

Molecular Genetics and Economics[†]

Jonathan P. Beauchamp, David Cesarini, Magnus Johannesson, Matthijs J. H. M. van der Loos, Philipp D. Koellinger, Patrick J. F. Groenen, James H. Fowler, J. Niels Rosenquist, A. Roy Thurik, and Nicholas A. Christakis

The question of how traits and behaviors pass from one generation to the next has been the subject of intense interest throughout the history of science. Simple parent–child correlations are open to multiple interpretations, as parents transmit both environment and genome to their children. Until recently, genotyping—or the direct measurement of variation in an individual’s DNA sequence through biological assays—was exorbitantly expensive; distinguishing the roles of genetics and environment was the realm of behavioral genetics, in which

■ *Jonathan P. Beauchamp just completed his Ph.D. in Economics at Harvard University, Cambridge, Massachusetts. David Cesarini is Assistant Professor of Economics, Center for Experimental Social Science, New York University, New York City, New York, and an Affiliated Researcher, Institute for Industrial Economics (IFN), Stockholm, Sweden. Magnus Johannesson is Professor of Economics, Stockholm School of Economics, Stockholm, Sweden. Matthijs J. H. M. van der Loos is a Ph.D. candidate in Applied Economics, Erasmus School of Economics, Rotterdam, Netherlands. Philipp D. Koellinger is Assistant Professor of Economics, Erasmus School of Economics, Rotterdam, Netherlands. Patrick J. F. Groenen is Professor of Statistics, Erasmus School of Economics, Rotterdam, Netherlands. James H. Fowler is Professor of Medical Genetics and Political Science, University of California at San Diego, La Jolla, California. J. Niels Rosenquist is Instructor at Harvard Medical School and the Massachusetts General Hospital’s Psychiatric and Neurodevelopmental Genetics Unit and a Research Fellow at the Institute for Quantitative Social Science at Harvard, Cambridge, Massachusetts. A. Roy Thurik is Professor of Economics and Entrepreneurship, Erasmus School of Economics, Rotterdam, Netherlands. Nicholas A. Christakis is Professor of Medicine and of Medical Sociology, Harvard Medical School, Cambridge, Massachusetts. The corresponding author is (beauchamp.jonathan.pierre@gmail.com).*

[†]To access the Appendix, visit <http://www.aeaweb.org/articles.php?doi=10.1257/jep.25.4.57>.
doi=10.1257/jep.25.4.57

samples of twin, adoption, or other pedigree data were analyzed. However, with the completion of the Human Genome Project in the early 2000s (Venter et al., 2001; Lander et al., 2001) and the advent of inexpensive, genome-wide scans of variation, it is now increasingly feasible to examine specific genetic variants that predict individual differences directly.

In fact, the costs of comprehensively genotyping human subjects have fallen to the point where major funding bodies, even in the social sciences, are beginning to incorporate genetic and biological markers into major social surveys. The National Longitudinal Study of Adolescent Health, the Wisconsin Longitudinal Study, and the Health and Retirement Survey have launched, or are in the process of launching, datasets with comprehensively genotyped subjects. Similar efforts are also underway in Europe, for example with the Biobank Project in the United Kingdom (Ollier, Sprosen, and Peakman, 2005) and the large-scale genotyping of subjects at several European twin registries. These samples contain, or will soon contain, data on hundreds of thousands of genetic markers for each individual in the sample as well as, in most cases, basic economic variables. How, if at all, should economists use and combine molecular genetic and economic data? What challenges arise when analyzing genetically informative data?

In this article, we lay out the terrain for such questions. We use the term “genoeconomics,” originally proposed by Benjamin et al. (2007), to describe the use of molecular genetic information in economics. To illustrate some of the challenges that researchers in this field are likely to encounter, we present results from a “genome-wide association study” of educational attainment, one of the first of its kind in economics.¹ This type of study involves analyzing hundreds of thousands of genetic markers and seeking to understand their association with some trait of interest. We use a sample of 7,500 individuals from the Framingham Heart Study. After quality controls, our dataset contains over 360,000 genetic markers per person. Despite some initially promising results, the main findings from this dataset fail to replicate in a second large replication sample of 9,500 people from the Rotterdam Study, suggesting that the original results were probably spurious. These findings are unfortunately typical in these types of studies of molecular genetics and therefore also cautionary.

The frequent replication failures in this molecular genetics literature are likely a result of several forces, the most important of which is probably that the samples used in research are too small to ensure that there is adequate power to detect true associations (Ioannidis, 2005, 2007). When true effect sizes are small, the power to detect true associations will of course be poor, and the ratio of true to false signals will hence be low. Indeed, an important implication of the genome-wide association study results reported in this paper is that they confirm

¹Preliminary results from a genome-wide association study of educational attainment have been reported by Posthuma et al. (2008), Beauchamp, Cesarini, Rosenquist, Fowler, and Christakis (2010), and most recently by Martin et al. (2011). A genome-wide association study of self-employment has been initiated by van der Loos, Koellinger, Groenen, and Thurik (2010).

that common genetic variants with large main effects are likely to be extremely rare for economic variables, which tend to be far removed from the molecular genetic variant in the chain of causation. We perform power analyses to demonstrate this point and show that under plausible assumptions about the effect sizes of a specific type of common variation in the human genome, samples in the tens of thousands, perhaps more, may be necessary to detect genetic influences on most complex economic variables in a robust manner. This insight suggests that most existing genoeconomic studies, which are based on samples in the hundreds, are dramatically underpowered and that we should expect a high false discovery rate until this is remedied. Our choice of educational attainment as the outcome variable of study was determined by the widespread availability of this characteristic in cohorts that have already been genotyped. An important next step of a successful genoeconomic research agenda is to start measuring more biologically proximate variables—such as preferences—in large samples. Variables which are less distant from the genome in the chain of causation are more likely to require smaller samples in order for genetic associations to be detected reliably, and any detected associations are more likely to have a biologically meaningful interpretation and economically meaningful implications.

The empirical results in this paper are also used to discuss several other methodological issues that arise in the analysis of molecular genetic data. Our overall assessment is cautiously optimistic: this new data source has the potential not only to complement traditional behavioral genetic studies but also to add a new dimension to our understanding of heterogeneity in economic behaviors and outcomes, especially when it comes to traits that are close to the underlying biology. But for this ample potential to be realized, researchers and consumers of this literature should be wary of the pitfalls that lie ahead (Benjamin *et al.*, 2007). The most urgent of these challenges is the difficulty of doing reliable inference when faced with multiple hypothesis problems, which are on a scale that has never before been encountered in social science.

Behavioral Genetics

Over the past few decades, behavioral geneticists have produced a compelling array of evidence that genetic endowments influence economic behaviors, outcomes, and preferences. The general approach in these studies is to make assumptions about the extent to which the different sibling types share genetic and environmental conditions and infer the fraction of variance that can be statistically accounted for by genetic variation (heritability, denoted h^2), rearing conditions (common environment, denoted c^2), and idiosyncratic factors (unique environment, denoted e^2). These studies often compare the resemblance of adoptees reared in the same family to that of biological siblings reared in the same family, or the resemblance of identical (monozygotic) twins, who share their entire genomes, to that of fraternal (dizygotic) twins, who share approximately half their genomes. Sacerdote (2010)

provides an accessible introduction for economists. A standard textbook is Plomin, DeFries, McClearn, and McGuffin (2008).

The simplest behavioral genetic model is based on a host of strong assumptions, including the independence of genetic and family effects, and functional form assumptions; it fails to take assortative mating and nonlinear genetic effects into account. In the 1970s, when the debate between environmentalists and hereditarians reached its peak, there was much controversy over whether the high heritability estimates, especially for IQ, were artifacts of the simplistic behavioral genetic framework that would go away in more elaborate designs and with better data. In response, behavioral geneticists have built much richer datasets and expanded their models, relaxing the various problematic assumptions. They have consistently found that personality, IQ, and most other traits remain highly, or at least moderately, correlated with genetic endowments (Bouchard and McGue, 2003). In fact, the consensus in behavioral genetics that virtually all traits are associated with genotype is so strong that it has been elevated to the status of a “law” (Turkheimer, 2000).²

Economic behaviors, preferences, and outcomes are no exception. Behavioral genetic methods originally made limited inroads into economics through the work of Taubman and coauthors (for example, Taubman, 1976), who demonstrated that genetically identical (monozygotic) twins exhibit greater similarity than fraternal (dizygotic) twins in both educational attainment and income. Since then, a number of papers have followed suit in applying behavioral genetic research designs to the study of economic outcomes. Many of these studies rely on quasi-experiments such as adoption (Sacerdote, 2007; Björklund, Lindahl, and Plug, 2006; Björklund, Jäntti, and Solon, 2005), twinning (Taubman, 1976; Lichtenstein, Pedersen, and McClearn, 1992), or comparisons of multiple sibling types (Björklund, Jäntti, and Solon, 2005). More recent work has also demonstrated that economic preferences elicited from incentivized experiments or surveys are heritable, with estimates typically in the 20–30 percent range (Wallace, Cesarini, Johannesson, and Lichtenstein, 2007; Cesarini, Dawes, Johannesson, Lichtenstein, and Wallace, 2009a, b). These estimates are biased downward because they do not take into account measurement error in the preference elicitation.³ Two other studies of portfolio choice data found heritability estimates of about 0.25–0.60 for various financial decision-making variables (Barnea, Cronqvist, and Siegel, 2010; Cesarini, Johannesson, Lichtenstein, Sandewall, and Wallace, 2010).

² Estimates of heritability based on family data, such as twins, have also recently been corroborated by techniques that utilize molecular genetic data in ingenious ways to estimate heritability (Visscher et al., 2006; Yang et al., 2010).

