

NBER WORKING PAPER SERIES

HIGH SCHOOL GENETIC DIVERSITY AND LATER-LIFE STUDENT OUTCOMES:
MICRO-LEVEL EVIDENCE FROM THE WISCONSIN LONGITUDINAL STUDY

C. Justin Cook
Jason M. Fletcher

Working Paper 23520
<http://www.nber.org/papers/w23520>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2017

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 Wisconsin Longitudinal Study 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by C. Justin Cook and Jason M. Fletcher. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

High School Genetic Diversity and Later-life Student Outcomes: Micro-level Evidence from the Wisconsin Longitudinal Study

C. Justin Cook and Jason M. Fletcher

NBER Working Paper No. 23520

June 2017

JEL No. J24,O4

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 novel question by subjecting the hypothesis to an alternative context that eliminates many alternative hypotheses by aggregating representative data to the high school level from a single 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 effect of genetic diversity on socioeconomic outcomes. Additionally, we check an existing mechanism and propose a new potential mechanism of the results for innovation: personality traits associated with creativity and divergent thinking.

C. Justin Cook

School of Social Sciences, Humanities and Arts

University of California, Merced

jcook9@ucmerced.edu

Jason M. Fletcher

University of Wisconsin-Madison

La Follette School of Public Affairs

1225 Observatory Drive

Madison, WI 53706

and NBER

jfletcher@lafollette.wisc.edu

Introduction

Novel and controversial 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.² Taken together, these two effects create a non-linear (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).

These findings have been the focus of a multidisciplinary debate (Ashraf and Galor 2013c, 2013d; Callaway 2012; d’Alpoim Guedes et al. 2013; and Gelman 2013), though much of the discussion has focused on the conceptual framing and interpretation of the original analysis rather than on specifying and attempting to rule out alternative hypotheses consistent with the results. In the current paper, we propose a strong test of the robustness of the relationship between genetic diversity and indicators of economic development 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 (i.e. confounding) (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 (iii) Clarifying

² 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.

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 (eliminating across-country confounders) comprised entirely of European ancestry (limiting issues of population stratification) to assess the robustness of the key findings in AG in an entirely different context. Thus, we consider our analysis to be a strong test of their original hypothesis, while ruling out several important alternative hypotheses consistent with their results. We also pursue initial findings in this literature by further exploring the mechanisms of genetic diversity. Specifically, we take advantage of measures of the Big Five personality traits in our sample 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.³ We also provide new evidence consistent with a prior hypothesis of genetic diversity's role in task specialization (Depetris-Chauvin and Özak 2016).

Our analysis will use survey-level data to measure both the genetic diversity of a high school and its impact on the socio-economic outcomes of its students. High school genetic diversity intends to measure the general social environment and interactions faced by the student and represents a time in life in which personality and beliefs are flexible (Costa & McCrae, 1994; McCrae & Costa, 1999; McCrae et al. 2000).⁴ Our hypothesis is that these early-life interactions help to shape personality and other beliefs in a similar way to other forms of diversity.

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. (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. In adapting this theory to the current work, we argue that exposure to diversity in

³ 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.

⁴ Additional findings suggest that personality may be 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.

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 changes to personality; hence, 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) as well as greater revenue from sales, more customers, increased market share, and larger relative profits (Herring 2009). 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 is tested in Table 7.

In contrast to identity-based diversity measures, we argue that genetic diversity more accurately measures differences within and across groups. 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 amongst individuals through measured differences in individual genetic markers. 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; e.g., the "one-drop rule" in the United States.

Data and Empirical Methodology

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 WLS, genetic biomarker data were collected for both WLS graduates and selected siblings, which consists of roughly 95 single nucleotide polymorphisms (SNPs). We are able to aggregate this individual-level survey data to the high school-level in order to create population-level measures for schools.⁵ 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 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 locus. Summary statistics of this variable and all other variables used are found in Table 1, and detailed descriptions and sources of each variable are found in the Variable Appendix.

One point that needs to be mentioned is the selection of SNPs within the WLS is not 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, BMI, 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.⁶ As a check of this potential source of bias, robustness exercises in Table 8 control for each individual SNP.

⁵ To increase sample size, school-level gene frequencies are calculated from genetic data for both graduates and siblings. Appendix Section 5 replicates all estimations using a genetic diversity score that is calculated only from WLS graduates.

⁶ 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).

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 is constructed from ancestral shares and the country-level genetic diversity measure from AG. The correlation coefficient between the two measures 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.⁷ Genetic biomarker data were collected in the 2004 wave, a time period in which 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.⁸ 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 amongst 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, parent economic well-being, parent education, and the WLS graduate's IQ, which intends to proxy more

⁷ This is also a potential issue in AG, who do not use population representative genetic data.

⁸ In recent work, Domingue et al. (2017) show correcting for mortality selection does not alter genetic associations.

favorable early life environments and heritable benefits that are unrelated to the high school environment.⁹

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} . These controls are entered in a piecemeal fashion for most analyses. The set of “Individual” controls include measures for IQ, sex, birth year, WLS graduate’s cohort’s size, and an indicator for being in either Madison or Milwaukee (urban areas with more than 100K in population). The set of “Family” controls an index of family socioeconomic status in 1957 and mother and father’s education.

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 represented by Z_j in the above equation. 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 100km 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 estimating equation by I_c .

Of note, our estimation strategy differs from that of AG in that we hypothesize a positive linear effect of genetic diversity instead of non-linear hump-shaped effect. The proposed linear association is due to our focus on the relatively homogenous population of Wisconsin. The population is composed

⁹ As shown in Appendix Table A2, genetic diversity is unrelated to IQ for our base specification.

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.¹⁰ 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. This narrow range of genetic diversity is further reduced by our empirical strategy of estimating within county effects.¹¹ 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)-(5) piecemeal add individual, family, and ancestral-origin controls to the estimation of column (2); and column (6) 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.

