UNIVERSITY OF CALIFORNIA SAN DIEGO

UNDERSTANDING THE GENETIC INFLUENCE ON TRAITS RELATED TO BRAIN AND BEHAVIOR IN ADULTS AND CHILDREN

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Cognitive Science

by

Robert J. Loughnan

Committee in charge:

Professor Terry L. Jernigan, Chair
Professor Anders M. Dale, Co-Chair
Professor Chun Chieh Fan
Professor Eran A. Mukamel
Professor Wesley K. Thompson
Professor Bradley Voytek

2021

The dissertation of Robert J. Loughnan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To Rachel Koester your love and guidance has always kept me grounded, motivated, and supported. And to my parents whose hard work and perseverance has afforded the opportunities available to me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

LIST OF SUPPLEMENTARY FILES

Regression result tables for all models run in Chapter 4.

ACKNOWLEDGEMENTS

Over the past 5 years I have been fortunate to have been supported and surrounded by individuals that have enriched my time at UCSD. Firstly, I would like to thank Terry Jernigan and Chun Chieh Fan for their unwavering guidance and support both personal and academic without which my studies would not have been the same. Anders Dale and Wesley Thompson have provided invaluable insight in statistical, neuroimaging and genetics techniques that have immensely benefited me and my work. I would also like to thank Eran Mukamel for his mentorship, thoughtful discussions of scientific rigor and thorough explanations of techniques. Bradley Voytek and the members of his lab, in particular Scott Cole, Richard Gao and Tom Donoghue, as well as Jisoo Park in the Ideker Lab, were instrumental in my learning and developing effective programming skills that were important to all aspects of my research. Trey Ideker, Daniel Carlin and Jianzhu Ma immeasurably helped develop my understanding of network genetics and computational techniques to model them. Clare Palmer, Carolina Makowski, Diliana Pecheva and Weiqi Zhao have all cultivated a lab environment that has been enjoyable and filled with many interesting discussions. I would like to thank Ole Andreassen for fostering cross-continental collaborations with exemplary researchers such as Dennis van der Mer, Alexey Shadrin and Oleksandr Frei. Mary Boyle's enthusiastic discussions taught me a great deal about metabolic disorders and their effects on the brain. I would also like to thank Cherisse Tompkins and Johnathan Ahern for being motivated and diligent undergraduate mentees, and I am glad to have had the good fortune of working with them. I am grateful for the Kavli Foundation's generous funding at UC San Diego through the Innovative Research Grant. Lastly, Marta Kutas' guidance has greatly improved my ability to communicate research and her support has enabled me to be steadfast in my goals.

Chapter 2, in full, is available as a preprint on MedRxiv and has been submitted for publication. Loughnan R, Ahern J, Tompkins C, Palmer C, Sugrue L, Iversen J, Thompson

W, Andreassen O, Jernigan T, Dale A, Boyle M, Fan C. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 3, in full, is available as a preprint on BioRxiv and has been submitted for publication. Loughnan R, Palmer C, Thompson W, Dale A, Jernigan T, Fan C. The dissertation/thesis author was the primary investigator and author of this paper.

Chapter 4, in full, is available as a preprint on MedRxiv and has been submitted for publication. Loughnan R, Palmer C, Makowski C, Thompson W, Barch D, Jernigan T, Dale A, Fan C. The dissertation/thesis author was the joint primary investigator and author of this paper.

Chapter 5, in full, is available as a preprint on BioRxiv and has been submitted for publication. Loughnan R, Shadrin A, Frei O, van der Meer D, Zhao, W, Palmer C, Thompson W, Makowski C, Jernigan T, Andreassen O, Chieh Fan C, Dale A. The dissertation/thesis author was the joint primary investigator and author of this paper.

VITA

2021	Doctor of Philosophy in Cognitive Science University of California San Diego
2016-2021	Graduate Student Researcher Teaching Assistant University of California San Diego
2015-2016	Research Assistant University College London
2010-2015	Masters of Physics University of Manchester
2013-2014	Visiting Student University of California Los Angeles

GRANTS

Kavli Innovative Research Grant 2019-2020 ($50,000), Primary Investigator: *Imputing Molecular Endophenotypes from Genome Wide Association Data to Understand Biological Pathways Leading to Psychiatric Disorders*

PUBLICATIONS

Genetic Risk for Hemochromatosis is Associated with Movement Disorders, *MedRxiv*, 2021 - ***Loughnan R***, Ahern J, Tompkins C, Palmer C, Sugrue L, Iversen J, Thompson W, Andreassen O, Jernigan T, Dale A, Boyle M, Fan C

Gene-experience correlation during cognitive development: Evidence from the Adolescent Brain Cognitive Development (ABCD) Study*, BioRxiv*, 2021 - ***Loughnan R***, Palmer C, Thompson W, Dale A, Jernigan T, Chieh Fan C

Unique prediction of developmental psychopathology from genetic and familial risk*, MedRxiv,* 2021 - ***Loughnan R***, Palmer C, Makowski C, Thompson W, Barch D, Jernigan T, Dale A, Fan C

Generalization of Cortical MOSTest Genome-Wide Associations Within and Across Samples*, BioRxiv*, 2021 - ***Loughnan R***, Shadrin A, Frei O, van der Meer D, Zhao, W, Palmer C, Thompson W, Makowski C, Jernigan T, Andreassen O, Chieh Fan C, Dale A

Distinct Regionalization Patterns of Cortical Morphology are Associated with Cognitive Performance Across Different Domains*, Cerebral Cortex*, 2021 - Palmer C, Zhao W, ***Loughnan***

R, Zou J, Fan C, Thompson W, Dale A, Jernigan T

Distinct Regionalization Patterns of Cortical Morphology are Associated with Cognitive Performance Across Different Domains*, Cerebral Cortex*, 2021 - Palmer C, Zhao W, ***Loughnan R***, Zou J, Fan C, Thompson W, Dale A, Jernigan T

Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology, *NeuroImage (in Review)*, 2021 - Shadrin A, Kaufmann T, van der Meer D, Palmer C, Makowski C***, Loughnan R***, Jernigan T, Seibert T, Hagler D, Smeland O, Chu, Y, Lin A, Chen W, Hindley G***,*** Thompson W, Roelfs D, Chieh Fan C , Holland D, Westlye L, Frei O, Andreassen O, Dale A

Genetic overlap between multivariate measures of human functional brain connectivity and psychiatric disorders, *Nature Genetics (in Review)*, 2021 – Roelfs D, van der Mer D, Alnæs D, Frei O, ***Loughnan R***, Chieh Fan C, Dale A, Andreassen O, Westlye L, Kaufmann T

The genetic architecture of human cortical folding*, Science Advances (in Review)* 2021 - van der Meer D, Kaufmann T, Shadrin A, Makowski C***,*** Frei O***,*** Roelfs D, Monereo Sanchez J, Linden D, Rokicki J, de Leeuw C, Thompson W, ***Loughnan R***, Chieh Fan C, Thompson P, Westlye L, Andreassen O, Dale A

Generalizing post-stroke prognoses from research data to clinical data*, Neuroimage Clinical*, *2019* - ***Loughnan R***, Lorca-Puls D, Lorca-Puls A, Gajardo-Vidal A, Espejo-Videla V, Gillebert C, Mantini D, Price C, Hope T

ABSTRACT OF THE DISSERTATION

UNDERSTANDING THE GENETIC INFLUENCE ON TRAITS RELATED TO BRAIN AND BEHAVIOR IN ADULTS AND CHILDREN

by

Robert J. Loughnan

Doctor of Philosophy in Cognitive Science

Professor Terry L. Jernigan, Chair
Professor Anders M. Dale, Co-Chair

Twin studies have established that the influence of genetics on human traits related to brain and behavior are pervasive. For a large majority of complex human traits uncovering which genetic variants are associated with phenotypic variations, by performing Genome Wide Association (GWA) studies, has been difficult due these traits being highly polygenic – many genetic variants with small effects that have a larger effect in aggregate. Conversely, some

traits have been shown to have a Mendelian genetic architecture – a single genetic variant imparting a large effect. In this thesis I explore the genetic contribution to variability of traits relating to brain and behavior in large GWA datasets for phenotypes of increasing complexity: a) Mendelian traits, b) polygenic traits and c) polygenic and multi-dimensional traits. First, I present analysis of the neurological impact of hereditary hemochromatosis, a Mendelian disorder that results in an excess of iron being absorbed by the body. Next, I present two projects investigating the genetic propensity/liability of i) cognitive performance and ii) psychopathology in a large sample of typically developing children aged 9-10 years old. Finally, I present a method for analyzing polygenic and multi-dimensional traits and apply it to the phenotype of human cortical morphology (cortical area, thickness and sulcal depth). In the age of large genomic databases this work may prove to be important for early detection of at risk groups as well as understanding the genetic determinants that give rise to complex human traits.

Chapter 1: *Overview*

Human genetics offers great promise from precision medicine to understanding the mechanisms of complex traits[1]. Since before the advent of genotyping technologies of the 21th century, it was understood that the genetic influence on human traits (e.g. height, intelligence, psychiatric disorders) was pervasive across phenotypes[2]. The degree to which the variability in a phenotype is attributable to genetics is defined as a phenotype's *heritability* and twin studies of the late 20th century enabled estimates of this quantity – termed twin heritability[3]. This type of heritability tells one the degree of contribution of genetics to a given trait, however it does not describe which genetic variants are contributing. Knowing which genetic variants contribute to a phenotype is important for understanding the mechanisms for that trait, and in the case that the trait is a disease or disorder this will help guide future therapies[4,5]. In order to locate which genetic variants are responsible for phenotypic variation the human genome needed to be sequenced and whole genome genotyping techniques need to be developed. The sequencing of the human genome in 2003[6], along with advances in whole genome genotyping (e.g. microarray[7]), enabled a type of study in the early 21th century known as Genome-Wide Association (GWA) studies. These studies enabled researchers to locate and quantify the association strength of genetic variants that contribute to variability in human phenotypes. Given the moderate to high heritability across many human traits[3] it was believed that GWA studies would enable us to quickly find which genetic variants were giving rise to variability in phenotypes[8]. However, aside from some notable early discoveries[9,10], a picture quickly began emerging that estimates of heritability from GWA studies were far lower than what was expected from twin studies. This divergence between i) heritabilities estimated

from GWA studies and ii) the expected heritability from twin studies was termed the 'missing heritability problem'[11]. Many explanations were given to explain this divergence such as epigenetics[12], epistatsis[13], and flaws in twin study design[14]. However, an overarching consensus that has been reached from the 'missing heritability problem' is that the vast majority of complex human traits are highly *polygenic* - they are determined by many small contributions dispersed across the genome that in aggregate explain some proportion of the overall heritability. As well as developing methods that can accurately model this polygenicity[15,16], the field of GWA studies has concluded that datasets of very large sample sizes will be needed to provide the statistical power to accurately estimate effects of single genetic variants. This has led to the construction of large biobanks with genotype and phenotype data on hundreds of thousands of individuals[17,18], as well as large consortia that aggregate subjects across studies[19,20]. The analysis of traits related to brain and behavior in these large datasets form the contents of this thesis. The chapters are structured in order of increasing genetic and phenotypic complexity and dimensionality. Chapter 2 deals with a phenotype determined by a single genetic variant – a *Mendelian* trait. Chapter 3 and 4 involve analysis of polygenic traits: i) measures of intelligence and ii) psychiatric disorders and mental health. Finally, Chapter 5 describes a methodological development for analyzing polygenic and multivariate traits – specifically applied to the genetics of human cortical morphology.

1.1 *Mendelian Traits*

Mendelian traits refer to traits that are determined by a single location in the genome, with eye color being the canonical example taught in school. GWA studies are best powered to detect discover variants for traits of this sort. Indeed one of the earliest landmark GWA studies was for age related macular degeneration[9], which although not strictly a Mendelian trait has a very low polygenicity (very few variants with large effects). Discovery of variants associated

with this trait was achieved in less than 200 people – a miniscule sample size by current GWA study standards. Mendelian traits also provide clear routes for therapeutic targets as the mechanism usually involves the disruption of a single gene. Examples of successful therapeutics for Mendelian disorders include elexacaftor/tezacaftor/ivacaftor for cystic fibrosis patients[4], Nusinersen for individuals with spinal muscular atrophy[5] and promising clinical trials of RNA therapies for Huntington's disease[21].

The specific Mendelian trait which I analyze in Chapter 2 is hereditary hemochromatosis (HH), an autosomal recessive genetic disorder that leads to iron overload in the body. It is understood the most common secondary disorders resulting from this condition are pathologies of the liver[22] caused by oxidative stress induced by iron overload. The neurological effects of this disorder have been studied in relatively small sample sizes or are case studies and have shown conflicting results[23–28]. In this portion of my thesis I conduct analysis of the genetic variant responsible for most cases of HH and neuroimaging measures that are sensitive to iron deposition, as well as enrichment for neurological disease associated with the neuroimaging findings. I conducted this analysis in a sample of 502,536 individuals from the UK Biobank – substantially larger than most other previous studies investigating HH and neurological deficits. The clinical benefit of studying this disorder is that there already exists treatment for HH that has been shown to be effective at reducing adverse secondary conditions if started early[22]. Furthermore, as a single gene is implicated in this disorder there is a good understanding of the biology and mechanisms resulting from the mutation.

1.2 *Polygenic Traits*

A major conclusion from the study of complex traits in human genetics is that the vast majority of phenotypes are polygenic in architecture[29]. This has motivated the wide adoption of a class of methods, known as Polygenic Risk Scores (PRS) or Polygenic Scores (PS), to predict liability/propensity of individual for a given polygenic trait. PS are usually generated in two stages: 1) training of the PS in a large cohort to learn the effects between genetic variants and the trait of interest – i.e. conducing a GWA study, 2) generating the PS in a (usually smaller) testing cohort by multiplying each genetic effect with the genotype of each individual and summing these across genetic variants. This process creates a single continuous score (PS) for each individual in the test set that describes the genetic liability/propensity for the trait of interest.

In Chapters 3 and 4 I generate PS's (i.e. step 2 above) in a sample of over 10,000 typically developing children aged 9-10 years old from the Adolescent Brain Cognitive Development (ABCD) Study – where step 1 was already performed in pervious published larger GWAS studies. The ABCD study contains a large battery of both cognitive and mental health assessments for each individual. In Chapter 3 I generate a PS for cognitive performance – termed Intelligence Polygenic Score (IPS) – and show how this is differentially associated with domains of cognitive performance measured for each child. Furthermore, I present analysis that some of the association between IPS and cognitive performance is mediated by the amount each child reads, possibly representing a mailable target for improving cognitive outcomes.

In Chapter 4 I present analysis after generating PRS for five major psychiatric traits in ABCD. We used these five PRS, as well as reported family history of ten traits, to capture liability for psychopathology. We sought to test how these 15 measures of liability related to

variability in a large battery of mental health assessments collected for each child. The importance of this work is in developing early detection for at risk individuals that may later go on to develop psychiatric disorders. This analysis was performed on the first time point of ABCD, when the children were aged 9-10 years old. With ABCD being a longitudinal study, Chapter 4 provides a baseline understanding of the relationship between psychopathology risk and mental health at this early age, from which we can understand evolving trajectories at later time points as the children develop.

1.3 *Polygenic Multi-dimensional traits*

Many of the techniques utilized and developed in the field of GWA studies are for single dimensional traits such as height[30], intelligence[31] or psychiatric diagnoses[20]. Cortical morphology (e.g. cortical area, depth and sulcal depth) represents a class of multi-dimensional traits that have importance for our understanding of the mind, as well as pathologies of the brain, and are known to be strongly determined by genetics[32,33]. Although, GWA studies have been conducted to investigate the genetics of cortical morphology[34], the techniques deployed are extensions to GWA methods for single dimensional traits that are not statistically well powered to detect effects of genetic variants that have effects that distributed across many regions of the cortex. This has motivated work in our research group to develop a GWA statistical test, known as the MOSTest (Multivariate Omnibus Statistical Test), that has greater power to detect associations distributed across many dimensions of a phenotypes (in this case cortical morphology)[35–37]. This preceding work is important for the *discovery* of genetic variants that are associated with variability of cortical morphology across many regions of the brain. In Chapter 5 I present a method and analysis for quantifying the *replication* performance of discovered genetic associations using MOSTest in an independent dataset – this method is referred to as a PolyVertex Score (PVS). This PVS has conceptual overlap with the method of

PRS, where PRS are aggregating the effects across multiple locations in the genome, PVS aggregates the effects across multiple locations (vertices) across the brain. The work in this chapter is important for validating the MOSTest method of discovery, as well as generating a framework for replicating GWA study findings of multivariate phenotypes of individuals from independent datasets. In the age of precision medicine these developments may prove to be important for individual level disease prediction[38].

1.4 *Kavli Project*

During the course of studies I was also fortunate enough to be the recipient of a Kavli Innovative Research Grant as a principal investigator to study “Imputing Molecular Endophenotypes from Genome Wide Association Data to Understand Biological Pathways Leading to Psychiatric Disorders”. For this work I spent time under the supervision of Trey Ideker to develop network genetics models for GWA datasets. The majority GWA studies focus on modelling linear and independent effects of genetic variants on phenotypes. For this project I worked on developing a computational model capturing how the interaction between genes can give rise to complex human phenotypes. This work build upon previous work in yeast[39] and in human cancer lines[40] in which neural network model was used with an architecture constrained by known network structures between genes from databases like Gene Ontology[41]. Constraining the neural network with known biology enables activations within a trained network to be interpreted as groups of genes that are working to form a higher order function. As a proof of principle I trained and applied this model to predict diabetes diagnoses and blood lipid levels from individual genetic data in a large biobank sample[17]. I learnt a great deal during this project about network biology, deep learning and effective computational practices, however I will not be presenting further details of this work in this thesis.

References

1 Ashley, E.A. (2016) Towards precision medicine. *Nat. Rev. Genet.* 17, 507–522

2 Eric Turkheimer (2000) Three laws of behavior genetics and what they mean. *Curr. Dir. Psychol. Sci.* 9, 160–164

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

4 Middleton, P.G. *et al.* (2019) Elexacaftor–Tezacaftor–Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. *N. Engl. J. Med.* 381, 1809–1819

5 Finkel, R.S. *et al.* (2017) Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* 377, 1723–1732

6 (2003) , International Consortium Completes Human Genome Project. . [Online]. Available: https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp

7 Schena, M. *et al.* (1995) Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science (80-.).* 270, 467 LP – 470

8 Bodmer, W.F. (1987) Human genetics: The molecular challenge. *BioEssays* 7, 41–45

9 Haines, J.L. *et al.* (2005) Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. *Science (80-.).* 308, 419 LP – 421

10 Polychronakos, C. and Li, Q. (2011) Understanding type 1 diabetes through genetics: Advances and prospects. *Nat. Rev. Genet.* 12, 781–792

11 Golan, D. *et al.* (2014) Measuring missing heritability: Inferring the contribution of common variants. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5272–E5281

12 Slatkin, M. (2009) Epigenetic inheritance and the missing heritability problem. *Genetics* 182, 845–850

13 Zuk, O. *et al.* (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1193–1198

14 Chaufan, C. and Joseph, J. (2013) The “Missing Heritability” of common disorders: Should health researchers care? *Int. J. Heal. Serv.* 43, 281–303

15 Yang, J. *et al.* (2011) GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82

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

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

18 All, T. *et al.* (2019) The “All of Us” Research Program. *N. Engl. J. Med.*

19 Pedersen, C.B. *et al.* (2018) The iPSYCH2012 case-cohort sample: New directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* 23, 6–14

20 Sullivan, P.F. *et al.* (2018) Psychiatric genomics: An update and an Agenda. *Am. J. Psychiatry* 175, 15–27

21 McColgan, P. and Tabrizi, S.J. (2018) Huntington’s disease: a clinical review. *Eur. J. Neurol.* 25, 24–34

22 Powell, L.W. *et al.* (2016) Haemochromatosis. *Lancet* 388, 706–716

23 Dekker, M.C.J. *et al.* (2003) Mutations in the hemochromatosis gene (HFE), Parkinson’s disease and parkinsonism. *Neurosci. Lett.* 348, 117–119

24 Guerreiro, R.J. *et al.* (2006) Association of HFE common mutations with Parkinson’s disease, Alzheimer’s disease and mild cognitive impairment in a Portuguese cohort. *BMC Neurol.* 6, 1–8

25 Aamodt, A.H. *et al.* (2007) Prevalence of haemochromatosis gene mutations in Parkinson’s disease. *J. Neurol. Neurosurg. Psychiatry* 78, 315–317

26 Akbas, N. *et al.* (2006) Screening for mutations of the HFE gene in Parkinson’s disease patients with hyperechogenicity of the substantia nigra. *Neurosci. Lett.* 407, 16–19

27 Buchanan, D.D. *et al.* (2002) The Cys282Tyr polymorphism in the HFE gene in Australian Parkinson’s disease patients. *Neurosci. Lett.* 327, 91–94

28 Kumar, N. *et al.* (2016) Movement disorders associated with hemochromatosis. *Can. J. Neurol. Sci.* 43, 801–808

29 Croucha, D.J.M. and Bodmer, W.F. (2020) Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proc. Natl. Acad. Sci. U. S. A.* 117, 18924–18933

30 Yengo, L. *et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649

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

32 Panizzon, M.S. *et al.* (2009) Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. *Cereb. Cortex* 19, 2728–2735

33 Eyler, L.T. *et al.* (2012) 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, 304–314

34 Grasby, K.L. *et al.* (2018) The genetic architecture of the human cerebral cortex.

35 van der Meer, D. *et al.* (2020) Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* 11, 3512

36 van der Meer, D. *et al.* (2021) The genetic architecture of human cortical folding. *bioRxiv* DOI: 10.1101/2021.01.13.426555

37 Shadrin, A.A. *et al.* (2021) Multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *bioRxiv* DOI: 10.1101/2020.10.22.350298

38 Denny, J.C. and Collins, F.S. (2021) Precision medicine in 2030—seven ways to transform healthcare. *Cell* 184, 1415–1419

39 Ma, J. *et al.* (2018) Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* 15, 290–298

40 Kuenzi, B.M. *et al.* (2020) Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell* 38, 672-684.e6

41 Ashburner, M. *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29

Chapter 2: Genetic Risk for Hemochromatosis is Associated with Movement Disorders

2.1 *Abstract*

Hereditary hemochromatosis (HH) is an autosomal recessive genetic disorder that can lead to iron overload, causing oxidative damage to affected organs. HH type 1 is predominantly associated with homozygosity for the mutation p.C282Y. Previous case studies have reported tentative links between HH and movement disorders, e.g. Parkinson's disease, and basal ganglia abnormalities on magnetic resonance imaging. We investigated the impact of p.C282Y homozygosity: on whole brain T2 intensity differences, a measure of iron deposition, and; on measures of movement abnormalities and disorders within UK Biobank. The neuroimaging analysis (154 p.C282Y homozygotes, 595 matched controls) showed that p.C282Y homozygosity was associated with decreased T2 signal intensity in motor circuits (basal ganglia, thalamus, red nucleus, and cerebellum; Cohen's d > 1) consistent with substantial iron deposition. Across the whole UK Biobank (2,889 p.C282Y homozygotes, 496,968 controls), we found a significant enrichment for movement abnormalities in male homozygotes (OR (95% CI) = 1.82 (1.27-2.61), p=0.001), but not females (OR (95% CI) = 1.10 (0.69-1.78), p=0.71). Among the 31 p.C282Y homozygote males with a movement disorder only 7 had a concurrent HH diagnosis. These findings indicate susceptibility to iron overload in subcortical structures in p.C282Y homozygotes, and confirmed an increased risk of movement

abnormalities and disorders in males. Given the effectiveness of early treatment in HH, screening for p.C282Y homozygosity in high risk individuals may offer a potential avenue to reduce iron accumulation in the brain and limit additional risk for the development of movement disorders among males.

2.2 *Introduction*

Hereditary hemochromatosis (HH) is a disorder that leads to iron overload in the body. HH type 1 is predominantly related to a HFE gene mutation, with 95% of cases being homozygote for p.C282Y (p.Cyst282Tyr) mutation[1]. The excess iron absorbed by the body leads to an accumulation of iron in organs, particularly in the liver leading to increased risk for liver disease and diabetes[2]. With a homozygosity rate of approximately 0.6% in northern European populations, it has been deemed the most prevalent genetic disorder in Europe[3,4]. The penetrance of HH and other associated diseases in p.C282Y homozygote individuals appears to be larger for males than females[5], leading researchers to believe that expelling excess iron through menstruation and pregnancy lowers disease burden and penetrance in females. The primary treatment for HH is phlebotomy which appears to be effective at reducing adverse clinical outcomes[3] if started early. As a result, some researchers have advocated for re-evaluating screening and early case ascertainment[2].

Although the impact of HH and p.C282Y homozygosity on the liver and heart is largely accepted[2,3], its effect on the central nervous system is still disputed. While some studies have reported a higher risk for Alzheimer's and Parkinson's disease in p.C282Y homozygote individuals[5,6], other studies have reported no risk[7,8] or a protective effect[9]. These conflicting results may be explained by modest sample sizes and pooled analysis across sexes. Case studies of HH individuals experiencing neurological deficits have confirmed neuroimaging abnormalities in the basal ganglia, substantia nigra, and cerebellum[10,11,12], all regions known to

have a substantial role in the control of movement[13]. However, these previous studies included only individuals diagnosed with HH, and therefore do not describe neurological deficits and abnormalities of p.C282Y homozygote individuals independent of HH diagnosis. A recent study performed an analysis of brain MRI data from 206 p.C282Y homozygote individuals taken from the UK Biobank[14]. Using analyses of $T2^*$ signal in predefined anatomical regions (Region Of Interest-based) they found indications of increased iron deposition in subcortical structures and cerebellum for p.C282Y homozygote individuals, and increased risk of dementia in p.C282Y homozygote males.

Since the implicated brain regions with p.C282Y homozygosity are known to have a strong involvement in motor functions[15] and reports of movement deficits in HH individuals[5,6,10,11], we investigated the relationship between p.C282Y homozygosity and movement disorders in 502,536 individuals from the UK Biobank, with improved imaging technology enabling greater granularity of associations across the brain. Specifically, we investigated: i) the impact of p.C282Y homozygosity on whole brain voxel-wise measures of iron deposits; ii) the sex stratified association of p.C282Y homozygosity with movement disorders and tested for overlap with HH diagnosis.

2.3 *Methods*

2.3.1 UK Biobank Sample

Genotypes, MRI scans, demographic and clinical data were obtained from the UK Biobank under accession number 27412, excluding participants who withdrew their consent. This resulted in a total sample of 502,536 individuals with a mean age of 57.0 years (standard deviation 8.1 years), 229,134 male. We used UK Biobank v3 imputed genotype data[30]. From

this sample, 2,889 individuals were identified as homozygote for p.C282Y (A/A at rs1800562), 1,293 male. As previous research does not indicate intermediate disease burden for p.C282Y heterozygotes[2], we coded control individuals as homozygote for no risk allele (G/G at rs1800562) or heterozygote (A/G at rs1800562) i.e., with a recessive model of inheritance. This resulted in 499,647 controls, 227,841 male.

2.3.2 Neuroimaging Analysis

2.3.2.1 ***Image acquisition***

T1 weighted and diffusion weighted scans were collected from three scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receiver head coils. For diffusion scans, multiple scans with no diffusion gradient were collected (b=0 s/mm^2) to fit diffusion models. The average of these b=0 scans was used as voxel-wise measures of T2 intensities. Diffusion weighted scans were collected using an SE-EPI sequence at 2mm isotropic resolution. T1 scans were collected using a 3D MPRAGE sequence at 1mm isotropic resolution.

2.3.2.2 ***Image Preprocessing***

Scans were corrected for nonlinear transformations provided by MRI scanner manufacturers[31,32], and T2 images were registered to T1 weighted images using mutual information[33]. Intensity inhomogeneity correction was performed by applying smoothly varying, estimated B1-bias field[34]. Images were rigidly registered and resampled into alignment with a pre-existing, in-house, averaged, reference brain with 1.0 mm isotropic resolution[34].

2.3.2.3 ***Atlas Registration***

To allow for voxel-wise analysis, subjects' imaging data were aligned using a multimodal nonlinear elastic registration algorithm. At the end of the preprocessing steps outlined in *Image Processing* and described in detail in Hagler et al. [34], subjects' structural images and diffusion parameter maps were aligned to an UK Biobank-specific atlas, using a custom diffeomorphic registration method.

2.3.2.4 ***Labelling regions of interest (ROI)***

Subcortical structures were labeled using Freesurfer 5.3[35]. Subjects' native space Freesurfer parcellations were warped to the atlas space and averaged across subjects. Additional subcortical nuclei, not available in the FreeSurfer segmentation, were labeled by registering readily available, downloadable, high spatial resolution atlases to our atlas space. The Pauli atlas was generated using T1 and T2 scans from 168 typical adults from the Human Connectome Project (HCP)[36]. The Najdenovska thalamic nuclei atlas was generated using a k-means algorithm taking as inputs mean fibre orientation density spherical harmonic coefficients from within a Freesurfer parcellation of the thalamus, using adult HCP data from 70 subjects[37]. All subcortical ROIs and abbreviations are listed in Supplementary Table 2.

2.3.2.5 ***Covariate Matched Controls***

From the full 2,889 p.C282Y homozygotes, only 154 had qualified imaging. As we did not want to have a large imbalance between the number of controls and p.C282Y homozygotes, we selected covariate matched controls at a ratio of 4:1 (controls:cases). We matched controls on age, sex, scanner, mean cortical area, and top ten components of genetics ancestry. These matched controls were generated through using the sample of 154

individuals with qualified imaging, fitting a logistic model to predict p.C282Y homozygosity from the listed covariates, and then generating propensity scores[38] for each individual. Controls were selected on having propensity scores with a threshold of $|s_1 - s_2| \leq threshold$, where s_1 and s_2 were the respective propensity scores of a p.C282Y homozygote and their matched control. The threshold was set to 1.5×10^{-4}. We performed this procedure by using the *pymatch* library (version 0.3.1) in python[39]. This resulted in 154 p.C282Y homozygotes (64 male) and 595 controls (248 male).

2.3.2.6 ***Statistical analysis***

General linear effect models were applied univariately to test the association between p.C282Y homozygosity and T2 intensities. Each voxel-wise T2 intensity was pre-residualized for age, sex, scanner, and top ten principal components of genetic ancestry. We then calculated Cohen's d effect sizes as the residualized voxel-wise differences between p.C282Y homozygotes and controls. As previous research has indicated a higher disease burden for p.C282Y homozygote males vs females, in supplementary results we additionally performed a sex-stratified analysis in 312 males (64 p.C282Y homozygote) and 437 females (90 p.C282Y homozygote).

2.3.3 Neurological Disease Burden Analysis

Given our neuroimaging findings of substantially lower T2 intensities of p.C282Y homozygotes in motor and gait circuits of the brain – indicative of iron deposition – we wanted to test if p.C282Y homozygosity imparted any risk for i) movement disorders, ii) gait disorders and iii) a broad category of neurological disorders. As completing imaging was not an inclusion criterion for this portion of the analysis, we included the entire sample listed above. We did not perform any covariate-matching of controls. Sex stratified logistic models fit 229,134 males

(1,293 p.C282Y homozygote) and 273,402 females (p.C282Y 1,596 homozygote) to predict diagnosis from p.C282Y homozygosity status controlling for age, sex, and top ten principal components of genetic ancestry. We fit three models for each sex to test the domains described above, predicting: movement disorders [ICD10: G20-G26], abnormalities of gait and mobility [ICD10: R26], and other disorders of the nervous system [ICD10: G90-99]. We fit an additional three models, two for the largest diagnoses of movement disorders; Parkinson's disease [ICD10: G20] and essential tremor [ICD10: G25], as well as a super-category of significantly associated diagnoses combining a diagnosis of either a) movement disorders or b) other disorders of the nervous system into a single outcome.

2.4 *Results*

Table 2.1 Sample size for each analysis.

	Neuroimaging Analysis	*Neurological Disease Analysis*
Sample Size	749 (312 males)	502,563 (229,134 males)
p.C282Y Homozygotes	154 (64 males)	2,889 (1,293 males)

2.4.1 Neuroimaging Analysis

We performed a voxel-wise analysis of T2 intensities (lower T2 intensity consistent with higher iron deposition[17]) of using p.C282Y homozygosity status as our predictor of interest – see Table 2.1 for sample sizes. We found evidence for iron deposition reflected in substantially lower T2 intensities localized to the basal ganglia, thalamus and cerebellum in p.C282Y homozygote individuals. p.C282Y homozygosity was associated with lower T2 intensities in bilateral caudate nucleus, putamen, thalamus (specifically the ventral anterior, ventral lateral

dorsal, ventral-lateral ventral and pulvinar nuclei), red nucleus, sub-thalamic nucleus, and both white and grey matter of the cerebellum (Figure 2.1). Additionally, we observed higher T2 intensities in the white matter of the superior cerebellar peduncle, possibly indicating gliosis in this region[18] which comprises the primary output pathway from the cerebellum to the thalamus and red nucleus. Supplementary sex-stratified analysis revealed similar effects in males and females, with p.C282Y homozygote females having on average approximately 30% smaller effects than males (Supplementary Figures 2.1 & 2.2).