³ Adjusting for noise appears to approximately double the heritability estimates (Beauchamp, Cesarini, and Johannesson, 2011). Test retest data (unpublished) has been collected for a sample of about 100 twins that participated in the experiments reported in Wallace, Cesarini, Johannesson, and Lichtenstein (2007) and Cesarini, Dawes, Johannesson, Lichtenstein, and Wallace (2009a, b). These data suggest a test retest correlation of about 0.5. Adjusting for measurement noise would thus approximately double the heritability estimates.

In interpreting heritability estimates, it is crucial to appreciate the possibility that genetic effects may operate via environmental effects, because genotypes may either evoke environmental responses or cause an individual to select a particular environment endogenously (Becker and Tomes, 1979; Dickens and Flynn, 2001; Fowler, Settle, and Christakis, 2011; Jencks, 1980; Ridley, 2003). This possibility has given rise to the expression “nature via nurture”—as opposed to “nature versus nurture.” Estimates of the behavioral genetic model can therefore be thought of as reduced form coefficients from a more general model in which some environments are endogenous to genotype (Dickens and Flynn, 2001; Jencks and Brown, 1977; Jencks, 1980; Lizzeri and Siniscalchi, 2008).

As pointed out by Jencks (1980), a common mistake is to equate “genetic” with “immutable”: the fact that a person’s DNA sequence is in some sense fixed does not mean that the effects of that sequence are fixed. Goldberger (1979) provides several examples of how the implications of heritability estimates have been misstated and notes that high estimates do not imply that interventions are doomed to failure. While genetic variation can statistically account for a moderate to large share of income in contemporary Western societies, this does not mean that it would be infeasible to use redistributive policies or policies that encourage human capital formation to change the distribution of income. Heritability is a population parameter that depends on both the environmental effects operating in a specific population at a certain point in time and on the genetic variation in that population. It says little about what would happen to the mean and variance of the trait were the environment to change. Therefore, there is no contradiction between observing a high heritability for height, say, and secular increases in height over time as the environment changes. Heritability estimates do not tell us how the genetic effects operate, of course, nor do they tell us much about whether the mechanisms are easy or hard to modify. But far from being useless, as has sometimes been asserted, heritability estimates tell us that for most traits, a sizable fraction of the within-family resemblance can ultimately be traced to shared DNA. We suspect that were it not for the impressive cumulative progress in behavioral genetics over the last couple of decades, the issue would still be contentious. Figuring out how and why genetic factors matter is an interesting scientific activity, and molecular genetic methods are an exciting tool to bring to bear on these questions.

Elementary Molecular Genetic Concepts

Molecular genetics is the branch of genetics that studies the structure and function of DNA at its most basic level. Recent decades have seen major advances, allowing researchers to better understand the numerous ways in which genomes vary between individuals. The human genome consists of 23 pairs of chromosomes that package DNA. One member of each pair of chromosomes is inherited from the mother and the other from the father. DNA itself consists of two strands of elementary building blocks that together form a double helix structure. The elementary

building blocks, called nucleotides, each contain one of four bases—A (adenine), C (cytosine), T (thymine), or G (guanine)—resulting in four distinct nucleotides. Due to a property of DNA called complementarity, a nucleotide with the base A is always paired with a nucleotide with the base T and a nucleotide with the base C is always paired with a nucleotide with the base G, forming so-called base pairs and holding the two strands of DNA together.

A locus is a specific position of a DNA sequence on a (pair of) chromosome(s). A locus thus refers to a *pair of base pairs* (or *nucleotide pairs*), one base pair coming from the paternal chromosome and the other base pair coming from the maternal chromosome. The human genome consists of approximately three billion such pairs of base pairs arranged into the 23 (pairs of) chromosomes. Because of complementarity, the second base of a base pair can be directly identified from knowledge of the first one, and so it is common practice to refer to a locus as consisting of two single bases rather than of a pair of base pairs. For example, the genotype AT-AT would be referred to as AA or as TT.⁴

Genes are sequences of nucleotide base pairs that code for some types of RNA products, many of which in turn code for proteins. These RNA products and proteins begin cascades of interactions that regulate bodily structures and functions. Only a small portion of the genome actually consists of genes, and both genetic variation in the genes and in the remaining portion of the genome can account for variation in behaviors and traits. However, because of genes' functional importance, many researchers have focused their attention on genetic variation in the genes; also, it is often said, loosely, that “a gene causes” a behavior or trait even though what is meant is that genetic variation at a given locus—often not even in a gene—accounts for some of the variation in the behavior or trait.

Humans share most, but not all, of their genetic material: approximately 99.6 percent of common genetic variants are the same when comparing any two unrelated individuals (Kidd et al., 2008). Genetic variation comes in many forms, but most can be traced to one of two types of mutation events. The simplest mutation event is a base substitution, in which the base pair of a nucleotide pair is substituted for another. Whenever a nucleotide varies at a specific locus across individuals in the population, it is said to be a *single nucleotide polymorphism*, or SNP (pronounced “snip”), with the different genetic variants of a SNP called “alleles.” Most other forms of genetic variation are due to repeated segments of DNA. In *variable number of tandem repeat* (VNTR) polymorphisms, there are differences across individuals in the number of times that particular short segments of DNA are repeated. In *copy number variation* (CNV) polymorphisms, there are differences in the number of repetitions of a long segment of DNA—of at least 1,000 base pairs and often many more.

Genotyping SNPs and other genetic variants is performed with technology that allows high-throughput typing of hundreds of thousands of genetic variants per individual. Current technologies type around 500,000 SNPs, but versions with over one

⁴ For an accessible introduction to the basic concepts in molecular genetics, see Strachan and Read (2003) or Carey (2003).

million SNPs and other variants are already available, and this number is expected to increase in the very near future. Within a decade, it will be possible to genotype entire genomes at relatively low cost. Because SNPs in the vicinity of each other are often highly correlated, it is generally possible to impute unobserved SNPs with high accuracy if a neighboring set of SNPs has been genotyped; for that reason, even though most arrays type only a minute fraction of the three billion base pairs in the human genome, they can in principle capture a large part of the relevant genetic variation.⁵

In some rare cases, a difference at a specific locus on a chromosome can single-handedly lead to a disease: Huntington's disease is an example. However, the vast majority of physical characteristics are "polygenic," meaning they are influenced by multiple genetic polymorphisms (Mackay, 2001). The fact that there are so many potentially etiologically relevant SNPs in the genome leads to huge challenges of inference in the face of multiple hypothesis testing, of a magnitude that social scientists have never before faced. Should gene-gene interactions also turn out to be important, the combinatorics would simply be staggering. However, as we later argue, the weight of the evidence suggests that a significant portion of the genetic variance in complex traits is "additive."

Many traits of interest to economists are several steps removed from the original genotype in the chain of causation. Therefore, it is unlikely that genetic variants with a *proximal* effect on, for example, socioeconomic status exist and will ever be found. However, simple association models between candidate SNPs and various outcomes are still useful because the main effects discovered could suggest areas for further exploration and point to the more proximate biological mechanism (for example, pointing to influences on the production of the neurotransmitter dopamine in the brain).

Candidate Gene Studies

Genetic association studies are becoming increasingly common in the social sciences, including economics. Many of the studies carried out to date have examined the relationship between some economic characteristics—typically an experimentally elicited preference parameter—and a relatively small number of genetic markers that are selected based on some a priori hypotheses derived from information about their biological function. The studies that follow this research design are therefore known as "candidate-gene" studies.

The first candidate-gene study in geneoeconomics of which we are aware was by Knafo et al. (2008), who reported that polymorphisms of the *AVPR1a* gene were associated with the level of generosity exhibited by a sample of 203 college students playing a standard dictator game (Forsythe, Horowitz, Savin, and Sefton, 1994).

⁵ When two SNPs are correlated, geneticists say that they are in "linkage disequilibrium." Even though the human genome is about 3 billion base pairs, half a million well-chosen markers are sufficient to cover ("tag") much of the genetic variation.

There have subsequently been many other reported associations between experimentally elicited preferences and genetic variants (Apicella et al., 2010; Crisan et al., 2009; Dreber et al., 2009; Israel et al., 2009; Kuhnen and Chiao, 2009; McDermott, Tingley, Cowden, Frazzetto, and Johnson, 2009; Roe et al., 2009; Zhong, Israel, Xue, Sham, Ebstein, and Chew, 2009; Zhong, Israel, Xue, Ebstein, and Chew, 2009). For a review of the research to date, see Ebstein, Israel, Chew, Zhong, and Knafo (2010).

The genetic markers used in these studies were typically selected based on evidence about their neurochemical function, and the studies test the hypothesis that these genetic markers are correlated with behavior. As an example, because the dopamine receptor *D4* (*DRD4*) gene affects dopamine receptors in the brain and because dopamine plays an important role in learning and the processing of reward (Schultz, Dayan, and Montague, 1997), many researchers initially suspected that variation in the *DRD4* gene is associated with variation in risk preferences. Two early geneoconomics studies, Dreber et al. (2009) and Kuhnen and Chiao (2009), both reported an association between a particular variant of *DRD4* and experimentally elicited risk preferences. However, Carpenter, Garcia, and Lum (2011) subsequently failed to replicate this finding, and in fact reported a borderline significant association in the opposite direction.

Yet another set of studies drew on pharmacological randomized control studies which demonstrated that exogenous administration of the neuropeptide oxytocin increases trust in humans relative to a placebo (Kosfeld, Heinrichs, Zak, Fischbacher, and Fehr, 2005). Such findings suggest that variation in the gene *OXTR*, which encodes the oxytocin receptor, may account for the heritable variation in prosocial behavior (Cesarini, Dawes, Johannesson, Lichtenstein, and Wallace, 2009a). Israel et al. (2009) reported an association between an *OXTR* variant and giving in the dictator game, but Apicella et al. (2010) failed to replicate this association. Similarly, Nicolaou, Shane, Adi, Mangino, and Harris (2011) reported an association between a variant of the dopamine receptor *D3* gene and self-employment, but the association did not replicate in a significantly larger sample (van der Loos et al., 2011).

