

High-school genetic diversity and later-life student outcomes: Micro-level evidence from the Wisconsin Longitudinal Study.<sup>1</sup>

C. Justin Cook  
University of California-  
Merced [jcook9@ucmerced.edu](mailto:jcook9@ucmerced.edu)

Jason M. Fletcher  
University of Wisconsin-  
Madison  
[jfletcher@lafollette.wisc.edu](mailto:jfletcher@lafollette.wisc.edu)

**Abstract** A novel hypothesis posits that levels of genetic diversity in a population may partially explain variation in the development and success of countries. Our paper extends evidence on this question by subjecting the hypothesis to an alternative context that eliminates many competing hypotheses. We do this by aggregating representative individual-level data for high schools from a single US state (Wisconsin) in 1957, when the population was composed nearly entirely of individuals of European ancestry. Using this sample of high school aggregations, we too find a strong association between school-level genetic diversity and a range of student socioeconomic outcomes. Our use of survey data also allows for a greater exploration into the potential mechanisms of genetic diversity. In doing so, we find positive associations between genetic diversity and indexes for openness to experience and extraversion, two personality traits tied to creativity and divergent thinking.

## **Keywords**

## **JEL classification**

---

<sup>1</sup> This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin–Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775, AG-21079, AG-033285, and AG-041868), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin–Madison. Since 1992, data have been collected by the University of Wisconsin Survey Center. A public use file of data from the WLS is available from the Wisconsin Longitudinal Study, University of Wisconsin–Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at <http://www.ssc.wisc.edu/wlsresearch/data/>. The opinions expressed herein are those of the authors.

## 1 Introduction

Novel research at the intersection of the macroeconomics, economic growth, and population genetics literatures posits that both contemporary and historic differences in country-level economic development could be tied to population-level genetic processes. Specifically, the hypothesis is that levels of genetic diversity, at the population- (i.e., country-) level, shape long-term patterns of economic development. The hypothesized mechanisms of this relationship are twofold: genetic diversity increases the ability of a society to innovate, but genetic variation also increases ethnic fractionalization and decreases trust within a population.<sup>2</sup> Taken together, these two effects create a nonlinear (hump shaped) effect of genetic diversity in determining economic success—a hypothesis supported in a theoretical framework and empirically in an initial and thorough analysis by Ashraf and Galor (2013a; hereafter AG) and further confirmed and extended by ongoing research (Ager and Brueckner 2016; Arbalti, Ashraf, and Galor 2015; Ashraf and Galor 2013b; Ashraf, Galor, and Kemp 2014 and 2015; and Depetris-Chauvin and Özak 2016).

In the current paper, we propose a robustness test of the relationship between genetic diversity and economic development indicators through leveraging a unique setting that limits several of the remaining empirical challenges to the results of AG. These challenges include (i) accounting for the potential of unmeasured country-level factors related to genetic diversity and economic success; (ii) accounting for the potential influence of population stratification—the conflation of impacts of genetics and race/ethnicity/ancestral origin on outcomes of interest, which could result in cryptic associations between genetic diversity and ancestry that confound the analysis; and (iii) clarifying and further exploring the proposed mechanisms, particularly increases in innovation from genetic diversity, which lack previous human evidence and are difficult to measure.

To accomplish our goals, we leverage a novel dataset from a single country (thus eliminating across-country confounders) composed entirely of European ancestry (thus limiting issues of population stratification) to assess the robustness of AG’s key findings but in an entirely different context. Therefore, we consider our analysis to be a unique test of their

---

<sup>2</sup> The more ultimate rationale for these two mechanisms is tied to the survival advantages of increased genetic diversity weighed against the resulting weakening of kin networks.

original hypothesis, while ruling out several important alternative hypotheses consistent with their results. Unlike AG, however, our study lacks a clean identification strategy, which limits our ability to estimate a causal effect of genetic diversity. Nevertheless, our baseline set of controls and subsequent robustness exercises should mitigate omitted variable bias.

Specifically, our primary estimations test the relationship between measured school-level genetic diversity and the school's students' socioeconomic status (SES) outcomes—education, occupational prestige, and income. In contrast to AG's quadratic effect of genetic diversity, our hypothesis is that genetic diversity will have a positive linear association with a student's SES. The reason is that our sample, composed entirely of people with European ancestry in 1950s Wisconsin, is unlikely to exhibit the negative channels of genetic diversity—mistrust and ethnic conflict— theorized by AG.<sup>3</sup> This argument is supported in the recent work of Ager and Brueckner (2016), who find linear increases in economic growth from genetic diversity of European immigrants to US counties.

We also explore AG's theorized positive mechanism of genetic diversity. To do so, we take advantage of indexes for the Big Five personality traits—openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism—to examine whether openness to experience and extraversion, which have associations with creativity and divergent thinking (Hirsh et al. 2009; Kaufman et al. 2016), may be affected by genetic diversity in a way consistent with the hypothesis of AG and the diversity literature more generally.<sup>4</sup> High school genetic diversity intends to measure the general social environment and interactions faced by the student and represents a life stage in which personality and beliefs are flexible (Costa & McCrae, 1994; McCrae & Costa, 1999; McCrae et al. 2000).<sup>5</sup> Our hypothesis is that these early-life interactions help to shape personality and other beliefs in a way similar to other forms of diversity. In exploring mechanisms, we also provide evidence consistent with a prior hypothesis of genetic diversity's role in task specialization (Depetris-Chauvin and Özak 2016).

---

<sup>3</sup> Roughly 75% of the population belongs to one of five ethnicities—British, Irish, Norwegian, German, and Polish— with 47% being derived solely from Germany.

<sup>4</sup> As reviewed by Pickering et al. (2016), neuroticism has routinely been shown to have a negative association with creativity. This relationship is also weakly supported in our analysis as shown in Panel C of Table 6.

<sup>5</sup> Additional findings suggest that personality may change over the life course (Srivastava et al. 2004); however, the personality traits that show change—conscientiousness, agreeableness, and neuroticism—are those that are not of primary interest for the current work.

Indeed, genetic diversity is shown to have a persistently positive, statistically significant association with both our set of SES outcomes and our personality measures of interest. To summarize our results, a one standard deviation increase in genetic diversity is associated with an increase of 1.75 months of schooling, a 0.03 standard deviation increase in occupational prestige, a 2% to 6% increase in family income, a 0.1 increase in the index for openness to experience and extraversion, and a 0.03 increase in task diversity. All estimated effects are statistically significant at conventional levels.

## **2 Background: The benefits of diversity**

Following the theory and evidence provided by AG, the beneficial effects of genetic diversity are similar to those of other identity-based diversity measures—e.g., sex, race, ethnicity, etc. (Alesina and La Ferrara 2005; Kemeny 2017). In short, diversity introduces novel perspectives that increases both the set and quality of solutions in problem solving (Hong and Page 2001, 2004).

Our motivating theory is derived from Hong and Page (2001), who show that the increased number of perspectives from a *group* of diverse individuals is more likely to lead to novel solutions to problem solving than a homogenous group. In adapting this theory to the current work, we argue that exposure to diversity in early-life increases the set of perspectives held by an individual, and that this increase in perspective operates in a similar manner at the individual level as Hong and Page's theory at the group level. Furthermore, this increase in perspective is likely to be measured by personality changes, which explains our focus on openness to experience and extraversion and the shared variation between the two personality traits.

Empirically, the benefits of diversity are explored in a recent study by Freeman and Huang (2015), who show that research articles from an ethnically diverse set of authors are more highly cited than those from ethnically homogeneous authors. This finding is extended into the private sector. Firms with more diverse workplaces are shown to have a greater quantity and diversity of patent applications (Parrotta et al. 2014), and Lazear (1999) proposes and finds evidence that the costs of the globalization of firms—i.e., cultural differences, etc.—are offset by beneficial cross-cultural complementarities.

The ideas of the current work are also echoed in the immigration and urban economics literature that routinely show beneficial effects from immigration, which is representative of an influx of ethnically diverse groups that is likely correlated with increased genetic diversity. While addressing obvious concerns of endogeneity, increased immigration (primarily of groups ethnically distinct from the native population) is shown to increase innovation, total factor productivity, wages, and the rental price of housing (Alesina et al. 2016; Hunt and Gauthier-Loiselle 2010; Ottaviano and Peri 2006; Peri 2012). Furthermore, Peri and Sparber (2009) find evidence that immigration spurs task diversity; this is similar to the hypothesis posed by Özak and Depetris-Chauvin (2017) that we test in Table 7.

In contrast to identity-based diversity measures, genetic diversity more accurately measures differences within and across groups (Arbatli, Ashraf, and Galor 2015; Ashraf and Galor 2013b, 2017). For example, traditional ethnic fractionalization measures do not account for the degree of difference between ethnic groups: ethnic diversity from a country (or other aggregation) with a 50/50 composition of people ancestral to the United Kingdom and Germany would be identical to a 50/50 composition of people ancestral to the United Kingdom and China. Genetic diversity accounts for the degree of difference of all individuals within the group. It does so by calculating root differences among individuals through measured differences in individual genetic markers.<sup>6</sup> Furthermore, ethnic-, or ancestry-, based measures of diversity may be biased by errors in the self-reporting of ethnicity. This is especially likely in countries with long-running histories of migration and intermingling. An example of potential bias is the “one-drop rule” in the United States, a racist construct that identified as African American anyone with any sub-Saharan African ancestry.

### **3 Data and empirical methodology**

#### **3.1 Data and limitations**

We use a unique sample, clustered at the level of secondary school, that contains data on a wide array of variables throughout the life course. The Wisconsin Longitudinal Study (hereafter WLS) is a prospective longitudinal data set collected from a one-third random sample of 1957 high school graduates in Wisconsin. Importantly, for the 2004 wave of the

---

<sup>6</sup> This idea is discussed more fully in Arbatli, Ashraf, and Galor (2015).

WLS, genetic biomarker data were collected for both WLS graduates and selected siblings, which consist of roughly 95 single nucleotide polymorphisms (SNPs), the simplest type of genetic variation between people. We aggregate this individual-level survey data to the high school-level in order to create population-level measures for schools.<sup>7</sup> This allows us to treat high schools as distinct aggregate groups in which to measure the effects of genetic diversity. In so doing, allele frequencies, which show the genetic diversity of a population, for each high school are used to calculate a population-level measure of expected heterozygosity, our primary measure of genetic diversity, which is defined as the probability that two randomly selected individuals possess different alleles for a specified gene locus. Summary statistics of this variable and all other variables we used are found in Table 1, and detailed descriptions and sources of each variable are found in the Variable Appendix.

An important note is that the selection of SNPs within the WLS is neither random nor genome-wide. Genetic biomarkers were chosen from previous candidate gene and genome-wide association studies that link particular loci, or SNPs, to a range of phenotypic outcomes (e.g., IQ and cognitive performance, body mass index, risk for certain types of cancer, etc.). Previous measures of expected heterozygosity are measured by *neutral* genetic variants in order to measure population differences that have arisen due to populations being separated over time (AG; Cavali-Sforza 2005). An important assumption of the current paper is that expected heterozygosity that is calculated from the non-random genetic variants found within the WLS serves as a valid proxy to heterozygosity calculated from neutral variants.<sup>8</sup> As a check of this potential source of bias, robustness exercises in Tables 9 and 10 control for each individual SNP.

As a check of this assumption, we compute a county-level measure of genetic diversity from the non-neutral SNPs of the WLS. We then compare this county-level measure to Ager and Brueckner's county-level data that are constructed from ancestral shares and the country-level genetic diversity measure from AG. The correlation coefficient between the two measures

---

<sup>7</sup> To increase sample size, school-level gene frequencies are calculated from genetic data for both graduates and siblings. Appendix Section 6 replicates all estimations using a genetic diversity score that is calculated only from WLS graduates.

