Exploring Molecular Genetic Influences on Prenatal Development: Evidence from a Genotyped Panel of Mothers and Children

Sam Trejo[1*]

1. Graduate School of Education, Stanford University
* Send correspondence to samtrejo@stanford.edu.

September 2019

Acknowledgements
This work has been supported by the National Science Foundation (DGE-1656518), the Institute of Education Sciences (R305B140009), and the Russell Sage Foundation (96-17-04). We are grateful to Benjamin Domingue, Liam Hill, and Dalton Conley for helpful comments. We also wish to thank Dan Mason, Rosie McEachan, and the Born and Bradford (BiB) staff for access to the restricted-use genetic data. BiB receives core infrastructure funding from the Wellcome Trust (WT101597MA) and the National Institute for Health Research (NIHR) under its Collaboration for Applied Health Research and Care (CLAHRC) for Yorkshire and Humber and Clinical Research Network (CRN) research delivery support. Further support for genome-wide and multiple 'omics measurements is from the UK Medical Research Council (G0600705), National Institute of Health Research (NF-SI-0611-10196), US National Institute of Health (R01 DK10324), and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 66954. All funders had no role in the design of the study, the collection, analysis, or interpretation of the data; the writing of the manuscript, or the decision to submit the manuscript for publication. Any opinions expressed are those of the authors alone.

1. Introduction

Over the last two decades, the social and biomedical sciences have witnessed an explosion of research on the *fetal origins hypothesis*,[1–4] also referred to as *the developmental origins of health and disease*. A large body of research demonstrates that the benefits of healthy prenatal development, commonly proxied by birth weight, cascade out across the life course and affect valued economic, educational, and health outcomes.[5–14] In this way, the circumstances of mothers in one generation can powerfully shape the distribution of resources and opportunity in the next.

Like virtually all human characteristics,[16] variation between individuals in prenatal development is produced by a combination of both genetic and environmental factors.[17] However, genetic factors largely remain a "black box" in existing models of life course development.[18] Genetic factors are typically relegated to the error term[19] or differenced out entirely using a quasi-experimental identification strategy.[20] In the fetal origins hypothesis literature, the predominant approach has been to study monozygotic twins with a family fixed effect model, thereby comparing differences in birth weight between twins to differences in outcomes between twins. Consistently, the heavier twin of the pair is more likely to have greater cognitive ability, higher wages, and improved adult health.

Understanding how environmental exposures and individual differences interact to shape neonatal health and human development has been a recent focus of empirical economics and sociology.[21,22] Policymakers tasked with adjudicating between potential public policies benefit not only from an understanding of the effects of specific interventions but also from a broader conceptualization of the underlying causal model at hand. While the prenatal development literature's now-common family fixed effects strategy using twins eliminates confounding from omitted genetic factors, it fails to identify *how* those genetic factors enter into the chain of events that produces birth weight and downstream outcomes. Thus, precious little is known about the ways in which genetic endowments influence prenatal development and whether genetic influences on prenatal development translate into positive life course development.

An increased understanding of the role that genetic factors play in the causal model that connects prenatal development to valued outcomes across the life course is useful to researchers and policymakers for a few reasons. First, genetic and environmental influences may not be additive and instead might interact in shaping an individual's prenatal development.[18,23,24] Second, genetic and environmental influences on prenatal development may have different implications for future life course development. For example, while increases in birth weight due to environmental factors have been shown to boost valued economic, educational, and health outcomes, increases in birth weight due to genetic factors may not induce the same benefits.

Historically, social scientists have been limited in their ability to empirically grapple with individual genetic factors, as the human genome was unobservable to researchers until the turn of the 20th century. Fortunately, recent decreases in the cost of DNA sequencing and computational developments in the processing of large amounts of genetic data have spurred new discoveries in the genetics of complex, behavioral traits. Novel insights from molecular genetics allow new avenues for social scientists to model genetic influences on human behavior.[26,27] This year, a genome-wide association study has led to the discovery of specific genetic variants that predict a child's birth weight.[17] Following the extant literature, we use birth weight as a proxy for healthy prenatal development.

