, white matter and cerebral spinal fluid boundaries(3), and tessellation of gray/white matter boundary(4). The resulting surfaces were registered to standard spherical brain template(5, 6), and then used to compute a range of surface-based measurements. Finally, vertices were smoothed at 10mm across the cortex full-width-at-half-maximum (FWHM) isotropic kernel(7).

Supplemental Table S1. *Descriptive Statistics for CESD, ICU, and ICU Subscales*

Scale	Mean	SD	Min	Max	Cronbach's Alpha	Shapiro-Wilk for Raw Scores	Shapiro-Wilk for Square Root Transformed Scores
CESD	8.786	8.477	0	50.000	0.902	0.834	0.979
ICU total	15.595	6.389	1	38.000	0.771	0.984	0.992
ICU callous	2.118	2.345	0	18.000	0.595	0.81	0.915
ICU unemotional	6.180	2.959	0	15.000	0.814	0.974	0.912
ICU uncaring	5.453	3.326	0	17.000	0.736	0.961	0.973

Note. Descriptive statistics for the the Center for Epidemiological Studies (CESD) total scale and the Inventory of Callous and Unemotional traits (ICU) total scale and subscales.

Supplemental Table S2. *Univariate Twin Models for Behavioral Measures*

	Model Fit						Standardized Paths			
Model	**χ^2**	**df**	***p***	**AIC**	**BIC**	**RMSEA**	**A**	**C**	**D**	**E**
CESD										
ACE	11.164	6	0.0834	1745.565	1760.230	0.077	0.597	0.000	--	0.802
ADE	10.509	6	0.1048	1744.910	1759.576	0.072	0.002	--	0.610	0.793
AE	**11.164**	**7**	**0.1316**	**1743.565**	**1754.564**	**0.064**	**0.597**	**--**	**--**	**0.802**
CE	33.427	7	<.0001	1765.828	1776.827	0.162	--	0.707	--	0.707
DE	10.509	7	0.1615	1742.910	1753.909	0.059	--	--	0.610	0.793
E	33.427	8	0.0001	1763.828	1771.161	0.148	--	--	--	1.000
ICU										
ACE	10.722	6	0.0974	1189.329	1203.939	0.074	0.574	0.244	NA	0.782
ADE	10.793	6	0.0950	1189.400	1204.010	0.075	0.627	--	0.000	0.779
AE	**10.793**	**7**	**0.1479**	**1187.400**	**1198.358**	**0.062**	**0.627**	**--**	**--**	**0.779**
CE	12.597	7	0.0826	1189.204	1200.161	0.075	--	0.575	0.818	0.818
DE	12.018	7	0.1000	1188.625	1199.582	0.071	--	--	0.631	0.776
E	38.447	8	<.0001	1213.054	1220.359	0.163	--	--	--	1.000

Note. Model fit for univariate models of Center for Epidemiological Studies-Depression scale (CESD) and the Inventory of Callous and Unemotional traits (ICU) total score. The CESD and ICU were residualized on mean thickness and sex. For each model, we tested what combination of A (Additive genetic), D (Dominance genetic), C (Common environment), or E (nonshared Environment) best fit each scale. Dashes indicates that the parameter was not estimated in that particular model. We used χ^2 difference testing, AIC, BIC and Root Mean Square Error of Approximation (RMSEA) as standards for model comparison. The preferred model is indicated in bold-face type. Although the AE and DE models fit similarly, the AE model was preferred, because D without A variance is biologically implausible. *N*= 285 same-sex twin pairs (142 monozygotic [MZ] and 143 dizygotic [DZ]). This sample was larger than that in the main analysis because it included twin pairs that were missing brain data but had behavioral assessments.

Supplemental Table S3. *Genetic Association Between CESD and Each ICU Subscale*

Var1	**Path1A**	***rG***	**Gr**	***Pvalue***
Callousness	0.456374	0.029352	0.00763	0.215
Uncaring	0.520091	0.491772	0.151531	0.02
Unemotional	0.790123	0.08316	0.038975	0.502
Total Scale	0.585954	0.351804	0.12266	0.036

Note. Center for Epidemiological Studies Depression scales (CESD) and Inventory of Callous and Unemotional traits (ICU) were residualized on sex and mean thickness. *r*G represents the genetic correlation and bivariate heritability represents the phenotypic correlation predicted by genetic covariance. All estimates were derived from the standard bivariate Cholesky decomposition. The A path represents each subscale's standardized A estimate; squaring this value yields the heritability. The *p*-value for the genetic associations was estimated by a 1-df chi-square difference test of the Cholesky cross path.

Supplemental Table S4. *Meta-analytic Terms for Right Hemisphere Clusters*

R-PreSMA		**R-Precuneus**		**R-PCC**		**R-Frontal Lateral Gyri**		**R-Frontal Sulci**	
Term	***R***	**Term**	***r***	**Term**	***r***	**Term**	**r**	**Term**	***r***
conflict	0.164	causality	0.119	inhibitory	0.143	None		reducing	0.3
distractors	0.145	precuneus posterior	0.106	Inhibit	0.119			relied	0.115
cortex anterior	0.077	experimentally	0.101	abnormality	0.108			middle cingulate	0.085
orienting	0.07	precuneus	0.096	prefrontal cortices	0.089			imagine	0.067
cortex acc	0.056	centered	0.09	chronic pain	0.075			amygdala anterior	0.039
acc	0.053	cortex precuneus	0.087	nervous	0.074			prefrontal cortex	0.02
dorsolateral prefrontal	0.036	theory	0.084	posterior medial	0.037			cingulate	0.018
anterior cingulate	0.035	deactivations	0.082	cingulate	0.023			prefrontal	0.018
task	0.035	midline	0.082	anterior insula	0.01				
anterior	0.03	spontaneous	0.082	brainstem	0.009				
cingulate cortex	0.029	thoughts	0.08	midbrain	0.008				
anterior insula	0.025	mode	0.077	pain	0.007				
cingulate	0.025	posterior cingulate	0.077						
supplementary	0.023	Pcc	0.075						
supplementary motor	0.023	default mode	0.073						
parietal cortex	0.022	mode network	0.072						
execution	0.02	connectivity networks	0.071						
basal ganglia	0.019	default	0.07						
ganglia	0.019	Self	0.068						
prefrontal	0.018	preserved	0.066						
motor	0.009	striking	0.066						
		mental	0.065						
		theory mind	0.065						
		deactivation	0.064						
		independent component	0.064						

Note. Top associated terms for a Neurosynth decoder analysis for each right hemisphere overlay cluster. Decoder analysis uses a meta-analytic method to find terms that appear frequently in papers that report effects at the coordinates input into the analyses. Each cluster was estimated as a mask and then put through decoder analysis separately. Each term was put into a single vector (excluding repeating terms for each mask) and the most common words across all areas were reported in the main text, using a wildcard* to account for like words.

Supplemental Table S5. *Meta-analytic Terms for Left Hemisphere Clusters*

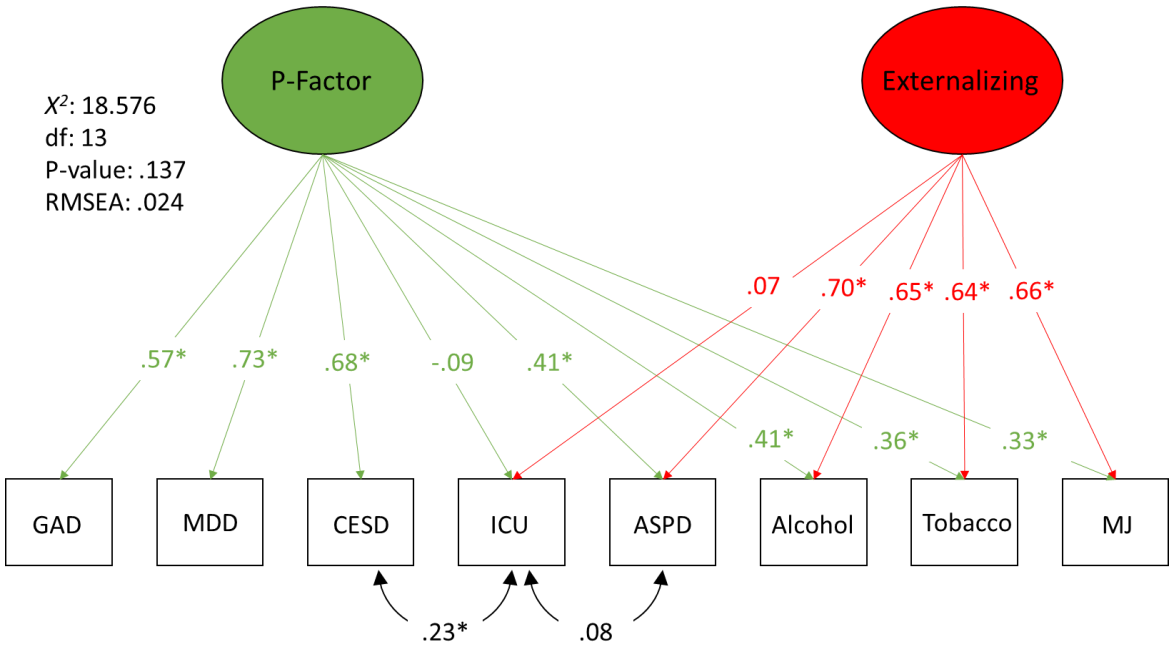
L-vSMA		**L-PTJ**		**L-OTJ**		**L-Precuneus**	
Term	***R***	**Term**	***r***	**Term**	***r***	**Term**	***r***
receiving	0.221	having	0.333	picture	0.295	personal	0.1
practice	0.18	indirect	0.284	researchers	0.109	sparse	0.069
facilitated	0.13	trained	0.138	videos	0.1	thinking	0.065
touch	0.117	failure	0.134	convergence	0.094	cortex posterior	0.061
sii	0.09	temporo parietal	0.126	actively	0.091	maintaining	0.061
tactile	0.09	inhibit	0.123	monkey	0.086	precuneus	0.051
si	0.088	temporo	0.113	extrastriate	0.075	cingulate cortices	0.043
sensory	0.079	recording	0.081	selectivity	0.075	states	0.043
induced	0.075	access	0.08	pictures	0.07	nervous	0.037
secondary somatosensory	0.075	online	0.076	scene	0.062	dorsomedial prefrontal	0.032
s1	0.074	response inhibition	0.076	motion	0.061	mental states	0.025
somatosensory cortex	0.072	valid	0.068	primary visual	0.061	mind tom	0.022
stimulation	0.068	mind	0.054	consecutive	0.06	posterior cingulate	0.022
somatosensory	0.067	reappraisal	0.054	middle occipital	0.052	tom	0.022
primary secondary	0.065	error	0.052	mt	0.052	autobiographical memory	0.018
lobule ipl	0.064	links	0.046	extrastriate visual	0.051	default mode	0.018
Painful	0.061	endogenous	0.044	plus	0.051	mode	0.017
primary somatosensory	0.06	theory mind	0.042	moving	0.048	mode network	0.016
discriminative	0.059	mind tom	0.04	v1	0.045	theory mind	0.016
sparse	0.058	successful	0.039	direction	0.044	episodic	0.015
motor control	0.057	group healthy	0.037	occipital	0.044	default	0.014
matrix	0.056	situation	0.036	body	0.043	autobiographical	0.012
motor task	0.056	actively	0.032	parieto	0.043	cingulate	0.011
secondary	0.053	mtg	0.032	social cognition	0.043		
pain	0.052	tom	0.032	vision	0.041		