<sup>8</sup> One potential source of bias from violating this assumption may be tied to favorable interactions amongst the population of possessing a particular trait tied to cognition. However, complicated phenotypes, such as those of differential economic wellbeing, are not likely to be linked to singular genetic variations (Chabris et al. 2012 and Chabris et al. 2013).

is 0.33 and is significant at the 1% level. This statistical relationship provides support for our assumption that genetic diversity from the non-neutral variants of the WLS serves as a valid proxy for a similar measure calculated from neutral genetic variants. Both county-level measures of genetic diversity are given in Figure 1. Going further, we will leverage this independent source of genetic diversity to semi-replicate our base findings in a sample of the WLS graduate's parents, who were of high school age for the 1920 measure of Ager and Brueckner's county genetic diversity.

Another major concern with our data is sample selection.<sup>9</sup> Genetic biomarker data were collected in the 2004 wave, when the graduates were approximately 65 years of age. Individuals possessing genetic variants that were favorable for economic or cognitive outcomes are likely to survive at a greater rate than those with alleles that provide no clear advantage.<sup>10</sup> For our purposes, however, this scenario likely leads to a reduction in heterozygosity as survivors will possess identical alleles for a particular locus, leading to bias in the opposite direction of the primary hypothesis of this paper. Additionally, our base set of controls includes a measure for individual IQ, accounting for this potential mechanism.

A secondary concern related to selection is that our variable of interest, high school heterozygosity, is positively correlated with other characteristics of the high school that may increase the average later-life income or wealth associated with a particular high school. Higher educated and higher earning parents self-select into neighborhoods and school districts (Sampson and Sharkey 2008). If this mechanism of selection is related to a particular genetic endowment shared among parents, and in turn offspring, the level of genetic diversity within a school is potentially accounting for other school-level factors that are associated with the economic outcomes of the school population. To partially account for this potential selection, our base set of controls include measures for county fixed effects, a set of controls for school district resources (e.g., mean teacher salary, education, and experience), parent economic well-being, parent education, and the WLS graduate's IQ, which intends to proxy more favorable early life environments and heritable benefits that

---

<sup>9</sup> This is also a potential issue in AG, who do not use population representative genetic data.

<sup>10</sup> In recent work, Domingue et al. (2017) show correcting for mortality selection does not alter genetic associations.

are unrelated to high school genetic diversity.<sup>11</sup>

### 3.2 Estimation strategy

While our measure of genetic diversity is at the school level, our base analysis is at the individual level with our primary estimating equation being of the following form:

$$y_{ij} = \beta_0 + \beta_1 \text{Genetic Diversity}_j + \beta_2' X_{ij} + \beta_3' Z_j + \gamma' I_c + \epsilon_{ij} \quad (1)$$

The primary analysis considers the effect of school-level genetic diversity on a range of economic outcomes for  $i$  individuals in school  $j$ . Our coefficient of interest is the measure of genetic diversity for  $j$  high schools. Individual- and family-level controls are denoted by  $X_{ij}$ , and school district and historic controls are denoted by  $Z_j$ .

The set of “Individual” controls includes measures for IQ, sex, and birth year. The set of “Family” controls includes an index of family socioeconomic status in 1957 and mother and father’s education. The set of “School” controls includes WLS graduate’s cohort’s size; an indicator for being in either Madison or Milwaukee (urban areas with more than 100,000 in population); and the school district means (c. 1954–1957) of teacher salary, teacher experience (total and district), teacher education, the number of school days, and classroom size (Olson and Ackerman 1998).

One way to account for the potential selection in our set of genes is to account for ancestral geo-climatic environments that may be associated with potentially omitted cultural traits that are correlated with both our outcomes of interest and our measure of genetic diversity. To create these variables in our data set, we first match country-level historic geo-climatic characteristics to the WLS graduate’s reported father’s ancestral country. We then average these variables at the high school level, creating our set of “Historic” controls. These include absolute latitude, fraction of arable land, mean temperature, mean precipitation, mean elevation, roughness index, mean distance to the coast or navigable river, and the fraction of land within 100 kilometers of the coast or navigable river. Finally, to further mitigate selection effects and to account for hard-to-measure spatial differences in culture and lifestyle across Wisconsin, our base analysis includes county fixed effects, which are denoted in the above

---

<sup>11</sup> As shown in Appendix Table A2, genetic diversity is unrelated to IQ for our base specification. Furthermore, Appendix Table A5 performs a propensity score matching exercise that attempts to account for potential selection into schools. As shown, the matched effects do not substantially differ from the OLS estimates.

estimating equation by  $I_C$ .

Our estimation strategy differs from that of AG in that we hypothesize a positive linear effect of genetic diversity instead of nonlinear hump-shaped effect.<sup>12</sup> The proposed linear association is due to our focus on the relatively homogenous population of Wisconsin. The population is composed nearly entirely from Northern European ancestry, so the negative channel of genetic diversity—mistrust and ethnic conflict—is likely to be substantially weakened in this setting.<sup>13</sup> This narrow range of genetic diversity is further reduced by our empirical strategy of estimating within-county effects.<sup>14</sup> Additionally, the population of Wisconsin in 1957 is likely to be assimilated into a common American-based culture, further limiting the possibility of ethnic-based conflict and mistrust.

Our analysis follows a standard form: column (1) simply regresses our outcome of interest on the measure for high school heterozygosity; column (2) includes county fixed effects; columns (3)–(6) piecemeal add individual, family, school, and ancestral-origin controls to the estimation of column (2); and column (7) includes all controls, comprising our baseline estimation. Within-county associations between genetic diversity and our socioeconomic and personality outcomes of interest are plotted in Figures 2–5. All standard errors are clustered at the county level.

## 4 Results

### 4.1 Socioeconomic effects of genetic diversity

The first outcome of interest we explore is years of schooling. In regards to this measure of education, WLS comprises substantial selection. This is due to the nature of the WLS, which again, is a random one-third sample of *high school graduates*, therefore the minimum level of education in our primary sample is high school graduates. This selection may lead to concerns about generalizability; however, we do not expect bias from this selection.<sup>15</sup> Our sample consists solely of high school graduates, so we are testing the effect of genetic diversity on

---

<sup>12</sup> As shown in Appendix Section 5, we observe no significant quadratic effect of genetic diversity.

<sup>13</sup> Roughly 75% of the population belongs to one of five ethnicities—British, Irish, Norwegian, German, and Polish, with 47% being derived solely from Germany.

<sup>14</sup> The within-county standard deviation of high school genetic diversity is 0.0138 compared to 0.0162 for the overall sample.

<sup>15</sup> Concerns of generalizability are addressed in part by the replication of Table 11.

years of schooling in excess of high school.

As shown in Table 2, high school genetic diversity has a statistically significant positive association with a WLS graduate's years of schooling. Going from the simple bivariate regression of column (1) to the inclusion of a large number of relevant controls in our baseline specification of column (6), the magnitude of the coefficient is slightly reduced; however, a positive, statistically significant effect is found throughout. From our baseline estimates of column (7), a one standard deviation increase in a high school's expected heterozygosity (0.01) results in an increase of roughly 1.75 months of schooling (~6% of a standard deviation). Going from the bottom 10% of genetically diverse schools to the top 10% (0.32 to 0.34) is associated with roughly a 3-month increase in schooling on average (~11% of a standard deviation).<sup>16</sup>

Following the findings of Table 2 and the previous literature on the benefits of diversity, we test in Table 3 whether individuals that were exposed to higher levels of genetic diversity in high school selected into more prestigious occupations. Occupations at the higher end of the prestige scale tend to be those that require higher years of schooling and provide higher earnings—e.g., architects, dentists and physicians, engineers, lawyers and judges, etc. Each panel of Table 3 considers a different measure of prestige for the WLS graduate's first job. Summary statistics for each measure are given in Table 1, and a fuller description and source of each variable is given in the Variable Appendix.

The pattern of Table 2 is seen again in the findings of Table 3: the point estimate reduces slightly with the inclusion of relevant controls, but genetic diversity has a consistent positive and statistically significant effect on each measure of job prestige.<sup>17</sup> For column (7), a one standard deviation increase in genetic diversity is associated with a 0.03 standard deviation increase in Duncan prestige, a similar 0.03 standard deviation increase in Seigel prestige, and a

---

<sup>16</sup> Following Ager and Brueckner (2017) and Nunn and Wantchekon (2011), we compare the coefficient from the restricted bivariate estimation of column (1) to our baseline unrestricted model in column (7) to examine the potential for omitted variable bias (Altonji, Elder, and Taber 2005). For years of schooling, the Altonji et al. ratio (hereafter AR) is 1.8, implying that the selection on unobservables would have to be roughly twice that as the selection on observables.

<sup>17</sup> For column (5) of Panels A and B, the coefficient of genetic diversity is not significant at conventional levels. We attribute this to reduction in confounding after adding school controls. The point estimate is similar across both panels for column (7), but the standard error is reduced from the addition of individual/family/historic controls that account for variation in the outcome.

0.05 standard deviation increase in the occupational education index. WLS graduates who went to high schools with more genetically diverse populations tended to seek out more cognitively challenging and more admired occupations.<sup>18</sup>

Building off Tables 2 and 3, Table 4 considers the effect of genetic diversity on the natural log of family income in two different waves of the WLS—1974 (respondent age ~ 35) and 1992 (respondent age ~ 53). For income in both waves, the coefficient of genetic diversity has a positive and significant effect (at conventional levels) for all specifications. For the baseline estimate of 1974 (column (7) of Panel A), increasing genetic diversity by one standard deviation is associated with roughly a 2% increase in income.<sup>19</sup>

A larger estimated coefficient is seen for the 1992 wave. This is likely due to cumulative advantages as incomes diverge over the life course. This is shown in the standard deviations of income for the two periods (Table 1). For the base estimation, increasing genetic diversity by a standard deviation is associated with a 6% increase in family income in the 1992 wave.<sup>20</sup> Though it is difficult to directly compare results across studies, Ager and Brueckner find a similar (10%) increase for a standard deviation increase in county-level genetic diversity.<sup>21</sup> The findings of Tables 2–4 taken together are highly supportive of the hypothesis posited by AG. Our estimations consistently find that WLS graduates that attended more genetically diverse schools had higher levels of education, more prestigious jobs, and larger earnings later in life.

#### 4.2 An examination of mechanisms: Personality and task diversity

Our use of a rich survey-level data set that is focused on a wide range of variables throughout the life course presents a unique opportunity to examine the possible mechanisms of genetic diversity in affecting differential economic outcomes. In particular, we are interested solely in the *beneficial* effect of diversity.<sup>22</sup> AG propose that the positive channel of genetic

---

<sup>18</sup> The AR is 0.42 for Duncan job prestige, 0.49 for Seigel job prestige, and 1.68 for the occupational education score. See footnote 15 for further discussion.

<sup>19</sup> The AR is 7.33 for the estimated relationship of genetic diversity with the natural log of family income in 1972. See footnote 15 for further discussion.

<sup>20</sup> The absolute value of the AR is 8.98 when comparing column (1) to column (7) in Panel B of Table 4. See footnote 15 for further discussion.

<sup>21</sup> The closest estimations to those of the current work are found in Table 2 (p.29), which regress contemporary county incomes on county-level genetic diversity for 1870 while controlling for income in 1870.

<sup>22</sup> A number of studies have focused on the negative effects of genetic diversity on income, paying particular

diversity is tied to a diversity in ideas that can increase technological innovation, raising the production possibilities frontier. In short, AG argue that more genetically diverse populations are likely to be more *creative* populations. This is similar to the broader diversity literature (covered in Section 2) that shows diversity is associated with novel solutions. As a test of this idea, we estimate the relationship between genetic diversity and personality traits tied to creativity: openness to experience and extraversion.

