Who Makes the Grade? A Gene-Environment Analysis of the Mechanisms of Educational Attainment in the 21st Century

David B. Braudt[1,2]

ABSTRACT

Although the role of social factors in the creation and maintenance of disparities in educational attainment has been the focus of decades of research, and nearly as much policy debate, doubt still remains as to the exact nature of the role of social-environmental factors due to a general inability to account for the association between genetic differences between individuals and educational attainment. Leveraging molecular genetic data from the National Longitudinal Study of Adolescent to Adult Health, this study investigates the differential role of social-environment factors net of genetics on educational attainment between three levels of parental education. Through a series of polygenic score analyses, genome-wide complex trait analyses, and behavior genetic analyses the results of this study depict a robust pattern in which social-environmental factors account for an increasing proportion of the total variance in educational attainment when individuals are born into more advantaged circumstances.

1 – Department of Sociology, University of North Carolina at Chapel Hill
2 – Carolina Population Center, University of North Carolina at Chapel Hill

INTRODUCTION

A well-established finding in social stratification research is the central role of educational attainment in both status attainment and social stratification processes. Starting as early as Blau and Duncan's (1967) path analysis formalizing the relationship between educational social origins and destinations and continuing through the various extensions of similar research by the Wisconsin school (e.g., Sewell, Haller, and Portes 1969) and others (e.g., Breen and Goldthorpe 1997), educational attainment has been shown to be one of, if not the most important factor in social reproduction in modern industrialized societies. A long standing limitation in much of the extant research in this area is the inability to account for the inherent endogeneity in the relationship between social origins and destinations due to the simultaneous transfer of social resources and genetics from parents to children. Fortunately, advances the measurement molecular genetics and the incorporation of these measures in national representative samples is now allowing researchers to address some of these limitations.

While there has been an increasing interest in the role of genetics in the processes of educational attainment (e.g., Conley et al. 2015; Domingue et al. 2015; Liu 2018), more often than not sociologists continue to conceptualize social reproduction in educational outcomes in purely socio-environmental terms. Failing to account for genetic effects can lead to overestimates of the relationship between educational attainment and various outcomes (for an example of the exact size and nature of the biases associated with omitting genetic factors in the association between education and health see Boardman, Domingue, and Daw 2015). Similarly, genetic analyses that only consider genetic effects may be biased in their reports of associations between social origins and destinations. However, by framing investigations of educational attainment in terms of how genetics and the socio-environmental factors interact in the complex processes connecting social origins and destinations, scientists can potentially address the shortcomings of both a purely genetic and/or purely social approach to the study of the links between social origins and destinations.

The integration of genetic and social science approaches to the study of complex outcomes such as educational attainment, i.e. sociogenomics, can be conducted on across variety of ontological levels. At the micro-level, sociogenomic methods, such as the estimation of individual-level genetic propensities for an outcome of interest, can be used to investigate the role of genetics in individual-level processes. At the meso-level, differences in the average estimated effect of genetics on social behaviors and traits (or vice versa)

between groups, stratified along important distinctions in social and environmental circumstances and/or experiences, can provide insight into how, or if, genes and the environment interact to produce the diversity of outcomes observed in society. The use of sociogenomic analyses and measures at the macro-level follows the same intuition as analyses at the meso-level, but with groups being defined as entire societies.

As alluded to above, an important advancement in sociogenomic research is the incorporation of gene-environment interactions (GxEs). While this term is evocative of a multiplicative interaction term common in generalized linear regression models, GxEs are not limited to purely multiplicative associations. Instead, GxEs represent the many ways in which a genetic effect is dependent on the environment and vice versa (Shanahan, Michael J. and Boardman, Jason D. 2009).

Not surprisingly, GxEs can occur at the micro-, meso-, and macro-level. At an individual-level, differences in the environments that people face in their daily lives (e.g., peer groups and family settings) are abundant and an increasing number of studies have begun to investigate the role of individual-level GxE in educational attainment (see for example Liu 2018; Conley et al. 2015; Belsky et al. 2016; Guo and Stern 2002). At the meso-level certain groups are more likely to encounter different sets or combinations of environmental effects (e.g., systematic discrimination and unequal access to resources) which may lead to different types of GxE within each group. Similarly, at the macro-level, differences in policies and/or government structures may produce different systems of gene-environment interactions in processes of social mobility, status attainment, and education between societies (Adkins and Guo 2008; Adkins and Vaisey 2009). However, many investigations of GxEs have either relied on a multipliciative operationalization (e.g., Liu 2018; Guo and Stern 2002) or omit explicit tests of GxE entirely (e.g. Nielsen 2006; Nielsen and Roos 2015).

A common misunderstanding of GxEs is that they represent conceptual ideas rather than statistical tests (Shanahan, Michael J. and Boardman, Jason D. 2009:229). Consequently, a single statistical test is likely insufficient to capture the non-random patterning of genetic and environmental factors associated with a GxE. Instead a consistent pattern of findings across statistical methods and related measures is needed to establish evidence of a GxE.

There are two general approaches to sociogenomic analysis: (1) quantitative genetic analysis and (2) molecular genetic analysis. Quantitative genetics relies on known population averages in genetic similarities

between biologically related individuals. The most common type of quantitative genetic analysis restricts analyses to monozygotic and dizygotic twins to decompose the variance in an outcome into genetic, common environmental, and unique environmental factors. Building on the twin design, the extended twin family design (ETFD) extends quantitative genetic analyses to include non-twin siblings (i.e., non-twin full-siblings, half-siblings, cousins, and siblings who aren't biologically related but who were raised in the same household). Molecular genetic analyses, on the other hand, rely on observed differences in genetic variation in allele frequencies (i.e., the number of G, C, T, or A nucleotide bases an individual has across both sets of chromosomes) at a specific single-nucleotide polymorphism, or SNP (pronounced "snip"), for hundreds of thousands, and in some cases millions, of SNPs for a single individual. With the decreased cost of individual genotyping, the inclusion of molecular genetic data has become increasingly common in nationally representative samples (e.g, the National Longitudinal Study of Adolescent to Adult Health). Leveraging the immense amount of data contained in molecular genetic datasets is complicated, but two types of analyses in particular are able to aggregate the large swaths of information in such datasets into more manageable and interpretable variables: genome-wide complex trait analysis (Domingue et al. 2016; Yang et al. 2011, 2017) and polygenic scores (Belsky and Israel 2014; Conley 2016; Dudbridge 2016; Plomin, Haworth, and Davis 2009). A more detailed discussion of quantitative and molecular genetic analysis and data is presented later in this article; but, for present purposes it is important to note that quantitative and molecular genetic approaches to sociogenomic research represent complementary strategies in investigations of the role of genetics in status attainment and social stratification processes (Diewald 2015:375).

This study applies the framework of GxEs to evaluate the degree to which social reproduction versus genetics can account for the observed variance in educational attainment across three social strata, defined as discrete levels of parental educational attainment, using data from a nationally representative sample of individuals completing their education during the first decade of the 21st century in the U.S. GxEs in educational attainment are investigated using quantitative genetic analysis (i.e. the extended twin family design) and molecular genetic analyses (i.e. genome-wide complex trait analyses and polygenic score analyses) in samples stratified by parental educational attainment.

EDUCATIONAL ATTAINMENT, STRATIFICATION, AND GENETICS

The study of social stratification, i.e. "the differential ranking of the human individuals who compose a given social system and their treatment as superior and inferior relative to one another in certain socially important respects"(Parsons 1940:841), has been, and continues to be, a central tenant of sociological inquiry. Furthermore, since at least the publication of Blau and Duncan's (1967) "The American Occupational Structure," educational attainment has held a central place in the study of status attainment (i.e., the differential treatment of individuals as superior/inferior based on relative social position referenced by Parson).Following Blau and Duncan's status attainment model, a proverbial deluge of research into potential mechanisms linking social origins and destinations in status attainment, and by association educational processes, established education as one of, if not the, fundamental aspect of stratification processes in modern industrialized societies (Sewell et al. 1969)

In the broadest sense, individual differences in educational attainment can be divided into two sources: genetics and socio-environmental factors; and failing to account for either source of variance will lead to biased estimates of the association between social origins and destinations. Diewald et al. (2015:374) summarizes the argument succinctly, "parents not only pass on resources and experiences to their children, but also their genetic predispositions… [thus] inequalities exist between individuals from birth on, not only in their social origins but also in their genetic endowments." For complex traits such as educational attainment, the genetic predispositions passed from parents to children don't follow simple Mendelian inheritance patterns of a single dominant or recessive gene, but, instead are influenced by a large number of genes from across the genome (Chabris et al. 2015; Dudbridge 2016; Plomin et al. 2009:200). While sociologists tend to focus almost exclusively on socio-environmental factors, a growing contingent of sociological research has begun to explore the role genetics in outcomes central to social stratification processes such as educational attainment (Conley et al. 2015; Domingue et al. 2015; Liu 2018; Nielsen 2006; Nielsen and Roos 2015; Shanahan et al. 2008) cognitive ability (Guo and Stearns 2002), and social status (Belsky et al. 2016). Additionally, theoretical frameworks linking genetics, social/institutional structures, and general stratification processes have become more prominent (Adkins and Guo 2008; Adkins and Vaisey 2009; Diewald et al. 2015; Nielsen 2016; Shanahan, Michael J. and Boardman, Jason D. 2009).