Note. Top associated terms for a Neurosynth decoder analysis for each left hemisphere overlay cluster. Decoder analysis uses a meta-analytic method to find terms that appear frequently in papers that report effects at the coordinates input into the analyses. Each cluster was estimated as a mask and then put through decoder analysis separately. Each term was put into a single vector (excluding repeating terms for each mask) and the most common words across all areas were reported in the main text, using a wildcard* to account for like words.

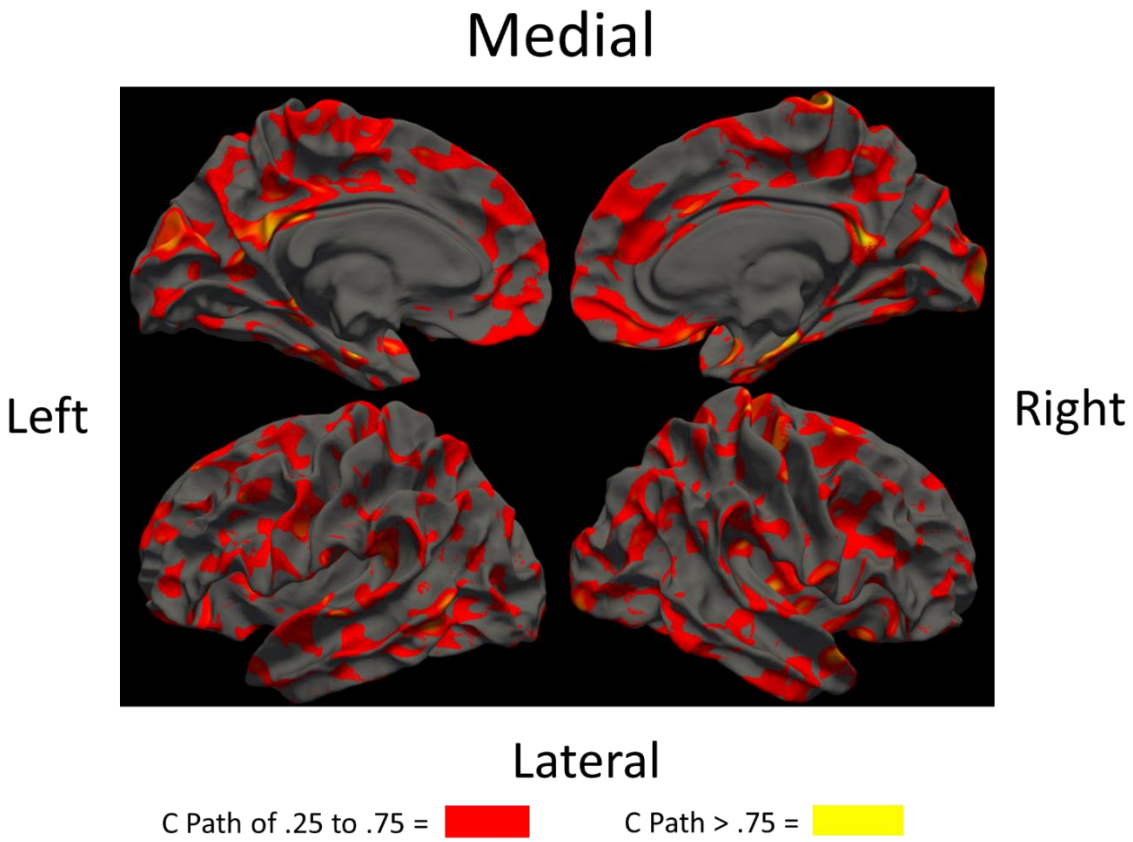
Supplemental Table S6. *Spatial Overlap of Overlap Clusters with the Yeo 7 Networks*

Overlap Cluster	**Network In Yeo 7**	**Direction**
L-OTJ	Visual	+
L-TPJ	Default	+
L-vSMA	Somatamotor	-
L-Medial Precuneus	Dorsal Attention	+
L-Lateral Precuneus	Default	+
R-DLPFC sulci	Default	+
R-DLPFC Gyri	Frontal	-
R-Posterior Cingulate	Frontal	-
R-PreSMA	Ventral Attention	+
R-Medial Precuneus	Default	-

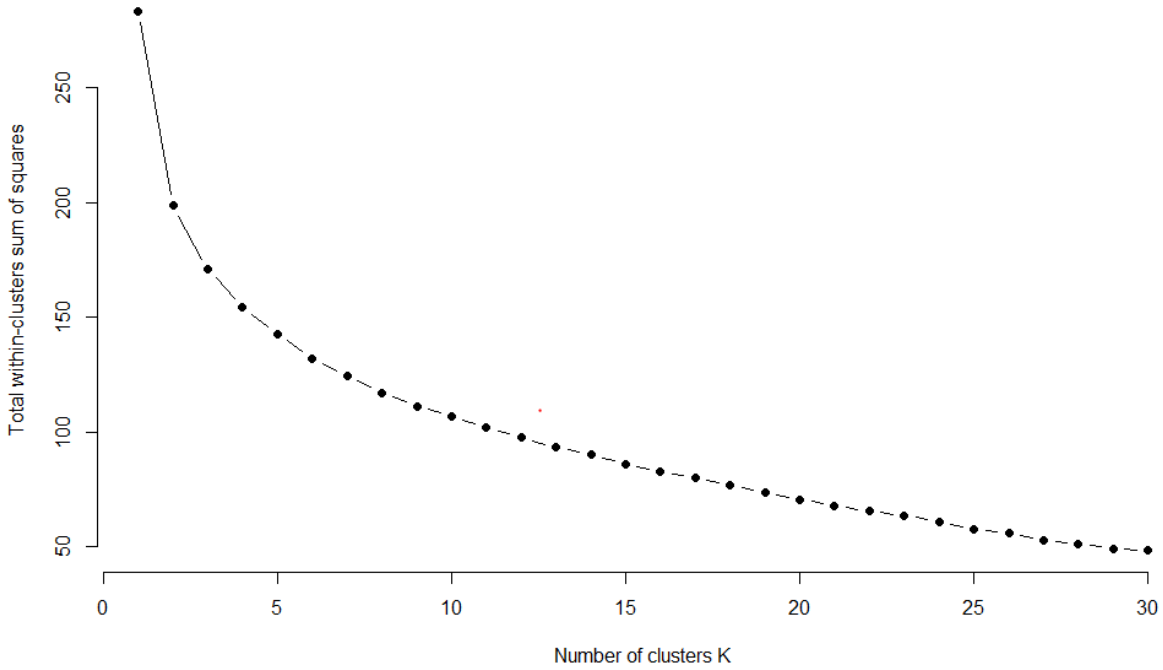
Note. Each cluster fell into only one network. The direction of effect in the original anatomical genetic brain map is shown to the right. Overlap is based on overlap of spatial coordinates, and not a statistical test of association. L- = left hemisphere and R- = right hemisphere. OTJ=Occipital Temporal Junction, TPJ = Parietal Temporal Junction, vSMA=ventral somatosensory Motor area, DLPFC = Dorsal Lateral Prefrontal Cortex, PreSMA = Pre-Somatosensory area.