Results

Socioeconomic Effects of Genetic Diversity

The first outcome of interest we explore is years of schooling. In regards to this measure of education, substantial selection is present in the WLS. 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.¹² Our sample consists solely of high school graduates, so we are testing the effect of genetic diversity on 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 (6), a one standard deviation

¹⁰ Roughly 75% of the population belongs to one of five ethnicities—British, Irish, Norwegian, German, and Polish, with 47% being derived solely from Germany.

¹¹ The within-county standard deviation of high school genetic diversity is 0.0138 compared to 0.0162 for the overall sample.

¹² Concerns of generalizability are addressed in part by the replication of Table 10.

increase in a high school's expected heterozygosity (0.01) results in an increase of roughly one month of schooling. 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.

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 that 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. WLS graduate's that went to high schools with more genetically diverse populations tended to seek out more cognitively challenging and more admired occupations.

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 (6) of Panel A), increasing genetic diversity by one standard deviation is associated with roughly a 2% increase in income.

A larger estimated coefficient is seen for the 1992 wave. This is likely due to differential earning profiles in later age, i.e., 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. 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.¹³

¹³ 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.

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.

An Examination of Mechanisms: Personality

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 in the beneficial effects of diversity.¹⁴ AG propose that the positive channel of genetic 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 diversity literature in Section 2 that shows diversity is associated with novel solutions. As a test of this idea, we test 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 amongst scientist are openness and extraversion (Feist 1998; Hirsh et al. 2009; Kaufman et al. 2016). Further studies also find beneficial effects from both extraversion and openness 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.¹⁵ Panel A tests our hypothesis of genetic diversity's role in influencing the personality trait of openness, 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. This reduction in

¹⁴ A number of studies have focused on the negative effects of genetic diversity on income, paying particular attention to the formation of ethnic groups and resulting ethnic conflict (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).

¹⁵ 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.

the coefficient of genetic diversity can be attributed to the inclusion of IQ, which has a strong relationship with openness (McCrae and John 1992).¹⁶

Panel B tests the association between the personality index of extraversion and our measure of high school genetic diversity. As with openness, a positive and statistically significant coefficient is estimated. Furthermore, the estimated relationship with extraversion is consistent across all specifications.

Given the role of both openness 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 significant effect is seen for all specifications, providing evidence for individual increases in creativity and divergent thinking from greater exposure to more diverse individuals.

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 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.

An Examination of Mechanisms: Task Diversity

In recent work, Depetris-Chauvin and Özak (2016) propose and provide evidence that genetic diversity has a positive association with the number of individualized tasks performed within a society prior to European colonization. The idea being that variation provides comparative advantages to differential

¹⁶ The indicator for Madison, which is the state capital and home to the University of Wisconsin, is also highly correlated with openness.

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 job diversity score is created.

Following the hypothesis of Depetris-Chauvin and Özak, a positive linear relationship is expected between our genetic diversity measure and task specialization, measured by our job diversity score. Table 7 presents evidence of this relationship. Table 7 follows a similar form as previous tables, but due to the job diversity score also being 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 the number of unique jobs its graduates find. This relationship holds for the previously mentioned specifications. Again, our unique and independent sample produces results consistent with prior evidence, providing support for the role of genetic diversity and a previously proposed mechanism.

Robustness

Table 8 examines the robustness of our previous findings to a set of ethnic and genetic controls. Column (1) replicates our baseline findings (col. (6) of previous tables). Column (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.¹⁷ In general, the inclusion of all individual specific genetic variants does not qualitatively change our estimated coefficients.¹⁸ In place of the gene-specific controls of column (5), column (6)

¹⁷ 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).

¹⁸ Conventional levels of statistical significance are lost, however, when considering family income in 1974 and all personality measures. This loss in statistical significance is not associated with a large change in p-values; rather, the p-value slightly increases above the 10% threshold.

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.

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 will 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 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.¹⁹

Table 10 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. Due to a lack of controls, all estimations (excluding Panel D) simply regress the parent's socioeconomic outcome on 1920's county genetic diversity. 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.

The baseline analysis shows a positive statistically significant effect of genetic diversity for all parental outcomes. In comparing the estimates of Panel A to those of the WLS graduates, a standard deviation increase in a county's genetic diversity is associated with roughly a 5.3 month increase in father's years of schooling and 1.6 months for mothers. This is in line with the estimates of WLS graduates in Table 2, which show that a standard deviation increase in high school diversity is associated with a 2.5 month increase in schooling for the simple within-county estimation and a 1.8 month increase

¹⁹ It is possible that those parents with high school graduates differ from those without high school graduates.

for our baseline estimation. In comparing Duncan occupational prestige, a respective one standard deviation increase in genetic diversity is associated with roughly 0.22 of a standard deviation increase in father's job prestige and a 0.07 standard deviation increase in the WLS graduate's job prestige. And finally, for occupational diversity a respective one standard deviation increase in genetic diversity leads to roughly identical increases for fathers and WLS graduates (~ 0.04 vs. 0.03).²⁰

A concern is that the WLS parents do not reside in the county where their children eventually graduate high school from in 1957. Given that the focus of the WLS is on the high school graduates and not their parents, no data exists that documents how long a particular family has lived in the 1957 county. To try and correct for measurement error associated with moving, Panels B-D make a few sample adjustments.

First, Panel B excludes the major urban areas of Wisconsin—Dane (Madison) and Milwaukee counties. As shown, excluding these two counties does not have a major effect on the estimated effect of genetic diversity. Panels C and D attempt to more accurately measure a county's propensity for migration. To do so, 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 five years past (Ruggles et al. 2015).²¹ Panel C excludes those counties in which 10% or more ($\sim 75^{\text{th}}$ percentile) of the county's population lived in a different county five years ago. Panel D simply includes our county-level fraction of migration to the estimation of Panel A. For both attempts at accounting for migration, we see increases in the coefficient of genetic diversity, suggesting classical measurement error that is somewhat alleviated by accounting for county migration.

Discussion

Using novel and independent populations that extend the previous literature, we find a statistically significant relationship between genetic diversity and socioeconomic outcomes that supports the hypothesis posited by AG. Our analysis provides a strong test of this hypothesis by considering micro-oriented aggregations that avoid potential confounders by country or ethnic level differences. The use

²⁰ 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.

