Lehrer, Steven F.; Ding, Weili

Article

Are genetic markers of interest for economic research?

IZA Journal of Labor Policy

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Lehrer, Steven F.; Ding, Weili (2017) : Are genetic markers of interest for economic research?, IZA Journal of Labor Policy, ISSN 2193-9004, Springer, Heidelberg, Vol. 6, Iss. 2, pp. 1-23,
https://doi.org/10.1186/s40173-017-0080-6

This Version is available at:
https://hdl.handle.net/10419/194377

Lehrer and Ding *IZA Journal of Labor Policy* (2017) 6:2
DOI 10.1186/s40173-017-0080-6

IZA Journal of Labor Policy

REVIEW Open Access

Are genetic markers of interest for economic research?

Steven F. Lehrer[1,2,3*] and Weili Ding[1,2]

* Correspondence: sl164@nyu.edu
[1]School of Policy Studies and Department of Economics, Queen's University, Kingston K7L3N6, Ontario, Canada
[2]NYU-Shanghai, 1555 Century Avenue Office 1127, Pudong New District, Pudong 200122, Shanghai, China
Full list of author information is available at the end of the article

Abstract

The idea that genetic differences may explain a multitude of individual-level outcomes studied by economists is far from controversial. Since more datasets now contain measures of genetic variation, it is reasonable to postulate that incorporating genomic data in economic analyses will become more common. However, there remains much debate among academics as to, first, whether ignoring genetic differences in empirical analyses biases the resulting estimates. Second, several critics argue that since genetic characteristics are immutable, the incorporation of these variables into economic analysis will not yield much policy guidance. In this paper, we revisit these concerns and survey the main avenues by which empirically oriented economic researchers have utilized measures of genetic markers to improve our understanding of economic phenomena. We discuss the strengths, limitations, and potential of existing approaches and conclude by highlighting several prominent directions forward for future research.

JEL Classification: I12, J19, I26

Keywords: Genetic markers, Gene–environment interactions, Genome-wide association studies, Candidate genes, Genetic instruments, Within-family variation

1 Introduction

It would not be an exaggeration to say that the mere mention of the word genetics to an economist a decade ago could cause alarm. This alarm may have been a response in part to one recalling the general response to the Harvard President Lawrence H. Summers' January 14, 2005, speech at an economics conference when discussing the under-representation of female scientists at elite universities.[1] This alarm perhaps was also triggered by memories of events approximately one decade earlier when Richard Herrnstein and Charles Murray attracted substantial controversy following the 1994 publication of "The Bell Curve," which was popularly (mis)interpreted as ascribing the link between race and IQ to genetic factors. [2] Even recently, economists working on issues related to genetic factors continue to attract interdisciplinary criticism. For example, Ashraf and Galor (2013) arguing for the importance of genetic diversity in explaining national income per capita drew a series of harsh responses from a long list of prominent scientists and anthropologists.[3] These three independent episodes occurred in a 20-year span have clearly indicated the controversy that one may encounter when interpreting or accounting for genetic factors within economic analyses. Thus, it would be unsurprising if individual researchers today would conclude that it is best to ignore genetic factors since the potential

costs from the subsequent criticism and potential damage to one's academic reputation could greatly outweigh any benefits one may receive from incorporating them. In short, genetic information becomes the hornet's nest that most economists stay away from.

In this paper, we argue that this would likely be the wrong response. Not only has the role of empirical work in economics increased sharply over the last 20 years but also has now a growing number of datasets that provide detailed information on genetic characteristics. Genetic markers are now being collected in multiple nationally representative social surveys that Conley (2009) argues can be deployed to (i) assess the direct impact of specific genetic influences on socioeconomic and behavioral outcomes, (ii) explore genetic–environmental interactions, and (iii) trace genealogies across time and space. Indeed, as we discuss below, economists have done substantial applied research related to the first two themes. Further, simultaneous to this underlying trend suggesting the growing importance of empirical economics has been the development of a multitude of econometric strategies that exploit various research designs to identify causal impacts. These applied econometric methods have transformed empirical practice not solely within economics but also in other disciplines such as political science and sociology. We suggest that as economists increase their familiarity with genetic data, it is likely that they can develop methodological tools to generate new strategies to shed new light on the role of genetic factors that will be of interest to those within economics and in many other scientific disciplines.

This paper can be viewed as an extension of the comprehensive reviews presented in Benjamin et al. (2007, 2012a) and Lehrer (2016) that explore the use of genetic markers in studies within economics. While Benjamin et al. (2007) coined the term "genoeconomics" for this field, the view that we advance is somewhat less ambitious. We argue that genetic markers are simply a new way to get inside the black box of individual permanent unobserved heterogeneity within numerous fields in economics. For example, in studies that explore labor supply, researchers often employ fixed effects to reflect permanent unobserved differences in tastes or preferences across individuals. Similarly, when estimating wage equations, researchers often employ fixed effects to capture permanent productivity characteristics of the individual. Genetic markers may be truly what is meant by permanent unobserved heterogeneity since they are assigned at conception and (with the sole exception of monozygotic twins) differ markedly across individuals.[4]

While some economists have begun to incorporate data on genetic markers in their empirical analyses, their use remains scattered and limited to a handful of specific applications. On the one hand, this is somewhat surprising given the long history of research that explores how numerous traits and behaviors pass from one generation to the next. With data on genetic markers, perhaps one can understand how the transmission of genetic factors influences the transmission of outcomes. We should also state explicitly that recent work by economists with genetic data has attracted significant positive acclaim by researchers in other disciplines. Thus, our true aims of this survey are to reduce entry costs and hopefully attract other labor economists to consider integrating genetic factors within their studies.

This paper is organized as follows. In the next section, we provide a brief scientific primer on genetic terminologies and then review the four major strands of research in economics that has used genetic data to date. Section 4 proposes three directions for future research among economists. A concluding section summarizes our arguments and draws links to how research using genetic data within economics is actually following trends in research within labor economics.

2 A primer on genetics

In the Oxford Dictionary, the word genome is defined to be a blend of the word gene and chromosome. The genome is contained in all cells that have a nucleus and consists of more than 3.2 billion DNA base pairs located on 23 pairs of chromosomes. To help visualize the human genome, consider an instruction manual composed of 23 chapters (chromosomes) that is over 3.2 billion letters. The length of each chapter varies between 48 and 250 million letters (A, C, G, T) without any spaces. This genome that lies within each cell in our body is formed at conception when one member of each pair of chromosomes is inherited from the mother and the other from the father.

Using genetic data requires undertaking a molecular genetic approach to understand variation between individuals in the genetic code itself. This differs sharply from the approach in behavioral genetics that categorizes much earlier research in economics that aimed to understand the role of genetic factors focused on using data collected from samples of twins or siblings. Briefly, interested readers are referred to Behrman (2016) for a recent review of research using this approach; researchers begin by assuming that all variation in the outcome being investigated could be decomposed into additively separable genetic and environmental contribution. That is, the variance of a behavior being investigated is decomposed into three orthogonal components: additive genetic effects (A), common environment effects (C), and unique environment effects (E), hence the acronym ACE models.[5]

By contrasting within twin correlation estimates that assumed an equal environment that had equal impacts between monozygotic twins who share the same hereditary and environmental variables thereby providing estimates of $A + C$, with estimates conducted on dizygotic twins who only share the same environmental variables and on average 50% of their genes, providing a direct estimate of $\frac{1}{2}A + C$, one could isolate the hereditary effect A, by taking twice the difference between identical and fraternal twin correlations. Similarly, C is then obtained by subtracting the estimated A from the identical twin correlation, whereas an estimate of E is given by subtracting the identical twin correlation from the number one. Within economics, Taubman (1976) is generally considered the first such study, which estimated that between 18 and 41% of variation in income across individuals was heritable.[6] Research using genetic data is now moving beyond variance decompositions between twins of different zygosity[7] and now focuses on analyzing the impacts of specific portions on the genetic code.

Molecular genetics is the branch of genetics that studies the structure and function of DNA. The sequencing of the human genome in 2001 (Venter et al. 2001) provided a means to measure genetic variation across individuals. One of the principal means through which genetic variation occurs is called a single nucleotide polymorphism (SNP) which is simply a mutation at a specific point in the genetic code where a single nucleotide is substituted (i.e., using the analogy before a single letter such as an A is substituted with a T at that point).[8] It has been estimated that there are only approximately two million sites on the genome where an SNP can be found and it is common to refer to the genetic variants of SNPs by the number of alleles. For example, at one of these specific locations, one's genotype can be denoted by the number of risky alleles (0, 1, or 2). Only a small minority of all of the known SNPs are thought to play important roles influencing the function and structure of the human body, and these could be selectively advantageous or disadvantageous. In other words, while the human genome

is over 3.2 billion chemical letters in length, less than 0.1% of these locations are believe to account for observed differences in socioeconomic outcomes.