AG's empirical evidence for genetic diversity's role in increasing creativity is shown by a strong positive effect from a country's level of genetic diversity on its total scientific publications. Relating this finding to personality, the two biggest personality predictors of creativity among scientists are openness to experience and extraversion (Feist 1998; Hirsh et al. 2009; Kaufman et al. 2016). Further studies also find beneficial effects from both extraversion and openness to experience for both creativity and divergent thinking (Furnham and Bachtiar 2008, Furnham and Chamorro-Premuzic 2004, King et al. 1996). Tables 5 and 6 regress indices for the Big Five personality traits on our measure of high school genetic diversity.<sup>23</sup> Panel A tests our hypothesis of genetic diversity's role in influencing the personality trait of openness to experience, from which a generally positive and statistically significant effect of genetic diversity is observed. The inclusion of our baseline set of controls reduces the coefficient of genetic diversity, but a positive effect is still estimated that is statistically significant at conventional levels.<sup>24</sup> Panel B tests the association between the personality index of extraversion and our measure of high school genetic diversity. As with openness, we estimate a positive and statistically significant coefficient. Furthermore, the estimated relationship with extraversion is consistent across all specifications. Given the role of both openness to experience and extraversion in creativity and divergent thinking, Panel C regresses the shared variation (i.e., the first principal component) between these two personality measures on our measure of high school genetic diversity. Again, a statistically

---

attention to the formation of ethnic groups and resulting ethnic conflict (Alesina and La Ferrara 2005; Arbalti, Ashraf, and Galor 2015; Ashraf and Galor 2013b). To our knowledge, only one other study focuses on the positive aspects of diversity (Depetris- Chauvin and Özak 2016).

<sup>23</sup> Personality indices are recorded for the 1992 and 2004 wave of the WLS. We take the average of each index for individuals that responded to both waves, while including individuals that contain data for only one of the waves.

<sup>24</sup> The estimates of column (5) in Panel A suggest high school genetic diversity is partially confounded by other school level variables. This confounding is addressed through the addition of our school-level controls. As with Panels A and B of Table 3, the point estimate remains smaller for column (7), but the standard error is reduced from the addition of individual/family/historic controls that account for variation in the outcome.

significant association is seen for all specifications, providing evidence for individual increases in creativity and divergent thinking from greater exposure to more diverse individuals.<sup>25</sup>

The effect on remaining personality traits is examined in Table 6. In general, genetic diversity exhibits no consistent relationship with these other personality measures. Of note is the negative bivariate relationships with agreeableness and neuroticism. The negative effect on agreeableness may be symptomatic of the hypothesized negative channel of diversity, which is associated with reduced cooperation or being less agreeable. And neuroticism has been previously found to have a negative association with creativity, which we hypothesize to have a positive association with diversity (Pickering et al. 2016). These two associations do not retain statistical significance when accounting for county fixed effects and our baseline set of controls.

The estimates of Tables 5 and 6 support genetic diversity's role in creativity. The two personality traits that have been routinely related to creativity and divergent thinking have statistically significant positive association with genetic diversity, while indices for other personality traits that have no associations with creativity are unrelated to genetic diversity.

Another potential mechanism is given by recent work of Depetris-Chauvin and Özak (2016), who provide evidence that genetic diversity has a positive association with the number of individualized tasks performed within a society prior to European colonization. They hypothesized that variation provides comparative advantages to differential tasks within a society, leading to unique specialization by members of the society. We propose an additional test of this potential mechanism by using the occupation code of first job for the WLS graduates.

Using this occupation code, a high school measure of task diversity is created. Specifically, we are interested in the selection by WLS graduates into diverse fields or industries. Therefore, our primary measure of task diversity is the high school mean of within-industry occupational diversity. In other words, we are not so much interested in the fact that individuals within a high school obtain different occupations, rather our focus is on the self-selection of individuals into occupations that are more likely to interact with others. The use of within-industry occupational diversity intends to measure this effect; however, as a replication exercise

---

<sup>25</sup>For Table 5, the AR is 0.46 for Panel A, 3.16 for Panel B, and 0.83 for Panel C. See footnote 15 for further discussion.

we will later examine the simple occupational diversity for each high school (Table 12). Following the hypothesis of Depetris-Chauvin and Özak, a positive linear relationship is expected between our genetic diversity measure and our within-industry occupational diversity score. Table 7 presents evidence of this relationship.

Table 7 follows a similar form as previous tables, but because the job diversity score also is an aggregate measure, the analysis is conducted at the high school level instead of for individuals. As is shown, a positive statistically significant relationship is found between a high school's genetic diversity and selection into more diverse industries. Although reduced in magnitude, a statistically significant relationship holds for all previously mentioned specifications.<sup>26</sup> Again, our unique and independent sample produces results consistent with the prior findings from Depetris-Chauvin and Özak, further supporting the role of genetic diversity in shaping economic outcomes.

As a further test of the proposed channels of genetic diversity, Table 8 includes personality measures of interest (openness and extraversion) and within-industry occupational diversity into our baseline estimating equation for SES outcomes considered in Tables 2–4.<sup>27</sup> We do so in two ways. Panel A controls for personality and task diversity, and Panel B examines the estimated residual-SES outcome after accounting for the proposed channels. Most SES outcomes are measured at the same time as the personality and task diversity measures, suggesting problems of reverse causality, and endogeneity more broadly, are likely present. That said, the estimates of Table 8 are viewed as suggestive. Of note, we expect a smaller mediating effect for years of schooling due to the contemporary timing of the education decision and the treatment (high school genetic diversity exposure).

For Panel A of Table 8, the magnitude of the coefficient of genetic diversity is reduced in nearly all cases: 13% for years of schooling, 25% for Duncan job prestige, 23% for Siegel job prestige, 10% for occupational education, and 17% for family income in 1974.<sup>28</sup> For Panel B, the association between genetic diversity and residual outcomes is lessened in all cases and statistically indistinguishable from zero for Duncan and Siegel job prestige and 1974 family income.

---

<sup>26</sup> The AR for Table 7 is 0.61. See footnote 15 for further discussion.

<sup>27</sup> We thank an anonymous reviewer for suggesting this analysis and discussion.

<sup>28</sup> The coefficient of genetic diversity *increases* by 8% when controlling for the proposed channels of personality and task diversity.

### 4.3 Robustness

Tables 9 and 10 examine the robustness of our previous findings to a set of ethnic and genetic controls. Column (1) replicates our baseline findings (column (7) of previous tables). Columns (2)–(4) include a number of ethnic-/ancestry-based controls. These include an indicator for the WLS graduate’s father’s ancestry in column (2), an ethnic fractionalization score calculated from this ancestry in column (3), and both in column (4). Comparing the estimate of column (1) to those of columns (2)–(4), the magnitude of the point estimates of genetic diversity is not significantly altered for any outcome, suggesting our results are not being driven by the presence of one ethnicity or a high school’s specific ethnic makeup. Columns (5) and (6) include gene-specific controls. First, column (5) includes an additive score for a particular variant (i.e., 0, 1, or 2) for each SNP used to calculate our high school measure of genetic diversity.<sup>29</sup> In general, the inclusion of all individual-specific genetic variants does not qualitatively change our estimated coefficients. In place of the gene-specific controls of column (5), column (6) controls for the shared variation across all SNPs. Controlling for the first principal component of each genetic variant does not substantially alter the coefficient of genetic diversity.

Columns (7) and (8) control jointly for the additional ethnic and genetic variables. Both columns include ethnic indicators and high school ethnic fractionalization, whereas column (7) includes additive markers for each SNP and column (8) includes the first principal component of these SNPs. As with the previous columns, the inclusion of these additional controls does not substantially alter the previously estimated coefficients of genetic diversity. We have no strong reason to believe that our results are driven by either a particular ethnicity or a particular genetic variant.

Table 10 replicates the analysis of Table 9 for the high school association between within-industry occupational diversity and genetic diversity. In addition to ethnic and genetic controls, the analysis of Table 10 also controls for industry diversity—separately in column (3) and jointly in column (4). As with Table 9, the inclusion of ethnic, genetic, and broader industry controls into our baseline estimating equation does not significantly alter the

---

<sup>29</sup> Each SNP is available for roughly 4,500 WLS graduates. To keep sample sizes maximized, we replace missing values with the mean. A dummy variable for those with missing values is also included in the estimation of columns (5) and (7).

coefficient of genetic diversity, implying the association between task and genetic diversity is not the byproduct of particular ethnicity, gene frequency, or industry.

#### 4.4 Replication

In an extension of our main findings, we explore the effects of genetic diversity in a similar but distinct sample: the WLS graduate's parents. In so doing, we use the county-level measure of genetic diversity for 1920 from Ager and Brueckner (2016). On average, WLS graduate's fathers were born in 1907 and mothers were born in 1911, so genetic diversity in 1920 is potentially representative of the environment faced by both parents in early life. The use of parents also generalizes our base findings, since the parents were not selected as being in a graduating high school class.<sup>30</sup>

Table 11 provides estimations that are similar to our baseline findings for the WLS graduates. In place of schools, the unit of observation for genetic diversity is the county and outcomes are given for each WLS graduate's family. Parental outcomes are denoted by column and include an index of family socioeconomic status in 1957, father's (Duncan) occupational prestige, father's education, mother's education, and a measure of county-level occupational diversity.

A concern is that the WLS parents do not reside in the same county as the one where their children eventually graduate high school, in 1957. Given that the focus of the WLS is on the high school graduates and not their parents, no data exist that document how long a particular family has lived in the 1957 county. To correct for measurement error associated with moving, we use the complete sample of the 1940 census to measure the fraction of a county's 1940 population that lived in a different county in the previous 5 years (Ruggles et al. 2015).<sup>31</sup> Therefore, we control for this measure in all specifications.<sup>32</sup> Other county controls attempt to account for 1920 characteristics that may be associated with either the genetic

---

<sup>30</sup> It is possible that those parents with high school graduates differ from those without high school graduates; however, when comparing the WLS parents to people of a similar age (within one standard deviation of mean year of birth) in the 1960 5% census sample we see very similar average years of schooling: 9.78 for WLS fathers and 9.49 for men in the 1960 census, and 10.47 for WLS mothers and 10.30 for women in the 1960 census.

<sup>31</sup> The 1940 Census is the most recent census with full coverage. Ideally, we'd use a census that is closer to the 1957 graduation date; however, 1% and 5% samples for the years 1950 and 1960 do not contain spatial data for areas with under 100,000 in population. This would exclude a majority of the rural counties that are central to our study.

<sup>32</sup> Omitting this control does not substantially alter the coefficient of interest.

diversity measure of parent outcomes. These include population density and the fraction born in Wisconsin, the United States (excluding Wisconsin), and a foreign country.<sup>33</sup>

The county analysis shows a positive statistically significant effect of genetic diversity for most parental outcomes; mother's education being the only outcome insignificantly associated with genetic diversity.<sup>34</sup> In comparing the estimates of Table 11 to those of the WLS graduates, a respective one standard deviation increase in genetic diversity is associated with roughly 0.32 of a standard deviation increase in father's job prestige and a 0.07 (bivariate; 0.03 baseline) standard deviation increase in the WLS graduate's job prestige.<sup>35</sup> For years of schooling, a standard deviation increase in a county's genetic diversity is associated with roughly an 8-month increase in father's years of schooling and 1 month for mothers (insignificant for mothers). This is in line with the estimates for WLS graduates in Table 2, which show that a standard deviation increase in high school diversity is associated with a 2.75-month increase in schooling for the bivariate estimation and a 1.75 month increase for our baseline estimation. And finally, for occupational diversity a respective one standard deviation increase in genetic diversity leads to roughly identical increases for fathers and WLS graduates (~0.06 vs. 0.11 [bivariate] and 0.04 [baseline]).<sup>36</sup>

Table 12 repeats the analysis of Table 7, considering an alternative measure of occupational diversity. Specifically, the outcome variable in Table 12 does not nest occupational diversity within industry, allowing for a measure of diversity independent of industry. From Table 12, a statistically significant and positive association is again observed between this alternative measure of task diversity and our high school measure of genetic diversity. As with the within-industry measure, the coefficient of interest is reduced in magnitude with the inclusion of our baseline controls but remains significant at the 1% level.