²¹ 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.

of schools in place of countries as the relevant “populations” 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.

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 non-cognitive 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. Our findings are immune to most previous criticisms of Ashraf and Galor’s initial empirical evidence. 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.

References

- 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.
- 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 (2013b). Genetic diversity and the origins of cultural fragmentation. *American Economic Review, Papers and Proceedings*, 103(3), 528-533.
- Ashraf, Q., & Galor, O. (2013c). Response to comments made in a letter by Guedes et al. on "The Out of Africa Hypothesis, Human Genetic Diversity and Comparative Development." http://www.econ.brown.edu/fac/Oded_Galor/pdf/Ashraf-Galor%20Response.pdf
- Ashraf, Q., & Galor, O. (2013d). The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development: A response to commonly raised criticisms. <https://docs.google.com/file/d/0Bw0xeVVYEG5XSGxDTms4ZlICb2M/edit>
- 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.
- Callaway, E. (2012). Economics and genetics meet in uneasy union. *Nature*, 490(7419), 154-155.
- 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
- d'Alpoim Guedes, J., Bestor, T., Carrasco, D., et al. (2013). Is poverty in our genes? A critique of Ashraf and Galor, "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review* (Forthcoming). *Current Anthropology*, 54(1), 71-79.
- 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.
- 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.
- Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74, 208.
- 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.

- Gelman, A. (2013). Ethics and statistics: They'd rather be rigorous than right. *Chance* 26(2), 45-49.
- 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.
- 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.

Stevenson, B., & Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin Paradox. *Brookings Papers on Economic Activity*, Spring.

Stevenson, B., & Wolfers, J. (2013). Subjective well-being and income: Is there any evidence of satiation? *American Economic Review, Papers and Proceedings*, 101(3), 598-604.

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 began collecting data on 96 single nucleotide polymorphisms. 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 for the WLS graduate's first job based on rankings by Duncan (1961). From the WLS variable *ocsx1u2*.

Siegel Job Prestige: A measure of job prestige for the WLS graduate's first job based on rankings by Siegel (1971). 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 Hirfendahl 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 *brxdy*.

High School Size: The size of the WLS graduate's graduating class. From WLS variable *hssize*.

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*.

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 100km 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 | 338 | 0.3209 | 0.0162 | 0.2687 | 0.3519 |
| County Sample (from Ager and Brueckner 2016) | 70 | 7.8970 | 0.8204 | 6.2801 | 11.43 |
| Baseline Controls | | | | | |
| 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 |
| High School Class Size | 8675 | 180.00 | 133.31 | 11 | 482 |
| 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 |
| Madison/Milwaukee Indicator | 8675 | 0.18 | 0.38 | 0 | 1 |
| Historic School-Level Controls | | | | | |
| 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) | 7444 | 73.852 | 178.96 | 0 | 9999.99 |
| 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 |
| Job Diversity | 338 | 0.91 | 0.05 | 0.61 | 0.98 |

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.

Table 2. Genetic Diversity’s Effect on Education

| Dependent Variable: Years of Schooling | | | | | | |
|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genetic Diversity | 17.6377*** (3.9636) | 16.1662*** (3.4429) | 13.5578*** (3.3630) | 10.1605*** (2.6310) | 14.8302*** (3.3960) | 11.5417*** (2.7988) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7703 | 7703 | 7703 | 7703 | 7703 | 7703 |
| R Sqr. | 0.0100 | 0.0063 | 0.2242 | 0.1574 | 0.0213 | 0.3009 |

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. Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father’s years of schooling, and mother’s years of schooling. 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)-(6). 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) |
| Panel A. Duncan Prestige Score | | | | | | |
| Genetic Diversity | 7.9318*** (1.8473) | 5.5102*** (1.2145) | 3.9997*** (1.3087) | 3.4324*** (1.0005) | 5.1104*** (1.1239) | 3.1425*** (1.0637) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.0107 | 0.0039 | 0.1563 | 0.0894 | 0.0128 | 0.1992 |
| Panel B. Seigel Prestige Score | | | | | | |
| Genetic Diversity | 6.4212*** (1.5691) | 4.6992*** (1.1137) | 3.4590*** (1.1675) | 2.5498*** (0.8357) | 4.4322*** (1.0172) | 2.7275*** (0.9753) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.0069 | 0.0028 | 0.1483 | 0.0949 | 0.0115 | 0.1939 |
| Panel C. Occupational Education | | | | | | |
| Genetic Diversity | 5.9120*** (1.4148) | 5.5408*** (1.1059) | 4.7123*** (1.1426) | 3.0573*** (0.8063) | 5.1307*** (0.9968) | 3.7882*** (0.9102) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.0058 | 0.0038 | 0.1561 | 0.1235 | 0.0160 | 0.2196 |

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. Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father’s years of schooling, and mother’s years of schooling. 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)-(6). 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) |
| Panel A. 1974 | | | | | | |
| Genetic Diversity | 2.6595** (1.1152) | 2.8973** (1.2847) | 2.4163* (1.2266) | 2.1835* (1.2301) | 2.7556** (1.2411) | 2.0548* (1.1843) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 8270 | 8270 | 8270 | 8270 | 8270 | 8270 |
| R Sqr. | 0.0013 | 0.0011 | 0.0179 | 0.0116 | 0.0018 | 0.0241 |
| Panel B. 1992 | | | | | | |
| Genetic Diversity | 6.3307** (2.4720) | 5.7768** (2.8556) | 6.7742** (2.5819) | 4.5701* (2.7092) | 6.0929** (2.8879) | 6.7962*** (2.5535) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7443 | 7443 | 7443 | 7443 | 7443 | 7443 |
| R Sqr. | 0.0009 | 0.0005 | 0.0509 | 0.0057 | 0.0015 | 0.0520 |

