Interactive Effects of in Utero Nutrition and Genetic Inheritance on Cognition: New Evidence Using Sibling Comparisons[1]

C. Justin Cook
Yale University

Jason M. Fletcher
Yale University

Abstract: Much research examining gene-environment interactions are not able to leverage exogenous variation in environmental exposure. This is potentially problematic if there is gene-environment correlation or gene-gene interactions. This paper explores the importance of this issue by extending previously findings of interaction between the *FADS2* gene and early life nutrition in explaining later-life IQ. Using sibling pairs in the Wisconsin Longitudinal Study we show that, while our baseline replication attempts are successful—we find similar results as the previous literature that genotype moderates the impact of early nutrition on later IQ—employing sibling comparisons shows the results and framework to be fragile to omitted family-level variables. The example has wider implications for the practice of investigating gene-environment interactions when the environmental exposure is not exogenous, and robust measures of the genome are not controlled in the analysis.

[1] WLS acknowledgement. Fletcher and Cook gratefully acknowledge research support from the NICHD (R21HD071884).

1. Introduction

It is well known that intellectual development is a product of both genetic and environmental factors. In particular, early nutrition (including in utero) has lifelong effects on a range of health and economic outcomes. Evidence from the Dutch Hunger Winter (Stein et al. 1975) has shown that famine conditions suffered in utero led to increases in adult obesity and mental illness. Related findings in the economics literature have shown that birth weight differences are associated with long term differences in IQ, education, and earnings (Black et al. 2007). In this study we seek to better understand the effects of early childhood nutrition for later life outcomes; in particular, we are concerned with the differential effects of early childhood environments from variation in the genome.

Understanding potential interactions between the "nature" and "nurture" domains has also been an increasingly common direction that has linked social and biological sciences and has led to novel findings that suggest focusing on "nature" or "nurture" in isolation misses important channels determining intellectual development. A key investigation along these lines is from Caspi et al (2007), who show a replicated interaction effect between early nutrition, as measured by breastfeeding, and a specific genetic variation thought to modify dietary fatty acids, which itself is potentially important in cognitive development. In particular, the authors interact two genetic variations in the *FADS2* gene with breastfeeding measures to predict childhood IQ outcomes. They find that, in two different study populations, individuals carrying the GG genotype of SNP rs174575 had no advantage or disadvantage from breastfeeding while those with at least one C-allele had a large (6.4 point) IQ advantage over individuals who were not breastfed. Although the authors check for common confounding influences, the potential that the genetic variants were correlated with the environmental exposure or that there could be gene-gene interaction rather than gene-environment interaction remains.

Previous studies that have focused on the potential interactions between a phenotype's inherent genetic endowment and the effect of this endowment in varied environments studies have been unable to fully control for unseen, or omitted, variables (i.e., unobserved genetic and environmental effects) that may lead to a spurious GxE relationship. To correct for this, we employ a sibling fixed effects model within the GxE framework. The intention is to mitigate potentially unseen gene-gene interactions or gene-environment correlation (rGE). In theory, siblings are identical in 50% of genes; therefore, the use of sibling fixed effects should eliminate 50% of the unobserved genetic variation that could be associated with the candidate gene in the GxE study. The use of sibling fixed effects does not completely eliminate the possibility for gene-gene interactions; however, the use of

sibling fixed effects should provide a viable check for potentially confounding GxG or rGE effects. Additionally, the use of sibling fixed effects allows for greater control of unobserved, shared family environments.

In this paper, we extend the GxE approach of Caspi et al. by using sibling comparisons from the Wisconsin Longitudinal Study (see Conley and Rauscher 2010 for the only other example of this strategy).[2] In particular, we examine the potential interactive effects between early favorable nutrition status, as measured by birth weight, and variation in the *FADS2* genotype in predicting young adult IQ. The main preliminary finding is that, while our baseline replication attempts are successful—we find similar results as the previous literature that genotype moderates the impact of early nutrition on later IQ—employing sibling comparisons shows the results and framework to be fragile to omitted family-level variables. The example has wider implications for the practice of investigating gene-environment interactions when the environmental exposure is not exogenous, and robust measures of the genome are not controlled in the analysis.