In this paper, we summarize these newly identified genetic variants related to birth weight into individual-level linear indices called polygenic scores. We combine these scores with a genetically informed longitudinal study of pregnant mothers and their resulting children to test a

series of hypothesis regarding genetic influences on birth weight. Specifically, we ask three main questions: First, do child genetic factors predict birth weight and does this association in part represent a causal effect? Second, do environmental factors, such as maternal body mass index and gestational age, moderate the relationship between child genetics and birth weight? Third, do child genetics related to birth weight predict socioemotional and cognitive development in childhood?

2. GWAS and Birth Weight

The advent and proliferation of a key tool, the genome-wide association study (GWAS), has ushered in remarkable developments to our understanding of the genetics of complex traits and the biology of human behavior and disease.[28,29] A GWAS probes the relationship between a trait and regions of the genome via large data sets containing individual-level information on genotype and phenotype.[30] A recent study conducted parallel GWAS on the maternal and fetal genetic influences on birth weight using data on over 500,000 individuals and revealed new insights about the genetic underpinnings of prenatal development.[17] The authors identified 190 independent regions of the genome that are significantly associated with birth weight at a statistical threshold of 5E-8 (representing a Bonferroni correction for approximately one million independent hypotheses, or genetic loci, tested). A child's polygenic score generated from these GWAS results explains roughly 4% of the variance in birth weight in independent samples (author's calculations).

Socials scientists have already begun unpacking GWAS findings for the genetic factors that influence other complex traits like education.[18,31,32] However, few studies have used polygenic scores to follow up on the recent birth weight GWAS findings. This paper is a first step in understanding how these genetic discoveries fit into the models of health and human development studied by social scientists. In section A1 of the appendix, we discuss some limitations with the current state of the genetics literature and the resulting interpretational challenges.

3. Data

3a. Born in Bradford

We use an analytic sample drawn from the Born in Bradford (BiB) study, a longitudinal multi-ethnic birth cohort study of the impact of environmental, psychological and genetic factors on maternal and child mental and physical health.[33] The full BiB cohort consists of ~12,500 women recruited from 2007 to 2010 across ~13,800 pregnancies, which resulted in ~13,900 children. The BiB cohort is roughly equal parts White British and Pakistani. A strength of the BiB study is its unusually rich, clinical and administrative measures, which begin during pregnancy and continue through when the children are in primary school. At birth, genetic samples were assayed from a subsample of consenting mothers and children. Clinical healthcare records were collected for mothers and longitudinal administrative educational records were collected for children.

3b. Sample Restriction

Although a strength of the cohort, ancestral diversity raises issues in studies of genetic prediction (Duncan et al., 2018). Because we generate polygenic scores for birth weight using results from genome-wide association studies of individuals of European ancestry,[17] we first restrict our sample to mothers and children who self-identified as British and were also of European ancestry (as identified via the first 2 principal components of individual genotype).[34] Our analytic sample was further restricted to mothers and children of European ancestry for who genetic data and test scores are available (N = ~3,000 mother-child dyads).

3c. Non-Genetic Measures

To capture the salient aspects of the child's prenatal environment, we use measures of the mother's physical health, mental health, and socioeconomic conditions during pregnancy from a maternal intake survey. We also employ measures of child health at birth using clinical medical records. Finally, we index child physical and cognitive development through in-school assessments and clinical growth measures. See Section A2 of the appendix for more details on the non-genetic measures used in this study.

3d. Genetic Measures

We used Illumina HumanCore Exome 12 and 24 BeadChip arrays (Version 1/1.1; Illumina, Hayward, CA) to assay common single-nucleotide polymorphism variation from the genomes of children in BiB using umbilical cord bloods. To calculated polygenic scores, we matched child genotypes from the Born in Bradford data with the most recent GWAS results for birth weight reported by the Early Growth Genomics Consortium[17]. For each genotype, we counted the number of birth weight-associated alleles (0, 1, or 2), multiplied this count by the effect size estimated in the original genome-wide association study, and then summed weighted counts across all genotypes to calculate each BiB participants' polygenic score. All genome-wide analyses were performed using Plink.[35]

4. Methods

4a. Are child genetic factors associated with birth weight and does this association represent, in part, a causal effect?

