CHAPTER 1

MOLECULAR ME

WELCOME TO THE COMING SOCIAL GENOMICS REVOLUTION

The social genomics revolution is now upon us. In a little more than one hundred years, a gene—a discrete unit of heredity—has been transformed from a vague notion among a small group of scientists to almost a ubiquitous news item and a consumer commodity. The black box of the genome has been cracked open by inexpensive DNA genotyping platforms that for a mere hundred dollars allow one to measure around a million of the base-pair differences between individuals. We can now all drink from a firehose of new data and research findings aimed at describing the underlying genetic architecture of human health and well-being. In addition to the large number of biologists and medical scientists working in this vein, a small but growing group of sociologists, political scientists, and economists have joined forces with statistical geneticists to make serious arguments about the role of genes in the broader domain of human social dynamics and inequalities.

A note on how to read this book: We wrote the main text with the general reader in mind. Endnotes and appendices to each chapter are meant for those who want to dig deeper into the technical details and scientific minutia.

This component of the genomics revolution is somewhat unusual in that it follows a long history during which geneticists and social scientists have been unwilling to work with each other on important scientific issues. Indeed, for most of history since Darwin penned *The Descent of Man*—his 1871 follow-up to *The Origin of Species* that focused on the evolution of human differences—there has been a contentious relationship between biologists and social scientists, with many examples of interdisciplinary exchanges having devastating impacts on society. There was Herbert Spencer, who applied natural selection as a metaphor to human society leading to the ideology of "social Darwinism," which justified inaction in addressing all manner of social ills and inequities. And there was Darwin's cousin, Francis Galton, who pioneered eugenics. Even Darwin himself became embroiled in a debate about whether blacks and whites constituted separate species.[1]

Part of the reason for trepidation by social scientists when it comes to the examination of genetics as it relates to human behavior is that it is commonly assumed that the answers obtained by looking at genetics are deterministic ones, which is at odds with much of the social science enterprise. Furthermore, to the extent that genes explain any social phenomenon, this fact "naturalizes" inequalities in that outcome. In other words, the degree to which IQ (or height) differences between people are genetically influenced might anchor beliefs that these outcomes are innate and thus immutable. If these "natural" differences between people extend to outcomes such as education levels or earnings (i.e., income inequality), such inequalities may be argued to be natural, unchangeable, and thus outside the realm of policy intervention—"naturalizing" or rationalizing inequity.[2]

One important example of the naturalization of social or economic inequality can be found in the best-selling book, *The Bell Curve: Intelligence and Class Structure in American Life*, in which Richard Herrnstein and Charles Murray argued that thanks to meritocracy, class stratification today is based on innate (i.e., genetic) endowment.[3] By selectively breeding with others of similar genetic stock, parents reinforce their offsprings' advantages or disadvantages. According to these authors, social policy to promote equal opportunity is counterpro-

ductive since individuals have reached the level of social status best suited to their native abilities. This conclusion is the nightmare of progressive social scientists and the primary reason most of them avoid genetic data.[4]

We, instead, head straight into this domain of inquiry with eyes wide open. The genetics of inequality is, in fact, a major theme of this book. Specifically, we ask how integrating molecular genetic information into social scientific inquiry informs debates about inequality and socioeconomic attainment (of individuals as well as of entire nations). We argue that there are three main ways.

First, by dealing directly with the contention that innate, inherited differences are the primary engine of social inequality, the integration of genetic markers shows the residual social inequalities in stark relief. By actively accounting for the portion of IQ, education, or income that is the result of genes, we can see more clearly the inequities in environmental inputs and their effects on individuals' chances in the game of life.

Second, we show how genotype acts as a prism, refracting the white light of average effects into a rainbow of clearly observable differential effects and outcomes. Our intuition is that genotype is a tool that will help us understand why, for example, childhood poverty wreaks havoc on some individuals whereas others are resilient to such traumas. Or, by explicitly integrating genetic information into social science, we can go from the adage that a gene for aggression lands you in jail if you are from the ghetto but in the boardroom if you are to the manor born, to a scientific research agenda showing how environmental and genetic effects mutually depend upon each other.