An Illustration of Geneoconomic Research

Though geneoconomics research has so far mainly focused on candidate-gene studies, another important research design is the genome-wide association study. In this approach, tens or hundreds of thousands of genetic markers are individually tested for association with a trait of interest. The molecular genetic approach is fundamentally different from the behavioral genetic approach in that the effect of a genotype is estimated directly, rather than inferred by contrasting the resemblance of different types of relatives. During the last decade, genome-wide association studies have emerged as a standard research tool in medical genetics. The approach has been somewhat successful in uncovering associations between common variation and complex diseases and traits, with over 1,000 published studies to date, a

figure that is rapidly growing. The advantages of genome-wide association studies are that they require no prior knowledge about the pathways between genotype and outcome or behavior of interest and that they scan all the available genome data rather than study a particular variant. This may lead to unanticipated findings. However, a potential downside is limited power, as proper adjustment for multiple hypothesis testing requires extremely stringent significance thresholds.

In this section, we describe a project in which we conducted a genome-wide association study seeking to identify SNPs associated with educational attainment. In the process, we describe the main steps generally required to work with genotypic data as well as some of the challenges that arise in this type of research. What follows is an overview of the main steps; additional technical details are available in an Appendix available online with this paper at <http://e-jep.org>. In the first stage of this study, we analyzed data on about 7,500 individuals who have been genotyped at over half a million SNPs, to search for specific SNPs associated with educational attainment. In the replication stage of our genome-wide association study, we attempted to replicate the 20 most significant associations from the first stage in an independent dataset with more than 9,500 genotyped individuals.

The First Stage: Data from the Framingham Heart Study

The data for the first stage of the genome-wide association study comes from the Framingham Heart Study (Dawber, Meadors, and Moore, 1951; Feinleib, Kannel, Garrison, McNamara, and Castelli, 1975), a longitudinal study which was started six decades ago by the U.S. Public Health Service to study cardiovascular diseases and which has played an important role in the medical literature (Levy and Brink, 2005). The study was initiated in 1948, when two-thirds of all adults domiciled in the town of Framingham were enrolled. The study was expanded in 1971 and 2002, when first- and second-generation descendants were enrolled, respectively. The spouses of first-generation descendants were also enrolled. Out of a total of 14,531 individuals included in the study, 8,496 had complete genotypic, education, and basic demographic data. Genotyping was conducted using the Affymetrix 500K chip, an array which contains 500,568 SNPs (Affymetrix, 2009). More details on the data used here, on the construction of the educational attainment variables, and on the genotyping are provided in the online Appendix.

As genome-wide association studies have become increasingly popular in medical genetics in the last few years, a number of methodological conventions have been developed, both to deal with the complexities of genotypic data and also with the issue of testing hypotheses in this context (Pearson and Manolio, 2008; Sullivan and Purcell, 2008). We followed these conventions here. We first applied a number of quality control measures to the Framingham sample comprising all 9,237 individuals with genetic data. We dropped 499 individuals from the sample because more than 5 percent of their genotypic data were missing; a high “missingness” can be suggestive that problems occurred in the genotyping procedure for the individual. Next, we only kept individual SNPs which satisfied each of three additional quality controls: 1) a missing data frequency less than 2.5 percent;

2) a test that the observed genotype frequencies are equal to their theoretical expectations under random mating (a test of “Hardy–Weinberg equilibrium”); and 3) the frequency of the *minor allele*, that is, the least common allele, greater than 1 percent. Failure to meet the first or second condition may indicate genotyping errors for the SNP; failure to meet the third generally leads to imprecisely estimated coefficients. From the 500,568 SNPs on our Affymetrix 500k array, 363,776 satisfied all three criteria.

Another important step for the researcher working with genotypic data is to control for population stratification—the differences in allele frequencies across subpopulations. Such differences can occur as a consequence of founder effects, genetic drift, and differences in natural selection pressures (Hartl, 1988). In population genetics, a founder effect is defined as the reduction in genetic variance that arises when a new population is established by a low number of isolated individuals. Genetic drift refers to changes in the allele frequencies in a population due to chance events. When both the frequencies of alleles and environmental factors affecting a trait of interest differ across subpopulations, failure to control properly for these differences can lead to omitted-variables bias and to spurious associations between those alleles and the trait. An illustration of population stratification was provided by Hamer and Sirota (2000), who asked their readers to entertain the thought experiment of looking for genetic markers for chopstick use in a sample comprising Caucasian and East Asian individuals. Without population stratification controls, markers which differ significantly in frequency between the Caucasian and Asian subpopulations will be found to be associated with chopstick use, but those associations will of course be due to cultural, not genetic, differences. Although the individuals in the Framingham Heart Study are almost all of European ancestry, population stratification has been shown to be a concern even in samples of European Americans (Campbell et al., 2005).

Several approaches have been proposed in the literature to control for population stratification. The most convincing solution to this problem is to only use within-family variation in genotype, using sibling fixed effects models. However, this comes at the cost of not using any of the between-family variation and thus, in most instances, of much reduced power. Instead, we employed a standard approach (the EIGENSTRAT method) developed by Price, Patterson, Plenge, Weinblatt, Shadick, and Reich (2006), which applies “principal component analysis” to the genotypic data. The first principal component of a set of variables is the linear combination of the variables with the coefficients chosen to capture as much of the sample variation as possible. The second principal component is obtained analogously, but subject to the constraint that it seek to capture whatever variation is remaining after the first principal component has been applied—and so on. We used the scores of each individual on each of the 10 first principal components as control variables in the main regression specification; in effect, these 10 values capture common variation across the population structure, and thus offer at least a partial control for population stratification. Consistent with standard practice, we also dropped individuals who were outliers from the sample, where outliers are defined as individuals whose score

is at least six standard deviations from the mean on one of the top ten principal components. As a result, 531 outliers were eliminated, leaving 8,207 individuals with satisfactory genotypic data. The final sample used for the genome-wide association study comprised 7,574 individuals with satisfactory data on genotype and on observed characteristics. The sample size for each regression was generally a bit smaller than that because for each SNP there were some individuals with missing genotypic information.

Association Analysis

For each individual SNP that passed our quality control standards, we specified the following standard regression model:

$$Edu = \beta_0 + \beta_1 \cdot SNP_k + \mathbf{PC} \cdot \beta_2 + \mathbf{X} \cdot \beta_3 + \varepsilon,$$

where Edu is years of education; SNP_k is the number of copies of the minor allele (0, 1, or 2) an individual has at SNP_k ; \mathbf{PC} is a vector of the individual's principal component scores on the 10 top principal components of the genome of the sample; and \mathbf{X} is a vector of controls.

A number of questions immediately arise in looking at this framework. For example, is the relationship between the genetic differences and educational attainment linear? (In genetic parlance, the model assumes that all genetic variation is “additive.”) A person can have at most two copies of the minor allele, one from each parent. The model is misspecified if, in expectation, the educational attainment of the *heterozygotes* (those with one copy of the minor allele) is in fact not the midpoint of the two *homozygotes* (those with zero or two copies of the minor allele). The main justification for specifying an additive model, besides parsimony, is that theory as well as converging evidence from behavioral genetics and animal breeding (Hill, Goddard, and Visscher, 2008) suggest that an additive, polygenic, model of inheritance fits the data surprisingly well for most complex traits. Among other observations, if higher-order interactions between genetic polymorphisms were important, one would observe a much more rapid decay of resemblance when comparing, say, monozygotic twins to full siblings and cousins.

Here, the coefficient estimate of β_1 cannot be regarded as an estimate of the causal effect of SNP_k on education; in particular, SNP_k can be correlated with neighboring genetic polymorphisms that causally affect education but are not included in the regression. Because the Framingham sample is family-based and related individuals share parts of their environments and large portions of their genomes, it is necessary to account for the non-independence of the error terms. The interested reader is referred to the online Appendix for additional details of the necessary calculations.

Inference under Multiple Hypotheses

A challenging issue that arises in genome-wide association studies is the huge number of potentially relevant genetic variables—that is, a very large number of hypotheses are being tested. Because of this, many SNPs will inevitably turn out to

have a statistically significant correlation with the dependent variables, at least at conventional levels of statistical significance, due to sampling variation and other chance events. Several methods have been proposed to draw statistical inferences in this context. The most commonly used—and the most stringent—solution is to use an approach named after the Italian mathematician Carlo Bonferroni, in which the conventional significance threshold is divided by the number of tests performed to obtain a Bonferroni-corrected significance threshold (or equivalently, all p -values are multiplied by the number of tests performed to obtain Bonferroni-corrected p -values). In the study with the Framingham data, 363,776 tests were performed (one for each SNP that passed the quality-control filters), thus yielding a Bonferroni-corrected significance threshold of $0.05/363,776 = 1.37 \times 10^{-7}$.

The Bonferroni approach is generally agreed to be overly conservative because SNPs that are close to one another are generally correlated, and thus not statistically independent; thus, the Bonferroni correction divides by too large a number. With this consideration in mind, one commonly used threshold in the early literature for genome-wide association studies based on 500,000-SNP array data was set by the Wellcome Trust Case Control Consortium at 5×10^{-7} (Wellcome Trust Case Control Consortium, 2007). The standard threshold is now 5×10^{-8} (McCarthy et al., 2008). For simplicity, we only report below the Bonferroni-corrected p -values, but several more sophisticated methods have been developed to deal with this situation.⁶ However, as we discuss below, despite the theoretical soundness of these methods, previous experience with false positives in the field of medical genetics has led researchers to be cautious in interpreting results that have not been replicated in an independent sample.