---

<sup>33</sup> County-level foreign-born country fractions are included only for those countries exceeding 10,000 migrants in the state of Wisconsin. These are Germany, Poland, Norway, Sweden, Russia (USSR), Canada, Czechoslovakia, Austria, England, Italy, and Hungary.

<sup>34</sup> When using spatially adjusted standard errors (50km), the association with father's within-industry occupational diversity is significant at the 1% level.

<sup>35</sup> We have no direct comparison between generations for the WLS's SES index. We report estimated effects for the bivariate and baseline estimations for WLS graduates to capture a range of outcomes that correspond to the parental estimates that have few controls.

<sup>36</sup> In creating comparisons, we use the coefficients from the simple within-county estimations (column 2 in specified table) for WLS graduates. The standard deviation for the individual sample is used for individual estimates—i.e., education and job prestige, and the standard deviation for the aggregate sample is used for occupational diversity.

## 5 Discussion

We show that a school-level measure of genetic diversity has a persistent positive and statistically significant association with a range of student-level socioeconomic outcomes. Students that attended more genetically diverse schools have higher levels of education (Table 2), more prestigious jobs (Table 3), and higher earnings later in life (Table 4). In exploring the mechanisms driving this association, we hypothesize and provide evidence that exposure to more diverse individuals shapes personality and role preferences during a formative time of the life cycle. Students who attend more diverse schools report higher later-life levels of openness to experience and extraversion (Table 5), two traits routinely shown to be associated with divergent thinking; these students are also more likely to select into industries with a more varied set of occupations (Table 7), supporting Depetris-Chauvin and Özak’s hypothesis. The association of genetic diversity with both SES outcomes and the proposed channels is robust to the inclusion of controls for ethnicity/ancestry and individual genetic differences (Table 9). Furthermore, using a different sample composed of the parents of the WLS cohort, we show a similar positive association between county-level diversity and parent-level SES outcomes (Table 11).

Our analysis provides a unique test of AG’s hypothesis by considering micro-oriented aggregations that avoid potential confounders by country or ethnic level differences. The use of schools in place of countries as the relevant “population” allows us to avoid potential confounding factors of institutional quality, human capital, and cultural differences that may be tied to both the level of genetic diversity and the economic outcomes across countries. The elimination of these potential confounders allows for more accurate measurement and interpretation of the effect proposed by AG. Given the level of separation between the country-level results reported in the literature and the findings of the current work, the statistically significant association between genetic diversity and socioeconomic outcomes does indeed appear to be non-spurious.

Ashraf and Galor do not ascribe their findings strictly to genomic differences; rather, they argue that the observed genetic variation may be accounting for unobserved cultural variation. We too cannot rule out that our measure of genetic diversity is simply accounting for cultural differences that are strongly tied to the measured genetic diversity score. However, our

use of relatively homogeneous (i.e., European ancestry), small, and local population aggregations as well as the inclusion of ethnic-specific controls does help reduce the overlap between the genetic and cultural channels, strengthening the argument of genetic diversity's role in influencing economic outcomes.<sup>37</sup>

The precise mechanism for the relationship between genetic diversity and economic productivity remains debatable; but, as first argued by AG, creativity and divergent thinking is a likely channel. A further potential mechanism is offered by Özak and Depetris-Chauvin. The use of survey data allows us to leverage ideas of societal interaction being driven by underlying genetic compositions to show evidence linking (local) genetic diversity with the development of noncognitive skills such as an orientation towards openness to new experiences or task specialization.

We, along with others (Ager and Brueckner 2016), provide a further step in validating the relationship between genetic diversity and economic development. The relevance of this research to policy applications will likely be controversial, and given the infancy of the empirical evidence, further replications and extensions are needed to verify this contentious relationship, with particular attention given to the precise mechanism through which genetic diversity is operating. However, the evidence so far suggests that the underlying genetic composition of groups does have an impact on behavior—personality and cooperation (Ashraf and Galor 2013b)—with a resulting impact on economic outcomes of interest.

---

<sup>37</sup> The limited ethnic setting of our study, while beneficial in limiting confounding factors associated with unobserved ethno-cultural associations, limits generalizing our findings to the global estimates of AG. Furthermore, the sample being restricted to European ancestry limits potential bias from European vs. non-European origins (Ager and Brueckner 2013; Easterly and Levine 2016).

## References

- Ager, P., & Brueckner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the Age of Mass Migration. *European Economic Review*, 64, 76-97.
- Ager, P., & Brueckner, M. (2016). Immigrants' genes: Genetic diversity and economic development in the US. Working Paper.
- Alesina, A., Harnoss, J., & Hillel, R. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2), 101-138.
- Alesina, A., & La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3), 762-800.
- Almund, M., Duckworth, A., Heckman, J., & Kautz, T. (2011). Personality, psychology, and economics. NBER Working Paper No. 16822.
- Arbalti, E., Ashraf, Q., & Galor, O. (2015). The nature of conflict. Working Paper. <https://ideas.repec.org/p/bro/econwp/2013-15.html>
- Ashraf, Q., & Galor, O. (2013a). The 'out of Africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1), 1-46.
- Ashraf, Q., & Galor, O. (2013b). Genetic diversity and the origins of cultural fragmentation. *American Economic Review, Papers and Proceedings*, 103(3), 528-533.
- Ashraf, Q., & Galor, O. (2017). The macrogenoeconomics of comparative development. NBER Working Paper No. 23199.
- Ashraf, Q., Galor, O., & Klemp, M. (2014). The out of Africa hypothesis of comparative development reflected by light intensity. Working Paper. <https://ideas.repec.org/p/bro/econwp/2014-4.html>
- Ashraf, Q., Galor, O., & Klemp, M. (2015). Heterogeneity and productivity. Working Paper. <https://ideas.repec.org/p/bro/econwp/2015-4.html>
- Boardman, J., Domingue, B., & Fletcher, J. (2012). How social and genetic factors predict friendship networks. *PNAS*, 109(43), 17377-17381.
- Chabris, C., Lee, J., Benjamin, D., Beuchamp, J., Glaeser, E., Borst, G., Pinker, S., & Laibson, D. (2013). Why is it hard to find genes that are associated with social science traits? Theoretical and empirical considerations. *American Journal of Public Health*, 103(S1), S152-S166.
- Chabris, C., Benjamin M. Hebert, Daniel J. Benjamin, Jonathan P. Beauchamp, David Cesarini, Matthijs J.H.M. van der Loos, Magnus Johannesson, Patrik K.E. Magnusson, Paul Lichtenstein, Craig
- S. Atwood, Jeremy Freese, Taissa S. Hauser, Robert M. Hauser, Nicholas A. Christakis, and David Laibson (2012). "Most Published Genetic Associations with General Intelligence Are Probably False Positives." *Psychological Science*, 23(11), 1314-1323. doi:10.1177/0956797611435528
- Comin, D., & Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5), 2031-2059.
- Costa, P. T., Jr., & McCrae, R. R. (1994). Set like plaster: Evidence for the stability of adult personality. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality*

- change? (pp. 21–40).  
 Washington, DC: American Psychological Association
- Depetris-Chauvin, E., & Özak, Ö. (2016). Population diversity, division of labor and comparative development. Working Paper.
- Domingue, B.W., Belsky, D.W., Harrati, A., Conley, D., Weir, D.R., & Boardman, J.D. (2017). Mortality selection in a genetic sample and implications for association studies. *International Journal of Epidemiology*, doi: 10.1093/ije/dyx041.
- Duncan, O.D. (1961). A Socioeconomic index for all occupations. In *Occupations and Social Status*. New York: Free Press of Glencoe.
- Easterly, W., & Levine, R. (2016). The European origins of economic development. *Journal of Economic Growth*, 21(3), 225-257.
- Feist, G.J. (1998). A meta-analysis of the impact of personality on scientific and artistic creativity. *Personality and Social Psychological Review*, 2, 290-309.
- Fowler, J., Settle, J., & Christakis, N. (2011). Correlated genotypes in friendship networks. *PNAS*, 108(5), 1993-1997.
- Furnham, A., & Bachtiar, V. (2008). Personality and intelligence as predictors of creativity. *Personality and Individual Differences*, 45, 613-617.
- Furnham, A., & Chamorro-Premuzic, T. (2004). Personality, intelligence, and art. *Personality and Individual Differences*, 36, 705-715.
- Hibbs, D., & Olsson, O. (2004). Geography, biogeography, and why some countries are rich and others are poor. *PNAS*, 101(10), 3715-3720.
- Hirsh, J.B., DeYoung, C.G., & Peterson, J.B. (2009). Metatraits of the big five differentially predict engagement and restraint of behavior. *Journal of Personality*, 77(4), 1085-1102.
- Hong, L., & Page, S. (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory*, 97, 123-163.
- Hong, L., & Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *PNAS*, 101, 16385-16389.
- Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation. *American Economic Journal: Macroeconomics*, 2, 31-56.
- Fogel, R.W. (2004). Health, nutrition, and economic growth. *Economic Development and Cultural Change*, 52(3), 643-658.
- Kaufman, S.B., Quilty, L.C., Grazioplene, R.G., Hirsh, J.B., Gray, J.R., Peterson, J.B., & DeYoung, C.G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, 84(2), 248-258.
- Kemeny, T. (2017). Immigrant diversity and economic performance in cities. *International Regional Science Review*, 40(2), 164-208.
- Lazear, E. (1999). Globalization and the market for teammates. *The Economic Journal*, 109, 15-40.
- King, L., Walker, L., & Broyles, S. (1996). Creativity and the five factor model. *Journal of Research in Personality*, 30, 189-203.
- McCrae, R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175-215.
- McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.),

- The five-factor model of personality: Theoretical perspectives (pp. 51–87). New York: Guilford Press.
- McCrae, R. R., Costa, P. T., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78, 173–186.
- Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *The Review of Economics and Statistics*, 94(1), 20-36.
- Olson, C., & Ackerman, D. (1998). “Wisconsin High School District Information for 1954-1957.” <https://www.ssc.wisc.edu/wlsresearch/documentation/supdoc/hsdistrict.txt>
- Ottaviano, G., and Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, 6, 9.
- Parrotta, P., Pozzoli, D., & Pytlikova, M. (2014). The nexus between labor diversity and firm’s innovation. *Journal of Population Economics*, 27, 303-364.
- Peri, G. (2012). The effect of immigration on productivity: Evidence from US states. *The Review of Economics and Statistics*, 94, 348-358.
- Peri, G., & Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1, 135-169.
- Pickering, A.D., Smillie, L.D., & DeYoung, C.G. (2016). Neurotic individuals are not creative thinkers. *Trends in Cognitive Science*, 20(1), 1-2.
- Preston, S.H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, 29(2), 231-248.
- Ruggles, S., Genadek, K., Goeken, R., Grover, J., & Sobek, M. (2015). Integrated Public Use Microdata Series: Version 6.0 [dataset]. Minneapolis: University of Minnesota. <http://doi.org/10.18128/D010.V6.0>.
- Siegel, P.M. (1971). Prestige in the American occupational structure. University of Chicago (doctoral dissertation).
- Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2), 469-529.
- Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature*, 51(2), 325-369.
- Acemoglu, D., Johnson, S., & Robinson, J. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.
- Srivastava, S., John, O., Gosling, S., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84(5), 1041-1053.