3 Categorizing research by economists with genetic data

The majority of databases that contain genetic information were collected by medical scientists. However, a growing number of longitudinal databases that were designed for social scientists are adding genetic information. For example, the Add Health study has collected information on a few SNPs for primarily the sibling sample, the Health and Retirement Study has recently begun to make this information available from consenting participants, and the UK Biobank has linked genetic information to participants in the 1958 birth cohort study. In general, to obtain measures of molecular genetic variation, a number of commercial entities have developed technologies that measure several hundred thousand human SNPs simultaneously from blood or saliva samples.[9] Over the last decade, there have been a multitude of technological breakthroughs that make it easier not only to genotype more SNPs and other genetic variants but also to do so at lower costs. That said, in many datasets, the genetic information provided is an imputed SNP that is calculated based on the high degrees of correlation between neighboring SNPs.[10] Prior to describing how this data is being utilized, it is important to point out that while some characteristics or health outcomes are known to be a unique result of a specific genetic difference, most characteristics that economists are interested in are polygenic, meaning they are influenced by multiple genetic polymorphisms.

Before proceeding further, a controversial issue that researchers in this area face is calling immutable characteristics such as SNPs "treatments." Many critics point to the impossibility of manipulating genetic traits that are fixed at conception in a manner that is analogous to administering a treatment in a randomized experiment. However, Greiner and Rubin (2011) noted that it is actually a matter of perception on those characteristics and perceptions are not immutable.[11] Even without going down to the level of perception, if two individuals are the same in all important characteristics (age, gender, education, family situations, residence, etc.) except for a specific SNP, then their difference in outcomes can still be attributed to this specific genetic difference, in which sense a specific SNP would be a legitimate treatment in the potential outcomes framework.

3.1 Candidate gene studies

A common refrain in many economic seminars is that one is constrained by data limitations. Indeed, much of the earliest work by economists using genetic data is subject to this limitation. Much of the early research was limited by the genetic information collected within the data being investigated. Generally, the initial genetic markers made available were those that were hypothesized to be the main importance. These markers are called candidate genes, and they were generally chosen to be genotyped since they were located in a particular chromosome region suspected of being involved in the outcome or its protein product may suggest that it could influence the outcome being investigated.

Numerous candidate gene studies in economics investigate whether specific SNPs correlate with measures of economic primitives such as risk aversion and delay discounting parameters, thus providing its biological microfoundation. Studies in this area are generally motivated by Cesarini et al.'s (2009) behavioral genetic investigation that

suggests that approximately one fifth of the variation in these measures is due to genetic factors. Initially, economists focused their candidate gene investigations on whether genes that are involved in the dopamine and serotonin system[12] in the brain's reward pathways represent primitives of behavior (e.g., Dreber et al. 2009; Kuhnen and Chiao 2009).[13] While these early studies found some statistically significant associations, they were not replicated in samples of adolescents (Gee 2014) and other samples analyzed by Carpenter et al. (2011) and Dreber et al. (2011).

Associations between candidate genes and other socioeconomic outcomes have been undertaken in situations where the genetic basis for variation in outcomes was not established. For instance, DeNeve and Fowler (2014) and Kuhnen et al. (2013) respectively explore if there are statistically significant links between specific genetic markers and credit card debt and the number of credit lines opened. Since decisions on i) whether or not to issue a credit card, and ii) the credit limit on a given credit card, are both made by lenders and not by individuals themselves, whether there is a biological plausible mechanism underlying any association should be justified. After all, in most candidate gene studies within genetic epidemiology, researchers explicitly explain how the putative candidate gene was chosen based on its relevance in the mechanism of the trait being investigated and it does not appear to have an ex-post justification.

In summary, studies that fall under the heading of candidate genes are likely undertaken based on convenience and have a poor track record when it comes to replication. Candidate gene studies also face concerns that they lack statistical power. Intuitively, if well-powered studies that search the entire genome for associations find only tiny effects, then the large effects found in many of these candidate gene studies with much smaller sample sizes are likely false positives.[14] We believe that despite the ease in which this research can be undertaken, candidate gene studies are unlikely to convince many in the research community.

An under-investigated aspect of candidate gene studies is whether the inclusion of genetic information changes the effects of other covariates. After all, if genetic factors are important, does their inclusion change estimates of other coefficients? In other words, is bias from omitted variables reduced and/or are certain covariates proxying for genetic factors? Answers to this question are important in understanding whether molecular genetic information is truly a valuable addition to many datasets.[15]

3.2 Moving beyond associations: genetic markers as instruments

Perhaps the area that has attracted the most amount of debate among economists is whether or not genetic data can provide a source of exogenous variation to identify the impact of specific health conditions on socioeconomic outcomes. This source of identifying variation was first introduced in economics by Ding et al. (2009) who essentially used candidate genes as instruments to understand the impact of health outcomes on academic performance. Ding et al.'s (2009) analysis underscores both the challenges researchers face when using genetic information as an instrument for specific health conditions and the need to investigate the sensitivity of one's conclusions to the identifying assumptions. We discuss these issues in further detail.

The concept of comorbidity is well known in the medical sciences and is defined as being the simultaneous presence of two chronic poor health conditions in an individual. In

empirical research, we are not provided with a single accurate measure of an individual's health but rather proxies such as specific diagnoses. In their analyses, Ding et al. (2009) show that using richer vectors of health conditions is important to understand the effects of a specific condition since poor physical and mental health conditions are positively correlated. By omitting the comorbid condition, different estimates may arise when using different estimators and specific instruments. This presence of the challenge of comorbidity has large implications for researchers aiming to identify the role of a specific health conditions and was first pointed out due to this investigation using genetic information. More generally, comorbidity also influences the general ways in which applied researchers select their instruments based on first stage relevance and whether they meet the exclusion restriction criteria.[16]

As with all studies that use instrumental variables to identify causal parameters, the plausibility of the (genetic) instrument comes into question. To a large extent, one will never know whether a specific candidate gene is a valid IV since one cannot randomly assign genes to humans or create human equivalents to knockout mice. In addition, the role of individual genetic markers in many socioeconomic outcomes is likely quite small and likely explains less than 1% of the variation in that phenotype. This suggests that individual markers are likely weakly correlated.[17] Further, in the presence of dynastic effects and without more detailed data on parental outcomes and family environments (as well as parental genes), we cannot separate out the portion of the impact that is uniquely brought on by the child's outcome.

Turning to the genetic marker itself, one may worry about population stratification[18] that there are subtle genetic differences between groups of individuals that are not accounted and the gene being investigated is correlated with a missing genetic marker that is driving the results. Similarly, this may happen since genes located close together on the same chromosome are sometimes inherited as a group, so one may not be attributing the effect to the correct polymorphism. Given these potential threats, researchers using genes as IVs should use the Conley et al. (2012) local to zero approximation sensitivity analysis.[19]

Conley (2009) points out that the phenomenon of pleiotropy presents a related challenge for the plausibility of a genetic instrument: since many genes code for proteins that may have multiple functions and effects, it is hard to know for certain that the instrument only affects outcomes through the endogenous regressor. Naturally, without random assignment, one may never be certain about the role of any specific genotype so this reinforces the need to investigate the robustness of results.

Among economists that use genetic markers as instruments, there are major differences in how these variables are included in the first stage. Ding et al. (2009) used a series of binary variables for each potential genetic polymorphism in the genes they investigated. A potential concern is that many instrument problems (Hausman et al. 2012) may result and, to date, no research has investigated using the LASSO in the first stage. Other researchers rather than use a set of discrete binary indicator variables choose to treat the genetic information as a continuous variable and include the count of the number of risk alleles. We suggest that using a count variable is not only more challenging for researchers to interpret first-stage relationships and assess if they are consistent with the scientific literature, but this additionally imposes a strong functional form relationship that first-stage outcomes are linear in the number of risk alleles. We would argue that by allowing for

nonlinear relationships through discrete indicator variable, one can easily test whether the linearity restriction is supported by the data. Second, the discrete variables truly shed more light on what features are driving the estimated effect and one can then get a better handle on if the relationships mimic those hypothesized in the scientific literature.

Studies that use genetic markers as instruments generally draw biological justification from results of published candidate gene studies, which as discussed are controversial.[20] The journal *Behavior Genetics* recently adopted strict standards for publication of candidate gene studies (Hewitt 2012). To be considered for publication, any candidate gene study must be well powered and make corrections in statistical inference for multiple testing and any new finding must be accompanied by a replication.[21] Thus, when searching for a plausible genetic instrument by reviewing the literature, researchers should also justify their choice by considering the statistical power of the study.

