Making History via DNA, Making DNA from History
Deconstructing the Race-Disease Connection in Admixture Mapping

Ramya Rajagopalan and Joan H. Fujimura

Admixture mapping is an approach in population genetics that has gained some recent notoriety because of the use of some of its tools in the growing industry of direct-to-consumer genealogy and ancestry testing, as well as in biomedical research. In particular, some social scientists have criticized the increasing focus on race in medical and genetics research, arguing that such projects essentialize race as biological, attributing health differences to innate differences among race groups. In this chapter, we explore the ways admixture mapping technologies (mis)use race and race groups and how they connect race groups to stories of continental origins. Focusing on the race-disease connections that are made through admixture mapping practices, this chapter examines and critiques the conceptual frameworks of admixture mapping (also known as mapping by linkage disequilibrium, or MALD), especially as used in biomedical studies of genetics and disease.

Our interest in admixture mapping stems from our ongoing project that examines the use of notions of population—including race and ancestry—in contemporary biomedical research. In the twenty-first century, the search for biological contributions to complex diseases has increasingly meant a turn within science to genetic readings of disease etiology. In the process, geneticists have cast bits of DNA in the genome as medically relevant, and these bits gain new weight and importance as they circulate inside and outside laboratories. Genomes, as biological information, are increasingly the focus of explanations for the disease conditions of bodies, and genomes as inherited substances are held accountable for the passing on of disease or (increasingly) disease risk across generations, from "ancestors" to "descendants."

Geneticists employ concepts of genetic history, ancestry, and theories of human genetic variation in various ways to search for disease susceptibility markers in human populations. In the last four years, for example, genome-wide association

"Making History via DNA, Making DNA from History: Deconstructing the Race-Disease Connection in Admixture Mapping."
In Genetics and the Unsettled Past: The Collision between DNA, Race, and History. (ed. K. Wailoo, C. Lee, A. Nelson)
Rutgers University Press, New Brunswick, NJ.
studies (GWAS) researchers have used the concept of “ancestry” to try to distinguish disease-related genetic differences from non-disease-related differences between groups, and they have developed methods to account for these different differences, using comparisons of genetic markers.

Although no method is devoid of cultural politics, some GWAS technologies designed to account for ancestry at least attempt to avoid the use of race categories (Fujimura et al. 2010; Fujimura and Rajagopalan, 2011). In contrast, admixture mapping tools and methods employ group categories generally considered by many in the United States to be socio-cultural race categories—categories whose biological and health relevance is and has been hotly contested. Our study identifies multiple problems and circular assumptions about the relationships linking race, population, and ancestry at each step in the theory and method of admixture mapping. What follows is an analysis and critique of how researchers in admixture mapping studies, attempting to identify disease-associated regions of the genome, deploy particular socio-cultural genealogical stories within the “ancestral origins” labels they attach to “chunks of DNA.”

We demonstrate how discourses of race and continental ancestry are deeply entangled in the construction of genetic technologies in this research.

**Genetics, Relatedness, and Race**

The concept of “relatedness” has been an important heuristic tool in geneticists’ search for disease genes, driven by the idea that DNA is inherited, passed on from one generation to the next. For over a hundred years, family linkage studies have followed groups of closely related individuals to track the inheritance of DNA that seems to be associated with particular phenotypic characteristics or disease symptoms. These studies have often been limited to diseases that were later shown to be strongly heritable, and thought to be caused by defects in one gene. Such single-gene hereditary diseases include, for example, Tay-Sachs disease, sickle-cell disease, and cystic fibrosis. The past decade has significantly changed this scene through the invention of methods and instruments of genetics that now allow researchers to begin to conduct studies of diseases on a genome-wide scale.

Coalescing around the major genomics projects at the turn of the century, these new technologies and genetic tools (such as faster and cheaper genotyping technologies, the Human Haplotype Map, and databases with human genome sequence and annotations) have facilitated the move from the study of single-gene diseases to the much more complicated analysis of diseases assumed to have multifactorial etiologies; that is, diseases that may be affected by many genetic factors as well as environmental factors. These genetic tools have been used in the search for genes related to common complex diseases like heart disease, cancer, and Type II diabetes. Significantly, common complex diseases tend to occur in all human groups, not just in particular families who share similar genetic profiles.