HOW THE STUDY OF GENETICS CAN ILLUMINATE THE STRUCTURE OF SOCIAL SYSTEMS

A common misconception concerning sociogenomic analyses is that the unit of analysis is restricted to individuals, when in fact sociogenomic methods and measures can be used to study the interplay between genetics and social structures for individuals, small to large groups, or entire social systems. For example, although often assumed to indicate individual-level associations, estimates of heritability (i.e. the proportion of variance in an outcome attributable to genetics) is a population-level measure of the average association between genetics and the outcome being studied, within a given society at particular time (Plomin et al. 2013). Consequently, estimates of heritability, can and do change in the face of different social structures. Due to the sensitivity of heritability to differences in social, institutional, and temporal configurations affecting an outcome, estimates of heritability can be used as indicators of meso- and macro-level social structures (Adkins and Guo 2008; Adkins and Vaisey 2009; Diewald et al. 2015; Nielsen 2016).

Following the definition of social stratification from Talcot Parson's (1940:841) mentioned above, Adkin's and colleagues (Adkins and Guo 2008; Adkins and Vaisey 2009) develop a unified stratification theory that incorporates estimates of heritability into traditional sociological theorizing of the status attainment processes. At its core, the unified stratification theory suggests that the strength of genetic influences e.g. estimates of heritability, are primarily moderated by two properties of social systems: (1) levels of resource inequality and (2) social mobility. In social systems defined by high levels of resource inequality and low social mobility genetic influences on outcomes related to status attainment processes will be low, if present at all. Conversely, when resource inequality is low and social mobility is high, the unified stratification theory posits that estimates of genetic effects on outcomes of central importance to status attainment will be high. Stated differently, these two postulates suggest that a social system can be defined as "open" or "closed" based on estimates of heritability for outcomes pertinent to status attainment (Diewald et al. 2015). For example, in modern

industrialized societies a higher heritability of educational attainment would indicate that individual level characteristics influenced by genetics (e.g., ability, intelligence, and conscientiousness) are of relatively greater importance in the social and institutional systems surrounding educational processes while lower estimates of the heritability of educational attainment would indicate social stratification, i.e. socio-environmental factors regulating access to resources beneficial to educational attainment, is of relative greater importance in determining the quantity and quality of education individuals obtain.

Additionally, while the unified stratification theory, and previous research (Nielsen and Roos 2015) emphasize inter-societal comparisons, the same theory and analytic design can be used to investigate intra-societal differences based on social origins. Specifically, "while the genome predisposes individuals to various developmental paths and vulnerabilities, the expression of these genetic potentialities is to some degree contingent on the social circumstances they [experience]. From this logic it follows that when social circumstances of individuals are widely divergent (i.e. inequality is great), the relative importance of the genome on ability formation will be diminished and that of social background will be increased" (Adkins and Guo 2008:242).

This line of reasoning leads to a series of testable hypotheses with respect to estimates of the heritability of educational attainment between social strata, defined by parental educational attainment. Form the unified stratification theory, we known that if social reproduction is the dominant mechanism in educational attainment, then estimates of heritability will be equal, or close to zero. Thus, the first hypothesis can be formulated as:

H1: For social strata in which social reproduction is the dominant mechanism linking parental educational attainment and children's educational attainment, estimates of the heritability of educational attainment will be close to zero (weak social closure), or equal to zero (strong social closure).

Conversely, the unified stratification theory suggests that if social mobility is the dominant mechanism in educational attainment, then estimates of heritability will be large. This leads to the second hypothesis:

H2: For social strata in which social mobility is the dominant mechanism linking parental educational attainment and children's educational attainment, estimates of the heritability of educational attainment will be close to one (weak social mobility), or equal to one (strong social mobility).

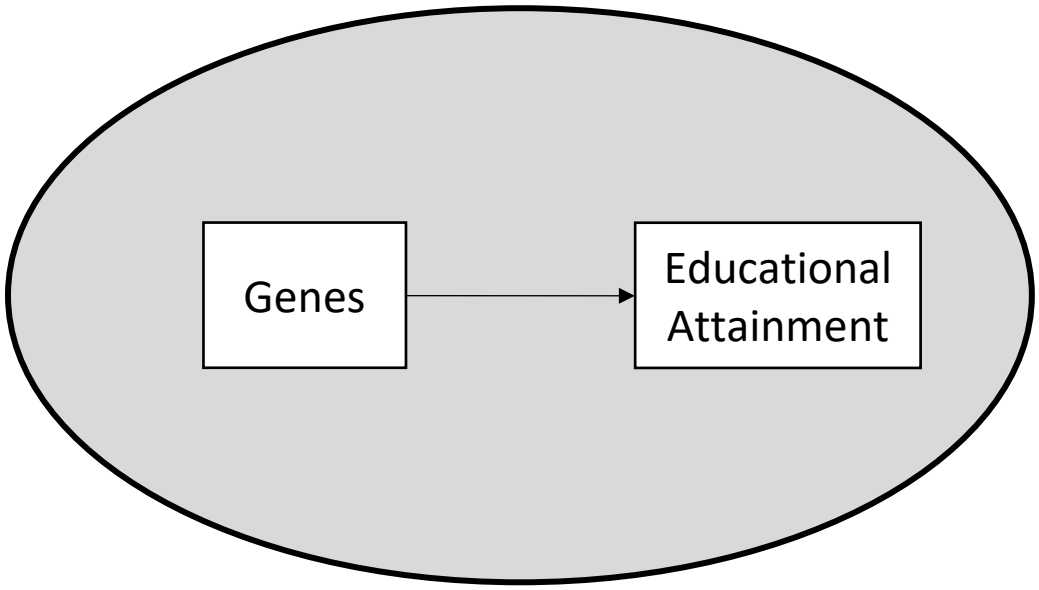
The unified stratification theory offers one perspective through which estimates of genetic influence on status attainment can be conceptualized, but it is not without limitation. The properties of heritability estimates that make them useful as relative descriptors of the degree of social reproduction, i.e. social closure, in a society, also obscure direct connections to potential policy decisions. With regards to educational attainment, (Morton 1974:327) argues that, "one would be quite unjustified in claiming that heritability is relevant to educational strategy." And while Morton is correct that by themselves heritability estimates of educational attainment are uninformative for micro-level educational strategy (e.g., pedagogical practices within the classroom), as comparative measures they provide a window through which researchers can observe the differential operation of higher order mechanisms in educational processes between groups who experience unequal levels of resource allocation. Gene-environment interactions provide a common framework in which intra-societal comparisons of the socio-environmental conditioning of genetic effects can be evaluated (Diewald et al. 2015; Nielsen 2006; Nielsen and Roos 2015).

GENE-ENVIRONMENT INTERACTIONS IN EDUCATIONAL ATTAINMENT

Genetic effects rarely represent immutable causal associations between the genome and an outcome; instead genetic effects are often conditioned by socio-environmental factors, and vice versa (Shanahan, Michael J. and Boardman, Jason D. 2009). Even in the rare case of 100% genetic determination of an outcome, socio-environmental interventions can often mitigate genetic effects. For example, assume that impaired vision is due entirely to genetics (it is not, this is just illustrative). If all cases of impaired vision could be completely counteracted by the use of corrective lenses, then despite the purely genetic causes, environmental interventions could completely mitigate a negative genetic propensity for impaired vision. In this example, estimates of the heritability of perfect eye sight would depend on the circumstances in which eye sight is measured. Specifically, if the eye sight for a given population is measured while individuals with impaired vision are wearing corrective lenses (again, assuming corrective lens can completely reverse the effect of impaired vision) the estimated heritability of eye sight would be zero. If, on the other hand, eye sight was measured without the use of corrective lenses the estimated heritability of perfect vision would be 1, i.e. perfectly heritable. For many measures of import to social stratification processes, including educational attainment, it is impossible to obtain measurements that have not been affected by socio-environmental factors. Children are not raised in a vacuum, and it is unethical to randomly constrain parental involvement in their children's lives. Consequently, situating estimates of genetic effects in the context of gene-environment interactions (GxEs) and gene-environment correlations (rGEs) is an important aspect of any analysis that attempts to incorporate genetic measures and estimate their effects.

GxEs refer to the general concept that socio-environmental factors may condition genetic effects on an outcome (Adkins and Vaisey 2009; Conley 2016; Shanahan, Michael J. and Boardman, Jason D. 2009). Figure 1 presents a conceptual model of GxEs in which a path diagram depicting the causal relationship between genetic influences and educational attainment is encompassed within, i.e. conditioned by, socio-environmental contexts. While the name, GxE, is suggestive of multipliciative

Figure 1: Gene-environment Interactions (GxEs) in Educational Attainment



interactions commonly applied in generalized linear models, GxEs come in many forms that include and extend beyond interaction terms.

There are four ideal types of GxEs: (1) triggering, (2) compensation, (3) social control, and (4) enhancement (Shanahan, Michael J. and Boardman, Jason D. 2009:219–22). Two of the four ideal types (triggering and compensation GxEs) refer to the modification of negative, or disadvantageous, genetic effects such as antisocial behavior or lower genetic propensities for advantageous traits/behaviors such as increased educational attainment. On the other hand, enhancement GxEs, refer to the modification of genetic effects by socio-environmental factors that accentuate positive genetic effects similar to how sporting equipment can enhance athletic performance. A simple example of an enhancement GxE is the use of specialty swimming suits by Olympic swimmers. Obviously, Olympic swimmers have some innate genetic propensities that help them swim faster than the rest of us (think of the elongated body, arms, and feet of Michael Phelps), but access to, and the use of, specialized equipment will enhance those differences. Lastly, social control GxEs refer to the suppression of genetic effects (positive or negative) by social norms and/or institutional structures.