The idea of using genetic information as a source of identifying variation also appears in the epidemiological literature where it is termed Mendelian randomization. Mendelian randomization was first proposed in Katan (1986) and applied with data in Davey Smith (2003). While not made explicit, studies using Mendelian randomization implicitly assume that there are no dynastic effects to invoke the term randomization. However, genes are inherited by design from one's parents who also transmit environments and numerous behaviors across generations. In effect, empirical economists can draw a parallel between the Mendelian randomization research design and the econometric analysis of a randomized experiment with noncompliance. Thus, under the assumption of no dynastic effects, Mendelian randomization is an encouragement design, and while randomization (experiments) is often associated with being the gold standard in medical research, we would suggest that these studies be more accurately recast as being a Mendelian encouragement design.

A final variant on the instrumental variable strategy was introduced by Fletcher and Lehrer (2009a, b, 2011) who exploit genetic inheritance within full biological siblings. Fletcher and Lehrer rename a family fixed effects instrumental variable estimator with genetic instruments for this sample as being the genetic lottery. This genetic lottery might truly be what is meant when one claims Mendelian randomization, since by controlling for the family fixed effects, one removes the dynastic effects (assuming they are constant) between full biological siblings. This strategy exploits variation in genetic inheritance and socioeconomic outcomes between full biological siblings and provides a means to test a key identifying assumption in a workhorse research design used in family and population economics that has been applied in almost every branch of empirical economics as well as behavioral genetics. That is, does the family fixed effects estimator fully solve the underlying endogeneity problem? By modifying the traditional Hausman test to compare a family fixed estimator to estimates from a family fixed effects IV (aka genetic lottery), one can find evidence that either refutes or is unable to reject the maintained assumptions. In each of their applications, Fletcher and Lehrer are able to reject that the family fixed effects estimator does not fully solve the endogeneity problem in health when estimating its effects on academic and early labor market outcomes. While labor economists have made substantial advances at estimating causal relationships, we believe that genetic information may hold more hope at identifying causal mechanisms, a topic we elaborate upon in our discussion of gene–environment interactions.

3.3 Economists' replication scientific studies: genome-wide association studies

Whereas research by economists treating genetic markers as instruments displays a new use of these data, economists have also ventured into using research methods common to medical science and geneticists. This work led by economists who established the Social Science Genetic Association Consortium (SSGAC) involved the development of large networks of researchers and pooling of multiple datasets containing genetic information.[22] The primary aim is to conduct large-scale genome-wide association studies on a number of training data sets and examine if the results replicate in other studies. Such analyses would provide robust evidence of the molecular genetic basis of outcomes of interest to economists, and this work strives to overcome many criticisms of candidate gene studies.[23]

To date, the best example of research in this strand of literature appears in Okbay et al. (2016).[24] The authors conduct a genome-wide association study (GWAS) of about 300,000 people (this is called the discovery sample)[25] and find 74 SNPs associated with educational attainment, where educational attainment is the amount of formal education completed. In aggregate, these 74 SNPs explain only 0.43% of the variation in educational attainment across individuals in the sample. The economic significance of each of the individual 74 SNPs is found to be quite small, since when comparing individuals with zero to two copies of the risky allele of the genetic variant with the reported strongest association is shown to predict (on average) roughly nine extra weeks of schooling. However, what is striking about the results from this study is that authors additionally conduct a replication with 110,000 individuals from the UK Biobank. In the replication, they remarkably find that 72 of the initially identified 74 SNPs remain significantly associated with educational attainment. Thus, they are confident they have identified the molecular genetic basis for educational attainment.

Okbay et al. (2016) are quite cautious in how one should interpret their findings since years of educational attainment is a complex phenomena and one cannot separate if these genes are truly related to educational attainment or do they explain the selection process that led one to complete more schooling.[26] Since there are more hypotheses of significant association than data points, one must make corrections for multiple testing, and they are careful to use an independent sample for the replication study. The authors take great care to convince the reader that the observed associations are unlikely to be spurious by both utilizing the latest quality control protocols in the medical genetics literature (Winkler et al. 2014) and carefully account for population stratification in their analysis. Specifically, the authors conduct a robustness check of their main analyses where they (i) ensure common support is imposed across samples by excluding dissimilar individuals, (ii) accounts for high levels of principal components as additional controls to capture potentially confounding genetic differences across samples,[27] and (iii) includes family fixed effects in the analysis.[28] This paper provides a comprehensive guide on how to undertake and report results from a GWAS.

Many economists' first reaction to a GWAS is that it is simply a data mining. After all, these studies are not motivated by any theory of why specific SNPs are being investigated and simply examine for an outcome of interest, whether it is associated with one or more of the (typically millions of) measured SNPs. Further, genetic researchers are generally solely interested in characterizing the variance of estimates of how much SNPs influence outcomes and point estimates are generally not the focus. While it is

often interesting to discover what percentage of the variation an individual SNP can account for, this is not how economists generally determine the relative importance of explanatory variables in wage regressions. In response, Rietveld et al. (2013a, 2014) suggest examining polygenic scores in future research and Papageorge and Thom (2016) present an early application of these scores in labor economics.

A polygenic score is constructed by adding up the individual alleles that are reliably related to this trait, where each allele is weighted by effect sizes estimated from a GWAS (Dudbridge 2013). The underlying idea is that from the GWAS results, we can give weights of relative importance to each SNP. Then, with a polygenic score, a researcher could exploit the joint predictive power of many SNPs when used as an input in an estimating equation. As an explanatory variable, these polygenic scores will explain more variation than individual SNPs and can provide clearer role on some combined genetic influence. The scores provide a means to identify individuals at high risk for certain outcomes. From an econometric perspective, this may reduce the chance of including irrelevant variables in a regression model and the resulting efficiency of estimates but comes at a cost of placing strong behavioral assumptions on the components of the score.[29] After all, the score is just a linear combination and implicitly makes assumptions about relative substitutability of effects of different SNPs.

GWAS research with replication samples is valuable to establish robust evidence of a main genetic effect. With this in place, studies using these specific genes either in a candidate gene approach or as an instrument would likely face significantly less opposition. Yet, we should point out that evidence of main effects from these large-scale GWAS require the genetic variant to have a similar effect across all samples, which likely differ on the basis of the environment and sample characteristics. It is reasonable to assume that specific genetic variants may only have significant effects in particular environments or with specific types of samples and be insignificant in all other cases. If either of the above scenarios hold, then a standard GWAS would never identify the main effect of this variant, despite the fact that there may be strong evidence of a significant heterogeneous impact of this variant on the environment.[30] Thus, the rationale for investigating gene by environment interactions is different, yet the need for replication across similar contexts is clearly of importance to provide robust evidence of interactive effect of genetic variants.

3.4 Gene by environment interactions

Recall that research in behavioral genetics began by assuming the absence of gene by environment interactions, henceforth G^*E. This assumption is now clearly rejected and researchers across a multitude of disciplines champion the importance of G^*E effects. Among labor economists, James Heckman is perhaps best known for arguing of the importance of G^*E effects in his arguments designed to convince policymakers to invest early in child development.[31]

To explore G^*E effects requires rich longitudinal data with clean variation in environmental exposure to interact with genetic factors. Modelling G^*E effects requires either exogenous variation in environmental factors or a clean econometric strategy that can identify unknown breakpoints in relationships between genetic factors and outcomes.[32] Rosenquist et al. (2015) undertake the latter approach by using the threshold regression estimator by Hansen (1999) to estimate an augmented version of a linear age–period–

cohort model to understand the source of *G*E* with longitudinal data collected between 1971 and 2008 in the offspring cohort of the Framingham Heart study. Specifically, they test whether the well-documented association between the rs993609 variant of the FTO gene and body mass index (BMI) varies across birth cohorts, time period, and the lifecycle. These models partition the time-related variation in obesity to the three distinct sources but cannot identify the specific environmental channel within the source.[33] A key feature of the analysis is statistically testing for a structural break of unknown timing across cohorts and checking the robustness of their finding by additionally controlling for family fixed effects. The selected breakpoint is based on the model that best fits the data using a grid search algorithm.[34]

Rosenquist et al. (2015) find that there is a robust relationship between birth cohort and the FTO risk allele with BMI, with an observed inflection point for those born after 1942.[35] Specification tests of the unrestricted model that additionally control for gene*cohort effects and gene*age effects provide evidence that the inclusion of gene*contemporaneous period effects is statistically insignificant. Only if one was to ignore gene*cohort effects, they would find evidence that *G*E* effects are due to contemporaneous events for FTO and BMI. Upon reflection, this result is unsurprising since environments are highly correlated over the lifecycle for most individuals and there is limited variation in environmental conditions experienced to affect the penetrance of genetic influences.

The results also have important implications for how one interprets evidence from GWAS that pools data across samples. As discussed in those studies, one accounts for population stratification but the findings in Rosenquist et al. (2015) raise the possibility that genetic associations may differ across birth cohorts due to variation in prevailing environmental contexts.[36] Thus, the low replication rates of many genome-wide association studies may also be due to differences in the period of time study subjects were born in and the historical moment researchers conduct their investigations, suggesting perhaps the need for environmental stratification.