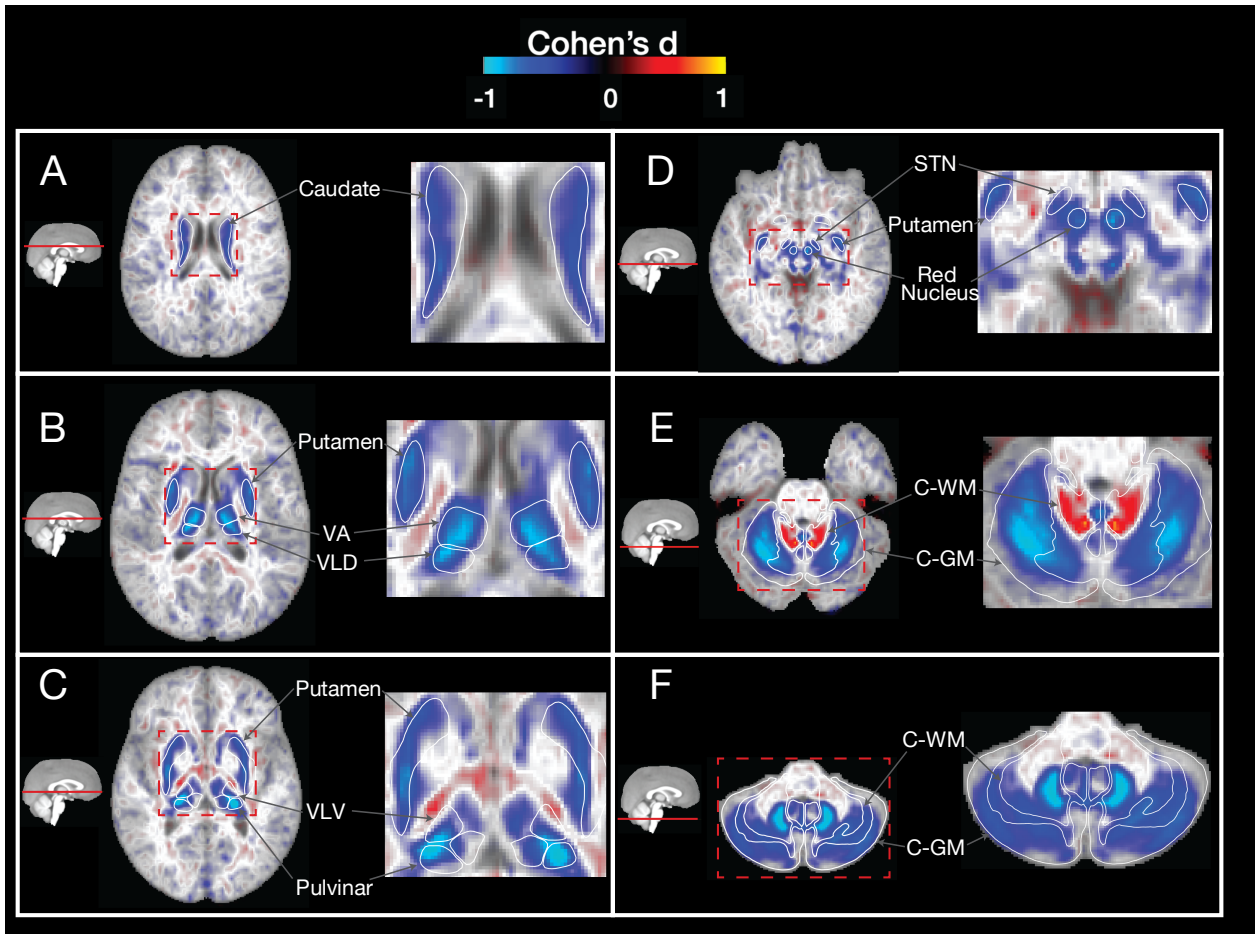


Figure 2.1 Voxelwise associations (cohen's d) of T2 intensities between matched controls vs p.C282Y homozygote individuals. Blue regions represent lower T2 intensities (indicating higher iron deposition) for p.C282Y homozygotes. Lower T2 intensities are observed for p.C282Y homozygotes in the A. caudate nucleus, B. putamen, ventral anterior and ventral lateral dorsal nuclei of the thalamus, C. ventral-lateral ventral and pulvinar nuclei of the thalamus, D. red nucleus, sub thalamic nucleus, E. and F. grey and white matter of the cerebellum. Higher T2 intensities are observed in E. in the superior cerebellar peduncle (primary output pathway connecting the cerebellum to the thalamus and red nucleus).

2.4.2 Neurological Disease Burden Analysis

The brain regions identified in our analysis have previously been implicated to have a strong involvement in motor control[15,19]. We aimed to determine if p.C282Y homozygote individuals had an enrichment for movement abnormalities, including movement diseases and gait/mobility, or other disorders of the nervous system. Table 2.1 displays the sample size for this analysis. As males appear to have a greater penetrance for HH and other associated diseases[2], we performed a sex-stratified analysis. We found that p.C282Y homozygote males had a higher chance of being diagnosed with a movement disorder (OR (95% CI) = 1.82 (1.27-2.61), p=0.001) and other disorders of the nervous system (OR (95% CI) = 1.51 (1.06-2.14), p=0.02). The International Classifications for Diseases (ICD) chapter of movement disorders includes Parkinson's disease and essential tremor, which were both associated with p.C282Y homozygosity in males (Parkinson's disease: OR (95% CI) = 1.78 (1.14-2.79), p=0.01 and essential tremor: OR (95% CI) = 1.92 (1.03-3.60), p=0.04). p.C282Y homozygote males did not have a higher chance of being diagnosed with gait or mobility disorders (OR (95% CI) = 0.96 (0.65-1.42), p=0.85). No significant associations were found for p.C282Y homozygote females for any diagnosis tested – see Figure 2.2 panel A and Supplementary Table 2.2.

Given the convergence of our neuroimaging and genetic associations on movement related circuits of the brain and movement disorders, p.C282Y homozygosity may lead to brain pathology that is undetected in sub-clinical HH cases. As the treatment for disorders like Parkinson's differs from that of hemochromatosis[3,20], we looked at the overlap of individuals with neurological diagnoses and a clinical diagnosis of hemochromatosis for p.C282Y homozygote males. We found that for p.C282Y homozygote males with a movement disorder diagnosis (31 individuals), most (24 individuals) did not have a concurrent HH diagnosis. We

also found a similar pattern for other nervous system disorders where 10 out of 33 men had a concurrent hemochromatosis diagnosis – see Figure 2.2 panel B.

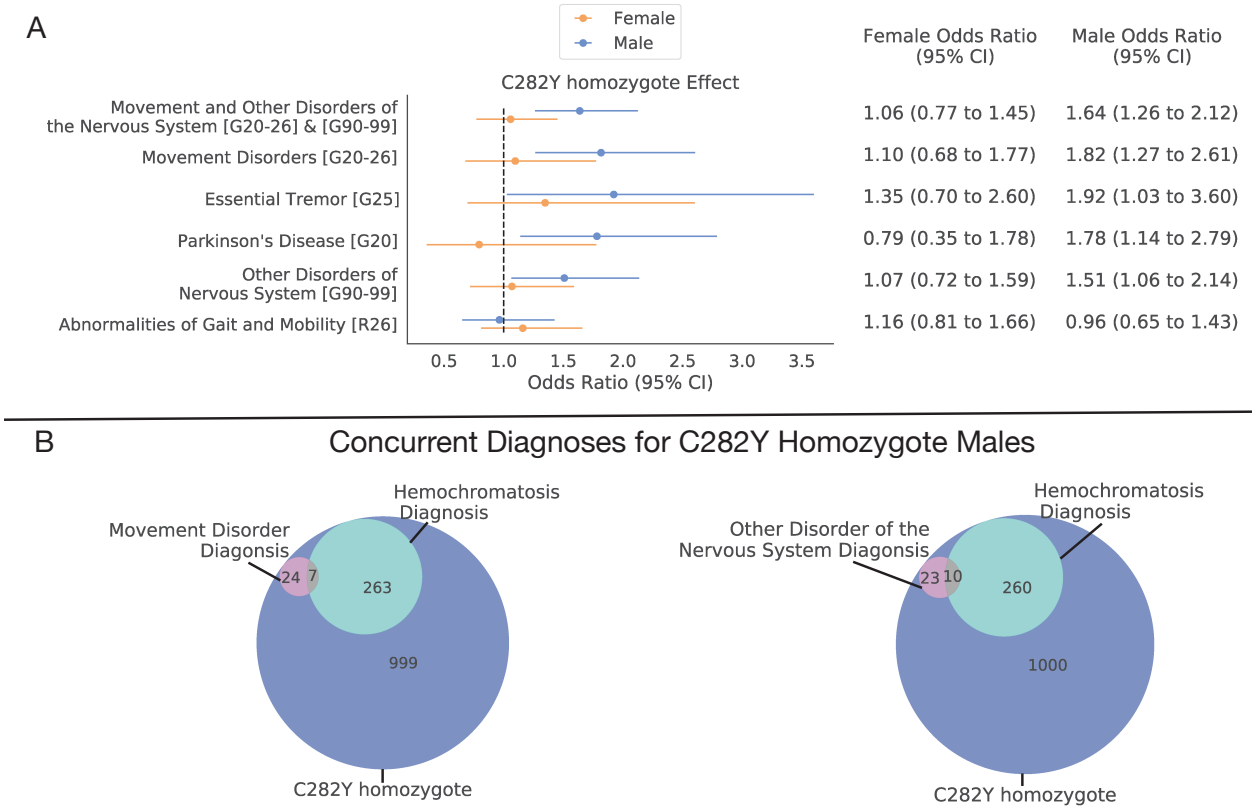


Figure 2.2 A. sex-stratified effect of C282Y homozygosity for neurological disorders (y axis), values in square brackets indicate ICD10 codes. Dotted vertical line indicates an odds ratio of 1 i.e., null effect. B. Venn diagrams indicating diagnosis overlap for C282Y homozygote males of hemochromatosis and i) movement disorders or ii) other disorders of the nervous system.

2.5 *Discussion*

We found the most prominent genetic risk factor for HH, p.C282Y homozygosity, was associated with substantially lower T2 intensities in brain regions related to motor control - consistent with iron deposition in these regions. Furthermore, p.C282Y homozygosity in males was associated with increased risk for movement-related disorders and other disorders of the nervous system, but not gait disorders. These results are consistent with both previous case reports of movement disorders in HH individuals[10,11], as well as a higher disease burden for

p.C282Y homozygote males vs. females[2,3]. Moreover, we found that most p.C282Y homozygote males did not have a concurrent hemochromatosis diagnosis when diagnosed with either i) a movement disorder or ii) other disorders of the nervous system. This is of importance given the difference in treatment for HH and movement disorders.

These results are consistent with a class of disorders termed 'Neurodegeneration with Brain Iron Accumulation' (NBIA) in which rare genetic mutations lead to iron deposition in the basal ganglia[21]. This iron deposition is believed to lead to oxidative damage of these brain regions impairing their function and resulting in movement deficits. Previous conflicting results regarding neurological manifestations of p.C282Y homozygosity[5,6,7,8,9] have meant that hemochromatosis has traditionally not been included as a cause of NBIA. These conflicting results are probably due to small sample sizes, no stratification based on sex, and biased subject ascertainment. We believe that our study addresses these issues by conducting disease associations in a sample 500 times larger than the previously listed studies, performing sex-stratified analysis, and selecting individuals on genotype, not disease status. Furthermore, our neuroimaging results provides strong support that p.C282Y homozygosity imparts a large, selective effect on the brain's motor circuits. The results presented here suggest revisiting p.C282Y homozygosity as a form of NBIA, albeit with reduced penetrance.

The globus pallidus is a region that appears to show large amounts of iron deposition in other NBIA disorders[21,22], with lower T2 intensities except in juvenile forms ('eye of the tiger'). Additionally, the previous study investigating the p.C282Y homozygosity effect on $T2^*$ intensities in the same sample as our study found a small to moderate effect in the pallidum[14]. However, our analysis did not reveal differences in T2 intensities in the globus pallidus (Figure 2.1). The lack of association in our analysis could be due to i) a genuine lack of iron deposition in the pallidum when compared other NBIA disorders or ii) a result of biological processes having opposing effects on T2 intensity (e.g. iron deposition decreasing intensities and gliosis

increasing intensities) averaging out. Differing sensitivities of T2-weighted vs T2*-weighted scans to minerals[23] likely explain the divergence in results of the pallidum between the previous study[14] and ours. Further research should be conducted to understand processes occurring in the pallidum of p.C282Y homozygotes and to determine if these truly do differ from other NBIA disorders.

Post mortem samples and in-vivo imaging of individuals diagnosed with Parkinson's display iron deposition in the brain regions we identified[24]. The most recent genome-wide association study (GWAS) of Parkinson's disease in males did not identify the variant at position p.C282Y as a risk factor ($p=0.16$)[25]. We hypothesize this is likely due to the study employing a GWAS standard additive model of inheritance, in which an additional copy of a risk allele imparts a dose-dependent risk. Consistent with findings from previous literature of HH[2], here we tested a recessive model of inheritance and observed a relatively sizeable increased risk for movement-related disorders in general (OR=1.82) and Parkinson's disease in particular (OR=1.78) – this is in the $5x10^{-6}$ percentile of effect sizes from the most significant previous GWAS of Parkinson's disease in males[26]. Further work needs to be done to verify the size and confidence of this effect in other ancestry groups and populations, particularly given the non-uniform distribution of p.C282Y across the globe (see Supplementary Figure 2.3).

Additionally, although we observe largest T2 intensity reductions (compatible with iron accumulation) in male p.C282Y homozygotes, we still observe a reduction in females and a non-significant increase risk for movement disorders. If, as others have argued[3], menstruation expels iron from the body, we may be observing the post-menopausal accumulation of iron that has not had enough time to impart the toxic effects compared to males. Indeed, the average age of the women in our neuroimaging study was 64 years old. Furthermore, estrogen may be playing an antioxidant role that is moderating the damaging effects of iron overload[27].

Although our neuroimaging results are consistent with iron accumulation in associated regions of the brain, lower T2 intensities may also indicate calcification. Indeed a recent case study reported an HH individual who displayed both iron and calcium accumulation for the same brain regions discovered in our analysis[28]. One possible explanation for this observed calcium deposition is that calcium and iron homeostasis are reciprocally connected within these regions. This hypothesis is supported by work demonstrating that iron, as Fe2+ ions, can lead to build of inorganic pyrophosphate with calcium[29]. Additional brain imaging of HH and p.C282Y homozygote individuals, using different imaging modalities, may further elucidate the mineral composition of the abnormalities we have observed.

The convergent evidence we have shown consistent with iron deposition in motor circuits of the brain and enrichment for movement and neurological disorders, in those who are homozygous for p.C282Y suggests considering p.C282Y homozygosity as a risk factor for movement disorders in males. Furthermore, given that early treatment of HH has been shown to be effective in preventing the negative health manifestations of the disease outside of the nervous system[3], our findings provide additional support for revisiting the public health implications of early screening of populations in which this variant is particularly prevalent.

Funding: R.L was supported by Kavli Innovative Research Grant under award number 2019-1624. C.F. was supported by grant R01MH122688 and RF1MH120025 funded by the National Institute for Mental Health (NIMH).

Conflict of Interest: A.M.D. reports that he was a Founder of and holds equity in CorTechs Labs, Inc., and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. He receives funding through research grants from GE Healthcare to UCSD. The terms of these arrangements have been reviewed by and approved

by UCSD in accordance with its conflict-of-interest policies. No other authors report a conflict of interest.

Acknowledgment

Chapter 2, in full, is available as on MedRxiv and has been submitted for publication. Loughnan R, Ahern J, Tompkins C, Palmer C, Sugrue L, Iversen J, Thompson W, Andreassen O, Jernigan T, Dale A, Boyle M, Fan C. The dissertation/thesis author was the primary investigator and author of this paper.

References

1. Adams, P. C. Epidemiology and diagnostic testing for hemochromatosis and iron overload. *Int. J. Lab. Hematol.* ***37***, 25–30 (2015).

2. Pilling, L. C. *et al.* Common conditions associated with hereditary haemochromatosis genetic variants: Cohort study in UK Biobank. *BMJ* ***364***, (2019).

3. Powell, L. W., Seckington, R. C. & Deugnier, Y. Haemochromatosis. *Lancet* ***388***, 706–716 (2016).

4. Bomford, A. Genetics of haemochromatosis. *Lancet* ***360***, 1673–1681 (2002).

5. Dekker, M. C. J. *et al.* Mutations in the hemochromatosis gene (HFE), Parkinson's disease and parkinsonism. *Neurosci. Lett.* ***348***, 117–119 (2003).

6. Guerreiro, R. J. *et al.* Association of HFE common mutations with Parkinson's disease, Alzheimer's disease and mild cognitive impairment in a Portuguese cohort. *BMC Neurol.* ***6***, 1–8 (2006).

7. Aamodt, A. H. *et al.* Prevalence of haemochromatosis gene mutations in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* ***78***, 315–317 (2007).

8. Akbas, N. *et al.* Screening for mutations of the HFE gene in Parkinson's disease patients with hyperechogenicity of the substantia nigra. *Neurosci. Lett.* ***407***, 16–19 (2006).

9. Buchanan, D. D., Silburn, P. A., Chalk, J. B., Le Couteur, D. G. & Mellick, G. D. The Cys282Tyr polymorphism in the HFE gene in Australian Parkinson's disease patients. *Neurosci. Lett.* ***327***, 91–94 (2002).

10. Kumar, N., Rizek, P., Sadikovic, B., Adams, P. C. & Jog, M. Movement disorders associated with hemochromatosis. *Can. J. Neurol. Sci.* ***43***, 801–808 (2016).

11. Nielsen, J. E., Jensen, L. N. & Krabbe, K. Hereditary haemochromatosis: A case of iron accumulation in the basal ganglia associated with a parkinsonian syndrome. *J. Neurol. Neurosurg. Psychiatry* ***59***, 318–321 (1995).

12. Berg, D. *et al.* The basal ganglia in haemochromatosis. *Neuroradiology* ***42***, 9–13 (2000).

13. Burn, D. *Oxford Textbook of Movement Disorders*. (Oxford University Press, 2013). doi:10.1093/med/9780199609536.001.0001

14. Atkins, J. L. *et al.* Hemochromatosis mutations, brain iron imaging, and dementia in the UK Biobank cohort. *J. Alzheimer's Dis.* ***79***, 1203–1211 (2021).

15. Burn, D. *Oxford Textbook of Movement Disorders*. (Oxford University Press, 2013). doi:10.1093/med/9780199609536.001.0001

16. Herrero, M. T., Barcia, C. & Navarro, J. M. Functional anatomy of thalamus and basal ganglia. *Child's Nerv. Syst.* ***18***, 386–404 (2002).

17. Gandon, Y. *et al.* Non-invasive assessment of hepatic iron stores by MRI. *Lancet* ***363***, 357–362 (2004).

18. McComb, R. D. Neuropathology: A Volume in the Foundations of Diagnostic Pathology Series. *Am. J. Surg. Pathol.* ***30***, (2006).

19. Takakusaki, K. Functional Neuroanatomy for Posture and Gait Control. *J. Mov. Disord.* ***10***, 1–17 (2017).

20. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Prim.* ***3***, 1–21 (2017).

21. Hayflick, S. J., Kurian, M. A. & Hogarth, P. Chapter 19 - Neurodegeneration with brain iron accumulation. in *Neurogenetics, Part I* (eds. Geschwind, D. H., Paulson, H. L. & Klein, C. B. T.-H. of C. N.) ***147***, 293–305 (Elsevier, 2018).

22. Kruer, M. C. *et al.* Neuroimaging features of neurodegeneration with brain iron accumulation. *Am. J. Neuroradiol.* ***33***, 407–414 (2012).

23. Labranche, R. *et al.* Liver iron quantification with MR imaging: A primer for radiologists. *Radiographics* ***38***, 392–412 (2018).

24. Wang, J. Y. *et al.* Meta-analysis of brain iron levels of Parkinson's disease patients determined by postmortem and MRI measurements. *Sci. Rep.* ***6***, 1–13 (2016).

25. Blauwendraat, C. *et al.* Investigation of Autosomal Genetic Sex Differences in Parkinson's disease. *medRxiv* 2021.02.09.21250262 (2021). doi:10.1101/2021.02.09.21250262

26. Blauwendraat, C. *et al.* Investigation of Autosomal Genetic Sex Differences in Parkinson's disease. *medRxiv* 2021.02.09.21250262 (2021).

doi:10.1101/2021.02.09.21250262

27. Zárate, S., Stevnsner, T. & Gredilla, R. Role of estrogen and other sex hormones in brain aging. Neuroprotection and DNA repair. *Front. Aging Neurosci.* ***9***, 1–22 (2017).

28. Scarlini, S. *et al.* Idiopathic brain calcification in a patient with hereditary hemochromatosis. *BMC Neurol.* ***20***, 1–5 (2020).

29. McCarty, D. J. & Pepe, P. F. Erythrocyte neutral inorganic pyrophosphatase in pseudogout. *J. Lab. Clin. Med.* ***79***, 277–284 (1972).

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

31. Jovicich, J. *et al.* Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* ***30***, 436–443 (2006).

32. Wald, L., Schmitt, F. & Dale, A. Systematic spatial distortion in MRI due to gradient non-linearities. *Neuroimage* ***13***, 50 (2001).

33. Wells, W. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* ***1***, 35–51 (1996).

34. Hagler, D. J. *et al.* Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* ***202***, (2019).

35. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* ***33***, 341–355 (2002).

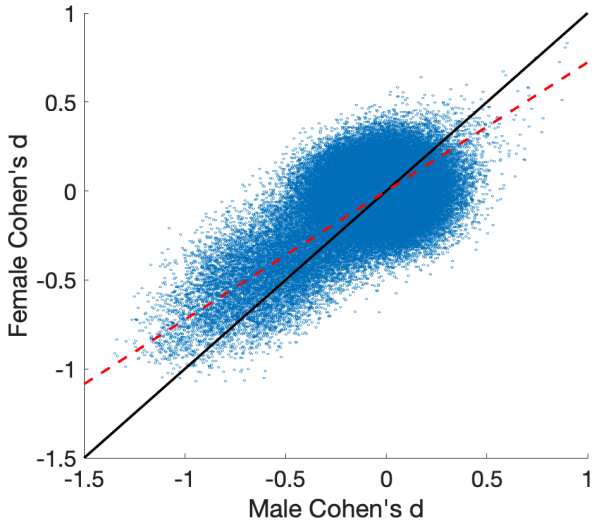
36. Pauli, W. M., Nili, A. N. & Michael Tyszka, J. Data Descriptor: A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* ***5***, 1–13 (2018).

37. Najdenovska, E. *et al.* In-vivo probabilistic atlas of human thalamic nuclei based on diffusion-weighted magnetic resonance imaging. *Sci. Data* ***5***, 1–11 (2018).

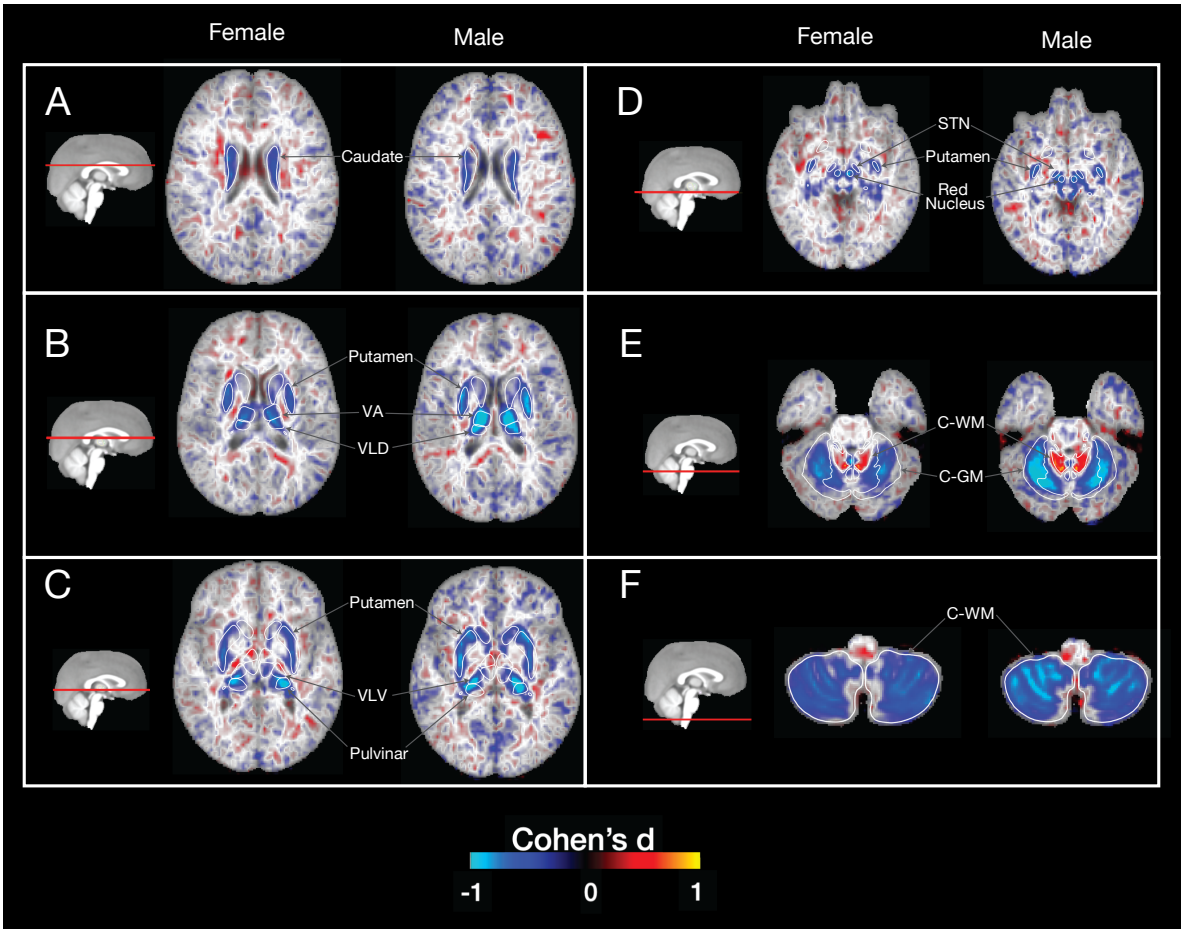
38. Rosenbaum, P. R. & Rubin, D. B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am. Stat.* ***39***, 33–38 (1985).

39. Benmiroglio. pymatch. Available at: https://github.com/benmiroglio/pymatch.

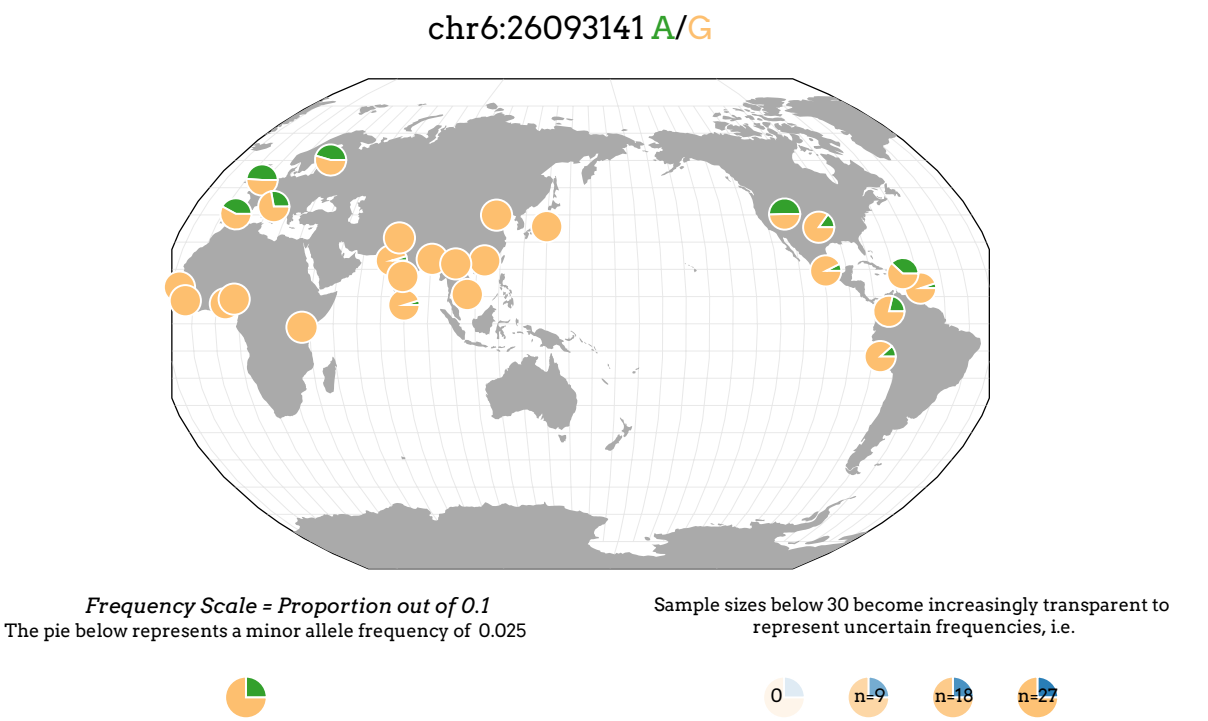
Appendix



Supplementary Figure 2.1 Sex stratified associations of p.C282Y homozygosity and T2 voxel intensities (Cohen's d), as scatter plot – each point represents a single voxel. Black line indicates y=x (i.e. equal effect sizes). Red line indicates best fitting line (for points of at least moderate effects $x^2 + y^2 > 0.2$), $\beta = 0.72$ indicating approximately a 28% reduction in T2 intensity brain associations for females compared to males.



Supplementary Figure 2.2 Maps of sex stratified associations of p.C282Y homozygosity and T2 voxel intensities (Cohen's d). Overall effects are larger in males vs females, with largest effects in males observed in the VA, VLD and pulvinar nuclei of the thalamus (B) as well as in the C-WM and C-GM (E). Abbreviations: VA – ventral anterior, VLD – ventral anterior dorsal, VLV – ventral anterior ventral, C-WM – cerebellum white matter, C-GM – cerebellum GM.



Supplementary Figure 2.3 Global distribution of p.C282Y (rs1800562). Allele frequencies are highest in Europe (particularly northern Europe) and parts of north America. Map generated from the Geography of Genetic Variants Browser[1].

Supplementary Table 2.1 ***Regions of interest (ROIs) labelled using 3 different methods.*** Column 1) automatic segmentation using FreeSurfer 5.3 applied to each subject's T1 image in atlas space[2]; Column 2) registration of the Pauli atlas of subcortical nuclei to the multispectral atlas[3]; Column 3) registration of the the Najdenovska thalamic nuclei atlas to our data[4].

FreeSurfer 5.3 segmentation T1	***Pauli, 2018 HCP T1 & T2***	***Najdenovska, 2018 HCP FODs***
Amygdala (Amg)	Putamen (Pu)	Anterior (tA)
Hippocampus (Hipp)	Caudate (Ca)	Ventral anterior (tVA)
Putamen (Pu)	Nucleus accumbens (NAcc)	Mediodorsal (tMD)
Caudate (Ca)	Extended amygdala (EA)	Ventral-latero-ventral (tVLV)
Globus pallidus (GP)	Substantia nigra pars compacta (SNpc)	Ventral-latero-dorsal (tVLD)
Accumbens area (NAcc)	Substantia nigra pars reticulata (SNpr)	Central-latero-lateral-posterior-medial-pulvinar (tC)
Thalamus (Thal)	Red nucleus (RN)	Pulvinar (tP)
Ventral Diencephalon (VDC)	Parabrachial pigmented nucleus (PBP)	
	Hypothalamus (Hyp)	
	Mamillary nucleus (MN)	
	Subthalamic nucleus (STN)	

Supplementary Table 2.2 Regression tables results for associating C282Y homozygote (+/+) cases with diagnoses of different neurological disorders. Models were fit separately for males and females.

Males

Diagnosis	Total Controls	Total Cases	C282Y Cases	OR	CI_lower	CI_upper	P
Abnormalities of Gait and Mobility [R26]	219012	4557	26	0.964119	0.65175	1.4262	0.854872
Other Disorders of Nervous System [G90-99]	219737	3832	33	1.5081	1.06499	2.13558	0.02063
Parkinson's Disease [G20]	221638	1931	20	1.78188	1.13941	2.78661	0.011341
Essential Tremor [G25]	222688	881	10	1.92304	1.02705	3.6007	0.0410158
Movement Disorders [G20-26]	220636	2933	31	1.8162	1.26605	2.60542	0.00118983
Movement and Other Disorders of the Nervous Sy...	217030	6539	61	1.63773	1.26297	2.12371	0.00019852

Females

Diagnosis	Total Controls	Total Cases	C282Y Cases	OR	CI_lower	CI_upper	P
Abnormalities of Gait and Mobility [R26]	260493	4226	31	1.16012	0.810591	1.66038	0.41681
Other Disorders of Nervous System [G90-99]	260857	3862	25	1.06932	0.719118	1.59005	0.740586
Parkinson's Disease [G20]	263551	1168	6	0.79401	0.354918	1.77633	0.574492
Essential Tremor [G25]	263662	1057	9	1.34698	0.696958	2.60323	0.375601
Movement Disorders [G20-26]	262280	2439	17	1.09684	0.678124	1.77411	0.706349
Movement and Other Disorders of the Nervous Sy...	258589	6130	40	1.05775	0.771525	1.45016	0.727282

References

1. Marcus, J. H. & Novembre, J. Visualizing the geography of genetic variants. *Bioinformatics* ***33***, 594–595 (2017).

2. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* ***33***, 341–355 (2002).

3. Pauli, W. M., Nili, A. N. & Michael Tyszka, J. Data Descriptor: A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* ***5***, 1–13 (2018).

4. Najdenovska, E. *et al.* In-vivo probabilistic atlas of human thalamic nuclei based on diffusion-weighted magnetic resonance imaging. *Sci. Data* ***5***, 1–11 (2018).