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. Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. 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)-(6). 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) |
| Panel A. Openness to Experience | | | | | | |
| Genetic Diversity | 26.4792*** (5.6082) | 19.0542*** (4.5326) | 10.5875* (5.5806) | 12.4703*** (3.9486) | 18.4017*** (4.3346) | 8.5416* (5.0169) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7148 | 7148 | 7148 | 7148 | 7148 | 7148 |
| R Sqr. | 0.0058 | 0.0022 | 0.0768 | 0.0510 | 0.0060 | 0.1005 |
| Panel B. Extraversion | | | | | | |
| Genetic Diversity | 13.8282** (5.8665) | 14.2696** (6.5161) | 13.5955** (6.5688) | 10.9299* (6.3029) | 12.3704* (6.3031) | 11.0301* (6.3743) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7159 | 7159 | 7159 | 7159 | 7159 | 7159 |
| R Sqr. | 0.0013 | 0.0010 | 0.0041 | 0.0089 | 0.0030 | 0.0148 |
| Panel C. p.c. of Openness and Extraversion | | | | | | |
| Genetic Diversity | 6.1735*** (1.5232) | 5.0368*** (1.4111) | 3.5951** (1.4869) | 3.5316*** (1.2753) | 4.6726*** (1.3502) | 2.9181** (1.3818) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7145 | 7145 | 7145 | 7145 | 7145 | 7145 |
| R Sqr. | 0.0049 | 0.0024 | 0.0299 | 0.0384 | 0.0061 | 0.0539 |

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. Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. 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)-(6). 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) |
| Panel A. Conscientiousness | | | | | | |
| Genetic Diversity | -2.7853 (3.3642) | -0.7780 (4.0783) | -1.3895 (4.0154) | -0.2638 (4.0004) | 0.1706 (4.3224) | -0.4516 (4.1623) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7170 | 7170 | 7170 | 7170 | 7170 | 7170 |
| R Sqr. | 0.0001 | 0.0000 | 0.0007 | 0.0003 | 0.0007 | 0.0018 |
| Panel B. Agreeableness | | | | | | |
| Genetic Diversity | -7.0674* (4.0366) | -1.9628 (4.4256) | -0.2853 (4.7672) | -1.3920 (4.6514) | -0.6895 (4.0611) | 0.2785 (4.6668) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7167 | 7167 | 7167 | 7167 | 7167 | 7167 |
| R Sqr. | 0.0005 | 0.0000 | 0.0582 | 0.0018 | 0.0012 | 0.0596 |
| Panel C. Neuroticism | | | | | | |
| Genetic Diversity | -8.8302*** (3.1307) | -4.6416 (3.8957) | -4.6592 (3.8180) | -2.6619 (3.9730) | -4.5488 (3.9087) | -3.9698 (3.7616) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual | N | N | Y | N | N | Y |
| Family | N | N | N | Y | N | Y |
| Historic | N | N | N | N | Y | Y |
| Observations | 7156 | 7156 | 7156 | 7156 | 7156 | 7156 |
| R Sqr. | 0.0006 | 0.0001 | 0.0351 | 0.0069 | 0.0009 | 0.0380 |

Summary & Notes: Additional personality indices are tested in Table 6. Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father’s years of schooling, and mother’s years of schooling. 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)-(6). 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: Occupational Diversity | | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Genetic Diversity | 2.0941*** (0.1857) | 1.9191*** (0.1627) | 1.3092*** (0.1555) | 1.7368*** (0.1589) | 1.9272*** (0.1590) | 1.2584*** (0.1383) |
| County Fixed Effects | N | Y | Y | Y | Y | Y |
| Individual/School Controls | N | N | Y | N | N | Y |
| Family Controls | N | N | N | Y | N | Y |
| Historic Controls | N | N | N | N | Y | Y |
| Observations | 338 | 338 | 338 | 338 | 338 | 338 |
| R Sqr. | 0.4051 | 0.3591 | 0.4989 | 0.4130 | 0.3869 | 0.5498 |

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). Individual/School controls include mean high school IQ, mean birth year, fraction of female students, high school size class size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include the high school mean of socioeconomic status in 1957, father’s years of schooling, and mother’s years of schooling. 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)-(6). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by *, **, and ***.

Table 8A. Robustness: Education and Job Prestige

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Panel A. Years of Schooling | | | | | | | | |
| Genetic Diversity | 11.5417*** (2.7988) | 11.5586*** (2.8043) | 11.5577*** (2.7619) | 11.5586*** (2.8043) | 11.1691*** (2.8936) | 11.5171*** (2.7963) | 11.2629*** (2.8807) | 11.5367*** (2.8027) |
| Observations | 7703 | 7703 | 7703 | 7703 | 7703 | 7703 | 7703 | 7703 |
| R Sqr. | 0.3009 | 0.3078 | 0.3009 | 0.3078 | 0.3193 | 0.3009 | 0.3260 | 0.3078 |
| Panel B. Std. Duncan Prestige | | | | | | | | |
| Genetic Diversity | 3.1425*** (1.0637) | 3.0723*** (1.0535) | 2.9601*** (1.0660) | 3.0723*** (1.0535) | 2.8276*** (1.0415) | 3.1685*** (1.0577) | 2.7542** (1.0417) | 3.0975*** (1.0482) |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.1992 | 0.2049 | 0.1994 | 0.2049 | 0.2180 | 0.1993 | 0.2240 | 0.2050 |
| Panel C. Std. Siegel Prestige | | | | | | | | |
| Genetic Diversity | 2.7275*** (0.9753) | 2.6516*** (0.9855) | 2.5827*** (0.9677) | 2.6516*** (0.9855) | 2.5043*** (0.9016) | 2.7749*** (0.9713) | 2.4292** (0.9172) | 2.6975*** (0.9821) |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.1939 | 0.1987 | 0.1940 | 0.1987 | 0.2156 | 0.1942 | 0.2206 | 0.1991 |
| Panel D. Std. Occupational Education | | | | | | | | |
| Genetic Diversity | 3.7882*** (0.9102) | 3.8964*** (0.9095) | 3.7996*** (0.8991) | 3.8964*** (0.9095) | 3.5684*** (0.8908) | 3.7991*** (0.9117) | 3.6615*** (0.8941) | 3.9087*** (0.9107) |
| Observations | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 | 8098 |
| R Sqr. | 0.2196 | 0.2252 | 0.2196 | 0.2252 | 0.2406 | 0.2196 | 0.2463 | 0.2252 |
| 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 8 examines the robustness of the previously estimated effect of genetic diversity. Column (1) replicates the baseline estimation (col. (6) in above tables). 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). Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. 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)-(6). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by *, **, and ***.