To date, the majority of work by social scientists evaluating *G*E* effects does not explicitly consider the endogeneity of the environmental variables that were selected by the individual. Perhaps the best example of research in this stream is Biroli (2015) who situates his analysis within an economic framework.[37] Biroli (2015) integrates genetic factors inside the canonical model of health production due to Grossman (1972), allowing genetic variants to both potentially differentially affect the health production function and preferences related to the incentives related to health investment faced by individuals. Using data from both the Framingham Heart study and Avon Longitudinal Study of Parents and Children, he finds evidence that genetic factors do change both the production function of BMI and the level of healthy investment. While this work extends our analysis of a workhorse model in health economics, the empirical analysis requires one to assume that caloric intake is exogenous and not a behavioral choice; otherwise, biased coefficients may result.

Studies that have tried to exploit genetic variation within families have the potential to provide more compelling evidence of candidate *G*E* effects. Similar to Fletcher and Lehrer (2011), the idea is to exploit within family differences in genetic code to remove biases from dynastic effects. For example, Thompson (2014) exploits within-family variation in genetic inheritance, to see if there are differential responses of household income on child education outcomes by variants of the MAOA genes. The results

indicate that the gradient is steeper for those with rarer variants.[38] However, Conley and Rauscher (2013) advise caution when interpreting evidence from studies estimating *G*E* effects exploiting within-family variation. This caution arises from their independent exploration of how genetic traits moderate the relationship between birthweight and several outcomes including high school GPA that exploits within twin-pair birthweight differences. The sole statistically significant *G*E* effect reported has a sign that is the opposite of what had been suggested by prior scientific research.

Economists are well aware of the benefits of comparative advantage. Thus, one can interpret the set of guidance provided in Conley and Rauscher (2013) as indicating with biomarker data that there are potential benefits from using an interdisciplinary research team, where a subset of members are better-equipped to assess whether the estimated effects are plausible in sign and magnitude.[39] Hatemi (2013) provides an illustration of a policy relevance of *G*E* effects by exploring how proximate events such as losing one's job, suffering a major financial loss, or getting a divorce can lead to a short-term change in one's economic policy attitudes that are consistent with maximizing self-interest. The results suggest that there are differential responses by genetic markers across individuals who lost a job, which are more likely to oppose policies that may have caused their change in economic situations such as immigration and capitalism.[40]

Dealing with potential environmental stratification is explicitly yet indirectly considered in what is known as genome-wide complex trait analysis (GCTA), which is a variant on the behavioral genetic approach to measure heritability between genetically dissimilar individuals.[41] Yang et al. (2010, 2011) suggest that genetic similarity between two individuals is essentially estimated as a weighted correlation of their genotypes on the included SNPs and the goal is to restrict the analysis to unrelated individuals. This restriction is motivated on the strong assumption that individuals who are more genetically related share a more similar environment than unrelated individuals. Using a restricted maximum likelihood estimator (commonly referred to as REML in the literature), one can obtain estimates of heritability without resorting to twins' data.[42] However, the identifying assumption appears strong to individuals trained as labor economists, and it is plausible that one can develop tests similar to understanding whether selection on observables leads to balance. We discuss further econometric and computational directions for economists in the next section.[43]

4 Potential future areas for economists to analyze genetic data

We suggest that there are, at a minimum, three main directions that economists can contribute to the literature using genetic data. First, economists can provide an understanding of the genetic mechanisms in a logically consistent framework. Economists should not fall victim to thinking that just because one's genetic code is fixed at conception, genetic-expression may vary across the lifecycle. Labor economists treat fixed characteristics such as gender and race as having time-varying effects in empirical analysis. Research is needed to understand whether polygenic scores are time varying and to what extent they are truly capturing a portion of what economists in longitudinal analysis refer to as permanent individual-specific unobserved heterogeneity. Clarifying what is meant by a genetic disposition is an area where economists can contribute strongly.

Further, and as Manski (2013) argues in his view of the incredible certitude taken by many in the public policy community, economists should help other research

communities become comfortable with embracing uncertainty in how genetic effects operate. To an extent, the work of the SSGAC evaluating genome associations is striving to reduce the degrees of certitude represented in any given candidate gene study. Future research if well-powered can also be used to understand why across the databanks collected by the SSGAC, divergent main effects are observed. That is, if genes influence one's behavior, can one's behavior also influence genetic expression? Indeed, if the genetic mechanisms are not activated or repressed by certain stimuli, then genotype at that given moment may not be relevant. However, specific life events may alter the magnitude of genetic expression on a given trait, or it may in fact instigate different genetic processes altogether.

Building on the above point that suggests potential environment interplay is something that should be emphasized more strongly in the genetics literature, a second area where economists can contribute is by developing tools and research designs to shed new light on the pathways through which genetic factors influence socioeconomic outcomes. Lehrer (2016) suggests that researchers should consider working with more aggregated environmental factors and perhaps exploit regional environmental changes. Indeed, there is a large history in empirical microeconomics of exploring differences in environmental conditions or policies across regions such as natural experiments and exploring genetic heterogeneity in the estimated effects which seem to be a simple extension. An early example is Okbay et al. (2016a) who compare cohorts prior to and post a suite of schooling reforms that, most importantly, extended mandatory schooling from 7 to 9 years. The authors find that the association between educational attainment and the polygenic score constructed from their GWAS is roughly half as large among Swedish individuals in later cohort, suggesting that the Swedish reforms reduced the effects of genetic variants in generating differences in educational attainment.

With genetic data, it may be possible to isolate biological mechanisms to shed light on why treatment effect heterogeneity is observed. To a large extent, researchers in empirical microeconomics already have sets of tools to explore whether interventions have different effects for subgroups defined on the basis of more aggregated predetermined characteristics such as gender and race.[44] Further, it will likely become necessary to move beyond linear models to study G^*E effects, and Conti and Heckman (2010) provide a more general framework to operationalize and interpret gene–environment interactions.

As a whole, there is tremendous scope in this stream for both empiricists and econometricians to collaborate and develop methodological tools for G^*E analyses and Lehrer (2016) also points out there may be a serious identification challenge. Researchers currently use G^*E to both describe situations where the effect of exposure to an environmental factor on a behavior is conditional upon a person's genotype, as well as situations when the genotype's effect is moderated by some environmental effect. While statistically separating these pathways is needed, policy audiences do need to understand what is being identified. Lehrer (2016) suggests that researchers use *G^*E responses* to refer to situations where the effect of exposure to an environmental factor on a behavior is conditional upon a person's genotype and *G^*E modifications* to refer to differential genetic reactions to environment. Personalized medicine and many policies that would target by genotype may be interested in G^*E modifications, whereas G^*E responses may be more interesting for researchers to study if they are interested in understanding the heterogeneity in environmental effects on outcomes across the

population. In summary, by improving methodological tools, more credible evidence from rigorous G^*E studies can be obtained which could subsequently reshape theories and they improve our understanding of the complex pathways that lead to various health and socioeconomic outcomes,

Third, economists need to shed light on the behavioral restrictions implicitly imposed by empirical methods used to both elucidate genetic associations and construct polygenic scores. Consider how Todd and Wolpin (2003) influenced researchers in the economics of education by highlighting the behavioral restrictions on an underlying model of human capital development that were implicitly made by researchers when estimating equations that proxied for these education production functions. Analogously, the socioeconomic outcomes being investigated in both candidate gene and GWAS studies are likely determined. As discussed in the section considering genes as instruments, should risk alleles enter these estimating equations as a count assuming a linear effect or as a series of indicator variables? The consequences of using imputed versus actual SNPs also require further evaluation. Further, it is worth stressing that much of the existing analysis in the scientific literature uses canned software that itself imposes additional assumptions on the underlying process generating the outcomes.

A number of potential methodological questions are worth considering related to GWAS including whether these equations be estimated for different health outcomes independently or as a system of equations framework allowing for correlations in the residuals? Similar to evidence of the importance of comorbidity in Ding et al. (2009), Boardman et al.'s (2015) GWAS investigation of the molecular basis of education and depression/self-rated health points out that one may wish to disentangle whether a given marker has an independent influence on outcomes or mediates the effect of these correlated outcomes on one another. It may also be interesting to explore the use of the LASSO estimator for GWAS studies as a means to shrink the variable set in place of REML estimators. Other directions include developing an optimal way to make corrections for multiple testing in settings where there are potentially more covariates (SNPs) than observations.

Labor economists have done much work developing methods to estimate both cross-sectional and panel data models with repeated cross sections. To an extent, work on GWAS is pooling many samples that did not choose individuals in the study via random sampling. In addition, between the different datasets, members of the population have an unequal probability of being observed. These sampling issues are normally not considered other than creating a form of balance in the samples by adding controls for population stratification. Issues related to nonparametric identification of population parameters in this setting when many choice-based samples are combined (and as noted earlier, Rosenquist et al. (2015) point out differences across environments in these studies is not considered) seem important to correctly interpret the resulting estimates. Further, given the combination of these nonrandom samples, there may be methods to conduct efficient estimation from a combination of biased samples. While much work has been done across disciplines on topics related to combining nonrandom datasets, the mix of types of datasets used in GWAS which range from case–control studies to random samples likely requires researchers to combine results to help improve both the estimation and interpretation of results from GWAS.