1.1 Early Nutrition Environment

A growing body of research is concerned with the theoretical underpinnings (i.e., natural selection) for the differential genetic response to particular environments. Foremost among these is the differential susceptibility hypothesis, which is more popularly known as "orchids and dandelions" (Belsky 2005). The main idea is that some individuals (orchids) are more sensitive to environmental cues, thriving in good environments while struggling in harmful environments. Other individuals (dandelions) are relatively uninfluenced by variation in the environment, achieving the same outcomes regardless of the environment in which they are placed. In other words, the association of a genetic variant to an outcome is dependent upon the environment in which this variant is placed, with some variants being sensitive (orchids) while others are robust (dandelions). This hypothesis provides a reason as to why "harmful" genetic variants have persisted into contemporary times.[3]

[2] Using the Add Health data set, Conley and Rauscher (hereafter CR) show that previous GxE findings are indeed suspect when controlling for a shared familial environment. CR show this for both MZ and DZ twin pairs. The use of MZ twins, while controlling for potential GxG effects within pairs, does not fully control for environmental differences between sibling pairs. To correct for this CR, use sibling fixed effects within DZ twins. The DZ analysis is similar to the current work, where DZ twins are as genetically similar as non-twin siblings.

[3] Where harmful in this case is determined by interaction with a particular environment (see e.g., Caspi et al. 2002, Caspi et al. 2003, Guo et al. 2008).

Considering the differential susceptibility hypothesis, the use of breastfeeding as an environmental difference is problematic. The main argument by Caspi et al. is that certain variants of the *FADS2* gene interact with variation in breastfeeding to produce differential IQ scores. This implies that specific genetic variants were advantageous in specific breastfeeding environments. But given that the environmental difference is a relatively recent occurrence (circa 1950; Castilho and Barros 2010), it may be unlikely that genetic variation has been selected to this particular environment over such a short time window. In other words, a counter hypothesis is that variation in *FADS2* is not strictly tied to differences in breastfeeding. This implies that the variants of *FADS2* may have differential effects for more general measures of the early nutrition environment. With this idea in mind, we use birth weight, not breastfeeding, as our primary measure for the early nutrition environment.[4]

The interaction between breastfeeding and *FADS2* is due to long-chained polyunsaturated fats (LCPUFAs). The main LCPUFAs under consideration are docosahexaenoic acid (DHA) and arachidonic acid (AA), which are associated with early cognitive development (McCann and Ames 2005). Breast milk contains high levels of these LCPUFAs, and the *FADS2* gene is associated with extracting LCPUFAs from the diet. This gives rise to the GxE found in Caspi et al. As stated earlier, we use birth weight as our environment in place of breastfeeding. Given that *FADS2* is associated with extracting LCPUFAs from the diet, birth weight should also have an interactive effect with FADS2 variants. Birth weight is simply a proxy for the early nutritional environment, and those infants who were exposed to better diets, including access to LCPUFAs, should have the same interactive effects with *FADS2* as breastfeeding. Furthermore, a number of studies have noted a correlation between LCPUFAs and birth weight (Leaf et al. 1992, Muthayya et al. 2009, Ramakrishnan et al. 2010). While the use of birth weight instead of breastfeeding does not allow us to truly replicate the findings of Caspi et al., we instead examine a broader and important measure of early nutritional status that, based on the findings above, could plausibly be hypothesized to interact with *FADS2* in a manner similar to that proposed by Caspi et al.

[4] The WLS data does not contain information on breastfeeding. As will be described in Sec. 2.1, the WLS data are primarily on Wisconsin high school seniors in 1957. Given that the use of infant formulas became widespread in the 1950s, it is safe to assume that the vast majority of our sample was breastfed.

2. Data and Empirical Strategy

2.1 Data

The Wisconsin Longitudinal Study (WLS) is a random sample of one-third of the 1957 high school graduates from Wisconsin. Information on graduates for a large number of individual and family characteristics were collected in 1957, 1964, 1975, 1992, and 2003, while information on selected siblings of the graduates began in 1977 with additional data collected in 1993 and 2004. For graduates, data were collected for 10,317 individuals, while for selected siblings, data were collected for 6,619 individuals.[5]

Our outcome variable of interest is IQ, which is mapped from a Henmon-Nelson test score from both the graduate's and sibling's junior year in high school. The independent variables of interest are birth weight, the allele frequency for two SNPs in the *FADS2* gene—rs174575 and rs1535, and the interaction (GxE) between these two variables. Following Caspi, important covariates include gender, race, age, mother's education, father's education, and a family-level score for socio-economic status in 1957. Summary statistics for our outcome variable and regressors of interest are given in Table 1.