To test if mother and child polygenic scores are independently associated with birth weight, we first use only unrelated individuals and fit the following model for child *i* in family *j*:

$$ln(BW_{ij}) = \hat{\alpha}_0 + \hat{\alpha}_1 \widehat{PGS}_j^{\mathrm{M}} + \hat{\alpha}_2 \widehat{PGS}_{ij}^{\mathrm{F}} + \boldsymbol{X_{ij}}\widehat{\boldsymbol{\Theta}} + \epsilon_{ij} \quad \text{(4a.i)}$$

ln (BW_{ij}): Natural logarithm of birth weight of child i in family j

$\widehat{PGS}_{ij}^{\mathrm{F}}$: Fetal birth weight polygenic score for child i in family j

$\boldsymbol{X_{ij}}$: Vector of individual i covariates including sex, age, and the first 10 PCs of genotype

In all of our analyses using polygenic scores, we follow the genetics literature and control for child sex, age and the first ten principal components of European ancestry genotype to account for population stratification and increase robustness of our findings.[34] The association between a child's polygenic score and birth weight cannot be interpreted as a causal effect. However, conditional on parental genotype, child genotype is randomly assigned through a process known as genetic recombination.[40] Thus, family fixed effect regression models that compare genetic differences in siblings to phenotypic differences in siblings are the gold standard for testing whether genetic differences are causally related to outcomes.[32,36–39] In that vein, we next use only sibling pairs and fit the following model for child *i* in family *j* with a family fixed effect:

$$\ln\,(BW_{ij}) = \hat{\alpha}_0 + \hat{\alpha}_1 \widehat{PGS}_{ij}^{\mathrm{F}} + \boldsymbol{X_{ij}}\widehat{\boldsymbol{\Theta}} + \gamma_j + \epsilon_{ij} \quad \text{(4a.ii)}$$

γ_j: Family j fixed effect

4b. Do environmental factors moderate the relationship between child polygenic score and birth weight?

To test which mechanisms might moderate the relationship between child polygenic scores and birth weight, we fit the model for child *i* in family *j* for a variety of candidate prenatal environmental features:

$$ln(BW_{ij}) = \hat{\alpha}_0 + \hat{\alpha}_1 \widehat{PGS}^{\mathrm{F}}_{ij} + \hat{\alpha}_2 Env_{ij} + \hat{\alpha}_3 (\widehat{PGS}^{\mathrm{F}}_{j} * Env_{ij}) + \boldsymbol{X_{ij}}\widehat{\boldsymbol{\Theta}} + \epsilon_{ij}$$

(4b.i)

The tested prenatal environmental features include maternal body mass index and gestational age. To the extent that $\hat{\alpha}_3$ in Equation 4b.i is statistically significant, we have evidence that the change in the conditional expectation associated with the change in a given neonatal environment on birth weight depends on child genetics (i.e. a gene-environment interaction).

4c. Do child genetic factors related to birth weight predict other developmental outcomes?

To test if child genetics predict other outcomes, we fit the model for child *i* in family *j*:

$$Dev_{ij} = \hat{\alpha}_0 + \hat{\alpha}_1 \widehat{PGS}^{\mathrm{F}}_{ij} + \boldsymbol{X_{ij}}\widehat{\boldsymbol{\Theta}} + \epsilon_{ij}$$

(4c.i)

Dev_{ij}: Developmental outcome for child *i* in family *j*

The tested child human capital development outcomes include measures of socioemotional development, cognitive development, and registered special educational needs. Socioemotional development is measured at age 5 and cognitive development is measure at age 7. To the extent that $\hat{\alpha}_1$ and in Equation 4c.i are positive and statistically significant, we have evidence that child genetic factors that influence birth weight are also positively associated with downstream human capital development.

Finally, to test if, child genetics predict other outcomes *net of birth weight*, we fit the following model for child *i* in family *j*:

$$Dev_{ij} = \hat{\alpha}_0 + \hat{\alpha}_1 \widehat{PGS}^{\mathrm{F}}_{ij} + \ln(BW_{ij}) + \boldsymbol{X_{ij}}\widehat{\boldsymbol{\Theta}} + \epsilon_{ij}$$

(4c.ii)

5. Preliminary Results

Our initial analyses show that a child's genetic factors significantly predict their birth weight. Sibling analyses demonstrate that a portion of this association represents a causal effect. Further, our analyses provide preliminary evidence for a gene-environment interaction between birth weight polygenic score and both maternal body mass index and gestational length. Specifically, for babies born to mothers with high body mass index, a marginal increase in polygenic score is associated with a larger increase in birth weight than babies born to mothers with low body mass index. In addition, for babies that spend a longer time in the womb, a marginal increase in polygenic score is associated with a larger increase in birth weight than babies that spend relatively less time in the womb. Figure 1 provides a graphical depiction of these results.