Each type of GxE can affect educational attainment estimated of the proportion of variance in educational attainment attributable to genetics in different ways. For example, a triggering GxE could include the onset of a genetic predisposition to depression due to experiences of school related violence such as bullying. If triggering GxEs systematically impact one group more than another, estimates of the proportion of variance in educational attainment attributable to genetics will be attenuated. Similarly, social control GxEs in educational attainment such as segregation and/or tracking based on an individual's race, class, and/or gender will attenuate estimates of the proportion of variance in educational attainment attributable to genetics. Compensation GxEs refer to the suppression of negative genetic effects by positive social and environmental settings and examples of compensation GxEs in educational attainment include the use of tutors, legacy admissions, and increased investments of parental time to counter act genetic propensities that adversely affect educational attainment (e.g. attention deficit disorder, dyslexia, etc.). Similar to triggering and social control GxEs, if the presence of compensation GxEs are systematically different between groups, estimates of the proportion of variance in educational attainment attributable to genetics will be attenuated. Lastly, examples of enhancement GxEs in educational attainment include, selection into "gifted" programs and/or attendance at intensive summer programs based on perceptions of above average ability. Unlike the other GxEs, if enhancement GxEs are pervasive estimates of the proportion of variance in educational attainment attributable to genetics will be artificially inflated.

A related concept to GxEs is gene-environment correlation (rGE). Whereas GxEs refer to the conditioning of the genetic effects on a behavior by socio-environmental factors, rGEs refer to dynamic processes linking genes to context to behavior (Shanahan and Boardman 2009:223). While the term is reminiscent of a Pearson correlation, and can be represented as such in certain cases, as shown in Figure 2, rGEs reflect causal processes in which socio-environmental factors intercede between genetic effects and the outcome with which they are associated. Stated differently, rGEs occur when genetic

Figure 2: Gene-environment Correlations (rGE) in Educational Attainment

propensities for, or against, a behavior lead to greater or lesser interaction with an environment that alters the likelihood of the same behavior respectively.

Failing to consider rGEs, can bias estimates of GxEs. For example, if a majority of people with a genetic propensity for succeeding in education also lack access to quality educational resources (e.g., instructional quality), then it would be difficult to test the hypothesis that access to higher quality educational resources attenuate genetic effects on educational attainment. In this hypothetical example, the estimate of a GxE would be inflated. Conversely, if nearly every one with a lower genetic propensity for completing more years of education came from more advantaged social origins, and thus had increased access to higher quality educational resources, then an examination of GxEs in educational attainment would be downwardly biased.

There are three generally accepted ideal types of rGEs: passive, evocative, and active (Plomin, DeFries, and Loehlin 1977; Shanahan, Michael J. and Boardman, Jason D. 2009:224). Passive rGEs refer to situations in which individuals inherit genetic factors and are exposed to socio-environmental factors that complement those genetic predispositions thus compounding genetic effects due to non-random genetic distributions in a population. For example, observed genetic clustering in schools (Domingue and Belsky 2017) may be reflective of genetic assortment into specific environments. This is similar to an enhancement GxE except that the connection between genetics and the environment is not purposeful. Evocative rGEs refer to situations in which genetic effects elicit differential responses

in social settings. For example, consider the hypothetical case of a child with a genetic propensity for antisocial behavior in a classroom(s) lead by poorly/inadequately trained instructors. An evocative rGE occurs when, the increased incidence of acting out by this student, due to an increased genetic predisposition for antisocial behaviors, are met by harsher disciplinary measures instead of instructional support by the teacher(s), thus leading to an increase in antisocial behavior by the child. Active rGEs, on the other hand, refer to the purposeful confounding of genetic effects by the selective choice of environments by the individual. An example of an active rGE in the case of educational attainment is when individuals with an increased genetic predisposition for succeeding in educational settings actively seek out environments that also increase educational attainment.

THE CASE FOR PARENTAL EDUCATIONAL ATTAINMENT AS A SUITABLE ENVIRONMENT IN GxEs

In order to accurately assess GxEs, measures of the environment, the "E" in GxEs, must meet certain criteria. Specifically, in order to evaluate a GxE in the face of potential rGEs, previous research suggests that the environmental factor(s) must: (1) have a causal relationship with the outcome, (2) show variability in individual responses to the environment, and (3) there needs be a plausible theoretical connection individuals' responses to the environment and genetics (Moffitt, Caspi, and Rutter 2005; Shanahan, Michael J. and Boardman, Jason D. 2009).

In the case of studying GxEs in educational attainment, parental educational attainment satisfies all three requirements. With regards to the first requirement, a long history of research has established a causal relationship between parental educational attainment and children's educational attainment (for an overview of this line of research see Breen and Jonsson (2005)). Importantly, the causal nature of parental educational attainment holds when genetic endowments are considered (see for example Belsky et al. 2016 and Liu 2018). With regards to the second requirement, while the correlation between social origins and destinations in educational attainment is high, the relationship is not

singular, i.e. not all individuals from highly educated parents complete Ph.D.s. Finally, differences in parental educational attainment are known to be highly correlated with resource inequality in early life environments (Link and Phelan 1995).

HYPOTHESES

Before discussing a series of potential hypotheses, a brief discussion of the architecture of genetic influences on educational attainment may be useful. When genetic effects are highly polygenic (i.e. there is no single, or dominant, genetic determinant of an outcome) genome-wide measures of genetic effects should be used instead of single candidate genes as was the case in much of the early work on GxEs (see for example the exemplary work by Shanahan et al. 2008) . The polygenic nature of genetic influences on educational attainment has been well documented (e.g., Okbay et al. 2016; Rietveld et al. 2013). Consequently, in order to accurately assess GxEs in educational attainment measures capable of accounting for genetic effects across the entire genome are required. Three examples of such measures include estimates of: narrow-sense heritability from quantitative genetic analyses, SNP-based heritability from genome-wide complex trait analysis, and polygenic penetrance from polygenic score analysis. While all three methods estimate the proportion of the observed variance in educational attainment attributable to genetic influences across the entire genome, they differ in the ways in which they measure genetic effects. A more detailed discussion of these measures and methods is provided in the data and methods section below.

Combining the theoretical foundations of the unified stratification theory, i.e. that the strength of genetic influences on an outcome are primarily moderated by levels of resource inequality within a social system, with the ideal typologies of GxEs, leads to the following three hypotheses:

H3: Estimates of the heritability of educational attainment will be lowest in the most advantaged social strata, due to a compensation GxE.

H4: Estimates of the heritability of educational attainment will be highest in the least advantaged social strata due to enhancement GxEs.

H5: Estimates of the heritability of educational attainment among all social strata will be significantly less than 1 due to the persistent effects of social and institutional structures on educational attainment processes.

While it is possible, and likely, that the unequal distribution of resources beneficial to educational attainment held by more educated parents operate as an enhancement GxE for some children, the more pervasive GxE within this advantaged social strata is likely that of compensation. The primacy of parental desires to avoid downward mobility among their children in the U.S. is well documented and in the extreme has been referred to as "opportunity hoarding" (Revees 2017). As described previously, compensatory and enhancement GxEs have opposite effects on estimates the proportion of variance in an outcome due to genetics, i.e. heritability. Consequently, if compensatory and enhancement GxEs are equally applied by more educated parents, estimates of heritability among this group should reflect average genetic effects. If, on the other hand, enhancement GxEs occur more frequently than compensatory GxEs, then, heritability estimates will be inflated. And conversely, if compensation GxEs occur more often among more advantaged families, estimates of heritability will be attenuated; and in the extreme will be indistinguishable from zero. H3 formalizes the idea that heritability estimates statistically indistinguishable from zero would provide evidence that despite the many policy interventions meant to equalize educational opportunities, substantial social reproduction in educational attainment continues to occur in the U.S. Importantly, the use of sociogenomic measures and methods to show this trend alleviates the endogeneity issue present in stratification research omitting genetic effects entirely.

Given the focus of many public policies adopted in the U.S. during the latter half of the 20th century to decreasing absolute inequality in educational attainment, it is likely that estimates of the heritability of educational attainment among the least advantaged social strata will be higher than those for groups from relatively more advantaged social origins. If such policies had the intended effect, estimates of the heritability of educational attainment will be higher among the least advanced social strata, reflecting an enhancement GxE as formalized in the fourth hypothesis.

Finally, the fifth hypothesis posits that for all social strata perfect social mobility based on individual level characteristics, i.e. heritability estimates equal to 1, is a myth. Stated in terms of absolute versus relative inequality, a heritability estimate of 1 is suggestive of perfect relative equality. And while some may find the idea of reducing relative inequality to zero, the continued opportunity hoarding of the middle class (Reeves 2017) makes it unlikely that estimates of the heritability of educational attainment will reflect a perfectly mobile social system for any class.