In order to measure the two genetic variants given by Caspi et al., we simply use an indicator for possessing two copies of the specified variant from Caspi et al. For SNP rs174575, we use an indicator for having two copies of the "C" variant, where having two copies of any one variant is referred to as homozygous.[6] The same indicator is used for the "A" variant of SNP rs1353. To correct for any potential benefits of containing just one copy of each variant, we also create an additive measure for each SNP that counts the number of variants an individual possesses—e.g., 0 for no copies of the "C" variant, 1 for heterozygotes, and 2 for homozygotes of the "C" variant.[7]

While we have IQ data for roughly 17,000 individuals, the number of observations are reduced through the collection of additional, necessary variables. First among these is birth weight. Birth weight is self-reported in the 2003 wave for graduates and the 2004 wave for the selected siblings. For

[5] The collected data for siblings differ in regards to the collection of other variables and other waves. For the 1977 wave, data were collected for 2,133 selected siblings. For the 1993 wave, mail questionnaires were completed for 4,036 siblings, while 4,804 completed at least part of the phone questionnaire. For the 2004 wave, 4,271 siblings completed at least part of a phone interview. The increase associated with IQ is due to the available data being present in high school roles, which is not directly sampled by the WLS.

[6] All individuals receive two copies of DNA: one from the mother and one from the father.

[7] Caspi et al. also find a favorable advantage for heterozygotes. We are unable, however, to replicate these results with the use of an indicator for heterozygous individuals.

graduates, 3,472 individuals were missing from the 2003 wave and 2,322 did not contain data for birth weight, giving birth weight data for 4,523 graduates. For siblings, 3,297 were missing from the 2004 wave and 1,226 are classified as inappropriate, giving data on birth weight for 2,096 siblings. Additionally, 113 graduates and 40 siblings are missing data for at least one covariate. A further sample reduction occurred from availability of biomarker data, which was collected in 2007 for graduates and 2008 for siblings. Complete biomarker data for the two *FADS2* variants exist for 4,455 graduates and 2,442 siblings. Overlapping these data with birth weight and other covariates results in the loss 1,732 graduates and 1,212 siblings. Finally, to perform the sibling fixed effects analysis we reduce the sample to only sibling pairs that have complete information for both the graduate and the selected sibling. In other words, all individuals are dropped from the sample unless they contain all available data and have a sibling who contains all available data. This results in a further reduction of 2,975 individuals, giving a sibling sample composed of 978 individuals from 489 sibling pairs.[8] Summary statistics for the differing samples are given in Table 1.

The primary cause for our sample truncations is due to a reduction in the sample for the 2003 (2004) wave. At this time, graduates were in their mid-sixties, and it's plausible that surviving until the 2003 (2004) wave is correlated with our dependent variable, IQ. To correct for this, we construct attrition weights by first regressing an indicator for being in the sibling-pair sample on IQ. The inverse of this probability is then used as a weight in estimation, correcting for possible sample selection due to IQ.

One other cause for concern in using the WLS is the generalizability of the sample. While being equally composed of both men and women, sibling-pair sample is composed almost entirely of peoples of European descent.[9] Out of the 489 sibling none are black, and no other ethnicities are represented. For our purposes in reproducing the findings of Caspi et al., however, the narrow focus of our sample is not a problem, as the sample of Caspi is also solely composed of European derived ethnicities.

2.2 Empirical Strategy

In an effort to replicate the findings of Caspi, we use birth weight as an indicator of early childhood (including in utero) nutrition. This approach, however, does not allow for an identical comparison to the gene-environment interaction of Caspi. Before replicating the findings of Caspi et al., we perform

[8] All graduates are linked wit data for only one sibling.
[9] This is not due to the sample truncations, but rather, the demographic composition of Wisconsin in the late 1950s.

additional estimations to check for the main, not interactive, effects of birth weight and the two SNPs of the *FADS2* gene on IQ, as well as check for a potential correlation between our genetic measures and our environment, birth weight.

The remainder of our study design is twofold. Firstly, we will replicate the GxE findings of Caspi et al. while substituting the indicator for being breastfed with a continuous measure of birth weight. Secondly, we will explore the replicated findings while conditioning on a shared sibling environment and genome.[10] More formally, our study will consider the following estimating equation:

$$\mathrm{IQ}_{ij} = \beta_0 + \beta_1 SNP_{ij} + \beta_2 BW_{ij} + \beta_3 SNP_{ij} \times BW_{ij} + \boldsymbol{\beta_4'} \boldsymbol{X_{ij}} + \boldsymbol{\beta_5'} \boldsymbol{Z_j} + u_{ij} \quad (1)$$

where β_3 is our coefficient of interest and represents the effect of gene-environment interaction for i individuals in j sibling pairs. OLS estimation of Equation (1) should replicate the main findings of Caspi et al., and the use of sibling pairs will allow us to control for unobserved, sibling-shared omitted variables.