## Variable appendix

### Regressors of interest

*High school genetic diversity* An expected heterozygosity score calculated from high school gene frequencies. Beginning in 2007, the Wisconsin Longitudinal Survey (WLS) began collecting data on 96 single nucleotide polymorphisms (SNPs). These data were collected for roughly half of the original WLS respondents and selected siblings (~7,000). In constructing our high school level genetic diversity score, we first tabulate the high school frequency of each SNP variant using all available genetic data (graduates and siblings). These high school-specific gene frequencies are then used to calculate expected heterozygosity as specified in Ashraf and Galor (2013). This gives us a high school-specific measure of genetic diversity.

*County genetic diversity* This measure is for 1920 and comes from Ager and Brueckner (2016). It is found by matching the ancestry of European immigrants to the estimated genetic diversity score for 1500 CE from Ashraf and Galor (2013).

### Outcomes

*Years of schooling* The number of completed years of education for the WLS graduate. From the WLS variable *rb003red*.

*Duncan job prestige* A measure of job prestige based on income, education, and surveyed perceptions of general social standing for certain occupations (Duncan 1961). Measured for the WLS graduate's first job. From the WLS variable *ocsx1u2*.

*Siegel job prestige* A measure of job prestige based on surveys that evaluated perceptions on the "general" or "social" standing of certain occupations (Siegel 1971). Measured for the WLS graduate's first job. From the WLS variable *ocpx1u2*.

*Occupational education score* A measure of job prestige for the WLS graduate's first job that is based on percentage of people in an occupation that completed one year of college. From the WLS variable *ocex1*.

*Family income, 1974* Total earnings for WLS graduate's family during 1974. From the WLS variable *yfam74*.

*Family income, 1992* Total earnings for the WLS graduate's family during 1992. From the WLS variable *rp044hef*.

*Openness to experience* An additive score from a series of questions intended to measure the WLS graduate's personality trait of openness. We use the average from two waves of the WLS—1992 and 2004. From the WLS variables *mh032rei* and *ih032rei*.

*Extraversion* An additive score from a series of questions intended to measure the WLS graduate's personality trait of extraversion. We use the average from two waves of the WLS—1992 and 2004. From the WLS variables *mh001rei* and *ih001rei*.

*Conscientiousness* An additive score from a series of questions intended to measure the WLS graduate's personality trait of conscientiousness. We use the average from two waves of the WLS—1992 and 2004. From the WLS variables *mh017rei* and *ih017rei*.

*Agreeableness* An additive score from a series of questions intended to measure the WLS graduate's personality trait of agreeableness. We use the average from two waves of the WLS—1992 and 2004. From the WLS variables *mh009rei* and *ih009rei*.

*Neuroticism* An additive score from a series of questions intended to measure the WLS graduate's personality trait of neuroticism. We use the average from two waves of the WLS—1992 and 2004. From the WLS variables *mh025rei* and *ih025rei*.

*High school occupational diversity* Using data on detailed occupation code, we construct high school frequency of each occupation. This high school occupational frequency is then used to calculate the high school's occupational diversity in an identical manner as genetic diversity (both being roughly identical to a Hirsch Index). From the WLS variable *ocx1u*.

## **Controls**

### **Individual**

*IQ* WLS graduate's IQ score mapped from raw Henmon-Nelson test score. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *gwiiq\_bm*.

*Female* An indicator for the WLS graduate's sex. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *sexrsp*.

*Birth year* The WLS graduate's year of birth. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *brdxdy*.

### **Family**

*SES, 1957* Index comprised of the WLS graduate's father's years of schooling, mother's years of schooling, father's Duncan job prestige, and parental income in 1957, the year of the initial wave of the WLS. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *ses57*.

*Father's years of schooling* WLS graduate's father's years of schooling. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *bmfaedu*.

*Mother's years of schooling* WLS graduate's mother's years of schooling. For the high school level sample of Tables 8 and 9, the high school mean is used. From WLS variable *bmmaedu*.

### **School**

*High school size* The size of the WLS graduate's graduating class. From WLS variable

*hssize*.

*Madison/Milwaukee indicator* An indicator for whether the WLS graduate lived in either Madison or Milwaukee during the 1957 collection wave. From WLS variable *pop57*.

*Teacher salary* The average teacher salary by district for years 1954–1957. High school resource data were compiled by Olson and Ackerman (1998). To preserve sample size, missing values are imputed by the county-level mean (an indicator for imputed observations is included); roughly 2% of observations have been imputed.

*Teacher experience* The average years of experience for teachers by district for years 1954–1957. Two measures of experience are used: within school district experience and teacher total experience. High school resource data were compiled by Olson and Ackerman (1998). To preserve sample size, missing values are imputed by the county-level mean (an indicator for imputed observations is included); roughly 2% of observations have been imputed.

*Teacher years of schooling* The average number of postsecondary years of schooling for teachers by district for years 1954–1957. High school resource data were compiled by Olson and Ackerman (1998). To preserve sample size, missing values are imputed by the county-level mean (an indicator for imputed observations is included); roughly 2% of observations have been imputed.

*Number of school days* The average number of school days by district for years 1954–1957. High school resource data were compiled by Olson and Ackerman (1998). To preserve sample size, missing values are imputed by the county-level mean (an indicator for imputed observations is included); roughly 2% of observations have been imputed.

*Classroom size* The average classroom size by district for years 1954–1957. The variable is created from the total number of enrolled of high school students (variables *en\_bt\_7* and *en\_gt\_7* in Olson and Ackerman 1998) divided by the total number of teachers across the same time frame (variable *te\_tea\_t*).

## **Historic**

All historic controls are found by matching the WLS graduate's father's ancestry to country-level data in Ashraf and Galor (2013). These data are then averaged at the high school level. The set of country-level variables includes absolute latitude, the fraction of arable land, the mean temperature, mean precipitation, mean elevation, an index of roughness, the mean distance to the coast or navigable river, and the fraction of land within 100 kilometers of the coast or a navigable river. All variables and prior sources are found in Ashraf and Galor (2013).

## **Robust**

*Indicator for father's nationality* The WLS graduate's reported father's ancestral nationality. For the high school level sample of Table 9, the high school fraction of each

reported ancestry is used. From the WLS variable *natfth*.

*Ethnic fractionalization* An ethnic fractionalization score derived from the high school frequency of father's nationality. Calculated in identical manner as high school genetic diversity and high school occupational diversity. From the WLS variable *natfth*.

*Genetic markers* An additive score (e.g., 0, 1, or 2) of the variant for each SNP in the WLS. These data are for roughly 4,500 WLS graduates. To prevent a loss in sample size, missing graduates are assigned the mean for each SNP; an individual-level indicator for those with missing genetic data is also included. For the high school level sample of Table 9, the high school frequency is used.

*Principal component of genetic markers* The first principal component of all SNPs used to calculate high school genetic diversity.

*Industry diversity* A measure of diversity based on the WLS graduate's industry associated with their first job. Graduates are assigned to one of twelve industry classifications. The fraction of the high school within each industry is then used to calculate a measure of diversity. From WLS variable *inmx1u*.

Table 1. Summary Statistics

Variable	N	Mean	Std. Deviation	Max	Min
Genetic Diversity					
Individual Sample	8675	0.3296	0.0129	0.2687	0.3519
High School Sample	332	0.3211	0.0161	0.2687	0.3519
County Sample (Ager and Brueckner 2016)	70	7.8970	0.8204	6.2801	11.43
Baseline Controls					
Individual					
IQ	8675	101.07	14.80	61	145
Female Indicator	8675	0.53	0.50	0	1
Birth Year	8675	1938.85	0.52	1930	1941
Family					
Family SES Index, 1957	8675	16.40	10.94	1	97
Mother's Years of Schooling	8675	10.47	2.79	0	21
Father's Years of Schooling	8675	9.78	3.40	0	26
School					
High School Class Size	8675	180.00	133.31	11	482
Madison/Milwaukee Indicator	8675	0.18	0.38	0	1
Teacher Salary	8675	4617.18	646.32	3064.29	5748.19
Teacher Experience (in district)	8675	6.0146	1.7818	1.2386	12.9243
Teacher Experience (total)	8675	7.2403	1.4289	2.1083	9.8262
Teacher Years of Schooling (post-secondary)	8675	4.5443	0.4382	2.8564	5.7314
Number of Sch. Days	8675	178.7428	4.2420	171.25	190
Classroom Size	8675	21.7863	2.5606	10.2671	28.0469
Historic					
Absolute Latitude	8675	52.20	1.82	41	61
Frac. Arable	8675	28.98	5.07	5.56	43.36
Avg. Temperature	8675	7.70	1.08	2.12	13.08
Roughness Index	8675	0.11	0.02	0.02	0.24
Land within 100 km of coast or river	8675	0.86	0.07	0.46	1
Avg. Dist. to Coast or River	8675	0.09	0.07	0.03	0.46
Avg. Elevation	8675	0.27	0.05	0.05	0.62
Precipitation	8675	69.97	5.99	54.27	91.66
Dependent Variables					
Years of Schooling	7703	13.65	2.27	12	21
Duncan Occupation Prestige	8098	436.97	231.32	20	960
Siegel Occupation Prestige	8098	405.55	143.13	141	812
Occupational Education	8098	321.44	294.85	8	995
Family Income, 1974 (in \$1,000s)	8270	20.62	17.15	0	501.79
Family Income, 1992 (in \$1,000s)	7443	73.85	178.96	0	8500
Openness to Experience	7148	21.60	4.47	6	36
Conscientiousness	7170	28.96	3.92	13	36
Extraversion	7159	22.86	5.02	6	36
Agreeableness	7167	28.60	4.11	7.5	36
Neuroticism	7156	15.61	4.51	5	30
Within-Industry Job Diversity	332	0.5821	0.1795	0	0.8912

**Summary & Notes:** This table provides summary statistics for all variables used within the empirical analysis. Variable descriptions and sources can be found in the Variable Appendix. The calculation of Ager and Brueckner's (2017) measure of county-level genetic diversity differs from that of AG and our high school measure. In short, Ager and Brueckner's measure is calculated from the number of immigrants and the immigrants source country AG genetic diversity.