Third, we think that public policy will have to deal with this new information as lay citizens get a hold of their own (and others') genetic data. There is much written about privacy, "genoism" (genetic discrimination by the likes of insurance companies), and personalized medicine. We shift attention from these topics to tackle the more traditional social policy domains of education, income support, economic development, and labor markets and interrogate the implications of genotype for those realms. We argue that, contrary to claims in *The Bell Curve*, we have not yet become a genotocracy (a society

ruled by the genetically advantaged), but that would not be such a far-fetched possibility once those with power and resources start to control their own genetic information and use it to selectively breed themselves. The social genomics revolution means that new forms of inequality may emerge based not only on genotype but also on whether individuals know their genotypes (and the genotypes of those around them) and can act on that information.

This revolution, however, ran into some stumbling blocks early. A few initial successes in finding "the gene for X" (such as age-related macular degeneration) produced false hope and optimism that uncovering the genetic basis for much common disease, and even socioeconomic outcomes, was within reach.[5] These successes were followed by failures—both null findings for some outcomes and (in hindsight) false insights into others. The lessons of statistical research—the need for adequate sample sizes and clear hypotheses—were learned anew the hard way by this emerging field. The result of these and other challenges was that measured genetic markers did not seem to explain the amount of variation in outcomes—ranging from schizophrenia to height to IQ—that was predicted to be due to genetics according to the prior generation of studies. For example, some earlier heritability estimates for schizophrenia reached higher than 80 percent,[6] yet some of the best studies that used DNA data produced heritability estimates closer to 3 percent—leading some scientists to dub this problem "missing heritability."[7] (We discuss heritability in chapter 2 and the mystery of missing heritability in chapter 3.)

Recently, geneticists began to reevaluate the reasons that this heritability was missing. Perhaps because the earlier studies used twins and family relationship data instead of actual DNA data, the role of genes was overestimated to begin with. Or maybe researchers had been looking in the wrong places in the genome. Genotyping companies have traditionally focused on measuring common genetic variants rather than rare ones. Perhaps there are still genes for "X" (e.g., schizophrenia) that remain undiscovered because they are rare in the population and not typically measured in the DNA data that are available. Slowly but surely, statistical geneticists have made significant

progress in solving this "missing heritability" mystery by using a range of newly developed tools.

The current and more widely accepted answer to this missingness relies on a paradigm shift—moving from a mindset of a "gene for X" to one of "many variants with small effects." Instead of a single important genetic variant (or allele), there are often hundreds or thousands that contribute to variation in a given outcome. But the "small effects" aspect of this paradigm has called for ever-larger data sets to find the needles in the genomic haystack because the needles are now thought to be much smaller than originally suspected.

Along with this shift, more national surveys have been asking respondents to spit into a cup, adding genotype data to the rich tapestry of social variables that economists, sociologists, and political scientists had previously used. It seems that genetics has finally gained a foothold in social science. And why not? Why should we be afraid of additional data that may help scientists better understand patterns of human behavior, enhance individuals' self-understanding, and design optimal public policy? Why be apprehensive, especially when the answers we get from carefully peering into the black box are not always—or even often—the kind that crudely reify existing inequalities, assumptions, and policies? As it turns out, new discoveries made by adding genetic data to social science are overturning many of our assumptions. For example, were Herrnstein and Murray correct in *The Bell Curve* when they argued that meritocracy has perversely resulted in more intransigent inequalities today because we are now sorted by genetic ability? Probably not. The data show that the magic of sexual reproduction and other genetic processes do as much (if not more) to upset existing inequalities (i.e., create social mobility) than they do to reinforce social reproduction. This molecular shake-up results from two main forces. First, while spouses are somewhat correlated in their genetic signatures, that correlation is weak enough to suggest that there is much dilution and resetting that occurs when someone at one end of the genetic distribution seeks a mate with which to reproduce: her offspring are likely going to revert to the mean. Second, to the extent that outcomes depend on particular

combinations of alleles (known as dominance or epistasis), mating disrupts extreme genotypes. We can describe this magic in terms that card players understand. In a common version of stud poker called Pass the Trash, players are dealt seven cards but then must pass three of them to their neighbor (and receive three from their other neighbor). You can be dealt a royal flush at first (i.e., be born with great genes), but once you have to join your cards (DNA) to someone else's, all your advantages can be reshuffled (i.e., your kids have no particular genetic advantage). This reshuffling can have big effects even if you have an advantaged neighbor (i.e., spouse) with whom to trade cards (DNA). You might lose your diamond royal flush when your queen of diamonds (which you were forced to pass to your neighbor) is replaced by his queen of hearts. In these two ways, then, the magic of sexual reproduction acts to reshuffle the deck of cards in each generation, preventing stable genocratic castes from emerging despite the well-observed affinity of like to marry like (we discuss trends in this kind of sorting along genetic and nongenetic lines in chapter 4).