Turning to the construction of polygenic scores,[45] should these be anchored in a metric that has economic content such as earnings? How should researchers account

for estimation error in these scores when including these measures as explanatory variables? Is there a partial identification approach to calculate polygenic scores? Labor economists have worked as applied econometricians for decades, and as one becomes more familiar not with just genetic data but with the methods used to measure and analyze this data, we believe there is a great scope to develop refinements that will have impacts not solely on the economics literature but on other disciplines that analyze these data.

Finally, as the scientific literature is also now moving beyond only considering main genetic effects, it is worth pointing out that gene–gene interactions almost certainly do exist.[46] Indeed, both Ding et al. (2006, 2009) and Fletcher and Lehrer (2009b, 2011) consider such two-way interactions in their instrument set, but there is not much information even in the behavioral genetics literature on how and why these interactions operate. In other words, understanding the genetic architecture of a particular trait is one of the main goals and this challenge mirrors the steps required when labor economists create empirical models to understand the underlying data generation process. With newer and richer data, future research will be able to additionally explore the interactive effects between genes themselves as well as with environmental interactions and genetic networks. With the likely continued increasing focus of labor economists at understanding the origins of economic inequality, we believe that molecular genetic data may help shed new light on understanding the sources of unobserved heterogeneity.

5 Conclusions

Many labor economists including Goldin (1999) have described their research strategy as first finding a topic that one is passionate about and then being the best detective one could be. Indeed, labor economists have for painstakingly long analyzed data not solely to help reveal trends and patterns but also to shed new light on drivers of human behavior. There is no question that genetic factors do play a role in nearly every socioeconomic outcome of interest to empirical economists and only recently have we begun to develop affordable and reliable technologies to measure this individual-level variation.

Over the past decade, a growing number of economists have begun to incorporate genetic markers in their empirical analyses. This area is quickly maturing, and it is likely that many of the low hanging applications of genetic data have already been undertaken. At this stage, and similar to trends within labor economics, we are witnessing a shift towards researchers using much larger datasets to asses genetic associations as well as researchers developing new econometric tools to understand heterogeneity in genetic effects. These trends parallel those within labor economics where a growing number of studies are relying on the use of rich administrative databases to draw credible evidence as well as the development of econometric tools to shed light on treatment effect heterogeneity.

Yet, there is likely much more that can be done particularly along entering the black box of individual-specific unobserved heterogeneity and exploring gene–environment interactions. Evidence from these studies can also be utilized to help refine theories and help. While the idea of developing a separate field within economics called genoeconomics is clearly appealing to those in the area, we believe that there is more potential from incorporating genetic data within existing fields such as labor economics. For example, in understanding educational attainment are genetic factors related to selection effects associated with education or with the true effect of achieving additional education.

We should caution that there are high start-up costs for researchers trained as economists in understanding genetics research, in part since research in other disciplines use different complex jargons and are generally less explicit about the implicit behavioral assumptions.[47] That said, there are likely high returns for economists to additionally develop richer empirically tractable models to investigate the role of genetic factors that can weaken the maintained behavioral assumptions. In summary, we are bullish on the future of genetic markers in economics but believe that their success will be achieved in a quicker fashion if their use is accompanied by solid scientific understanding of the mechanisms through which they influence socioeconomic outcomes.

Endnotes

[1]A national media frenzy erupted focusing on an explanation that this underrepresentation may stem in part from "issues of intrinsic aptitude" differences between men and women, without considering the context in which the remarks were made. The remarks made at the NBER Conference on Diversifying the Science & Engineering Workforce are posted online at http://www.harvard.edu/president/speeches/summers_2005/nber.php.

[2]Despite the hysteria in academic debates at the time concerning this book and the link between IQ and genes, the authors (on page 311) made clear that while this explanation may hold water, they had no idea as to the importance. Specifically, (on page 311) in the concluding section, they write, "If the reader is now convinced that either the genetic or environmental explanation has won out to the exclusion of the other, we have not done a sufficiently good job of presenting one side or the other. It seems highly likely to us that both genes and the environment have something to do with racial differences. What might the mix be? We are resolutely agnostic on that issue; as far as we can determine, the evidence does not yet justify an estimate".

[3]For example, d'Alpoim Guedes et al. (2012, 2103) present critiques against claims within the paper and Ashraf and Galor (2012) replied to the first critique.

[4]Jencks (1980) may have been the first to point out that "genetic" does not imply "immutable." Thus, it may be the case that the effects of specific portions in an individual's DNA sequence on specific outcomes vary over the lifecycle, perhaps due to environmental stimuli. In a traditional fixed effects estimating equation, both the impact and tock of unobserved heterogeneity are assumed to be fixed over the period in which data is being analyzed.

[5]We should point out that while not discussed due to space constraints, there are many other approaches used to estimate heritability with data on twins under alternative assumptions including regression-based methods (e.g., DeFries and Fulker 1985), structural equation models (Boker et al., 2011), and generalized linear mixed models (Rabe-Hesketh et al., 2008).

[6]Jensen (1967) previously conducted a study published in a general interest scientific journal that tried to isolate the role of heredity, environment, and luck in earnings.

[7]More recent research has shifted from using data collected in traditional surveys to using data from either incentivized experiments or surveys, to explore the heritability in different measures of economic preferences (e.g., Wallace et al. 2007; Cesarini et al. 2008, 2009, 2010, 2012). See Kohler et al. (2011) for a discussion of how to leverage twin studies to model unobserved genetic endowments and causal pathways. Last, we discuss genome-wide complex trait analysis (GCTA) a method that uses restricted

maximum likelihood estimation to estimate heritability from molecular genetic data in the section on gene by environment interactions.

[8]Other sources of genetic variation are due to mutations affecting repeated segments of DNA and include what are known as variable number of tandem repeat polymorphisms and copy number variation (CNV) polymorphisms. Since these will not be discussed further, the interested reader is referred to a molecular genetics text.

[9]There are methods that target the whole genome and others that are more targeted. In general, the resulting data quality is heavily dependent upon the average number of times each base in the genome is actually "read" during the sequencing process.

[10]The consequences of imputation have received scant attention, and this clearly generates measurement error, which is something that labor economists have a rich set of tools to offer to other researchers.

[11]Specifically, Greiner and Rubin (2011) focus on circumstances under which race/gender can be appropriately called treatments. They argue that what causally explains gaps in outcomes between groups are not the groups themselves but rather are perceptions of the groups. In a genetic marker context, similar arguments could be made if employers or health insurers make decisions based on perceptions of the genetic characteristics of workers. While laws do exist in many countries including the Genetic Information Nondiscrimination Act (GINA) in the USA prohibit employment discrimination based on genetic information and forbid employers from asking about individuals' genetic information, including information about family members' health status, or family history, there are many reports that individuals can voluntarily provide this information to help in the development of workplace-based wellness programs that Baicker et al. (2010) made a case would provide large benefits to employers.

[12]Dopamine and serotonin are two powerful neurotransmitters that affect one's mood and happiness. In general, neurotransmitters are chemical messengers which neurons use to tell other neurons that they have received an impulse.

[13]See Knafo et al. (2008), Mertins et al. (2011), and Zhong et al. (2009) for studies that link specific genetic variants to outcomes measured in the laboratory.

[14]Chabris et al. (2012) discuss the importance and challenge of replicating results from studies using genetic data since this is needed to verify that the reported associations are not false positives.

[15]As we will discuss later, research focused on identifying genetic associations with outcomes of interest to economists has moved from using data collected on a few candidate genotypes to those measuring variation across the full genome. These studies have also proposed calculating genetic risk scores which are single variable measures that capture information contained in a multitude of SNPs. These covariates may help provide some preliminary information on the value of genetic information as a covariate but present difficulties in their interpretation.

[16]Using genes as instruments has been subject to criticism as outlined in Cawley et al. (2011) and Fang (2013), among others.

[17]For example, Wehby et al. (2011) use two independent samples from Norway and the USA to conduct IV analyses and find weak correlations between maternal smoking and the genetic variant instrument sets.

[18]We discuss the term population stratification in further detail in the section on genome-wide association studies, but the general idea is that there might be systematic

differences in the frequency of risky alleles between groups, thereby leading to a form of omitted variable bias.

[19]This analysis involves making an adjustment to the asymptotic variance matrix, thereby directly affecting the standard errors. That is, a term that measures the extent to which the exogeneity assumption is erroneous constructed from prior information regarding plausible values of the impact of genetic factors on second-stage outcomes is added to the variance matrix.