Table 2. Genetic Diversity's Effect on Education

Dependent Variable: Years of Schooling							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Genetic Diversity	17.6377*** (3.9636)	16.1662*** (3.4429)	13.5900*** (2.6196)	10.1605*** (2.6310)	12.2673** (5.0713)	14.8302*** (3.3960)	11.3359*** (2.9096)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7703	7703	7703	7703	7703	7703	7703
R Sqr.	0.0100	0.0063	0.2238	0.1574	0.0146	0.0213	0.3017

**Summary & Notes:** Our primary hypothesis is that attending more genetically diverse high school is positively associated with later-life socioeconomic outcomes. To test this hypothesis, Table 2 regresses an individual's years of schooling on the genetic diversity score of their 1957 Wisconsin high school. Variables that comprise each set of controls are given in Table 1 and the Variable Appendix. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 3. Genetic Diversity's Effect on Job Prestige

Dependent Variable: Std. Occupational Prestige for First Job							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Duncan Prestige Score							
Genetic Diversity	7.9326*** (1.8475)	5.5108*** (1.2147)	5.0022*** (1.0903)	3.4328*** (1.0006)	2.5246 (1.5311)	5.1109*** (1.1240)	2.3676** (0.9562)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.0107	0.0039	0.1557	0.0894	0.0131	0.0128	0.2007
Panel B. Seigel Prestige Score							
Genetic Diversity	6.4083*** (1.5659)	4.6898*** (1.1115)	4.1020*** (0.9062)	2.5446*** (0.8341)	2.1764 (1.6639)	4.4233*** (1.0152)	2.1178** (1.0049)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.0069	0.0028	0.1478	0.0949	0.0108	0.0115	0.1949
Panel C. Occupational Education							
Genetic Diversity	5.8600*** (1.4024)	5.4921*** (1.0962)	4.8392*** (0.8919)	3.0304*** (0.7992)	3.9145** (1.7873)	5.0855*** (0.9881)	3.6701*** (1.0018)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.0058	0.0038	0.1558	0.1235	0.0123	0.0160	0.2207

**Summary & Notes:** Exposure to diverse individuals is hypothesized to increase creativity and aptitude associated with a more prestigious occupation. As a test of this, Table 3 regresses an individual's job prestige on the genetic diversity score of their 1957 Wisconsin high school. Panel A considers the measure of prestige from Duncan (1961); Panel B considers a similar measure from Siegel (1971); and Panel C considers the fraction with one or more years of college for a given occupation. Variables that comprise each set of controls are given in Table 1 and the Variable Appendix. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 4. Genetic Diversity's Effect on Income

Dependent Variable: Natural Log of Family Income in Specified Wave							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. 1974							
Genetic Diversity	2.6595** (1.1152)	2.8973** (1.2847)	2.6972** (1.2103)	2.1835* (1.2301)	2.4678* (1.3796)	2.7556** (1.2411)	2.3667* (1.2392)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	8270	8270	8270	8270	8270	8270	8270
R Sqr.	0.0013	0.0011	0.0176	0.0116	0.0027	0.0018	0.0248
Panel B. 1992							
Genetic Diversity	6.3307** (2.4720)	5.7768** (2.8556)	4.4193* (2.4968)	4.5701* (2.7092)	6.7241* (3.5678)	6.0929** (2.8879)	7.1242** (3.2256)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7443	7443	7443	7443	7443	7443	7443
R Sqr.	0.0009	0.0005	0.0506	0.0057	0.0018	0.0015	0.0525

**Summary & Notes:** Table 3 examines the relationship between an individual's family income in two waves of the WLS on the genetic diversity score of their 1957 Wisconsin high school. Panel A considers the natural log of family income in the 1974 wave of the WLS, while Panel B considers an equivalent measure for the 1992 wave. Variables that comprise each set of controls are given in Table 1 and the Variable Appendix. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 5. Genetic Diversity's Effect on Personality

Dependent Variable: Index of Specified Personality Trait							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Openness to Experience							
Genetic Diversity	26.4792*** (5.6082)	19.0542*** (4.5326)	15.4727*** (3.8635)	12.4703*** (3.9486)	8.5289 (6.1148)	18.4017*** (4.3346)	8.3387* (4.6258)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7148	7148	7148	7148	7148	7148	7148
R Sqr.	0.0058	0.0022	0.0762	0.0510	0.0080	0.0060	0.1015
Panel B. Extraversion							
Genetic Diversity	13.8282** (5.8665)	14.2696** (6.5161)	14.5255** (6.4986)	10.9299* (6.3029)	13.1575* (6.6110)	12.3704* (6.3031)	10.5053* (6.1930)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7159	7159	7159	7159	7159	7159	7159
R Sqr.	0.0013	0.0010	0.0040	0.0089	0.0017	0.0030	0.0153
Panel C. p.c. of Openness and Extraversion							
Genetic Diversity	6.1863*** (1.5249)	5.0449*** (1.4118)	4.5144*** (1.2877)	3.5365*** (1.2757)	3.2045** (1.5711)	4.6808*** (1.3509)	2.8041** (1.2371)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7145	7145	7145	7145	7145	7145	7145
R Sqr.	0.0049	0.0024	0.0298	0.0385	0.0056	0.0061	0.0549

**Summary & Notes:** As a test of the mechanisms behind genetic diversity's beneficial role in later-life economic well-being, Table 5 tests the relationship between high school genetic diversity and individual personality traits. Particular attention is given to the traits of openness and extraversion, which have been routinely associated with individual creativity and divergent thinking—two traits at the root of AG's proposed beneficial effect of diversity. Panel A considers the index for openness to experience; Panel B considers the index for extraversion; and Panel C considers the first principal component between openness and extraversion: this shared variation attempts to more accurately measure divergent thinking. Variables that comprise each set of controls are given in Table 1 and the Variable Appendix. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 6. Genetic Diversity and Other Personality

Dependent Variable: Index of Specified Personality Trait							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Conscientiousness							
Genetic Diversity	-2.7853 (3.3642)	-0.7780 (4.0783)	-0.5844 (4.0603)	-0.2638 (4.0004)	-2.0769 (4.2743)	0.1706 (4.3224)	-1.2845 (4.3367)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7170	7170	7170	7170	7170	7170	7170
R Sqr.	0.0001	0.0000	0.0007	0.0003	0.0005	0.0007	0.0023
Panel B. Agreeableness							
Genetic Diversity	-7.0674* (4.0366)	-1.9628 (4.4256)	-1.3228 (4.4797)	-1.3920 (4.6514)	-2.2180 (5.0232)	-0.6895 (4.0611)	-2.8821 (4.7182)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7167	7167	7167	7167	7167	7167	7167
R Sqr.	0.0005	0.0000	0.0580	0.0018	0.0010	0.0012	0.0602
Panel C. Neuroticism							
Genetic Diversity	-8.8302*** (3.1307)	-4.6416 (3.8957)	-2.8528 (4.0678)	-2.6619 (3.9730)	-4.7050 (4.1374)	-4.5488 (3.9087)	-5.5814 (4.3503)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	7156	7156	7156	7156	7156	7156	7156
R Sqr.	0.0006	0.0001	0.0348	0.0069	0.0015	0.0009	0.0391

**Summary & Notes:** Additional personality indices are tested in Table 6. As shown, no statistical relationship exists for personality measures other than those tested in Table 5. Variables that comprise each set of controls are given in Table 1 and the Variable Appendix. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 7. Genetic Diversity's Effect on Task Diversity

Dependent Variable: Within-Industry Occupational Diversity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Genetic Diversity	6.7061*** (0.6746)	5.6217*** (0.5669)	5.4426*** (0.5825)	4.9200*** (0.5410)	3.0387*** (0.6493)	5.6899*** (0.5125)	2.5359*** (0.6228)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	332	332	332	332	332	332	332
R Sqr.	0.3608	0.3058	0.3370	0.3755	0.5224	0.3281	0.5759

**Summary & Notes:** As a second test of the mechanisms behind genetic diversity's beneficial role in later-life economic well-being, Table 7 tests the relationship between high school genetic diversity and a measure for high school job diversity. This test follows the framework originally proposed by Depetris-Chauvin and Özak (2015). Individually controls include mean high school IQ, mean birth year, and the fraction of female students. Family controls include the high school mean of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. School controls include high school size class size, an indicator for being in either Madison or Milwaukee, mean teacher salary, mean teacher experience, mean teacher education, school days, and classroom size. Historic controls include ancestral high school averages of absolute latitude, fraction of arable land, mean temperature, mean precipitation, mean elevation, roughness index, mean distance to the coast or navigable river, and the fraction of land within 100km of the coast or navigable river. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 8. Moderation from Proposed Mechanisms?

Dependent Variable:	Years of Schooling (1)	Duncan Occ. Prestige (2)	Seigel Occ. Prestige (3)	Occupational Education (4)	Fam. Inc. 1974 (5)	Fam. Inc. 1992 (6)
Panel A. "Horse Race"						
Genetic Diversity	9.8861*** (3.2352)	1.7759* (0.9833)	1.6297 (1.1357)	3.2966*** (1.1234)	1.9587 (1.2410)	7.7149** (3.7050)
Openness to experience	0.1226*** (0.0059)	0.0223*** (0.0029)	0.0277*** (0.0031)	0.0369*** (0.0035)	0.0006 (0.0028)	0.0282*** (0.0086)
Extraversion	-0.0236*** (0.0048)	-0.0030 (0.0020)	-0.0027 (0.0024)	-0.0070*** (0.0025)	0.0148*** (0.0024)	0.0017 (0.0064)
W/in-Ind. Job Diversity	0.2797 (0.3248)	0.1552 (0.1459)	0.1170 (0.1340)	0.0544 (0.1341)	0.0475 (0.1415)	-0.5281 (0.4396)
County Fixed Effects	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
Observations	7703	8098	8098	8098	8270	7443
R Sqr.	0.3451	0.2080	0.2059	0.2393	0.0310	0.0647
Panel B. Residuals						
Genetic Diversity	7.0414** (2.8068)	-0.3279 (0.9692)	-0.0581 (0.9774)	1.9260** (0.9525)	1.3785 (1.2736)	5.8607* (3.0932)
County Fixed Effects	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
Observations	7703	8098	8098	8098	8270	7443
R Sqr.	0.2256	0.1526	0.1447	0.1625	0.0202	0.0414

**Summary & Notes:** This table considers the association between genetic diversity and SES outcomes from Tables 2-5 when accounting for the proposed personality and task diversity mechanisms. Panel A controls for all mechanisms, while Panel B considers the residuals of SES outcomes when regressing each on openness, extraversion, and within-industry job diversity. County fixed effects are included for each estimation, and baseline controls include individual, family, school, and historic controls (outlined in Table 1). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 9A. Robustness: Education and Job Prestige

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Years of Schooling								
Genetic Diversity	11.3359*** (2.9096)	11.3980*** (2.9334)	11.4244*** (2.8669)	11.3980*** (2.9334)	11.0944*** (3.0428)	11.3138*** (2.9093)	11.2038*** (3.0170)	11.3781*** (2.9333)
Observations	7703	7703	7703	7703	7703	7703	7703	7703
R Sqr.	0.3017	0.3086	0.3017	0.3086	0.3201	0.3017	0.3269	0.3086
Panel B. Std. Duncan Prestige								
Genetic Diversity	2.3676** (0.9562)	2.4182** (0.9610)	2.2813** (0.9585)	2.4182** (0.9610)	2.0489** (0.9436)	2.3959** (0.9517)	2.0851** (0.9466)	2.4461** (0.9572)
Observations	8098	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.2007	0.2062	0.2008	0.2062	0.2195	0.2008	0.2254	0.2063
Panel C. Std. Siegel Prestige								
Genetic Diversity	2.1178** (1.0049)	2.1430** (1.0439)	2.0520** (1.0050)	2.1430** (1.0439)	1.9190** (0.9361)	2.1697** (1.0040)	1.9373** (0.9668)	2.1943** (1.0433)
Observations	8098	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.1949	0.1997	0.1950	0.1997	0.2167	0.1953	0.2217	0.2000
Panel D. Std. Occupational Education								
Genetic Diversity	3.6701*** (1.0018)	3.7846*** (1.0102)	3.6832*** (0.9853)	3.7846*** (1.0102)	3.4579*** (0.9856)	3.6826*** (1.0068)	3.5492*** (0.9874)	3.7990*** (1.0149)
Observations	8098	8098	8098	8098	8098	8098	8098	8098
R Sqr.	0.2207	0.2262	0.2207	0.2262	0.2417	0.2207	0.2475	0.2263
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y	Y
Ind. for Father's Nat.	N	Y	N	Y	N	N	Y	Y
Ethnic Frac.	N	N	Y	Y	N	N	Y	Y
Genetic Markers	N	N	N	N	Y	N	Y	N
P.C. of Genetic Markers	N	N	N	N	N	Y	N	Y