Or take the case of the most sensitive issue of all when it comes to human genetics: race. What if we said that both of the authors, European mutts of mixed white ancestry, are genetically more similar to, say, a Mongolian, than the Luo tribe of Kenya is to the Kikuyu ethnic group of Kenya? You may not believe me (or your eyes), but it is true. Looks can deceive when it comes to racial classification. In fact, the entire community of non-African (and non–African American) human beings collectively can display the same level of genetic similarity as the population of a single region of sub-Saharan Africa (namely, the Rift Valley, where humans originated and which remains the deepest wellspring of human genetic diversity). That result is because the population that first left Africa to populate the rest of the world was at one point as small as 2,000 individuals, creating a bottleneck that filtered out much genetic diversity. It is not clear whether that 2,000 figure was the result of mass starvation from an original intrepid caravan of 10,000 or whether the 2,000 was the total number of adventurous eccentrics who thought it a good idea to cross the Sahara or the Red Sea in search of their fortune. Either way, the result of such a bottleneck is to reduce diversity. So, while it may be obvi-

ous why the new (mostly humorous) iPhone app "Wait! Don't Fuck Her, She's Your Cousin" has cropped up in an island community like Iceland, the real question is why we all have not downloaded it. Indeed, new research by us and others shows that the typical marriage in the United States is between people who are the genetic equivalent of second cousins (see chapter 4).

The upshot is that our very notion of race—often based on "natural" physiognomic differences such as eye shape, hair type, or skin tone—is, for lack of a better word, just plain *wrong* in genetic terms, as we discuss in chapter 5. Indeed, thanks to the window onto human history that genomic analysis provides, we are now able to resolve all sorts of mysteries of human prehistory, ranging from whether we did it with Neanderthals (yes, if you are of European or Asian descent) to how fertile Genghis Khan was (very) to when and how humans populated New Zealand (not all that long ago, it turns out). From a contemporary perspective, our newfound understanding of human migration and genetic segregation can explain some conundrums. This book will tiptoe through the minefield of race and genetics to confront unspoken beliefs head on. What does a reconstituted understanding of race look like in light of surprising genetic information? How should policy deal with this stubborn concept?

It is not just racial difference (and sameness) that is informed by genetic analysis. The genetic prism can help us understand the rise and fall of entire nations. In chapter 6, we step back and take a macro view of how genetic theories and findings are infiltrating a broader class of global questions; for example, a fundamental domain of inquiry in macroeconomics is why some countries have thrived while others have stagnated during the past several hundred years. There is a long-standing set of hypotheses suggesting that long-ago events and circumstances had lasting impacts that have shaped the tremendous differences in economic success and growth that exist around the world today. Everything, from the north-south versus east-west continental orientations to the shapes carved into the Earth during the last ice age, seems to matter for the wealth of nations. A newcomer to this discussion of "deep determinants" of economic success is population genetics. A new breed of macroeconomists has posited

that genetic diversity within countries is a key to development. In 2013, Quamrul Ashraf and Oded Galor published a paper claiming that a "Goldilocks" level of genetic diversity within countries might lead to higher incomes and better growth trajectories.[8] The authors discuss the observation that there have been many societies with low diversity (e.g., Native American civilizations) and populations with high diversity (e.g., many sub-Saharan African populations) that have experienced pallid economic growth, whereas many societies with intermediate diversity ("just right"—i.e., European and Asian populations) have been conducive to development in the precolonial as well as the modern eras. The researchers hypothesize that the Goldilocks advantage results from disadvantages at the extremes of genetic diversity—very low levels of diversity lead to a lack of innovation because everyone is the same, but populations with very high levels of diversity cooperate less because no one is the same.

In addition to calculating the Goldilocks level of genetic diversity with respect to economic development, economists have considered the role of population genetics as it interacts with environmental resources to affect growth patterns across countries. Justin Cook has shown that populations in early human history with the (genetic) ability to digest milk after weaning were conferred large advantages in population density around 1500 CE.[9] Because other studies have shown that historical differences in economic development have been remarkably persistent into the present day, the implication of Cook's study is that (relatively) small changes in the genome at the right time and in the right place (i.e., during the Neolithic Revolution in areas able to domesticate cattle) can lead to large, persistent, and accumulating differences in economic development across countries. But these genes confer advantages only when they occur in environments that have the ability to foster agriculture. With no cows, goats, or other domesticable mammals, the gene confers no population advantage.