[20]The scientific literature is populated with conflicting findings from candidate gene studies, and many early studies failed to replicate since, initially, researchers did not adjust for population stratification. Further, studies in the literature suffer from low statistical power and coupled with potential publication bias as well as undisclosed pretesting which could have led to too many false positives appearing in press.

[21]Chabris et al. (2013) illustrate several points related to the limits of candidate gene studies by trying to replicate previously identified candidate genes using data from three independent longitudinal studies. Their results are disappointing from a replication perspective since they found fewer significant associations than a traditional power analyses would have ex ante predicted.

[22]The economists who established the SSGAC are Dan Benjamin, David Cesarini, and Phil Kollinger, and the work of the SSGAC is characterized by its ambition and may have been motivated by The Wellcome Trust Case Control Consortium attempts to improve the understanding of the aetiological basis of several major causes of global disease by pooling databases collected by individual research teams. This approach has yielded important findings in the medical sciences, particularly in understanding the genetics of autism (Glessner et al. 2014) and schizophrenia (Ripke et al. 2014).

[23]The declining cost of genotyping and technological advances include the availability of canned software packages to do the analyses also likely played a large role in their growth. See McCarthy et al. (2008) among others for early examples of work in this area. Other work involves using what is termed genomic-relatedness-matrix restricted maximum likelihood (GREML) that for a sample of unrelated individual pairs estimates what portion of the total fraction of variance in a trait is attributable to the average effects of SNPs. That is, does genetic similarity predicts phenotypic similarity? We return to how genetic similarity is measured in the section on gene by environment interactions.

[24]Two examples of earlier papers by the SSGAC include Rietveld et al. (2013a) who combined data on 42 cohorts providing over 100,000 individuals to study which of approximately two million single nucleotide polymorphisms influences measures of educational attainment such as college completion and years of education. This research suggested three specific genetic variants. Subsequently, Rietveld et al. (2014) verified the robustness of these findings using data from three new sources, as well as used exploited only genetic variation within families.

[25]The discovery sample pools numerous datasets and contains information from participants in 15 different countries.

[26]Instrumental variable estimators of the effects of years of schooling generally identify the causal effect of years of schooling only for the subsample whose behavior was influenced by the instrument. A popular example is compulsory schooling, and often, the resulting estimate compares individuals with 11 to 12 years of schooling. At

present, with GWAS, we do not know where in the decision process the individual markers operate on individual behavior.

[27]Since data is pooled from different studies, the principal components off the gene chip (i.e., the correlation matrix of all the assayed SNPs) are measured. To control for population stratification, generally, the first four of these components are used to identify geographic ancestry within the sample.

[28]By exploiting variation within siblings, one controls for dynastic factors and any differences in genetic factors do not come from differences in sample composition. Since there is less variation and a smaller sample size, the effects are noisier relative to the discovery sample but the effect sizes are remarkably similar on average, enhancing confidence in the initial GWAS results.

[29]Debates about the relevance of polygenic scores exist outside of economics. Purcell et al. (2009) list concerns on their likely usefulness, whereas Belsky et al. (2012, 2013) are empirical examples illustrating potential benefits.

[30]We are grateful to Pietro Biroli for the discussion that clarified why gene by environment interactions should not solely be motivated by results from GWAS.

[31]Specifically, Heckman (2007) writes "Third, the nature versus nurture distinction, although traditional, is obsolete. The modern literature on epigenetic expression and gene environment interactions teaches us that the sharp distinction between acquired skills and ability featured in the early human capital literature is not tenable (Rutter, (2006), Gluckman and Hanson (2005), Rutter et al. (2006)). Additive "nature" and "nurture" models, although traditional and still used in many studies of heritability and family influence, mischaracterize gene-environment interactions. Recent analyses in economics that break the "causes" of birthweight into environmental and genetic components ignore the lessons of the recent literature. Genes and environment cannot be meaningfully parsed by traditional linear models that assign unique variances to each component. Abilities are produced, and gene expression is governed by environmental conditions (Rutter, (2006), Rutter et al., (2006). Behaviors and abilities have both a genetic and an acquired character. Measured abilities are the outcome of environmental influences, including in utero experiences, and also have genetic components."

[32]Fletcher and Conley (2013) argue that G^*E interactions are most meaningful when they are based on exogenous environmental measures that are not themselves a function of genes. Pushing further, van IJzendoorn and Bakermans-Kranenburg (2012) advocate using randomized controlled trials to study how environmental changing interventions have differential effects as a function of genetic endowments.

[33]In other words, aggregate macro-environmental conditions and not person-specific conditions as their environmental influences are explored. Due to the data being collected in one small geographic area, biases due to sorting across regions based on environmental conditions due to unobservables are reduced.

[34]This breakpoint does not necessarily mean that the relationship is most different but rather this is a point where the difference between birth cohorts explains the most variation in the data.

[35]Consistent with Rosenquist et al. (2015), Biroli (2015) finds that the estimated interaction between the FTO genotype and caloric intake is stronger for individuals born in later cohorts.

[36]This criticism is not viewed favorably among geneticists and is more natural to economists who understand that GWAS just reports associations and not causal or structural parameters.

[37]As with candidate gene studies, concerns related to low statistical power due to a combination of potential pre-testing and publication bias likely hold some validity in what we will term candidate gene*environment interactions. For example, Caspi et al. (2002) report the effects of self-reported childhood maltreatment on adolescent anti-social behavior varied based on one's MAOA gene. Shanahan et al. (2008) report in explaining educational outcome that there is a significant interaction between a variant of the DRD2 dopamine receptor gene with factors such as having a parent that belongs to the PTA and how often parents discuss school related issues with the student.

[38]Thompson (2014) also points out that parents may make "compensating" investments in which more resources are allocated to the less able sibling to promote equality. Thus, one cannot rule out with the data that MAOA variants are correlated with the environmental conditions children receive from their parents after conception. Future research is needed to see if a child's MAOA status induces differential treatment from their parents' investments in their children's human capital and, if so, to what signals of MAOA status do parents respond to, given that they are unlikely to have genotyped their children.

[39]There are numerous examples of successful interdisciplinary collaborations reviewed in this chapter including the multiple papers produced by the SSGAC that was lauded in an editorial *Nature* (Hayden 2013), Rosenquist et al. (2015), among others.

[40]This analysis implicitly assumes that the changes in personal circumstances are exogenous and do not decompose the variation in the sources of G^*E effects between proximate and global distal environmental factors since birth. Smith et al. (2012) suggest this decomposition may be important since distal factors have little effect on genetic or environmental variance component estimates of political attitudes, but additionally note that the evidence base is very weak.

[41]Benjamin et al. (2012a, b) use this approach in their analysis to explain heritability of economic preferences.

[42]As an example, Rietveld et al. (2013b) point out that while twins' studies suggest that genetic factors may account for as much as 30–40% of the variance in subjective well-being measures, additive effects of genetic polymorphisms that are common in the population can only explain between 5 and 10% of the variation in these measures. While subjective well-being measures are not accurately measured, if one accounts for measurement error in this analysis, it was found to only increase the amount of explained variation of additive genetic effects to range from 12 to 18%.

[43]More recent developments for sequencing and linkage analysis that have been introduced in the genetics literature include Ott et al. (2015) and Pabinger et al. (2014), but to the best of our knowledge have yet to be used by economists.

[44]For example, Lee and Shaikh (2014) and Lehrer et al. (2016) provide a set of methodological tools to analyze heterogeneity in causal effects that can additionally incorporate corrections for multiple testing.

[45]At present, most researchers who construct these scores rely on canned software routines such as PRSice for convenience and do not directly discuss the statistical and behavioral restrictions embedded.

[46]See Lazopoulou et al. (2015), Huang et al. (2011), among others for evidence of significant gene–gene interactions in obesity. In behavioral genetics literature, additive genetic effects are associated with a narrow sense of heritability and broad-sense heritability refers to the proportion of trait variation that can be attributed to all types of genetic effects, including dominance, epistatic interaction, and additive effects. The estimates are generally believed to provide a lower bound since it only contains SNPs.

[47]Similar to how the training of econometricians and microeconomic theorists rely on developing stronger backgrounds in specific branches of mathematics and statistics, economists will need to become more familiar with the genetics literature. This is not unique to this field and currently many economists are now learning machine learning tools to analyze large datasets (Athey 2015), whereas many behavioral economists need to keep track of developments in the psychology and neuroscience literatures.