Chapter 3: *Gene-experience correlation during cognitive development: Evidence from the Adolescent Brain Cognitive Development (ABCD) Study℠*

3.1 *Abstract*

Findings in adults have shown more culturally sensitive “crystallized” measures of intelligence have greater heritability, these results were not able to be shown in children. In the current study we aimed to test if the genetic propensity for high cognitive performance in children had a larger association with more culturally sensitive “crystallized” measures of intelligence. Furthermore, if such a relationship was found we sought to test if this could be partially explained by a larger statistical mediating effect of reading. With data from 8,518 participants, aged 9 to 11, from the Adolescent Brain Cognitive Development (ABCD) Study®, we used polygenic predictors of intelligence test performance (based on genome-wide association meta-analyses of data from 269,867 individuals) and of educational attainment (based on data from 1.1 million individuals), associating these predictors with neurocognitive performance. We then assessed the extent of mediation of these associations by a measure of recreational reading. More culturally sensitive ‘crystallized’ measures were more strongly associated with the polygenic predictors than were less culturally sensitive ‘fluid’ measures. This mirrored heritability differences reported previously in adults and suggests similar

associations in children. Recreational reading more strongly statistically mediated the genetic associations with crystallized than those with fluid measures of cognition. This is consistent with a prominent role of gene-environment correlation in cognitive development measured by “crystallized” intelligence tests. Such experiential mediators may represent malleable targets for improving cognitive outcomes.

3.2 *Introduction*

Scores on cognitive tests in both children and adults have been linked to long term outcomes and to genetic variation(1–4). Some cognitive tests, e.g., those requiring literacy and mathematical skills, depend upon and are more sensitive to variability in cultural and socio-economic factors. These measures are often referred to as ‘crystallized’ intelligence measures. In contrast, other tests that tap the capacity to solve novel problems, or process novel information, often referred to as ‘fluid’ measures, are less culturally sensitive and are less strongly related to socio-economic variables(5,6). A recent review reported systematic differences in heritability (an estimate of trait variability attributable to genetic variation) of the traits measured by these different kinds of cognitive measures(7). Surprisingly, in studies of adult twins, more culturally sensitive tests exhibited higher, rather than lower, heritability; which runs counter to predictions from conventional models of intelligence. The authors described similar trends in the twin studies of children, but increased heritability of crystallized relative to fluid measures have not yet been established for children, in whom intellectual functions are continuing to mature.

The finding that the measures most strongly influenced by cultural factors exhibit higher heritability is perhaps counterintuitive; however previous authors have noted that genetic variation can be associated with environmental, cultural, or experiential (ECE) factors that themselves amplify effects of a genotype on the phenotype, a phenomenon often referred to as

rGE (gene-environment correlation). These associations between genotypes and ECE factors could influence the development of cognitive and intellectual abilities in several ways. As an example, if others in the social environments of children recognize traits, e.g., precocious behavior, in those with a genetic propensity for a given cognitive ability, they may begin to treat such individuals differently, rewarding them disproportionately for intellectual pursuits, investing more in their instruction, and/or placing them in environments that drive learning more effectively. Alternatively, the associations can be driven by the motivation of the children themselves if for example they develop greater enthusiasm for intellectual activities for which they have been more frequently rewarded, and which they then pursue more assiduously, thus enjoying beneficial effects of the increased practice associated with these activities. In either case, the genetically advantaged abilities are disproportionately enhanced by these mediating ECE factors. Of course, individuals with less advantageous genotypes may experience the converse of these social and motivational effects, resulting in languishing, or in the worst case suppressed, intellectual development, even within similar environments. Such rGE effects can increase variance in intellectual phenotypes and increase estimates of heritability using both epidemiological and genomic methods(8). The important implication is that a component of this increased heritability requires the mediating ECE effects for its expression. In essence, more direct biological effects of the genotype *and* associated differences in the environments or experiences of the child are both contributing causal factors influencing the mature phenotype, but they act through dissociable mechanisms.

Heritability is a population statistic frequently measured using a twin design. For this study, we used polygenic scores to examine variation in genetic and experiential factors and their relationship to trait measures of cognitive function. Polygenic scores have the advantage that they can be used to index relevant genetic factors in samples of unrelated individuals by leveraging the statistical power of meta-analysis results from large Genome Wide Association

Studies (GWAS). Using neurocognitive test scores, genomic data, and a measure of parent-reported recreational reading assessed in a large sample of 8,618 children, aged 9 to11, from the ABCD Study®, we used polygenic scores of intelligence test performance (based on GWAS of 269,867 individuals(9)) and educational attainment, sometimes considered a proxy for intellectual ability (based on 1.1 million individuals(10)), to ask 3 questions: First, do these genomic predictors account for more of the variability in estimates of culturally sensitive crystallized traits than fluid traits in children, as might be expected from reports of higher heritability in adult twins? Second, does a parent-reported estimate of the time their children spend reading for pleasure mediate the relationship between a genomic predictor and measures of cognitive performance, consistent with a role of this experiential enhancer of performance in increasing heritability? Third, if mediation is observed, is this mediating effect larger for the culturally sensitive crystallized than the fluid measures of cognitive performance, consistent with a role for rGE in the higher heritability of these measures?

In additional analyses, we examined the degree to which the findings in the ethnically diverse ABCD sample were similar between the subgroup of children with high genomic European ancestry (EurA) and a remaining subgroup of children who were from diverse ancestry groups (DivA). Finally, using simulations, we tested whether our observed findings may be due to previously reported differences in test-retest reliabilities (for crystallized vs fluid measures).

3.3 *Materials and Methods*

3.3.1 Data available in the ABCD data release 2.0.1

The ABCD study (http://abcdstudy.org) enrolled the families of 11,875 children aged 9 or 10 years at baseline(11). This longitudinal study follows the development of these children at 21 sites across the US for ten years. The cohort exhibits a large degree of sociodemographic diversity. Exclusion criteria were limited to: 1) lack of English proficiency; 2) the presence of severe sensory, neurological, medical or intellectual limitations that would inhibit the child's ability to comply with the protocol; 3) an inability to complete an MRI scan at baseline. The study protocols are approved by the University of California San Diego Institutional Review Board(12). Parent/caregiver permission and child assent from each participant were obtained. Here, our data were drawn from the baseline assessments shared in ABCD release 2.0.1 (NDAR DOI: 10.15154/1504041).

3.3.2 Cognitive Measures

Seven of the 10 cognitive tasks were subtests from The NIH Toolbox Cognition Battery® (NTCB) in the version recommended for ages 7+(http://www.nihtoolbox.org)(13). The average time to complete this battery is approximately 35 minutes. The NTCB was administered in English(14), using an iPad, with support from a research assistant when needed. The battery yields individual test scores measuring specific constructs and composite scores that have been shown to be highly correlated with 'gold standard' measures of intelligence in adults(15) and children(5). Here, all 7 individual test scores and 2 composite scores were examined: the Crystallized Cognition Composite Score (derived from scores on the Picture Vocabulary and Oral Reading Recognition measures) and the Fluid Cognition Composite Score (derived from

the remaining measures). Additionally, three neurocognitive tasks were used that were not components of the NTCB: Rey-Auditory Learning Task, Little Man Task and Matrix Reasoning. Please see supplementary materials for a description of each task.

3.3.3 Latent Neurocognitive Factors

Thompson et al. derived a three factor, varimax rotated, solution for the latent structure across the neurocognitive battery in ABCD using Bayesian Probabilistic PCA(16). The final latent factor solution included the measures described above, except for the Matrix Reasoning task which had very little effect on the solution. The factors will be referred to as Bayesian Factors (BF) 1-3. Language tasks loaded most heavily on BF1, which was highly correlated with the Crystallized Composite (r=0.93); executive functioning tasks loaded most heavily on BF2; and learning/memory tasks loaded heavily on BF3.

3.3.4 Recreational Reading

Parents of ABCD participants were asked to complete a survey of their children's activities. One question asked, "Does your child read for pleasure?" The follow-up question was, "About how many hours per week does your child read for pleasure?". This estimate of number of hours of recreational reading was log transformed due to skewness. To confine the analyses of this variable to a homogenous group of children who read for pleasure, we included only children whose parents answered 'yes' to the first question.

2.1.4 Genetic Data and Computing Polygenic Scores

Using genotype data we derived genetic ancestry using fastStructure(17) with four ancestry groups. Genetic principal components were also calculated using PLINK. Variants were imputed using the Michigan Imputation Server(18). Polygenic scores were computed using PRSice(19). The Intelligence Polygenic Score (IPS) was trained on 269,867 individuals by Savage et al.(9), and focused on neurocognitive tests considered to gauge fluid intelligence. The Education Attainment Polygenic Score (EAPS) was generated from 1.1 million individuals, predicting the phenotype of number of years of schooling completed. Please see supplementary materials for further details on genetic data and analysis.

We were primarily focused on studying the IPS association with cognitive tests in ABCD, due to it being trained on a more directly relevant phenotype. However, we additionally examined EAPS as a secondary analysis for comparison as it has been previously used as a proxy for cognitive ability and has a discovery sample size four times the size of the IPS.

3.3.5 Analytic Methods

3.3.5.1 ***Ancestry Group Analyses***

Training and testing polygenic scores in different ancestry groups has been shown to reduce predictive power(20–22). Given the ancestry differences between the polygenic score discovery samples (predominantly European) and the ABCD study (multiple ancestry groups), we wanted to confirm our main results in the full samples were not driven by population structure. As such we additionally performed analyses in two subsamples: 1) children with estimated proportion of European ancestry higher than 90% (EurA) and 2) a group of the remaining children with diverse ancestry, which included those from other or mixed ancestry (DivA).

3.3.5.2 ***Statistical Model for Genomic Prediction of Behavioral Measures***

To assess the association between the polygenic scores and cognitive performance in ABCD, we fit Generalized Linear Mixed-Effects Models (GLMMs) using the gamm4 package(23) in R. Each model predicted performance on a different cognitive measure. Continuous variables were z-scored before model fitting to allow coefficients to be interpreted as standardized effect sizes. To test if regression coefficients differed between regressions we performed a z-test on the difference between coefficients, based on the propagated standard error for the two regression coefficients as the sum of the error of variances for each measure. This test assumes the standard errors are uncorrelated and so provides a conservative estimate of significance. Please see supplementary materials for details and covariates used.

3.3.5.3 ***Differential Mediation Analysis***

To assess whether recreational reading is a plausible ECE factor increasing heritability of crystallized cognition, through rGE effects, we performed a mediation analysis. Specifically, we compared the statistical mediation effects of recreational reading experience on the associations between the IPS and both the Crystallized Composite and Fluid Composite, respectively. We achieved this by calculating the proportion of mediation of recreational reading on the i) IPS-crystallized and the ii) IPS-fluid associations using an average causal mediation effect model(24) across 10,000 bootstrap samples. With bootstrapped samples we tested if the mediation effect of recreational reading on the IPS-crystallized association was greater than that of IPS-fluid , by performing a Welch t-test on the samples. Mediation analysis was performed using general linear models in the mediation package in R(25), see supplementary materials for details.

3.4 *Results*

3.4.1 Demographics

Figure *3.1* illustrates a flow-chart for sample selection. For the final analysis we have 8,518 individuals in the full sample, 4,885 in the EurA sample and 3,633 in the DivA sample.

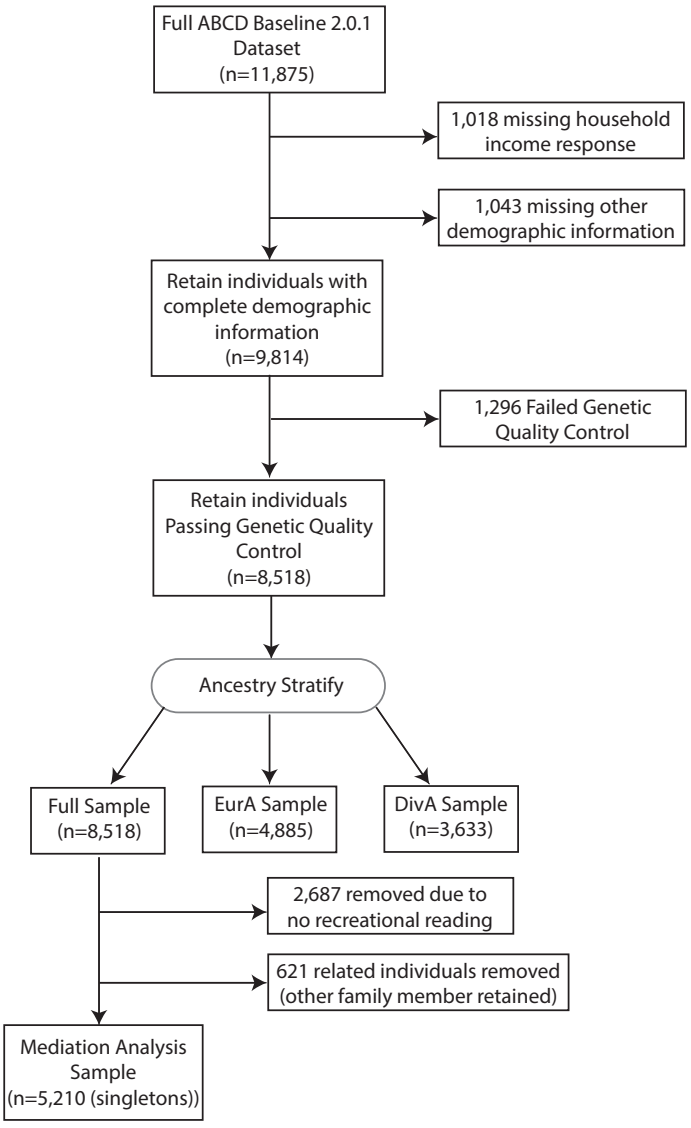


Figure 3.1 Flow chart of sample selection and exclusion.

Table 3.1 Summary of demographics for individuals included in the full sample for the present genomic prediction analyses, and for the genomic European Ancestry and genomic Other Ancestry subgroups.

	Full Sample	*European Ancestry*	*Other Ancestry*
Total N	8518	4885	3633
		Mean (SD)	
Age - months	119.05 (7.48)	119.21 (7.49)	118.85 (7.47)
		N (%)	
Sex Male	4438 (52.1)	2576 (52.7)	1862 (51.3)
Parent Married = Yes	6024 (70.7)	4066 (83.2)	1958 (53.9)
Parental Education			
< HS Diploma	302 (3.5)	21 (0.4)	281 (7.7)
HS Diploma/GED	649 (7.6)	138 (2.8)	511 (14.1)
Some College	2149 (25.2)	899 (18.4)	1250 (34.4)
Bachelor	2318 (27.2)	1548 (31.7)	770 (21.2)
Post Graduate Degree	3100 (36.4)	2279 (46.7)	821 (22.6)
Household Income			
[<50K]	2353 (27.6)	596 (12.2)	1757 (48.4)
[>=50K & <100K]	2444 (28.7)	1471 (30.1)	973 (26.8)
[>=100K]	3721 (43.7)	2818 (57.7)	903 (24.9)
Race			
White	5715 (67.7)	4750 (97.4)	965 (27.1)
Black	1129 (13.4)	1 (0.0)	1128 (31.7)
Asian	199 (2.4)	0 (0.0)	199 (5.6)
Other	1397 (16.6)	128 (2.6)	1269 (35.6)
Hispanic			
Hispanic	1628 (19.1)	131 (2.7)	1497 (41.2)

3.4.2 Behavioral Measures and Sociocultural Factors

Table 3.2 Mean (SD) and median for each behavioral measure in the full sample, estimated % variance explained by sex, age, and the set of socio-cultural covariates (parental marital status, parental education, household income, genetic ancestry PCs and Hispanic/non-Hispanic).

	Mean (SD)	Median	Sex	Age	Sociocultural
Crystallized Composite	86.87 (6.93)	87	0.01	9.51	21.57
Fluid Composite	92.18 (10.43)	93	0.32	7.17	10.28
Reading	91.23 (6.73)	91	0.01	5.97	13.18
Picture Vocabulary	85.04 (8.02)	84	0.07	7.67	20.14
Pattern	88.29 (14.47)	88	0.57	4.81	1.90
List	97.43 (11.81)	97	0.13	2.04	9.47
Picture	103.33 (12.01)	103	0.51	1.17	5.46
Flanker	94.42 (8.83)	96	0.03	3.21	3.73
Cardsort	92.97 (9.26)	94	0.48	3.76	5.22
Rey Auditory Verbal	43.78 (9.96)	44	1.25	2.23	8.57
Matrix Reasoning	18.13 (3.74)	18	0.34	2.74	9.15
Little Man Task	0.60 (0.17)	0.56	0.48	5.13	6.25
Bayesian Factor 1	0.05 (0.76)	0.06	0.28	9.63	20.85
Bayesian Factor 2	0.02 (0.76)	0.06	0.22	5.49	2.65
Bayesian Factor 3	0.04 (0.70)	0.04	0.89	1.59	7.83
Recreational Reading (hours)	6.5 (10)	4	0.45	0.18	0.86

Mean performance, standard deviation (SD), median and estimates of variance explained by age, sex, and the set of socio-cultural covariates (parental marital status, highest education level of parent/caregiver, household income, ethnicity, genetic principal components) are given for each behavioral measure examined in Table 3.2. Consistent with previous reports, there are substantial differences in the degree to which socio-cultural factors account for variability in these measures. The Crystallized Composite, its constituent Picture Vocabulary and Reading Recognition measures, and BF1, on which these measures of language and literacy load heavily, all exhibit higher levels of association with socio-cultural variables. This pattern persisted when controlling for IPS (Supplementary Table 3.2). Sex, age and socio-cultural factors explained little variability in recreational reading. Partial correlations

between the individual cognitive task measures controlling for covariates (Figure 3.2), suggest that performance on the different tasks is modestly correlated across children (rs=.08-.41) in this sample. Correlations peak in the .3 range within Fluid Composite measures, and the highest correlation is observed between the two Crystallized Composite measures (Picture Vocabulary and Oral Reading r=.41).

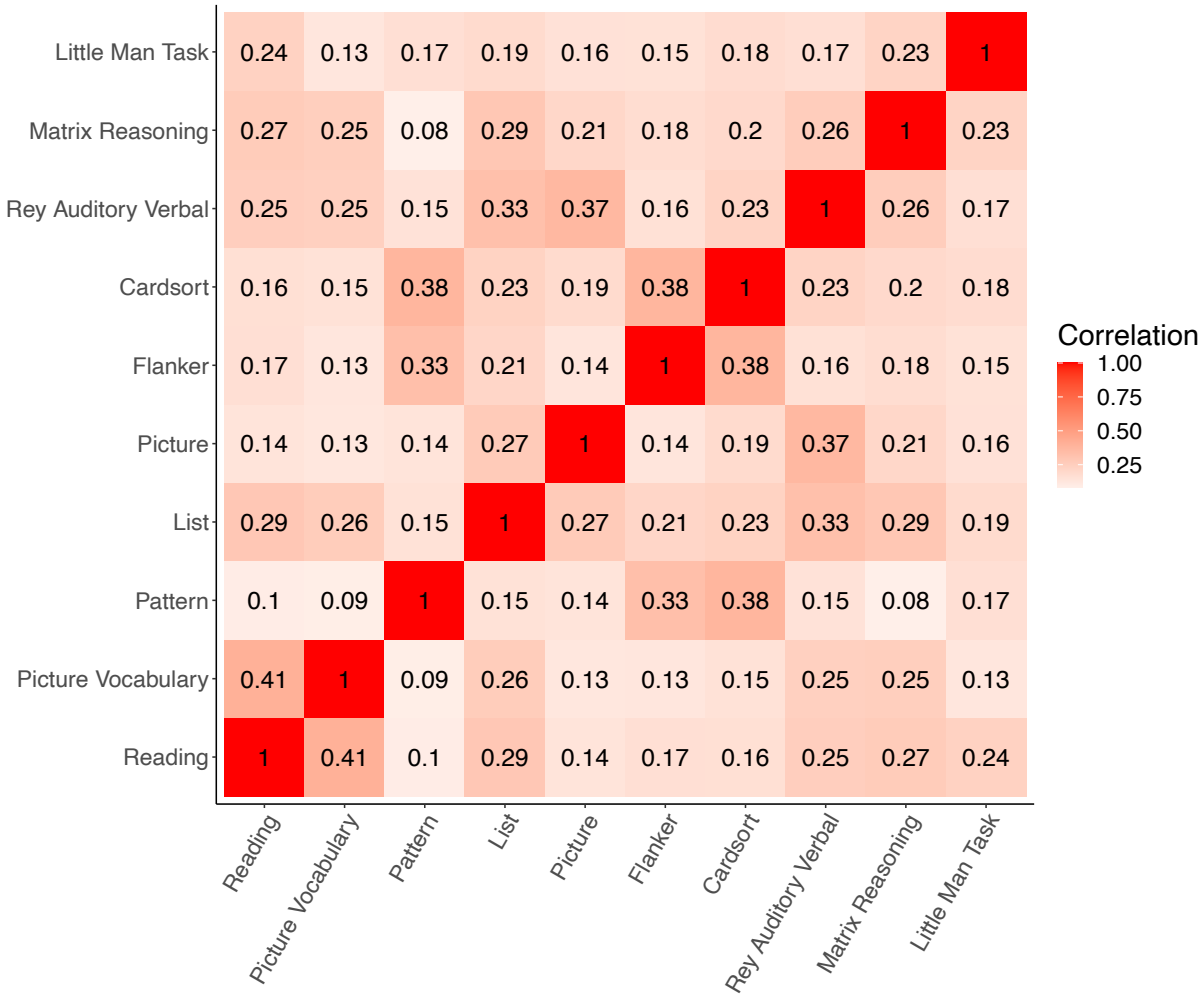


Figure 3.2 Partial correlation matrix showing intercorrelations among individual task performance measures (controlling for age, sex, parental marital status, parental education, household income, principal components of genetic ancestry and Hispanic status) in the full sample included in the present study of genomic predictors.

3.4.3 Genomic Prediction of Crystallized and Fluid Cognition Measures

Table 3.3 summarizes the regression results for predicting the Crystallized and Fluid Composites with IPS or EAPS in the full sample, and separately in the EurA and DivA subsamples. The IPS was a significant predictor of both measures in all analyses. Importantly, the standardized regression coefficient was significantly higher for the Crystallized than the Fluid Composite regardless of ancestry group (full sample: z=4.8, $p=1.8x10^{-6}$, EurA: z=4.6, $p=5.1x10^{-6}$ and DivA: z=2.5, $p=1.4x10^{-2}$).

Table 3.3 Regression results for GLMMs associating IPS (top) and EAPS (bottom) with Crystallized Composite and Fluid Composite of the NIH Toolbox

Sample	*Fluid Composite*				*Crystallized Composite*			
	Std. ß	t	P value	% Var. Explained	Std. ß	t	P value	% Var Explained
	IPS							
Full Sample	0.28	8.03	1.14E-15	0.75	0.50	15.82	1.31E-55	2.86
EurA	0.11	7.53	6.10E-14	1.15	0.21	14.48	1.44E-46	4.13
DivA	0.20	3.41	6.52E-04	0.32	0.40	7.34	2.68E-13	1.47
	EAPS							
Full Sample	0.11	7.23	5.26E-13	0.61	0.19	14.21	2.56E-45	2.32
EurA	0.09	6.60	4.66E-11	0.89	0.18	12.95	9.36E-38	3.34
DivA	0.08	3.38	7.28E-04	0.32	0.15	6.66	3.24E-11	1.21

In no case did the EAPS, despite a much larger training sample size, appear to account for more of the variance in the neurocognitive measures than did IPS. However, across ancestry groups and for both composite scores, combining both genomic predictors explained significantly more variance in behavior than IPS alone (supplementary results). IPS + EAPS explained 5.8% variance ($p=4.5x10^{-64}$) in the Crystallized Composite for EurA (a 40% increase compared to IPS alone). Supplementary Table 3.3-3.8 show regression results for each behavior using IPS, EAPS and IPS + EAPS within each ancestry group.

Fitting separate regression models for each individual task in the neurocognitive battery, we found that the IPS was a significant predictor for each cognitive measure for the full

sample and the EurA subsample (all p values<10^{-3}), surviving the Bonferroni-corrected significance threshold of 0.05/10=0.005. Within the DivA subsample only six of the ten tasks were individually significantly predicted by the IPS (Supplementary Table 3.8). Figure 3.3 shows the standardized regression coefficients of IPS predicting performance on each task, as well as the Crystallized and Fluid Composite measures from the NTCB and Bayesian Factors 1-3(16), in the full sample. Individual cognitive measures included in the Crystallized Composite have consistently higher IPS standardized regression weights than the measures included in the Fluid Composite. Other neurocognitive tasks from the ABCD battery (shaded in gray) showed similar associations to the Fluid Composite. The results for the Bayesian Factors mirrored these results: BF1, on which 'Crystallized" measures had the highest factor loadings (Supplementary Figure 3.1)(16), displayed a stronger association with IPS than BF2 and BF3 on which 'Fluid", executive function and memory measures had higher loadings. The results in the subsamples (EurA and DivA) are provided in the appendix.

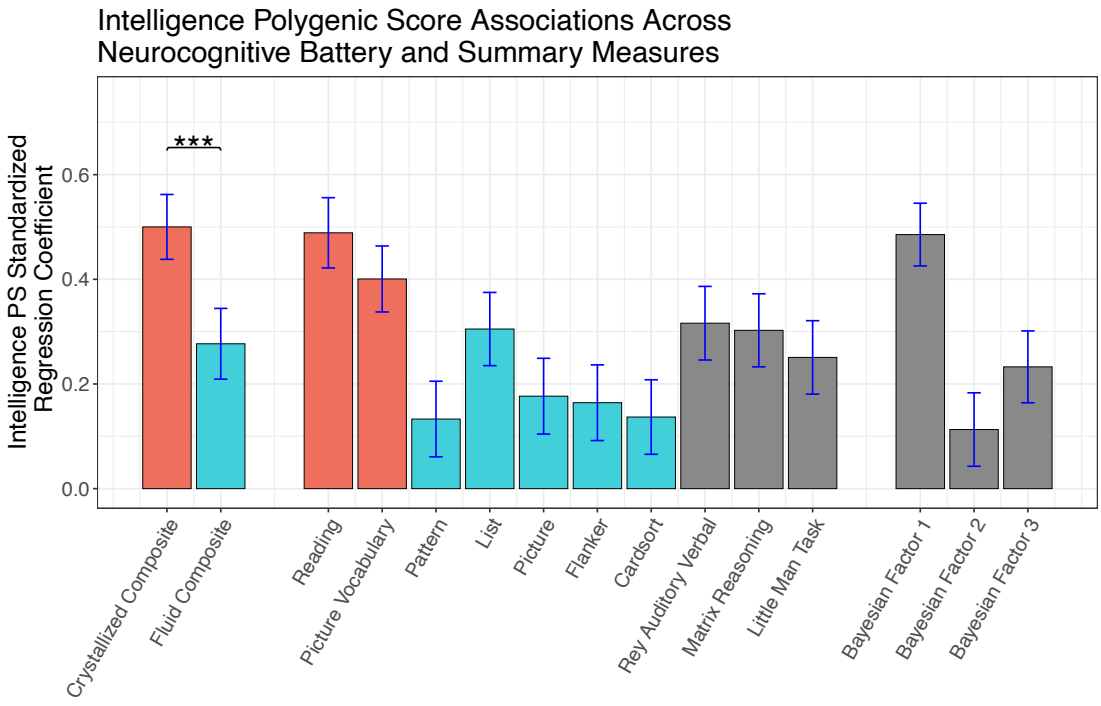


Figure 3.3 Standardized regression coefficients of IPS for fitting linear mixed models to performance on Fluid and Crystallized Composites, each individual task from the NTCB, additional measures from the ABCD neurocognitive battery, and Bayesian (latent) Factors 1-3, in the full sample. Prediction of the Crystallized Composite is significantly stronger than for the Fluid Composite. Tasks included in the Fluid Composite (shaded in blue) have consistently lower regression coefficients than those included in the Crystallized Composite (shaded in red). Additional measures from the neurocognitive battery exhibit associations with IPS more similar to the Fluid Composite than to the Crystallized Composite, however Bayesian Factor 1, on which the verbal tasks load heavily, exhibits an association similar to the Crystallized Composite. Error bars show estimates of 95% confidence intervals as 1.96 standard error.

3.4.4 Differential Mediation Results

The mediation analysis showed that recreational reading partially mediated associations between IPS and both composite measures, proportions of mediation: fluid 0.084 (95% CI: 0.047-0.14, $p<10^{-16}$), crystallized 0.12 (95% CI: 0.088-0.16, $p<10^{-16}$). However, the differential mediation analyses revealed a highly significant difference between the large degree of attenuation of the association between IPS and the Crystallized Composite relative to that between IPS and the Fluid Composite (Welch t-test: t=125, df=19053, $p<10^{-16}$), shown in Figure 3.4.

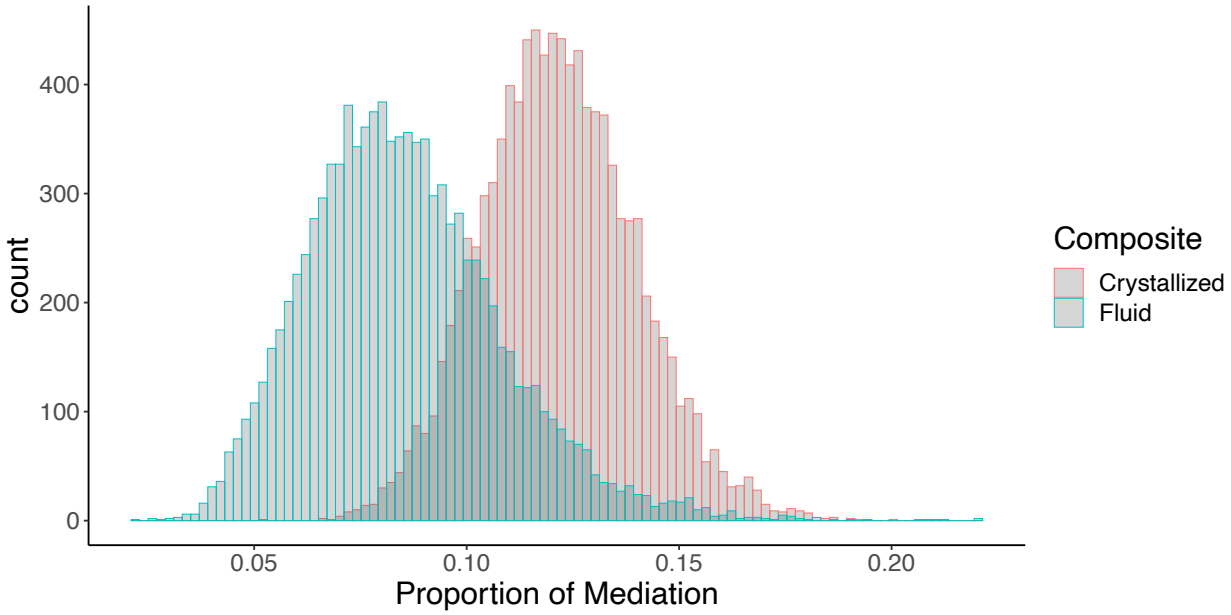


Figure 3.4 Differential mediation analysis in singletons (N= 5,210): histograms shows 10,000 bootstrap estimates for proportion of mediation of recreational reading on: i) IPS and Crystallized Composite (red) and ii) IPS and Fluid Composite (blue). Recreational reading attenuates the relationship between the IPS and the Crystallized Composite to a significantly greater degree.

3.4.5 Sensitivity Analyses to Address Test Reliability

A previous study reported the test-retest reliability for the Fluid Composite from the NTCB (.76) was somewhat lower than for the Crystallized Composite (.85)(5), raising questions about whether differences in the strength of their associations with IPS could be attributed to more noise in the Fluid Composite measure. In supplementary sensitivity analyses we demonstrate that our results are robust to the addition of simulated noise to the Crystallized Composite that mimics this difference in test reliability. At this level of simulated noise we estimated 1.0 power (alpha=0.05) to detect $Cryst_{noise}$ having a significantly greater IPS standardized regression coefficient than the Fluid Composite. Moreover, additional sensitivity analyses indicate that the observed differences in the mediation effects of recreational reading are similarly robust against potential measurement error modelled as random noise. These analyses are described in detail in the appendix.

3.5 *Discussion*

We have shown that polygenic predictors of intelligence test performance and of educational attainment are associated with neurocognitive performance in this large group of children from diverse backgrounds. These results are consistent with previous findings demonstrating that virtually all behavioral traits, including cognitive and intellectual phenotypes, are heritable(26). Moderate estimates of heritability of many behavioral phenotypes also establish that a substantial portion of the variability is due to independent environmental influences. Given that behavioral phenotypes emerge through interactions between children and their physical, social, and cultural environments, much attention has been paid to how these environmental factors modify the phenotypes, since they are presumably the malleable factors. However, recently, more attention has been focused on the possible roles of mediating nongenetic (ECE) factors that, through their statistical association with genetic variation (rGE), may amplify heritability(7,8).

We found that a culturally dependent estimate of crystallized cognitive functions, the Crystallized Composite measure from the NTCB, is more strongly associated with the best available polygenic predictor of intelligence test performance than is the Fluid Composite measure, consistent with earlier findings in adults of heritability differences(7) and polygenic score performance(27) across similar measures. This is despite the IPS being based on a large meta-analysis of GWAS combining cognitive measures that were described by the authors as primarily “fluid intelligence” measures(9). Indeed, the relative size of the IPS association across the 15 measures examined here (Figure 3.3) closely mirrored the relative percent variance explained in these measures by socio-cultural variables (Table 3.2), a pattern that persists after accounting for IPS (Supplementary Table 3.2). Moreover, for children who read for pleasure, the extent of recreational reading was found to partially mediate the associations between IPS

and both composite measures, but to a significantly greater degree for the crystallized than for the fluid measure, consistent with a more prominent role of rGE in the development of abilities tapped by measures that are both more heritable, and apparently more sensitive to socio-cultural variables. In other words, even when controlling for *independent* contributions of more global sociocultural variables, how often a child reads for pleasure more strongly mediates the association between IPS and crystallized rather than in fluid performance.