In chapter 7, we further expand our discussion by more explicitly bringing the environment back into the conversation and discuss the many complications that emerge when one takes an integrative view of genes and environments. Indeed, genetics and environmental fac-

tors seem to interact with one another in a complex, dynamic feedback loop that further explains several aspects of the behavior of humans and entire societies. One strand of this research asks whether genetic factors may predispose people to be particularly sensitive to environmental variation. The idea is that some people are orchids and will thrive (or wilt) based on environmental enrichment (or environmental disadvantage), and other people are dandelions and are relatively immune to their surroundings, for better or worse. If we can tell the difference between orchids and dandelions at a young age, should we use this information in assigning them to their teachers, classrooms, after-school programs, and the like?

This question is evoked by the very mixed success rate that social policies have had up to the present time. Some interventions are successful for some people or during some periods of time, but not others. New evidence merging genetics and public policy has started to uncover *why* we see such different impacts of the same policy for different people and how future policies might be adjusted accordingly, thus extending the concept of "personalized medicine" to allow for "personalized policy." If for genetic reasons some people do not respond to health policies, such as taxes on sugary drinks or cigarettes, should we still make them pay the tax?

Evidence has also suggested that some educational interventions have greater or lesser effects depending on the targeted students' genotypes. Should some students be targeted for future interventions while the resources are diverted from other students? What if we find out that the Earned Income Tax Credit spurs low-wage workers to labor more or less depending on their genes? Whether we want to go down the path toward genetically personalized policy is one of the topics of our conclusion.

While the implications of the genetic revolution for social policy are being debated, a second implication of the social genomics revolution has been the democratization of genetic information. Nowadays, scientists can no longer conceal this information from the public or shape the information to fit their research agendas or political predilections. The cost of human genotyping is dropping faster than Moore's law predicts for the price of computer chips. As a result,

Americans are genotyping themselves in record numbers using consumer services like 23andme, Navigenics, and Knome. And those persons are acting on the information they receive, sometimes even when that information only weakly indicates a predisposition: they are asking their doctor if a procedure or test might be right for them. This is a new direction of consumer-driven medicine, in which patients no longer need to get their medical data from their doctors. A longer-term trend starting with home pregnancy tests and diabetic blood sugar tests has shifted abruptly to assessing immutable health traits—your genotype. Some individuals (such as Angelina Jolie) are seeking preventive bilateral mastectomies to reduce their risk of breast cancer. Couples learn about their status as carriers for a hereditary form of hearing loss and make family-planning decisions based on the health of their unconceived children. In our epilogue we outline one dystopian scenario of genotocracy that could emerge from this brave new world of self-directed eugenics—or what the biologist Lee Silver calls reprogenetics.[10]

In short, this book interrogates how new findings from genetics inform our understanding of social inequalities. We survey the literature (including much of our own research), synthesize findings, highlight problems, put forth new hypotheses, and stir in a lot of guesswork. We show how new genetic discoveries may be both disruptive as well as transformational. Indeed, these discoveries often overturn our preexisting notions of social processes. They also show how limited our current understanding of genetic effects is and that our attempts to cleanly separate genetic and environmental influences face huge challenges. But rather than wallowing in this perceived intractability, we pivot toward a richer integration of social science with genetics to enhance both fields of study. The potential implications are farreaching. Will we extend ongoing efforts in personalized medicine to create personalized policy? Will we use our new molecular understandings of human difference to fundamentally transform how we view and use race/ethnicity categories around the world? How will we insure that these new discoveries do not further widen social inequalities, as the rich take advantage of genetic information at a faster

pace than the poor? We discuss these and many other issues in the pages that follow.

While progress in this burgeoning field is happening at a breakneck pace (at least as compared to typical social science), we feel we must also note that it is still early days. We think the nature-nurture debate that dominated the twentieth century and a portion of the nineteenth is now over, as evidenced by the proliferation of social science surveys that now collect genetic data. That said, the tools for integrating these data are only just being developed and suffer from many limitations that may or may not be overcome in the next decade or two. Sample sizes are still too small. Biological mechanisms are hard to pin down. And social systems have a tricky way of learning and adapting, thereby undermining the stability of findings. Despite these challenges, we think this is an exciting new field worth sharing with you. We hope you agree. The genetics revolution may be well underway, but the social genomics revolution is just getting started.