[Insert Figure 1 Here]

Our analyses provide preliminary evidence that the genetics of birth weight are associated with healthy cognitive and socioemotional development in early childhood. Babies with higher birth weight polygenic scores have increased socioemotional scores at age five, increased cognitive measures at age seven, and are less likely to be categorized by their school as special education. Figure 2 provides a graphical depiction of these results.

[Insert Figure 2 Here]

References

1. Barker D. The fetal and infant origins of adult disease. *BMJ Br Med J*. 1990;301(156):1111. doi:10.1136/bmj.301.6761.1111.
2. Barker D. The fetal and infant origins of disease. *Eur J Clin Invest*. 1995. doi:10.1111/j.1365-2362.1995.tb01730.x.
3. Victora CG, Adair L, Fall C, et al. Maternal and child undernutrition: consequences for adult health and human capital. *Lancet*. 2008. doi:10.1016/S0140-6736(07)61692-4.
4. Almond D, Currie J. Killing Me Softly: The Fetal Origins Hypothesis. *J Econ Perspect*. 2011;25(3):153-172. doi:10.1257/jep.25.3.153.
5. Conley D, Bennett NG. Is biology destiny? Birth weight and life chances. *Am Sociol Rev*. 2000;65(3):458. doi:10.2307/2657467.
6. Behrman JR, Rosenzweig MR. Returns to Birthweight. *Rev Econ Stat*. 2004;86(2):586-601. doi:10.1162/003465304323031139.
7. Almond D, Chay K, Lee D. The Costs of Low Birth Weight. *Q J …*. 2005;120(3):1031-1083. doi:10.1162/003355305774268228.
8. Conley D, Strully KW, Bennett NG. Twin differences in birth weight: The effects of genotype and prenatal environment on neonatal and post-neonatal mortality. *Econ Hum Biol*. 2006. doi:10.1016/j.ehb.2005.12.001.
9. Black SE, Devereux PJ, Salvanes KG. From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *Q J Econ*. 2007;122(February):409-439. doi:10.1162/qjec.122.1.409.
10. Oreopoulos P, Stabile M, Walld R, Roos LL. Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins. *J Hum Resour*. 2008;43:88-138. doi:10.1353/jhr.2008.0003.
11. Royer H. Separated at girth: US twin estimates of the effects of birth weight. *Am Econ J Appl Econ*. 2009. doi:10.1257/app.1.1.49.
12. Torche F, Echevarría G. The effect of birthweight on childhood cognitive development in a middle-income country. *Int J Epidemiol*. 2011;40(4):1008-1018. doi:10.1093/ije/dyr030.
13. Figlio D, Guryan J, Karbownik K, Roth J. The effects of poor neonatal health on children's cognitive development? *Am Econ Rev*. 2014. doi:10.1257/aer.104.12.3921.
14. Boardman J, Powers D, Padilla Y, Hummer R. Low birth weight, social factors, and developmental outcomes among children in the United States. 2002;39(2):353-368.
15. Rossin-Slater M. Promoting health in early childhood. *Futur Child*. 2015. doi:10.1353/foc.2015.0002.
16. Polderman TJC, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47(7):702-709. doi:10.1038/ng.3285.
17. Warrington N, Beaumont R, Horikoshi M, Day F, Helgeland O. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet*. 2019;(Ld). doi:10.1038/s41588-019-0403-1.
18. Papageorge N, Thom K. Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study. 2017;055654. doi:10.2139/ssrn.2982606.
19. Lillard LA, Willis RJ. Dynamic Aspects of Earning Mobility. *Econometrica*. 1978. doi:10.2307/1911432.
20. Wooldridge JM. *Introductory Econometrics : A Modern Approach*.; 2015.
21. Almond D, Currie J. *Human Capital Development before Age Five*. Vol 4.; 2011.