What distinguishes these new approaches in genetics is that they have not required that the sampled participants be selected on the basis of their familial

relations, and an important question is, why not? Part of the reason is because some of these studies make conclusions about similarities and differences in DNA based on assumptions of relatedness within and between groups, rather than families.

Admixture mapping is a research approach that in some sense occupies a middle ground between family linkage studies that use classical genetics methods and tools to look at a few loci in the genome, and resource-intensive approaches that look more comprehensively across the genome and across many unrelated individuals (Cooper et al. 2008). It is important to note that, while admixture mapping incorporates analysis of genetic markers spread throughout the genome, it examines far fewer points in each genome than, for example, GWAS studies. It looks at individuals who are not necessarily familial related, but who self-report as belonging to the same race category.

This is related to two other points critical to our analysis. First, admixture mapping methods assume a particular form of relatedness and unrelatenedness among and between the individuals they study. That is, they use socio-culturally defined categories of race to infer a higher level of relatedness among people in one race category and a lower level of relatedness between people in different categories. Unlike family linkage studies, admixture mapping researchers do not require that their samples come from related individuals. Nevertheless, they do still assume that distant relationships among people who identify in the same race category will make their search for medically relevant genetic loci more efficient. In this effort, they have delineated a set of assumptions and practices that weave notions of continental ancestry and race with socio-cultural ideas about geohistorical origins.

Second, admixture scientists inextricably link particular groups of purportedly related humans to particular diseases. This by itself does not set it apart from other kinds of genetic studies. But in admixture mapping the link between disease and group is a product of the scientists’ assumptions about relatedness within race groups. We explore and deconstruct these assumptions below.

**Deconstructing “Admixture” in Admixture Mapping**

In an admixture mapping study, particular groups become linked to particular complex diseases, both before the study begins, through the practices and processes of the study, and through the study outcomes. Researchers have designed admixture mapping studies around group-disease correlations, which have often been generated by epidemiological studies that suggest that particular groups have an elevated risk or incidence of a particular disease. For example, epidemiological data has been mobilized in admixture mapping to argue that African Americans have an elevated risk of prostate cancer over white Americans (Freedman et al. 2006), and Hispanic Americans have an elevated risk of asthma over non-Hispanic Americans (Burchard et al. 2003). Genetic researchers have argued that socioeconomic or environmental factors can explain some, but not
all, of the elevated risk of such diseases, and much of the work of contemporary medical genetics is devoted to finding genetic contributors to these and other common complex diseases.

Epidemiological data are often collected and analyzed according to U.S. census categories of race and ethnicity, for a variety of historical and institutional reasons. Admixture mapping researchers use these categories as they try to construct appropriate study groups. They base their study on two premises. The first is that if a potential study group—defined by race—appears to have an elevated incidence of a disease compared to another race group, then genetics may be part of the explanation. This is where the categorization of samples by race or ethnicity in epidemiological studies has implications for the study of genetics and disease. Second, they only study groups that they describe as genetically “admixed,” with a mix of DNA inherited from two “ancestral” groups. We will next interrogate the ideas of genetic admixture that admixture mapping researchers work with.

Admixture mapping geneticists have cast “admixture” and “admixed populations” in particular ways. (Here, it is necessary to emphasize that admixture as a term has a long history of its own, and this essay will focus only on uses of the concept in the context of American biomedical genetics.) Although encounters among many peoples throughout human history have resulted in mixtures of genetic material, researchers do not consider all groups as suitable subjects for the admixture mapping approach. They only work with study subjects that they feel comfortable describing as members of populations descended from encounters between two previously “geographically isolated” groups.

Further, admixture mapping geneticists distinguish “recent” admixture from other kinds of admixture, and study only “recently admixed” groups. This choice is based on the theory of DNA exchange occurring every generation between the two copies of each chromosome in a human genome, one copy of which is inherited from the maternal ancestors and the other of which is inherited from paternal ancestors. The theory holds that over a large number of generations, the DNA exchange process makes it increasingly difficult to trace any particular piece of DNA, let alone whole chromosomes, to either the maternal or paternal side. For the purposes of distinguishing the lineages that contribute to present-day genomes, admixture mapping geneticists have decided that they have to examine admixtures that occurred in the last twenty generations. They thus specify “recently admixed” groups as those whom they believe are descendants of two “ancestral” and “isolated” populations that “mixed” within the last twenty generations.