The Replication Stage: Data from the Rotterdam Study

The data for the replication stage of the genome-wide association study comes from the Rotterdam Study (Hofman et al., 2009), a study that currently consists of three cohorts. Recruitment for the first, second, and third cohorts began in 1990, 2000, and 2006, respectively. Together, these three cohorts contain data on 14,926 individuals from the well-defined Ommoord district in Rotterdam, all of whom were 45 years old or more when the third cohort was recruited. Of these, 9,535 individuals have complete genotypic, education, and basic demographic data. Genotyping was done with the Illumina 550K and 610K arrays.

In the replication stage, we attempted to replicate the 20 most significant associations from the first stage using the Rotterdam data as an independent sample.⁷

⁶ Some possibilities include permutation-based approaches (Churchill and Doerge, 1994); Bayesian approaches that incorporate prior biological knowledge and SNP characteristics (Stephens and Balding, 2009); estimation and control of the false discovery rate (the expected proportion of significant associations that are false positives) (Hochberg and Benjamini, 1990; Sabatti, Service, and Freimer, 2003); and approaches that calculate the effective number of independent tests (Gao, Becker, Becker, Starmer, and Province, 2010).

⁷ An alternative approach would have been to pool all the coefficient estimates from the two samples and then meta-analyze the results. This approach is typical in consortium studies.

The procedures we followed were very similar to those outlined above for the first stage, but many of the 20 SNPs were not directly available in the Rotterdam data and so had to be imputed since the Framingham Heart Study and the Rotterdam Study used different genotyping platforms. Imputation is performed by using the correlation structure of an independent, more densely genotyped reference sample to infer the expected genotypes at the SNPs that have not been genotyped in the sample of interest. Though this procedure inevitably results in measurement error for the imputed SNPs (and thus to downward bias in their coefficient estimates), this effect is relatively small here as all 20 SNPs were imputed with a R^2 greater than 0.92. We used the HapMap CEU sample (International HapMap Consortium, 2003) as reference sample. Again, the online Appendix provides more detail on this procedure.

Results

Table 1 reports results for the 20 SNPs that attained the highest statistical significance in the first-stage analysis of the Framingham dataset. The first column gives the “rs number”—the most widely used SNP identifier—of each SNP with the chromosome on which it is located in parentheses. The second column shows a letter of the base pair corresponding to the minor allele of the SNP; this can be useful when comparing the direction of the estimated coefficients across studies. The third column indicates whether the SNP is close (less than 100,000 base pairs away) to any known genes (as mentioned above, SNPs need not be within genes).

The fourth column shows the estimated regression coefficients; these are clustered around 0.25 for most SNPs, meaning that in our sample, the difference between the two homozygotes (those with zero or two copies of the minor allele) is about 0.5 years in educational attainment. However, these estimates are subject to upward bias because of a “winner’s curse” type of selection bias (Zhong and Prentice, 2008). Put simply, the significance of a SNP depends in part on the size of the regression coefficient in the sample studied such that the most significant SNPs are likely to have actual regression coefficients that are inflated relative to the actual population parameter. In fact, the similarity in the magnitude of our reported coefficient estimates may suggest that the bias may count for much more than the underlying true coefficients. The estimated effect sizes are usually smaller in follow-up studies than in the original study, even when replication attempts are successful (Ioannidis, Ntzani, Trikalinos, and Contopoulos-Ioannidis, 2001).

The fifth column reports the raw p -value of each SNP. Four of the SNPs are significant at the 5×10^{-7} level and none of them at the 5×10^{-8} level. As shown in the sixth column, none of the SNPs survive a Bonferroni correction at the 10 percent level, the two lowest Bonferroni-corrected p -values being 0.11. The top two hits are in the vicinity of several known genes, the closest being the IER2 gene, which is located a little over 40,000 base pairs away from the two SNPs.⁸ As a robustness

⁸ In Table V of the online Appendix, we report the SNPs (from the set of 20 in Table 1) that are in or near any known genes, along with the distances in base pairs.

Table 1

20 Most Significant Associations from First Stage and Results from Replication Stage

SNP (chromosome)	Minor allele	Nearby genes?	First stage			Replication stage	
			β	p -value	Bonferroni	β	p -value
rs11758688† (6)	T	Yes	-0.253	$2.97 \cdot 10^{-7}$	0.107	-0.0674	0.1209
rs12527415† (6)	T	Yes	-0.253	$3.03 \cdot 10^{-7}$	0.109	-0.0689	0.1138
rs17365411 (2)	C	No	0.26	$3.73 \cdot 10^{-7}$	0.134	-0.0314	0.4553
rs7655595 (4)	G	No	-0.266	$3.99 \cdot 10^{-7}$	0.144	0.0007	0.9877
rs17350845 (1)	C	Yes	-0.291	$6.22 \cdot 10^{-7}$	0.224	0.0175	0.7162
rs12691894 (2)	G	No	-0.246	$6.67 \cdot 10^{-7}$	0.24	0.0698	0.0998
rs9646799 (2)	T	Yes	0.271	$7.41 \cdot 10^{-7}$	0.267	-0.0139	0.7579
rs11722767 (4)	C	No	-0.257	$7.77 \cdot 10^{-7}$	0.28	0.0007	0.9877
rs10947091† (6)	T	Yes	-0.245	$9.03 \cdot 10^{-7}$	0.325	-0.0674	0.1229
rs6536456 (4)	C	No	0.23	$1.32 \cdot 10^{-6}$	0.474	0.0272	0.4917
rs1580882 (4)	T	No	0.229	$1.43 \cdot 10^{-6}$	0.516	0.026	0.5105
rs6536463 (4)	G	No	0.229	$1.48 \cdot 10^{-6}$	0.533	0.0209	0.5962
rs1502720 (4)	C	No	0.228	$1.66 \cdot 10^{-6}$	0.56	0.026	0.5213
rs10436961 (1)	A	Yes	-0.268	$1.82 \cdot 10^{-6}$	0.657	0.0173	0.7196
rs4845129 (1)	G	Yes	-0.265	$2.07 \cdot 10^{-6}$	0.745	0.0175	0.7164
rs17365432 (2)	G	No	0.257	$2.32 \cdot 10^{-6}$	0.836	-0.0326	0.4675
rs11225388 (11)	G	Yes	0.261	$2.51 \cdot 10^{-6}$	0.904	-0.0648	0.1476
rs7743593 (6)	C	Yes	0.301	$2.68 \cdot 10^{-6}$	0.965	0.0172	0.7397
rs10028331 (4)	G	No	-0.259	$2.93 \cdot 10^{-6}$	1	-0.0226	0.6317
rs11964691 (6)	T	Yes	0.307	$3.29 \cdot 10^{-6}$	1	-0.0441	0.4137

Notes: This panel reports the 20 SNPs with the highest statistical significance from the first stage of the GWAS as well as results for those SNPs from both the first and the replication stages. The “Nearby genes?” column indicates whether the SNP is less than 100,000 base pairs away from any known genes. The “ p -value” columns report the (uncorrected) p -values from the first and from the replication stages, respectively. The “Bonferroni” column reports the Bonferroni-corrected p -values for the first stage, with corrections for 363,776 tests.

† These three SNPs are the least likely to be false positives, and future work in still other replication datasets may clarify what role, if any, they play.

check, we also computed standard errors by clustering at the level of the family. In general, the clustered standard errors were considerably smaller than the standard errors used to compute the p -values reported in Table 1, and in this calculation eight of the top 20 hits survived the Bonferroni correction at the 10 percent level.

The last two columns of Table 1 report the results of the replication attempt with the independent Rotterdam data, of the 20 most significant associations from the first stage. As shown, none of the top 20 SNPs has a statistically significant association with educational attainment in the Rotterdam data. In fact, the signs of the estimated beta coefficients from the first stage and the replication stage are identical for only nine of the 20 SNPs, and the correlation between the 20 regression coefficients in the two samples is 0.01. In other words, although our first-stage results might have seemed somewhat promising, the attempt to replicate them in an independent sample was a complete failure. Although this may be a false

nonreplication of true effects, we argue below that it is more likely that these initial results were spurious.

Multiple Testing, False Positives, and Power Considerations

Our examination of the Framingham data revealed a number of markers with “suggestive” associations with educational attainment, but an attempt to replicate these results in the independent Rotterdam sample failed. This experience points to a number of valuable lessons for economists interested in studying the molecular genetic basis of economic behaviors and outcomes. As noted earlier, a number of studies using the “candidate gene” approach have also found interesting associations, which then failed to replicate in independent samples.

This pattern of failed replications is not unique to economics or the social sciences. After the decoding of the Human Genome Project in the early 2000s, there was a rush to find the molecular genetic correlates of important diseases; hundreds of studies, many with strikingly small samples and often not properly controlling for multiple testing, were published to report genetic associations with some traits and diseases. However, later meta-analyses and review studies revealed that most of these associations failed to replicate (Ioannidis, 2005, 2007). Though it is possible that some of the published associations were true only in the particular populations in which they were originally found because of treatment effect heterogeneity or that they were falsely nonreplicated (Ioannidis, 2007), the most plausible explanation is that the initial findings were spurious. We are not suggesting that replications are always necessary: one can imagine cases where the corrected p -values are so small that the results can be judged convincing on the basis of one sample alone. However, as a practical matter, replication has emerged as an important method for evaluating work in this area, and we think it is advisable that economists adopt this as a convention, at least in the sense that the onus is on those who wish to depart from the default to provide a convincing motivation.

More generally, it would greatly benefit the geno-economics enterprise if economists and other social scientists learned from the mistakes that have been made in genetic research in the past decade, rather than repeat them in our own discipline. In our judgment, the plethora of false positives can be attributed to two main factors: studies that are statistically underpowered because sample sizes are too small; and bias toward publishing results that seem to show positive and significant correlations. We will discuss these in turn.