MEASURING GxEs

GxEs represent a patterning of genetic associations across environments, and as such cannot be captured in a single statistical test tests (Shanahan and Boardman. 2009:229). Thus, it's important to evaluate GxEs through a variety of methodological approaches. Furthermore, since all analytic methods have strengths and weaknesses, if observed patterns of GxEs are robust to varied empirical investigations, a stronger case for a significant GxE can be made.

In this study, the investigation of potential GxEs in educational attainment based on social origins, as defined by the highest parental educational attainment of either parent, is undertaken using three methodological strategies that differ in the manner in which they estimate genetic effects: (1) the extended twin family design, (2) genome-wide complex trait analysis, and (3) polygenic score analysis used to estimate the polygenic penetrance of a polygenic score for educational attainment. The extended twin family design comes from the quantitative genetic tradition which relies on known

population averages of genetic similarities between biological siblings to decompose variation in educational attainment between sibling pairs into unobserved measures of additive genetics, i.e. narrow-sense heritability, common environmental factors, and unique socio-environmental factors. Genome-wide complex trait analysis and polygenic score analyses, on the other hand, leverage molecular genetic data, i.e. observed differences on genotyped SNPs, to formulate estimates of heritability comparable to estimates of the narrow-sense heritability (Conley et al. 2016; Yang et al. 2017). Estimates of heritability from genome-wide complex trait analysis, referred to as SNP-based heritability due to the fact that SNPs are the primary independent variables, are derived from linear mixed models in which genetic influences for each individual are estimated as a single random effect (Dudbridge 2016:270). Estimates of heritability from polygenic score analysis are referred to as polygenic penetrance, and are estimated as the incremental R-squared of a polygenic score for the outcome being studied (Conley et al. 2016).

When considered jointly, the strengths and weaknesses of the extended twin family design, genome-wide complex trait analysis, and polygenic penetrance complement one another. For example, a common critique of quantitative genetic models such as the ETFD, is the reliance on strict model assumptions such as the assumption that the experience of common environments (e.g. home environments) is invariant between siblings (see for example (Daw, Guo, and Harris 2015). Despite the fact that quantitative genetic models have been shown to be robust to violations of the equal environment assumption (Conley et al. 2013), and the slight relaxation of the equal environment assumption in the extended twin family design (Keller et al. 2009; Nielsen and Roos 2015), this potential weakness is absent in the other two methodologies. Genome-wide complex trait analysis and polygenic score analyses purposefully restrict their samples to unrelated individuals, thus removing the common environment completely from the analyses. On the other hand, estimates of SNP-based heritability and polygenic penetrance, are based on the available genotyped SNPs in the molecular genetic data, and thus represent a subset of the possible causal genetic variants associated within an

outcome. Consequently, by definition, estimates of SNP-based heritability and polygenic penetrance will always be smaller than estimates of narrow-sense heritability from quantitative genetic analyses (Domingue et al. 2016; Yang et al. 2017). Similarly, while estimates of SNP-based heritability provide an upper limit of the genome-wide significant heritability of SNPs included in the molecular genetic data (Yang et al. 2017), estimates of polygenic penetrance are more conservative due to their reliance on beta-weights from an independent sample (Dudbridge 2013, 2016).

Similar to investigating GxEs through multiple methodological approaches, another within sample sensitivity analysis for identifying potential GxEs, is to examine if the observed patterning of GxEs extends to related measures and concepts (Shanahan and Boardman 2009:230). In results not reported in this article, the analytical strategy employed in this study is applied to three variant measures of educational attainment as well as two other operationalizations of parental educational attainment: maternal educational attainment and paternal educational attainment.

DATA AND MEASURES

This study uses data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is an ongoing nationally-representative longitudinal study of adolescents in the U.S. who were in grades 7-12 in 1994-5. Wave I (79% response rate), collected in 1994-5, included a sample of 20,745 adolescents in 80 high schools selected to ensure representativeness of US schools with respect to region, urbanicity, school size, school type, and racial and ethnic demographics. Wave IV (80% response rate) was collected in 2008. This study uses data on parental educational attainment and reported sibling relationships from Wave I and data on children's educational attainment and molecular genetics from Wave IV, when respondents were between 24-32 years old.

Beyond the large sampling frame and longitudinal nature of the data, at Wave I Add Health over sampled twins and siblings raised in the same household. Siblings residing in the same household at Wave 1 are classified as monozygotic twins (MZ), dizygotic twins (DZ), full siblings (FS), half siblings (HS), cousins (CO), and non-related (NR).

As part of the Wave IV data collection, saliva samples were obtained from consenting participants (96% of Wave IV respondents). Approximately 12,200, or 80% of those participants, consented to long-term archiving and were consequently eligible for genotyping. After quality control procedures, genotyped data are available for 9,975 individuals on 609,130 single-nucleotide polymorphisms (SNPs).

Measures

The primary dependent variable in all analyses is the total years of education completed, or educational attainment, of respondents at Wave IV. To harmonize the measurement of educational attainment in Add Health with the operationalization of education in the latest genome-wide association study of educational attainment, each category of educational attainment reported by Add Health is transformed into years of education, resulting in a continuous variable with a range of 8 – 22 years of education.

Highest parental educational attainment is based on the highest level of parental educational attainment reported by either parent at Wave I. All analyses are stratified by parental educational attainment, thus the continuous measure of parental educational attainment is divided into three categories representing: 1) the completion of high school (or equivalent) by either parent either parent; 2) attending, but not completing college; and 3) the completion a four-year college degree or more education by either parent. This tricotomization of parental educational attainment represents important substantive distinction in potential resources parents can provide in educational settings. Additionally, the division of both the Add Health sibling sample and Add Health molecular genetic samples according to these categories results in nearly equal sample sizes across the three levels of parental educational attainment, thus partially ensuring similar statistical power within each group.

Common socio-demographic controls including race/ethnicity, and sex are included in all analyses. Race/ethnicity is based on respondents self-identified race/ethnicity at Wave I and is split into four dichotomous variables: representing 1) non-Hispanic White, 2) non-Hispanic Black, 3) Hispanic, and 4) Asian/other race/ethnicity. Respondents' biological sex is based on interviewer reports of respondents' sex at Wave I, with 1 = female.

Polygenic Scores (PGSs) are calculated using beta-weights from the most recent genome-wide association study of educational attainment (Okbay et al. 2016) using the PRSice R wrapper (XXXX) in PLINK (Chang et al. 2015; Purcell et al. 2007) following (Dudbridge 2013) as shown in Equation 1 below:

$$PGS_{EA_i} = \sum_{j=1}^{k} \beta_j SNP_{ij} \quad (1)$$

where, SNP_{ij} is the allele frequency of the j^{th} SNP for the i^{th} individual in the Add Health molecular genetic sample, β_j is the estimated regression coefficient for SNP j as reported by Okbay et al. (2016), and PGS_{EA_i} is the raw PGS for educational attainment for the i^{th} individual.

Once the weighted sum is calculated for each individual in the Add Health molecular genetic sample, the PGS is standardized to have a mean of 0 and a standard deviation of 1. All SNP associations, from Okbay et al. (2016) are included in the calculation of the PGS independent of p-value and/or linkage disequilibrium. PGSs created following this procedure follow the same pattern as other PGSs using different combinations of p-value thresholding and pruning due to linkage disequilibrium, while outperforming other PGSs in respect to their predictive power vis-à-vis educational attainment (Ware et al. 2017). A detailed discussion of the calculation of the PGSs is presented in the technical appendix. For a recent review of PGSs see (Conley 2016).

ANALYTIC METHODS

In order to test the robustness of findings, I use three sociogenomic methods to estimate the heritability of educational attainment within the three mutually categories of highest parental educational attainment: (1) completion of four-year college degree or more; (2) attending, but not completing college; and (3) the completion of high school or less. In the first stage of the analysis I estimate a series of quantitative genetics models following the ETFD. In the second stage of analysis, I use molecular genetic data on the Add Health GWAS sample to estimate the SNP-heritability of educational attainment using a series of genome-wide complex trait analyses (GCTA). Then, in the third stage of analysis, I estimate the polygenic penetrance (PGP)

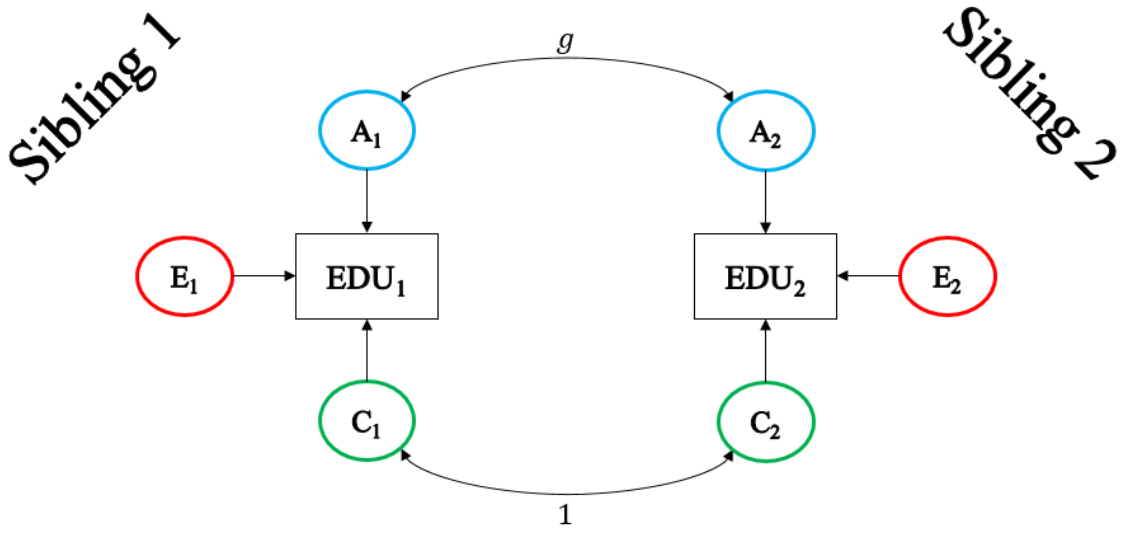
of the PGS for educational attainment. At every stage of analysis, the estimated models include controls for sex, age, and race/ethnicity. Additionally, the GCTAs and estimates of PGP include controls for the first 10 principal components of the genetic data.