Table 8B. Robustness: Income and Personality

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--|-----------|----------|----------|----------|----------|-----------|----------|----------|
| Panel A. ln Family Income, 1974 | | | | | | | | |
| Genetic Diversity | 2.0548* | 1.9758 | 1.9145* | 1.9758 | 1.9618 | 2.0592* | 1.8676 | 1.9795 |
| | (1.1843) | (1.1930) | (1.1452) | (1.1930) | (1.1922) | (1.1808) | (1.1961) | (1.1900) |
| Observations | 8270 | 8270 | 8270 | 8270 | 8270 | 8270 | 8270 | 8270 |
| R Sqr. | 0.0241 | 0.0322 | 0.0242 | 0.0322 | 0.0424 | 0.0241 | 0.0509 | 0.0322 |
| Panel B. ln Family Income, 1992 | | | | | | | | |
| Genetic Diversity | 6.7962*** | 6.5392** | 6.3451** | 6.5392** | 5.4366** | 6.7986*** | 5.2202** | 6.5484** |
| | (2.5535) | (2.6311) | (2.6522) | (2.6311) | (2.4801) | (2.5519) | (2.5320) | (2.6302) |
| Observations | 7443 | 7443 | 7443 | 7443 | 7443 | 7443 | 7443 | 7443 |
| R Sqr. | 0.0520 | 0.0577 | 0.0521 | 0.0577 | 0.0803 | 0.0520 | 0.0861 | 0.0577 |
| Panel C. Openness to Experience | | | | | | | | |
| Genetic Diversity | 8.5416* | 8.4021 | 8.0581 | 8.4021 | 6.2780 | 8.5077* | 6.0689 | 8.3899 |
| | (5.0169) | (5.1642) | (5.1855) | (5.1642) | (4.9560) | (4.9823) | (5.1781) | (5.1384) |
| Observations | 7148 | 7148 | 7148 | 7148 | 7148 | 7148 | 7148 | 7148 |
| R Sqr. | 0.1005 | 0.1082 | 0.1005 | 0.1082 | 0.1255 | 0.1005 | 0.1328 | 0.1082 |
| Panel D. Extraversion | | | | | | | | |
| Genetic Diversity | 11.0301* | 11.0514* | 10.8324 | 11.0514* | 8.0204 | 11.0399* | 7.7389 | 11.0566* |
| | (6.3743) | (6.6075) | (6.6593) | (6.6075) | (6.3758) | (6.3937) | (6.5713) | (6.6266) |
| Observations | 7159 | 7159 | 7159 | 7159 | 7159 | 7159 | 7159 | 7159 |
| R Sqr. | 0.0148 | 0.0211 | 0.0148 | 0.0211 | 0.0398 | 0.0148 | 0.0466 | 0.0211 |
| Panel E. Principal Component of Open. and Ext. | | | | | | | | |
| Genetic Diversity | 2.9181** | 2.8995** | 2.8164* | 2.8995** | 2.1406 | 2.9141** | 2.0667 | 2.8984** |
| | (1.3818) | (1.4362) | (1.4482) | (1.4362) | (1.3720) | (1.3800) | (1.4373) | (1.4356) |
| Observations | 7145 | 7145 | 7145 | 7145 | 7145 | 7145 | 7145 | 7145 |
| R Sqr. | 0.0539 | 0.0613 | 0.0540 | 0.0613 | 0.0787 | 0.0539 | 0.0858 | 0.0613 |
| 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 8 examines the robustness of the previously estimated effect of genetic diversity. Column (1) replicates the baseline estimation (col. (6) in above tables). 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). Individual controls include IQ, sex, birth year, high school size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include an index of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. 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)-(6). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by *, **, and ***.

Table 9. Genetic Diversity and Task Diversity: Robustness

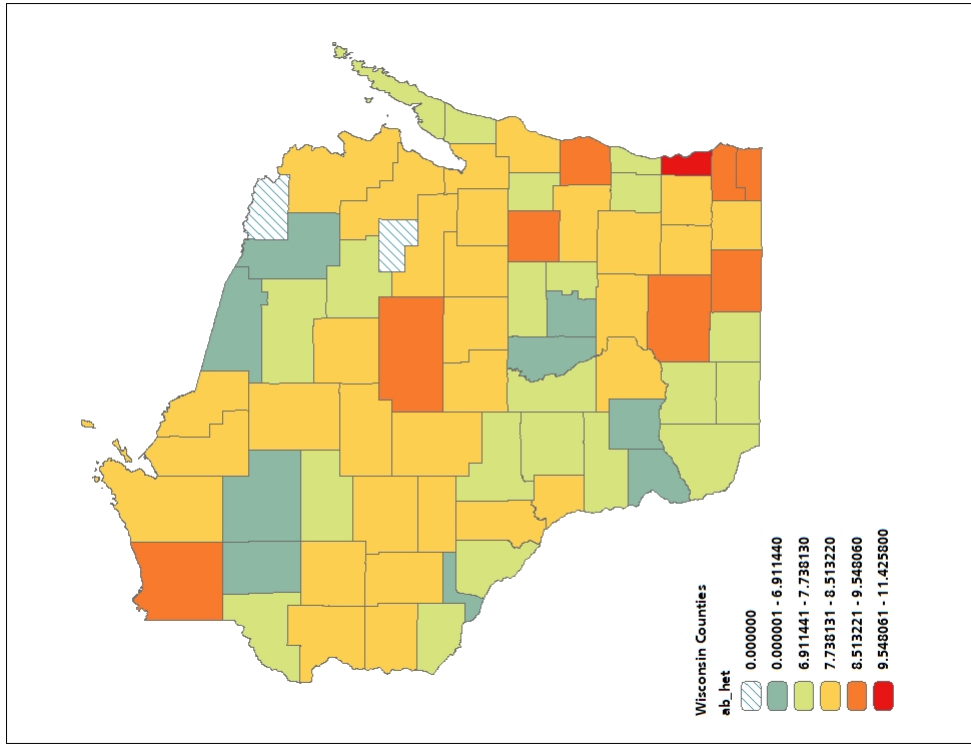
| Dependent Variable: Occupational Diversity | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) |
| Genetic Diversity | 1.0895*** (0.1423) | 2.5456*** (0.2289) | 1.2124*** (0.1441) | 2.3535*** (0.2614) |
| County Fixed 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 | 338 | 338 | 338 | 338 |
| R Sqr. | 0.6705 | 0.8138 | 0.5771 | 0.8744 |