The potential for omitted variable bias can be seen in the composition of the error term, u_{ij}. In addition to a random component, the error term is composed of unseen sibling-shared variation, both genetic and environmental, as well as individual specific variation. This is shown by:

$$u_{ij} = s_j + g_{ij} + e_{ij} + \varepsilon_{ij} \quad (2)$$

where s_j represents the unobserved, sibling-shared genetic and environmental effects. The use of sibling fixed effects would correct for potential omitted variable bias due to the unobserved, sibling-shared effects. In other words, if $cov(SNP_{ij} \times BW_{ij}, u_{ij}) \neq 0$ due to the $cov(SNP_{ij} \times BW_{ij}, s_j)$, then the use of a sibling fixed effects model will eliminate this potential source of bias.[11] Correcting for this bias will give a more accurate effect of the GxE interaction resulting from variants of the *FADS2* gene and early childhood nutrition.

The structure of our tables will follow the following form. Column (1) performs OLS estimation of Equation (1) with the largest possible sample. Column (2) repeats the estimation of column (1) but uses the sibling pair sample. Column (3) weights the estimation of column (2) by the inverse of the probability of being in the sibling pair sample, where this probability is calculated by IQ. Finally, column (4) controls for sibling fixed effects. The primary comparison we wish to make is between columns (2)

[10] Siblings share 50% of the genome passed from parents.
[11] The possibility of unobserved, individual specific genomic or environmental bias remains, however.

and (4) of Table 4, where the estimation of column (2) is intended to replicate the main findings of Caspi et al. and the estimation of column (4) includes sibling fixed effects. All tables are broken into two panels with panel A using the rs174575 SNP and panel B using the rs1535 SNP. The next section discusses our findings.

3. Results

3.1 Preliminary Estimation

Table 2 begins by exploring the main effects of the CC genotype for SNP rs174575, the AA genotype for SNP rs1535, and birth weight. All estimations of Table 2 control for gender, race, age, mother's education, father's education, and socio-economic status of the family in 1957 with standard errors clustered at the family level.[12] Using as large as possible a sample in column (1) shows that each variant of *FADS2*—rs174575 and rs1353—is insignificantly associated with high school IQ scores, while birth weight has a positive and significant association, with roughly each standard deviation increase in birth weight being associated with a one point increase in IQ. Column (2) restricts the sample to sibling pairs, leading to no major change in the associations of column (1); however, the magnitude of the coefficient of standardized birth weight increases from 0.88 to 1.55 while remaining significant at the 1% level. This effect is consistent in both Panel A and Panel B, which estimate the effect of the rs174575 locus and the rs1353 locus, respectively. The estimation of column (3) weights the OLS estimation of column (2) by the inverse probability of being in the sibling-pair sample. This weighting causes no consequential change in the magnitude or significance of the coefficient of birth weight or each measure of the *FADS2* gene: the effect of birth weight remains positive and significant, while *FADS2* has no direct effect on IQ. Finally, in column (4) we control for sibling level fixed effects. The inclusion of sibling fixed effects does not cause a consequential change in the coefficient of either *FADS2* or birth weight. Each variant of *FADS2* has an insignificant association with IQ, while the effect of birth weight is consistent with previous findings.

The estimates of Table 2 support previous findings that birth weight is indeed a significant source of variation in later-life IQ (Black et al. 2007), while each locus of *FADS2* has no direct effect. Given the consistency of the coefficient of standardized birth weight, we have little reason to suspect

[12] When employing sibling fixed effects race, mother's education, father's education, and socio-economic status are omitted due to these being shared controls amongst siblings.

this to be a spurious relationship. From Caspi et al., breastfeeding is associated with 5-6 point increase in IQ. From the estimations of Table 2, this large of an effect on IQ would be associated with a 3-4 standard deviation difference in birth weight. The 5-6 point difference in IQ found between being breastfed and not should be taken with caution. Breastfeeding is a choice made by a mother, and this choice may be associated with other choices that influence later life IQ (Fletcher 2011).[13] In other words, the effect of breastfeeding may be biased by unobserved heterogeneity.[14] The early nutrition environment is may also have a larger impact on IQ in earlier years. The IQ measure of Caspi et al. is found by averaging IQ from 7-13 years of age, a period significantly earlier than that measured in the current work.