Most published association studies, especially in economics and the social sciences, are seriously underpowered given what we now know about the plausible effect sizes of common genetic variants. Consequently, the probability that an association study will detect a true signal is vanishingly small. In other words, with the small samples that are typical of many studies, the probability that an association in the data is due to a true signal is low relative to the probability that it is due to noise (in Bayesian language, the posterior odds of a true association are low). This

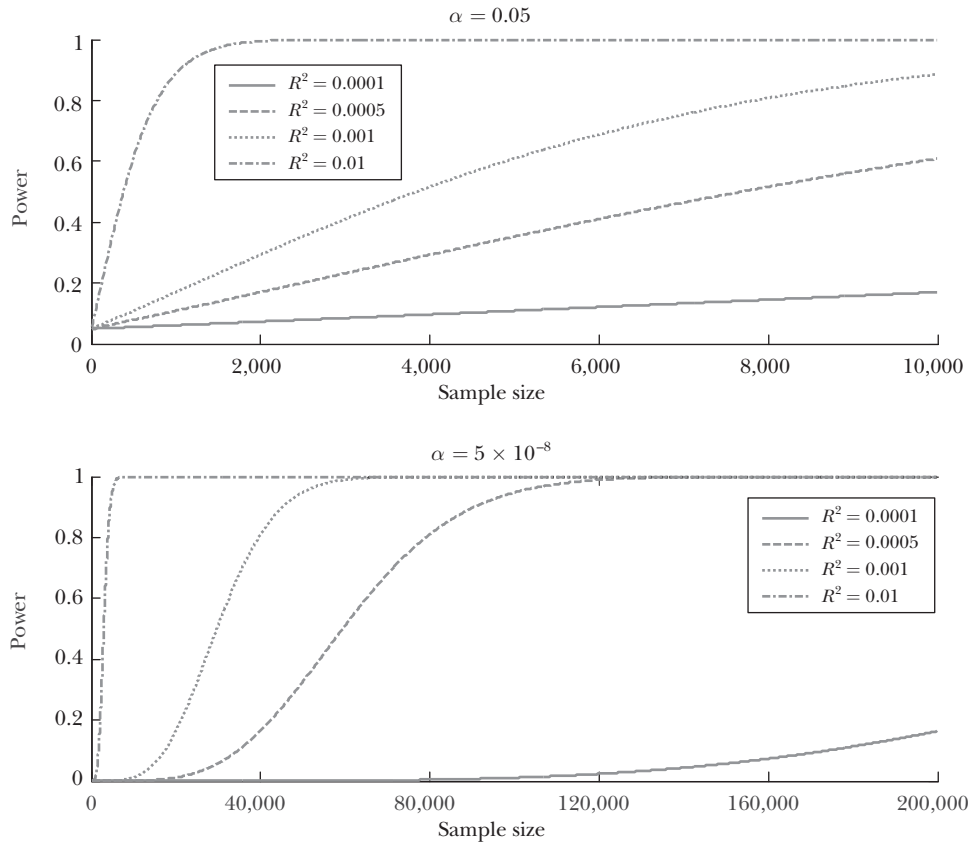
point is now close to universally accepted in the molecular genetics community and is illustrated by our above results from our genome-wide association study of educational attainment. Accumulating evidence from the genetics literature suggests that for most complex traits, the genetic variance is highly diffuse throughout the genome, that most true causal variants have very small effect sizes, and that genetic markers with an R^2 greater than 1 percent are rare (Visscher, 2008). In the case of height, which is one of the most studied traits, with a very high heritability of around 80 percent (Visscher, Hill, and Wray, 2008), the 180 loci identified by the largest study to date, with more than 180,000 subjects, only explain a total of about 10 percent of the variance in this characteristic (Allen et al., 2010).

Several nonreplicated published results notwithstanding, there is no reason to expect traits of interest to economists to be different from medical traits with respect to the distribution of the effect sizes of the genetic markers. In fact, the situation is likely to be even more difficult for economists, because socioeconomic traits are generally very distal from the genome in the chain of causation and because of the low reliability of many measured economic traits. For example, it is not clear how well behavior in economic laboratory experiments captures attitudes toward altruism, cooperation, or risk.

For most traits, thus, the genetic markers that have been found so far generally explain only a small fraction of the population variance. In contrast, estimates of the effect of heritability from behavioral genetics often hover around 30 to 50 percent or more. This gap has given rise to a debate in the genetics community about the causes of this “missing heritability” (Eichler et al., 2010; Manolio et al., 2009). Many explanations have been proposed, the main ones being about structural variants such as: CNVs and VNTRs (discussed earlier), which until recently were not typed by the main platforms; low power to detect multi-gene interactions; very rare causal variants that are not captured by the main genotyping platforms; and—perhaps the most plausible to us—very large numbers of causal variants, each with very small effect sizes. A recent paper provided support for the latter explanation as well as some hope to the genetic community: Yang et al. (2010) showed that considering all the SNPs on some common genotyping chips simultaneously—as opposed to testing each SNP individually with some stringent significance threshold (the usual approach in a genome-wide association study)—could account for 45 percent of the variance in human height. The 45 percent figure is bound to rise with more comprehensive genotyping platforms and the advent of whole genome sequencing. Similar results were recently documented for IQ (Davies et al., 2011).

With the newest genotyping platforms now integrating structural variants and increasing their coverage to include rarer SNPs, very large samples will be needed to have adequate statistical power to make progress in mapping the genetic architecture of complex traits. To illustrate this point more formally, Figure 1 shows power graphs for the conventional 5 percent level of significance often used in the social sciences and for the much more stringent 5×10^{-8} level of statistical significance. Statistical power, measured on the vertical axis, is the probability that the null hypothesis of no association will be rejected at a given significance threshold. This

Figure 1
Power Graphs



Notes: Each curve in these graphs shows the statistical power to detect a true marker–trait association as a function of the sample size. Statistical power is the probability that the null hypothesis of “no association” will be rejected at a given significance threshold. The four different curves in each panel correspond to markers that truly explain (that is, their true population R^2 's are) 0.01, 0.05, 0.1, and 1 percent of the variance of the trait. The two panels correspond to two different significance thresholds ($\alpha = 0.05$ and $\alpha = 5 \times 10^{-8}$). These power graphs are for single tests—that is, without correction for multiple hypothesis testing.

probability of rejecting the null hypothesis varies with the sample size, shown on the horizontal axis. The four different curves in each panel correspond to markers that truly explain (that is, their true population R^2 are) 0.01, 0.05, 0.1, and 1 percent of the variance of the trait. Thus, these graphs show the statistical power to detect a true marker–trait association as a function of the sample size. In the first panel, for the $\alpha = 0.05$ level of statistical significance—a significance threshold that is much higher than what is generally relevant in association studies given the multiple testing involved—a sample of about 4,000 subjects is required to have power of 50 percent to detect a marker with a true R^2 of 0.1 percent (with no correction for

multiple testing). For a marker with a true R^2 of 0.01 percent and with a sample of 10,000 individuals, the power of detecting the association is less than 20 percent. Even for a marker with a very large true R^2 of 1 percent—larger than the R^2 s of all the markers that have been found to predict height—a sample of almost 800 subjects is needed to have power of 80 percent. As can be seen in the second panel, for the $\alpha = 5 \times 10^{-8}$ level of statistical significance, sample sizes in the tens of thousands are needed to have confidence in statistical estimates if the marker has an R^2 of .01 percent, 0.05 percent, or 0.1 percent. If the marker has an R^2 of 0.01 percent, even sample sizes of 200,000 will have power of less than 20 percent.

As a result of such considerations, large multi-sample consortia are rapidly becoming the norm in medical genetics, with total sample sizes in the tens of thousands and sometimes exceeding 50,000; examples are the Wellcome Trust Case Control Consortium (2007), CHARGE (Psaty et al., 2009), and GEFOS (Rivadeneira et al., 2009). The Entrepreneur Consortium (van der Loos, Koellinger, Groenen, and Thurik, 2010), aimed at finding genetic predictors of entrepreneurship in a sample of over 60,000 individuals, is one of the first such consortia in economics; we and some collaborators are also in the process of forming a similar consortium for educational attainment. We believe consortia of this type to be the appropriate constructive response to the otherwise quite dismaying findings reported here.

The second main factor behind finding so many false positives in the literature is publication bias. It is much easier to publish “positive results” than properly powered negative findings, and this creates incentives for “data mining” and selective reporting. Researchers invest significant effort in assembling datasets with rich genotypic and phenotypic information. With little theory to discipline the empirical work, and with a still-scant biological background regarding the genetics of socioeconomic traits, the temptation is great to experiment with different specifications on different traits and markers before settling on the “right” tests to run. Unless the p -values in the subsequently published manuscript fully account for the multiple testing and the uncertainty stemming from the specification search, the resulting p -values will be too low. For example, if additive, nonadditive, and sex-specific models are estimated, but only findings from the model with lowest p -values are reported, the resulting inference will of course be incorrect. It is difficult to assess the magnitude of these problems in geno-economics, but the reported p -values in many published papers are likely to be grossly misleading. The lack of convincing replications and the known failed replications in the field are consistent with this interpretation. Journal editors and referees also have a responsibility to avoid publishing headline-grabbing results, at least when experience suggests that such results are likely to evaporate under later replication attempts. And there needs to be some reward to carefully executed and well-powered studies with negative findings. These conclusions resonate with a recently published editorial policy of the leading field journal *Behavior Genetics*, a policy written in response to the frequent replication failures in human molecular genetics (Hewitt, 2011).

In our view, the problem is not the fact that the approach of the genome-wide association study is a largely atheoretical exercise. Exploratory data mining is a perfectly legitimate scientific endeavor when it is identified and treated as such. Many of the replicable findings that thus far have emerged from genome-wide association studies of medical outcomes were unexpected and unlikely to have been uncovered from candidate-based approaches. But data mining must be done with discipline, meaning a complete reporting of and correction for the number of hypotheses tested, including any pretesting of the model specification.