Measuring the Heritability of Educational Attainment via the Extended Twin Family Design

Quantitative genetic analyses, also referred to as twin and/or ACE models, rely on known population averages in genetic similarity between biologically related individuals to decompose variation in an observed outcome into genetic, common or shared environmental influences, and unique environmental influences such as individual peer groups (respectively labeled as A, C, and E in Figure 3). In this study, I include all sibling pairs who are: monozygotic twins (MZ), dizygotic twins (DZ), full siblings (FS), half siblings (HS), cousins (CO), and non-related (NR). The extension of twin models to other sibling types is referred to as the extended twin family design (ETFD) (Keller et al. 2009). Prior to the estimation of the ETFD, respondents' educational attainment is residualized on age, sex, and race/ethnicity so as to remove potential confounding due to differences in age, sex, and race/ethnicity. Figure 3 presents a path diagram of the behavior genetic analyses used in this study.

Four variations of the ETFD are estimated within each of the three social strata and then evaluated based on overall model fit following the same criteria as applied to typical structural equation models with latent variables. Model 1 includes all three latent variables A, C, and E; Model 2 includes only the latent variables A and E; Model 3 includes only the latent variables C and E; and Model 4 restricts the relationship to the single latent variable of E. Model 1 represents the typical variance decomposition of ACE models as described above; Model 2 hypothesizes that the influence of the common environment on the residualized measure of educational attainment is zero, while the influence of additive genetic and unique environmental factors are significantly different from zero; Model 3 hypothesizes that the influence of additive genetics is zero, while the influence of the common and unique environmental factors are significantly different from zero; and Model 4 hypothesizes that the only factor with a significant influence on the total years of education an individual completes, net of age, sex, and race/ethnicity, are environmental factors unique to the individual.

Figure 3: Extended Twin Family Model of Educational Attainment*



A – Genetic component
C – Shared environment
E – Unique environment

g = 1.0 (MZ); 0.5 (DZ); 0.25 (HS); 0.125 (CO); 0 (NR)

*EDU_1 a represents a residualized measure of the total years of education completed by the first sibling in a sibling pair while EDU_2 represents a residualized measure of the total years of education completed by the second sibling in each sibling pair.

As mentioned earlier, each of the four models is estimated separately by parental educational attainment, resulting in a total of 12 models. Within each social strata, a best fitting model is identified by comparing various overall model fit indices including the confirmatory fit index (CFI), the Tuker-Lewis index (TLI), the Basian Information Criterion (BIC) following the formulation of the BIC in Raftery (1995), and the root-mean squared error of approximation (RMSEA). Parameter estimates from each model are then used to calculate the narrow-sense heritability of educational attainment within each social strata.

Measuring the Heritability of Educational Attainment via Genome-wide Complex Trait Analysis

Genome-wide complex trait analysis (GCTA), "directly quantifies the proportion of phenotypic variance explained by all SNPs [in the genome-wide data]" (Yang et al. 2017:1305) by comparing the estimated genetic relatedness between unrelated individuals to similarities between their reported total years of education

completed (Yang et al. 2011). GCTA has been widely applied to social demographic outcomes and has been shown to be robust to the number of SNPs available in the genome-wide data (as long as that number of SNPs are > 100,000), the exclusion of SNPs thought to be causally related to the phenotype, and heteroscedastic errors (Domingue et al. 2016). A detailed outline of GCTA is beyond the scope of this paper, but can be found in Yang et al. (2010), Yang et al. (2011), Domingue et al. (2016), and Yang et al. (2017).

GCTA results in a population estimate of SNP-based heritability similar to narrow-sense heritability from the ETFD. The difference between SNP-based heritability and narrow-sense heritability is that the former relies on observed differences in genome-wide data, i.e. SNPs, while the later relies on an unobserved latent measure of genetics.

Measuring the Heritability of Educational Attainment via Polygenic Penetrance

A comparable measure to narrow-sense heritability and SNP-based heritability is the polygenic penetrance (PGP) of educational attainment, i.e. the incremental R-square of a PGS for educational attainment predicting the total years of educational attainment completed (Conley et al. 2016). Analytically, estimates of polygenic penetrance reflect the incremental R-squared due to the addition of the PGS for educational attainment. To maintain comparability to the estimates of narrow-sense heritability and SNP-based heritability the estimates of PGPG include the same set of control variables as in the other analyses.

Patterns versus Statistical Tests

As mentioned previously, when considered jointly, the strengths and weaknesses of the ETFD, GCTA, and PGP complement one another. For example, the potential to underestimate GxE effect sizes in PGP due to a number of possibilities including, insufficiently powered genome-wide association studies and/or the incomplete genotyping of all possible SNPs that may affect educational attainment in the analytic sample, is counterbalanced by the ETFD in which genetic variants are not measured, but instead are statistically derived using latent variables. Furthermore, sitting in-between these two poles, estimates of SNP-based heritability from the GCTAs provide an upper limit of heritability measured based on the set of genotyped SNPs available in the Add Health molecular genetic data (Yang et al. 2017).

PATTERNS OF SOCIAL REPRODUCTION IN QUANTITATIVE GENETIC ANALYSES

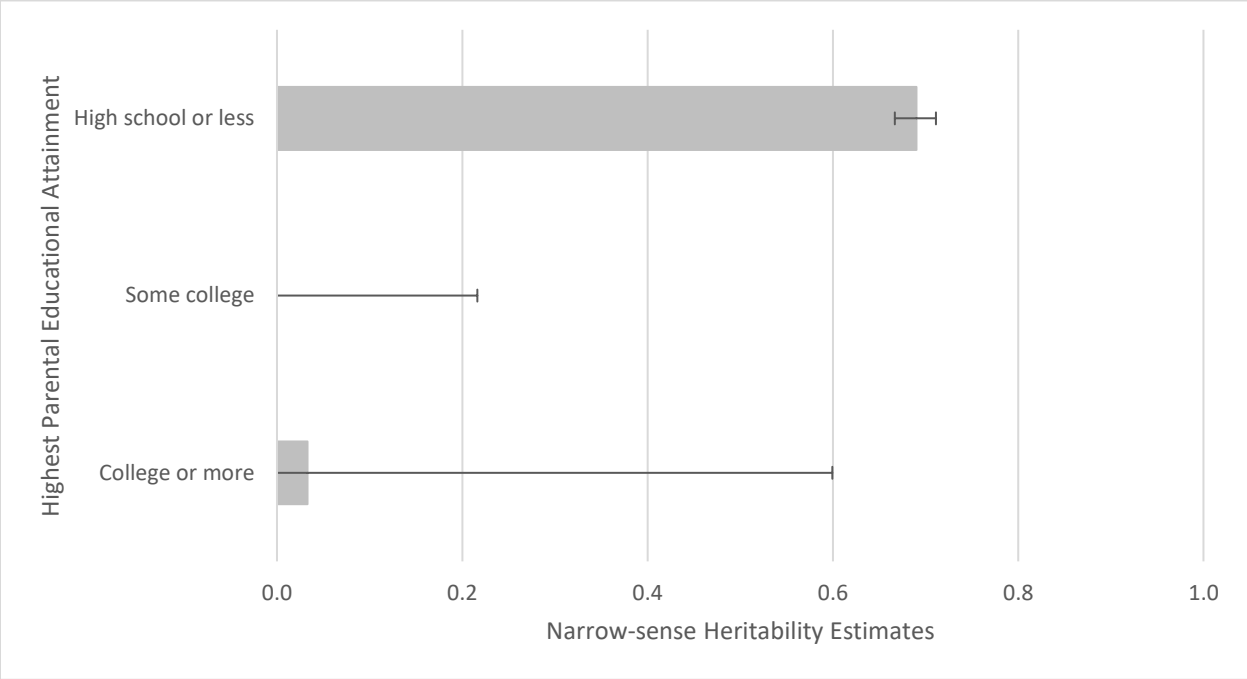
The quantitative genetic analyses paint a clear picture of social reproduction among individuals raised in household in which at least one parent completed some college or more. Table 1 reports the model fit indices for each of the four models, across the three levels of parental educational attainment. The column labeled "ACE" depicts the model fit indices for the model that estimated the additive genetic, common environmental, and unique environmental factors. The column labeled "AE" corresponds to the model that constrains the effect of the common environment to zero. The column labeled "CE" corresponds to the model that constrains the effect of additive genetic influences on educational attainment to zero. Finally, the column labeled "E" corresponds to models in which the effect of the common environment and additive genetics are both assumed to be zero. As shown in Table 1 the best fitting model for individuals raised in homes in which either parent obtained a college degree or more, and for individuals raised in homes in which either parent attended college but didn't finish a 4-year degree is the "CE" model, suggesting that for these groups there is no significant association between genetics and educational attainment. Conversely, the best fitting model for individuals from the least advantaged social origins, i.e. the highest educational attainment of either parent is at most a high school diploma or equivalent, is the "AE" model, in which the effect of the common environment is constrained to zero.