In short, scientists’ constructions of the populations they deem to be well-suited to these studies illustrate how their views of history, and the timescale of that history, enter into and shape scientific practice. In the latter part of the twentieth century, professional historians argued that histories may differ, depending on the available evidence and the storyteller. Biological scientists are no different from the rest of us in their reliance on these stories of the past. Thus, for example, admixture mapping scientists employ particular preestablished definitions of groups and categories to produce genetic understandings of these groups. They gain these ideas of groups and categories from stories in other disciplines. For example, they use sociocultural discourses on geography, history, and migration of human groups written by archaeology, physical anthropology, linguistic anthropology, and social anthropology. Genetics then is not a stand-alone discipline, basing its research only on its statistical techniques. Biological scientists’ knowledge and their research outcomes are also steeped in fields of research that study pieces of bones, pottery, and language phonemes. And just as there are debates in anthropology about how to make meaning of these materials, so too must we address these issues with respect to the field of population genetics.

How do these stories become embedded in the materials of population genetics? The idea motivating admixture mapping is that, if the admixture is “recent,” then some large chunks of the genome can be inherited virtually unchanged, because of the small number of generations since admixture. The contentious issue arises when scientists trace these chunks back through time to one of two “continental lines of descent.” Typically, they label these DNA chunks with identifiers that describe continental origins, such that some chunks become labeled, for example, “African,” and others “European.” In this way, stories of human migration patterns become etched into descriptions of DNA, even though these stories often simplify very complicated exchanges and “mixing” of DNA throughout history.

Equally thorny is the description of the “ancestral populations,” which are thought to have mixed, giving rise to the admixed population. Researchers define ancestral populations as having been largely geographically, and thus reproductively and genetically, isolated from each other prior to the historical events that facilitated large-scale “mixing” of these groups. For example, admixture mapping researchers believe that people who self-identify as African Americans possess both African and European ancestry, as a result of colonial encounters in which Europeans brought people of African ancestry to the Americas and subsequent mixing between these groups during the period of slavery in the United States. We note that this research does not consider any other kind of “mixing” that could have happened prior to this large-scale transfer of peoples, including mixing within Africa, where population geneticists hypothesize that “ancestral” peoples had much greater genetic variation compared to groups on any other continent.

Further, admixture mapping researchers assume two ancestral contributions, for the sake of simplicity. Thus, they consider African Americans to have combinations of segments of “African” inherited genes and “European” inherited genes. When studying groups that researchers assume to have multiple ancestries (for example Hispanic groups, whom they describe as descended from white Europeans, native/indigenous groups in America, and Africans), these scientists typically examine only two of these ancestries at a time.

Our respondents (including population geneticists and genetic epidemiologists) felt they could define admixture and admixed populations because they felt comfortable distinguishing ancestral populations from each other. Of course,
argued that the search for biological and/or genetic contributions to racially stratified health outcomes may distract clinicians and public health officials from more proximate social causes and more effective social interventions to redress these inequities (Cooper et al. 2003; Ossorio and Duster 2005). Other critics have warned that, while race may be expedient as a proxy for anything from diet to ancestry, contemporary biomedical uses are often incorrect, misleading, or able to cause more harm than good (Sankar and Cho 2002). Their primary concern was that the conflation of race categories (especially in the United States) with biomedical research categories has the potential to reinstitute and reify these race categories as biological (Duster 2003; Kahn 2006; Marks 1995; Wailoo and Pemberton 2006). Others have suggested that, while race as a sociopolitical construct is appropriate for monitoring health disparities, race and race categories are not appropriate analytical constructs for genetic studies of complex diseases (Feldman 2006; Fujimura and Rajagopalan 2011; Shields et al. 2005). For these reasons, the link that admixture mapping makes between social histories and the genetics of disease not only may potentially re-stabilize race categories, but may be misleading for understanding human disease.