Summary & Notes: Table 9 mirrors the robustness exercises of Table 8 except that the unit of observation is the high school. Individual/School controls include mean high school IQ, mean birth year, fraction of female students, high school size class size, and an indicator for being in either Madison or Milwaukee—the two major urban areas of Wisconsin. Family controls include the high school mean of socioeconomic status in 1957, father's years of schooling, and mother's years of schooling. 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)-(6). Standard errors are clustered by county. Statistical significance at the 10%, 5%, and 1% levels is respectively denoted by *, **, and ***.

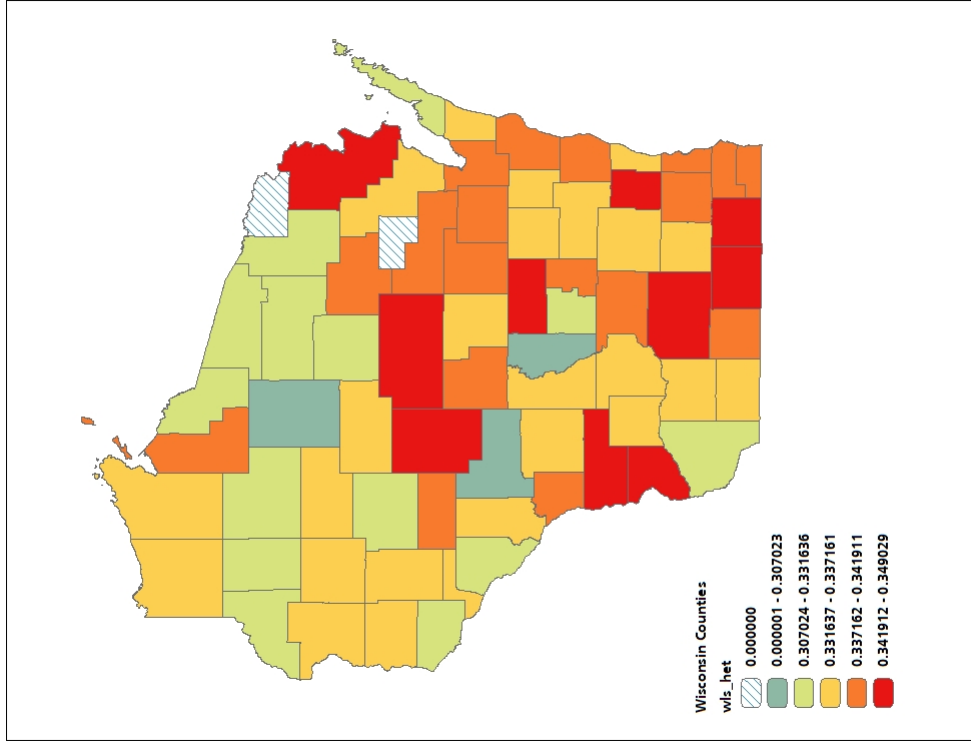
Table 10. County Replication with WLS Parents

| Dependent Variable: | SES | | Father's | | Mother's | | Father's | | |
|--|-------------------------------------|-----------------------------------|------------------------------------|------------------------------------|-----------------------------------|------------|---------------|---------------|--|
| | 1957 (1) | Job Prestige (2) | YOS (3) | YOS (4) | YOS (4) | YOS (4) | Job Diversity | Job Diversity | |
| Panel A. Base Analysis | | | | | | | | | |
| County Genetic Diversity | 1.7279*** (0.1473) [0.1776] | 0.1591*** (0.0083) [0.0112] | 0.3196*** (0.0322) [0.0524] | 0.0947*** (0.0255) [0.0357] | 0.0536*** (0.0123) [0.0170] | | | | |
| Observations | 8690 | 7682 | 8690 | 8690 | 70 | | | | |
| R Sqr. | 0.0501 | 0.0550 | 0.0170 | 0.0022 | 0.1490 | | | | |
| Panel B. Excluding Dane and Milwaukee Counties | | | | | | | | | |
| County Genetic Diversity | 1.7603*** (0.4465) [0.4454] | 0.1763*** (0.0287) [0.0256] | 0.2995** (0.1185) [0.1417] | 0.0539 (0.1108) [0.0989] | 0.0593*** (0.0158) [0.0240] | | | | |
| Observations | 6539 | 5745 | 6539 | 6539 | 68 | | | | |
| R Sqr. | 0.0118 | 0.0156 | 0.0034 | 0.0002 | 0.1303 | | | | |
| Panel C. Excluding High Immigration Counties | | | | | | | | | |
| County Genetic Diversity | 1.8099*** (0.1039) [0.1160] | 0.1658*** (0.0073) [0.0070] | 0.3450*** (0.0271) [0.0366] | 0.1202*** (0.0240) [0.0272] | 0.0630*** (0.0145) [0.0127] | | | | |
| Observations | 7499 | 6613 | 7499 | 7499 | 57 | | | | |
| R Sqr. | 0.0574 | 0.0646 | 0.0217 | 0.0039 | 0.1942 | | | | |
| Panel D. Controlling for County Immigration | | | | | | | | | |
| County Genetic Diversity | 2.2254*** (0.1517) [0.1308] | 0.1913*** (0.0079) [0.0070] | 0.4485*** (0.0479) [0.0385] | 0.2002*** (0.0362) [0.0320] | 0.0633*** (0.0125) [0.0134] | | | | |
| County Frac. of Immigrants, 1940 | 53.2400*** (12.0030) [6.7648] | 3.4317*** (0.5843) [0.3880] | 13.7871*** (4.1823) [1.7530] | 11.2915*** (3.2300) [1.8271] | 1.3191*** (0.3578) [0.2699] | | | | |
| Observations | 8690 | 7682 | 8690 | 8690 | 70 | | | | |
| R Sqr. | 0.0612 | 0.0625 | 0.0262 | 0.0114 | 0.2205 | | | | |