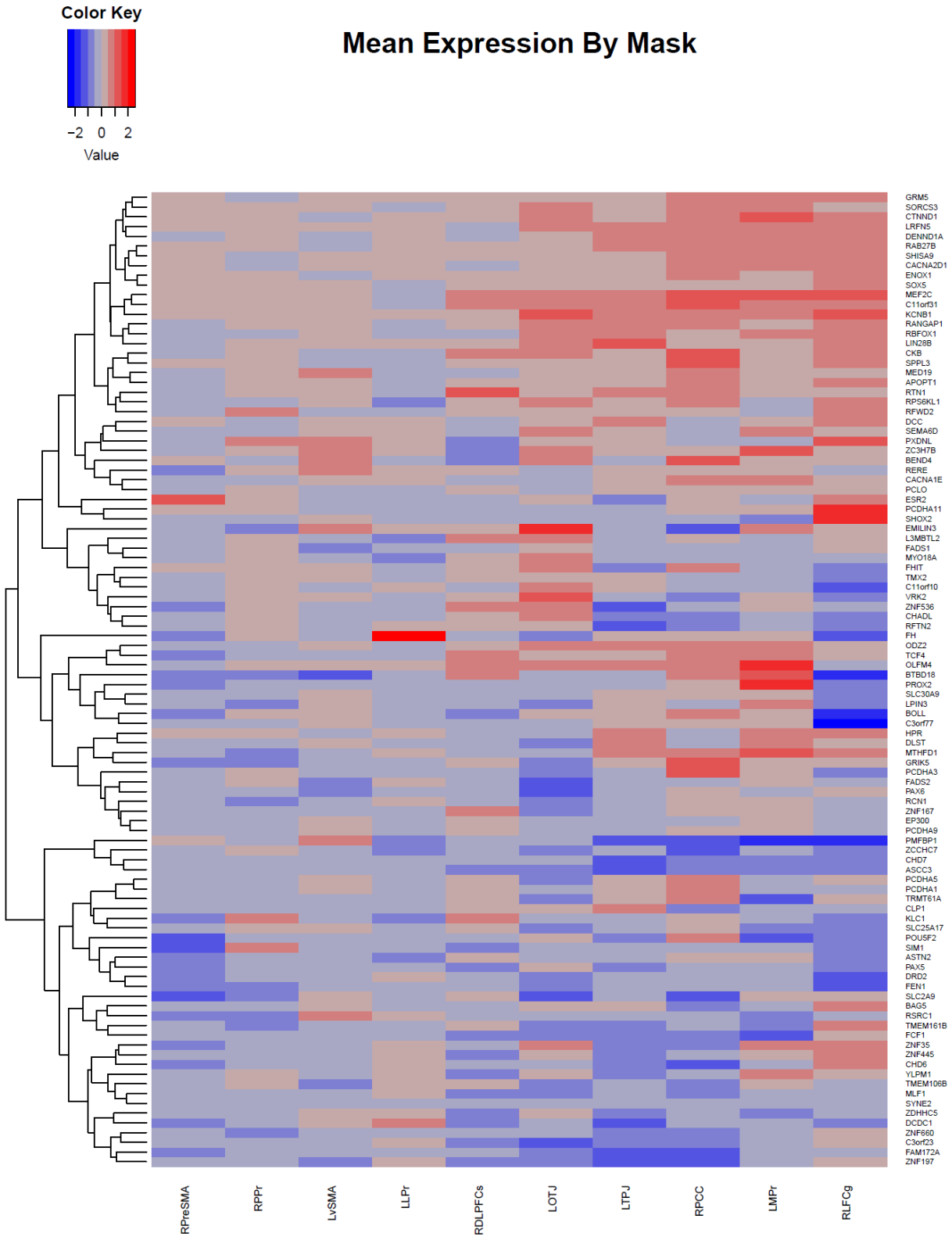
Supplementary Figure S1. P-factor model. The indicators for Generalized Anxiety Disorder (GAD), Major Depressive Disorder (MDD), Antisocial Personality Disorder (ASPD), Alcohol addiction (Alcohol), Tobacco addiction (Tobacco) and Marijuana addiction (MJ) were lifetime DSM-4 diagnosis, based on structured clinical interviews at age 23(8). The diagnosis variables were coded as 0 for no symptoms, 1 for symptoms but no diagnosis, and 2 for diagnosis. We treated these variables as ordinal with a threshold model, estimated with the weighted least squares, means and variance adjusted (WLSMV) estimator. Fit statistics for this model are shown to the left. We tested for a specific association (above and beyond the P-factor) between the CESD and ICU by adding a residual correlation. We also added a residual correlation between the ICU and ASPD, as these variables have been associated in the past. The model includes a P-factor and Externalizing factor only because the two measures that would normally load on an Internalizing factor left the model empirically underidentified when allowed to load on their own factor due to low factor loadings. *$p < .05$.



Supplemental Figure S2. Map of the shared environmental (C) effects for thickness across the whole cortex, thresholded by path estimate. C effects range from explaining 5 to over 50% of the variance across the brain. These estimates were taken from the same map as the Cholesky decomposition for the CESD to demonstrate the substantial C those models found across the cortex.

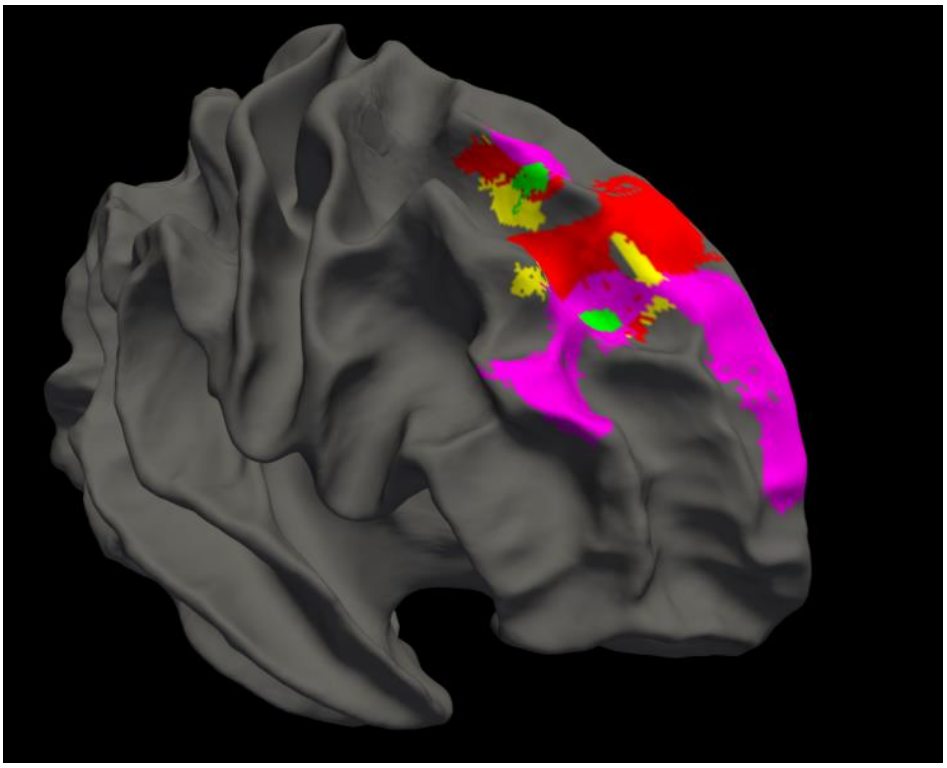


Supplementary Figure S3. Sum of squared error reduction with increased number of K clusters in K-means clustering. This plot is used for the "elbow-method" to determine number of clusters of an expression matrix.

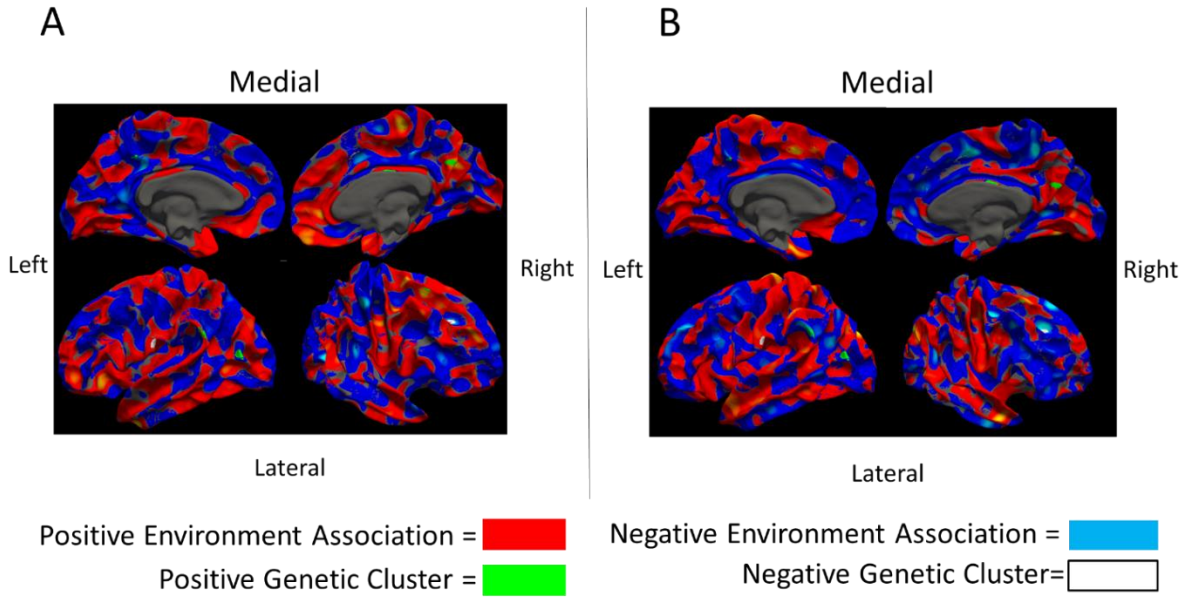


Supplemental Figure S4. Hierarchical clustering of expression patterns of depression genes in derived clusters. Color scale is the *z*-score for the degree of expression of that gene in the