It is perhaps unsurprising that recreational reading more strongly mediates the association between IPS and culturally sensitive measures of intelligence since such measures are generally sensitive to educational factors. Indeed, a measure of oral reading proficiency loads highly on both measures of crystallized functions examined here, Crystallized Composite and BF1. One can imagine that children with neurobehavioral phenotypes advantageous for learning to read might be more likely to develop the habit of reading for pleasure than those with other neurobehavioral phenotypes, for a variety of reasons. However, these results imply that choosing to read for pleasure at 10 years of age is associated with having a genotype linked to intellectual functions most dependent on reading, and that an estimate of the frequency of reading behavior mediates that link. This is consistent with previous descriptions of rGE effects, and with analyses by Beam and Turkheimer(8), who showed that increasing rGE over time could explain observed increases in the heritability of measures of cognitive function through development. The ABCD study will provide an opportunity to measure changes in heritability at later time points of this longitudinal study. Importantly, despite the lower test-retest reliability of the fluid compared to the crystallized composite score from the NTCB(5), our supplementary analyses show that this difference in test reliability is unlikely to explain our findings.

Though recreational reading would appear to be an enhancing mediator of intellectual development, it is important to note that genotype-correlated ECE factors can also suppress

intellectual development. As an example, early struggles to read by children with less advantageous genotypes for reading may decrease the likelihood that these children will choose reading activities, leading to slower progression of these faculties. Worse, if children's early reading attempts are experienced very negatively, these children may develop avoidant responses to reading, which could result in active suppression of developing literacy. Importantly, when these kinds of differences originate with differences in children's genotypes, they can increase heritability and exaggerate disparities. Identifying ECE factors that contribute to heritability of cognitive and intellectual phenotypes is important because it can point to practices that better adapt to neurogenetic diversity among children. Innovative pedagogical practices may lead to approaches that increase "enhancing" ECE effects in the subset of children disadvantaged by current practices and reduce ECE effects that suppress intellectual development and academic achievement, which may lead to more equitable educational outcomes.

Limitations and Caveats

The proportion of variance in the cognitive measures accounted for by the genomic predictors was larger in the EurA participants than in the DivA group(

Supplementary Table 3.5 & Supplementary Table 3.7), as would be expected given the discovery samples were in individuals of European ancestry. However, the patterns were generally similar in the DivA group. This suggests similar genetic architecture for these cognitive phenotypes across ancestry groups and supports the validity of the results from the full sample. Analyses in all three groups included as covariates the top ten genetic principal components derived from the full sample. Because of broad ancestral diversity in the ABCD cohort, there is limited power for comparing the effects in different ancestry groups. As has been discussed in genetics generally(28,29), the lower predictive performance in the DivA group once again underscores the importance of collecting genetic data from ancestrally diverse populations and developing methods that can be used across ancestry groups.

One may have predicted that EAPS would have been a more powerful predictor of cognitive measures in ABCD than IPS, due to it having over 4 times the discovery sample size. However, we found generally the IPS had stronger associations (Supplementary Table 3.3-3.8), perhaps because the phenotype is a better match between training and testing. This contrasts with results of a previous study of adults, where EAPS explained 7-10% of the variance in cognition(10), while IPS explained only 2-5%(9). This discrepancy may be due to methodological differences, alternatively the young age of the cohort may be the key difference. Educational attainment, while clearly related to scores on cognitive tests, may be influenced by other genetically influenced traits (e.g. personality) that may contribute to greater persistence in formal education; thus the EAPS is likely to reflect to a greater degree these traits. Such pleiotropic of EAPS effects has been observed in adults(30). When we include both EAPS and IPS in a single model, together they explain 5.8% of the variability in the Crystallized Composite (EurA, Supplementary Table 3.6), substantially more than IPS alone explains (4.1%), indicating that these genomic predictors capture unique sources of the relevant variance, and are likely measuring different (relevant) constructs.

These results are consistent with previous evidence for a role of genetic variation in developing cognitive functions, and they strengthen the evidence for rGE during cognitive development. However, it should be emphasized that the genomic predictors (together) account for only 4.15% of cognitive performance variance in the full sample. Furthermore, this was observed for the Crystallized Composite measure, the culturally sensitive measure hypothesized to exhibit increased genetic association because of rGE effects. The additive effects of potentially confounding sociocultural covariates, even controlling for IPS, accounted for 13.2% of the variability. For the Fluid Composite the genomic predictors together accounted for only 1.1% of the variance, and sociocultural covariates accounted for almost 5%. Of note, even with the narrow 2 year age range in the cohort, age alone accounts for 10% of the variability in the Crystallized Composite and 7% in the Fluid Composite. These effects may reveal clues about a highly dynamic process of cognitive and intellectual development in children.

Finally, though the results of the mediation analysis focusing on recreational reading strengthen the plausibility that such ECE mediators associate with genotypes and increase genetic effects, these results do not prove a causal explanation, and none should be inferred. In the context of an observational study such as ABCD it is always possible that confounding variables not accounted for in the analysis are responsible for the mediation effect we observed.

Acknowledgment

Chapter 3, in full, is available as a preprint on BioRxiv and has been submitted for publication. Loughnan R, Palmer C, Makowski C, Thompson W, Dale A, Jernigan T, Fan C. The dissertation/thesis author was the primary investigator and author of this paper.

References

1. Deary IJ, Gottfredson LS. Intelligence Predicts Health and Longevity, but Why? Curr Dir Psychol Sci. 2004;13(1):1–4.

2. Kuh D, Richards M, Hardy R, Butterworth S, Wadsworth MEJ. Childhood cognitive ability and deaths up until middle age: A post-war birth cohort study. Int J Epidemiol. 2004;33(2):408–13.

3. Kell HJ, Lubinski D, Benbow CP. Who Rises to the Top? Early Indicators. Psychol Sci. 2013;24(5):648–59.

4. Polderman TJC, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet [Internet]. 2015;47(7):702–9. Available from: http://dx.doi.org/10.1038/ng.3285

5. Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen S. NIH Toolbox Cognitive Function Battery (CFB): Composite Scores of Crystallized, Fluid, and Overall Cognition. Monogr Soc Res Child Dev. 2013;78(4):119–32.

6. Akshoomoff N, Newman E, Thompson WK, Mccabe C, Frazier JA, Gruen JR, et al. The NIH Toolbox Cognition Battery: Results from a Large Normative Developmental Sample (PING). Neuropsychology. 2014;28(1):1–10.

7. Kan KJ, Wicherts JM, Dolan C V., van der Maas HLJ. On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent. Psychol Sci. 2013;24(12):2420–8.

8. Burgess A, Shah K, Hough O, Hynynen K. Phenotype–environment correlations in longitudinal twin models. Dev Psychopathol. 2013;25(1):7–16.

9. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50(7):912–9.

10. Zacher M, Nguyen-viet TA, Bowers P, Sidorenko J, Linnér RK. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. 2018;50(August).

11. Morgan GD, Deeds BG, Wargo EM, Gordon JA, Bianchi DW, Riley WT, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. Dev Cogn Neurosci [Internet]. 2017;32(April 2017):4–7. Available from: https://doi.org/10.1016/j.dcn.2017.10.002

12. Auchter AM, Hernandez Mejia M, Heyser CJ, Shilling PD, Jernigan TL, Brown SA, et al. A description of the ABCD organizational structure and communication framework. Dev Cogn Neurosci [Internet]. 2018;32(November 2017):8–15. Available from: http://dx.doi.org/10.1016/j.dcn.2018.04.003

13. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox Cognition assessment using the NIH Toolbox. 2013;

14. Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, et al. Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. J Int Neuropsychol Soc. 2015;21(5):378–91.

15. Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, et al. Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. J Int Neuropsychol Soc. 2014;20(6):588–98.

16. Thompson WK, Barch DM, Bjork JM, Gonzalez R, Nagel BJ, Nixon SJ, et al. The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. Dev Cogn Neurosci [Internet]. 2018;(December):100606. Available from: https://doi.org/10.1016/j.dcn.2018.12.004

17. Raj A, Stephens M, Pritchard JK. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89.

18. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet [Internet]. 2016 Aug 29;48:1284. Available from: https://doi.org/10.1038/ng.3656

19. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466–8.

20. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet [Internet]. 2017;100(4):635–49. Available from: http://dx.doi.org/10.1016/j.ajhg.2017.03.004

21. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun [Internet]. 2019;10(1). Available from: http://dx.doi.org/10.1038/s41467-019-11112-0

22. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, et al. Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. PLoS Biol. 2013;11(9).

23. Wood S, Scheipl F. gamm4: Generalized additive mixed models using mgcv and lme4. R package version 0.2-3. 2014.

24. Imai K, Tingley D, Yamamoto T. Experimental designs for identifying causal mechanisms. J R Stat Soc Ser A Stat Soc. 2013;176(1):5–51.

25. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. J Stat Softw. 2014;59(5):1–38.

26. Eric Turkheimer. Three laws of behavior genetics and what they mean. Curr Dir Psychol

Sci. 2000;9(5):160–4.

27. Genç E, Schlüter C, Fraenz C, Arning L, Nguyen HP, Voelkle MC, et al. Polygenic Scores for Cognitive Abilities and their Association with Different Aspects of General Intelligence – a Deep Phenotyping Approach. bioRxiv [Internet]. 2020 Jan 1;2020.06.03.131318. Available from: http://biorxiv.org/content/early/2020/08/20/2020.06.03.131318.abstract

28. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet [Internet]. 2019;51(4):584–91. Available from: http://dx.doi.org/10.1038/s41588-019-0379-x

29. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. Genome Biol [Internet]. 2016;17(1):16–8. Available from: http://dx.doi.org/10.1186/s13059-016-1016-y

30. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. Proc Natl Acad Sci [Internet]. 2014 Oct 21 [cited 2020 Nov 20];111(42):15273–8. Available from: https://www.pnas.org/content/111/42/15273

Appendix

Description of Each Cognitive Task used in ABCD

NIH Toolbox Cognition Battery® Tasks

The *Toolbox Oral Reading Recognition Task®* measured language decoding and reading. Children were asked to read aloud single letters or words presented in the centre of an iPad screen. The research assistant marked pronunciations as correct or incorrect. Extensive training was given prior to administering the test battery. Item difficulty was modulated using computerized adaptive testing (CAT).

The *Toolbox Picture Vocabulary Task®*, a variant of the Peabody Picture Vocabulary Test (PPTV), measured language and vocabulary comprehension. Four pictures were presented on an iPad screen as a word was played through the iPad speaker. The child was

instructed to point to the picture, which represented the concept, idea or object name heard. CAT was implemented to control for item difficulty and avoid floor or ceiling effects.

The *Toolbox Pattern Comparison Processing Speed Test*® measured processing speed. Children were shown two images and asked to determine if they were identical or different by touching the appropriate response button on the screen. This test score is the sum of the number of items completed correctly in the time given.

The *Toolbox List Sorting Working Memory Test*® measured working memory. Children heard a list of words alongside pictures of each word and were instructed to repeat the list back in order of their actual size from smallest to largest. The list started with only two items and a single category (food or animals). The number of items increased with each correct answer to a maximum of seven. The child then progressed to the next stage in which the two different categories were interleaved. At this stage children were required to report the items back in size order from the first category followed by the second category. Children were always given two opportunities to repeat the list correctly before the experimenter scored the trial as incorrect.

The *Toolbox Picture Sequence Memory Test*® measured episodic memory. On each trial, children were shown a series of fifteen pictures in a particular sequence. The pictures illustrated activities or events within a particular setting (e.g., going to the park), and as each appeared on the screen a pre-recorded narrative briefly described the content of the picture. Participants were instructed to arrange the pictures in the original sequence in which they were shown. The Rey-Auditory Verbal Learning Task was also included in the ABCD neurocognition battery as a more comprehensive measure of episodic memory.

The *Toolbox Flanker Task*® measured executive function, attentional and inhibitory control. This adaptation of the Eriksen Flanker task(1) captures how readily a participant is influenced by the congruency of stimuli surrounding a target. On each trial a target arrow was

presented in the center of the iPad screen facing to the left or right and was flanked by two additional arrows on both sides. The surrounding arrows were either facing in the same (congruent) or different (incongruent) direction to the central target arrow. The participant was instructed to push a response button to indicate the direction of the central target arrow. Accuracy and reaction time scores were combined to produce a total score of executive attention, such that higher scores indicate a greater ability to attend to relevant information and inhibit incorrect responses.

The *Toolbox Dimensional Change Card Sort Task*® measured executive function and cognitive flexibility. On each trial, the participant was presented with two objects at the bottom of the iPad screen and a third object in the middle. The participant was asked to sort the third object by matching it to one of the bottom two objects based on either colour or shape. In the first block participants matched based on one dimension and in the second block they switched to the other dimension. In the final block, the sorting dimension alternated between trials pseudorandomly. The total score was calculated based on speed and accuracy.

Other Neurocognitive Tasks

Rey-Auditory Verbal Learning Task. This task measures auditory learning, recall and recognition. Participants listened to a list of 15 unrelated words and were asked to immediately recall these after each of five learning trials. A second unrelated list was then presented and participants were asked to recall as many words as possible from the second list and then recall words again from the initial list. Following a delay of 30 minutes (during which other non-verbal tasks from the cognitive battery are administered), longer-term retention was measured using recall and recognition. This task was administered via an iPad using the Q-interactive platform of Pearson assessments(2). In the current study, the total

number of items correctly recalled across the five learning trials was summed to produce a measure of auditory verbal learning.

Little Man Task. This task measures visuospatial processing involving mental rotation with varying degrees of difficulty(3). A rudimentary male figure holding a briefcase in one hand was presented on an iPad screen. The figure could appear in one of four positions: right side up vs upside down and either facing the participant or with his back to the participant. The briefcase could be in either hand. Participants indicated which hand the briefcase was in using one of two buttons. Performance across the 32 trials was measured by the percentage of trials in which the child responded correctly. This was divided by the average reaction time to complete the task (in seconds) to produce a measure of efficiency of visuospatial processing. This was the dependent variable analysed for this task.

Matrix Reasoning. Nonverbal reasoning was measured using an automated version of the Matrix Reasoning subtest from the Weschler Intelligence Test for Children-V(4). On each trial the participant was presented with a series of visuospatial stimuli, which was incomplete. The participant was instructed to select the next stimulus in the sequence from four alternatives. There were 32 possible trials and testing ended when the participant failed three consecutive trials. The total raw score, used in the current study, was the total number of trials completed correctly.

Genetic Data and Computing Polygenic Scores

Saliva and blood samples were collected and sent to Rutgers University Cell and DNA Repository for DNA isolation. Genotyping was performed using the Smokescreen array(5), calling 646,247 genetic variants. Pre-variant imputation, quality control (QC) on the genotyping was performed to ensure each genetic variant had been successfully called in more than 95%

of the sample, and that missingness for each individual was lower than 20%. After QC, 517,724 SNPs and 10,659 individuals remained. Based on genotype data, we derived genetic ancestry using fastStructure(6) with four ancestry groups. Genetic principal components were also calculated using PLINK.

We performed imputation using the Michigan Imputation Server(7) with the hrc.r1.1.2016 reference panel, Eagle v2.3 phasing and multiethnic imputation. PLINK(8) was used to convert dosage files to PLINK files using a best guess threshold of 0.9 for each locus. After PLINK conversion, we used post imputation variant QCs of minor allele frequency above 5%, Hardy-Weinberg threshold of 10^{-6} and no greater than 10% missing SNPs for each individual. These QC processes resulted in 38,900,342 SNPs and 10,659 individuals remaining.

We computed polygenic scores usine PRSice(9). After variant imputation we performed clumping and pruning of SNPs with a clumping window of 250 kb and r^2 of 0.1. SNPs from the major histone compatibility complex were removed from the analysis. This resulted in 692,685 SNPs remaining. The polygenic scores for each individual were then computed as a sum of their SNPs weighted by the variant effect size in the discovery samples(10,11), with no p-value thresholding of summary statistics. Since part of the EAPS discovery sample was 23andMe participants, only the top 10,000 SNPs were included in summary statistics for this polygenic score.

Statistical Models for Genomic Prediction of Behavior Measures

Associations between polygenic scores and behavior measures were assessed using Generalized Linear Mixed-Effects Models (GLMMs). In addition to the IPS or EAPS, all models included the fixed effects of age, sex at birth, parental marital status, education level of parent/caregiver, household income, ethnicity (Hispanic/non-Hispanic) and the top ten genetic

principal components. Data collection site and family were included as random effects. To assess variance explained by predictor(s) we computed either from t statistics and degrees of freedom (i.e. $R^2 = t^2/(t^2 + df)$) or as the computed change in R^2 based on a log likelihood ratio test between a full and reduced model.

Differential Mediation Analysis

Due to the computational burden of fitting GLMMs for bootstrapped estimates we instead fit general linear models (GLMs) with the same fixed effects as above and adding study site. To control for family we restricted this analysis to randomly selected singletons (one member from each family). By randomly selecting singletons we believe we obtain a conservative estimate of associations as we lose power when compared to GLMMs (where all family members are retained) whilst controlling family relatedness in ABCD.

	Factor 1	*Factor 2*	*Factor 3*
Reading	0.82	0.12	0.12
Picture Vocabulary	0.75	0.07	0.19
Pattern	0.02	0.81	0.09
List	0.47	0.15	0.49
Picture	0.01	0.14	0.86
Flanker	0.21	0.71	0.07
Card Sort	0.21	0.71	0.23
Rey Auditory Verbal	0.31	0.13	0.71
Little Man Task	0.50	0.30	0.07

Supplementary Figure 3.1 Loadings of Bayesian factors from (12), mean of posterior distributions.

Supplementary Table 3.1 Variables used for associations, with NDA (National Institute of Mental Health Data Archive) data dictionary names.

Assessment	*Variables Analyzed*	*NDA Data Dictionary Name*	*Informant*
NIH Toolbox®	Crystallized Composite Fluid Composite Reading Picture Vocabulary Pattern List Picture Flanker Cardsort	nihtbx_cryst_uncorrected nihtbx_fluidcomp_uncorrected nihtbx_reading_uncorrected nihtbx_picvocab_uncorrected nihtbx_pattern_uncorrected nihtbx_pattern_uncorrected nihtbx_list_uncorrected nihtbx_flanker_uncorrected nihtbx_cardsort_uncorrected	Youth
Little Man Task	Percentage correct	lmt_scr_perc_correct	Youth
Pearson Scores	Rey Auditory Verbal total correct WISC-V Matrix Reasoning	pea_ravlt_sd_trial_[i-v]_tc (sum i-v) pea_wiscv_trs	Youth
Bayesian Latent Factors	Bayesian Factor 1 Bayesian Factor 2 Bayesian Factor 3	neurocog_pc1.bl neurocog_pc2.bl neurocog_pc3.bl	Youth (derived)
Sports and Activity Involvement Questionaries	Reading hours	sports_activity_read_hours_p	Caregiver

Supplementary Table 3.2 Mean (SD) and median for each behavioral measure in the full sample, estimated % variance explained by sex, age, and the set of socio-cultural covariates (parental marital status, parental education, household income, genetic ancestry PCs and Hispanic/non-Hispanic), ***after accounting for IPS***.

	Mean (SD)	Median	Sex	Age	Sociocultural
Crystallized Composite	86.87 (6.93)	87	0.01	9.96	13.06
Fluid Composite	92.18 (10.43)	93	0.33	7.32	4.97
Reading	91.23 (6.73)	91	0.01	6.27	9.24
Picture Vocabulary	85.04 (8.02)	84	0.07	7.94	10.67
Pattern	88.29 (14.47)	88	0.57	4.85	1.09
List	97.43 (11.81)	97	0.13	2.11	4.82
Picture	103.33 (12.01)	103	0.51	1.20	2.26
Flanker	94.42 (8.83)	96	0.03	3.26	2.13
Cardsort	92.97 (9.26)	94	0.48	3.80	2.54
Rey Auditory Verbal	43.78 (9.96)	44	1.26	2.32	3.34
Matrix Reasoning	18.13 (3.74)	18	0.34	2.83	4.54
Little Man Task	0.60 (0.17)	0.56	0.48	5.23	3.80
Bayesian Factor 1	0.05 (0.76)	0.06	0.30	10.05	12.73
Bayesian Factor 2	0.02 (0.76)	0.06	0.22	5.53	1.50
Bayesian Factor 3	0.04 (0.70)	0.04	0.90	1.64	2.94
Recreational Reading (hours)	6.5 (10)	4	0.45	0.21	1.65

Simulating Noise on Crystallized Composite Score

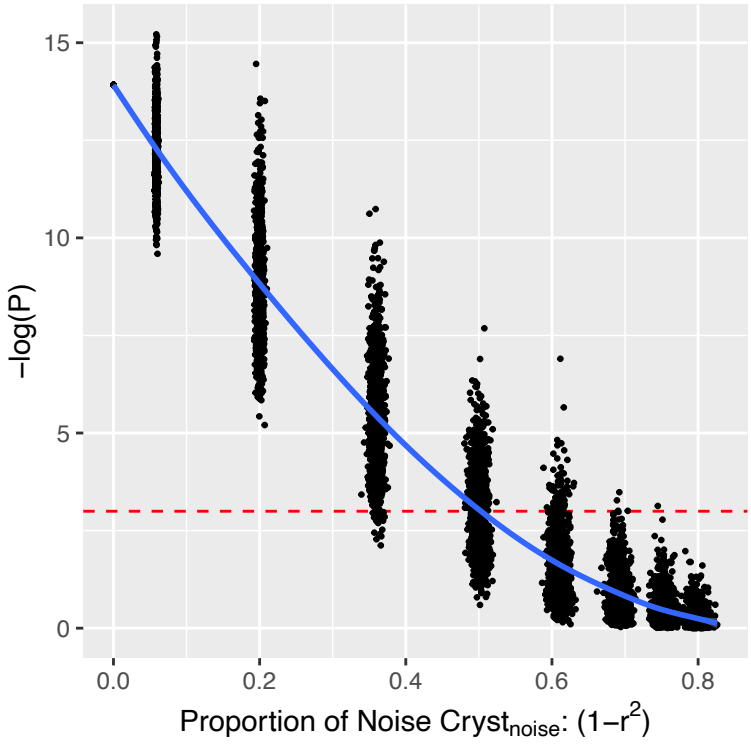
A previous study noted that crystallized composite of NIH toolbox shows greater test-retest reliability than fluid domains of intelligence(13). This is important to consider in the current study as it is possible that the observed findings, of IPS being stronger predictors of crystallized domains of intelligence, may be due to differences in noise between fluid and crystallized measures. To test the robustness of the results to different degrees of noise we simulated additive noise on the crystallized composite score as:

$$Cryst_{noise} = Cryst + N(0, \sigma)$$

In which we created samples of $Cryst_{noise}$ for different values of noise, σ and repeated the main behavioral analyses with IPS.

IPS Predictive Performance on $Cryst_{nosie}$ vs Fluid Composite

We simulated 1,000 samples of $Cryst_{noise}$ at each of 9 different values of σ from 0 to 2 in increments of 0.25 resulting in 9x1000=9000 samples of $Cryst_{noise}$. Next we performed the same z-test on performed in the main analysis, now between the standardized regression coefficients of $Cryst_{noise}$ and Fluid composite. We calculated the proportion of noise in signal for $Cryst_{noise}$ as $1-r^2$, where r = cor(Cryst, $Cryst_{noise}$). Supplementary Figure 3.2 shows the negative log(p) value of the z-test as a function of noise, with the red dotted line indicating a significance threshold of 0.05. A previous validation study estimated the test-retest reliability of Fluid to be 0.76 and Crystallized to be 0.85(13), this means $Cryst_{noise}$ would need to have an $r \approx 0.76/0.85=0.89$ with *Cryst* or $1-r^2 \approx 0.2$. At this level of noise we have 1.0 power (alpha=0.05) to detect $Cryst_{noise}$ having a significantly greater IPS standardized regression coefficients than Fluid composite. This demonstrates that our findings are robust to the difference in noise between crystallized and fluid composite measures.

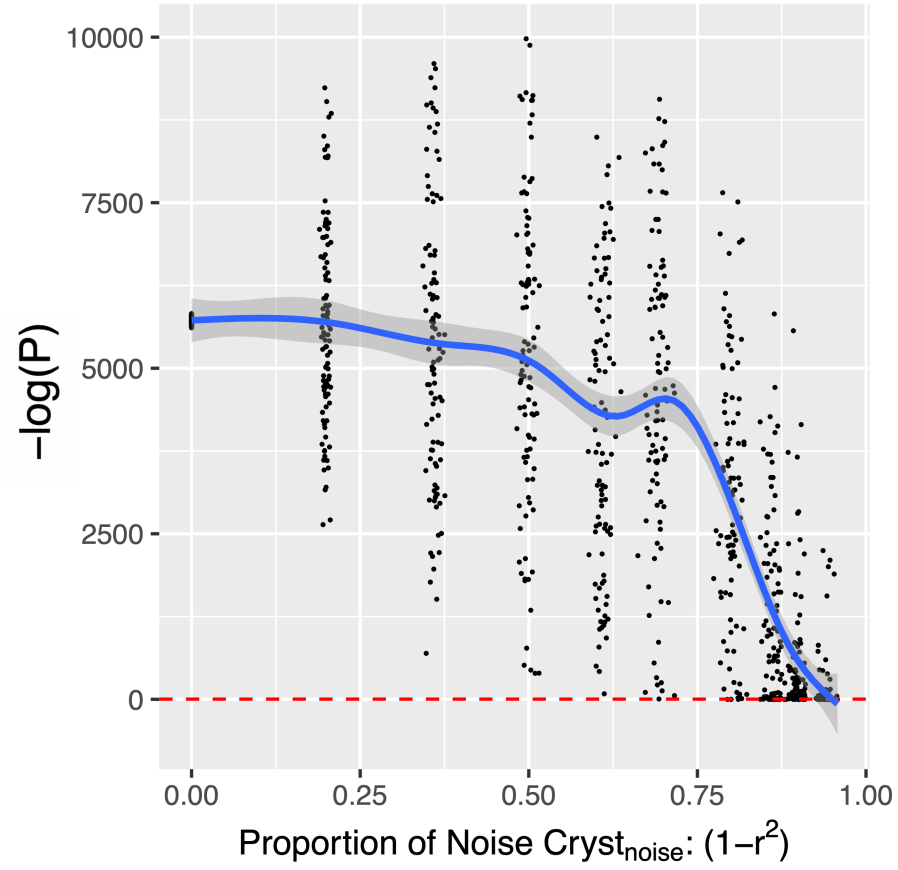


Supplementary Figure 3.2 Significance of difference in IPS standardized regression coefficient between predicting i) Cryst$_{noise}$ and ii) Fluid composite (y axis) as in main Figure 3.3, as a function of proportion of noise in Cryst vs Cryst$_{noise}$ (x-axis). Red dotted line indicates alpha=0.05.

Differential Mediation of Reading Between i) IPS and Cryst$_{nosie}$ and ii) IPS and Fluid Composite

To test the robustness of the mediation results in the main analysis, we repeated the same process with 100 samples of Cryst$_{noise}$ at each of 10 different values of σ (0, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 4) resulting in 1,000 different samples of Cryst$_{noise}$. Repeating a similar mediation analysis to the main text except instead using the more classical framework of Baron and Kenny(14), due to computational feasibility, which we have found to be convergent with the mediation model(15) presented in the main text. We again looked at 5,210 singletons from the full sample to test the differential mediation of reading between IPS and Cryst$_{noise}$ vs Fluid composite using 10,000 bootstrap samples for each Cryst$_{noise}$, we found once again a similar

robustness to the level of noise simulated. Supplementary Figure 3.3 shows the negative log(p) for the same test performed for the difference in distributions shown in Figure 3.4 at the 1,000 different simulated samples of $Cyrst_{noise}$ as a function of noise ($1-r^2$). We see that when $Cryst_{noise}$ has a value of noise making it comparable to the test retest reliability of fluid composite ($1-r^2 \approx 0.2$), all the samples remain significant. Again these simulated noise analyses demonstrate that our main results are robust to a higher degree of noise than the observed difference in test-retest reliability between Crystallized and Fluid composite measures in the toolbox.



Supplementary Figure 3.3 Welch t-test for differential mediation between i) IPS – reading - Fluid and ii) IPS – reading $Cryst_{noise}$ (y-axis), as a function of proportion of noise in Cryst vs $Cryst_{noise}$ (x-axis). Red dotted line indicates alpha=0.05.

IPS, EAPS and EAPS + IPS Cognition Associations

Supplementary Table 3.3 Full sample: behavioral associations between polygenic scores i) IPS and ii) EAPS in two separate regression for each behavior using GLMMs and controlling for variables of no interest.

	IPS				***EAPS***			
Behavior	coeff	Var. Explained (%)	t	pval	coeff	Var. Explained (%)	t	pval
Crystallized Composite	0.50	2.86	15.82	1.31E-55	0.19	2.32	14.21	2.56E-45
Fluid Composite	0.28	0.75	8.03	1.14E-15	0.11	0.61	7.23	5.26E-13
Reading	0.49	2.34	14.27	1.12E-45	0.16	1.39	10.95	9.85E-28
Picture Vocabulary	0.40	1.79	12.46	2.59E-35	0.17	1.87	12.74	7.75E-37
Pattern	0.13	0.15	3.61	3.10E-04	0.05	0.15	3.51	4.44E-04
List	0.31	0.85	8.55	1.50E-17	0.11	0.63	7.33	2.47E-13
Picture	0.18	0.27	4.79	1.73E-06	0.09	0.41	5.88	4.30E-09
Flanker	0.16	0.23	4.45	8.70E-06	0.03	0.04	1.91	5.63E-02
Cardsort	0.14	0.17	3.77	1.67E-04	0.06	0.18	3.96	7.70E-05
Rey Auditory Verbal	0.32	0.91	8.81	1.50E-18	0.11	0.66	7.52	6.23E-14
Matrix Reasoning	0.30	0.84	8.49	2.34E-17	0.14	0.98	9.18	5.36E-20
Little Man Task	0.25	0.57	7.00	2.69E-12	0.07	0.24	4.55	5.50E-06
Bayesian Factor 1	0.49	2.88	15.88	5.82E-56	0.17	1.93	12.92	7.47E-38
Bayesian Factor 2	0.11	0.12	3.15	1.62E-03	0.03	0.04	1.80	7.19E-02
Bayesian Factor 3	0.23	0.52	6.64	3.37E-11	0.11	0.67	7.55	4.80E-14

Supplementary Table 3.4 Full sample: Combined (EAPS + IPS) regression on behavior. Left 3 columns show effect of EAPS + IPS, right 3 columns show EAPS effect after controlling for IPS.

Behavior	***EAPS + IPS effect***			***EAPS effect beyond IPS alone***		
	Var. Explained (%)	χ^2	pval	Var. Explained (%)	χ^2	pval
Crystallized Composite	4.15	360.91	4.26E-79	1.33	113.72	1.50E-26
Fluid Composite	1.09	93.46	5.08E-21	0.34	29.09	6.90E-08
Reading	3.01	260.70	2.46E-57	0.69	58.99	1.59E-14
Picture Vocabulary	2.93	253.05	1.12E-55	1.16	99.06	2.44E-23
Pattern	0.24	20.28	3.95E-05	0.08	7.25	7.09E-03
List	1.19	101.62	8.56E-23	0.34	28.68	8.54E-08
Picture	0.54	46.18	9.37E-11	0.27	23.25	1.42E-06
Flanker	0.24	20.50	3.54E-05	0.01	0.66	4.17E-01
Cardsort	0.28	23.91	6.44E-06	0.11	9.67	1.87E-03
Rey Auditory Verbal	1.25	107.51	4.51E-24	0.35	30.02	4.27E-08
Matrix Reasoning	1.46	124.94	7.40E-28	0.62	52.99	3.36E-13
Little Man Task	0.67	57.33	3.56E-13	0.10	8.26	4.05E-03
Bayesian Factor 1	3.88	337.50	5.16E-74	1.04	88.74	4.50E-21
Bayesian Factor 2	0.13	11.07	3.94E-03	0.01	1.11	2.91E-01
Bayesian Factor 3	0.95	81.48	2.03E-18	0.44	37.44	9.44E-10

Supplementary Table 3.5 European ancestry sample: behavioral associations between polygenic scores i) IPS and ii) EAPS in two separate regression for each behavior using GLMMs and controlling for variables of no interest.

	IPS				***EAPS***			
Behavior	coeff	Var. Explained (%)	t	pval	coeff	Var. Explained (%)	t	pval
Crystallized Composite	0.21	4.13	14.48	1.44E-46	0.18	3.34	12.95	9.36E-38
Fluid Composite	0.11	1.15	7.53	6.10E-14	0.09	0.89	6.60	4.66E-11
Reading	0.20	3.49	13.25	2.05E-39	0.14	2.03	10.04	1.64E-23
Picture Vocabulary	0.16	2.56	11.30	2.90E-29	0.16	2.71	11.64	6.71E-31
Pattern	0.05	0.24	3.41	6.65E-04	0.04	0.18	2.94	3.27E-03
List	0.12	1.21	7.72	1.41E-14	0.10	0.95	6.83	9.50E-12
Picture	0.07	0.38	4.29	1.84E-05	0.07	0.50	4.96	7.42E-07
Flanker	0.08	0.50	4.95	7.72E-07	0.04	0.12	2.37	1.80E-02
Cardsort	0.06	0.27	3.64	2.74E-04	0.05	0.25	3.49	4.83E-04
Rey Auditory Verbal	0.11	1.13	7.44	1.17E-13	0.10	1.03	7.11	1.32E-12
Matrix Reasoning	0.11	1.10	7.35	2.24E-13	0.11	1.08	7.27	4.10E-13
Little Man Task	0.09	0.74	6.03	1.75E-09	0.05	0.22	3.31	9.45E-04
Bayesian Factor 1	0.20	4.12	14.46	1.88E-46	0.16	2.75	11.73	2.43E-31
Bayesian Factor 2	0.05	0.22	3.27	1.09E-03	0.02	0.06	1.65	9.89E-02
Bayesian Factor 3	0.08	0.64	5.61	2.10E-08	0.10	0.90	6.64	3.37E-11

Supplementary Table 3.6 European ancestry sample: Combined (EAPS + IPS) regression on behavior. Left 3 columns show effect of EAPS + IPS, right 3 columns show EAPS effect after controlling for IPS.