While the patterning of the model fit indices suggests a zero-order association between genetics and educational attainment for the two most advantaged groups, and a zero-order association between the shared environment of siblings and educational attainment for the least advantaged group, the second best fitting model for each group is the full "ACE" in which the additive genetic, the common environment, and unique environment factors are all allowed to influence educational attainment. As shown in Figure 4, evaluating the proportion of observed variance in educational attainment, as estimated in the full "ACE" model, across all three groups yields a similar, finding.

XTable 1: Model Fit of the Behavior Genetics Models by Highest Parental Educational Attainment

	College Plus				Some college				High school or less			
Fit Statistics	ACE	AE	**CE**	E	ACE	AE	**CE**	E	ACE	**AE**	CE	E
Chi-Sqaure	37.93	46.24	**37.96**	77.79	21.96	31.42	**21.96**	51.81	23.64	**23.68**	32.67	74.52
df	15	16	**16**	17	15	16	**16**	17	15	**16**	16	17
p-value	0.001	0.000	**0.002**	0.000	0.109	0.012	**0.144**	0.000	0.071	**0.097**	0.008	0.000
1-RMSEA	0.83	0.81	**0.84**	0.74	0.91	0.87	**0.92**	0.81	0.9	**0.91**	0.87	0.77
CFI	0.39	0.20	**0.00**	0.00	0.76	0.46	**0.79**	0.00	0.85	**0.87**	0.71	0.00
TLI	0.76	0.70	**0.42**	0.43	0.90	0.80	**0.92**	0.57	0.94	**0.95**	0.89	0.65
BIC	-48.92	-46.40	**-54.68**	-20.64	-66.16	-62.57	**-72.03**	-48.07	-65.06	**-70.94**	-61.94	-26.01

Figure 4: Narrow-sense Heritability of Educational Attainment by Highest Parental Education*



* All estimates of narrow-sense heritability include controls for age, sex, and race/ethnicity. Error bars represent 95% confidence intervals.

The pattern depicted in Figure 4 suggests a substantial gene-by-environment interaction in which educational attainment for individuals from the most advantaged social strata is due entirely to socio-environmental factors. While the analyses cannot determine the exact type of the GxE operating in the most advantaged strata, the stark differences in the role of social resources in educational processes among the most advantaged is a likely candidate given that among the most advantaged, the role of innate individual level characteristics typically identified by the additive genetic latent variable are statistically indistinguishable from zero.

Combined with the first and second hypotheses, the observed patterning of findings across social strata from the ETFD provide preliminary evidence that, on average, social reproduction is the primary mechanism of educational attainment for individuals from the most advantaged social origins while mobility, i.e. the leveraging of innate individual level characteristics, plays an important, albeit imperfect

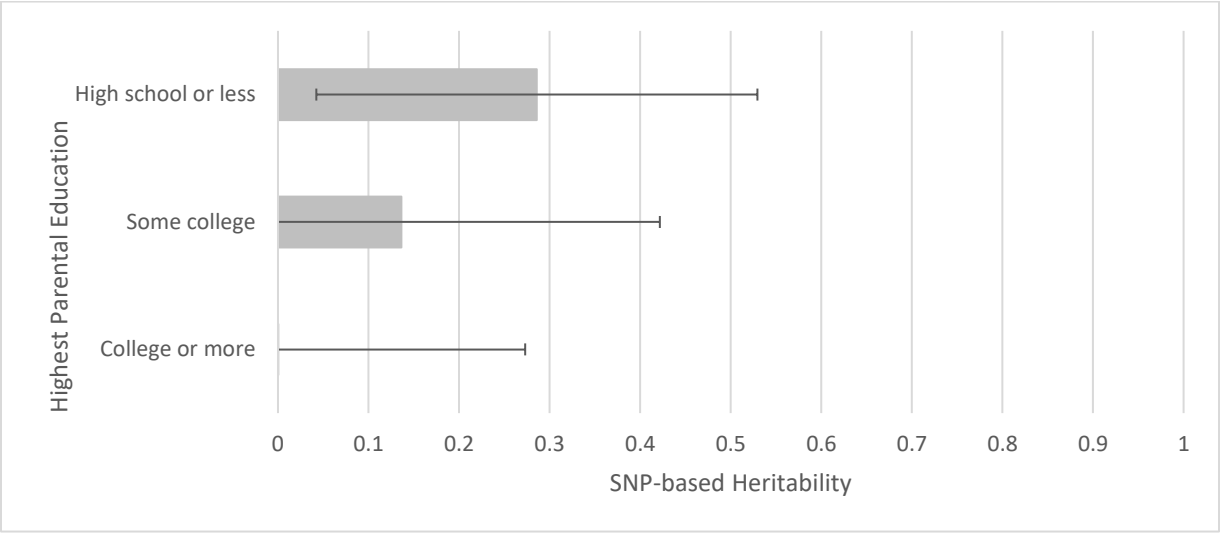
role in educational attainment among individuals from the least advantaged social strata. Recall that H1 posits that for social strata in which social reproduction is the dominant mechanism of educational attainment, estimates of the heritability of educational attainment will be close, or equal to zero. Conversely, H2 posits that for social strata in which social mobility is the dominant mechanism in educational attainment, estimates of the heritability of educational attainment will be close, or equal to one. While the results fail to show perfect mobility in educational attainment among individuals from the least advantaged social strata, this finding is in line with hypothesis five that estimates of the heritability of educational attainment for the least advantaged social strata will be significantly less than 1 due to persisting inequality of opportunity in educational attainment. The results also confirm hypotheses three and four. Specifically, these hypotheses related to the relative estimates of the heritability of educational attainment between social strata.

PERSISTING PATTERNS OF SOCIAL REPRODUCTION IN MOLECULAR GENETIC ANALYSES

The patterns of social reproduction depicted in the quantitative genetic analyses of educational attainment are mirrored by the results from the GCTAs and PGP analyses. The results from the GCTA stratified by parental educational attainment are presented in Figure 5. As expected estimates of SNP-based heritability are smaller than the estimates of narrow-sense heritability estimated in the ETFD, but the pattern of results in which the proportion of variance in the total years of education completed attributable to genetics is only statistically significant for individuals from the least advantaged backgrounds remains unchanged; again suggesting that social reproduction is the primary mechanism driving the amount of education individuals from more advantaged backgrounds obtain.

One caveat to the GCTA results is that, unlike the behavior genetic analyses, and despite the fact that the point estimate for only the least advantaged group is statistically significant, the 95% confidence

Figure 5: SNP-based Heritability of Educational Attainment by Highest Parental Education*

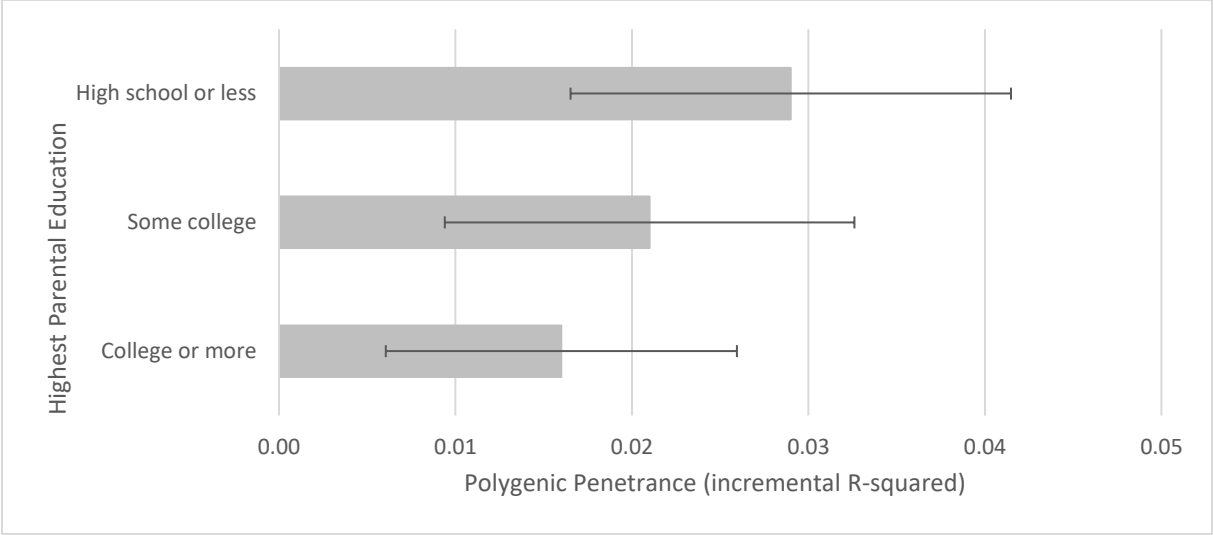


*All estimates of SNP-based heritability include age, sex, race/ethnicity, and the first five principal components of all observed SNPs as covariates. Error bars represent 95% confidence intervals.

intervals of the estimates of SNP-based heritability for the three groups overlap. Thus, it is possible that the true estimates of SNP-based heritability within the three groups may not be statistically distinct. While this is a statistical possibility, the overlapping 95% confidence intervals are likely an artifact of sampling error. As observed by Yang et al. (2010:568), in relation to the performance of GCTA estimates of SNP-based heritability with changes in sample size, "The average estimates of variance explained by all SNPs are not affected by sample size, but as expected, the sampling error increases as sample size decreases." But, as implied by Shanahan and Boardman (2009), when identifying potential GxEs the statistical significance of a single test is less important than the robust patterning of results.