Table 3 reproduces the estimation strategy of Table 2 but replaces IQ with standardized birth weight. The purpose of Table 3 is to show that no significant association exists between our gene and environment. A significant association between the gene and the environment would cause the gene-environment interaction to be suspect. It may not be the interaction, but rather the gene that is causing both selection into a particular environment and the outcome of interest. This concern is eliminated by the estimates of Table 3, which finds an insignificant relationship between each SNP of *FADS2* and birth weight, our proxy for the early nutrition environment.

3.2 Gene x Environment

Figure 1 shows the differential effect of birth weight from the variants of each *FADS2* SNP. Panel A of Figure 1 plots the effect of birth weight for those with two copies of C-allele of SNP rs174575 versus those without two copies of the C-allele. As is shown, homozygotes of the C-allele have a responsive, positive association between birth weight and IQ, while those without two copies of the C-allele exhibit no association between birth weight and IQ. The same is also true for A-allele homozygotes of SNP rs1535 in Panel B. This differential association with birth weight exhibits a text book example for the

[13] Birth weight is also influenced by choices of the mother during pregnancy. However, the use of sibling fixed effects should control for consistent choices across siblings. And given the consistency in the coefficient of birth weight, this is unlikely to be a source of bias.

[14] Caspi et al. do mention the possibility of bias estimation due to SES status and maternal education. Their analysis, however, is problematic. Firstly, their measure of SES status is imprecise, simply grouping individuals into one of three SES classifications. Secondly, Caspi et al. don't directly control for SES status in their estimations. Instead they argue that since variants of *FADS2* don't have a significant interaction with SES in explaining IQ that it is unlikely that SES status is accounting for the effects of breastfeeding. The variation in breastfeeding behavior within each SES class introduces measurement error into the estimation, leading to the insignificant interaction between *FADS2* and SES. To accurately measure the effect of breastfeeding, SES status should be included into the estimation.

differential susceptibility hypothesis (Belsky 2005), in which some individuals are sensitive to the environment—homozygotes of the C-allele for SNP rs174575 and homozygotes of the A-allele for SNP rs1535—while others are robust.

The gene-environment interaction is tested in Table 4. Table 4 repeats the estimations of Table 2 while including the GxE interaction into all regressions. The results of Table 4 are fairly stark. A significant GxE interaction is found in all estimations until the inclusion of sibling fixed effects. When sibling fixed effects are included into the estimation the coefficient of the GxE interaction is reduced by roughly half and becomes statistically insignificant. The large reduction in the GxE coefficient implies that unobserved heterogeneity shared between siblings is a source of bias for the previously estimated coefficients. A major candidate for this unobserved heterogeneity is other, unmeasured genetic variation, implying that the significant interaction between *FADS2* and early nutrition environments is the product of a gene-gene interaction, not the previously hypothesized gene-environment interaction.

As a check for the findings of Table 4, Table 5 re-estimates the findings of Table 4 with an additive measure for the sensitive variant of each *FADS2* SNP. The additive measure is intending to capture degrees of difference in the number of sensitive variants, and also controls for any beneficial effects of heterozygotes versus homozygotes for the robust alleles.[15] The results of Table 5 are consistent with those of Table 4: A significant GxE interaction exists in standard OLS estimation, but this effect is substantially reduced by the inclusion of sibling fixed effects, resulting in an insignificant coefficient for the GxE interaction.

The estimates of tables 4 and 5 suggest that the previous findings of Caspi et al., and similar results from papers that are unable to control for shared family factors, should be taken cautiously. For one, the estimated effect of breastfeeding may be biased upwards due to inadequately controlling for SES status, which is positively correlated with breastfeeding and IQ. Furthermore, the inability by Caspi et al. to control for unobserved genetic heterogeneity leaves open the possibility that the interaction between *FADS2* and breastfeeding is the spurious byproduct of an unseen gene-gene interaction. The use of sibling fixed effects allows us to partially control for unobserved genetic differences. Doing so, results in an insignificant effect for the previously robust interaction between *FADS2* and the early nutrition environment and calls into consideration the true nature of this relationship.

[15] As mentioned in Sec. 2.1 and footnote 5, Caspi et al. find a significant interaction between heterozygotes and breastfeeding.