As we will show next, these qualitative assumptions of admixture become embedded in and quantified in the tools and algorithms of admixture mapping that are used to link continental ancestry and disease during the analysis of genotypes in disease-affected patients.

How Race Is Read into DNA: The Making of Ancestry-Informative Markers (AIMs)

We have described how admixture mapping researchers study populations that they define as having two ancestral lineages. In order to differentiate two ancestral contributions to the DNA of study subjects, admixture mapping scientists have devised a method for generating genetic markers they label as “ancestry-informative.” The production of these “ancestry-informative markers” or AIMS, is a clear example of how admixture mapping begins with socio-cultural categories and groups and produces genetic understandings of these groups.

The AIMS are a subset of genomic markers known as SNPs, or single nucleotide polymorphisms. They are sites in the genome that vary between different people. Geneticists estimate that there are about 10 million SNPs in a human genome, and of these, researchers have deemed a few thousand informative for ancestry, which they call “AIMs.” Some population geneticists and some genetic epidemiologists consider subsets of these AIMS to be “informative” for different ancestries, for example, European ancestry, African ancestry, or Native American ancestry.

Researchers have designated AIMS as “special genetic markers” that are able to distinguish between groups, as one respondent described: “An ancestry informative marker has more information about ancestry on average than a random
marker on the genome. And you can say how much more information it has about ancestry. We can quantify that.

How do researchers make AIMS? Admixture mapping researchers begin with an “admixed population”—with all the qualifications we discussed above—in which to study disease, and postulating the two ancestral populations that gave rise to it. Then they construct a particular set of AIMS to use in studying that admixed population. They call this set an “admixture map” for that population. For example, to study disease in African Americans, a collaboration of scientists at over twenty institutions created a map they believed to be specific for African American groups. This map has since been used in many recent studies (Smith et al. 2004). An admixture map, then, is a set of AIMS that are specifically built and put together for the purposes of studying disease in a predefined admixed population. We argue that this is a circular process and demonstrate this circularity next by showing in detail how an admixture map is constructed.

As described above, AIMS are a subset of SNPs, so-called polymorphic sites in the DNA at which more than one of the DNA “bases” (abbreviated as “A,” “T,” “G,” and “C”) appears when human genomes are compared. For example, at a non-SNP position in the genome, all people will have the same letter, but at a SNP position, some people will have one base, others will have another base. These different base possibilities are known as “variants.” For the purposes of building an admixture map, researchers focus their efforts on determining the frequencies of SNP variants. For example, at a SNP position at which some people have “A” and others have “C,” the frequency of the “A” variant is the percentage of people in a group who have “A” at that position. Many SNPs in human genomes have been estimated to occur with similar variant frequencies in different groups. This means that, at many SNP markers, variant frequencies are believed to not differ across groups. In contrast, however, admixture mapmakers claim that AIMS have comparatively “large” variant frequency differences between groups. They argue that at an AIM, one of the two possible variants is much more common (or much less common) in certain populations compared to other populations. It is these sorts of statistical assessments that admixture mapping researchers use to label markers as diagnostic or “informative” of a particular continental ancestry at particular points in the genome.

Admixture mapping researchers believe that such “large” frequency differences can be used statistically to differentiate “ancestral populations.” But the meaning of “large” matters here, since admixture scientists distinguish AIMS from other SNPs based on their frequency differences. For example, in constructing the previously mentioned admixture map used to study African American groups, researchers were searching for AIMS to differentiate between groups they called West Africans and European Americans. They constructed AIMS with variants that, on average, differed in frequency by about 56 percent between their two groups of samples.

Thus, admixture mapping, the construction of groups and the construction of markers for distinguishing those groups are tightly connected. We have already discussed some concerns about the practices of delineating admixed and ancestral groups in admixture mapping. How do researchers’ assumptions about these groups get built into the technologies of the admixture map and reinforce ideas of difference between groups?