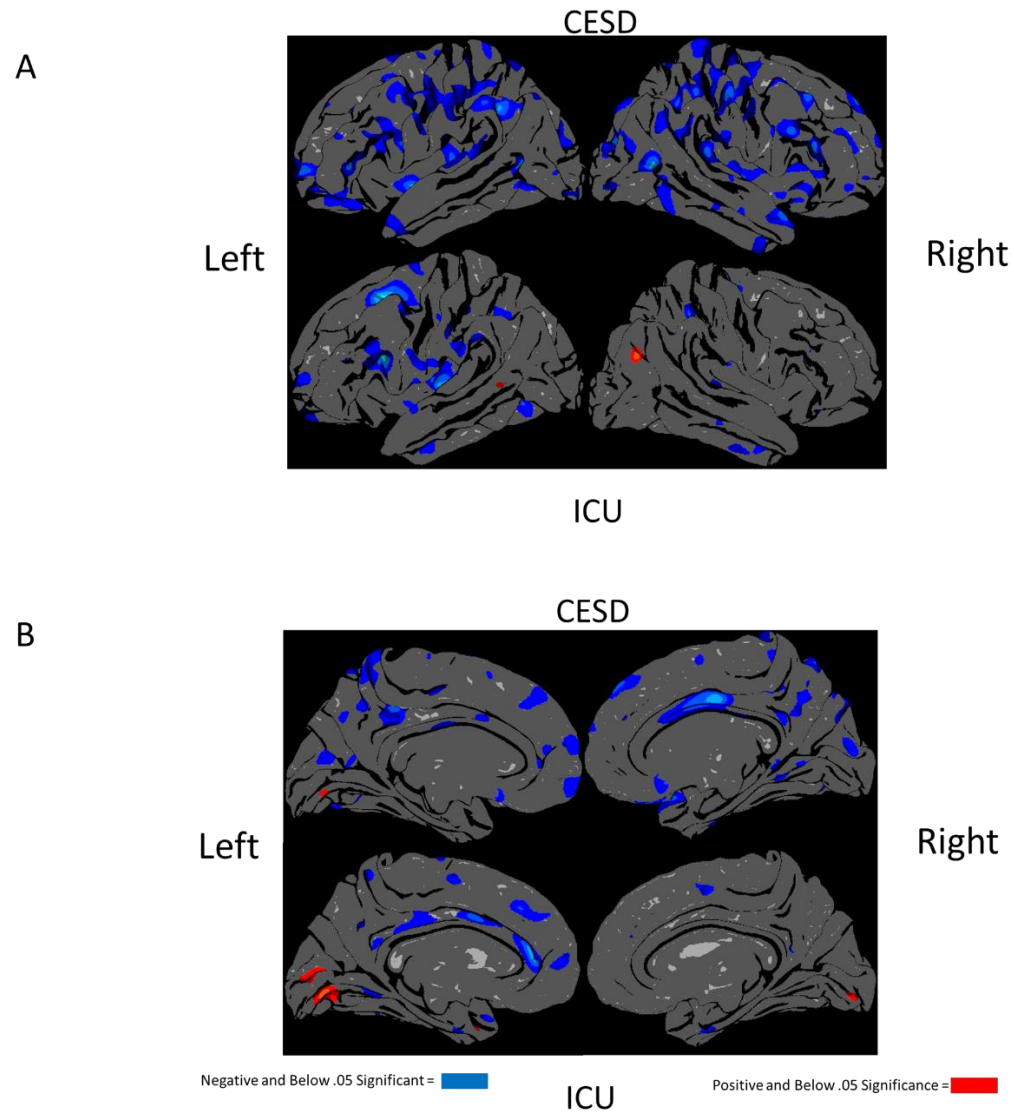
derived area mask compared to the whole cortex. Depression genes were obtained from the Psychiatric Genomics Consortium GWAS gene-burden tests, excluding genes from the major histocompatibility complex region(1). Gene expression values were recovered from Neurosynth-gene, which processed data from the Allen Brain Atlas, Human Brain Atlas. R = right hemisphere clusters; L = left hemisphere clusters. RDLPFCs = right dorsal lateral prefrontal cortex sulci, RLFCg = Right Lateral frontal gyri, LLPr = Left Lateral Precuneus, LMPr = Left Medial Precuneus, LOTJ = Left Occipital Temporal Junction, RPreSMA = Right Pre-Somatosensory Area, RPCC = Right Posterior Cingulate Cortex, LvSMA = Left Ventral somatosensory, LTPJ = Left temporoparietal Junction, and RPPr= Right Posterior Precuneus.



Supplementary Figure S5. Overlap across the split half replication in the right hemisphere. for the Inventory of Callous and Unemotional traits (ICU) and Center for Epidemiological Studies Depression scale (CESD). Yellow = CESD sample A, Green = CESD sample B, red = ICU Sample A, purple = ICU sample B. The more frontal clusters were all positively associated contiguous clusters and the more posterior clusters were all negative. These clusters overlap with the results from the full sample and are used to establish our "high confidence" clusters. The only split half-replicated results were in the right hemisphere.



Supplementary Figure S6. Environmental heatmap of Center for Epidemiological Studies Depression scale (CESD; Panel A) and Inventory of Callous and Unemotional traits (ICU: Panel B), with genetic overlap clusters visualized over the environmental association pattern. Positive environmental associations are red; negative are blue. The genetic clusters from our main analysis show the positively associate clusters in white and the negatively associated clusters in green. None of the genetic and environmental associations were in the same direction on this map.



Supplemental Figure S7. Phenotypic brain map results a standard qdec Freesurfer analysis of the Center for Epidemiological Studies Depression scale (CESD) and Inventory of Callous and Unemotional traits (ICU). Lateral views of the brain are in Panel A and medial views are in Panel B. Clusters represent significance at nominal significance ($p < .05$). We chose a liberal threshold for this plot so we would have more likelihood of finding similarity between phenotypic and genetic clusters, as a key premise of our method is that they might differ.

Online Methods

Participants. Participants were 501,826 individuals in the UK Biobank study who had completed at least one cognitive assessment at the time that the data were released to us in April, 2017[1,2]. The UK Biobank includes a total of 502,544 participants (54.4% female) aged 56.5 years (SD=8.1, range=37-73) at recruitment. Cognitive data were collected at up to four time points: an initial assessment visit (2006-2010) during which participants completed cognitive function tests on a touchscreen computer, a repeat assessment (2012-2013), an imaging visit (2014+), and a cognitive online follow-up (2014+). Sample sizes differed for each visit and task described below, as detailed in Table 1. Notably, the sample had dense assessment at cognitive online follow-up (93,024 individuals overall).

We restricted genetic analyses to 427,037 individuals of European ancestry as determined by principle components analysis (mean age=56.849, SD = 8.009, 54% female, 46% male) whose genotypes were imputed to the Haplotype Reference Consortium (McCarthy et al. 2016 Nature Genetics)[3], 1000 Genomes, and UK10K reference panels by the UK Biobank[2]. Subjects were genotyped on a UK BiLEVE array or the UKBiobank axiom array. After removing individuals with mismatched self-reported and genetic sex, we filtered imputed SNPs using a Hardy-Weinberg equilibrium P-value threshold of $>1\times10^{6}$, variant missingness > 0.05, imputation quality score (INFO) >0.95, and minor allele frequency (MAF) above 0.01, retaining 7,391,068 SNPs. More information is available in Bycroft et al.[3]

To guarantee consistent SNP effects across variably phenotyped subsets of individuals (Table S1) and to replicate SNP effects, we had three phases of GWAS analysis. First, we conducted our GWAS of cEF in the full sample (*n*=427,037, mean age=56.849, SD = 8.009, 54% female). Then, to evaluate consistency across subsamples, we divided the sample into a "Trails-

Online" or simply “Trails+” assessment sample (*n*=93,024, mean age=56.065, SD = 7.657, 55% female) and an "Trails-" sample (final Trails- sample *n*=256,135, mean age=56.996, SD = 8.050, 54% female). To be in the Trails+ sample, individuals must have completed at least the trail-making task. We pruned for related individuals before splitting into the Trails+ and Trails- samples (using Plink's "greedy" algorithm[4]) so SNP discovery and the genetic correlation between Trails+ and Trails-sample would not be biased by related individuals . Because of consistency and independent SNP replication between the Trails+ and Trails- samples we conducted all follow-up analysis on the full sample to increase power.

EF Measures

The cognitive battery in the UK Biobank contains one classic neuropsychological EF task, the trail making task, which requires shifting between sets of numbers and letters. The other cognitive measures were not tasks that are commonly used to assess particular EFs, but a number of them have EF components that can be extracted through our structural modeling approach. These tasks were symbol digit-symbol substitution, digit span, prospective memory, and pairs matching [be sure that these are in some table with the field IDs clearly listed]. We reasoned that a common factor extracting shared variance across these tasks and the trail making task would be closely related to the Common EF factors examined in smaller studies[5–7], two of which also used the trail making task[6,8].

Trail making (online). Participants clicked the computer mouse to join sets of circles "as quickly and accurately" as possible. The first set (numeric) consisted of 25 circles enclosing the numbers 1-25, which participants joined in ascending order. The second set (alphanumeric) consisted of 25 circles enclosing numbers and letters (1-13 and A-L), which participants joined in alternating order (1, A, 2, B, 3, C, etc.). Each set was preceded by a practice set of 8 circles. The total time

in seconds taken to correctly complete each set was recorded, starting when the participant clicked the first item. Only correct answers were accepted; incorrect answers were recorded but not analyzed here, because the time to commit errors was included already in the total time. The alphanumeric set is a classic neuropsychological EF measure because it requires participants to avoid the prepotent tendency to join stimuli in ascending order; rather, participants must switch between two sets of stimuli and maintain and update information about the current position within each set. The numeric set is used as a control condition to assess variation due to basic processing and motor speed. The dependent measure (DM) was the log-transformed time to complete the alphanumeric set (field 20157) residualized on the log-transformed time to complete the numeric set (field 20156).

Symbol-digit substitution (online). Participants saw a grid with 8 symbols above the digits 1-8, presented left to right, at the top of the screen. Underneath that, they saw the symbols re-arranged, and had to place the numbers 1-8 underneath them using the keypad, "as quickly and accurately" as possible. After a practice set, participants had 1 minute to complete as many grids as possible. The DM was the number of symbol-digit matches made correctly (field 20159); data from individuals who "abandoned" (field 20245) the test were treated as missing. Although this test is often used as a processing speed measure, its requirements to avoid the prepotent tendency to enter numbers in order and to update which symbol is paired with which number across grids are somewhat executive in nature. Supporting this conceptualization, trail making performance, a classic measure of EF that controls for basic speed, correlated more strongly with symbol-digit substitution (r= –.34) than with the simple reaction time (RT) measures we describe below (r= .16 to .18 for the 3 RT assessments). Thus, we included this test in both the EF (loading = .606) and RT factor score models (loading = .383).

Prospective Memory (initial visit, repeat assessment, imaging visit). Part 1: Before any other cognitive tests, participants saw the following text, "At the end of the games we will show you four coloured shapes and ask you to touch the Blue Square. However, to test your memory, we want you to actually touch the Orange Circle instead." Part 2: After they completed the other cognitive tests, they saw the following text, "That's the last game. Just one more thing left to do…" After they press "Next", they saw a screen with four shapes (blue square, pink star, gray cross, and orange circle), along with the instruction, "Please touch the Blue Square then touch the 'Next' button." If they pressed "Next" without touching a shape, they were prompted to touch a shape. The symbol they touched was surrounded by a yellow box. If they touched any symbol besides the blue square, the test ended, but if they touched the blue square, they received the following instructions, "At the start of the games we asked you to remember to touch a different symbol when this screen appeared. Please try to remember which symbol it was and touch it now." This prompt repeated each time they touched the blue square, until they touched any other symbol, which ended the test. The DM was whether they touched the orange circle on the first try (field 20018). Data from individuals who "abandoned" (field 4287) the test were treated as missing. If they never touched the orange circle or touched it on the second try, it was scored as incorrect. Although this test assesses memory, we judged it to also have an EF component because that memory is for a goal that must be used to override the more salient current instruction to touch the blue square.