Behavior	***EAPS + IPS effect***			***EAPS effect beyond IPS alone***		
	Var. Explained (%)	χ^2	pval	Var. Explained (%)	χ^2	pval
Crystallized Composite	5.80	291.74	4.46E-64	1.74	85.82	1.97E-20
Fluid Composite	1.58	78.02	1.15E-17	0.44	21.38	3.76E-06
Reading	4.33	216.21	1.12E-47	0.88	43.24	4.85E-11
Picture Vocabulary	4.08	203.45	6.62E-45	1.56	76.94	1.76E-18
Pattern	0.32	15.78	3.74E-04	0.08	4.17	4.11E-02
List	1.68	82.56	1.18E-18	0.47	23.08	1.55E-06
Picture	0.68	33.48	5.37E-08	0.30	14.96	1.10E-04
Flanker	0.52	25.47	2.94E-06	0.02	0.95	3.29E-01
Cardsort	0.40	19.76	5.13E-05	0.13	6.46	1.10E-02
Rey Auditory Verbal	1.66	82.04	1.53E-18	0.54	26.67	2.41E-07
Matrix Reasoning	1.68	82.92	9.89E-19	0.59	28.93	7.51E-08
Little Man Task	0.80	39.07	3.28E-09	0.05	2.64	1.04E-01
Bayesian Factor 1	5.39	270.80	1.57E-59	1.33	65.36	6.25E-16
Bayesian Factor 2	0.23	11.28	3.55E-03	0.01	0.56	4.53E-01
Bayesian Factor 3	1.21	59.41	1.26E-13	0.57	27.83	1.32E-07

Supplementary Table 3.7 Diverse ancestry sample: behavioral associations between polygenic scores i) IPS and ii) EAPS in two separate regression for each behavior using GLMMs and controlling for variables of no interest.

	IPS				***EAPS***			
Behavior	coeff	Var. Explained (%)	t	pval	coeff	Var. Explained (%)	t	pval
Crystallized Composite	0.40	1.47	7.34	2.68E-13	0.15	1.21	6.66	3.24E-11
Fluid Composite	0.20	0.32	3.41	6.52E-04	0.08	0.32	3.38	7.28E-04
Reading	0.39	1.21	6.64	3.61E-11	0.12	0.66	4.90	1.01E-06
Picture Vocabulary	0.32	0.89	5.70	1.30E-08	0.14	1.02	6.11	1.13E-09
Pattern	0.10	0.07	1.58	1.14E-01	0.05	0.10	1.88	6.08E-02
List	0.26	0.47	4.13	3.64E-05	0.09	0.30	3.30	9.87E-04
Picture	0.12	0.09	1.83	6.76E-02	0.08	0.25	2.98	2.92E-03
Flanker	0.09	0.05	1.39	1.64E-01	0.01	0.01	0.44	6.58E-01
Cardsort	0.10	0.07	1.59	1.13E-01	0.06	0.13	2.19	2.83E-02
Rey Auditory Verbal	0.28	0.56	4.53	6.13E-06	0.08	0.23	2.90	3.70E-03
Matrix Reasoning	0.26	0.50	4.28	1.93E-05	0.14	0.87	5.62	2.07E-08
Little Man Task	0.21	0.31	3.37	7.64E-04	0.08	0.24	2.97	2.96E-03
Bayesian Factor 1	0.40	1.57	7.59	4.10E-14	0.14	1.03	6.12	1.04E-09
Bayesian Factor 2	0.06	0.03	0.96	3.39E-01	0.02	0.02	0.74	4.57E-01
Bayesian Factor 3	0.19	0.27	3.14	1.73E-03	0.09	0.32	3.39	7.03E-04

Supplementary Table 3.8 Diverse ancestry sample: Combined (EAPS + IPS) regression on behavior. Left 3 columns show effect of EAPS + IPS, right 3 columns show EAPS effect after controlling for IPS.

Behavior	***EAPS + IPS effect***			***EAPS effect beyond IPS alone***		
	Var. Explained (%)	χ^2	pval	Var. Explained (%)	χ^2	pval
Crystallized Composite	2.27	83.46	7.54E-19	0.81	29.70	5.04E-08
Fluid Composite	0.54	19.64	5.44E-05	0.22	7.94	4.84E-03
Reading	1.60	58.42	2.06E-13	0.39	14.33	1.53E-04
Picture Vocabulary	1.62	59.33	1.31E-13	0.73	26.79	2.27E-07
Pattern	0.14	5.16	7.56E-02	0.07	2.65	1.03E-01
List	0.65	23.75	6.95E-06	0.19	6.64	9.97E-03
Picture	0.29	10.56	5.10E-03	0.20	7.24	7.12E-03
Flanker	0.05	1.99	3.69E-01	0.00	0.04	8.43E-01
Cardsort	0.17	6.35	4.17E-02	0.10	3.80	5.12E-02
Rey Auditory Verbal	0.69	25.04	3.65E-06	0.12	4.51	3.36E-02
Matrix Reasoning	1.17	42.75	5.22E-10	0.67	24.39	7.86E-07
Little Man Task	0.47	17.19	1.85E-04	0.16	5.79	1.61E-02
Bayesian Factor 1	2.21	81.27	2.25E-18	0.66	23.84	1.05E-06
Bayesian Factor 2	0.03	1.27	5.30E-01	0.01	0.34	5.57E-01
Bayesian Factor 3	0.50	18.16	1.14E-04	0.23	8.35	3.86E-03

References

1. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys. 1974 Jan;16(1):143–9.

2. Daniel MH, Wahlstrom D, Zhang O. Equivalence of Q-interactive™ and Paper Administrations of Cognitive Tasks: WISC ®-V Q-interactive Technical Report 8. 2014.

3. Acker W. A computerized approach to psychological screening - The Bexley-Maudsley Automated Psychological Screening and The Bexley-Maudsley Category Sorting Test. Int J Human-computer Stud ⁄ Int J Man-machine Stud. 1982;17:361–9.

4. Weschler D. Weschler Intelligence Scale for Children, 5th ed. V. Bloomington, MN: Pearson; 2014.

5. Baurley JW, Edlund CK, Pardamean CI, Conti D V., Bergen AW. Smokescreen: A targeted genotyping array for addiction research. BMC Genomics [Internet]. 2016;17(1):1–12. Available from: http://dx.doi.org/10.1186/s12864-016-2495-7

6. Raj A, Stephens M, Pritchard JK. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89.

7. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet [Internet]. 2016 Aug 29;48:1284. Available from: https://doi.org/10.1038/ng.3656

8. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience [Internet]. 2015;4(1). Available from: https://doi.org/10.1186/s13742-015-0047-8

9. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466–8.

10. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50(7):912–9.

11. Zacher M, Nguyen-viet TA, Bowers P, Sidorenko J, Linnér RK. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. 2018;50(August).

12. Thompson WK, Barch DM, Bjork JM, Gonzalez R, Nagel BJ, Nixon SJ, et al. The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. Dev Cogn Neurosci [Internet]. 2018;(December):100606. Available from: https://doi.org/10.1016/j.dcn.2018.12.004

13. Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen S. NIH Toolbox Cognitive Function Battery (CFB): Composite Scores of Crystallized, Fluid, and Overall Cognition. Monogr Soc Res Child Dev. 2013;78(4):119–32.

14. Baron R, Kenny D. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. J Peronality Soc Psychol. 1986;51(6):1173–82.

15. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. J Stat Softw. 2014;59(5):1–38.

Chapter 4: *Unique prediction of developmental psychopathology from genetic and familial risk*

4.1 *Abstract*

Early detection is critical for easing the rising burden of psychiatric disorders. However, the specificity of psychopathological measurements and genetic predictors is unclear among youth. We measured associations between genetic risk for psychopathology (polygenic risk scores (PRS) and family history (FH) measures) and a wide range of behavioral measures in a large sample (n=5204) of early adolescent participants (9-11 years) from the Adolescent Brain and Cognitive Development (ABCD) Study[SM]. Associations were measured both with and without taking into consideration shared variance across measures of genetic risk. Polygenic risk for Attention Deficit Hyperactivity Disorder (ADHD) and depression (DEP) shared many significant associations with externalizing, internalizing and psychosis-related behaviors. However, when accounting for all measures of genetic and familial risk these two PRS also showed clear, unique patterns of association: the DEP PRS showed significantly stronger associations with somatic complaints and depression symptoms; whereas the ADHD PRS showed stronger associations with ADHD symptoms, impulsivity and prodromal psychosis. The Schizophrenia PRS showed a unique negative association with performance on cognitive tasks measuring fluid abilities, such as working memory and executive function, that was not accounted for by other measures of genetic risk. FH accounted for unique

variability in behavior above and beyond PRS and vice versa with FH measures explaining a greater proportion of unique variability compared to the PRS. Our results indicate that, among youth, many behaviors show shared genetic influences; however, there is also specificity in the profile of emerging psychopathologies for individuals with high genetic risk for particular disorders. This may be useful for quantifying early, differential risk for psychopathology in development.

4.2 Introduction

Psychiatric disorders place a huge burden on those affected, their families and society. Identifying risk for psychopathology in developmental samples may offer an opportunity for early detection and intervention. Nearly all psychiatric disorders have a heritable component, with twin heritability estimates ranging from 33-84% across affective, psychotic and developmental disorders[1]. Lifetime prevalence rates of several disorders are higher among first degree biological relatives of individuals with a diagnosis compared to families of individuals with no diagnosis[2]. Therefore, estimating genetic liability for psychiatric disorders presents one avenue for identifying at risk individuals and probing differential and transdiagnostic risk factors. Here we sought to determine: 1) if increased genetic risk within a large, typically developing and demographically diverse developmental sample would be associated with symptoms of psychopathology, related individual difference factors, and cognitive function; and, 2) whether there was any evidence for specificity in the behavioral measures predicted by different genetic markers.

Large-sample analyses of results from genome-wide association studies (GWAS) have revealed the highly polygenic architecture of complex behavioral phenotypes, with many variants in the genome additively accounting for substantial heritability, but individually exerting only very small effects. Models using effect sizes at single nucleotide polymorphisms (SNPs)

estimated from large-scale independent GWAS, can be used to compute polygenic risk scores (PRS), which are aggregate scores of an individual's genetic risk for a trait. Notably recent powerful, cross-disorder meta-analyses [3,4] reveal high genetic correlation and widespread pleiotropy across psychiatric disorders, consistent with overlapping genetic architecture. Indeed, polygenic risk for depression has been shown to positively associate with childhood psychopathology across behavioral domains[5].

Family history (FH) is a clinically used factor for predicting psychiatric risk[6], yet there has been a lack of direct comparisons of associations between PRS and FH of psychopathology in childhood and adolescence. SNP heritabilities (h^2_{SNP}) based on effects across the genome are lower than twin heritabilities, which suggests there are genetic factors driving psychiatric phenotypes that are not fully captured with common variants at current GWAS sample sizes. Indeed, FH likely reflects a complex combination of genetic and environmental factors. Due to the differential information that PRS and FH measures may provide, it is important to determine whether they explain independent or overlapping variance in developmental psychopathology and cognition. For example, in large cohorts, both family history of schizophrenia [7] and polygenic risk for schizophrenia [8] have been shown to associate with developmental psychopathology; however, the unique contribution of polygenic risk above and beyond FH is unclear.

For this study we used behavioral and genetic data from 9-11 year-old children from the Adolescent Brain and Cognitive Development (ABCD) Study[SM]. We generated five PRS that were trained on large independent datasets. We used these PRS and measures of FH of psychopathology both independently and within the same models to predict a large array of both caregiver and youth-reported phenotypes thought to reflect behavioral risk for developing psychiatric disorders. These measures included both dimensional and diagnostic assessments of psychopathology, individual difference measures of impulsivity and behavioral approach and

inhibition, prodromal psychosis and behaviors associated with mania and prosocial behavior. We additionally measured associations with cognitive measures from the NIH Toolbox given documented associations and genetic overlap between cognitive impairment and schizophrenia and bipolar disorder[9,10]. Using this approach, we aimed to uncover variability across early signs of psychopathology across domains that is uniquely associated with each genetic/familial predictor. This research is an essential first step in this large longitudinal study to determine whether we can identify early signs of specificity in genetic-behavior associations in development, which can then be tracked to determine their potential predictive power for future diagnoses.

4.3 Methods and Materials

4.3.1 Sample

The ABCD study is a longitudinal study across 21 data acquisition sites in the United States following 11,880 children starting at 9-11 years. This paper uses baseline data from the NIMH Data Archive ABCD Collection Release 2.0.1 (DOI: 10.15154/1504041). The ABCD cohort was recruited to ensure the sample was as close to nationally representative as possible, therefore the cohort includes individuals across different racial and ethnic backgrounds and socioeconomic groups [11]. There is an embedded twin cohort and many siblings. As the chosen PRS were all trained on European individuals, the main associations in this study were conducted in a predominantly European ancestry sample (n=5204). Supplementary analyses were conducted in those with non-European ancestry (n=3964) and the full sample (n=9168). Table 4.1 outlines the demographics of the European sample used in the main analysis. Prevalence rates of diagnoses based on the Kiddie

Schedule for Affective Disorders and Schizophrenia (KSADS) semi-structured interview for the caregiver and youth are in Supplementary Table 4.1& Supplementary Table 4.2.

4.3.2 ABCD Baseline Mental Health Battery

The Mental Health Battery in ABCD is an extensive battery of questionnaires and semi-structured interviews assessing diagnostic and dimensional measures of psychopathology and individual difference factors. Both youth and their caregivers provided responses at baseline using divergent and overlapping measures. Motivation behind selecting these assessments is outlined here: [12]. Supplementary Table 4.3 lists variables used from the ABCD public release.

4.3.2.1 ***DIAGNOSTIC ASSESSMENTS***

Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS). Participants completed a semi-structured, self-administered, computerized version of the validated and reliable KSADS-5[13]. Research Assistants had extensive training to support youth completing this assessment. Caregivers and youth completed modules on depression, bipolar disorder, generalized anxiety disorder, social anxiety disorder, suicidality and sleep. Only caregivers completed psychosis, obsessive-compulsive disorder (OCD), ADHD, oppositional defiant disorder (ODD), conduct disorder (CD), panic disorder and eating disorders modules. Symptom scores were the sum of lifetime symptoms endorsed in each module and were scored and analyzed separately for each respondent. The total symptom score was a sum across modules.

4.3.2.2 *DIMENSIONAL ASSESSMENTS*

Child Behavior Checklist (CBCL). Caregiver-reported CBCL[14] has eight syndrome scales: anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule breaking behavior and aggressive behavior, and a total problems score.

General behavior inventory. Caregiver-report ten-item Mania Scale[15] derived from the 73-item General Behavior Inventory (PGBI) for Children and Adolescents[16].

Prosocial Behavior Survey. Caregivers and youth were asked three questions about how helpful and considerate the youth was in general, with summed scores for both caregiver and youth.

Prodromal Questionnaire Brief (PQ-B). Youth-report measure, modified for use in children in our age range, consisting of a 21-item scale assessing subclinical manifestations of psychosis[17–19]. The prodromal psychosis severity score is the sum of the number of symptoms endorsed weighted by how distressing the symptoms were.

UPPS-P for children short scale. Youth-report impulsive behavior scale, which includes five sub-scales that measure four factors of impulsivity: positive and negative urgency, lack of perseverance, premeditation, and sensation seeking[20].

Behavioral inhibition and behavioral activation (BISBAS scale). Youth-report measure of approach and avoidance behaviors [21,22] that produces scores for drive, fun seeking, reward responsiveness, and behavioral inhibition.

NIH Toolbox Cognition Battery®. Widely used battery of cognitive tests that measures a range of different cognitive domains[23–25]. We analyzed the uncorrected composite scores broadly measuring fluid and crystallized intelligence that are generated from the NIH Toolbox® and have been validated against gold-standard measures [26–28]. The fluid composite score

includes performance on the flanker task, picture sequence memory task, list sorting memory task, pattern comparison processing speed and dimensional change card sort task. The crystallized composite score includes performance on the oral reading recognition task and picture vocabulary task.

4.3.3 Genetic & Familial Measures

4.3.3.1 ***Family History Assessment***

Caregivers were given a questionnaire asking about family history (FH) of 10 behaviors associated with psychopathology: alcohol use; drug use; depression; mania; psychosis; conduct problems; nerves; seen a therapist; hospitalized for a mental health problem; and, suicide. For each question the caregivers were asked specifically if any *blood* relative had experienced any of the described behaviors (more detail in Supplementary Table 4.4). Importantly, these variables do not indicate clinical diagnoses associated with these behaviors.

4.3.3.2 **Polygenic Risk Scores (PRS)**

PRS were estimated from summary statistics for ADHD[29], Autism Spectrum Disorder (ASD)[30], Bipolar Disorder (BPD)[31], Schizophrenia (SCZ)[32] and Depression (DEP)[33] from the Psychiatric Genetics Consortium (https://www.med.unc.edu/pgc/results-and-downloads). Additional details of preprocessing genetic data and PRS estimation are in supplementary materials.

4.3.4 Statistical Analysis

Generalized Linear Models (GLMs) were fit to measure the association between i) each of the 41 behavioral phenotypes and ii) FH and PRS. Univariate models included one independent variable of interest (PRS or FH) in each model (i.e. behavior ~ PRS_i + covariates or behavior ~ FH_i + covariates). Multivariable models included all PRS and FH measures in the same model (i.e. behavior ~ PRS_1 + PRS_2 … + FH_1 + FH_2 … + covariates). Each behavioral phenotype was modelled separately with the same set of predictors and covariates. Fixed nuisance covariates included age, sex, top 10 genetic principal components, household income, highest parental education and data collection site. ΔR^2 was reported as change in R^2 from a reduced model (covariates only) to a full model (including the predictor of interest)[34,35]. Supplementary analyses were conducted without controlling for household income and parental education to understand the impact of socioeconomic status (SES). Family relatedness was controlled for using a random subset of the sample that only included singletons. Significant associations were determined using a false discovery rate (FDR) and reported p-values are FDR adjusted (p-adj). See appendix for additional analysis details. Additional models were implemented to measure pairwise spearman correlations across all of the DVs and IVs in the European ancestry sample after residualizing for the covariates of no interest (Supplementary Figure 4.3 & Supplementary Figure 4.4). Behavioral measures were categorized by behavioral domain (see Supplementary Table 4.3) in order to determine whether associations with each genetic predictor were enriched for measures within domains.

4.4 *Results*

4.4.1 Unique behavioral associations with PRS across domains

In the univariate models (measuring the association between each PRS and each behavioral variable in separate models), controlling for SES, the ADHD and DEP PRS showed the largest and greatest number of associations across internalizing, externalizing and psychosis-related measures (Figure 4.1, left panel). The ADHD PRS significantly associated with CBCL rule-breaking (ΔR^2=0.0071, p-adj=5.7x10^{-6}), inattentive (ΔR^2=0.0064, p-adj=4.6x10^{-8}) and aggressive (ΔR^2=0.0031, p-adj=5.2x10^{-4}) behaviors, prodromal psychosis severity (ΔR^2=0.0062, p-adj=2.4x10^{-5}), and caregiver reported KSADS oppositional/conduct disorder (ΔR^2=0.0042, p-adj=7.7x10^{-5}) and ADHD (ΔR^2=0.0030, p-adj=9.2x10^{-4}) symptoms, followed multiple youth and caregiver reported measures of impulsivity, depression and suicidality symptoms, bipolar and psychosis related measures and developmental social problems. The DEP PRS showed largest significant associations with CBCL somatic complaints (ΔR^2=0.0054, p-adj=1.9x10^{-6}), KSADS symptoms of oppositional/conduct disorder (ΔR^2=0.0039, p-adj=1.6x10^{-4}) and CBCL anxious/depressive (ΔR^2=0.0031, p-adj=2.5x10^{-4}), aggressive (ΔR^2=0.0031, p-adj=5.1x10^{-4}), and rule-breaking (ΔR^2=0.0029, p-adj=4.8x10^{-6}) behaviors. These were followed by caregiver reported KSADS symptoms of suicidality (ΔR^2=0.0027, p-adj=1.8x10^{-3}), bipolar disorder (ΔR^2=0.0027, p-adj=2.0x10^{-3}) and anxiety (ΔR^2=0.0020, p-adj=8.4x10^{-3}) and youth reported KSADS depression symptoms (ΔR^2=0.0027, p-adj=2.2x10^{-3}), as well as other measures of negative urgency, developmental social problems, behavioral inhibition and bipolar and psychosis related behaviors. Both the ADHD and DEP PRS were also associated with the CBCL Total Problems and KSADS Total Symptoms scores.

The BPD and SCZ PRS were not significantly associated with any bipolar or psychosis-related measures; however, they did significantly associate with CBCL rule-breaking with a smaller effect size compared to ADHD and DEP (BPD: ΔR^2=0.0016, p-adj=4.3x10^{-2}; SCZ: ΔR^2=0.0022, p-adj=1.5x10^{-2}). In addition, the SCZ PRS negatively associated with the fluid composite score from the NIH Toolbox® (ΔR^2=0.0027, p-adj=2.2x10^{-3}). The ASD PRS was associated with CBCL inattention (ΔR^2=0.0022, p-adj=1.9x10^{-2}), youth reported KSADS total symptoms scale (ΔR^2=0.0028, p-adj=1.4x10^{-2}) and the CBCL total symptoms scale (ΔR^2=0.0014, p-adj=3.1x10^{-2}).

Multivariable models determined the specificity of these associations by covarying for all PRS and FH predictors simultaneously. In these models, PRS associations were attenuated and showed greater specificity for the ADHD and DEP PRS (Figure 4.1, right panel). The ADHD PRS predicted unique variance across externalizing and psychosis-related measures not predicted by other measures of genetic risk, whereas the DEP PRS predicted unique variance across a different set of internalizing and externalizing behaviors. Additionally, in the multivariable models, only the ADHD PRS was significantly associated with CBCL rule-breaking. The ASD PRS also no longer showed any significant associations. The SCZ PRS association with the fluid composite score from the NIH Toolbox® remained significant when controlling for all other measures of genetic risk (ΔR^2=0.0029, p-adj=7.7x10^{-3}).

When not controlling for SES, behavioral associations were slightly larger and the overall pattern of associations was similar (Supplementary Figure 4.5, Supplementary Figure 4.6). However, there were additional significant associations with cognitive measures for the ADHD and BPD PRS. The SCZ association with the fluid composite score was the only association with cognitive performance that was robust to controlling for SES.

We categorized each behavior into a domain based on the construct measured in order to highlight the different types of behavioral measures predicted by each PRS (Figure 4.2;

variables in each domain are highlighted in Supplementary Table 4.3). Across both the univariate and multivariable models, the largest associations with the ADHD PRS were with externalizing and psychosis-related measures; whereas for the DEP PRS, associations encompassed a mix of internalizing and externalizing measures. In the multivariable models, the specificity in the unique pattern of behaviors predicted by these PRS across domains was clarified due to the removal of any shared variance across the genetic predictors.

4.4.2 Unique behavioral associations with FH across domains

Behavioral associations with FH measures were larger than with PRS (Figure 4.3, left panel). Given the large number of overlapping univariate associations, we will focus on the associations from the multivariable models. When controlling for all other genetic predictors (PRS and FH) in the multivariable models, FH of conduct problems, depression and anxiety/stress showed the largest effects with some specificity across the behavioral measures (Figure 4.3, right panel). FH of conduct problems significantly associated with the CBCL subscales particularly with rule-breaking (ΔR^2=0.0081, p-adj=1.8x10^{-5}), as well as KSADS symptoms related to both externalizing and internalizing disorders (ΔR^2range=0.0023-0.0072), and mania (ΔR^2=0.0057, p-adj=8.0x10^{-3}). FH of depression significantly associated with the total problems scales from the CBCL (R2=0.0046, p-adj=4.5x10^{-4}) and KSADS (ΔR^2=0.0043, p-adj=5.8x10^{-4}), as well as internalizing and externalizing measures across the KSADS and CBCL (ΔR^2range=0.0018-0.0040). This pattern was similar to DEP PRS, however, unlike the DEP PRS, FH of depression only associated with caregiver-reported measures in the multivariable models. FH of anxiety/stress showed several associations across domains with the largest effects for caregiver-reported KSADS anxiety symptoms (ΔR^2=0.0081, p-adj=5.8x10^{-7}) and the CBCL anxious/depressive subscale (ΔR^2=0.0071, p-adj=5.8x10^{-7}).

FH of use of professional health services was most strongly associated with CBCL somatic complaints (ΔR^2=0.0040, p-adj=5.6x10^{-4}), thought problems (ΔR^2=0.0028, p-adj=7.0x10^{-3}) and the total problem score (ΔR^2=0.0036, p-adj=2.2x10^{-3}), and also showed a positive association with the crystallized composite score (ΔR^2=0.0027, p-adj=1.0x10^{-2}). Interestingly, when controlling for all other measures of genetic risk, FH of drug and alcohol abuse associated with differential behaviors, with FH of drug abuse explaining unique variance in CBCL rule-breaking (ΔR^2=0.0034, p-adj=1.2x10^{-2}) and KSADS PTSD symptoms (ΔR^2=0.0029, p-adj=7.6x10^{-3}), and FH of alcohol abuse explaining unique variance in CBCL social problems (ΔR^2=0.0020, p-adj=4.9x10^{-2}) and anxious/depressive behaviors (ΔR^2=0.0017, p-adj=3.4x10^{-2}). FH of hospitalization showed several negative associations with caregiver-reported internalizing behaviors, which were positive in the univariate models. This sign flip of effects may be due to collinearity across the genetic risk measures (Supplementary Figure 4.4) when used in a single model.

Figure 4.4 displays the enrichment of FH associations by behavioral domain. For the univariate models, the FH measures associated with behaviors across several domains. These patterns became more specific towards particular domains in the multivariable models. For example, FH of depression or anxiety/stress were significantly associated with internalizing behaviors, whereas FH of conduct disorder was significantly associated with externalizing behaviors.

Finally, we quantified the variance in each behavior predicted by the set of PRS and set of FH measures when controlling for the other set of genetic predictors. Supplementary Table 4.5 shows that, in all cases, each set independently predicted unique variance over and above the other set of genetic predictors. The maximum variance explained by the FH and PRS measures combined was ΔR^2=0.062 of CBCL Total Problems scale, of which ΔR^2=0.053 was uniquely predicted by FH and ΔR^2=0.0061 was uniquely predicted by PRS. The maximum

unique variance explained collectively by PRS was ΔR^2=0.0071 of the variability in oppositional/conduct disorder symptoms. These results further demonstrate that PRS and FH predict unique, non-overlapping variance across different domains of behavior in youth with PRS predicting a smaller proportion of variability than FH.

4.5 *Discussion*

Polygenic risk and FH of psychopathology predict both overlapping and unique variability in behavior across domains in 9-11 year old youth. Several externalizing and internalizing behaviors were associated with multiple measures of genetic risk highlighting shared genetic influences underlying variability in developmental psychopathology. However, when controlling for shared variance across PRS and FH measures, polygenic risk for ADHD and depression predicted unique variance across differential externalizing, internalizing and psychosis-related behaviors with the strongest associations for ADHD and depression symptomatology. Moreover, the SCZ PRS specifically and uniquely predicted variability in cognitive performance. This highlights that these PRS are signaling differential behavior related to specific disorders. FH of psychopathology explained additional unique variance in behavior, independent of the PRS, indicating additional genetic and environmental influences on behavior and recapitulating results in adults demonstrating the complementary information provided by PRS and FH[36,37]. Using the combined information across these genetic and familial measures and the dense behavioral phenotyping in the ABCD study, we have identified several, specific patterns of behavior associated with genetic risk for psychopathology that may be useful for quantifying early risk across different disorders during development.

Of the PRS analyzed, the ADHD and DEP PRS showed univariate associations across largely overlapping behavioral measures. In particular, both PRS predicted variability in externalizing behaviors (e.g. rule-breaking, aggression and conduct problems), internalizing

behaviors (e.g. suicidality and youth reported depression), psychosis-related behaviors (e.g. prodromal psychosis, bipolar symptoms and thought problems), and inattentive and social problems. Given the correlation between behavioral problems in youth, this supports evidence that these frequently comorbid behaviors across different domains have shared genetic influences[5,38,39]. This indicates a common pathway that may contribute to the development of psychopathology. Indeed, suicidality and depression are common across individuals with several different psychiatric disorders and there is evidence that externalizing behaviors in childhood may indicate risk for both externalizing and internalizing disorders in adulthood[40,41]. However, there was some specificity in the behaviors predicted by the ADHD and DEP PRS. The ADHD PRS specifically associated with behavioral approach subscales, impulsivity, ADHD symptoms and mania; whereas the DEP PRS associated with somatic complaints, insomnia, anxiety symptoms, impulsivity specifically in response to negative affect and behavioral inhibition. When controlling for other measures of genetic risk in the same model, the PRS-behavior associations became more specific. These findings highlight potentially distinct pathways associated with the development of these unique profiles of behaviors.

Our results replicated previous findings, with a similar magnitude of effects, showing that ADHD PRS significantly associated with hyperactive and inattentive traits in a developmental sample[42–45]. Across the PRS, ADHD and ASD were moderately correlated, and when controlling for the other genetic predictors ASD no longer associated with behavioral problems on the CBCL highlighting the genetic overlap between these disorders in development[46]. There may be additional factors that contributed to the lack of unique associations of ASD PRS to youth behaviors. Not attending mainstream school classes and an inability to carry out the ABCD protocol, which includes a two-hour MRI scan, was an exclusion criterion; therefore, many individuals with low functioning ASD would have been ineligible for the study. This suggests that the prevalence of ASD symptoms in the ABCD

cohort is likely small and restricted to only part of the autism spectrum, which may have a larger overlap with ADHD. Moreover, only 17% of the phenotypic variance in ASD is thought to be attributable to common genetic variants at current sample sizes[47] and our PRS was not sensitive to important rare chromosome deletions in ASD[48].

Interestingly, in our sample, the ADHD PRS predicted many bipolar-related behaviors and psychotic-like symptoms. Symptom profiles for pediatric BPD and ADHD are very similar and there is high comorbidity across these disorders[49]. Other studies have shown that childhood ADHD is often a premorbidity for the later development of schizophrenia and relatives of individuals with schizophrenia have higher rates of ADHD than the general population[50–52]. Given the low correlation between ADHD, SCZ and BPD PRS in this study, the ADHD PRS may highlight individuals at risk for developing psychosis-related disorders that may be etiologically distinct from those with high SCZ or BPD scores.

Despite previous studies showing that the SCZ PRS associates with several markers of general psychopathology in adolescence[8,43,53,54], we did not find any associations of SCZ or BPD PRS with psychopathology in our models. This could be driven by differences in the statistical approach, the demographics of the samples or the phenotypes measured – which can impact the stability of results across adolescent samples[55]. Nevertheless, as hypothesized, we did identify a significant negative association between the SCZ PRS and the fluid composite score from the NIH Toolbox®, which remained after controlling for sociodemographic factors and was unique to the SCZ PRS. Cognitive impairment is a core feature of several psychiatric disorders, particularly those that include psychotic symptoms. Neurodevelopmental studies have highlighted premorbid cognitive impairment across domains in patients with schizophrenia and bipolar disorder[56,57]. Indeed, there is a large genetic overlap across schizophrenia, bipolar disorders and general intelligence[9,10], which suggests there are shared etiological mechanisms that affect psychopathology and cognition.

Perturbations in the cognitive domain may indicate early risk for schizophrenia more readily than other behavioral manifestations.

There were differences in associations across caregiver and youth reported behaviors, particularly with genetic risk for depression. In the multivariable models, youth-reported depression symptom scores were more associated with the DEP PRS, whilst caregiver-reported depression was associated with a FH of depression. Informant discrepancies between caregiver and child-reported measures have been widely reported[58] and we see relatively low correlations between youth and caregiver reported measures in the current study. Negative biases from caregivers, particularly due to caregiver depression, can also impact behavioral reports[16,59]. An awareness of a history of depression within the youth's family may have biased the informant's report about the youth's depression, leading to a stronger relationship of FH of depression with caregiver compared to youth reported measures. Future time points are required to delineate which informant-reported measures are more accurate at predicting later diagnoses.

FH of anxiety/stress and conduct problems showed the greatest number of associations across different behavioral domains, supporting a role for anxiety or sensitivity to stress and delinquent behavior as transdiagnostic traits. However, in the multivariable models in particular, there were subtle differences in the pattern of FH-behavior associations across domains. For example, FH of drug abuse explained unique variance in rule-breaking behaviors; whereas FH of alcohol abuse explained unique variance in social problems and anxious/depressive behaviors. This highlights that the pattern of behaviors associated with a predictor when accounting for other measures of genetic risk is likely important for understanding predictive specificity across disorders. Inherent to FH measures are implicit genetic and environmental influences that are difficult to separate. It remains to be seen whether the additional variance in behavior explained by FH measures above and beyond PRS

reflects environmental or additional genetic influences. Together FH and PRS measures predicted ~6% of the variability in the CBCL Total Problems score. These analyses highlight the utility of measuring multiple markers of genetic risk.