**Summary & Notes:** Table 9 examines the robustness of the previously estimated association of genetic diversity. Column (1) replicates the baseline estimation (col. (7) in above Tables 2-7). Columns (2)-(4) introduce controls to account for potential broad-based ethnic differences. These include controlling for an indicator of ancestral background (col. 2) and the level of high school fractionalization from this indicator (col. 3). Columns (5) and (6) attempt to account for gene-specific effects by controlling for each individual's genetic markers that are used to construct the diversity scores (col. 5) and the shared variation of these markers (col. 6). County fixed effects are included for each estimation, and baseline controls include individual, family, school, and historic controls (outlined in Tables 1 and the Variable Appendix). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 9B. Robustness: Income and Personality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. ln Family Income, 1974								
Genetic Diversity	2.3667*	2.3223*	2.2632*	2.3223*	2.1852*	2.3708*	2.1134*	2.3259*
	(1.2392)	(1.2659)	(1.2098)	(1.2659)	(1.2321)	(1.2354)	(1.2520)	(1.2627)
Observations	8270	8270	8270	8270	8270	8270	8270	8270
R Sqr.	0.0248	0.0329	0.0249	0.0329	0.0430	0.0248	0.0515	0.0329
Panel B. ln Family Income, 1992								
Genetic Diversity	7.1242**	7.0061**	6.8399**	7.0061**	5.6229*	7.1255**	5.5835*	7.0153**
	(3.2256)	(3.3230)	(3.3253)	(3.3230)	(3.0332)	(3.2218)	(3.1340)	(3.3216)
Observations	7443	7443	7443	7443	7443	7443	7443	7443
R Sqr.	0.0525	0.0583	0.0526	0.0583	0.0809	0.0525	0.0867	0.0583
Panel C. Openness to Experience								
Genetic Diversity	8.3387*	8.2075*	8.1888*	8.2075*	6.1337	8.3021*	5.9860	8.1943*
	(4.6258)	(4.6624)	(4.6845)	(4.6624)	(4.7371)	(4.5697)	(4.8195)	(4.6159)
Observations	7148	7148	7148	7148	7148	7148	7148	7148
R Sqr.	0.1015	0.1093	0.1015	0.1093	0.1264	0.1015	0.1338	0.1093
Panel D. Extraversion								
Genetic Diversity	10.5053*	10.6123*	10.3769	10.6123*	7.8333	10.5117*	7.6157	10.6140*
	(6.1930)	(6.2778)	(6.2993)	(6.2778)	(6.2539)	(6.2106)	(6.3308)	(6.2953)
Observations	7159	7159	7159	7159	7159	7159	7159	7159
R Sqr.	0.0153	0.0216	0.0153	0.0216	0.0403	0.0153	0.0471	0.0216
Panel E. Principal Component of Open. and Ext.								
Genetic Diversity	2.8016**	2.7956**	2.7603**	2.7956**	2.0538	2.7967**	1.9974	2.7938**
	(1.2363)	(1.2511)	(1.2610)	(1.2511)	(1.2588)	(1.2302)	(1.2853)	(1.2463)
Observations	7145	7145	7145	7145	7145	7145	7145	7145
R Sqr.	0.0547	0.0621	0.0547	0.0621	0.0794	0.0547	0.0865	0.0621
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y	Y
Ind. for Father's Nat.	N	Y	N	Y	N	N	Y	Y
Ethnic Frac.	N	N	Y	Y	N	N	Y	Y
Genetic Markers	N	N	N	N	Y	N	Y	N
P.C. of Genetic Markers	N	N	N	N	N	Y	N	Y

**Summary & Notes:** Table 9 examines the robustness of the previously estimated association of genetic diversity. Column (1) replicates the baseline estimation (col. (7) in Tables 2-7). Columns (2)-(4) introduce controls to account for potential broad-based ethnic differences. These include controlling for an indicator of ancestral background (col. 2) and the level of high school fractionalization from this indicator (col. 3). Columns (5) and (6) attempt to account for gene-specific effects by controlling for each individual's genetic markers that are used to construct the diversity scores (col. 5) and the shared variation of these markers (col. 6). County fixed effects are included for each estimation, and baseline controls include individual, family, school, and historic controls (outlined in Tables 1 and the Variable Appendix). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 10. Genetic Diversity and Task Diversity: Robustness

Dependent Variable: Within-Industry Occupational Diversity				
	(1)	(2)	(3)	(4)
Genetic Diversity	1.8544** (0.7889)	2.5359*** (0.6228)	2.6259*** (0.4863)	2.0428*** (0.6094)
County Fixed Effects	Y	Y	Y	Y
Base Controls	Y	Y	Y	Y
Indicator for Father's Nationality	Y	N	N	Y
Ethnic Fractionalization	Y	N	N	Y
Gene Frequencies	N	Y	N	Y
Industry Diversity	N	N	Y	Y
Observations	332	332	332	332
R Sqr.	0.6548	0.5759	0.6740	0.7351

**Summary & Notes:** Table 10 estimates similar robustness checks as Table 9, testing the association between within-industry occupation diversity and genetic diversity of Table 7. Column (1) includes ancestral fractions and ethnic fractionalization; column (2) controls for the high school frequency of each SNP; column (3) controls for broad-based industry diversity; and column (4) includes all robust controls. County fixed effects are included for each estimation, and baseline controls include individual, family, school, and historic controls (outlined in Table 1 and the Variable Appendix). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.

Table 11. County Replication with WLS Parents

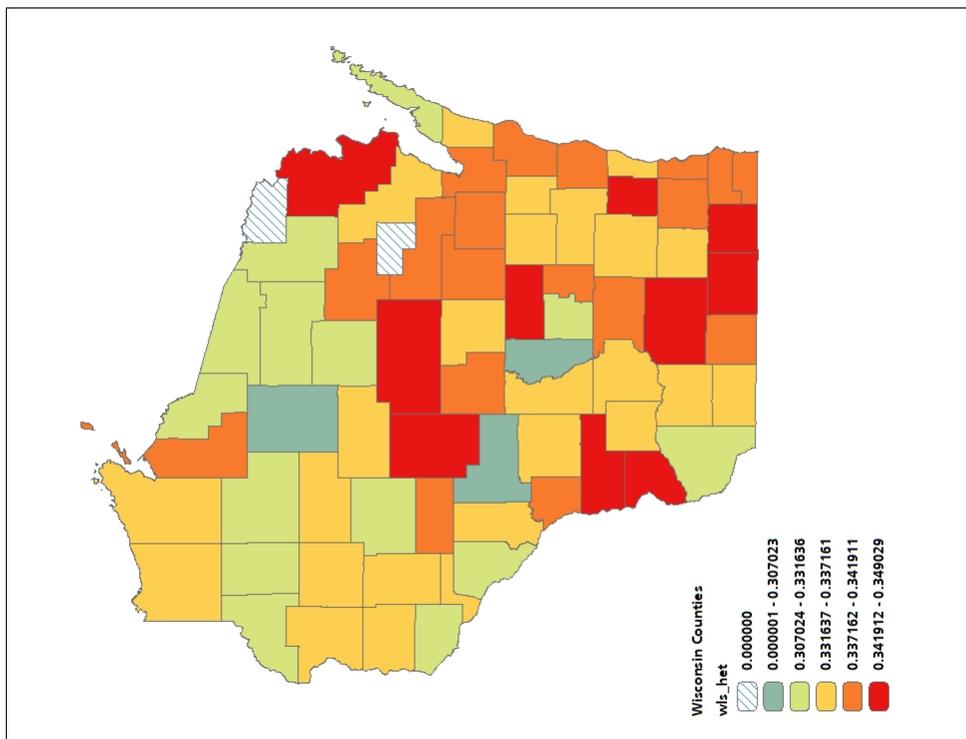
Dependent Variable:	SES 1957 (1)	Father's Job Prestige (2)	Father's YOS (3)	Mother's YOS (4)	Father's Job Diversity (5)
Genetic Diversity	2.4029*** (0.4681) [0.4268]	0.2271*** (0.0253) [0.0290]	0.4798*** (0.1326) [0.1215]	0.0612 (0.1036) [0.0679]	0.0752*** (0.0517) [0.0217]
Frac. Born outside of County, 1940	46.3326*** (9.4784)	3.5717*** (0.5517)	11.7207*** (3.1212)	5.6165** (2.1249)	1.9816 (1.2513)
Population Density, 1920	0.0017 (0.0010)	0.0001 (0.0001)	0.0005** (0.0003)	0.0008*** (0.0002)	-0.0000 (0.0001)
Frac. born in WI, 1920	-4.2510 (15.9746)	-0.3850 (1.4526)	-2.0830 (3.3815)	-2.7188 (2.6636)	-0.3182 (0.8922)
Frac. born in US (other than WI), 1920	-8.3825 (15.5015)	-0.5176 (1.3722)	-3.1868 (3.2739)	-2.6951 (2.7802)	-0.7230 (0.9559)
Frac. foreign born, 1920	Y	Y	Y	Y	Y
Observations	8690	7682	8690	8690	70
R Sqr.	0.0705	0.0666	0.0356	0.0256	0.3653

**Summary & Notes:** Table 11 explores the relationship between 1920 county-level genetic diversity (from Ager and Brueckner 2016) and outcomes for the WLS graduate's parents, who were of high school age in 1920. This analysis is seen as an independent replication for a sample individuals that were not selected by educational status. County-level population fractions come the full 1920 and 1940 censuses. The fraction of migrants in 1940 is the number of individuals who reported living in a different county in the last 5 years and is intended to control for the possibility that the district WLS parents reside in 1957 is different than from the 1920 residence. For those born outside the US, we consider only country origins with greater than 10,000 immigrants in 1920. These are Germany, Poland, Norway, Sweden, Russia (USSR), Canada, Czechoslovakia, Austria, England, Italy, and Hungary. Standard errors clustered by county are given in parenthesis; spatially adjusted standard errors (50km) are given in brackets. Statistical significance is calculated for spatially adjusted standard errors and is denoted at the 10%, 5%, and 1% levels respectively by \*, \*\*, and \*\*\*.

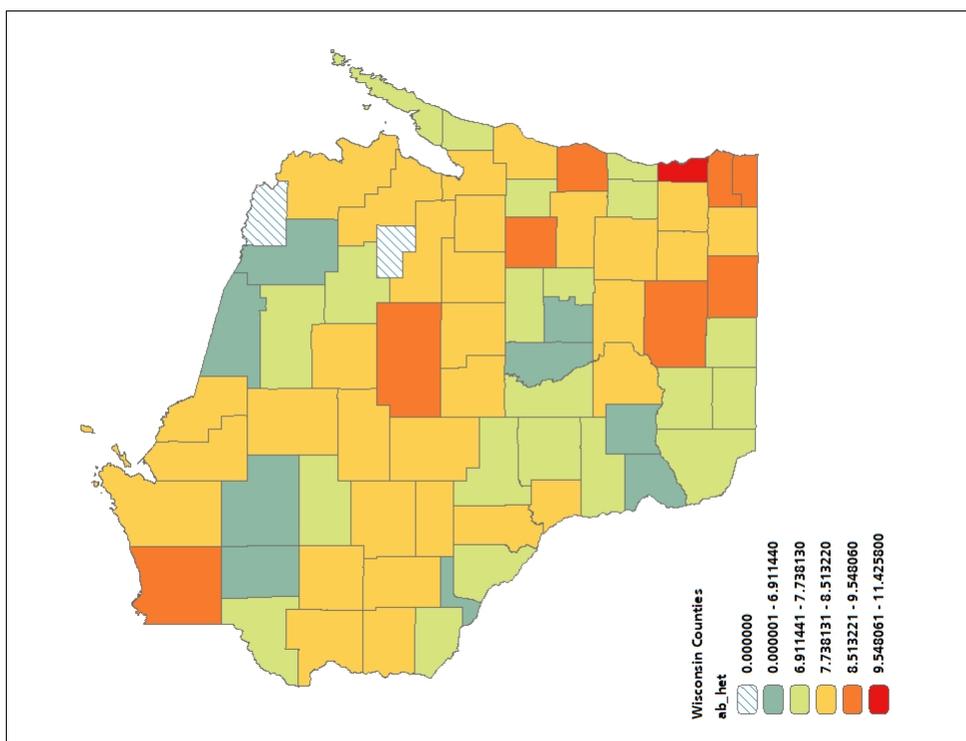
Table 12. Genetic Diversity’s Effect on an Alternative Measure of Task Diversity