Similar to the GCTA results, estimates of the PGP of educational attainment follow the same pattern of decreasing significance in explaining the total variance in individuals' own educational attainment as parental educational attainment increases. Figure 6 plots the PGP of educational attainment across the three levels of parental educational attainment. As expected the magnitude of the estimates of PGP are smaller than the estimates of SNP-based heritability from the GCTAs., but the pattern is nearly identical. The exact estimates of polygenic penetrance for the three groups are: 0.016 (s.e. 0.005) for the

Figure 6: Polygenic Penetrance (PGP) of Educational Attainment by Highest Parental Educational Attainment*



*All estimates of include age, sex, race/ethnicity, and the first five principal components of all observed SNPs as covariates. Error bars represent 95% confidence intervals. The x-axis has been shortened to increase visibility of the between patterning of PGP.

group with at least one parent completing a four-year college degree or more, 0.021 (s.e. 0.006) for the group in which at least one parent attended college but never received a degree, and 0.029 (s.e. 0.006) for the group with neither parent continued their education beyond high school. As is evident from the 95% confidence intervals shown in Figure 6, and the standard errors reported above, within the Add Health molecular genetic sample, the pattern is consistent with that observed in both the ETFD and the GCTAs. While some difference in the statistical distinction between point estimates exist, the resilience of the patterning of estimates of heritability across all three methods is striking. As with the estimates of SNP-based heritability, the size of the standard errors of the PGP estimates is likely due in part to sampling fluctuations and would be expected to decrease with in larger sample sizes.

As a sensitivity analysis, all analyses described above were repeated using three variant measures of educational attainment as well as two other operationalizations of parental educational attainment (i.e.

mother's education and father's education). In all cases, the pattern of GxEs identified above remained constant (results available upon request).

DISCUSSION AND CONCLUSIONS

Results across the three methodological approaches show a consistent patterning of results indicative of a robust finding of a GxE in educational attainment among individuals completing their education during the early years of the 21st century in the U.S. Revisiting the hypotheses stated earlier, a narrative of differing environmental effects and importance of genetic factors in educational attainment based on an individual's social origins can be observed.

Recall that the first two hypotheses (H1 and H2), stemming from (Adkins and Vaisey 2009) unified stratification model, posit that the relative openness and closure of a social system can be interrogated by examining estimates of heritability. Specifically, H1 suggests that heritability estimates indistinguishable from 0 indicate the "strong" closure of social systems while heritability estimates close to, but not equal to, 0 indicate "weak" social closure. On the other end of the spectrum, H2 suggests that heritability estimates indistinguishable from one indicate "strong" mobility, or openness, in a social system while heritability estimates significantly below one indicate "weak" social mobility. The repeated finding of a clear gradient in heritability estimates of educational attainment based on social origins, suggests that educational processes are defined by "weak" (as reported in the GCTA and PGP analyses) and possible "strong" (as indicated by the ETFD results) social closure. Stated differently, the evidence suggests that social reproduction, independent of individual differences in genetic propensities for ability, personality, and other characteristics potentially beneficial in obtaining more educational attainment, is the dominate mechanism in determining how much education an individual obtains for the more advantaged social strata (i.e., the middle and upper classes).

Conversely, the results for individuals from the most disadvantaged social origins (i.e. the working class) indicate at most weak mobility in educational attainment.

In the evaluation of the third hypothesis (H3), that estimates of the heritability of educational attainment for the most advantaged group will be the lowest of all three groups, likely indicative of a compensation GxE, the continued patterning of results suggests that socio-environmental factors are indeed suppressing genetic effects for this group. While this study did not explicitly test which elements of the social environment are leading to the suppressing of genetic effects, the degree of suppression provides a compelling argument for a compensation GxE. Compensation GxEs refer to the suppression of negative genetic effects by positive social and environmental settings (Shanahan and Boardman 2009:219-222). Examples of compensation GxEs in educational attainment include the use of tutors, legacy admissions, and increased investments of parental time to help children who have a genetic propensity that adversely affects educational attainment. Such genetic effects include adverse genetic propensities for any trait that is causally associated with educational attainment (e.g., an increased genetic propensity for anti-social behavior or a lower genetic propensity for cognitive ability). In each case, the compensation GxE posits that over time parents become aware of lower performance among one, or more, of their children and actively seek to attenuate the effect of any negative innate propensities by increasing their relative investments in that child's education.

The results also indicate that hypothesis four (H4), i.e. that estimates of the heritability of educational attainment will be highest for individuals in the least advantaged social strata due to weak upward mobility and policies aimed at decreasing absolute inequality in educational attainment. Again, this study did not explicitly test which elements of the social environment are associated with the enduring higher estimates of the heritability of educational attainment across

the three methods used in the analyses, but combined with our current understanding socio-environmental effects in educational attainment among the working class (for a recent overview see (Reeves 2017) the results are indicative of a decrease in absolute inequality in educational attainment. Evidence of decreases in relative inequality would include similar heritability estimates between social strata. The obstinate patterning of results identified in this study, combined with H3, suggest that changes in relative inequality are minor at best and non-existent at worst. This finding in and of itself is not novel, but the use of genetic measures in its documentation represents an important advance in our understanding of educational inequality in the U.S. at the beginning of the 21st century. Furthermore, previous research has not considered inter-class differences in heritability (e.g., (Nielsen and Roos 2015).

Lastly, the results are suggestive that hypothesis five (H5), i.e. that estimates of the heritability of educational attainment among the least advantaged social strata will be significantly less than 1 due to the lingering effects of social and institutional structures limiting the degree of educational mobility afforded to the for the most disadvantaged, is confirmed. Similar to the other hypotheses, a direct test of the causal association between explicit measures of the social environment and estimates of heritability are beyond the scope of this study, but the results indicate that, at most, the heritability of educational attainment for the most disadvantaged social strata is 0.718 (with a 95% confidence interval of 0.723 and 0.713). Note that these estimates are net of the effect of race/ethnicity on educational attainment. But, that does not mean that the effects of social and institutional structures on estimates of heritability are invariant between race/ethnic groups. While such analyses are beyond the scope of this study, future research should investigate if the degree of educational mobility, operationalized as the heritability if educational attainment, is invariant between race/ethnic groups within the working class.

In concluding, the results presented in this study conform to the state of the art understanding of GxEs in that a consistent patterning of differential genetic effects is observed based on meso-level environmental differences (Shanahan and Boardman 2009). Likely determinants of the robust patterning of genetic effects documented above include resource inequalities based on social class (e.g. differences in the quality of classroom instruction; increased access to computer labs, AP courses, etc.; and an increased availability of parental time for involvement in school events and assistance with school assignments), and the general focus in public policies concerning educational attainment on decreasing absolute inequality in while neglecting the importance of relative inequality. These results challenge popular belief in the myth that the potential benefits of increased educational attainment are open to the "best" and "brightest," independent of their social origins. And invite inquiries into the degree to which higher educational attainment is primarily purchased by those from more advantaged social classes, independent of individual level characteristics.

REFERENCES

Automatic citation updates are disabled. To see the bibliography, click Refresh in the Zotero tab.

Adkins, Daniel E. and Stephen Vaisey. 2009. “Toward a Unified Stratification Theory: Structure, Genome, and Status across Human Societies.” *Sociological Theory* 27(2):99–121.

Belsky, Daniel W. et al. 2016. “The Genetics of Success How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development.” *Psychological Science* 0956797616643070.

Belsky, Daniel W. and Salomon Israel. 2014. “Integrating Genetics and Social Science: Genetic Risk Scores.” *Biodemography and Social Biology* 60(2):137–55.

Blau, Peter and Otis Dudley |. Duncan. 1967. *The American Occupational Structure*. New York, N.Y.: Wiley.

Boardman, Jason D., Benjamin W. Domingue, and Jonathan Daw. 2015. “What Can Genes Tell Us about the Relationship between Education and Health?” *Social Science & Medicine (1982)* 127:171–80.

Breen, Richard and John H. Goldthorpe. 1997. “EXPLAINING EDUCATIONAL DIFFERENTIALS: TOWARDS A FORMAL RATIONAL ACTION THEORY.” *Rationality and Society* 9(3):275–305.

Breen, Richard and Jan O. Jonsson. 2005. “Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility.” *Annual Review of Sociology* 31(1):223–43.

Chabris, Christopher F., James J. Lee, David Cesarini, Daniel J. Benjamin, and David I. Laibson. 2015. “The Fourth Law of Behavior Genetics.” *Current Directions in Psychological Science* 24(4):304–312.