4. Conclusion

This paper finds evidence that the interaction between the *FADS2* gene and early life nutrition may have no effect on later-life IQ. Rather, this previously found GxE association may be the product of unobserved, familial characteristics. In order to identify an interaction between a single environment and a single allele, other observed and unobserved differences in the genome and the environment must be accounted for. While not being able to control for all genetic and environmental differences, sibling fixed effects provides an important robustness check for the influence of unobserved, sibling-shared genes and environments. Along with Conley and Rauscher (2010), the current work calls into question many previously found GxE associations.

In summary, we argue that birth weight is a viable proxy for the early nutrition environment and could interact with *FADS2* in a similar manner to breast feeding. The GxE interaction between *FADS2* and birth weight is shown to have a positive and statistically significant effect on later-life IQ, a finding that echoes the interaction effect of *FADS2* and breastfeeding from Caspi et al. The statistical significance of this GxE interaction, however, dissipates with the inclusion of sibling fixed effects. The example has wider implications for the practice of investigating gene-environment interactions when the environmental exposure is not exogenous, and robust measures of the genome are not controlled in the analysis. More specifically, our findings also question previous results linking early nutrition and FADS2 genotype with IQ.

5. References

Belsky (2005). Differential susceptibility to rearing influences: An evolutionary hypothesis and some evidence. In B. Ellis & D. Bjorklund (Eds.), Origins of the social mind: Evolutionary Psychology and Child Development (pp. 139–163). New York: Guildford

Black et al. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. Quarterly Journal of Economics, 122(1), 409-439.

Castilho and Barros (2010). The history of infant nutrition. Journal de Pediatria, 86(3), 179-188.

Conley and Rauscher (2010). Genetic Interactions with prenatal social environment: Effects on academic and behavioral outcomes. NBER Working Paper 16026. Cambridge, MA: National Bureau of Economic Research (May).

Caspi et al. (2007). Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. PNAS, 104 (47), 18860-18865.

Fletcher (2011). Long term effects of health investments and parental favoritism: The case of breastfeeding. Health Economics, 20(11), 1349–1361

Leaf et al. (1992). Long chain polyunsaturated fatty acids and fetal growth. Early Human Development, 30(3), 183-191.

Muthayya et al. (2009). The effect of fish and ω-3 LCPUFA intake on low birth weight in Indian pregnant women. European Journal of Clinical Nutrition, 63(3), 340-346.

McCann and Ames (2005). Is docosahexaenoic acid, an n3 long-chain polyunsaturated fattyacid, required for development of normal brain function? An overview of evidence from cognitive and behavioral tests in humans and animals. The American Journal of Clinical Nutrition, 82(2), 281-295.

Ramakrishnan et al. (2010). Effects of docosahexaenoic acid supplementation during pregnancy on gestational age and size at birth: Randomized, double-blind, placebo-controlled trial in Mexico. Food and Nutrition Bulletin, 31(2), S108-S116.

Stein et al. (1975). Famine and human development: The Dutch hunger winter of 1944-1945. New York, NY, US: Oxford University Press.

6. Tables and Figures

Table 1: Summary Statistics

	N	Mean	Std. Dev.	Min	Max
Siblings with DNA and Birth Weight (Base Sample)					
IQ in High School	978	104.32	14.81	61	145
Birth Weight (in grams)	978	3364.58	610.49	992.23	7455.92
SNP rs174575 (homozygote for "C" variant)	978	0.61	0.49	0	1
SNP rs1535 (homozygote for "A" variant)	978	0.48	0.50	0	1
Individuals with DNA and Birth Weight					
IQ in High School	3953	103.72	14.82	61	145
Birth Weight (in grams)	3953	3379.96	629.45	538.64	7767.76
SNP rs174575 (homozygote for "C" variant)	3953	0.59	0.49	0	1
SNP rs1535 (homozygote for "A" variant)	3953	0.46	0.50	0	1
Individuals with Birth Weight					
IQ in High School	6466	102.60	14.84	61	145
Birth Weight (in grams)	6466	3374.42	638.03	538.64	8051.26
All Individuals					
IQ in High School	16936	101.00	15.32	61	145

Table 2. Main Effects of Birth Weight and *FADS2* on IQ

Dependent Variable: High School IQ				
Sample	All	Siblings		
	(1)	(2)	(3)	(4)
	Panel A: FADS2 SNP = rs174575			
Homozygous for "C" Variant	-0.2108	-0.3564	-0.8504	1.4707
	(0.4622)	(0.9597)	(0.9773)	(1.4764)
Standardized Birth Weight	0.8861***	1.5587***	1.5857***	1.4163**
	(0.2432)	(0.4916)	(0.5076)	(0.6622)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1004	0.1194	0.1269	0.7042
	Panel B: FADS2 SNP = rs1535			
Homozygous for "A" Variant	0.1740	-0.2973	-0.7633	0.9629
	(0.4539)	(0.9300)	(0.9360)	(1.3253)
Standardized Birth Weight	0.8854***	1.5548***	1.5765***	1.4409**
	(0.2432)	(0.4917)	(0.5068)	(0.6603)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1004	0.1193	0.1268	0.7038