Acknowledgements
We are extremely grateful to David Cesarini, Jason Fletcher, Susumu Imai, and J. Niels Rosenquist for the often heated conversations over the past 12 years that have helped us to think critically on whether and how economists should incorporate molecular genetic data within their analysis. We are also grateful to Arnaud Chevalier for his patience and support in preparing this draft. Similar to an Academy Awards acceptance speech, we do not have the space to list others with whom we have had less heated debates including anonymous reviewers and the editor. Lehrer also would like to thank SSHRC for the research support. We are responsible for all errors, glaring omissions, and interpretations of evidence in the literature.
Responsible editor: Juan Jimeno

Competing interests
The IZA Journal of Labor Policy is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Author details
[1]School of Policy Studies and Department of Economics, Queen's University, Kingston K7L3N6, Ontario, Canada. [2]NYU-Shanghai, 1555 Century Avenue Office 1127, Pudong New District, Pudong 200122, Shanghai, China. [3]NBER, National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge 02138, MA, USA.

Received: 19 September 2016 Accepted: 16 February 2017
Published online: 21 April 2017

References

Ashraf Q, Galor O (2012) Response to comments made in a letter by d'Alpoim Guedes et al. on "The Out of Africa hypothesis, human genetic diversity and comparative development." http://www.brown.edu/Departments/Economics/Faculty/Oded_Galor/pdf/Ashraf-galor-Response.pdf

Ashraf Q, Galor O (2013) The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. Am Econ Rev 103(1):1–46

Athey S (2015) Machine learning and causal inference for policy evaluation, In Proc. 21st ACM SIGKDD Intl. Conf. Knowl. Disc. Data Min., ACM Press, New York. p. 5-6.

Baicker K, Cutler D, Song Z (2010) Workplace wellness programs can generate savings. Health Aff 29(1):1–8

Behrman JR (2016) In: Komlos J, Rashad I (eds) Twin studies in economics in the Oxford handbook of economics and human biology. Oxford University Press, p. 385-404

Belsky DW, Moffitt TE, Houts R, Bennett GG, Biddle AK, Blumenthal JA, … Caspi A (2012) Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. Arch Pediatr Adolesc Med 166:515–521

Belsky DW, Moffitt TE, Baker TB, Biddle AK, Evans JP, Harrington H, … Caspi A (2013) Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. JAMA Psychiatry 70:534–542

Benjamin DJ, Chabris CF, Glaeser EL, Gudnason V, Harris TB, Laibson DI, Launer L, Purcell S (2007) Genoeconomics. In: Weinstein M, Vaupel JW, Wachter KW (eds) Biosocial Surveys, Committee on population, division of behavioral and social sciences and education. The National Academies Press, Washington

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, Guðnason V, Harris TB, Launer LJ, Purcell S, Smith AV, Johannesson M, Magnusson PKE, Beauchamp JP, Christakis NA, Atwood CS, Hebert B, Freese J, Hauser RM, Hauser TS, Grankvist A, Hultman CM, Lichtenstein P (2012a) The promises and pitfalls of genoeconomics. Annu Rev Econ 4: 627–662

Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, Magnusson PKE, Chabris CF, Conley D, Laibson DI, Johannesson M, Visscher PM (2012b) The genetic architecture of economic and political preferences. Proc Natl Acad Sci 109(21):8026–8031

Biroli P (2015) Genetic and economic interaction in the formation of human capital: the case of obesity, Mimeo. University of Zurich

Boardman JD, Domingue BW, Daw J (2015) What can genes tell us about the relationship between education and health? Soc Sci Med 127:171–180
Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, … Fox J (2011) OpenMx: an open source extended structural equation modeling framework. Psychometrika 76:306–317
Carpenter J, Garcia J, Lum J (2011) Dopamine receptor genes predict risk preferences, time preferences, and related economic choices. J Risk Uncertain 42:233–261
Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, Taylor A, Poulton R (2002) Role of genotype in the cycle of violence in maltreated children. Science 297(5582):851–854
Cawley J, Han E, Norton E (2011) The validity of genes related to neurotransmitters as instrumental variables. Health Econ 20(3):884–888
Cesarini D, Dawes CT, Fowler J, Johannesson M, Lichtenstein P, Wallace B (2008) Heritability of cooperative behavior in the trust game. Proc Natl Acad Sci 105:3271–3276
Cesarini D, Dawes CT, Johannesson M, Lichtenstein P, Wallace B (2009) Genetic variation in preferences for giving and risk-taking. Q J Econ 124:809–842
Cesarini D, Johannesson M, Lichtenstein P, Sandewall O, Wallace B (2010) Genetic variation in financial decision-making. J Financ 65:1725–1754
Cesarini D, Johannesson M, Magnusson P, Wallace B (2012) The behavioral genetics of behavioral anomalies. Manag Sci 58(1):21–34
Chabris CF, Lee JJ, Benjamin DJ, Beauchamp JP, Glaeser EL, Borst G, Pinker S, Laibson DI (2013) Why is it hard to find genes that are associated with social science traits? Theoretical and empirical considerations. Am J Public Health 103(S1):S152–S166
Chabris CF, Hebert BM, Benjamin DJ, Beauchamp J, Cesarini D, van der Loos M, … Laibson D (2012) Most reported genetic associations with general intelligence are probably false positives. Psychol Sci 23(11):1314–1323
Conley D (2009) The promise and challenges of incorporating genetic data into longitudinal social science surveys and research. Biodemography Soc Biol 55(2):238–251
Conley D, Rauscher E (2013) Genetic interactions with prenatal social environment: effects on academic and behavioral outcomes. J Health Soc Behav 54(1):109–127
Conley TG, Hansen CB, Rossi PE (2012) Plausibly exogenous. Rev Econ Stat 94:260–272
Conti G, Heckman JJ (2010) Understanding the early origins of the education–health gradient: a framework that can also be applied to analyze gene–environment interactions. Perspect Psychol Sci 5:585–605
d'Alpoim Guedes J, Reich D, Herzfeld M, Patterson N, Bestor T, Lieberman D, Comaroff J et al (2012) Response to Ashraf and Galor "The Out of Africa hypothesis, human genetic diversity and comparative economic development"., http://ssrn.com/abstract=2155060
d'Alpoim Guedes J, Bestor TC, Carrasco D, Flad R, Fosse E et al (2013) Is poverty in our genes? Curr Anthropol 54:71–79
Davey Smith G (2003) "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32(1):1–22
DeFries JC, Fulker DW (1985) Multiple regression analysis of twin data. Behav Genet 15(5):467–473
DeNeve J-E, Fowler J (2014) Credit card borrowing and the monoamine oxidase A (MAOA) gene. J Econ Behav Organ, 107(B):428–39
Ding W, Lehrer SF, Rosenquist NJ, Audrain-McGovern J (2006) The impact of poor health on education: new evidence using genetic markers, NBER Working paper 12304
Ding W, Lehrer SF, Rosenquist JN, Audrain-McGovern J (2009) The impact of poor health on academic performance: new evidence using genetic markers. J Health Econ 28(3):578–597
Dreber A, Apicella CL, Eisenberg DTA, Garcia JR, Zamore RS, Lum JK, Campbell B (2009) The 7R polymorphism in the dopamine receptor D4 gene (DRD4) is associated with financial risk taking in men. Evol Hum Behav 30(2):85–92
Dreber A, Rand DG, Wernerfelt N, Garcia JR, Vilar MG, Lum JK, Zeckhauser RJ (2011) Dopamine and risk choices in different domains: findings among serious tournament bridge players. J Risk Uncertain 43:19–38
Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. PLoS Genet 9:e1003348
Fang MZ (2013) Violating the Monotonicity condition for instrumental variable—Dimorphic patterns of gene–behavior association. Economics Letters 122(1): 59–63
Fletcher JM, Conley D (2013) The challenge of causal inference in gene-environment interaction research: leveraging research designs from the social sciences. Am J Public Health 103(suppl 1):S42–S45
Fletcher JM, Lehrer SF (2009a) Using genetic lotteries within families to examine the causal impact of poor health on academic achievement, National Bureau of Economic Research Working Paper Series No. 15148
Fletcher JM, Lehrer SF (2009b) The effects of adolescent health on educational outcomes: causal evidence using genetic lotteries between siblings. Forum Health Econ Policy 12(2):Article 8
Fletcher JM, Lehrer SF (2011) Genetic lotteries within families. J Health Econ 30:647–659
Gee (2014) All the time in the world: an examination of time preferences using monetary delay discount rates. MA Research paper, Queen's University
Glessner JT, Connolly JJ, Hakonarson H (2014) Genome-wide association studies of autism. Curr Behav Neurosci Rep 1(4):234–241
Gluckman PD, Hanson M (2005) The fetal matrix: evolution, development, and disease. Cambridge University Press, Cambridge
Goldin C (1999) The economist as detective. In: Szenberg M (ed) Passion and craft: economists at work
Greiner J, Rubin D (2011) Causal effects of perceived immutable characteristics. Rev Econ Stat 93:775–785
Grossman M (1972) On the concept of health capital and the demand for health. J Polit Econ 80(2):223–255
Hansen BE (1999) Threshold effects in non-dynamic panels: estimation, testing, and inference. J Econ 93:345–368
Hatemi PK (2013) The influence of major life events on economic attitudes in a world of gene-environment interplay. Am J Polit Sci 57(4):987–1000
Hausman JA, Newey WK, Woutersen T, Chao JC, Swanson NR (2012) Instrumental variables estimation with heteroskedasticity and many instruments. Quant Econ 3(2):211–255
Hayden EC (2013) Dangerous work. Nature 502:5–6, 03 October 2013