Pairs Memory (initial visit, repeat assessment, imaging visit, online). In the first round of this task, participants had 3s to memorize an array of 6 cards with 3 symbols (i.e., 3 pairs) displayed in a random order. The cards were then shown face-down, and participants had to select the matching pairs. After they touched 2 cards, they were turned over. If they matched, the pair

disappeared and the participant selected another pair. If they did not match, they were turned face-down again and the participant touched another pair. This continued until all pairs were correctly identified. After the 6-card round, participants completed a 12-card (6 pairs) round. The DM was the log-transformed number of incorrect matches in the round +1 (field 399 for in-person and 20132 for online), summed across the 6- and 12-card rounds. Data were treated as missing if participants did not match all of the pairs in a round (e.g., if they abandoned the task; field 398 for in-person and 20131 for online), and the DM was only created for participants who had complete data for both rounds. Although this task taps visuospatial memory, it also requires working memory maintenance and updating ability, particularly as incorrect pairs are revealed, so we included it as an EF measure.

Digit span (initial visit, imaging visit, online). The digit span test (called "numeric memory" by UK Biobank) required participants to recall numbers with increasing numbers of digits (from 2 to 12). Participants were shown a number for 2s + 500ms*number of digits (e.g., 3s for a 2-digit number). The number disappeared and after 3s, the participant entered the number. After pressing "Next," the entry was removed and the next number was presented 600ms later, or the test was ended. The number length increased by 1 digit with each correct answer, and the test was terminated after 2 successive incorrect answers (if 3 or more digits), or 5 successive incorrect answers if 2 digits. The keyboard was deactivated when entry screen was not present. Each number was different than the previous number and the previous but one number. The DM was the maximum number of digits remembered correctly (field 4282 for in-person and 20240 for online); data from individuals who "abandoned" (coded as a score of –1 and/or "abandoned" in field 4281 for in-person) the test was treated as missing. The digit span is a classic test of verbal short-term memory, which is not generally considered executive (unless the backward

span is used, which requires working memory to re-arrange); typically, complex working memory span tasks that have a simultaneous processing requirement are used to tap working memory EF processes. However, as argued by Unsworth and Engle (2007)[9], simple and complex working memory tasks seem to measure similar processes (e.g., working memory maintenance, updating, and controlled retrieval) but differ in the extent to which those processes operate. Moreover, as the number of digits exceeds short-term memory span (supraspan), the task becomes more predictive of higher-order cognition. Thus, we included it as an EF measure.

Generation of common EF model

We used Mplus version 8 for the confirmatory factor analyses and extraction of the cEF factor score. Values for continuous measured tasks that were greater or less than 4 SDs from the mean were replaced with values equal to 4 SDs from the mean, after log-transforming skewed variables. This trimming procedure had little influence on the correlations or model results due to the large sample size, but it improved the normality of the distributions (see Table S1), which is an assumption of structural equation modeling. Variables were rescaled to have variances close to 1 to avoid ill-scaled matrices, which can cause model non-convergence, and variables were reversed so that for all measures, higher scores indicate better performance.

The prospective memory scores were categorical (pass/fail, Table 1 presents tetrachoric and biserial correlations with prospective memory performance); so we analyzed this data with a threshold model. The threshold model assumes these categories (of prospective memory) reflected an underlying normal distribution of probability of remembering. To incorporate the threshold model, used a means- and variances-adjusted weighted least-squares estimator (WLSMV, only pairwise deletion is available).

Table 1 presents the demographic information for each task. Figure 1A presents the phenotypic zero-order correlations among the cognitive measures used in the cEF factor models, as well as the cEF factor scores. The confirmatory factor analyses used to obtain the cEF factor score is shown in Figure 1B. All tasks loaded on the cEF factor, and orthogonal task-specific factors were used to account for repeated measurement of some tasks (prospective memory, pairs memory, and digit span). We did not analyze factor scores for the specific factors, which can be considered to reflect a combination of method variance as well as variance due to processes specific to that paradigm (i.e., uncorrelated with the other tasks in the model). The fit of this model was good, $\chi^2(44)=1786.53$, $p<.001$, CFI=.980, RMSEA=.009. With this sample size (total n=490,588), a large chi-square statistic for model fit is expected, but the model fit well by other fit criteria, particularly a CFI>.95 and RMSEA<.06[24]. When additional factors were added to capture time-specific effects, they had non-significant loadings, so they were not included.

Genome-Wide Association Analysis

In the full GWAS and both the Trails+ and Trails-subsets we followed the same GWAS procedure. We ran a test of association using a leave-one-chromosome-out Bayesian approximation of a linear mixed effect model using BOLT-LMM, controlling for age, age^2, sex, first 20 principal components (PCs), batch, and site. BOLT-LMM is a faster and more statistically powerful procedure for running GWAS in large samples compared to standard software like PLINK[4] and GCTA[10], and has demonstrated high efficiency with the UKBiobank[11]. Further the LMM procedure allows us to better account for stratification/cryptic relatedness and family structure than using only fixed-effect PCs. The summary statistics in analyses of SNP effects used BOLT's LMM infinitesimal model P-values[12]. Genome-wide results were

characterized using the FUMA/MAGMA[13] pipeline[14], LD score regression[15], and PrediXcan[16] (method details are below).

Characterization of SNPs

Annotation through FUMA/MAGMA pipeline. Using data from the 1000 Genomes Project (1000G)[17] phase 3 European (EUR) population as a reference, LD structure (r^2) of pairwise SNPs and minor allele frequencies (MAFs) were pre-computed. Independent significant SNPs ($r^2 < 0.6$) having a genome-wide significant *p-value* (5×10^{-8}) were distinguished. All SNPs available in the 1000G[17] EUR reference panel, in LD with the independent significant SNPs ($r^2 \geq 0.6$, max 1Mb window, MAF ≥ 0.01) were defined as candidate SNPs for association with EF. Of the identified significant SNPs, those independent at $r^2 < 0.1$ were defined as lead SNPs. LD blocks of all the identified independent significant SNPs and lead SNPs that were < 250 kb apart were combined and characterized as genomic risk loci. Because the power of GWAS is dependent upon how well causal variants are tagged, annotation is extended beyond independent significant SNPs to incorporate all candidate SNPs. Thus, candidate SNPs are functionally annotated and used for our gene prioritization analyses, while lead SNPs with the greatest significance (lowest *p-value*) are used to represent their respective genomic loci.

Functional Characterization of Lead Independent SNPs. To determine the functional consequences of SNPs significantly associated with cEF, ANNOVAR[18] was run on candidate SNPs located within the independent genomic loci to determine their functional consequences in FUMA (defaults: r2 ≥ 0.6, $p < 0.05$, MAF ≥ 0.01). These SNPs were matched according to chromosomal location, base pair position, reference and non-reference alleles and then annotated accordingly. To map candidate SNPs significantly associated with cEF to genes, two different strategies were applied based on Ensembl genes (build 85) using FUMA[14]. First, SNPs on or

near genes, determined via ANNOVAR[18] annotation, were positionally mapped to genes based on their physical distance (< 10 kb) from protein-coding genes.

Second, to determine if significantly associated SNPs related to gene expression, SNPs were further annotated by FUMA expression quantitative trait loci (eQTLs) status. SNPs that significantly affect gene expression were extracted from the GTeX sample[19] and BRAINEAC[20] sample. SNPs were mapped to genes within a 1 Mb window, known as cis-eQTLs, and were limited to only significant SNP-gene pairs (false discovery rate (FDR) ≤ 0.05, default in FUMA). GTEx v7 Brain and BRAINEAC database of brain-tissue-specific gene expression data were utilized to perform eQTL mapping of the following brain tissues: amygdala, anterior cingulate cortex BA24, caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex BA9, hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia, cervical (c-1) spinal cord, and substantia nigra. BRAINEAC eQTLs were used for the following brain tissues: cerebellar cortex, frontal cortex, hippocampus, inferior olivary nucleus, occipital cortex, putamen, substantia nigra, temporal cortex, thalamus, intralobular white matter.

Regulatory Elements of Intronic SNPs: CADD and Regulome scoring and 3D Chromatin Interaction. To determine whether intronic independent SNPs served a possible regulatory functioning, we annotated significant independent SNPs for their Combined Annotation Dependent Depletion score (CADD), Regulome score and for chromatin-chromatin interaction via 3D chromatin interaction (Hi-C)[21]. CADD scoring shows the likelihood that the variant is deleterious. Regulome scores intronic SNPs based on the likelihood that they have cis-regulatory function. Hi-C examines whether SNPs represent long-range enhancer-promotor associations. For Hi-C data from the following pre-existing builds were used in FUMA: dorsolateral PFC, hippocampus, neural progenitor cells.

Gene-level Regression Analyses via MAGMA

Gene-based analysis. To determine what genes are significantly associated with cEF and create a prioritized list of genes based on the degree of association, MAGMA[22] (Multi-marker Analysis of GenoMic Annotation) gene analysis was performed in FUMA. MAGMA Uses a multiple regression model run on GWAS summary statistics designed to incorporate LD between genetic variants and detect the aggregated effects of multiple weakly associated variants (. MAGMA combines the *p-values* of SNPs, mapped to protein-coding genes (< 10 kb of gene), to generate a gene-based *p-value*, in addition to genetic correlations between neighboring genes. This produces a prioritized list of significantly associated, protein-coding genes to quantify the level of association between the identified genes and EF. We ran this analysis in discovery sample, the replication sample and the full sample.