Summary & Notes: Table 10 explores the relationship between 1920 county-level genetic diversity (from Ager and Brueckner) 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 of individuals that were not selected by educational status. Panel A estimates the simple relationship; Panel B drops the two large urban areas of Wisconsin; Panel C restricts the sample by immigration; and Panel D controls for county immigration. County-level immigration fractions come from the full 1940 census and measure the fraction of individuals who reported living in a different county in the last 5 years. Standard errors clustered by county are given in parentheses; spatially adjusted standard errors (100km) are given in brackets. 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 ($p < 0.01$).

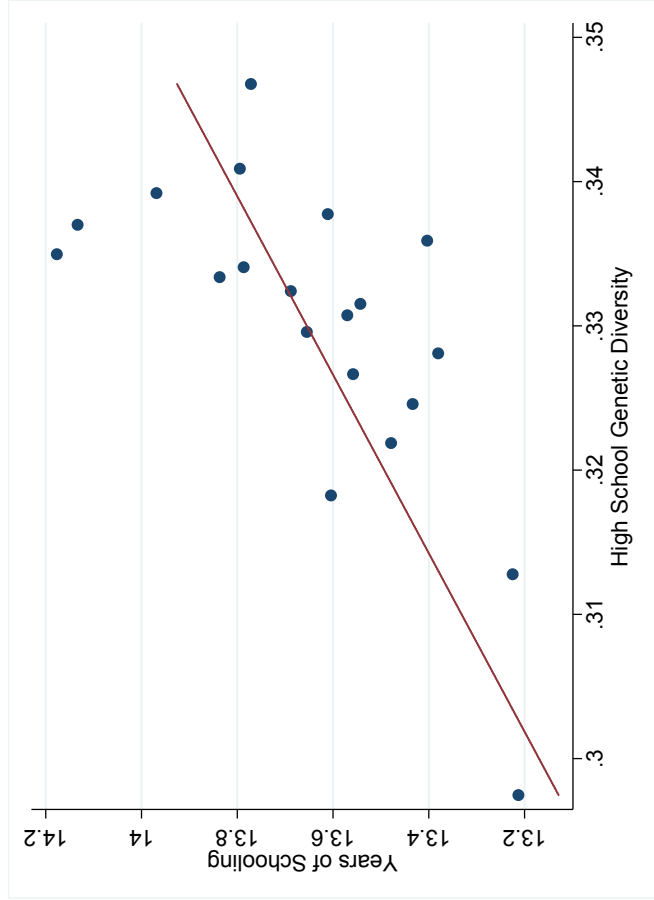
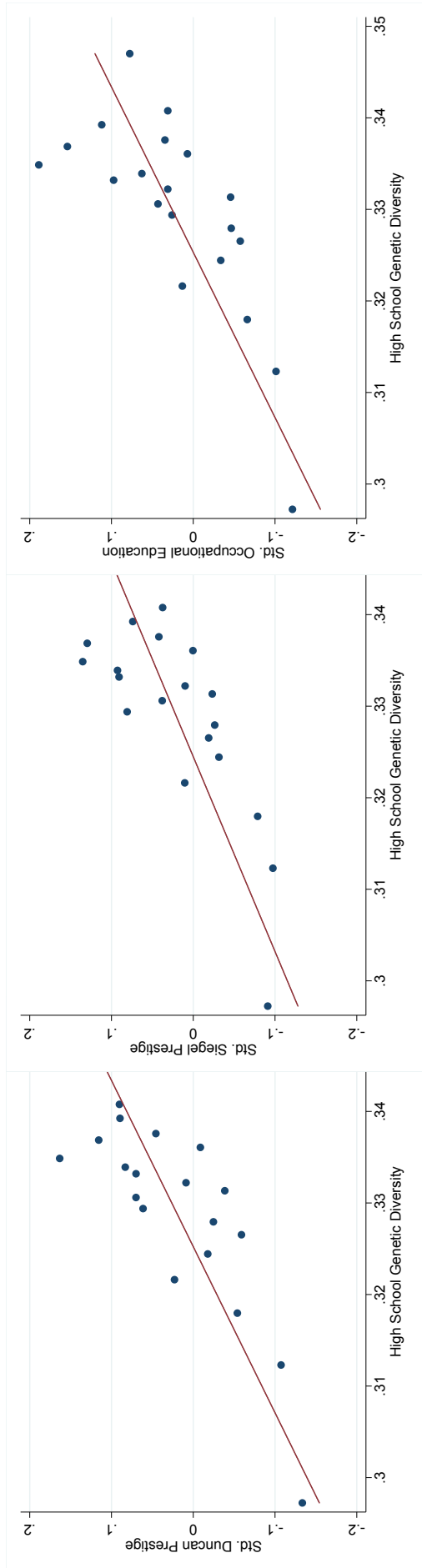