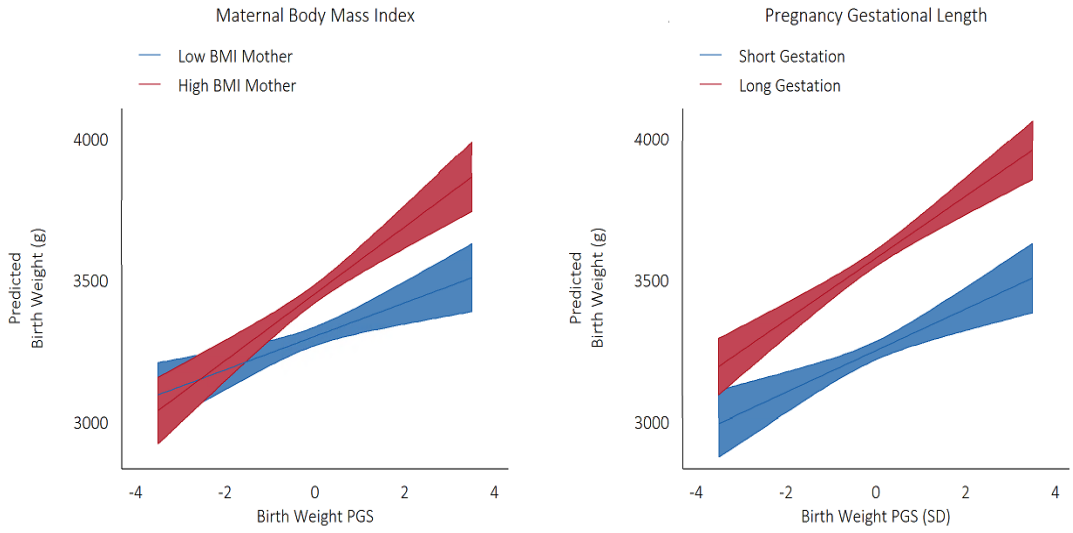
doi:10.1016/S0169-7218(11)02413-0.
22. Almond D, Currie J, Duque V. Childhood Circumstances and Adult Outcomes: Act II. *J Econ Lit*. 2018. doi:10.1257/jel.20171164.
23. Trejo S, Belsky D, Boardman J, et al. Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Heath. *Sociol Sci*. 2018. doi:10.15195/v5.a22.
24. Barcellos SH, Carvalho LS, Turley P. Education can reduce health differences related to genetic risk of obesity. *Proc Natl Acad Sci*. 2018. doi:10.1007/BF00871674.
25. Armstrong-Carter E, Trejo S, Hill L, Crossley K, Mason D, Domingue BW. The Earliest Origins of Genetic Nurture: Prenatal Environment Mediates the Association Between Maternal Genetics and Child Development. *psyarxiv*. 2019. doi:10.22201/fq.18708404e.2004.3.66178.
26. Beauchamp JP, Cesarini D, Johannesson M, et al. Molecular Genetics and Economics. *J Econ Perspect*. 2011. doi:10.1257/jep.25.4.57.
27. Benjamin DJ, Cesarini D, Chabris CF, et al. The Promises and Pitfalls of Genoeconomics. *Annu Rev Econom*. 2012. doi:10.1146/annurev-economics-080511-110939.
28. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90(1):7-24. doi:10.1016/j.ajhg.2011.11.029.
29. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017. doi:10.1016/j.ajhg.2017.06.005.
30. Pearson T a, Manolio T a. How to interpret a genome-wide association study. *JAMA*. 2008;299(11):1335-1344. doi:10.1001/jama.299.11.1335.
31. Barth D, Papageorge N, Thom K. Genetic Endowments and Wealth Inequality. 2018:1-94. doi:10.3386/w24642.
32. Belsky DW, Domingue BW, Wedow R, et al. Genetic analysis of social-class mobility in five longitudinal studies. *Proc Natl Acad Sci*. 2018. doi:10.1073/pnas.1801238115.
33. Wright J, Small N, Raynor P, et al. Cohort profile: The born in bradford multi-ethnic family cohort study. *Int J Epidemiol*. 2013. doi:10.1093/ije/dys112.
34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006. doi:10.1038/ng1847.
35. 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. doi:10.1186/s13742-015-0047-8.
36. Rietveld CA, Conley D, Eriksson N, et al. Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychol Sci*. 2014;25(11):1975-1986. doi:10.1177/0956797614545132.
37. Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD. Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health. *AERA Open*. 2015;1(3):1-13. doi:10.1177/2332858415599972.
38. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018. doi:10.1038/s41588-018-0147-3.
39. Trejo S, Domingue BW. Genetic Nature or Genetic Nurture? Quantifying Bias in Analyses Using Polygenic Scores. *bioRxiv*. 2019;(96):524850. doi:10.1101/524850.