There are several limitations in the current study. PRS association strength is limited by the phenotype's heritability and the training sample used[60,61]. DEP had the largest discovery sample (Supplementary Figure 4.2) and the lowest SNP heritability, yet displayed some of the largest associations in our sample. This may be due to depression having relatively greater population prevalence compared to the other psychiatric disorders measured, therefore compared to other disorders, risk alleles may be well represented in our sample. The correlations between the PRS generated in this study were much lower than the genetic correlations determined in the original GWAS, which may be because this cohort is not enriched for individuals with risk alleles. Many psychiatric disorders have increased penetrance during adolescence, therefore the lack of variance in psychopathology symptoms at this age may explain the limited associations between behavior and the SCZ/BPD PRS. Moreover, the GWAS used to produce the PRS in this study were all conducted on European samples. The ABCD sample is demographically diverse, however PRS trained in different ancestry groups do not validly predict phenotypes in admixed or different ancestry samples. This highlights the limited predictive capacity of European-only GWAS for admixed populations and emphasizes the need for conducting GWAS in different ancestry groups. Finally, the magnitude of the genetic-behavior effects detected was very small; the development of psychopathology is extremely complex and genetic risk as estimated with polygenic predictors appears to only account for a small proportion of variability in behavior at this age.

Here we have shown that different PRS and FH measures predicted unique patterns of symptoms of psychopathology, related individual difference factors and cognitive function in a

large sample of 9 to 11-year-old children. The unique associations controlling for other genetic measures provides encouraging evidence that genetic data may be useful alongside FH in identifying specific risk for psychiatric disorders. Longitudinal analyses will further elucidate the specificity of these associations and can track these coexisting patterns of behavior to determine the differential predictive utility for PRS and FH measures.

Acknowledgments

The authors wish to thank the youth and families participating in the Adolescent Brain Cognitive Development (ABCD) Study and all ABCD staff. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/Consortium_Members.pdf. ABCD consortium investigators designed and implemented the study and/or provided data but did not all necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The data was downloaded from the NIMH Data Archive ABCD Collection Release 2.0.1 (DOI:

10.15154/1504041). C.F. was supported by grant R01MH122688 funded by the National Institute for Mental Health (NIMH).

Chapter 4, in full, is available as a preprint on MedRxiv and has been submitted for publication. Loughnan R, Palmer C, Makowski C, Thompson W, Barch D, Jernigan T, Dale A, Fan C. The dissertation/thesis author was the joint primary investigator and author of this paper.

Conflict of Interest

A.M.D. reports that he was a Founder of and holds equity in CorTechs Labs, Inc., and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc. He receives funding through research grants from GE Healthcare to UCSD. The terms of these arrangements have been reviewed by and approved by UCSD in accordance with its conflict-of-interest policies. No other authors report a conflict of interest.

Table 4.1 ***Sociodemographic breakdown for the European ancestry sample analyzed in this study***. All main analyses were conducted with a European ancestry only sample. Compared to the full ABCD sample had higher proportions of individuals from households with higher income and a higher parental education level and self-identifying as White. Age, sex, household income, parental education, data collection site and the top 10 genetic principal components were controlled for in the main analyses. Supplementary analyses were conducted without controlling for SES and in the Full sample and Non-European ancestry sample.

	European Ancestry Sample
Total N	5,204
Age - months	119.18 (7.48)
Sex	
M	2744 (52.7)
Parental Education	
< HS Diploma	22 (0.4)
HS Diploma/GED	149 (2.9)
Some College	963 (18.5)
Bachelor	1650 (31.7)
Post Graduate Degree	2420 (46.5)
Household Income	
[<50K]	642 (12.3)
[>=50K & <100K]	1590 (30.6)
[>=100K]	2972 (57.1)
Race Ethnicity	
White	4934 (94.9)
Black	1 (0.0)
Hispanic	137 (2.6)
Asian	0 (0.0)
Other	126 (2.4)

Figure 4.1 ***Univariate (left) and multivariable (right) associations for each behavioral phenotype predicted by the PRS.*** Effect sizes for each association are displayed as the partial variance explained, R2, (as a percentage) multiplied by the sign of the beta estimate to indicate the magnitude and sign of the association (red=positive, blue=negative). Each row represents a model with the dependent variable along the y-axis and each PRS on the x-axis. In the univariate models (left) only a single genetic/familial predictor was included in each model (each cell = 1 model) – i.e. behavior ~ PRS + covariates. In the multivariable models (right) all genetic/familial predictors were included in each model including all PRS and FH measures (each row = 1 model) – *i.e.* behavior ~ PRS_1 +PRS_2 … +FH_1 + FH_2 … + covariates. Along the x-axis from left to right: the five PRS measured (Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorder (ASD), Bipolar Disorder (BPD), Depression (DEP), Schizophrenia (SCZ)). All models controlled for covariates of age, sex, the top 10 Principal Components of the genetic data, household income, highest parental education and data collection site. Effects represent the median across 100 subsamples of singletons to control for family relatedness. Dots indicate FDR significant associations.

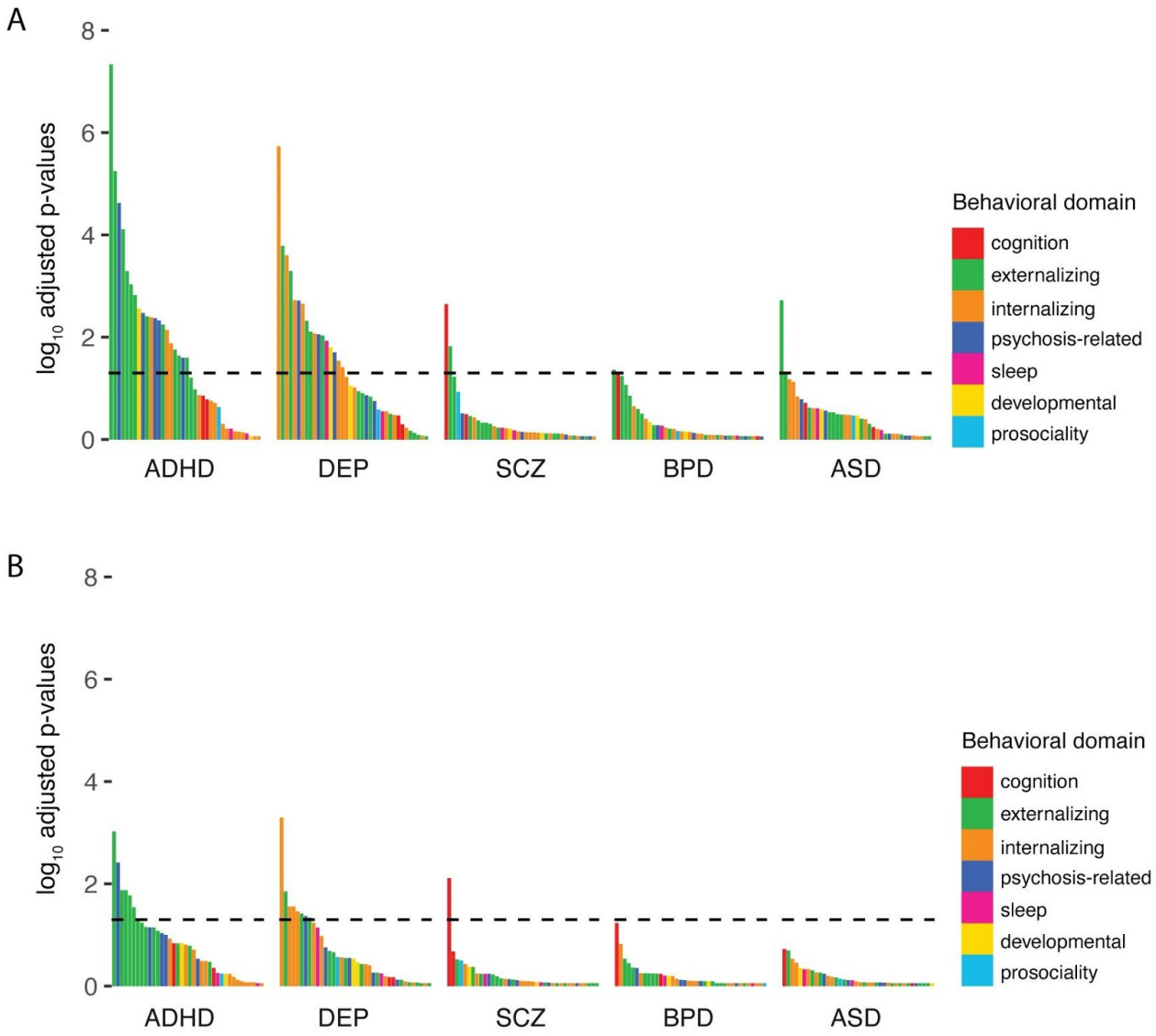


Figure 4.2 ***Enrichment of PRS associations across different behavioral domains.*** FDR adjusted p-values (logged) for all of the associations shown in Figure 4.1 for both the univariate (A) and multivariable models (B). Each association is colored based on the behavioral domain the dependent variable was categorized within (see Supplementary Table 4.3). Associations are ordered with the most significant effect to the far left. The horizontal line represents the FDR adjusted significance threshold (p=0.05). All models controlled for covariates of age, sex, the top 10 Principal Components of the genetic data, household income, highest parental education and data collection site. Effects represent the median across 100 subsamples of singletons to control for family relatedness.

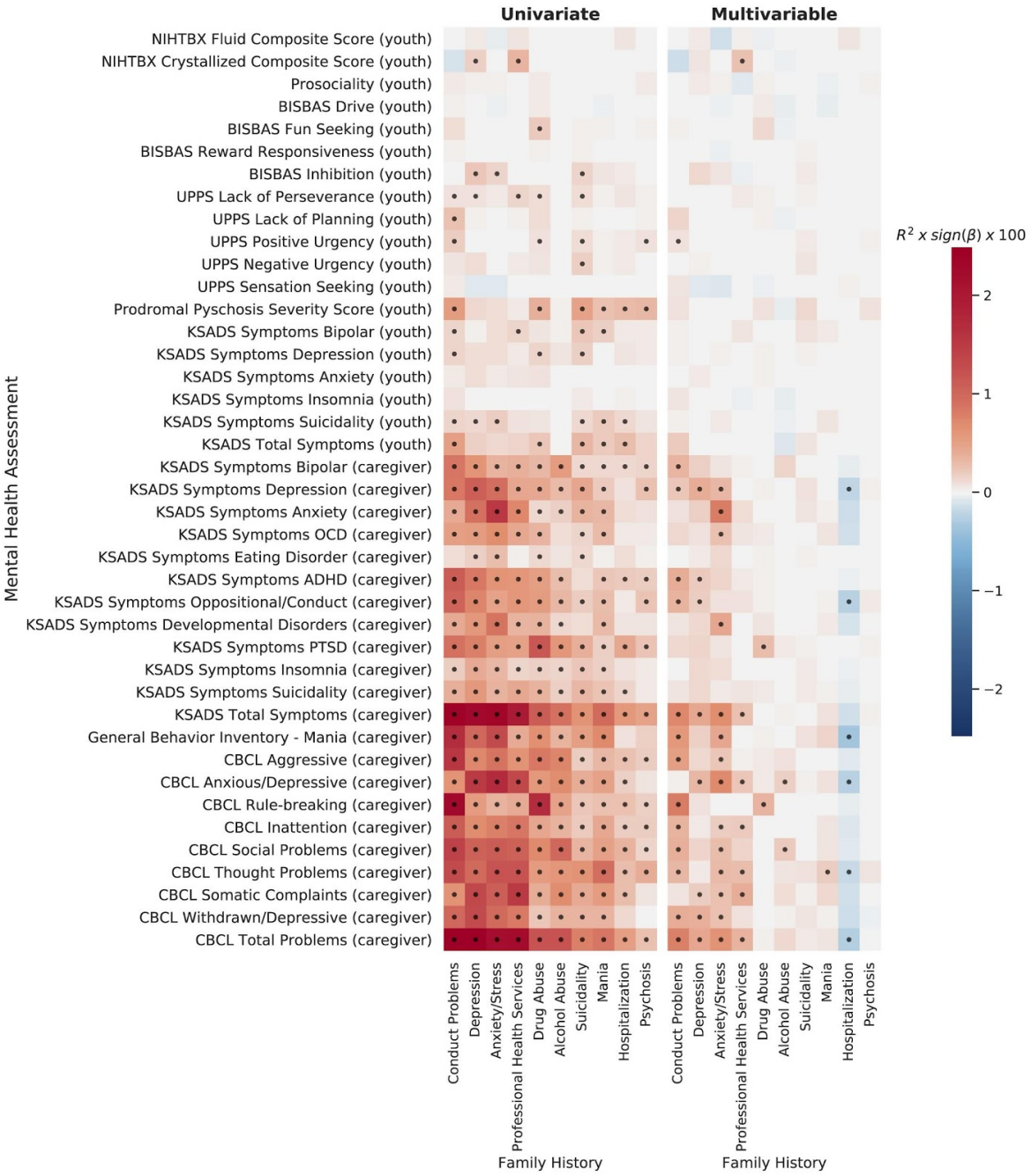


Figure 4.3 ***Univariate (left) and multivariable (right) associations for each behavioral phenotype predicted by FH.*** Effect sizes for each association are displayed as the partial variance explained, R2, (as a percentage) multiplied by the sign of the beta estimate to indicate the magnitude and sign of the association (red=positive, blue=negative). Each row represents a model with the dependent variable along the y-axis and each FH measure on the x-axis. In the univariate models (left) only a single genetic/familial predictor was included in each model (each cell = 1 model) – *i.e.* behavior ~ FH_i + covariates. In the multivariable models (right) all genetic/familial predictors were included in each model including all PRS and FH measures (each row = 1 model) – *i.e.* behavior ~ PRS_1 +PRS_2 … +FH_1 + FH_2 … + covariates. All models controlled for covariates of age, sex, the top 10 Principal Components of the genetic data, household income, highest parental education and data collection site. Effects represent the median across 100 subsamples of singletons to control for family relatedness. Dots indicate FDR significant associations.

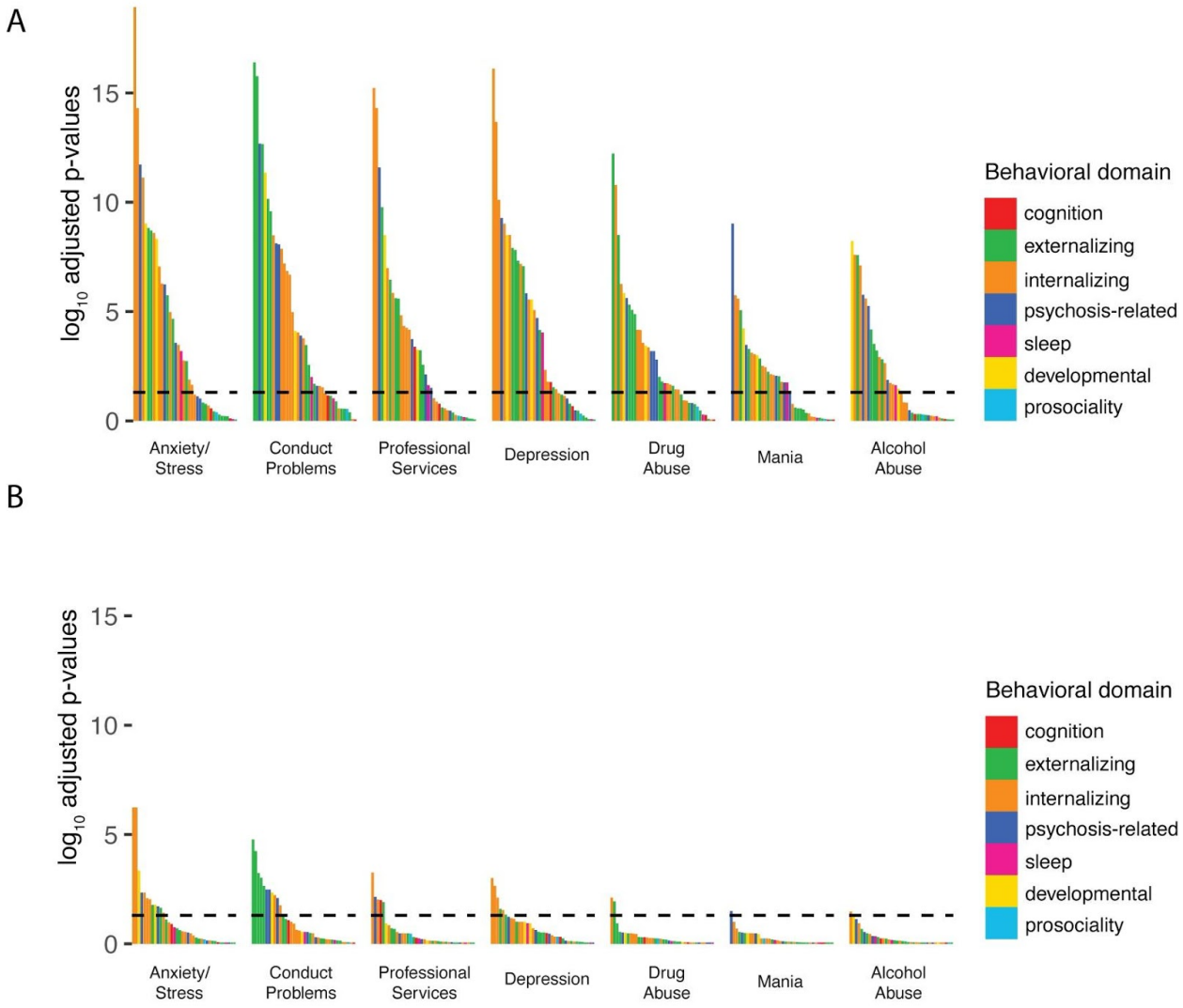


Figure 4.4 ***Enrichment of FH associations across different behavioral domains.*** FDR adjusted p-values (logged) for all of the associations shown in Figure 4.3 for both the univariate (A) and multivariable models (B). Each association is colored based on the behavioral domain the dependent variable was categorized within (see Supplementary Table 4.3). Associations are ordered with the most significant effect to the far left. The horizontal line represents the FDR adjusted significance threshold (p=0.05). All models controlled for covariates of age, sex, the top 10 Principal Components of the genetic data, household income, highest parental education and data collection site. Effects represent the median across 100 subsamples of singletons to control for family relatedness.

References

1. Kendler, K. S. Twin studies of psychiatric illness: An update. *Archives of General Psychiatry* vol. 58 1005–1014 (2001).

2. Kendler, K. S., Davis, C. G. & Kessler, R. C. The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity survey: A family history study. *Br. J. Psychiatry* ***170***, 541–548 (1997).

3. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* ***360***, eaap8757 (2018).

4. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* ***179***, 1469-1482.e11 (2019).

5. Akingbuwa, W. A. *et al.* Genetic Associations between Childhood Psychopathology and Adult Depression and Associated Traits in 42998 Individuals: A Meta-analysis. *JAMA Psychiatry* ***77***, 715–728 (2020).

6. Blacker, D. *et al.* Psychiatric Genetics: A Survey of Psychiatrists' Knowledge, Opinions, and Practice Patterns. *J. Clin. Psychiatry* ***66***, 821–830 (2010).

7. Taylor, J. H. *et al.* Characteristics of youth with reported family history of psychosis spectrum symptoms in the Philadelphia Neurodevelopmental Cohort. *Schizophr. Res.* ***216***, 104–110 (2020).

8. Nivard, M. G. *et al.* Genetic Overlap Between Schizophrenia and Developmental Psychopathology: Longitudinal and Multivariate Polygenic Risk Prediction of Common Psychiatric Traits During Development. *Schizophr. Bull.* ***43***, 1197–1207 (2017).

9. Smeland, O. B. *et al.* Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol. Psychiatry* ***25***, 844–853 (2020).

10. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* ***9***, (2018).

11. Garavan, H. *et al.* Recruiting the ABCD sample: Design considerations and procedures. *Dev. Cogn. Neurosci.* ***32***, 16–22 (2018).

12. Barch, D. M. *et al.* Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev. Cogn. Neurosci.* ***32***, 55–66 (2018).

13. KAUFMAN, J. *et al.* Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *J. Am. Acad. Child Adolesc. Psychiatry* ***36***, 980–988 (1997).

14. Achenbach, T. M. The Achenbach System of Empirically Based Asessment (ASEBA): Development, Findings, Theory and Applications. *Burlington, VT Univ. Vermont Res.*

Cent. Child. Youth Fam. (2009).

15. Youngstrom, E. A., Frazier, T. W., Demeter, C., Calabrese, J. R. & Findling, R. L. Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *J. Clin. Psychiatry* ***69***, 831–9 (2008).

16. Youngstrom, E. A., Findling, R. L., Danielson, C. K. & Calabrese, J. R. Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychol. Assess.* ***13***, 267–276 (2001).

17. Ising, H. K. *et al.* The Validity of the 16-Item Version of the Prodromal Questionnaire (PQ-16) to Screen for Ultra High Risk of Developing Psychosis in the General Help-Seeking Population. *Schizophr. Bull.* ***38***, 1288–1296 (2012).

18. Karcher, N. R. *et al.* Assessment of the prodromal questionnaire-brief child version for measurement of self-reported psychoticlike experiences in childhood. *JAMA Psychiatry* ***75***, 853–861 (2018).

19. Loewy, R. L., Pearson, R., Vinogradov, S., Bearden, C. E. & Cannon, T. D. Psychosis risk screening with the Prodromal Questionnaire--brief version (PQ-B). *Schizophr. Res.* ***129***, 42–6 (2011).

20. Whiteside, S. P., Lynam, D. R., Miller, J. D. & Reynolds, S. K. Validation of the UPPS impulsive behaviour scale: a four-factor model of impulsivity. *Eur. J. Pers.* ***19***, 559–574 (2005).

21. Carver, C. S. & White, T. L. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *J. Pers. Soc. Psychol.* ***67***, 319–333 (1994).

22. Jorm, A. F. *et al.* Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Pers. Individ. Dif.* ***26***, 49–58 (1998).

23. Luciana, M. *et al.* Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev. Cogn. Neurosci.* ***32***, 67–79 (2018).

24. Gershon, R. C. *et al.* NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology* ***80***, S2--S6 (2013).

25. Palmer, C. E. *et al.* Distinct Regionalization Patterns of Cortical Morphology are Associated with Cognitive Performance Across Different Domains. *Cereb. Cortex* ***00***, 1–16 (2021).

26. Akshoomoff, N. *et al.* VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monogr. Soc. Res. Child Dev.* ***78***, 119–132 (2013).

27. Hodes, R. J., Insel, T. R., Landis, S. C. & NIH Blueprint for Neuroscience Research. The NIH Toolbox: Setting a standard for biomedical research. *Neurology* ***80***, S1--S1 (2013).

28. Heaton, R. K. *et al.* Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. *J. Int. Neuropsychol. Soc.* ***20***, 588–598 (2014).

29. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* ***51***, 63–75 (2019).

30. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* ***51***, 431–444 (2019).

31. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* ***51***, 793–803 (2019).

32. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* ***511***, 421–427 (2014).

33. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* ***22***, 343–352 (2019).

34. Cox, D.R. and Snell, E. . *Analysis of Binary Data. 2nd Edition*. (1989).

35. Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika* ***78***, 691–692 (1991).

36. Agerbo, E. *et al.* Risk of Early-Onset Depression Associated with Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA Psychiatry* ***78***, 387–397 (2021).

37. Lu, Y. *et al.* Genetic risk scores and family history as predictors of schizophrenia in Nordic registers. *Psychol. Med.* ***48***, 1201–1208 (2018).

38. Subbarao, A. *et al.* Common genetic and environmental influences on major depressive disorder and conduct disorder. *J. Abnorm. Child Psychol.* ***36***, 433–444 (2008).

39. Cosgrove, V. E. *et al.* Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *J. Abnorm. Child Psychol.* ***39***, 109–123 (2011).

40. Kwong, A. S. F. *et al.* Genetic and Environmental Risk Factors Associated with Trajectories of Depression Symptoms from Adolescence to Young Adulthood. *JAMA Netw. Open* ***2***, 1–14 (2019).

41. Loth, A. K., Drabick, D. A. G., Leibenluft, E. & Hulvershorn, L. A. Do childhood externalizing disorders predict adult depression? A meta-analysis. *J. Abnorm. Child Psychol.* ***42***, 1103–1113 (2014).

42. Brikell, I. *et al.* The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. doi:10.1038/s41380-018-0109-2.

43. Jansen, P. R. *et al.* Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *J. Child Psychol. Psychiatry Allied Discip.* ***59***, 39–47 (2018).

44. Martin, J., Hamshere, M. L., Stergiakouli, E., O'Donovan, M. C. & Thapar, A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol. Psychiatry* ***76***, 664–671 (2014).

45. Groen-Blokhuis, M. M. *et al.* Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *J. Am. Acad. Child Adolesc. Psychiatry* ***53***, 1123-1129.e6 (2014).

46. Pinto, R., Rijsdijk, F., Ronald, A., Asherson, P. & Kuntsi, J. The Genetic Overlap of Attention-Deficit/Hyperactivity Disorder and Autistic-like Traits: an Investigation of Individual Symptom Scales and Cognitive markers. *J. Abnorm. Child Psychol.* ***44***, 335–345 (2016).

47. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* ***45***, 984–994 (2013).

48. Gamsiz, E. D. *et al.* Discovery of Rare Mutations in Autism: Elucidating Neurodevelopmental Mechanisms. *Neurotherapeutics* ***12***, 553–571 (2015).

49. Frías, Á., Palma, C. & Farriols, N. Comorbidity in pediatric bipolar disorder: Prevalence, clinical impact, etiology and treatment. *Journal of Affective Disorders* vol. 174 378–389 (2015).

50. Dalsgaard, S. *et al.* Association between attention-deficit hyperactivity disorder in childhood and schizophrenia later in adulthood. *Eur. Psychiatry* ***29***, 259–263 (2014).

51. Dalteg, A., Zandelin, A., Tuninger, E. & Levander, S. Psychosis in adulthood is associated with high rates of ADHD and CD problems during childhood. *Nord. J. Psychiatry* ***68***, 560–566 (2014).

52. Peralta, V. *et al.* The meaning of childhood attention-deficit hyperactivity symptoms in patients with a first-episode of schizophrenia-spectrum psychosis. *Schizophr. Res.* ***126***, 28–35 (2011).

53. Jones, H. J. *et al.* Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry* ***73***, 221–228 (2016).

54. Riglin, L. *et al.* The impact of schizophrenia and mood disorder risk alleles on emotional problems: Investigating change from childhood to middle age. *Psychol. Med.* ***48***, 2153–2158 (2018).

55. Akingbuwa, W. A., Hammerschlag, A. R., Bartels, M. & Middeldorp, C. M. Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. *J. Am. Acad. Child Adolesc. Psychiatry* ***00***, 1–16 (2021).

56. Sheffield, J. M., Karcher, N. R. & Barch, D. M. Cognitive Deficits in Psychotic Disorders: A Lifespan Perspective. *Neuropsychology Review* vol. 28 509–533 (2018).

57. Mollon, J. & Reichenberg, A. Cognitive development prior to onset of psychosis. *Psychological Medicine* vol. 48 392–403 (2018).

58. De Reyes, A. L. & Kazdin, A. E. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychol. Bull.* ***131***, 483–509 (2005).

59. Fergusson, D. M., Lynskey, M. T. & Horwood, L. J. The effect of maternal depression on maternal ratings of child behavior. *J. Abnorm. Child Psychol.* ***21***, 245–269 (1993).

60. Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E. & Visscher, P. M. Complex trait prediction from genome data: Contrasting EBV in livestock to PRS in humans. *Genetics* ***211***, 1131–1141 (2019).

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

Appendix

Genetic Data Preprocessing

At the baseline visit blood or saliva samples of participants were collected and sent to Rutgers University Cell and DNA Repository for DNA isolation and storage. Genotyping was performed on 646,247 genetic variants using the smokescreen array[1]. 1,221 individuals and 128,523 markers were removed due to missing genetics or failing to meet QC of greater than 5% minor allele frequency and less than 20% of the sample missing for each marker. We derived genetic ancestry factors (GAFs) using fastStructure with four ancestry factors[2] and genetic relatedness was computed using PLINK[3]. Imputation was performed using the Michigan Imputation Server[4] with hrc.r1.1.2016 reference panel, Eagle v2.3 phasing and multiethnic imputation process. Best guess conversion at a threshold of 0.9 was used to convert dosage files to plink files using PLINK[3]. Post imputation QC criteria was a Hardy-Weinberg threshold of 10^{-6}. This QC filtering was performed using PLINK[3] and resulted in 38,900,342 remaining markers and 10,659 individuals.

PRS estimation (additional details)

Due to linkage disequilibrium nearby SNPs are often correlated with one another, as such these are removed before polygenic scoring, this process is known as clumping and pruning. After genetic imputation and post imputation QC, we performed clumping of SNPs using PRSice[5] with a clumping r^2 of 0.1 , clumping window of 250 kb. We did not use a p-value threshold for calculating the polygenic scores, except for DEP which only included the top 10k SNPs due to legal stipulations of the 23andMe sample – this induced an effective threshold of ~0.003. Additionally, variants part of the major histone compatibility (MHC) region

(chromosome 6 28MB-34MB) were removed from the analysis due to its highly variable LD structure[6]. However, we retained the C4 locus (situated in the MHC region) due to its strong association with schizophrenia[7,8]. Indels and multi-allelic SNPs were excluded. The PRS for each participant was calculated as the dot product of the allele value at each loci multiplied by its effect size. Due to some polygenic scores having skewed distributions we decided to rank normalize[9] each score to ensure they followed a normality.

Statistical Analysis (additional details)

Due to convergence issues when using mixed effects models with a random effect of family, we controlled for family relatedness using a random subset of the sample that only included singletons. To ensure the stability of our findings we ran all models in 100 random subsamples of singletons and took the median of effect sizes across all iterations. We calculated effect sizes as R^2 type III sum of squares using Nagelkerke's correction to Cox and Snell's formulation[10,11]. P-values were calculated using a log likelihood ratio test and significant associations determined using a false discovery rate (FDR) significance threshold calculated using the Benjamini-Hochberg method[12]. This correction was made within each ancestry group and model type (univariate or multivariate) – i.e. correcting for 15 genetic/familial predictors and 41 behavioral assessments = 615 multiple tests. GLMs were implemented using the R stats package. Model output for all averaged models is downloadable as csv files (see Statistical Data Tables).

The distribution of each of the DVs fell into three categories a) normal, b) right skewed, zero inflated or c) binary, and we appropriately modelled each of these distributions differently. A) For normal distributions we further ensured normality by rank normalizing[9] (as was performed for PRS), and fit GLMs using the default gaussian family. B) For the right skewed,

zero inflated distributions we fit using a gamma distribution with a log link function, first ensuring that each distribution was non-negative to ensure correct bounds for the link function. C) For binary variables we performed logistic regression. KSADS symptom scores (except for the youth and caregiver reported total symptom score) were binarized using a median split, as they exhibited convergence issues when treated as continuous.

Associations across ancestry strata

As PRS were trained on European individuals and ABCD has high ancestral admixture, the main analyses were performed in a European only sample (European Genetic Ancestry Factor (EUR-GAF)>0.9) and supplementary analyses were conducted in the full sample and a non-European sample (EUR-GAF<0.9) to check for consistency across ancestral groups. Allele frequency differences across ancestral groups can lead to spurious results when PRS trained on a single ancestral group are applied to samples of different or mixed genetic ancestry. Supplementary Figure 4.1 shows the same univariate and multivariable PRS and FH associations in the full sample (n=9,168 with complete genetic data) and a non-European sample (n=3,964). There was a similar pattern of associations for the full sample compared to the Europeans, but with far fewer significant associations for the non-European sample, despite very similar prevalence rates across KSADS diagnoses for the three samples (Supplementary Table 4.1 and 4.2). Supplementary Figure 4.7 shows broadly consistent analogous associations between European and non-European groups; however, there was moderate dispersion observed between estimated effect sizes between groups. These effects are difficult to interpret as the discovery sample only included European individuals. These inconsistencies once again demonstrate the issues of portability of GWAS between ancestry groups[13–15].

Supplementary Table 4.1 ***Prevalence rates of diagnoses based on the KSADS diagnostic interview with the caregiver.*** Diagnoses are reported based on the specified ICD codes. An 'X' indicates that the ICD codes includes all subtypes. Percentage of participants who reached clinical threshold for the described diagnoses are reported for both the full sample (n=9168), the European only sample (n=5204) and the non-European sample (n=3964).