The nature of molecular genetic data makes the multiple hypothesis testing problem particularly acute, but there may also be a silver lining. In particular, it is generally relatively easy to attempt a replication of a genoeconomic finding in independent samples. Moreover, awareness of these issues is growing among genoeconomists. Indeed, we believe that economists have much to learn from the different scientific culture that prevails in the medical sciences in general and in genetics in particular. Though the problems of multiple hypothesis testing and publication bias have received some attention in economics (for example, Anderson, 2008; Card and Krueger, 1995; DeLong and Lang, 1992; Kling, Liebman, and Katz, 2007), they are generally overlooked. A culture that encourages—when this is technically feasible—collaboration across groups of authors to pool data and to replicate results across datasets would doubtless contribute to this desirable goal.

Molecular Genetic Data Providers

Most available genotyped samples have been collected for medical research purposes. These studies vary in sample size from a few hundred to several thousand observations. In addition to medical outcomes, the studies usually also collect some background information about their participants that are interesting for economists, such as education, profession, labor market experience, and income. Many of the large data-providers have formed research consortia, the most important of which are the CHARGE and GIANT consortia. As the costs of genotyping technologies are falling rapidly, social science data providers are expressing greater interest in adding molecular genetic information to their lists of variables. The first major U.S. data provider to include such information was the Longitudinal Study of Adolescent Health, which, as part of its third wave of data collection between 1999 and 2001, collected saliva samples and genotyped approximately 2,500 individuals for six genetic markers on or near genes linked to functions of the neurotransmitters dopamine and serotonin. A major provider of genoeconomic data in the coming years will be the U.S. Health and Retirement Survey, which is presently being expanded to include genome-wide association data—over one million genetic markers—in approximately 13,000 subjects. Given the large sample, and the rich set of phenotypic data on health, psychological well-being, and economic and financial behavior, we anticipate that this data will emerge as

a valuable resource for researchers in the social sciences interested in studying human genetic variation.

We anticipate that a steady stream of new datasets with economic variables and comprehensively genotyped subjects will be made available to researchers over time. One promising approach for the near future is to focus on the collection of additional, standardized socioeconomic variables in samples that already contain genotyped data. Over time, adding genotype information to existing socioeconomic datasets should provide an important new source of data to economists interested in studying the molecular genetic associates of economic behaviors and outcomes. Socioeconomic datasets already contain a wealth of relevant variables and are unlikely to be matched by data collected for other purposes in terms of measurement accuracy and breadth or duration of longitudinal follow-up. We further anticipate the formation of geno-economic research consortia for the social sciences. An obvious challenge with this approach is that for research consortia to be effective, everyone will have to measure the same set of traits using reliable batteries that are uniform across surveys. The National Human Genome Research Institute is funding the PhenX project to develop consensus measures for phenotypes and exposures to be used in genome-wide association studies (<https://www.phenx.org/>).

Potential Benefits of Molecular Genetic Data

We have thus far discussed a number of methodological challenges that face researchers who use molecular genetics to reliably identify genetic associates of economic traits. In addition, the possibility of genome-based information and interventions raise a host of privacy and ethical issues that we have not pursued here. Despite those difficulties, we believe there are several reasons, beyond sheer intellectual curiosity, why the use of molecular genetic data will ultimately benefit the economic sciences.

First, knowledge of the biological mechanisms might suggest additional policies or interventions that had previously not been anticipated. There are already examples in the medical literature of genome-wide association studies leading to unanticipated findings that in turn have helped inspire new therapies (Hirschhorn, 2009). It is not a necessary condition that the genetic markers identified have large effects on the outcome, only that the biological pathway they implicate provides useful cues about possible prevention mechanisms. For some behavioral outcomes, genetic markers may also be of diagnostic and predictive utility, regardless of whether or not they shed direct light on biological mechanisms. For example, Benjamin (2010) points out that if dyslexia can eventually be predicted sufficiently well by genetic screening, children with dyslexia-susceptibility markers could be taught differently how to read from a very young age. More generally, if molecular genetic data can be used to predict which individuals are at high “genetic risk” for adverse outcomes, then it would in principle be possible to use such information

to evaluate the costs and benefits of taking preventive measures targeted at helping at-risk individuals.

Given the current state of knowledge, the predictive utility of the genetic markers that have been identified thus far (with very few, but important, exceptions) is too weak for these hopes to be realized any time soon for the overwhelming majority of potential outcomes. The initial promises that genomics would revolutionize “personalized medicine” have not yet been realized. If anything, the main use of genotyping technology has proven to be prognostic, not therapeutic, as anticipated by some (Christakis, 1999, p.196). Nonetheless, in many applications where the purpose is risk prediction, several statistical techniques may yield informative predictions without having precise estimates of the effect of individual markers, by making efficient use of all the available genotypic information, using so called polygenic risk scores (Purcell et al., 2009). For such approaches to have practical value in the social sciences, larger sample sizes than those presently available will be needed because prediction accuracy is presently too poor for complex traits. As increasingly rich genetic information becomes available, and as new statistical techniques are developed, the predictive utility of the genetic data will doubtless rise.

Second, analysis of molecular genetic data may also help empirical economists to estimate causal effects. As genetic endowments become increasingly easy to measure, it will be possible to control for genetic determinants of the dependent variable, thereby reducing omitted variable bias (Bernheim, 2009). For example, a successful genome-wide association study of educational attainment might be able to identify genetic markers that influence the general likelihood of high intellectual performance, delayed gratification, and persistence. Such markers would be a valuable control variable in empirical studies on a number of economic questions, including income, labor supply, occupational choice, or savings decisions (though the absence of these variables is not a source of bias in a properly executed randomized evaluation).

Third, a more speculative empirical use of genetic data is to use markers as instruments in empirical settings where reverse causality or omitted variable bias is otherwise a concern. This use of genetic markers was anticipated by Davey Smith and Ebrahim (2003) and was pioneered in economic analyses by Ding, Lehrer, Rosenquist, and Audrain-McGovern (2006), who used molecular genetic data to instrument for health in a regression of academic outcomes on health. To serve as valid instrument, the marker Z needs to impact the outcome Y only through its impact on the intervening variable X . This chain of causation will be exceedingly, perhaps impossibly, difficult to establish in many cases that interest economists (Conley, 2009; Cawley, Han, and Norton, 2011). There are two major challenges. First, many markers are pleiotropic, meaning that they influence multiple outcomes directly. If the marker Z also affects Y through other unobserved channels for which the researcher cannot control, then the so-called exclusion restriction fails. Second, even if the marker does not have pleiotropic effects, it may be correlated (in linkage disequilibrium) with a genetic marker that affects Y through the unobservables. Economists will likely be in a better position to assess the plausibility that a given marker satisfies the exclusion restriction once its biological function is intricately

understood. Conley (2009) argues that it is very unlikely that any markers which satisfy the requirements of a valid instrument will ever be found.

Finally, and perhaps most importantly, molecular genetic data may prove helpful in understanding variation in policy response across individuals. Plausibly uncovering such treatment-effect heterogeneity will require a detailed research design where plausibly exogenous environmental variation is used to estimate the gene–environment interactions (Conley, 2009). The most frequently cited example of a gene–environment interaction is the work of Caspi et al. (2002). These authors studied the determinants of a composite index of antisocial behavior in a sample of 1,037 male children in New Zealand. In a regression of their index on a particular polymorphism in a gene called MAOA, childhood maltreatment, and the interaction of these two variables, they found a significant interaction and concluded that the MAOA gene modifies the influence of the childhood maltreatment. While this paper is a useful illustration of the possibilities in this area, it is important to note that childhood maltreatment is not necessarily exogenous but may be correlated with other unobserved determinants of poor outcomes. Also, a recent meta-study (Risch et al., 2009) was unable to corroborate the Caspi et al. (2002) findings. This is not too surprising because the study is vulnerable to exactly the same criticisms that we have made against candidate genes studies in economics. Economists have a powerful toolkit for the estimation and identification of causal effects, which has the potential to contribute to the current gene–environment interaction literature. Presently, this literature suffers from the limitation that the environmental variables used may be endogenous, as existing studies are not based on experimentally or quasi-experimentally generated exogenous variation in environment (Conley, 2009).

The cost of genotyping has now fallen to a level where it is feasible to directly study molecular genetic variation, and many social science surveys are in the process of incorporating this information. We believe that such data will add a valuable new understanding to one dimension of heterogeneity in economic behavior. If researchers adhere to high empirical standards, we are confident that geno-economics will generate a number of durable scientific insights that will ultimately lead to a richer and more comprehensive economic science.

■ *The authors wish to thank the JEP editorial team, along with Anders Björklund, Markus Jäntti, Sandy Jencks, and Peter Visscher, for comments on an earlier version of this paper. We are also grateful to Sarah Medland and Peter Visscher for the intellectual generosity they have shown us in responding to a number of methodological queries. Given that the list of authors is a lengthy one for an economics paper, we spell out contributions: Beauchamp and Cesarini were the primary authors for the paper, while Koellinger and van der Loos contributed to the writing. Beauchamp designed the research protocol and analyzed the Framingham data. Van der Loos analyzed the Rotterdam data. Beauchamp, Cesarini, Christakis, and Johannesson conceived the project and provided impetus. Christakis, Johannesson, Fowler, Groenen, Rosenquist, and Thurik critically reviewed the manuscript. As always, any views expressed here are those of the authors alone, and we retain responsibility for any errors.*