40. Conley D, Fletcher J. *The Genome Factor What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*.; 2017.
41. Martschenko D, Trejo S, Domingue BW. Genetics and Education: Recent Developments in the Context of an Ugly History and an Uncertain Future. *AERA Open*. 2019;5(1):1-15. doi:10.1093/ije/dyx041.
42. Hamer DH, Sirota L. Beware the chopsticks gene. *Mol Psychiatry*. 2000. doi:10.1038/sj.mp.4000662.
43. Novembre J, Johnson T, Bryc K, et al. Genes mirror geography within Europe. *Nature*. 2008. doi:10.1038/nature07331.
44. Mostafavi H, Harpak A, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *bioRxiv*. 2019. doi:10.1101/629949.
45. Duncan L, Shen H, Gelaye B, et al. Analysis of Polygenic Score Usage and Performance in Diverse Human Populations. *Nat Commun*. 2019. doi:10.1101/398396.
46. Bertrand M, Mullainathan S. Are Greg and Emily more employable then Lakisha and Jamal? A field experiment on labor market discrimination. *Am Econ Rev*. 2004.
47. Fryer RG. An Empirical Analysis of Racial Differences in Police Use of Force. *J Polit Econ*. 2019. doi:10.1086/701423.
48. Aizer A, Currie J. The intergenerational transmission of inequality: Maternal disadvantage and health at birth. *Science (80-)*. 2014. doi:10.1126/science.1251872.
49. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. *Psychol Med*. 1979. doi:10.1017/S0033291700021644.

Figures

Figure 1

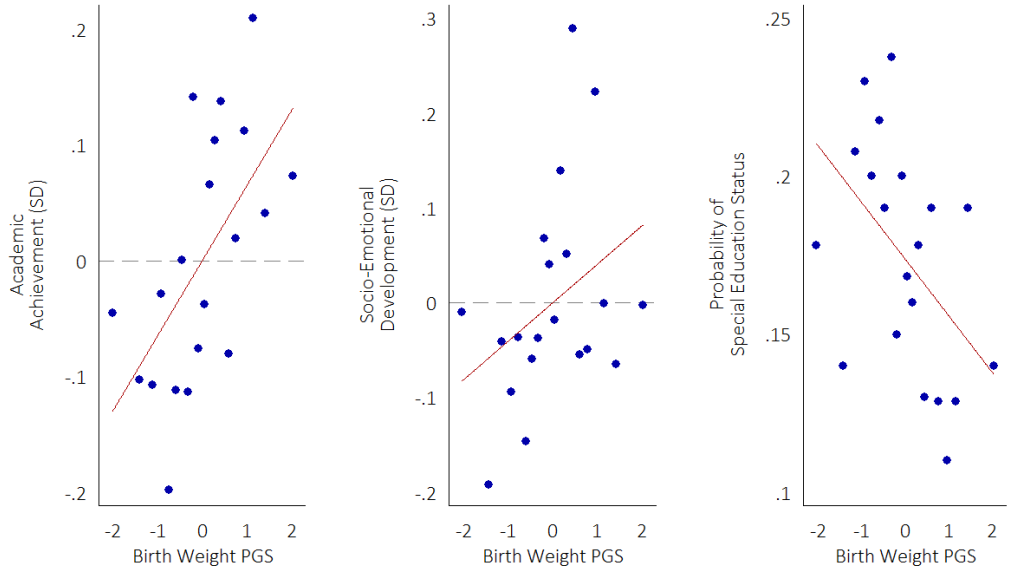
Gene-Environment Interactions



Underlying regression model includes controls for age of mother and 10 PCs.

Figure 2

The Genetics of Birth Weight Predicts Childhood Development



Each bin represents roughly 100 children. Birth Weight PGS was residualized on sex and 10 PCs.