Notes: (i) "Homozygous for 'C' Variant" is an indicator for the CC genotype for SNP rs174575, whereas "Homozygous for 'A' Variant" is an indicator for the AA genotype for SNP rs1535. (ii) Demographic and family controls include race, sex, birth year (age), mother's education, father's education, and a score for family SES in 1957. (vi) Standard errors are clustered at the family level with *, **, and *** representing significance at the 10, 5, and 1% significance level, respectively.

Table 3. Effects of *FADS2* on Birth Weight

Dependent Variable: Standardized Birth Weight				
Sample	All	Siblings		
	(1)	(2)	(3)	(4)
	Panel A: FADS2 SNP = rs174575			
Homozygous for "C" Variant	-0.0014	0.0698	0.0352	0.1055
	(0.0314)	(0.0639)	(0.0663)	(0.0961)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.0293	0.0337	0.0298	0.6259
	Panel B: FADS2 SNP = rs1535			
Homozygous for "A" Variant	0.0163	0.0336	-0.0056	0.0243
	(0.0312)	(0.0639)	(0.0669)	(0.0906)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.0293	0.0328	0.0295	0.6252

Notes: (i) "Homozygous for 'C' Variant" is an indicator for the CC genotype for SNP rs174575, whereas "Homozygous for 'A' Variant" is an indicator for the AA genotype for SNP rs1535. (ii) Demographic and family controls include race, sex, birth year (age), mother's education, father's education, and a score for family SES in 1957. (vi) Standard errors are clustered at the family level with *, **, and *** representing significance at the 10, 5, and 1% significance level, respectively.

Table 4. Effect of Interaction between Birth Weight and *FADS2* on IQ

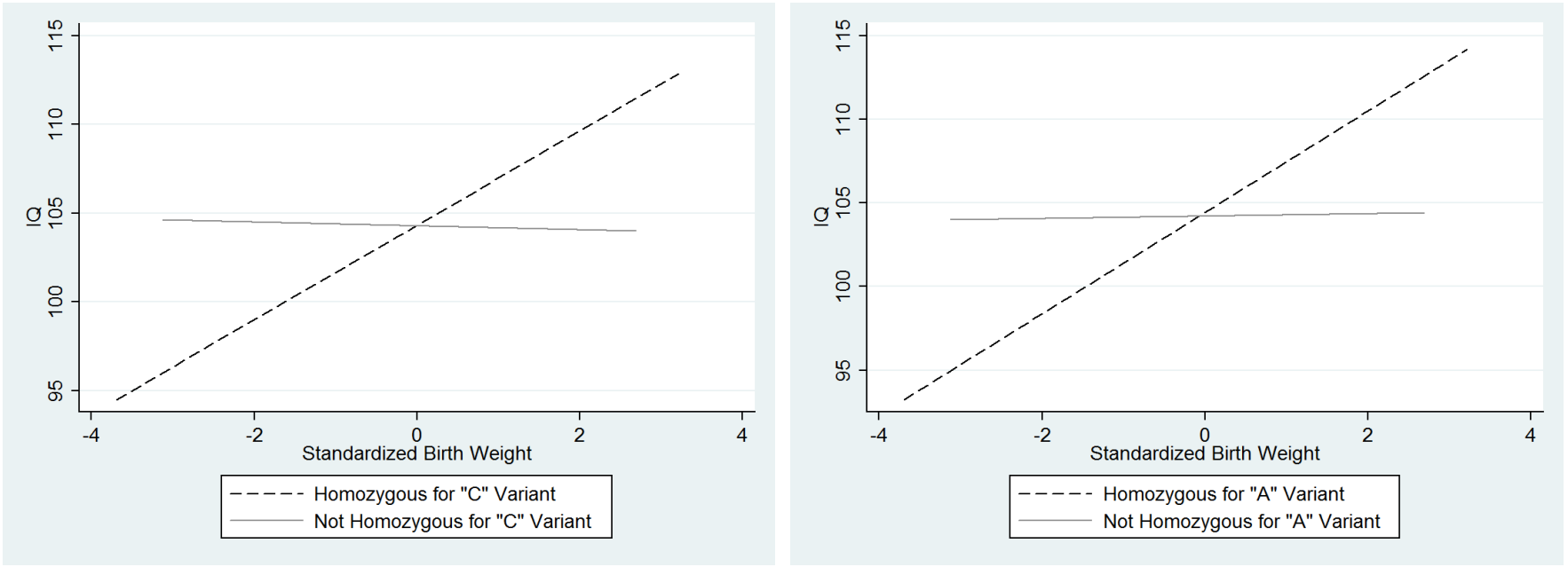
Dependent Variable: High School IQ				
Sample	All	Siblings		
	(1)	(2)	(3)	(4)
	Panel A: FADS2 SNP = rs174575			
Homozygous for "C" Variant	-0.2214	-0.2853	-0.7592	1.4679
	(0.4618)	(0.9552)	(0.9742)	(1.4717)
Standardized Birth Weight	0.3831	-0.0124	0.0982	0.6109
	(0.3751)	(0.7639)	(0.7973)	(0.9728)
GxE	0.8387*	2.4533**	2.3128**	1.2241
	(0.4823)	(0.9625)	(0.9915)	(1.2269)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1011	0.1250	0.1321	0.7048
	Panel B: FADS2 SNP = rs1535			
Homozygous for "A" Variant	0.1561	-0.2666	-0.6914	0.9197
	(0.4532)	(0.9252)	(0.9323)	(1.3192)
Standardized Birth Weight	0.2788	0.2332	0.3035	0.5829
	(0.3331)	(0.6825)	(0.7156)	(0.8662)
GxE	1.2538***	2.4415***	2.3299**	1.4814
	(0.4738)	(0.9383)	(0.9621)	(1.2029)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1021	0.1254	0.1324	0.7048