Diagnosis (caregiver report)	ICD Code	Full sample (%)	European sample (%)	Non-European sample (%)
Major depressive disorder (MDD)	F32/F33	2.53	2.17	3.00
Major depressive disorder (MDD) in partial remission	F32.4	0.20	0.17	0.23
Unspecified depressive disorder	F32.9	3.58	3.52	3.66
Persistent depressive disorder, dysthymia	F34.1	0.16	0.15	0.18
Bipolar II disorder	F31.81	1.06	0.81	1.39
Bipolar I disorder, current episode depressed	F31.3X	0.23	0.17	0.30
Bipolar I disorder, currently hypomanic	F31.0	0.01	0.02	0.00
Bipolar I disorder, currently manic	F31.1X	2.53	1.61	3.73
Unspecfied bipolar and related disorder	F31.9X	3.58	2.71	4.72
Disruptive mood dysregulation disorder	F34.8	0.07	0.10	0.03
Unspecified schizophrenia spectrum and other related psychotic disorders	F29	0.71	0.58	0.88
Other specified anxiety disorder	F41.8	9.26	10.68	7.39
Panic disorder	F41.0	0.25	0.23	0.28
Agoraphobia	F40.00	0.57	0.65	0.45
Separation anxiety disorder	F93.0	0.39	0.33	0.48
Social anxiety disorder	F40.10	5.25	5.65	4.72
Specific phobia	F40.2X	31.46	30.46	32.77
Generalised anxiety disorder (GAD)	F41.1	4.97	5.88	3.78
Obsessive compulsive disorder (OCD)	F42.X	11.30	10.59	12.24
Other specified OCD related disorder	F42.8	2.08	2.38	1.69
Binge eating disorder	F50.81	0.96	0.71	1.29
Other specified eating disorder	F50.89	9.58	9.45	9.74
Anorexia nervosa restricting subtype	F50.01	0.12	0.10	0.15
Anorexia nervosa binge eating subtype	F50.02	0.00	0.00	0.00
Bulimia nervosa	F50.2	0.04	0.00	0.10
Attention deficit hyperactivity disorder (ADHD)	F90.X	21.20	21.06	21.39
Unspecified ADHD	F90.9	1.01	1.04	0.98
Oppositional defiant disorder (ODD)	F91.3	14.47	16.06	12.39
Conduct disorder (CD) childhood onset	F91.1	3.44	3.07	3.91
Conduct disorder (CD) adolescent onset	F91.2	0.57	0.31	0.91
Other specified neurodevelopmental disorder, not Autism Spectrum Disorder	F88	26.96	26.52	27.55
Post-traumatic stress disorder (PTSD)	F43.1	2.03	1.59	2.60
Other specified trauma-and stressor related disorder	F43.8	3.01	2.52	3.66
Sleep problems	G47	10.96	12.45	9.01

Supplementary Table 4.2 ***Prevalence rates of diagnoses based on the KSADS diagnostic interview with the youth.*** Diagnoses are reported based on the specified ICD codes. An 'X' indicates that the ICD codes includes all subtypes. Percentage of participants who reached clinical threshold for the described diagnoses are reported for both the full sample (n=9168), the European only sample (n=5204) and the non-European sample (n=3964)

Diagnosis (youth report)	ICD Code	Full sample (%)	European sample (%)	Non-European sample (%)
Major depressive disorder (MDD)	F32/F33	2.52	2.04	3.15
Major depressive disorder (MDD) in partial remission	F32.4	0.17	0.10	0.28
Unspecified depressive disorder	F32.9	2.33	2.21	2.50
Persistent depressive disorder, dysthymia	F34.1	0.04	0.02	0.08
Bipolar II disorder	F31.81	1.30	1.04	1.64
Bipolar I disorder, current episode depressed	F31.3X	0.46	0.37	0.58
Bipolar I disorder, currently hypomanic	F31.0	0.03	0.04	0.03
Bipolar I disorder, currently manic	F31.1X	3.23	2.90	3.66
Unspecfied bipolar and related disorder	F31.9X	5.69	5.28	6.23
Disruptive mood dysregulation disorder	F34.8	0.13	0.08	0.20
Other specified anxiety disorder	F41.8	2.31	2.13	2.55
Social anxiety disorder	F40.10	0.86	0.77	0.98
Generalised anxiety disorder (GAD)	F41.1	0.82	0.98	0.61
Sleep problems	G47	17.93	17.31	18.74

Supplementary Table 4.3 . DEAP variable names for all behavioral variables analyzed in this study. The KSADS symptom scores were calculated by summing all of the symptom items in each KSADS module indicated in the table. Symptom items were all binary (1 or 0). Both past and present symptoms were included in the summary scores, therefore summary scores represent a lifetime assessment of symptoms associated with a particular disorder. Symptom summary scores were calculated by summing the symptoms within each module (the exception was for ODD and CD which were summed together). All KSADS symptom scores were then coded as binary using a median split for statistical modelling. Behavioral domain represents broad categories of construct similarity used for data visualization. R code to extract and process the variables used for this analysis will be published on the ABCD study GitHub: https://github.com/ABCD-STUDY/

Questionnaire	Variables Analyzed	DEAP Variable Names	Informant	Domain
Child Behavior Checklist (CBCL)	CBCL Aggressive CBCL Anxious/Depressive CBCL Rule-breaking CBCL Inattention CBCL Social Problems CBCL Thought Problems CBCL Somatic Complaints CBCL Withdrawn/Depressive CBCL Total Problems	cbcl_scr_syn_aggressive_r cbcl_scr_syn_anxdep_r cbcl_scr_syn_rulebreak_r cbcl_scr_syn_attention_r cbcl_scr_syn_social_r cbcl_scr_syn_thought_r cbcl_scr_syn_somatic_r cbcl_scr_syn_withdep_r cbcl_scr_syn_totprob_r (sum of all sub-scales)	Caregiver	Externalizing Internalizing Externalizing Externalizing Developmental Psychosis-related Internalizing Internalizing NA
General Behavior Inventory	General Behavior Inventory - Mania	pgbi_ss_score_p	Caregiver	Psychosis-related
Prosocial Behavior Survey (youth)	Prosociality	prosocial_ss_mean	Youth	Prosociality
Prodromal Questionnaire Brief Version (PQ-B)	Prodromal Pyschosis Severity Score	prodrom_psych_ss_severity_score	Youth	Psychosis-related
UPPS-P for Children Short Scale	UPPS Lack of Perseverance UPPS Lack of Planning UPPS Positive Urgency UPPS Negative Urgency UPPS Sensation Seeking	upps_ss_lack_of_perseverance upps_ss_lack_of_planning upps_ss_positive_urgency upps_ss_negative_urgency upps_ss_sensation_seeking	Youth	Externalizing Externalizing Externalizing Externalizing Externalizing

Behavioral inhibition and behavioral activation (BISBAS) scale	BISBAS Drive BISBAS Fun Seeking BISBAS Reward Responsiveness BISBAS Inhibition	bisbas_ss_bas_drive bisbas_ss_bas_fs bisbas_ss_bas_rr bisbas_ss_bis_sum	Youth	Externalizing Externalizing Externalizing Internalizing
Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) categorical diagnostic assessments	KSADS Symptoms Depression KSADS Symptoms Bipolar KSADS Symptoms Anxiety KSADS Symptoms OCD KSADS Symptoms Eating Disorder KSADS Symptoms ADHD KSADS Symptoms Oppositional/Conduct KSADS Symptoms Developmental Disorders KSADS Symptoms PTSD KSADS Symptoms Insomnia KSADS Symptoms Suicidality KSADS Total Symptoms	Modules 1 (depressive disorder) & 3 (disruptive mood dysregulation) Module 2 (all bipolar subtypes) Module 8 (social anxiety) & 10 (general anxiety) Module 11 (OCD) Module 13 (eating disorders) Module 14 (ADHD) Module 15 (oppositional defiant) & 16 (conduct disorder) Module 18 (other developmental disorder NOT autism) Module 21 (PTSD) Module 22 (sleep problems) Module 23 (suicidality) Summary score across all modules	Caregiver & Youth Caregiver & Youth Caregiver & Youth Caregiver Caregiver Caregiver Caregiver Caregiver Caregiver & Youth Caregiver & Youth Caregiver & Youth	Internalizing Psychosis-related Internalizing Internalizing Internalizing Externalizing Externalizing Developmental Internalizing Sleep Internalizing NA
NIH Toolbox®	NIH Toolbox® Fluid Composite Score NIH Toolbox® Crystallized Composite Score	nihtbx_cryst_uncorrected nihtbx_fluidcomp_uncorrected	Youth Youth	Cognition Cognition

Supplementary Table 4.4 Description of the family history variables used and the questions asked. If the participant had ANY blood relative who endorsed the described behavior the “DEAP variable name Lead Q” was endorsed with a 1 (if not = 0).

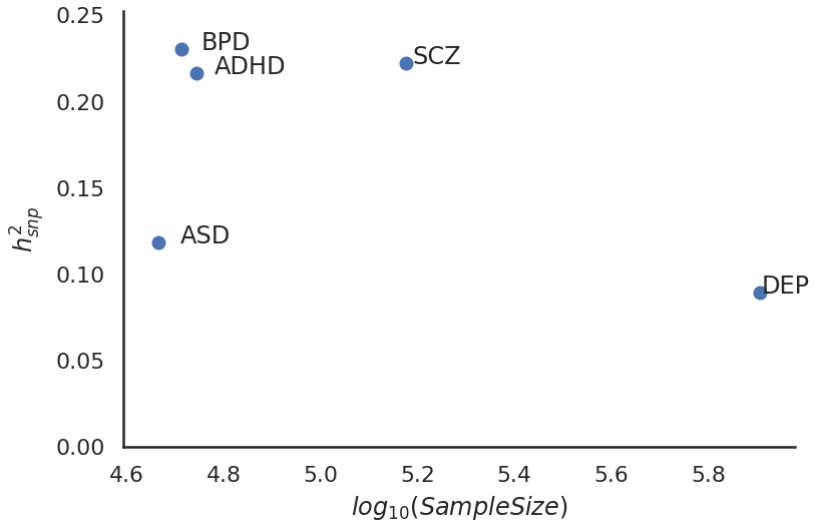
Behavior	DEAP Variable Name Lead Q	Lead Question
Alcohol use	***famhx_4_p***	Has ANY blood relative of your child ever had any problems due to alcohol, such as: Marital separation or divorce; Laid off or fired from work; Arrests or DUIs; Alcohol harmed their health; In an alcohol treatment program; Suspended or expelled from school 2 or more times; Isolated self from family, caused arguments or were drunk a lot.
Drug use	***fam_history_5_yes_no***	Has ANY blood relative of your child ever had any problems due to drugs, such as: Marital separation or divorce; Laid off or fired from work; Arrests or DUIs; Drugs harmed their health; In a drug treatment program; Suspended or expelled from school 2 or more times; Isolated self from family, caused arguments or were high a lot.
Depression	***fam_history_6_yes_no***	Has ANY blood relative of your child ever suffered from depression, that is, have they felt so low for a period of at least two weeks that they hardly ate or slept or couldn't work or do whatever they usually do?
Mania	***famhx_7_yes_no***	Has ANY blood relative of your child ever had a period of time when others were concerned because they suddenly became more active day and night and seemed not to need any sleep and talked much more than usual for them?
Psychosis (visions)	***famhx_8_yes_no***	Has ANY blood relative of your child ever had a period lasting six months when they saw visions or heard voices or thought people were spying on them or plotting against them?
Conduct problems (trouble)	***famhx_9_yes_no***	Has ANY blood relative of your child been the kind of person who never holds a job for long, or gets into fights, or gets into trouble with the police from time to time, or had any trouble with the law as a child or an adult?
Nerves	***famhx_10_yes_no***	Has ANY blood relative of your child ever had any other problems with their nerves, or had a nervous breakdown?
Seen a therapist (professional)	***famhx_11_yes_no***	Has ANY blood relative of your child ever been to a doctor or a counselor about any emotional or mental problems, or problems with alcohol or drugs?
Hospitalized	***famhx_12_yes_no***	Has ANY blood relative of your child ever been hospitalized because of emotional or mental problems, or drug or alcohol problems?
Suicide	***famhx_13_yes_no***	Has ANY blood relative of your child ever attempted or committed suicide?

Supplementary Table 4.5 Unique and shared variance across behaviors predicted by PRS and FH. Change in R^2% is calculated by comparing the R^2 from a reduced model with covariates of no interest and genetic predictors of no interest, and a full model including the nested reduced model plus the genetic predictors of interest. The χ^2 statistic and p-value from the likelihood ratio test comparing these models is also displayed. The parenthesis in the DV column indicate youth (Y) or caregiver (C) report

	PRS Effect (beyond FH + covariates)			***FH Effect (beyond PRS + covariates)***			***PRS + FH Effect (beyond covariates)***		
DV	ΔR^2	χ^2	$p(\chi^2)$	ΔR^2	χ^2	$p(\chi^2)$	ΔR^2	χ^2	$p(\chi^2)$
BISBAS Drive (Y)	0.14%	4.92	1.08E-01	0.20%	7.21	2.11E-01	0.35%	12.37	9.07E-02
BISBAS Fun Seeking (Y)	0.16%	7.64	2.22E-01	0.26%	12.25	3.42E-01	0.43%	20.63	2.20E-01
BISBAS Reward Responsiveness (Y)	0.37%	33.05	6.39E-03	0.12%	10.33	8.88E-01	0.49%	43.81	1.24E-01
BISBAS Inhibition (Y)	0.25%	10.37	5.57E-02	0.41%	17.08	5.87E-02	0.66%	28.04	1.52E-02
UPPS Lack of Perseverance (Y)	0.17%	6.07	3.05E-02	0.34%	12.18	5.86E-03	0.54%	19.33	5.85E-04
UPPS Lack of Planning (Y)	0.35%	13.74	9.14E-03	0.30%	11.78	2.17E-01	0.68%	26.66	1.31E-02
UPPS Positive Urgency (Y)	0.10%	4.31	1.61E-01	0.33%	14.73	2.55E-03	0.46%	20.35	1.11E-03
UPPS Negative Urgency (Y)	0.30%	11.98	2.35E-02	0.27%	10.91	2.96E-01	0.60%	24.25	3.50E-02
UPPS Sensation Seeking (Y)	0.24%	10.92	6.17E-02	0.26%	11.79	3.30E-01	0.49%	22.31	1.22E-01
CBCL Total Problems (C)	0.61%	21.00	6.50E-05	5.27%	190.82	1.86E-46	6.17%	225.69	2.73E-52
CBCL Aggressive (C)	0.32%	22.45	1.12E-02	2.34%	169.31	2.51E-19	2.81%	204.09	2.72E-21
CBCL Anxious/Depressive (C)	0.18%	11.42	9.41E-02	3.12%	197.88	8.37E-30	3.43%	218.15	2.96E-30
CBCL Rule Breaking (C)	0.69%	47.10	3.10E-04	2.66%	185.05	3.21E-15	3.73%	261.84	3.43E-20
CBCL Inattention (C)	0.65%	44.36	2.05E-06	2.03%	140.32	1.02E-18	2.85%	198.87	4.03E-25
CBCL Social Problems (C)	0.26%	18.12	9.82E-02	2.62%	189.85	1.90E-16	3.04%	220.84	3.97E-17

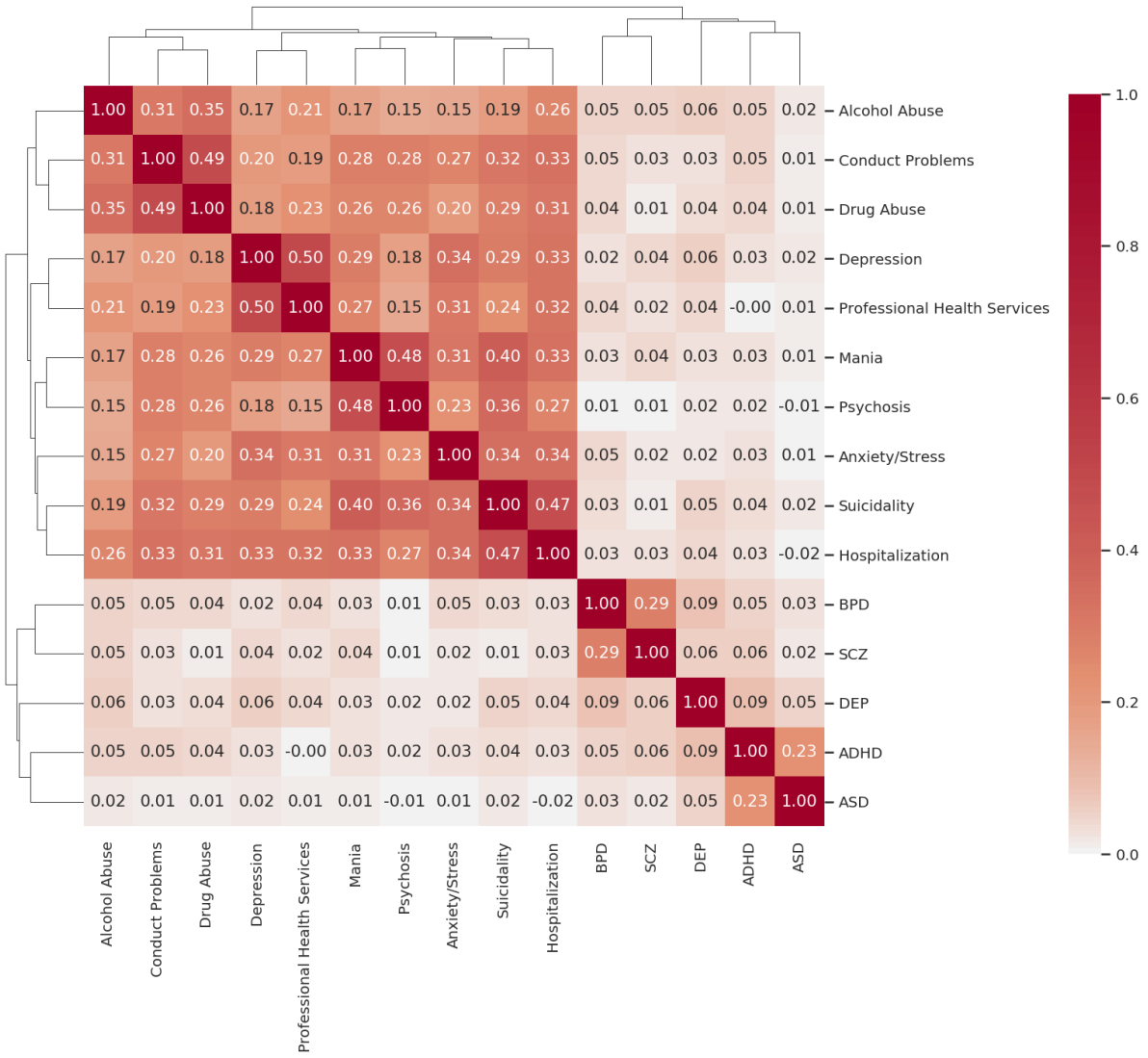
CBCL Thought Problems (C)	0.26%	15.50	3.54E-02	3.06%	188.48	3.28E-26	3.43%	212.51	3.99E-27
CBCL Somatic Complaints (C)	0.42%	26.50	1.25E-03	2.65%	171.39	6.51E-23	3.17%	205.64	2.30E-25
CBCL Withdrawn/Depressive (C)	0.23%	16.31	1.84E-01	2.31%	168.37	1.43E-12	2.59%	189.32	3.09E-12
Psychosis Severity Score (Y)	0.59%	61.65	1.25E-03	1.00%	104.03	2.04E-04	1.72%	180.49	4.47E-07
Mania (C)	0.23%	18.83	4.06E-01	2.99%	250.72	1.24E-10	3.41%	286.29	2.19E-10
Prosociality (Y)	0.20%	6.24	1.33E-01	0.20%	6.42	5.60E-01	0.42%	13.10	2.75E-01
KSADS Symptoms Depression (C)	0.09%	4.11	5.33E-01	1.99%	88.73	9.55E-15	2.15%	95.81	8.07E-14
KSADS Symptoms Bipolar (C)	0.36%	16.14	6.47E-03	1.39%	61.88	1.60E-09	1.86%	82.91	2.05E-11
KSADS Symptoms Anxiety (C)	0.30%	13.43	1.97E-02	2.35%	105.28	4.76E-18	2.64%	118.11	4.39E-18
KSADS Symptoms OCD (C)	0.06%	2.43	7.87E-01	1.47%	65.43	3.35E-10	1.52%	67.70	1.14E-08
KSADS Symptoms Eating Disorder (C)	0.04%	1.87	8.66E-01	0.42%	18.51	4.69E-02	0.46%	20.33	1.60E-01
KSADS Symptoms ADHD (C)	0.28%	12.26	3.14E-02	1.92%	85.45	4.26E-14	2.29%	102.48	4.42E-15
KSADS Symptoms Oppositional Conduct (C)	0.71%	31.27	8.28E-06	1.92%	85.52	4.13E-14	2.80%	125.43	1.66E-19
KSADS Symptoms Developmental Disorders (C)	0.12%	5.30	3.81E-01	1.55%	68.84	7.43E-11	1.67%	74.51	6.93E-10
KSADS Symptoms PTSD (C)	0.09%	4.00	5.50E-01	2.09%	93.50	1.08E-15	2.27%	101.22	7.65E-15
KSADS Symptoms Insomnia (C)	0.15%	6.83	2.34E-01	0.69%	30.75	6.46E-04	0.87%	38.57	7.43E-04
KSADSS Symptoms Suicidality (C)	0.35%	15.39	8.82E-03	0.98%	43.54	3.98E-06	1.41%	62.75	8.45E-08
KSADS Total Symptoms (C)	0.41%	27.68	2.67E-03	5.22%	365.99	4.06E-46	5.80%	408.99	1.23E-48

KSADS Symptoms Depression (Y)	0.40%	17.66	3.41E-03	0.30%	13.47	1.99E-01	0.77%	34.24	3.15E-03
KSADS Symptoms Biploar (Y)	0.27%	11.73	3.87E-02	0.48%	21.31	1.91E-02	0.79%	34.87	2.57E-03
KSADS Symptoms Anxiety (Y)	0.02%	0.71	9.82E-01	0.20%	8.70	5.61E-01	0.21%	9.18	8.68E-01
KSADS Symptoms Insomnia (Y)	0.12%	5.37	3.73E-01	0.14%	6.37	7.83E-01	0.27%	11.92	6.85E-01
KSADS Symptoms Suicidality (Y)	0.24%	10.48	6.27E-02	0.38%	16.83	7.83E-02	0.67%	29.71	1.30E-02
KSADS Total Symptoms (Y)	0.55%	59.87	9.73E-03	0.78%	84.84	1.80E-02	1.51%	166.08	2.22E-04
NIHTBX Fluid Composite Score (Y)	0.48%	18.47	7.79E-04	0.33%	12.64	1.54E-01	0.80%	30.71	2.40E-03
NIHTBX Crystalized Composite Score (Y)	0.43%	15.14	2.03E-03	0.66%	23.14	1.32E-03	1.11%	39.32	1.74E-05

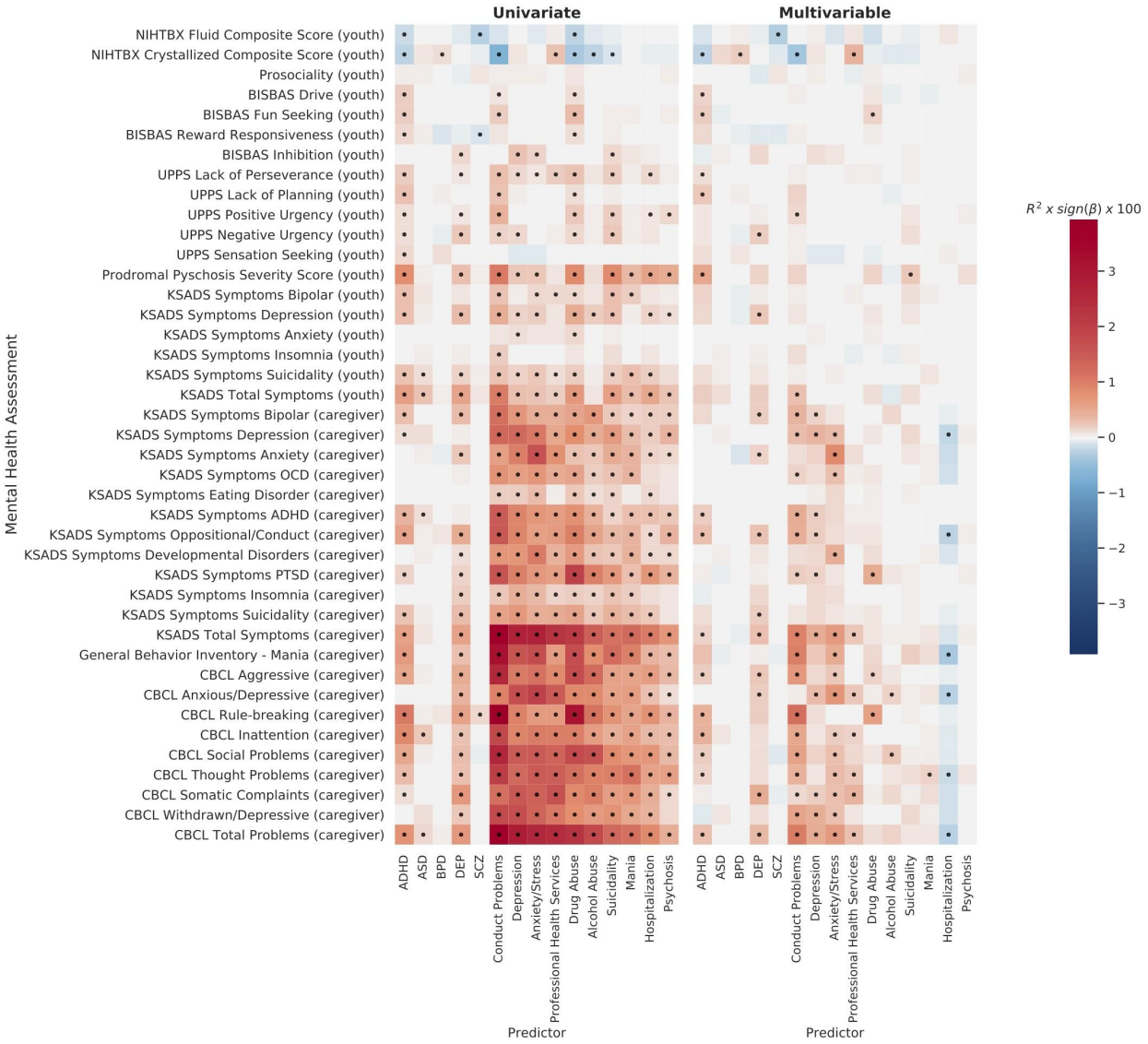


Supplementary Figure 4.2 Plot of SNP heritability (h2snp) and log(sample size) for each discovery GWAS sample, as reported in discovery papers, used to calculate PRS.

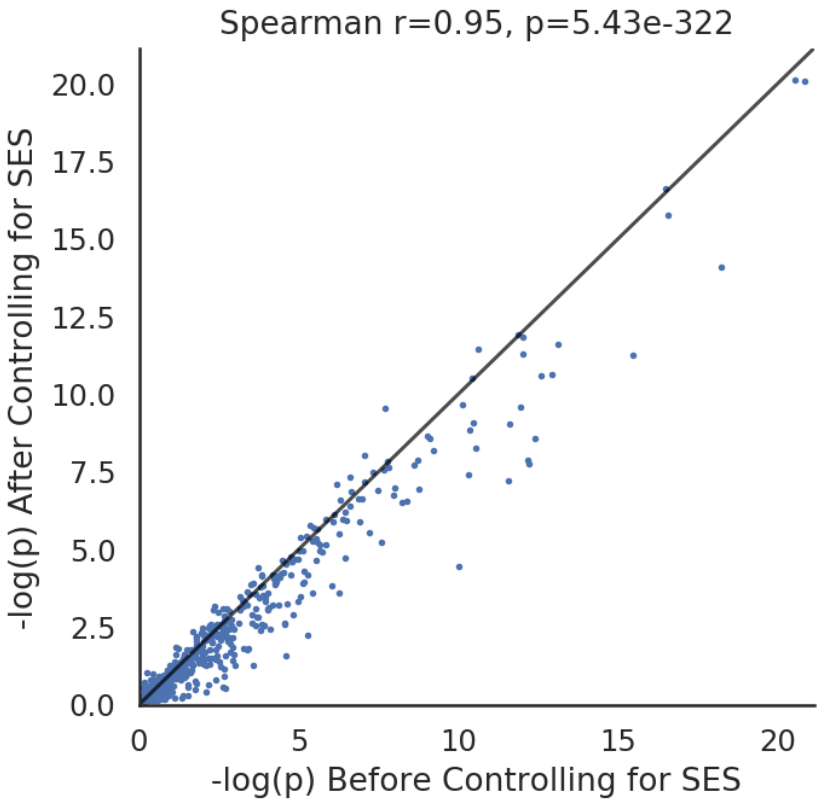
Supplementary Figure 4.3 Pairwise spearman correlations between all of the behavioral phenotypes in the European sample pre-residualized for covariates of no interest including socioeconomic status (SES). Caregiver reported measures were more strongly associated with each other than youth reported measures.



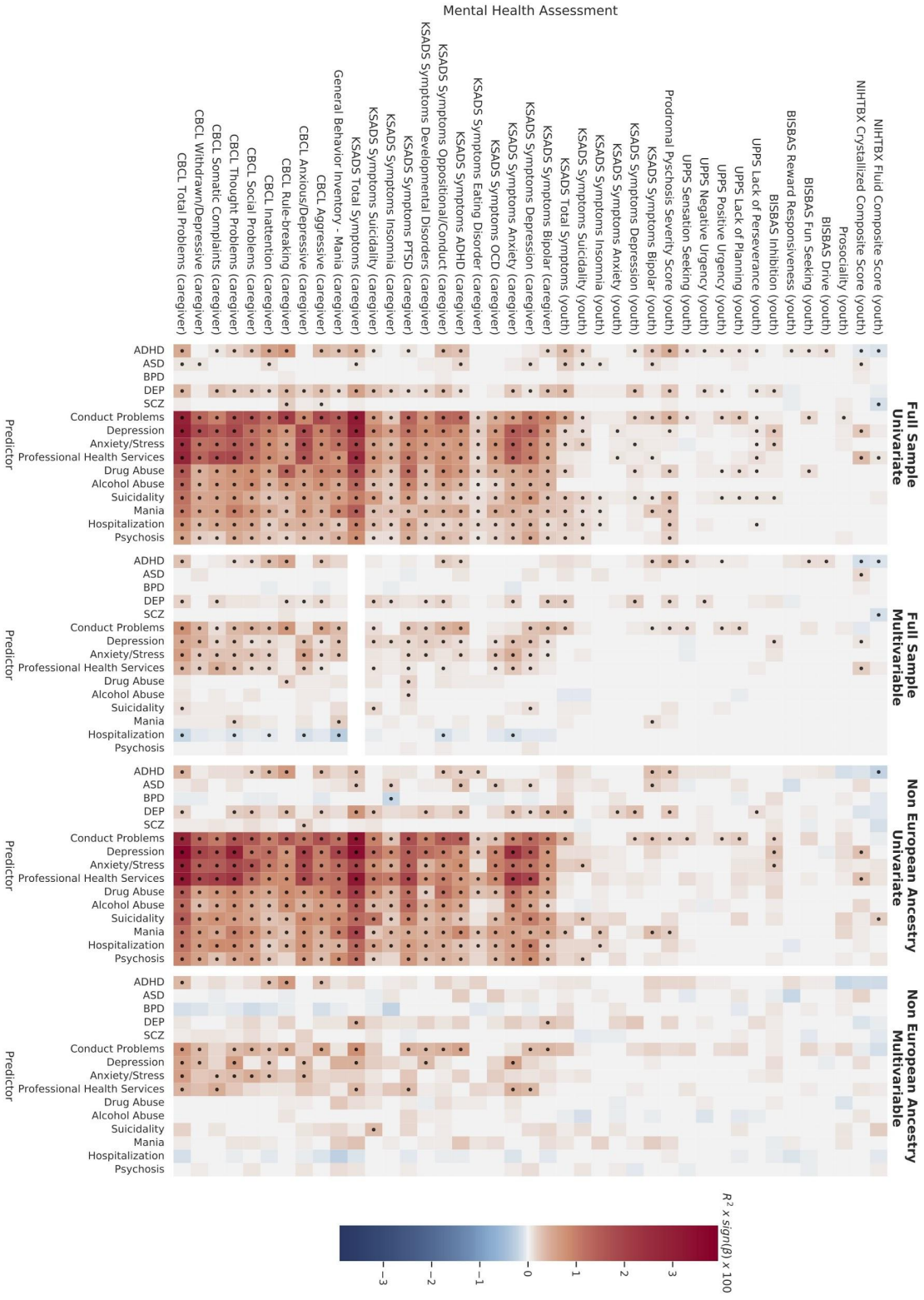
Supplementary Figure 4.4 Pairwise spearman correlations between all of the genetic risk measures in the European sample pre-residualized for covariates of no interest including SES. FH measures were more strongly associated with each other than the PRS. The FH measures were moderately correlated with one another. There were limited associations between FH and PRS.



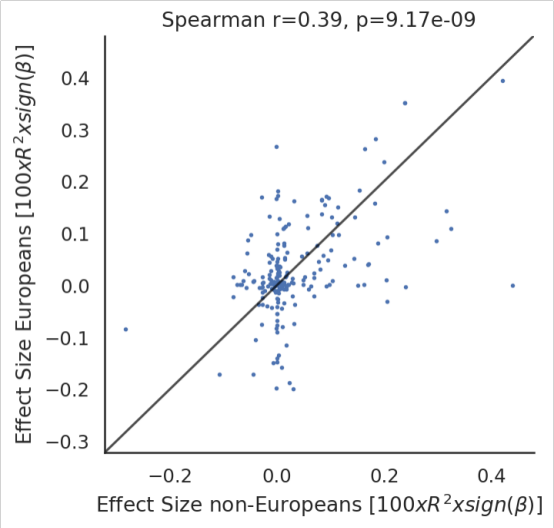
Supplementary Figure 4.5 ***Behavioral associations in the European sample without controlling for SES.*** Univariate (left) and multivariable (right) associations between the genetic predictors (PRS and Family History) and the behavioral phenotypes, controlling for covariates of no interest, but not controlling for SES. The main differences across the associations with and without controlling for SES were found with the cognitive performance measures. Many associations between genetic risk for psychopathology and cognitive function only reached threshold for statistical significance when SES was not taken into account due to the shared variance between sociodemographic factors and cognition.



Supplementary Figure 4.6. -log(P-values) for all multivariable associations after controlling for SES (Y-axis) and before controlling for SES (X-axis) in the European sample. The pattern of associations was very similar with and without controlling for SES, however controlling for SES attenuated many of the associations



Supplementary Figure 4.7 ***Behavioral associations across ancestry strata.*** Univariate and multivariable associations between the genetic predictors (PRS and Family History) and the behavioral phenotypes for the full ABCD sample and the non-European ancestry sample, controlling for covariates of no interest including SES. Models that did not converge after 10,000 iterations were left blank.