Gene-set analysis. To detect biological pathways significantly associated with cEF, we used MAGMA to run a competitive gene-set analysis and cell-type specific gene-set analysis. This analysis also accounts for potential confounding variables, such as gene density and size[23]. A competitive gene-set analysis is a gene-level linear regression model designed to determine whether genes within a gene-set have a significantly greater association with cEF than all other genes outside of the gene-set. Gene-sets are determined by shared biological and functional characteristics between genes defined by the datasets in MBsig6.1[24]. For the cell-type specific analysis, we annotated our findings with QTL information from RNA cell-type specific studies of human postmortem cortex[25], hippocampus[26] and frontal cortex[27] (during prenatal development).

Gene-property analysis to determine tissue specificity. To answer the question of what tissues SNP effects across the whole genome are likely expressed in, MAGMA gene-property analysis

was performed in FUMA. This gene-property analysis was performed on 30 general and 53 specific GTEx v7[19] tissue types.

Heritability and Genetic Correlations. To calculate cEF univariate heritability, we used BOLT-REML with a single variance component and considering all variants simultaneously[28]. We calculated genetic correlations of cEF with psychiatric, personality, neurological, and health related outcomes via LD Hub[29] with the GWAS summary statistics from the full sample. LD Hub[29] is a database of publicly available GWAS summary statistics and automated pipeline of LD (linkage disequilibrium) score regression analysis that is utilized to estimate heritability of a trait of interest and genetic correlations with that trait and other relevant traits.

Transcription Patterns. To identify genetic transcription patterns implicated by the whole-genome SNP effects and eQTL results, we ran a Transcriptome-Wide Analysis (TWAS) using PrediXcan[30] PrediXcan imputes gene expression from SNPs via an elastic net model trained in an eQTL sample, in this case the GTeX sample[19]. We chose PrediXcan because it uses summary statistics and allows variability across tissue types[30]. We ran PrediXcan with the summary statistics for our cEF GWAS separately for each tissue that was significant by MAGMA gene-property analysis.

Drug Relabeling Based on Known Associations. Because TWAS offers a predicted transcriptional profile, this profile can be compared to other datasets from computational pharmacogenomics. So et al. 2017[31] expanded on PrediXcan TWAS using the connectivity Map (cMAP) Library[32] to infer what drugs mimic the implicated transcription pattern. The cMAP is a dataset of experimental transcription changes in stem cell lines after exposure to a pharmaceutical substance[32]. We entered all significant inferred transcripts post-Bonferroni

correction and their predicted differential expression *z*-scores into the Connectivity map toolbox to obtain the top 15 substances predicted to reverse the pattern of transcription[31].

1. 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 Med.* **12,** e1001779 (2015).

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

3. Consortium, the H. R. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).

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

5. Miyake, A. & Friedman, N. P. The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Curr. Dir. Psychol. Sci.* **21,** 8–14 (2012).

6. Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P. & Tucker-Drob, E. M. Genes Unite Executive Functions in Childhood. *Psychol. Sci.* **26,** 1151–1163 (2015).

7. Gustavson, D. E. *et al.* Genetic and Environmental Influences on Verbal Fluency in Middle Age: A Longitudinal Twin Study. *Behav. Genet.* **48,** 361–373 (2018).

8. Gustavson, D. E., Miyake, A., Hewitt, J. K. & Friedman, N. P. Genetic relations among procrastination, impulsivity, and goal-management ability: implications for the evolutionary origin of procrastination. *Psychol. Sci.* **25,** 1178–88 (2014).

9. Unsworth, N. & Engle, R. W. The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychol. Rev.* **114,** 104–132 (2007).

10. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

11. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).

12. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46,** 100–106 (2014).

13. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11,** e1004219 (2015).

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

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

16. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47,** 1091–1098 (2015).

17. Gibbs, R. A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

18. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

19. GTEx Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–5 (2013).

20. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17,** 1418–1428 (2014).

21. Schmitt, A. ., Hu, M., Jung, I. & Xu, Z. et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17,** 2042–2059 (2016).

22. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11,** e1004219 (2015).

23. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–71 (2013).

24. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).

25. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 7285–90 (2015).

26. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14,** 955–958 (2017).

27. Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555,** 524–528 (2018).

28. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47,** 1385–1392 (2015).

29. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability

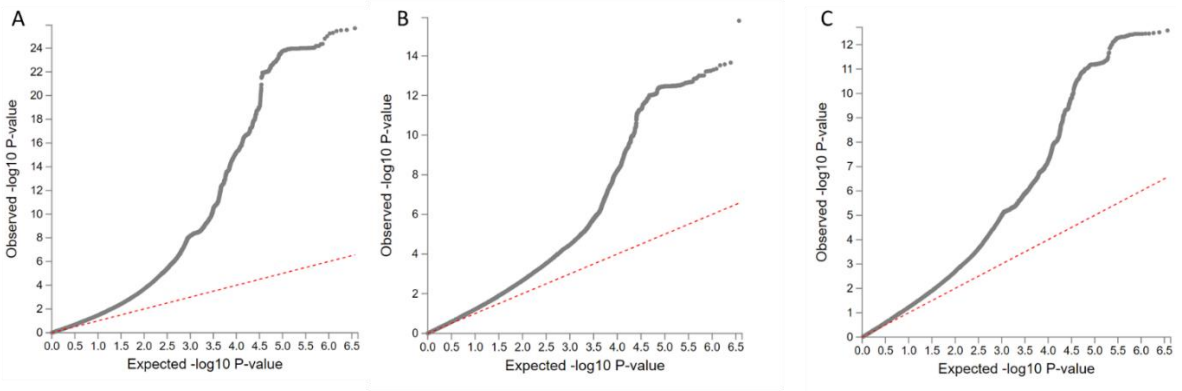
and genetic correlation analysis. *Bioinformatics* **33,** 272–279 (2017).

30. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv* 045260 (2017). doi:10.1101/045260

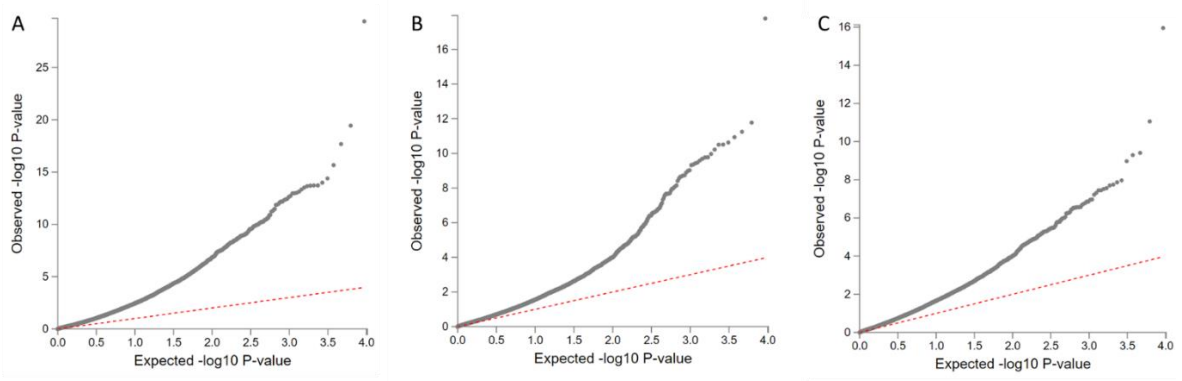
31. So, H.-C. *et al.* Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat. Neurosci.* **20,** 1342–1349 (2017).

32. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–35 (2006).