Notes: (i) "Homozygous for 'C' Variant" is an indicator for the CC genotype for SNP rs174575, whereas "Homozygous for 'A' Variant" is an indicator for the AA genotype for SNP rs1535. (ii) Demographic and family controls include race, sex, birth year (age), mother's education, father's education, and a score for family SES in 1957. (vi) Standard errors are clustered at the family level with *, **, and *** representing significance at the 10, 5, and 1% significance level, respectively.

Table 5. Effect of Interaction between Birth Weight and Additive *FADS2* on IQ

Dependent Variable: High School IQ				
Sample	All	Siblings		
	(1)	(2)	(3)	(4)
	Panel A: FADS2 SNP = rs174575			
Number of "C" Variants	-0.5010	-0.2705	-0.5685	1.1630
(Additive Measure)	(0.3802)	(0.8340)	(0.8548)	(1.2852)
Standardized Birth Weight	-0.0578	-1.6574	-1.5643	0.2585
	(0.6214)	(1.3046)	(1.3729)	(1.6093)
GxE	0.6118	2.0099**	1.9687**	0.7172
	(0.3752)	(0.7797)	(0.8143)	(0.9652)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1013	0.1246	0.1320	0.7044
	Panel B: FADS2 SNP = rs1535			
Number of "A" Variants	0.0856	0.2523	0.0127	0.9396
(Additive Measure)	(0.3461)	(0.7295)	(0.7578)	(1.1353)
Standardized Birth Weight	-0.0711	-0.8067	-0.6605	-0.2552
	(0.5256)	(1.1072)	(1.1398)	(1.2662)
GxE	0.6938**	1.6171**	1.5395**	1.1370
	(0.3420)	(0.7072)	(0.7214)	(0.8389)
Controls				
Demographic and Family SES	Y	Y	Y	Y
Sibling Fixed Effects	N	N	N	Y
Estimation				
Weighting by Prob. of Being in Sib Sample	N	N	Y	N
N	3953	978	978	978
R Sqr.	0.1013	0.1237	0.1304	0.7049

Notes: (i) "Number of 'C' Variants" is a count for the number of C alleles an individual has at SNP rs174575, whereas "Number of 'A' Variants" is a count for the number of A alleles an individual has for SNP rs1535. (ii) Demographic and family controls include race, sex, birth year (age), mother's education, father's education, and a score for family SES in 1957. (vi) Standard errors are clustered at the family level with *, **, and *** representing significance at the 10, 5, and 1% significance level, respectively.



(a) rs174575 **(b)** rs1535

Figure 1
Differential Effects of Birth Weight for *FADS2* Variants