Chang, Christopher C. et al. 2015. “Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets.” *GigaScience* 4:7.

Conley, Dalton et al. 2015. “Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?” *Sociological Science* 2:82–105.

Conley, Dalton. 2016. “Socio-Genomic Research Using Genome-Wide Molecular Data.” *Annual Review of Sociology* 42(1):275–99.

Conley, Dalton, Thomas M. Laidley, Jason D. Boardman, and Benjamin W. Domingue. 2016. “Changing Polygenic Penetrance on Phenotypes in the 20th Century Among Adults in the US Population.” *Scientific Reports* 6:30348.

Conley, Dalton, Emily Rauscher, Christopher Dawes, Patrik K. E. Magnusson, and Mark L. Siegal. 2013. “Heritability and the Equal Environments Assumption: Evidence from Multiple Samples of Misclassified Twins.” *Behavior Genetics* 43(5):415–26.

Daw, Jonathan, Guang Guo, and Kathie Mullan Harris. 2015. “Nurture Net of Nature: Re-Evaluating the Role of Shared Environments in Academic Achievement and Verbal Intelligence.” *Social Science Research* 52:422–39.

Diewald, Martin, Tina Baier, Wiebke Schulz, and Reinhard Schunck. 2015. “Status Attainment and Social Mobility.” *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 67(1):371–95.

Domingue, Benjamin W. et al. 2016. “Genome-Wide Estimates of Heritability for Social Demographic Outcomes.” *Biodemography and Social Biology* 62(1):1–18.

Domingue, Benjamin W. and Daniel W. Belsky. 2017. “The Social Genome: Current Findings and Implications for the Study of Human Genetics” edited by G. P. Copenhaver. *PLOS Genetics* 13(3):e1006615.

Domingue, Benjamin W., Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D. Boardman. 2015. “Polygenic Influence on Educational Attainment: New Evidence from the National Longitudinal Study of Adolescent to Adult Health.” *AERA Open* 1(3):2332858415599972.

Dudbridge, Frank. 2016. “Polygenic Epidemiology: Polygenic Epidemiology.” *Genetic Epidemiology* 40(4):268–72.

Dudbridge, Frank. 2013. “Power and Predictive Accuracy of Polygenic Risk Scores” edited by N. R. Wray. *PLoS Genetics* 9(3):e1003348.

Guo, G. and E. Stearns. 2002. “The Social Influences on the Realization of Genetic Potential for Intellectual Development.” *Social Forces* 80(3):881–910.

Keller, Matthew C. et al. 2009. “Modeling Extended Twin Family Data I: Description of the Cascade Model.” *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies* 12(1):8–18.

Liu, Hexuan. 2018. “Social and Genetic Pathways in Multigenerational Transmission of Educational Attainment.” *American Sociological Review* 0003122418759651.

Moffitt, Terrie E., Avshalom Caspi, and Michael Rutter. 2005. “Strategy for Investigating Interactions between Measured Genes and Measured Environments.” *Archives of General Psychiatry* 62(5):473–81.

Morton, N. E. 1974. “Analysis of Family Resemblance. I. Introduction.” *American Journal of Human Genetics* 26(3):318–30.

Nielsen, François. 2006. “Achievement and Ascription in Educational Attainment: Genetic and Environmental Influences on Adolescent Schooling.” *Social Forces* 85(1):193–216.

Nielsen, François. 2016. “The Status-Achievement Process: Insights from Genetics.” *Frontiers in Sociology* 1. Retrieved November 22, 2016 (http://journal.frontiersin.org/Article/10.3389/fsoc.2016.00009/abstract).

Nielsen, François and J. Micah Roos. 2015. “Genetics of Educational Attainment and the Persistence of Privilege at the Turn of the 21st Century.” *Social Forces* 94(2):535–61.

Okbay, Aysu et al. 2016. “Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment.” *Nature* 533(7604):539–42.

Parsons, Talcott. 1940. “An Analytical Approach to the Theory of Social Stratification.” *American Journal of Sociology* 45(6):841–62.

Plomin, Robert, J. C. DeFries, and John C. Loehlin. 1977. “Genotype-Environment Interaction and Correlation in the Analysis of Human Behavior.” *Psychological Bulletin* 84(2):309–22.

Plomin, Robert, John C. DeFries, Valerie S. Knopik, and Jenae M. Neiderhiser. 2013. *Behavioral Genetics*. 6 edition. New York: Worth Publishers.

Plomin, Robert, Claire MA Haworth, and Oliver SP Davis. 2009. “Common Disorders Are Quantitative Traits.” *Nature Reviews Genetics* 10(12):872.

Purcell, Shaun et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American Journal of Human Genetics* 81(3):559–75.

Raftery, Adrian E. 1995. “Bayesian Model Selection in Social Research.” *Sociological Methodology* 25:111.

Reeves, Richard. 2017. *Dream Hoarders : How the American Upper Middle Class Is Leaving Everyone Else in the Dust, Why That Is a Problem, and What to Do about It*. Washington, D.C.: Brookings Institution Press.

Rietveld, Cornelius A. et al. 2013. “GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment.” *Science* 340(6139):1467.

Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969. “The Educational and Early Occupational Attainment Process.” *American Sociological Review* 34(1):82.

Shanahan, Michael J. and Boardman, Jason D. 2009. “Genetics and Behavior in the Life Course: A Promising Frontier.” Pp. 215–35 in *The Craft of Life Course Research*. New York, N.Y.: The Guilford Press.

Shanahan, Michael J., Stephen Vaisey, Lance D. Erickson, and Andrew Smolen. 2008. “Environmental Contingencies and Genetic Propensities: Social Capital, Educational Continuation, and Dopamine Receptor Gene *DRD2*.” *American Journal of Sociology* 114(S1):S260–86.

Ware, Erin B. et al. 2017. “Heterogeneity in Polygenic Scores for Common Human Traits.” *BioRxiv* 106062.

Yang, Jian et al. 2010. “Common SNPs Explain a Large Proportion of the Heritability for Human Height.” *Nature Genetics* 42(7):565–69.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. “GCTA: A Tool for Genome-Wide Complex Trait Analysis.” *The American Journal of Human Genetics* 88(1):76–82.

Yang, Jian, Jian Zeng, Michael E. Goddard, Naomi R. Wray, and Peter M. Visscher. 2017. “Concepts, Estimation and Interpretation of SNP-Based Heritability.” *Nature Genetics* 49(9):1304–10.

TECHNICAL APPENDIX

Polygenic Scores (PGSs), sometimes referred to as polygenic risk scores or genetic risk scores, represent the influence of additive genetics on a specific trait/behavior, commonly referred to as a "phenotype." The calculation of PGSs relies on summary statistics from a genome-wide association study (GWAS) to create a weighted sum of the associations between allele frequencies at individual single-nucleotide polymorphisms (SNPs, pronounced snips) and the phenotype for which the GWAS was conducted. This process yields a hypothesis free measure of the additive genetic influences on the phenotype being studied. PGSs are hypothesis free because they aggregate the individual associations between SNPs and the phenotype, thus removing the possibility of investigating links between specific biological pathways and the phenotype.

The PGS calculated for this study uses regression weights from the most recent, and highest powered, GWAS of educational attainment available at the time of submission (Okbay et al. 2016). The regression weights for each SNP from Okbay et al. (2016) are derived from a series of linear regression models estimating the association (Beta) between the allele frequency at each SNP, and the total years of education an individual completed (edu) in a sample of 293,723 individuals of western European descent. Equation 1 provides a formal description of the analyses form Okbay et al. (2016) from which beta weights for the PGS were estimated; subscripts for individuals and studies have been dropped to reduce clutter.

$$edu = \alpha + \beta_j SNP_j + PC\gamma + C\eta + S\theta + \varepsilon \quad (1)$$

where SNP is the allele frequency for the jth SNP, **PC** is a vector of the first ten principal components of the genome-wide data used to capture genetic population stratification; **C** is a vector of individual-level controls including a third-order polynomial of age, a dichotomous variable for being female, and their interactions; and **S** is a vector of study specific controls including dichotomous variables for any major

event such as policy changes to wars that may affect access to education (Okbay et al. 2016 Supplemental Materials).

The regression coefficients, standard errors, and p-values for each iteration of Equation 1 over the j SNPs are then used to calculate the PGS for educational attainment in the Add Health sample as:

$$PGS_{EA_i} = \sum_{j=1}^{k} \beta_j SNP_{ij} \quad (2)$$

where, SNP_{ij} is the allele frequency of the j^{th} SNP for the i^{th} individual and β_j is the estimated regression coefficient for SNP j derived form Equation 1 and as reported by Okbay et al. (2016). Once a weighted sum is calculated for each individual in the dataset, the PGS is standardized to have a mean of 0 and a standard deviation of 1 (Dudbridge 2013). All SNP associations, from Okbay et al. (2016) are included in the calculation of the PGS independent of p-value and/or linkage disequilibrium, i.e. no pruning or trimming. PGSs created following this procedure follow the same pattern as other PGSs using different combinations of p-value thresholding and accounting of linkage disequilibrium, while outperforming other PGSs in respect to their predictive power vis-à-vis their associated phenotype (Ware et al. 2017). For a detailed outline of the creation of PGSs and their potential uses in sociology see Conley (2016).