Dependent Variable: School Occupational Diversity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Genetic Diversity	2.0889*** (0.1881)	1.9191*** (0.1628)	1.8687*** (0.1728)	1.7368*** (0.1590)	1.1720*** (0.1338)	1.9272*** (0.1592)	0.9746*** (0.1201)
County Fixed Effects	N	Y	Y	Y	Y	Y	Y
Individual	N	N	Y	N	N	N	Y
Family	N	N	N	Y	N	N	Y
School	N	N	N	N	Y	N	Y
Historic	N	N	N	N	N	Y	Y
Observations	332	332	332	332	332	332	332
R Sqr.	0.4023	0.3591	0.3891	0.4130	0.5890	0.3869	0.6484

**Summary & Notes:** As a second test of the mechanisms behind genetic diversity’s association with task diversity, Table 12 estimates the relationship between high school genetic diversity and an alternative measure for high school job diversity: the overall measure of occupational diversity, as opposed to the *within-industry* measure of Table 7. This test confirms the analysis of Table 7. Individually controls include mean high school IQ, mean birth year, and the fraction of female students. Family controls include the high school mean of socioeconomic status in 1957, father’s years of schooling, and mother’s years of schooling. School controls include high school size class size, an indicator for being in either Madison or Milwaukee, mean teacher salary, mean teacher experience, mean teacher education, school days, and classroom size. Historic controls include ancestral high school averages of absolute latitude, fraction of arable land, mean temperature, mean precipitation, mean elevation, roughness index, mean distance to the coast or navigable river, and the fraction of land within 100km of the coast or navigable river. County fixed effects are included in columns (2)-(7). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by \*, \*\*, and \*\*\*.



(a) Ager and Brueckner, 1920



(b) WLS, 1957

Figure 1. Two Measures of Genetic Diversity in Wisconsin

**Summary & Notes:** This figure plots two independent measures of county-level genetic diversity. The correlation coefficient between the two measures is 0.33 ( $p < 0.01$ ).

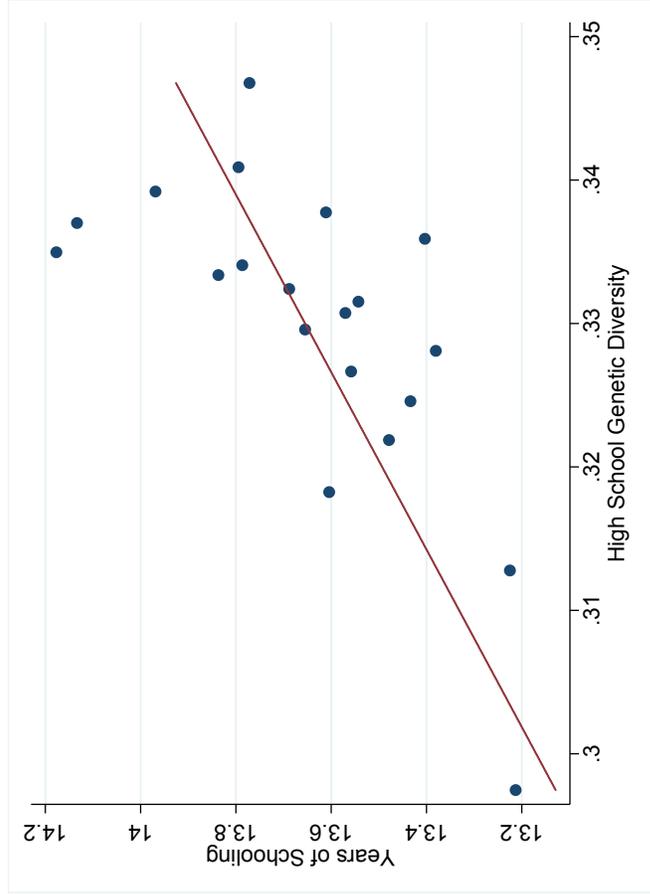
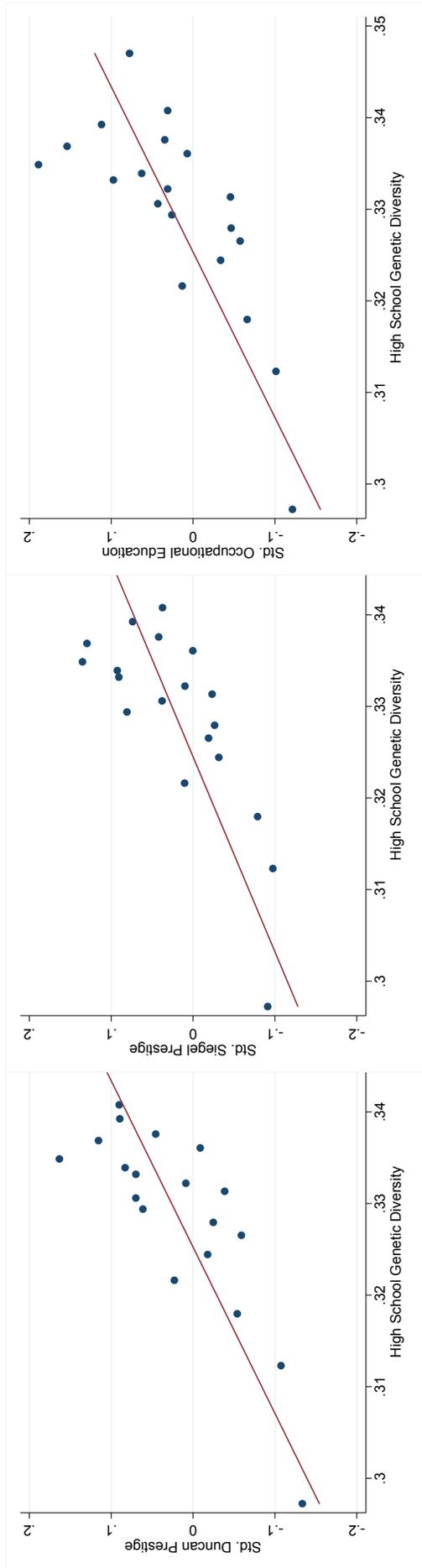


Figure 2. Years of Schooling and Genetic Diversity

**Summary & Notes:** This figure plots the within county relationship between years of schooling and high school genetic diversity for 20 equally sized bins.



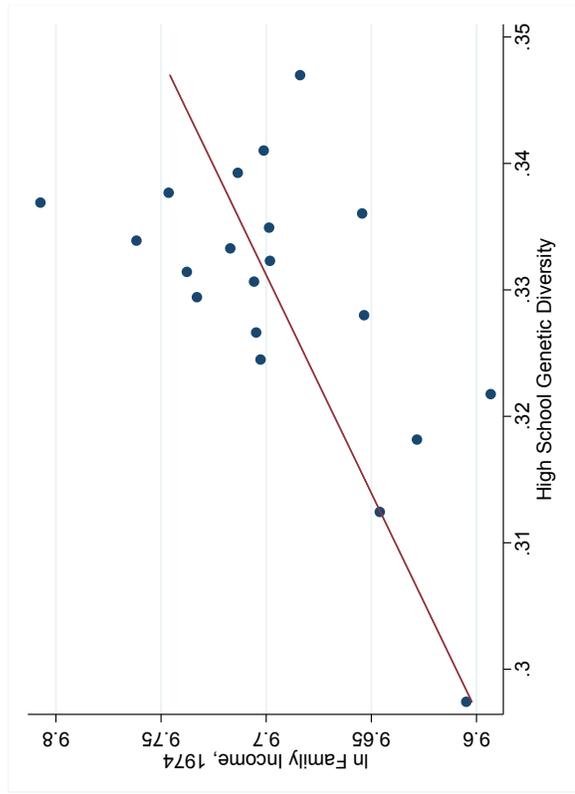
(a) Duncan

(b) Siegel

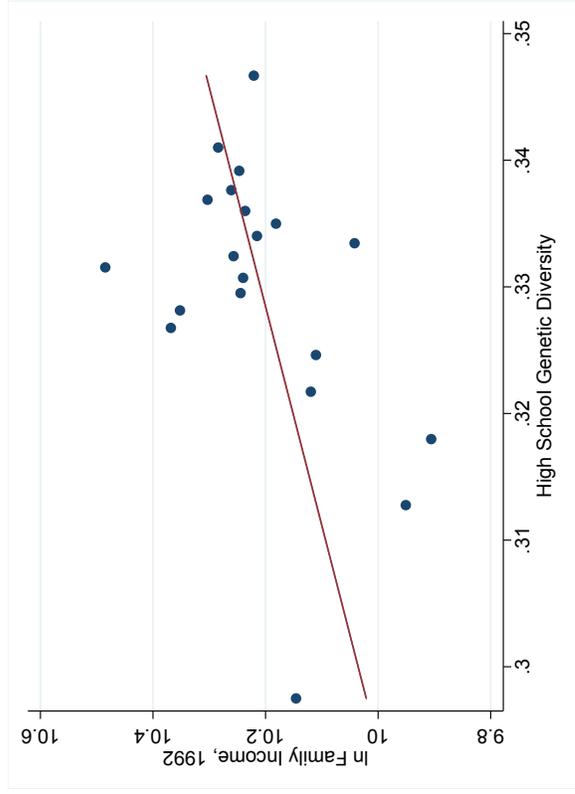
(c) Occ. Edu.

Figure 3. Occupational Prestige and Genetic Diversity

**Summary & Notes:** This figure plots the within county relationship between occupational prestige for the WLS graduate's first job and high school genetic diversity for 20 equally sized bins.



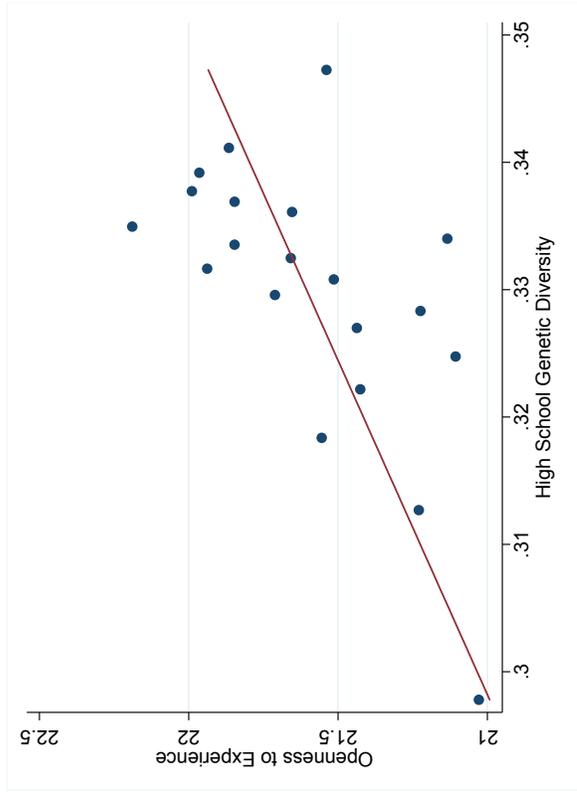
(a) ln Family Income, 1974



(b) ln Family Income, 1992

Figure 4. Family Income and Genetic Diversity

**Summary & Notes:** This figure plots the within county relationship between later-life family income and high school genetic diversity for 20 equally sized bins.



(a) Openness to Experience