Supplementary Figure 4.8 Signed Effect Sizes for all multivariable PRS associations for the European sample (X-axis) and non-European sample (Y-axis) , controlling for covariates of no interest and SES. The estimated associations were broadly consistent between European and non-European groups, however, there was moderate dispersion observed between estimated effect sizes, which highlights difficulties in using PRS generated in certain ancestry groups to other ancestry groups.

Statistical Data Tables

For completeness and transparency of reporting, we have generated additional data tables as .csv files with the results from all statistical models. Each table includes the beta estimate, t statistic, signed percentage variance explained (R^2) and p-value for each predictor in each model. These results are the averaged statistics across 100 samples of singletons as outlined in the Supplementary Methods. The following data tables are included:

- EUR Ancestry Results
- Non EUR Ancestry Results
- Full Sample Results

Each file contains a separate sheet for:

1 Univariate SES covaried

2 Multivariable SES covaried multivariable

3 Univariate (not controlling for SES)

4 Multivariable (not controlling for SES)

References

1. Baurley JW, Edlund CK, Pardamean CI, Conti D V., Bergen AW. Smokescreen: A targeted genotyping array for addiction research. BMC Genomics. 2016;17(1):1–12.

2. Raj A, Stephens M, Pritchard JK. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89.

3. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1).

4. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Aug;48:1284.

5. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466–8.

6. Miretti MM, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am J Hum Genet. 2005;76(4):634–46.

7. Sekar A, Bialas AR, De Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016;530(7589):177–83.

8. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421–7.

9. Pilia G, Chen W-M, Scuteri A, Orrú M, Albai G, Dei M, et al. Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. PLOS Genet. 2006 Aug;2(8):e132.

10. Benjamini, Yoav and Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. J R Stat Soc. 1995;57(1):289–300.

11. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10(1).

12. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, et al. Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. PLoS Biol. 2013;11(9).

13. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635–49.

Chapter 5: *Generalization of Cortical MOSTest Genome-Wide Associations Within and Across Samples*

5.1 *Abstract:*

Genome-Wide Association studies have typically been limited to single phenotypes, given that high dimensional phenotypes incur a large multiple comparisons burden: ~1 million tests across the genome times the number of phenotypes. Recent work demonstrates that a Multivariate Omnibus Statistic Test (MOSTest) is well powered to discover genomic effects distributed across multiple phenotypes. Applied to cortical brain MRI morphology measures, MOSTest has resulted in a drastic improvement in power to discover loci when compared to established approaches (min-P). One question that arises is how well these discovered loci replicate in independent data. Here we perform 10 times cross validation within 35,644 individuals from UK Biobank for imaging measures of cortical area, thickness and sulcal depth (>1,000 dimensionality for each). By deploying a replication method that aggregates discovered effects distributed across multiple phenotypes, termed PolyVertex Score (PVS), we demonstrate a higher replication yield and comparable replication rate of discovered loci for MOSTest (# replicated loci: 348-845, replication rate: 94-95%) in independent data when compared with the established min-P approach (# replicated loci: 31-68, replication rate: 65-80%). An out-of-sample replication of discovered loci was conducted with a sample of 8,336 individuals from the Adolescent Brain Cognitive Development® (ABCD) study, who are on

average 50 years younger than UK Biobank individuals. We observe a higher replication yield and comparable replication rate of MOSTest compared to min-P. This finding underscores the importance of using well-powered multivariate techniques for both discovery and replication of high dimensional phenotypes in Genome-Wide Association studies.

5.2 *Introduction*

Performing Genome Wide Association Studies (GWAS) on high dimensional phenotypes incurs a large multiple comparisons burden (number of independent genetic tests by number of phenotypes) using traditional approaches, which can result in low power to detect associations. Vertex-wise measures of cortical morphology (area, thickness and sulcal depth) represent high dimensional phenotypes (>1000 dimensions) and, from twin studies, are known to have high heritabilities of up to 90% and 50% for total and regional area respectively, and 80% and 60% for mean and regional thickness respectively[1,2]. Our group has previously developed a novel Multivariate Omnibus Test (MOSTest) [3–5], which aggregates the effect of a genomic variant across the cortex. This method significantly boosts discovery of genetic loci linked to cortical morphology, with an up to 10x increase in number of loci discovered – when compared to an established approach (min-P) deployed for the same phenotypes[5]. Additionally, discovered loci show strong enrichment with pathways involved in neurogenesis and cell differentiation. Two main benefits of MOSTest over established techniques, like min-P, are: 1) its ability to aggregate pleiotropic effects into a single statistical test and 2) it drastically reduces the multiple comparison burden across the dimensionality of phenotypes, while still accounting for genome-wide multiple comparisons correction. Given such a dramatic increase in discovery of genomic loci, it is of interest to understand how well these discoveries replicate in independent data.

Here we perform 10-times cross validation with brain imaging data taken from the UK Biobank, and randomly split the sample into ⅔ training and ⅓ replication splits. For the training samples we perform discovery of vertex-wise measures of area, thickness and sulcal depth as in [4]. Having discovered genomic loci in training folds, we perform replication of these loci in the test sets. To perform replication for each SNP we calculate a PolyVertex Score (PVS) (similar to [6,7]) from imaging data in the test set for each MOSTest discovered locus. This PVS aggregates the distributed effects across the cortex by taking a weighted sum across all vertices using mass univariate z statistics as weights from the training set. This approach is similar to the widely used method of Polygenic Risk Scores (PRS) in genetics[8], where instead of predicting a phenotype we are predicting a single genomic variant and instead of using distributed effects across the genome as predictors we use the distributed effects across the cortex, estimated in the training set. For each discovered training set we generate a PVS for each individual, which represents a continuous prediction of the genotype in the test set. We then correlate each PVS with its corresponding measured genomic variant in the test set to test how well these discovered loci replicate (one tailed t test, $p<0.05$). We test this MOSTest discovery and PVS replication, against an established GWAS approach (min-P)[9]. Figure 5.1 displays a schematic of how replication of how min-P and MOSTest differs for a single discovered variant. We confirm a higher replication yield and comparable replication rate MOSTest versus min-P. Finally, we test the generalization of loci discovered in UK Biobank to a developmental cohort of 9-11 year old children from the Adolescent Brain Cognitive Development® (ABCD; https://abcdstudy.org) Study, where we see a higher yield of replicated loci for MOSTest versus min-P.

5.3 *Results*

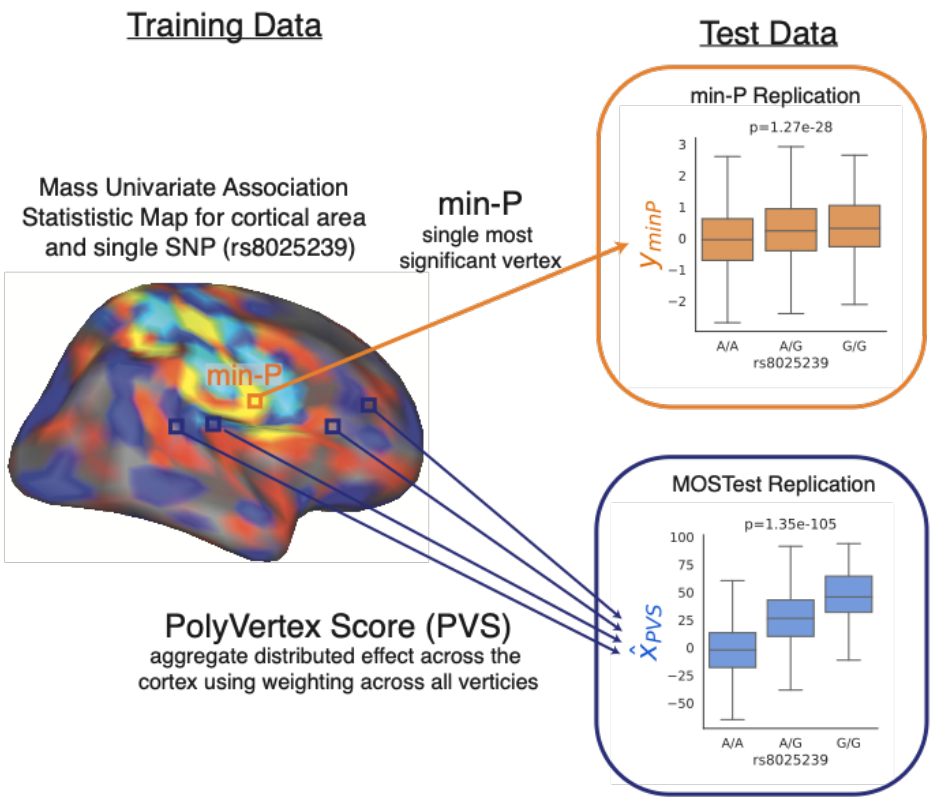


Figure 5.1. Schematic of replication process for a single SNP. Variant rs8025239 is discovered in training fold and has mass univariate map of association statistics with cortical area. Min-P replication (indicated by orange box and arrow) takes most significant vertex and associates that vertex with variant rs8025239 in test data. MOSTest replication (indicated by blue box and arrow) computes a PolyVertex Score (PVS) in test data which aggregates all effects across the cortex by taking a weighted sum (using association statistics from training set) across all vertices – the PVS is then correlated with the variant rs8025239. This process is repeated for all discovered variants in training set with a separate PVS being generated for each MOSTest discovery. Replication of a variant is defined as $p<0.05$ in one tailed t-test.

Across training folds, the UK Biobank sample, we confirm that MOSTest confers up to a 10-fold increase in discovered loci over min-P (area: min-P_{yield}=52, MOSTest$_{yield}$=433, thickness: min-P_{yield}=42, MOSTest$_{yield}$=367 and sulcal depth min-P_{yield}=85, MOSTest$_{yield}$=890). When replication of loci is defined at the nominal level ($p<0.05$, see methods) we see a higher number of replicated loci, as well as comparable replication rate for MOSTest (area:94%, thickness: 95%, sulcal depth: 95%) vs min-P (area:65%, thickness: 73%, sulcal depth: 80%) – see Figure 5.2. Averaged across cross-validation folds, we found that the lead SNP of the top locus

accounted for more variance in the replication set with MOSTest (R^2= area:0.037 σ=3.8x10^{-3}, thickness: 0.059 σ=1.4x10^{-2}, sulcal depth: 0.052 σ=4.0x10^{-3}) compared to min-P (R^2= area: 0.011 σ=1.2x10^{-3}, thickness: 0.011 σ=2.2x10^{-3}, sulcal depth: 0.015 σ=1.6x10^{-3}). If replication is defined more conservatively with significance corrected for the number of discovered loci (p<0.05/# of discovered loci), we again find that MOSTest confers a comparable replication rate (area: 69%, thickness: 70%, sulcal depth: 68%) to min-P (area: 41%, thickness: 55%, sulcal depth: 50%).

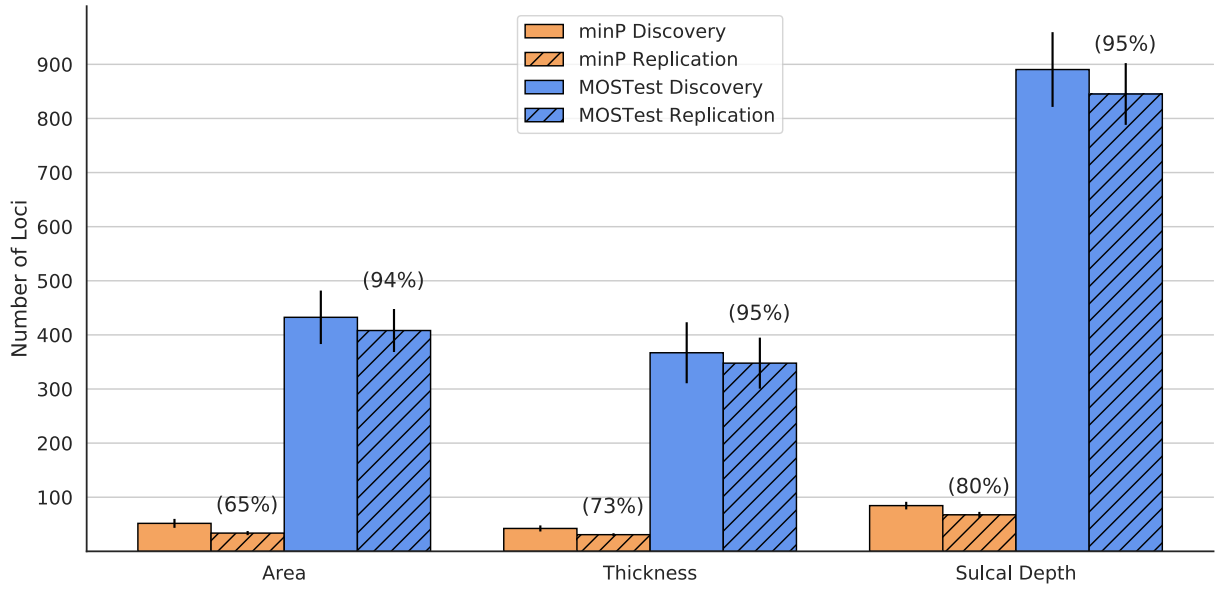


Figure 5.2 Cross-validation discovery and replication yield within 10-times cross validation within UK Biobank for cortical morphometry measures. Solid bars represent the number of genome wide significant loci associated with each measure. Hashed bars represent the number of loci that replicate in test folds at a nominal significance level (p<0.05). Error bars are standard deviations across 10 cross-validation repetitions. Numbers in parentheses represent replication rate (# of discovered loci / # replicated loci) for each method-phenotype pair.

Next, we tested the generalization performance of loci discovered in each training fold of UK Biobank to a developmental cohort of adolescents from the Adolescent Brain Cognitive Development study. Here we once again see a higher absolute number of replicated loci (nominal p<0.05 level), as well as a comparable replication rate for MOSTest (area: 74%, thickness: 69%, sulcal depth: 72%) to min-P (area: 51%, thickness: 48%, sulcal depth: 57%) -

see Figure 5.3. Again, the variance explained by the lead SNP of the top locus (averaged across cross-validation folds) accounted for more variance for MOSTest (R^2= area: 1.6x10^{-2} σ=9.0x10^{-4}, thickness: 2.9x10^{-2} σ=8.4x10^{-3}, sulcal depth: 2.1x10^{-2} σ=1.5x10^{-3}) than for min-P (R^2= area: 7.1x10^{-3} σ=1.0x10^{-4}, thickness: 4.8x10^{-3} σ=1.1x10^{-3}, sulcal depth: 9.1x10^{-3} σ=5.0x10^{-4}).

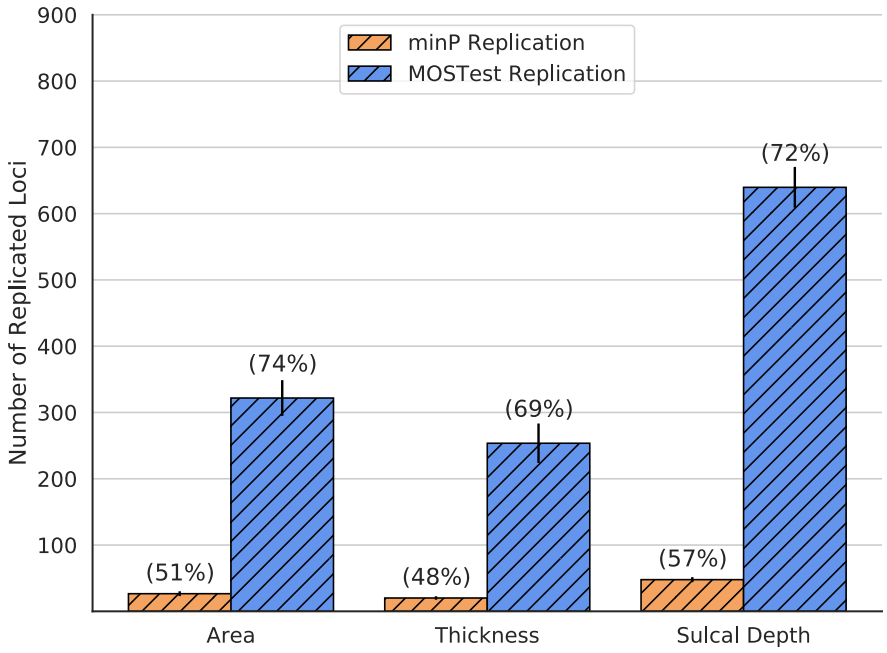


Figure 5.3 Replication yield within the ABCD dataset across 10 training folds of UK Biobank for cortical morphometry measures. Bars represent the number of loci that replicate in ABCD at a nominal significance level (p<0.05). Error bars are standard deviations across 10 training sets of UK Biobank. Numbers in parentheses represent replication rate (# of discovered loci / # replicated loci). ABCD: Adolescent Brain Cognitive Development Study.

5.4 *Discussion*

Here we have confirmed the increased power of using a MOSTest across training folds of UK Biobank. Further, through the deployment of PVS we show that loci discovered with MOSTest result in a higher replication yield and comparable replication rate to independent data than established approaches. The comparable replication rate for MOSTest loci (94-95% vs 65-80% for min-P) indicates that the difference in absolute number of replicated loci for

MOSTest vs min-P is not merely a result of MOSTest discovering a higher number of loci. Furthermore, we still see a comparable replication rate when we penalize the replication significance threshold by the number of loci discovered by each method (i.e. $p<0.05/$ # of discovered loci). This underscores the distributed effects of the genome across the cortex, which multivariate methods are better powered to capture and in turn, will display stronger generalization to independent data.

Additionally, we have shown that genetic-cortical morphology associations learned within an adult population (mean age 64 years) of individuals from the UK generalize out of sample to adolescents aged 9-11 years old in the United States of America taken from the ABCD study. There are marked differences between the training sample of UK Biobank and validation sample of ABCD including: large age differences, a high degree of ancestry admixture in ABCD, different scanners used, imaging protocols and the number of individuals in validation sets. In spite of these differences we observe a high replication rate in ABCD of discoveries found within UK Biobank via MOSTest. We see higher replication for cortical area and sulcal depth in ABCD than for cortical thickness. Cortical thickness changes more dynamically over the lifespan[10], therefore, given the large age disparity between the two samples, perhaps it is not a surprise to see that cortical thickness is the measure that exhibits the largest reduction in replication rates in ABCD when compared across cross-validation folds of UK Biobank for MOSTest (69% vs 95%). We may expect that the replication rate of discovered cortical thickness loci to increase as the children develop, a hypothesis that can be tested as more longitudinal ABCD data is collected. Despite differences across these datasets we observe greater replication of UK Biobank discovered loci in ABCD when taking into account the multivariate nature of associations across the cortex (i.e. MOSTest and PVS).

Furthermore, we demonstrated that lead MOSTest discoveries explained a notable amount of variance out of sample, by GWAS standards: 3-6% in UK Biobank and 1-3% in

ABCD. Methods, such as MOSTest and PVS, that result in high replication yield and out of sample variance explained may support precision medicine efforts[11]. In particular if these methods are deployed on disorders of the brain they may provide complimentary predictive power to well established models such as Polygenic Risk Scores.

The training data used here to detect loci and train PVS projections weights were taken from individuals of European ancestry from the UK Biobank. We may expect that the genetic architecture of cortical morphology to differ between ancestry groups[12]. We also acknowledge that our use of PVS to predict genotypes out of sample is just one possible projection weighting scheme, which may not provide optimal out of sample prediction. Here we have demonstrated the high generalization performance of cortical morphology discoveries using MOSTest to independent data. This was shown both within study (UK Biobank) and across studies (UK Biobank to ABCD) despite substantial age differences of participants. This work underscores the importance of deploying well powered multivariate methods when performing GWAS on high dimensional phenotypes, both for discovery and replication.

5.5 *Methods*

The UK Biobank sample and methods used for min-P and MOSTest discovery overlap with previous work[4,5].

5.5.1 UK Biobank Sample

Genotypes, MRI scans, demographic and clinical data were obtained from the UK Biobank under accession number 27412, excluding participants who withdrew their consent. For this study we selected white British individuals (as derived from both self-declared ethnicity and principal component analysis) who had undergone the neuroimaging protocol. The resulting

sample contained 35,644 individuals with a mean age of 64.4 years (standard deviation 7.5 years), 18,433 female. T_1-weighted MRI scans were collected from three scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receive head coils. We used UK Biobank v3 imputed genotype data[13].

5.5.2 Adolescent Brain Cognitive Development® (ABCD) Sample

The ABCD study is a longitudinal study across 21 data acquisition site following 11,878 children starting at 9 and 10 years old. This paper analyzed the full baseline sample from data release 3.0 (NDA DOI:10.151.54/1519007). The ABCD study used school-based recruitment strategies to create a population-based, demographically diverse sample with heterogeneous ancestry. T_1-weighted MRI scans were collected using Siemens Prisma, GE 750 and Phillips 3T scanners. Scanning protocols were harmonized across 21 acquisition sites. Genetic ancestry factors were estimated using fastStructure[14] with four ancestry groups. Genotype data was imputed at the Michigan Imputation Server[15], using the HRC reference panel as described in[16,17]. We selected individuals who had passed neuroimaging and genetic quality control checks, resulting in 8,336 individuals with a mean age of 9.9 years (standard deviation 0.62 years), 3,974 female.

5.5.3 Data processing

T1-weighted structural MRI scans were processed with the FreeSurfer v5.3 standard "recon-all" processing pipeline[18] to generate 1,284 non-smoothed vertex-wise measures (ico3 downsampling with the medial wall removed) summarizing cortical surface area, thickness and sulcal depth. All measures were pre-residualized for age, sex, scanner site, the first ten genetic principal components. In contrast to previous MOSTest work[3,5] we did not pre-residualize for global measures specific to each set of variables (total cortical surface area or mean cortical

thickness) as there is no clear corollary for sulcal depth, nor did we control for Euler number. Subsequently, a rank-based inverse normal transformation was applied to the residualized measures. For genomic data we carried out standard quality-checks as described previously[3], setting a minor allele frequency threshold of 0.5% and finding the intersecting variants between UK Biobank and ABCD, leaving 8 million variants. Variants were tested for association with cortical surface area, cortical thickness and sulcal depth at each vertex using the standard univariate GWAS procedure. Resulting univariate p-values and effect sizes were further combined in the MOSTest and min-P analyses to identify area, thickness and sulcal depth associated loci.

5.5.4 Cross validation

We performed 10 times cross validation within UK Biobank with approximately ⅔ training, ⅓ testing splits, performed randomly except for related individuals were kept together. Due to relatedness in the sample we wanted to ensure that individuals who were highly genetically related were not split across training and testing folds. We estimated relatedness using 'plink –genome' and from this defined relatedness clusters with individuals who were related as 3rd degree relatives (threshold>1/8). This resulted in 34,813 clusters. Across the 10 folds, the mean training sample size was 24,471 individuals (S.D. =9.5).

5.5.5 MOSTest Discovery

Consider *N* variants and *M* (pre-residualized) phenotypes. Let $z_{i,j}$ be a z-score from the univariate association test between ith variant and jth (residualized) phenotype and $\boldsymbol{z}_j$ be the vector of z-scores of the ith variant across phenotypes. Let $\boldsymbol{Y}$ be a matrix of (pre-residualized) phenotypes with *I* (individuals) rows and *M* (phenotypes) columns, and $\boldsymbol{R}$ be its correlation matrix. $\boldsymbol{R}$, being Hermitian, can be decomposed using singular valued decomposition as $\boldsymbol{R} = \boldsymbol{USU}^T$ (**U**

– unitary matrix, $\boldsymbol{S}$ – diagonal matrix with singular values on its diagonal). Consider the regularized version of the correlation matrix $\boldsymbol{R_r} = \boldsymbol{US_rU}^T$, where $\boldsymbol{S_r}$ is obtained from $\boldsymbol{S}$ by keeping r largest singular values and replacing the remaining with r^{th} largest. The MOSTest statistics for i[th] variant (scalar) is then estimated as $\chi_i = \boldsymbol{z}_i\boldsymbol{R}_\boldsymbol{r}^{-1}\boldsymbol{z}_i^T = \boldsymbol{z}_i\boldsymbol{US}_\boldsymbol{r}^{-1}\boldsymbol{U}^T\boldsymbol{z}_i^T$, where regularization parameter is selected separately for cortical area, thickness and sulcal depth to maximize the yield of genome-wide significant loci. As established in previous work[3–5] the largest yield for cortical surface area is obtained with r =10; the optimal choice for cortical thickness and sulcal depth was r =20. The distribution of the test statistics under null ($CDF_{\text{null}}^{\text{most}}$) is approximated from the observed distribution of the test statistics with permuted genotypes, using the empirical distribution in the 99.99 percentile and Gamma distribution in the upper tail, where shape and scale parameters of Gamma distribution are fit to the observed data. The p-value of the MOSTest test statistic for the i[th] variant is then obtained as $p_{most} = CDF_{\text{null}}^{\text{most}}(\chi_i)$.

5.5.6 min-P Discovery

Similar to the MOSTest analysis, consider *N* variants *M* and pre-residualized phenotypes. Let $z_{i,j}$ be a z-score from the univariate association test between i[th] variant and j[th] (residualized) phenotype and $\boldsymbol{z}_i$ be the vector of z-scores of the i[th] variant across phenotypes. The min-P statistics for the i[th] variant is then estimated as $y_i = 2\Phi\left(-\max_{j=1\ldots M}\left(\left|z_{i,j}\right|\right)\right)$, where Φ is a cumulative distribution function of the standard normal distribution. The distribution of the min-P test statistics under null ($CDF_{\text{null}}^{\text{minP}}$) is approximated from the observed distribution of the test statistics with permuted genotypes, using the empirical distribution in the 99.99[th] percentile and Beta distribution in the upper tail, where shape parameters of Beta distribution (α and β) are fit to the observed data. The p-value of the min-P test statistic for the i[th] variant is then obtained as $p_{minP} = CDF_{null}^{minP}(y_i)$.

5.5.7 Locus definition

Independent significant SNPs and genomic loci were identified in accordance with the PGC locus definition, as also used in FUMA SNP2GENE[19]. First, we select a subset of SNPs that pass genome-wide significance threshold $5x10^{-8}$, and use PLINK to perform a clumping procedure at LD r2=0.6, to identify the list of independent significant SNPs. Second, we clump the list of independent significant SNPs at LD r2=0.1 threshold to identify lead SNPs. Third, we query the reference panel for all candidate SNPs in LD r2 of 0.1 or higher with any lead SNPs. Further, for each lead SNP, it's corresponding genomic loci is defined as a contiguous region of the lead SNPs' chromosome, containing all candidate SNPs in r2=0.1 or higher LD with the lead SNP. Finally, adjacent genomic loci are merged if they are separated by less than 250 KB. Allele LD correlations are computed from EUR population of the 1000 genomes Phase 3 data. Obtained clumps of variants were considered as independent genome-wide significant genetic loci.

5.5.8 Replication of Discovered Variants

A schematic displaying the difference between min-P and MOSTest replication is displayed in Figure 5.1. For genome-wide significant loci defined in the training folds, we performed replication in test folds of UK Biobank, as well as the whole sample of ABCD. Let $\boldsymbol{X^{test}}$ represent the genotype matrix of individuals in the test set of *I* individuals and *N* variants and $\boldsymbol{Y^{test}}$ represent the phenotype matrix of *I* individuals and *M* (pre-residualized) phenotypes. Replication was performed in one of two ways, depending on whether the genetic variant was discovered using min-P or MOSTest. Firstly, for a min-P discovery, implicated by the association statistic $z_{i,j}$, the i^{th} variant, $\boldsymbol{x}_i^{test}$, is associated with the j^{th} (residualized) phenotype

$\boldsymbol{y}_j^{test}$, in the test set. Secondly, for MOSTest validation the i[th] discovered loci corresponds to a vector of mass univariate association statistics across all vertices $\boldsymbol{z_i}$ - these are used to generate projection weights to create a PolyVertex Score (PVS) [7], $\boldsymbol{x}_{PVS,i}^{test}$. This approach largely mirrors the use of polygenic scores used in genetics, where here we are aggregating effects of vertices across the cortex. For polygenic scores, it is well known that the correlation structure (i.e. linkage disequilibium) across the genome can result in suboptimal out of sample performance. This has motivated techniques like LD-Pred[20] and PRSice[21] to first account for this genomic correlation before generating scores. Similarly, we decorrelate the association statistics, $\boldsymbol{z}_i$, as $\boldsymbol{w}_i = \widetilde{\mathbf{R}}_r \mathbf{z}_i$ using the regularized correlation matrix $\widetilde{\mathbf{R}}_r$that was learned in the training fold. We then generate the polyvertex score for the i[th] genomic variant as the dot product of $\boldsymbol{w}_i$ with the (pre-residualized) phenotype matrix, $\boldsymbol{Y}^{test}$, in the test set: $\widehat{\boldsymbol{x}}_{PVS,i}^{test} = \boldsymbol{w}_i \boldsymbol{Y}^{test}$.

To perform the association in test set, both of min-P and MOSTest/PVS replications, we used linear mixed-effect models (LMMs) to control for genetic/family relatedness – this is particularly relevant for the ABCD dataset which has a high degree of family relatedness. We used a single fixed effect of the discovered variant, x_i, and a random effect intercept using a grouping, *c*, of either: i) genetic relatedness cluster (defined above) for UK Biobank replication or ii) family id (rel_family_id) for ABCD replication. The response variable was either a) the most significant vertex for min-P validation, $y_{minP,i}$, or b) the computed PVS, $\widehat{\boldsymbol{x}}_{PVS,i}^{test}$, for MOSTest. For min-P replication:

$$y_{minP,i} \sim x_i + (1|c)$$

And for MOSTest replication:

$$\hat{x}_{PVS,i} \sim x_i + (1|c)$$

As the phenotype matrix, $\boldsymbol{Y}^{test}$, was pre-residualized for covariates before taking the most significant vertex (min-P) or computing the PVS (MOSTest) we did not need to control for other covariates. We fit an LMM for each discovered locus in training set. For both min-P and MOSTest validation, we one-tailed p values from t statistics of the fixed effect as we assume the effect to be in the same direction for training folds and test sets. To define replicated loci we use a nominal p value threshold of 0.05 for associations. Due to the higher number of discovered loci for MOSTest vs min-P, we additionally report the number of loci validated at a Bonferroni corrected threshold, where this number of independent tests is taken to be the number of discovered loci in the training set. This corrected threshold penalizes MOSTest to a greater extent than min-P for discovering a larger number of loci. We calculate the variance explained by the single lead i[th] variant in the replication sample from t statistics of x_i from fitted LMMs and degrees of freedom (*df*) as: $R^2 = \frac{t^2}{(t^2 + df)}$. We report the average and standard deviation (σ) of this value across training folds.

Funding

This work was supported by Kavli Innovative Research Grant under award number 2019-1624, and grant R01MH122688 and RF1MH120025 funded by the National Institute for Mental Health (NIMH).

Acknowledgement

Data used in the preparation of this article were obtained from the ***Adolescent Brain Cognitive Development℠ Study*** (***ABCD Study®***) (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000

children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers:
U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147

A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators.html. ABCD Study consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD Study consortium investigators.
The ABCD data repository grows and changes over time. The ABCD data used in this came from [NIMH Data Archive Digital Object Identifier (10.151.54/1519007)].

Chapter 5, in full, is available as a preprint on BioRxiv and has been submitted for publication. Loughnan R, Shadrin A, Frei O, van der Meer D, Zhao, W, Palmer C, Thompson W, Makowski C, Jernigan T, Andreassen O, Chieh Fan C, Dale A. The dissertation/thesis author was the joint primary investigator and author of this paper.

Conflict of Interest Statement

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

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

References

1. Panizzon, M. S. *et al.* Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. *Cereb. Cortex* ***19***, 2728–2735 (2009).

2. 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***, 304–314 (2012).

3. van der Meer, D. *et al.* Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* ***11***, 3512 (2020).

4. van der Meer, D. *et al.* The genetic architecture of human cortical folding. *bioRxiv* 2021.01.13.426555 (2021) doi:10.1101/2021.01.13.426555.

5. Shadrin, A. A. *et al.* Multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *bioRxiv* 2020.10.22.350298 (2021) doi:10.1101/2020.10.22.350298.

6. Zhao, W. *et al.* The Bayesian polyvertex score (PVS-B): a whole-brain phenotypic prediction framework for neuroimaging studies. *bioRxiv* 813915 (2019) doi:10.1101/813915.

7. Zhao, W. *et al.* Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. *Cereb. Cortex* ***31***, 1478–1488 (2021).

8. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* ***12***, 44 (2020).

9. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. (2018).

10. Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M. & Brown, T. T. Through Thick and Thin: a Need to Reconcile Contradictory Results on Trajectories in Human Cortical Development. *Cereb. Cortex* ***27***, 1472–1481 (2017).

11. Denny, J. C. & Collins, F. S. Precision medicine in 2030—seven ways to transform healthcare. *Cell* ***184***, 1415–1419 (2021).

12. Fan, C. C. *et al.* Modeling the 3D Geometry of the Cortical Surface with Genetic Ancestry. *Curr. Biol.* ***25***, 1988–1992 (2015).

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

14. Raj, A., Stephens, M. & Pritchard, J. K. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* ***197***, 573–589 (2014).

15. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* ***48***, 1284 (2016).

16. Loughnan, R. J. *et al.* Gene-experience correlation during cognitive development: Evidence from the Adolescent Brain Cognitive Development (ABCD) Study. *bioRxiv* 637512 (2021) doi:10.1101/637512.

17. Palmer, C. E. *et al.* Delineating genetic and familial risk for psychopathology in the ABCD study. *medRxiv* 2020.09.08.20186908 (2020) doi:10.1101/2020.09.08.20186908.

18. Dale, A. M. *et al.* Automatically Parcellating the Human Cerebral Cortex. *Cereb. Cortex* ***14***, 11–22 (2004).

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

20. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* ***97***, 576–592 (2015).

21. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* ***15***, 2759–2772 (2020).