Appendix

A1. Limitations and Interpretational Challenges

Given the fraught history of discourses regarding individual genetic differences in human health and abilities [41], we aim to follow the good example set by Papageorge and Thom (2017) and be explicit about the interpretational limitations of our molecular genetic analyses. Specifically, we note three important caveats regarding our use of polygenic scores for birth weight.

First, the GWASs from which we derive our allelic weights are associational in nature. Thus, our polygenic scores do not capture only the causal effect of an individual's genes. In particular, GWAS are at risk for confounding from omitted variables related to what geneticist call *population stratification*. Population stratification refers to the fact that an individual's genome contains fine grain information about their place in the intricate structure of a population[42–44], meaning that GWASs for complex traits may partly identify genes related to confounding environmental variables such as ancestry, geography, or socioeconomic status. Including the first ten principal components of individual genotype has been shown to be an effective strategy for reducing confounding due to population stratification[34], but some concerns remain.

Second, we restrict our sample to mothers and children of European ancestry. We do this because interpreting polygenic score analyses in diverse samples is not currently feasible (Martin et al. 2017). Moreover, because the birth weight GWAS from which we derive our polygenic scores was conducted in European individuals, predictions in other groups has been shown to be diminished [45]. While we recognize the importance of research in more diverse samples, we can only responsibly move as quickly as the developments in human genomics. We therefore consider our analysis as a first step in considering the role of the genes in the production of birth weight and human capital. Because both environmental exposures[46,47] and the distribution of birth weight[48] vary across ancestries, we emphasize that the results in this paper apply only to European ancestry individuals.

Third, our polygenic scores contain large amounts of measurement error. It is therefore impossible to use the current birth weight polygenic score to draw conclusions about the relative importance of genetic endowments versus environmental factors for birth weight and downstream human capital accumulation. Nonetheless, as others have pointed out,[18] polygenic scores allow for the testing of the direction of interaction effects. As sample sizes for genetic discovery increase and geneticists hone their statistical methodologies, the amount of measurement error in polygenic scores will approach zero. At that point, polygenic scores may be useful tools for controlling out genetic factors in studies of the environment or decomposing the variance between genetic and environmental factors. Given that the causal processes that connect prenatal development to human capital accumulation within families may differ from the causal processes between families, results from family fixed effect designs using monozygotic twins may fail to fully generalize. Thus, future advances in molecular genetic data may play an important role in controlling for genetic factors in studies of unrelated individuals.

A2. Non-Genetic Measures

Maternal Prenatal Characteristics

Body Mass Index: Maternal body mass index was calculated using height and weight measurements that were directly assessed by medical staff at 26-28 weeks gestation.

Socioeconomic Status: Maternal socioeconomic status is measured by the first principle component of five maternal characteristics measured at 26-28 weeks gestation: educational

attainment, receipt of mean-tested governmental benefits, neighborhood Index of Multiple Deprivation, partner cohabitation, and home ownership.

Health at Birth

Birth Weight: Child birth weight was clinically assessed by hospital staff at delivery and recorded in grams.

Gestational Age: Child gestational age was obtained from maternal clinical pregnancy records.

Size for Gestational Age: A child was deemed small or large for their gestational age if their birth weight was below 10th percentile or above 90th percentile birth weight for their gestational age, respectively.

Child Development.

Subjective Child Socioemotional Development: Subjective child development is measured through the Early Years Foundation Stage Profile (EYFSP), a teacher-led observational assessment conducted towards the end of the child's first year at school. The profile is completed at the end of the child's first year in school, when they are usually five years old. The profile measures children's attainment in seven main areas of learning: communication and language; expressive arts and design; literacy, mathematics; physical development; personal, socio and emotional development and understanding the world. Teachers completed the assessment for each child based on their knowledge and observations of that child. Subscales scores were summed and standardized within sample to be mean 0 and standard deviation 1.

Objective Child Cognitive Development: Objective academic achievement was assessed using the Key Stage 1 Assessment, a standardized test towards administered at the end of a child's third year in school when the children are typically seven years old. The Key Stage 1 Assessment includes math, reading and science subscales. Subscale scores were summed and standardized within sample to be mean 0 and standard deviation 1.

Child Special Needs: A binary indicator for whether a child has a documented special educational need was extracted from the child's educational records when the child was seven years old.