Figure 2. Years of Schooling and Genetic Diversity

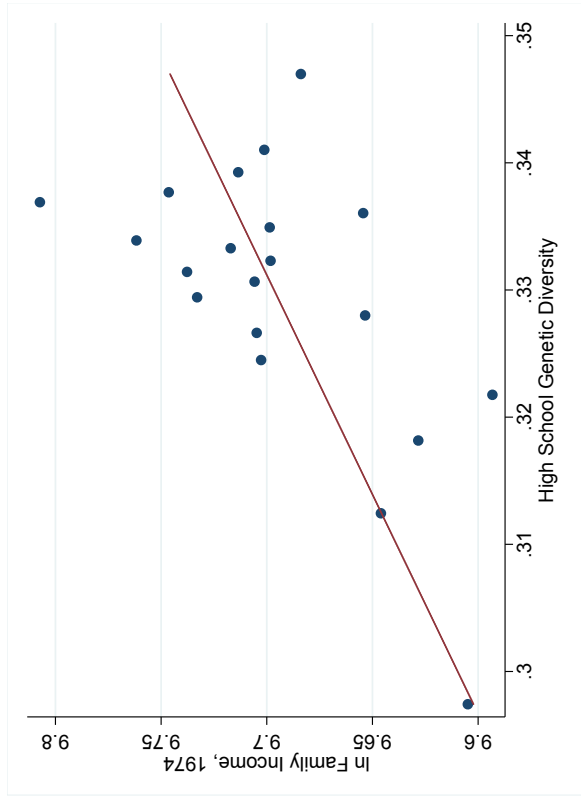
Summary & Notes: This figure plots the within county relationship between years of schooling and high school genetic diversity for 20 equally sized bins.



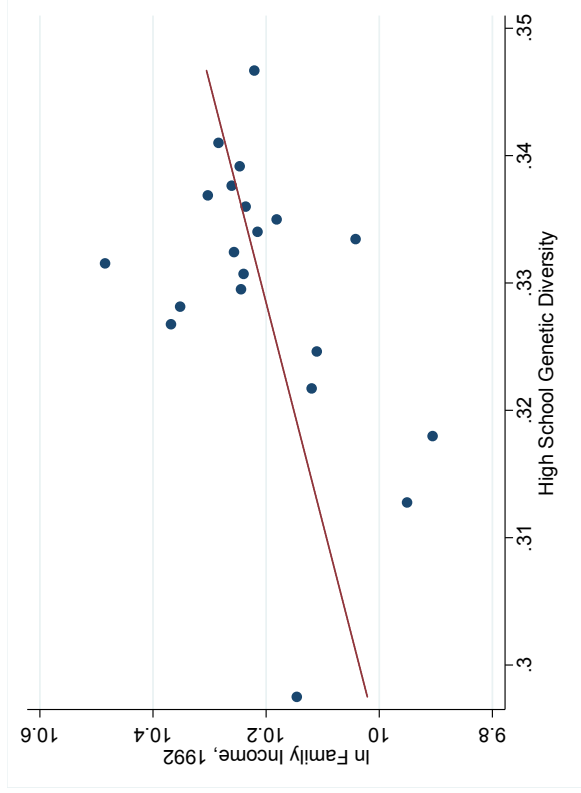
(a) Duncan
 (b) Siegel
 (c) Occ. Edu.

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.



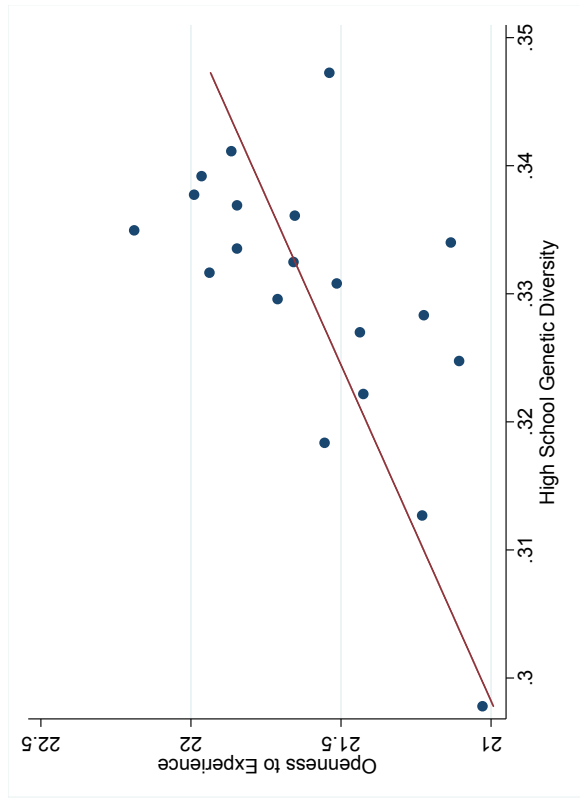
(a) ln Family Income, 1974



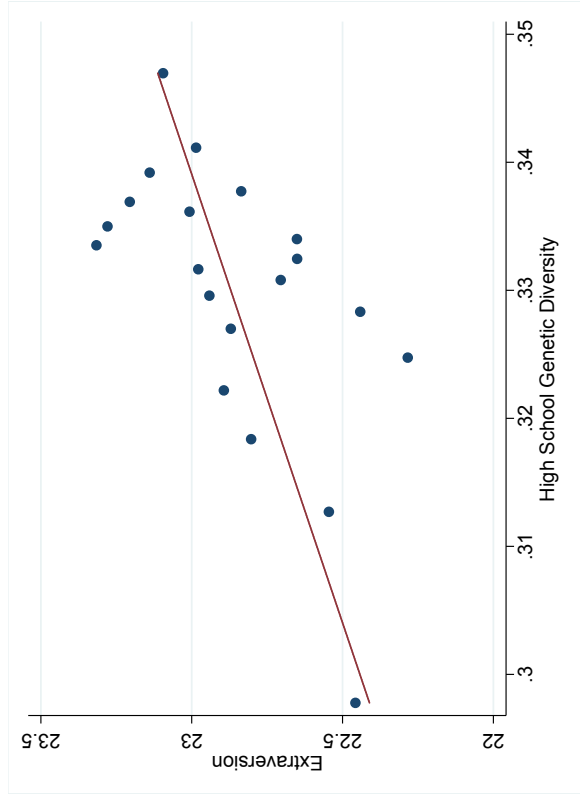
(b) ln Family Income, 1992

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



(b) Extraversion

Figure 5. Personality and Genetic Diversity

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