References

- Affymetrix.** 2009. "Affymetrix® Genome-Wide Human SNP Array 5.0." Data Sheet. http://www.affymetrix.com/support/technical/datasheets/genomewide_snp5_datasheet.pdf.
- Allen, Hana Lango, et al.** 2010. "Hundreds of Variants Clustered in Genomic Loci and Biological Pathways Affect Human Height." *Nature*, 467(7317): 832–838.
- Anderson, Michael L.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–95.
- Apicella, Coren L., David Cesarini, Magnus Johannesson, Christopher T. Dawes, Paul Lichtenstein, Björn Wallace, Jonathan Beauchamp, and Lars Westberg.** 2010. "No Association between Oxytocin Receptor (OXTR) Gene Polymorphisms and Experimentally Elicited Social Preferences." *PLoS ONE*, 5(6): e11153.
- Barnea, Amir, Henrik Cronqvist, and Stephan Siegel.** 2010. "Nature or Nurture: What Determines Investor Behavior?" *Journal of Financial Economics*, 98(3): 583–604.
- Beauchamp, Jonathan P., David Cesarini, and Magnus Johannesson.** 2011. "The Psychometric Properties of Measures of Economic Risk Preferences." Unpublished paper, Harvard University.
- Beauchamp, Jonathan, David Cesarini, J. Niels Rosenquist, James H. Fowler, and Nicholas A. Christakis.** 2010. "A Genome-Wide Association Study of Educational Attainment." Working paper, presented at IZA conference *Genes, Brains and the Labor Market*.
- Becker, Gary S., and Nigel Tomes.** 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *The Journal of Political Economy*, 87(6): 1153–89.
- Benjamin, Daniel J., Christopher F. Chabris, Edward L. Glaeser, Vilmundur Gudnason, Tamara B. Harris, David I. Laibson, Lenore J. Launer, and Shaun Purcell** 2007. "Genoecconomics." In *Biosocial Surveys*, eds. Maxine Weinstein, James W. Vaupel, and Kenneth W. Wachter, 192–289. Washington: National Academies Press.
- Benjamin, Daniel J.** 2010. "A White Paper." Chapter 6 of *Genes, Cognition, and Social Behavior: Next Steps for Foundations and Researchers*, the final report for the NSF Workshop on Genes, Cognition, and Social Behavior, prepared by Arthur Lupia. Link to report: http://economics.cornell.edu/dbenjamin/NSF_Report_Final.pdf.
- Bernheim, B. Douglas.** 2009. "On the Potential of Neuroeconomics: A Critical (But Hopeful) Appraisal." *American Economic Journal: Microeconomics*, 1(2): 1–41.
- Björklund, Anders, Markus Jäntti, and Gary Solon.** 2005. "Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types in Sweden." In *Unequal Chances: Family Background and Economic Success*, eds. Samuel Bowles, Herbert Gintis, and Melissa Osborne Graves. Princeton: Princeton University Press.
- Björklund, Anders, Mikael Lindahl, and Erik Plug.** 2006. "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data." *Quarterly Journal of Economics*, 121(3): 999–1028.
- Bouchard, Thomas J. Jr., and Matt McGue.** 2003. "Genetic and Environmental Influences on Human Psychological Differences." *Journal of Neurobiology*, 54(1): 4–45.
- Campbell, Catarina D. et al.** 2005. "Demonstrating Stratification in a European American Population." *Nature Genetics*, 37(8): 868–72.
- Card, David, and Alan B. Krueger.** 1995. "Time-Series Minimum-Wage Studies: A Meta Analysis." *American Economic Review*, 85(2): 238–43.
- Carey, Gregory.** 2003. *Human Genetics for the Social Sciences*. Thousand Oaks: Sage Publications.
- Carpenter, Jeffrey, Justin Garcia, and J. Koji Lum.** 2011. "Dopamine Receptor Genes Predict Risk Preferences, Time Preferences, and Related Economic Outcomes." *Journal of Risk and Uncertainty*, 42(3): 233–61.
- Caspi, Avshalom, Joseph McClay, Terrie E. Moffitt, Jonathan Mill, Judy Martin, Ian W. Craig, Alan Taylor, and Richie Poulton** 2002. "Role of Genotype in the Cycle of Violence in Maltreated Children." *Science*, 297(5582): 851–54.
- Cawley, John, Euna Han, and Edward C. Norton.** 2011. "The Validity of Genes Related to Neurotransmitters as Instrumental Variables." *Health Economics*, 20(8): 884–88.
- Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace.** 2009a. "Genetic Variation in Preferences for Giving and Risk-Taking." *Quarterly Journal of Economics*, 124(2): 809–41.
- Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace.** 2009b. "Heritability of Overconfidence." *Journal of the European Economic Association*, 7(2–3): 617–27.
- Cesarini, David, Magnus Johannesson, Paul Lichtenstein, Örjan Sandewall, and Björn Wallace.**

2010. "Genetic Variation in Financial Decision-Making." *Journal of Finance*, 65(5): 1725–54.
- Christakis, Nicholas A.** 1999. *Death Foretold: Prophecy and Prognosis in Medical Care*. Chicago: University of Chicago Press.
- Churchill, Gary A., and Rebecca W. Doerge.** 1994. "Empirical Threshold Values for Quantitative Trait Mapping." *Genetics*, 138(3): 963–71.
- Conley, Dalton.** 2009. "The Promise and Challenges of Incorporating Genetic Data into Longitudinal Social Science Surveys and Research." *Biodemography and Social Biology*, 55(2): 238–51.
- Crisan, Liviu G., Simona Pana, Romana Vulturar, Renata M. Heilman, Raluca Szekely, Bogdan Druga, Nicolae Dragos, and Andrei C. Miu.** 2009. "Genetic Contributions of the Serotonin Transporter to Social Learning of Fear and Economic Decision Making." *Social Cognitive and Affective Neuroscience*, 4(4): 399–408.
- Davey Smith, George, and Shah Ebrahim.** 2003. "Mendelian Randomization: Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease." *International Journal of Epidemiology*, 32(1): 1–22.
- Davies, Gail, et al.** 2011. "Genome-Wide Association Studies Establish that Human Intelligence is Highly Heritable and Polygenic." *Molecular Psychiatry*, August 9.
- Dawber, Thomas R., Gilcin F. Meadors, and Felix E. Moore, Jr.** 1951. "Epidemiological Approaches to Heart Disease: The Framingham Study." *American Journal Public Health*, 41(3): 279–286.
- DeLong, Bradford J., and Kevin Lang.** 1992. "Are All Economic Hypotheses False?" *Journal of Political Economy*, 100(6): 1257–72.
- Dickens, William T., and James R. Flynn.** 2001. "Heritability Estimates versus Large Environmental Effects: The IQ Paradox Resolved." *Psychological Review*, 108(2): 346–69.
- Ding, Weili, Steven F. Lehrer, J. Niels Rosenquist, and Janet Audrain-McGovern.** 2006. "The Impact of Poor Health on Education: New Evidence Using Genetic Markers." *NBER Working Paper* w12304.
- Dreber, Anna, Coren L. Apicella, Daniel T. A. Eisenberg, Justin R. Garcia, Richard Zamore, J. Koji Lum, and Benjamin Campbell.** 2009. "The 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4) Is Associated with Financial Risk-Taking in Men." *Evolution and Human Behavior*, 30(2): 85–92.
- Ebstein, Richard P., Salomon Israel, Soo Hong Chew, Songfa Zhong, and Ariel Knafo.** 2010. "Genetics of Human Social Behavior." *Neuron*, 65(6): 831–44.
- Eichler, Evan E., Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M. Leal, Jason H. Moore, and Joseph H. Nadeau.** 2010. "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease." *Nature Reviews Genetics*, 11(6): 446–50.
- Feinleib, Manning, William B. Kannel, Robert J. Garrison, Patricia M. McNamara, and William P. Castelli.** 1975. "The Framingham Offspring Study: Design and Preliminary Data." *Preventive Medicine*, 4(4): 518–52.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton.** 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6(3): 347–69.
- Fowler, James H., Jaime Settle, and Nicholas A. Christakis.** 2011. "Correlated Genotypes in Friendship Networks." *PNAS*, 108(5): 1993–97.
- Gao, Xiaoyi, Lewis C. Becker, Diane M. Becker, Joshua D. Starmer, and Michael A. Province.** 2010. "Avoiding the High Bonferroni Penalty in Genome-Wide Association Studies." *Genetic Epidemiology*, 34(1): 100–105.
- Goldberger, Arthur S.** 1979. "Heritability." *Economica*, 46(184): 327–47.
- Hamer, Dean H., and Leo Sirota.** 2000. "Beware the Chopstick Gene." *Molecular Psychiatry*, 5(1): 11–13.
- Hartl, Daniel L.** 1988. *A Primer of Population Genetics*. Sunderland: Sinauer Associates.
- Hewitt, John K.** 2011. "Editorial Policy on Candidate Gene Association and Candidate Gene-by-Environment Interaction Studies of Complex Traits." *Behavior Genetics*. Published online September 18. doi: 10.1007/s10519-011-9504-z.
- Hochberg, Yosef, and Yoav Benjamini.** 1990. "More Powerful Procedures for Multiple Significance Testing." *Statistics in Medicine*, 9(7): 811–18.
- Hofman, Albert, et al.** 2009. "The Rotterdam Study: 2010 Objectives and Design Update." *European Journal of Epidemiology*, 24(9): 553–72.
- Hill, William G., Michael E. Goddard, and Peter M. Visscher.** 2008. "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." *PLoS Genetics*, 4(2): e1000008.
- Hirschhorn, J. N.** 2009. "Genomewide Association Studies—Illuminating Biologic Pathways." *New England Journal of Medicine*, 360(17): 1699–1701.
- International HapMap Consortium.** 2003. "The International HapMap Project." *Nature*, 426(6968): 789–96.
- Ioannidis, John P. A.** 2005. "Why Most Published Research Findings Are False." *PLoS Medicine*, 2(8): e124.
- Ioannidis, John P. A.** 2007. "Non-Replication and Inconsistency in the Genome-Wide Association Setting." *Human Heredity*, 64(4): 203–13.