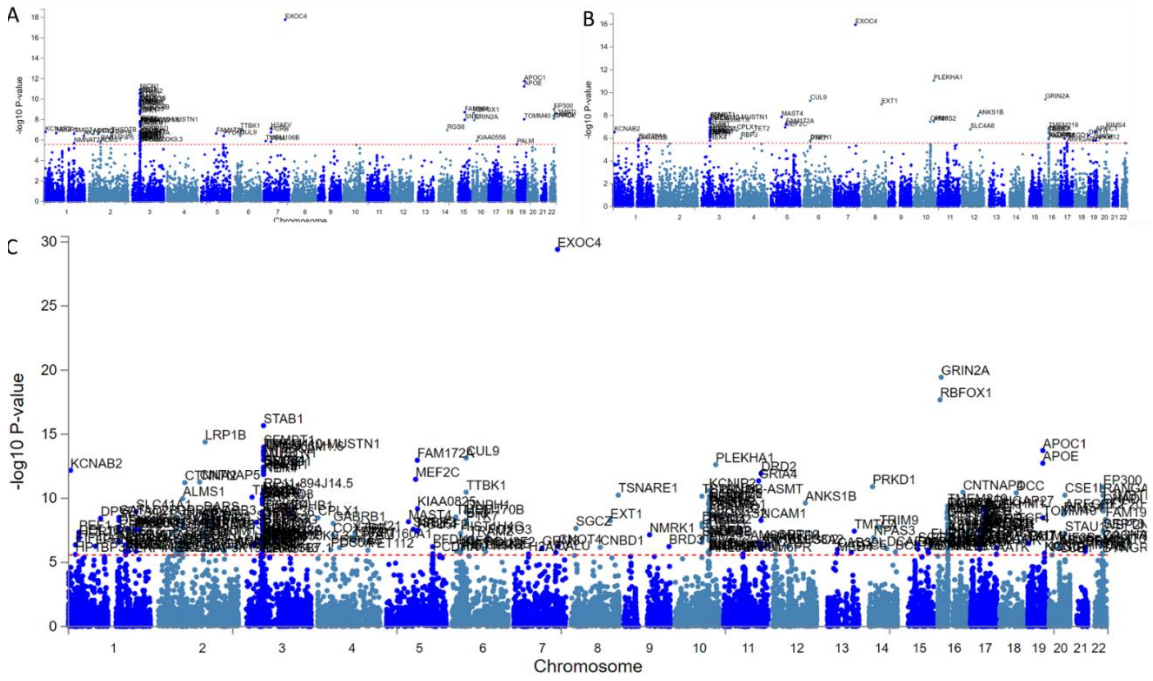
Online Figures



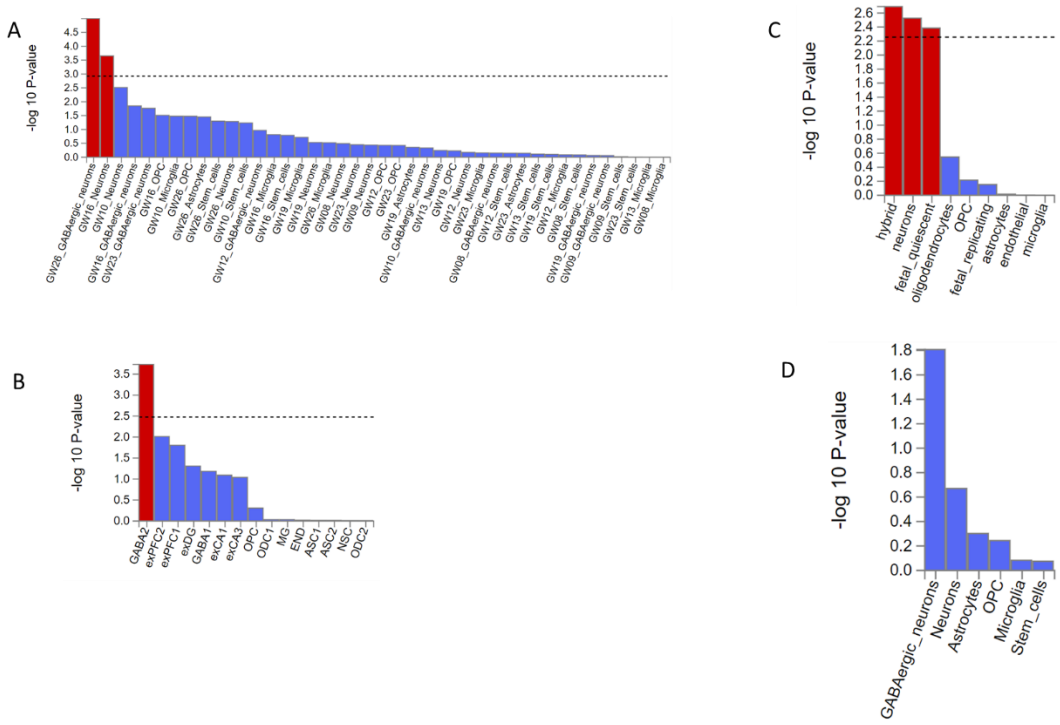
Online Figure S1. QQ plots of SNP p-values. Black dots represent the observed p-values plotted against the y axis on the -log10 scale, red dots represent the expected p-values plotted on the x-axis. P-values deviated substantially from expected. (A) QQ plot of p-values from the full sample. (B) represents p-values in online sample. (C) Offline sample p-values QQ plot, excluded individuals related to individuals in the online sample.



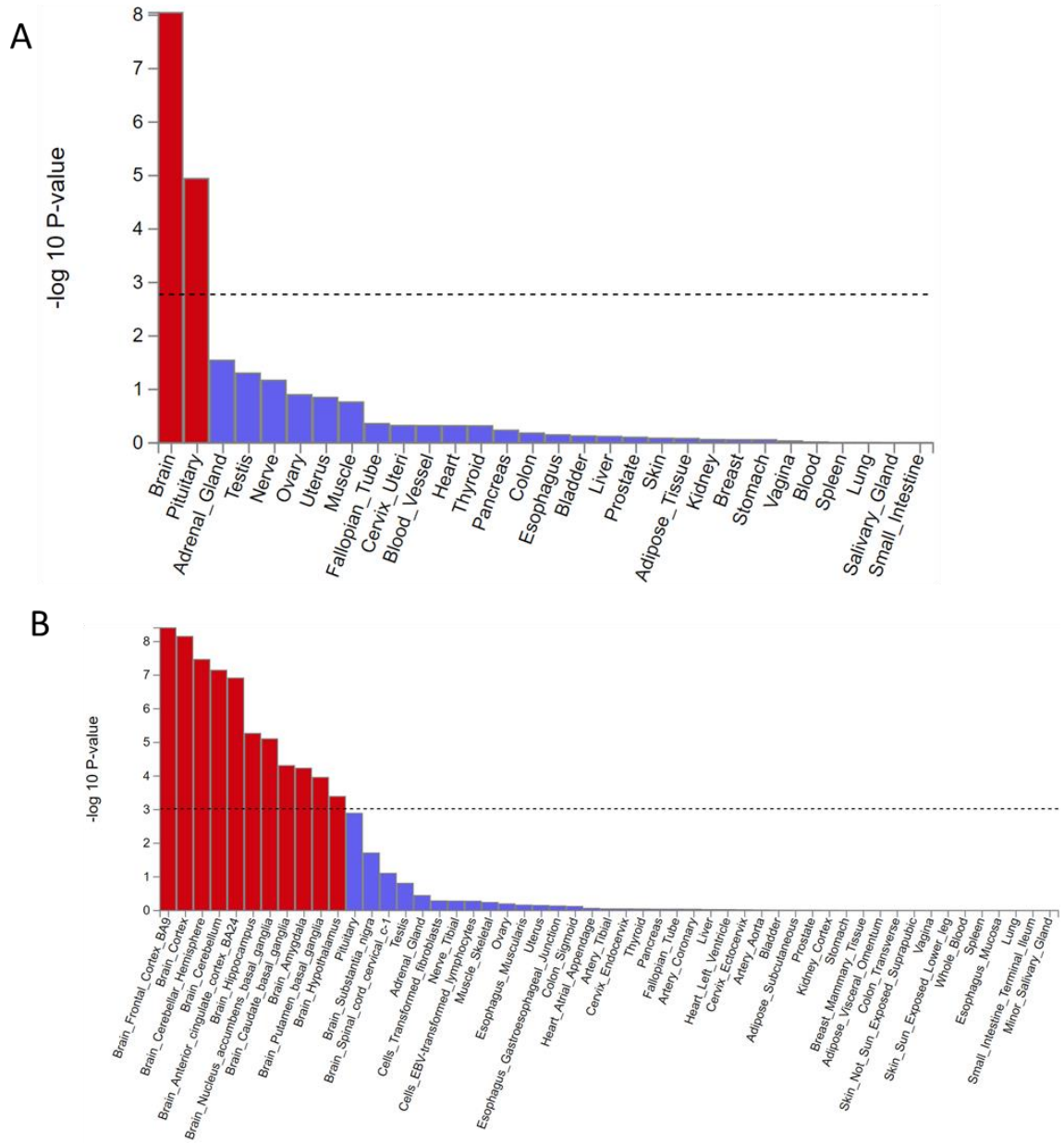
Online Figure S2. QQ plots of Gene-wise p-values. Black dots represent the observed p-values plotted against the y axis on the -log10 scale, red dots represent the expected p-values plotted on the x-axis. P-values deviated substantially from expected. (A) QQ plot of p-values from the full sample. (B) represents p-values in online sample. (C) Offline sample p-values QQ plot, excluded individuals related to individuals in the online sample.



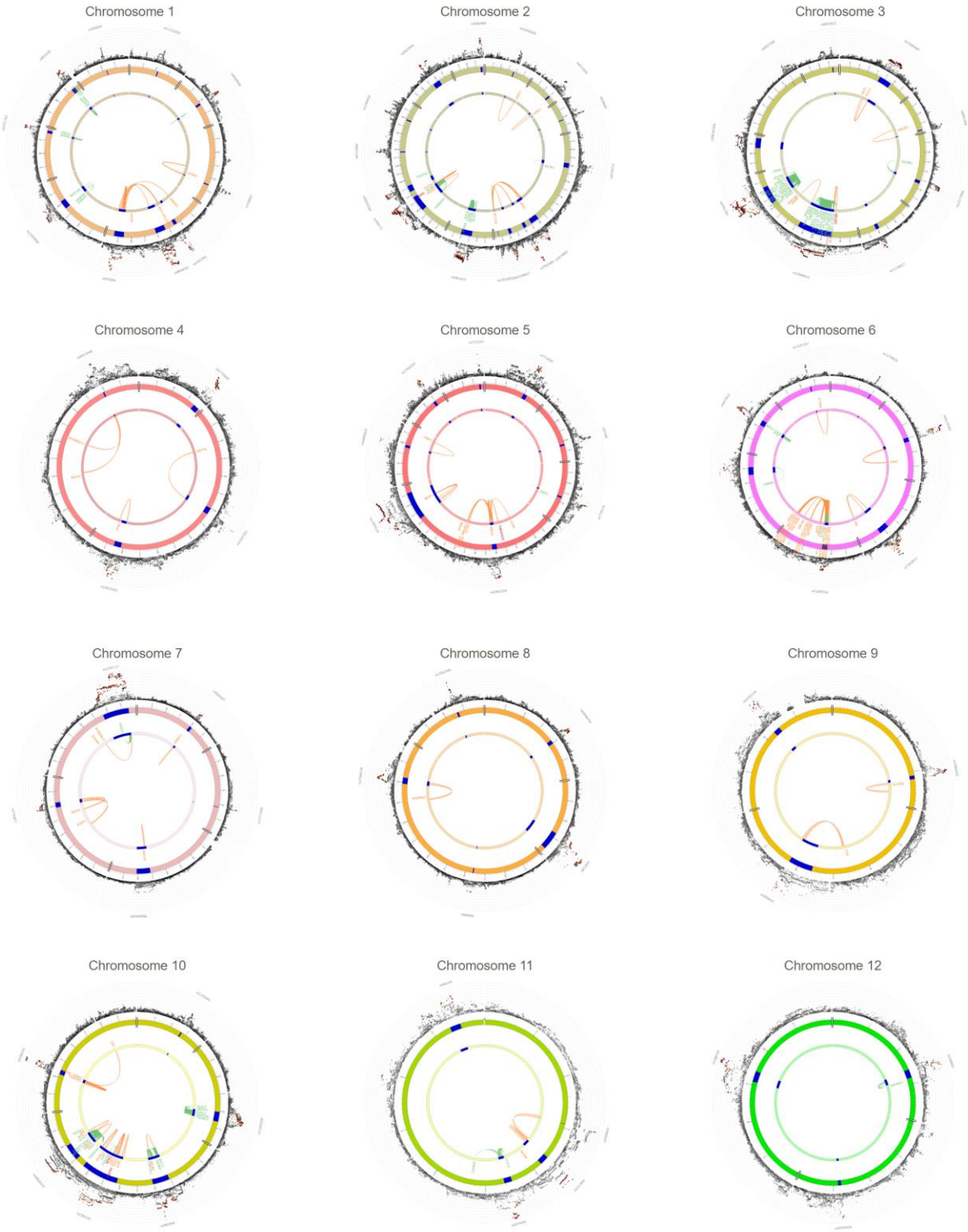
Supplemental Figure S3. Manhattan Plots for All Three Gene-Wise Association Tests of cEF. The UKBiobank was split into a discover sample that had dense online assessment (Panel A) from all tasks and an offline dataset that had missing on tasks for individuals (Panel B). Relatives were pruned from the Offline sample to ensure gene associations were not due to inflation by cryptic relatedness. The genes must have been associated by GWAS Bonferroni significance (P = 0.05/18739 = 2.668e-6). (C) Both samples were combined, and related individuals were included to increase power. All models were run using Bolt-LMM to account for polygenicity and (for panel C) family structure.