Heckman JJ (2007) The economics, technology and neuroscience of human capability formation. Proc Natl Acad Sci 104:13250–13255
Herrnstein RJ, Murray C (1994) The bell curve. The Free Press, New York
Hewitt JK (2012) Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. Behav Genet 42(1):1–2
Huang W, Sun Y, Sun J (2011) Combined effects of FTO rs9939609 and MC4R rs17782313 on obesity and BMI in Chinese Han populations. Endocrine 39(1):69–74
Jencks C (1980) Heredity, Environment, and Public Policy Reconsidered. Am Sociol Rev 45(5):723–36
Jensen AR (1967) Estimating the limits of hereditability of traits by comparison of monozygotic and dizygotic twins. Proc Natl Acad Sci 58:149–156
Katan MB (1986) Apolipoprotein E isoforms, serum cholesterol and cancer. Lancet 327:507–508
Knafo A, Israel S, Darvasi A, Bachner-Melman R, Uzefovsky F, Cohen L, Feldman E, Lerer E, Laiba E, Raz Y (2008) Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1 a receptor RS 3 promoter region and correlation between RS 3 length and hippocampal mRNA. Genes Brain Behav 7(3):266–275
Kohler H-P, Behrman JR, Schnittker J (2011) Social science methods for twins data: integrating causality, endowments, and heritability. Biodemography Soc Biol 57(1):88–141
Kuhnen CM, Chiao JY (2009) Genetic Determinants of Financial Risk Taking. PLoS ONE 4(2): e4362
Kuhnen CM, Samanez-Larkin GR, Knutson B (2013) Serotonergic genotypes, neuroticism, and financial choices. PLOS ONE 8:e54632
Lazopoulou N, Gkioka E, Ntalla I, Pervanidou P, Magiakou AM, Roma-Giannikou E, Chrousos GP, Papassotiriou I, Dedoussis G, Kanaka-Gantenbein C (2015) The combined effect of MC4R and FTO risk alleles on childhood obesity in Greece. Hormones 14(1):126–133. doi:10.14310/horm.2002.1524
Lee S, Shaikh A (2014) Multiple testing and heterogeneous treatment effects: re-evaluating the effect of progress on school enrollment. J Appl Econ 29:612–626
Lehrer SF (2016) In: Komlos J, Rashad I (eds) Biomarkers as inputs in the Oxford handbook of economics and human biology. Oxford University Press, 339-365
Lehrer SF, Pohl VR, Song K (2016) Targeting policies: multiple testing and distributional treatment effects. NBER Working Paper No. 22950
Manski C (2013) Public policy in an uncertain world: Analysis and Decisions. Harvard University Press, Cambridge
McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9(5):356–369
Mertins V, Schote AB, Hoffeld W, Griessmair M, Meyer J (2011). Genetic susceptibility for individual cooperation preferences: the role of monoamine oxidase A gene (MAOA) in the voluntary provision of public goods. PLoS ONE, 6(6):e20959
Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P,…, Visscher PM, Esko T, Koellinger PD, Cesarini D, Benjamin DJ (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533:539–542, 26 May 2016
Ott J, Wang J, Leal SM (2015) Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 16:275–284
Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z (2014) A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform 15:256–278
Papageorge NW, Thom K (2016) Genes, education, and labor market outcomes: evidence from the health and retirement study. Mimeo, John Hopkins university
Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, … Fraser G (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460(7256):748–752
Rabe-Hesketh S, Skrondal A, Gjessing HK (2008) Biometrical modeling of twin and family data using standard mixed model software. Biometrics 64:280–288
Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, …, Conley D, Davey-Smith G, Franke L, Groenen PJF, Hofman A, Johannesson M, Kardia SLR, Krueger RF, Laibson D, Martin NG, Meyer MN, Posthuma D, Thurik AR, Timpson NJ, Uitterlinden AG, van Duijn CM, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340(6139):1467–1471
Rietveld CA, Cesarini D, Benjamin DJ, Koellinger PD, De Neve J-E, Tiemeier H, Johannesson M, Magnusson PKE, Pedersen NL, Krueger RF, Bartels M (2013b) Molecular genetics and subjective well-being. Proc Natl Acad Sci 110(24):9692–9697
Rietveld CA, Esko T, Davies G, Pers TH, Turley PA, Beben B, Chabris CF, Emilsson V, Johnson AD, Lee JJ, de Leeuw C, Marioni RE, Medland SE, Miller MB, Rostapshova O, Van der Lee SJ, Vinkhuyzen AAE, Amin N, Dalton C, Derringer J, van Duijn CM, Fehrmann R, Franke L, Glaeser EL, Hansell NK, Hayward C, Iacono WG, Ibrahim-Verbaas CA, Jaddoe V, Karjalainen J, Laibson D, Lichtenstein P, Liewald DC, Magnusson PKE, Martin NG, McGue M, McMahon G, Pedersen NL, Pinker S, Porteous DJ, Posthuma D, Rivadeneira F, Smith BH, Starr JM, Tiemeier H, Timpson NJ, Trzaskowski M, Uitterlinden AG, Verhulst FC, Ward ME, Wright MJ, Smith GD, Deary IJ, Johannesson M, Plomin R, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD (2014) Common genetic variants associated with cognitive performance identified using proxy-phenotype method. Proc Natl Acad Sci 111(38):13790–13794
Ripke S, members of the Schizophrenia Working Group of the Psychiatric Genomics Consortium et al (2014) Biological insights from 108 schizophrenia-associated genetic loci. Nature 511(7510):421–427
Rosenquist JN, Lehrer SF, Malley AJO, Zaslavsky AM, Smoller JW, Christakis NA (2015) Cohort of birth modifies the association between FTO genotype and BMI. Proc Natl Acad Sci 112(2):354–359
Rutter M (2006) Genes and behavior: nature–nurture interplay explained. Blackwell, Oxford
Rutter M, Moffitt TE, Caspi A (2006) Gene-environment interplay and psychopathology: multiple varieties but real effects. J Child Psychol Psychiatry 47(3-4):226–261
Shanahan MJ, Vaisey S, Erickson LD, Smolen A (2008) Environmental contingencies and genetic propensities: social capital, educational continuation, and dopamine receptor gene DRD2. Am J Semiotics 114(86):S260–S286

Smith K, Alford JR, Hatemi PK, Eaves LJ, Funk C, Hibbing JR (2012) Biology ideology, and epistemology: how do we know political attitudes are inherited and why should we care. Am J Polit Sci 56(1):17–33
Taubman P (1976) The determinants of earnings: genetics, family, and other environments: a study of white male twins. Am Econ Rev 66:858–870
Thompson O (2014) Economic background and educational attainment the role of gene-environment interactions. J Hum Resour 49(2):263–294
Todd PE, Wolpin KI (2003) On the specification and estimation of the production function for cognitive achievement. Econ J 113(1):3–33
Van IJzendoorn MH, Bakermans-Kranenburg MJ (2012) A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. Psychoneuroendocrinology 37:438–443
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Yuan Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Lai Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F.,May, D., McCawley, S., McIntosh, T.,McMullen, I., Moy, M.,Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N., Tse, S., Vech, C.,Wang, G.,Wetter, J., Williams, S., Williams, M.,Windsor, S., Winn-Deen, E.,Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, Kejariwal, B., Mi, A.H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R.,Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L. , Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y., Coyne, M., Dahlke, C., Deslattes Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros,W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E.,Wang, M.,Wen, M.,Wu, D.,Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304–1351
Wallace B, Cesarini D, Lichtenstein P, Johannesson M (2007) Heritability of ultimatum game responder behavior. Proc Natl Acad Sci 104:15631–15634
Wehby G, Fletcher JM, Lehrer SF, Moreno LM, Murray JC, Wilcox A, Lie RT (2011) A genetic instrumental variables analysis of the effects of prenatal smoking on birth weight: evidence from two samples. Biodemography Soc Biol 57(1):3–32
Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, Fall T, Graff M, Justice AE, Luan J, Gustafsson S, Randall JC, Vedantam S, Workalemahu T, Kilpeläinen TO, Scherag A, Esko T, Kutalik Z, the GIANT consortium, Heid IM, Loos RJF (2014) Quality control and conduct of genome-wide association meta-analyses. Nature Protocols 9(5):1192–1212
Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42(7):565–569
Yang JA, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics 88:76–82
Zhong S, Israel S, Xue H, Ebstein RP, Chew SH (2009) Monoamine oxidase a gene (maoa) associated with attitude towards longshot risks. PLoS One 4(12):e8516