- Ioannidis, John P. A., Evangelia E. Ntzani, Thomas A. Trikalinos, and Despina G. Contopoulos-Ioannidis.** 2001. "Replication Validity of Genetic Association Studies." *Nature Genetics*, 29(3): 306–309.
- Israel, Salomon, et al.** 2009. "The Oxytocin Receptor (OXTR) Contributes to Prosocial Fund Allocations in the Dictator Game and the Social Value Orientations Task." *PLoS ONE*, 4(5): e5535.
- Jencks, Christopher.** 1980. "Heredity, Environment, and Public Policy Reconsidered." *American Sociological Review*, 45(5): 723–36.
- Jencks, Christopher, and Marsha Brown.** 1977. "Genes and Social Stratification." In *Kinometrics*, ed. Paul Taubman, 169–233. Amsterdam: North-Holland.
- Kidd, Jeffrey M. et al.** 2008. "Mapping and Sequencing of Structural Variation from Eight Human Genomes." *Nature*, 453(7191): 56–64.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz.** 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.
- Knafo, Ariel, et al.** 2008. "Individual Differences in Allocation of Funds in the Dictator Game Associated with Length of the Arginine Vasopressin 1a Receptor (AVPR1a) RS3 Promoter Region and Correlation between RS3 Length and Hippocampal mRNA." *Genes, Brain and Behavior*, 7(3): 266–75.
- Kosfeld, Michael, Marcus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr.** 2005. "Oxytocin Increases Trust in Humans." *Nature*, 435(7042): 673–76.
- Kuhnen, Camelia M., and Joan Y. Chiao.** 2009. "Genetic Determinants of Financial Risk Taking." *PLoS ONE*, 4(2): e4362.
- Lander, Eric S., et al.** 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature*, 409(6822): 860–921.
- Levy, Daniel, and Susan Brink.** 2005. *A Change of Heart: How the People of Framingham, Massachusetts Helped Unravel the Mysteries of Cardiovascular Disease*. New York: Knopf.
- Lichtenstein, Paul, Nancy L. Pedersen, and Gerald E. McClearn.** 1992. "The Origins of Individual Differences in Occupational Status and Educational Level." *Acta Sociologica*, 35(1): 13–31.
- Lizzeri, Alessandro, and Marciano Siniscalchi.** 2008. "Parental Guidance and Supervised Learning." *Quarterly Journal of Economics*, 123(3): 1161–95.
- Mackay, Trudy.** 2001. "The Genetic Architecture of Quantitative Traits." *Annual Review of Genetics*, vol. 35, pp. 303–39.
- Manolio, Teri A., et al.** 2009. "Finding the Missing Heritability of Complex Diseases." *Nature*, 461(7265): 747–53.
- Martin, Nicholas W.** 2011. "Educational Attainment: A Genome Wide Association Study in 9538 Australians." *PLoS One*, 6(6): e20128. doi: 10.1371/journal.pone.0020128.
- McCarthy, Mark I., Goncalo R. Abecasis, Lon R. Cardon, David B. Goldstein, Julian Little, John P. A. Ioannidis, and Joel N. Hirschhorn** 2008. "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges." *Nature Review Genetics*, 9(5): 356–69.
- McDermott, Rose, Dustin Tingley, Jonathan Cowden, Giovanni Frazzetto, and Dominic D. P. Johnson.** 2009. "Monoamine Oxidase A Gene (MAOA) Predicts Behavioral Aggression Following Provocation." *PNAS*, 106(7): 2118–23.
- Nicolaou, Nicos, Scott Shane, Georgina Adi, Massimo Mangino, and Juliette Harris.** 2011. "A Polymorphism Associated with Entrepreneurship: Evidence from Dopamine Receptor Candidate Genes." *Small Business Economics*, 36(2): 151–55.
- Ollier, William, Tim Sprosen, and Tim Peakman.** 2005. "UK Biobank: From Concept to Reality." *Pharmacogenomics*, 6(6): 639–46.
- Pearson, Thomas A., and Teri A. Manolio.** 2008. "How to Interpret a Genome-Wide Association Study." *JAMA*, 299(11): 1335–44.
- Plomin, Robert, John C. DeFries, Gerald E. McClearn, and Peter McGuffin.** 2008. *Behavioral Genetics*, 5th ed. New York: Worth Publishers.
- Posthuma, Danielle, et al.** 2008. Abstract from "Behavioral Genetical Association 38th Annual Meeting Abstracts" titled "A Genome-Wide Association Study for Educational Attainment." *Behavior Genetics*, 38(6): 643–44.
- Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich** 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics*, 38(8): 904–909.
- Psaty, Bruce M., Christopher J. O'Donnell, Vilmundur Gudnason, Kathryn L. Lunetta, Aaron R. Folsom, Jerome I. Rotter, André G. Uitterlinden, Tamara B. Harris, Jacqueline Witteman, and Eric Boerwinkle, on behalf of the CHARGE Consortium.** 2009. "Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies from Five Cohorts." *Circulation: Cardiovascular Genetics*, 2(1): 73–80.
- Purcell, Shaun M., et al.** 2009. "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder." *Nature*, 460(7256): 748–52.
- Ridley, Matt.** 2003. *Nature via Nurture: Genes, Experience, and What Makes Us Human*. New York: Harper Collins Publishers.

- Rivadeneira, Fernando, et al.** 2009. "Twenty Bone-Mineral-Density Loci Identified by Large-Scale Meta-Analysis of Genome-Wide Association Studies." *Nature Genetics*, 41(11): 1199–206.
- Risch, Neil, et al.** 2009. "Interaction between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression: A Meta-Analysis." *JAMA*, 301(23): 2462–71.
- Roe, Brian E., Michael R. Tilley, Howard H. Gu, David Q. Beversdorf, Wolfgang Sadée, Timothy C. Haab, Audrey C. Papp.** 2009. "Financial and Psychological Risk Attitudes Associated with Two Single Nucleotide Polymorphisms in the Nicotine Receptor (CHRNA4) Gene." *PLoS ONE*, 4(8): e6704.
- Sabatti, Chiara, Susan Service, and Nelson Freimer.** 2003. "False Discovery Rate in Linkage and Association Genome Screens for Complex Disorders." *Genetics*, 164(2): 829–33.
- Sacerdote, Bruce.** 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *Quarterly Journal of Economics*, 122(1): 119–57.
- Sacerdote, Bruce.** 2010. "Nature and Nurture Effects on Children's Outcomes: What Have We Learned from Studies of Twins and Adoptees?" In *Handbook of Social Economics*, vol 1A, ed. Jess Benhabib, Matthew Jackson, and Alberto Bisin, 1–30. Amsterdam: North Holland.
- Schultz, Wolfram, Peter Dayan, and P. Read Montague.** 1997. "A Neural Substrate of Prediction and Reward." *Science*, 275(5306): 1593.
- Stephens, Matthew, and David J. Balding.** 2009. "Bayesian Statistical Methods for Genetic Association Studies." *Nature Reviews Genetics*, 10(10): 681–90.
- Strachan Tom, and Andrew Read.** 2003. *Human Molecular Genetics*, 3rd ed. London: Garland Science/Taylor & Francis Group.
- Sullivan, Patrick F., and Shaun Purcell.** 2008. "Analyzing Genome-Wide Association Study Data: A Tutorial Using PLINK." In *Statistical Genetics; Gene Mapping through Linkage and Association*, ed. Benjamin M. Neale, Manuel A. R. Ferreira, Sarah Medland, and Danielle Posthuma, 355–94. New York: Taylor & Francis Group.
- Taubman, Paul.** 1976. "The Determinants of Earnings: Genetics, Family and Other Environments: A Study of White Male Twins." *American Economic Review*, 66(5): 858–70.
- Turkheimer, Eric.** 2000. "Three Laws of Behavior Genetics and What They Mean." *Current Directions in Psychological Science*, 9(5): 160–64.
- van der Loos, Matthijs J. H. M., et al.** 2011. "Candidate Gene Studies and the Quest for the Entrepreneurial Gene." *Small Business Economics*, 37(3): 269–75.
- van der Loos, Matthijs J. H. M., Philipp D. Koellinger, Patrick J. F. Groenen, and A. Roy Thurik.** 2010. "Genome-Wide Association Studies and the Genetics of Entrepreneurship." *European Journal of Epidemiology*, 25(1): 1–3.
- Venter, J. Craig, et al.** 2001. "The Sequence of the Human Genome." *Science*, 291(5507): 1304–51.
- Visscher, Peter M.** 2008. "Sizing up Human Height Variation." *Nature Genetics*, 40(5): 489–90.
- Visscher, Peter M., William G. Hill, and Naomi R. Wray.** 2008. "Heritability in the Genomics Era Concepts and Misconceptions." *Nature Reviews Genetics*, 9(4): 255–66.
- Visscher, Peter M., Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, and Nicholas G. Martin.** 2006. "Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings." *PLoS Genetics*, 2(3): e41.
- Wallace, Björn, David Cesarini, Magnus Johannesson, and Paul Lichtenstein.** 2007. "Heritability of Ultimatum Game Responder Behavior." *PNAS*, 104(40): 15631–34.
- Wellcome Trust Case Control Consortium.** 2007. "Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls." *Nature*, 447(7145): 661–78.
- Yang, Jian, et al.** 2010. "Common SNPs Explain a Large Proportion of the Heritability for Human Height." *Nature Genetics*, 42(7): 565–69.
- Zhong, Hua, and Ross L. Prentice.** 2008. "Bias-Reduced Estimators and Confidence Intervals for Odds Ratios in Genome-Wide Association Studies." *Biostatistics*, 9(4): 621–34.
- Zhong, Songfa, Salomon Israel, Hong Xue, Richard P. Ebstein, and Soo Hong Chew.** 2009. "Monoamine Oxidase A Gene (MAOA) Associated with Attitude towards Longshot Risks." *PLoS ONE*, 4(12): e8516.
- Zhong, Songfa, Salomon Israel, Hong Xue, Pak C. Sham, Richard P. Ebstein, and Soo Hong Chew.** 2009. "A Neurochemical Approach to Valuation Sensitivity over Gains and Losses." *Proceedings of the Royal Society B: Biological Sciences*, 276(1676): 4181–88.