Supplemental Figure S4. RNA-cell type specific enrichment in Human post-mortem brain samples. To identify cellular mechanisms of cEF SNP associations we used MAGMA gene-set analysis to predict 4 different post-mortem datasets. A. Cortex expression patterns across fetal development. (B) Cell-type specific analysis in human post-mortem hippocampal tissue. C) combined cortex expression across time (adults and fetal tissue) to predict what cell-types (neuronal/immune) related to cEF. (D) Cell-type specific analysis in the human post-mortem prefrontal cortex tissue.



Supplemental Figure S5. Significant Tissue Enrichment for snp eQTLs in the GTeX v7 sample. Enrichment was estimated in MAGMA using a gene-level regression controlling for gene-size and population structure. Line represents Bonferroni significance. (A) Enrichment by broad general tissue types. (B) Enrichment by 53 specific tissue types.


Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12

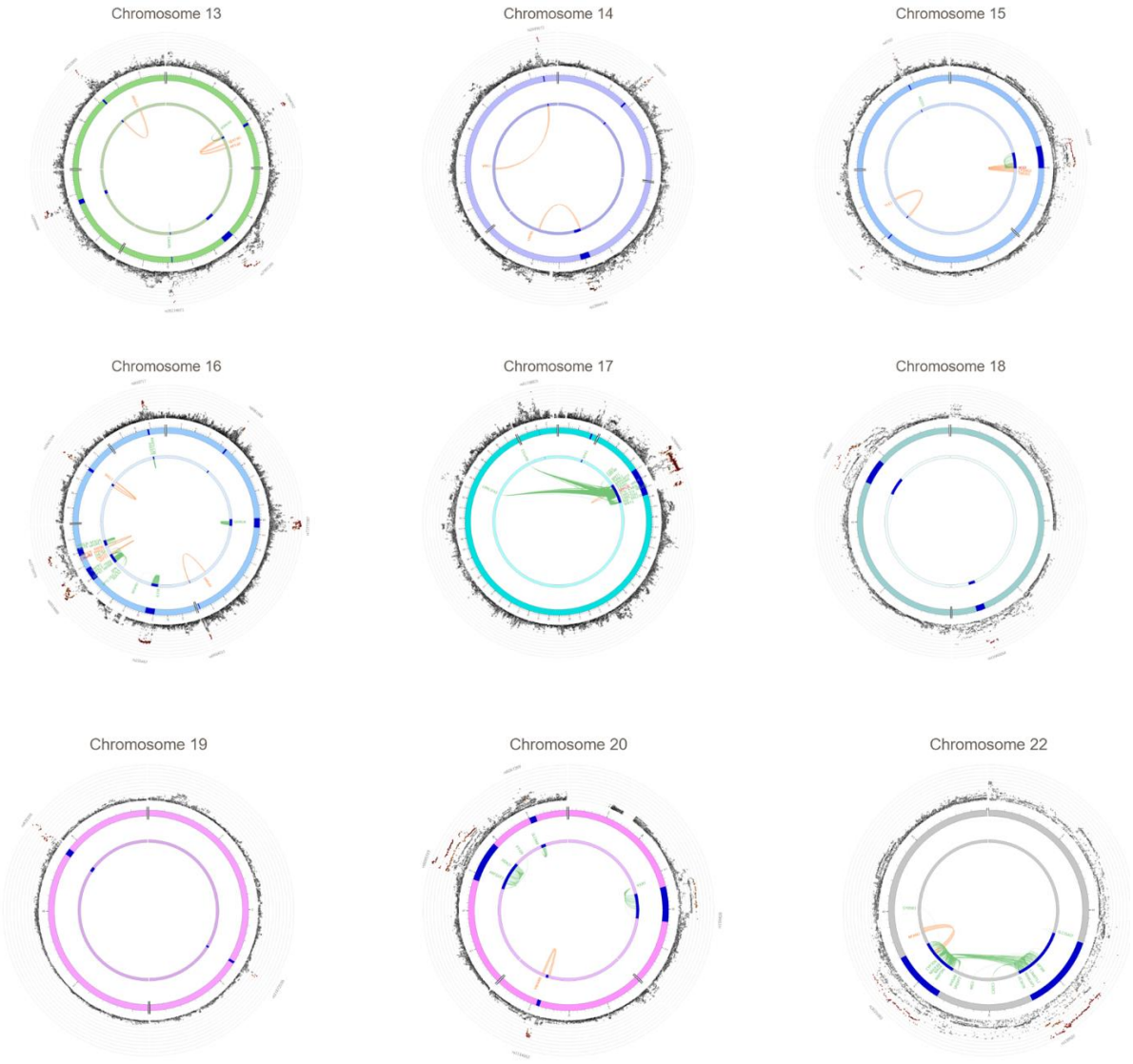


Figure S6. Circos plots by chromosome. Only genomic risk loci or eQTLs are mapped, the third layer shows whether it is an eQTL, it is green, orange for genomic risk loci and red for both. The outermost layer represents a manhattan plot with SNPs P <.05 displayed for genomic risk loci. The genomic loci are colored based on maximum r^2 to an independent significant SNP, red ($r^2 > 0.8$), orange ($r^2 > 0.6$), green ($r^2 > 0.4$) and blue ($r^2 > 0.2$), and gray means > .2. Along the chromosome ring (second layer) the risk loci are blue. The inner most layer (third layer) represents the interactions between SNPs.

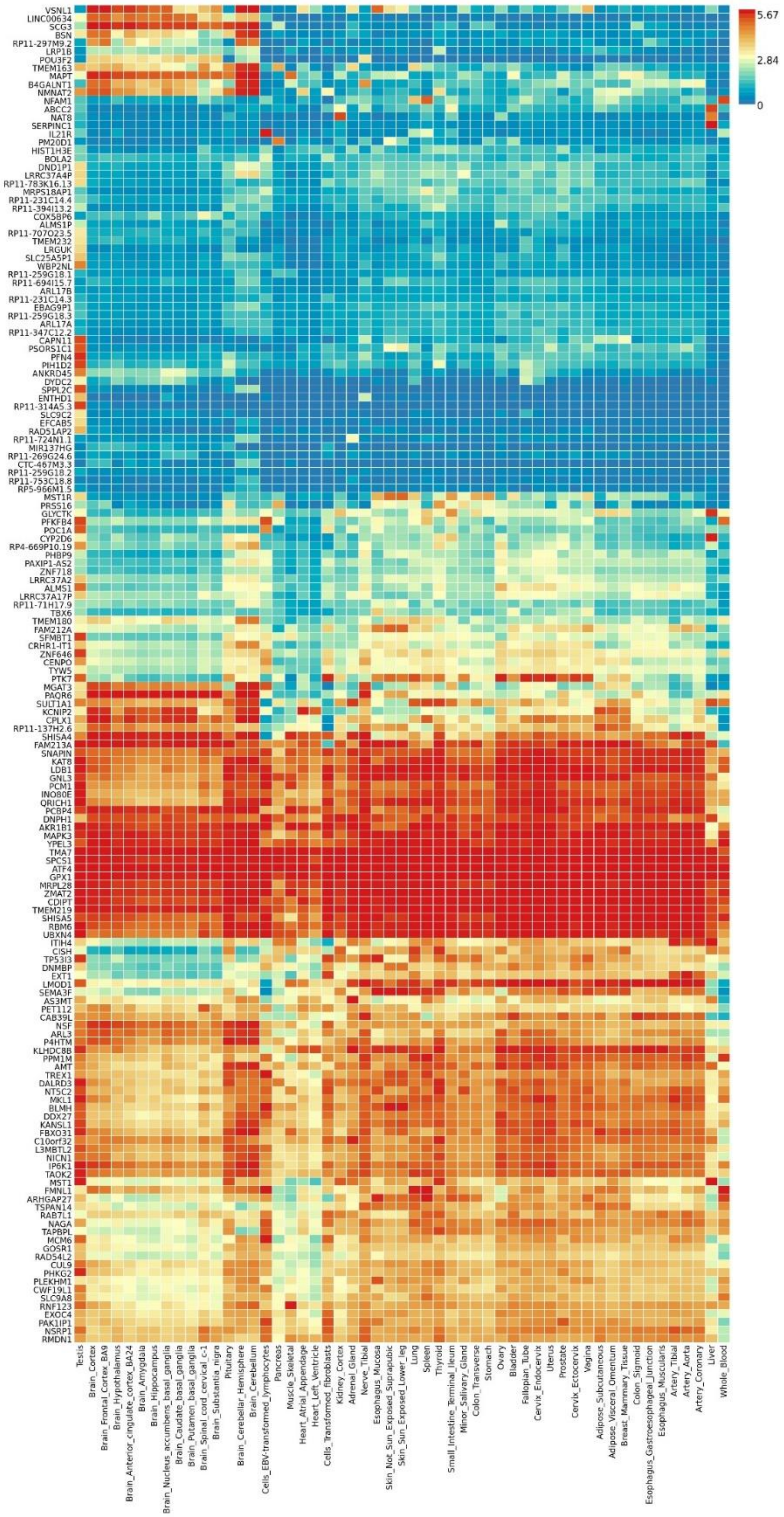


Figure S7. Transcriptional Profile cEF genes as a co-expression heatmap across 53 specific GTeX tissues. Results are clustered based on tissue and gene expression.