For one, the process of admixture mapmaking illuminates the contingent ways in which researchers conceive of “ancestral” groups, as well as “ancestral” and contemporary DNA. In practice, ancestral DNA samples do not exist or are inaccessible. So, admixture mapping researchers use contemporary samples to estimate the frequencies of SNP variants, and select those SNPs with “large” frequency differences as AIMS for their admixture maps. We note that SNP frequencies are always estimates, calculated either by genotyping purported representatives of a group, or computationally estimated. For example, in selecting AIMS for the previously mentioned admixture map, researchers estimated the frequencies of SNP markers in samples collected from contemporary individuals who identified as white. The researchers called these samples “European American” and used their estimated SNP frequencies as a stand-in for marker frequencies in “ancestral Europeans.”

Similarly, under the assumption that some ancestors of contemporary African Americans came from West Africa, they collected and genotyped DNA from individuals currently residing in West or sub-Saharan Africa. They then used the marker frequencies estimated from these samples to stand in for “ancestral” African frequencies. In a puzzling twist, for some SNPs they genotyped DNA from contemporary African Americans and then created statistical genetics formulas for computationally estimating “ancestral African” frequencies from these contemporary frequencies. Finally, by comparing the SNP frequencies they had constructed for “ancestral European” and “ancestral African” groups, they selected those markers with large frequency differences between the two “ancestral populations.” These markers are now known as AIMS, and they circulate as such among researchers as part of an “admixture map for African Americans.” At each SNP in the admixture map, the variant that appears to have a higher frequency in “ancestral West Africans” (by their methods) is thought to be indicative of “African” ancestry, and the other variant specifies “European” ancestry.

Thus, admixture mapping accomplishes both a geographical elision and a generational elision, assuming similarities and differences between groups separated both in time and space. The researchers assume that SNP frequencies in contemporary “European American” samples can be treated as equivalent to the frequencies in the “ancestral European” peoples that supposedly contributed DNA to contemporary African Americans. Indeed, labels for groups such as “European” and “European American” were sometimes treated as interchangeable by some respondents, even though both of these so-called groups can be considered to be very heterogeneous in themselves. In admixture map publications, researchers’ reasoning for their assumptions about groups is often left unstated. For example, why should those identifying as European American among the geneticists’ samples be thought of as able to directly represent “an” ancestral population from
Europe that came to America, as if such a static population existed or as if people living in Europe five hundred years ago were genetically isomorphic with people who identify as white in America today? Many peoples from many countries in present-day Europe colonized present-day America. This treatment of ancestral and contemporary “European groups” as indistinguishable, and essentially unchanged during or since encounters with other peoples, implies an assumption about the fixity of whiteness.

Similarly, researchers regarded contemporary West African samples as a good approximation of the gene pools of groups in Africa who were brought to the Americas and contributed DNA to present-day African Americans. On the other hand, frequencies calculated in contemporary African Americans were somehow seen as “admixed.” They required further transformation (by statistical genetic algorithms) before they could become a good approximation for frequencies in the “ancestral African” population that supposedly contributed DNA to today’s African American groups. Again, as with the genetic diversity of Europe described above, these kinds of manipulations collapse the genetic diversity of Africa. They also tell a particular story about how researchers view the relations between and among various groups separated in time and space. The geneticists conceptualized these African American samples as admixtures of both “African” and “European” DNA for the sake of the overall study of disease, but more African than European for the sake of making a map. Thus, they used these samples in a circular way to estimate the frequencies of DNA variants in their putative “African” ancestors.

The admixture mapping algorithms include statistical tests to assess how closely the European American and West African “populations” correspond to the “true” ancestral populations of African Americans (Smith et al. 2004). Nevertheless, these algorithms that are used to validate the mapmaking process have themselves been designed based on statistical formulas from population genetics that estimate measures of genetic drift.6 These measurements are dependent on probabilistic and statistical models that have been developed using only a small number of contemporary samples. Thus, the conceptions of ancestral populations in this research collapse many local histories. For example, the variation and diversity of peoples in West Africa five hundred years ago becomes circumscribed by a few markers sampled from a few individuals in the present, and assembled into a “set” of markers that are then mobilized to tell researchers something about what it means (in terms of disease) to have West African ancestry. The admixture map, which contains the AIMS markers that are central to admixture mapping, is thus built by an amalgam of circular logics and assumptions that reinforce a particular story (and thus legitimate each other) at every step. Genetic stories are interwoven with social histories, giving rise to a complex and contingent tapestry of mutual reinforcement, sometimes serving as legitimization for the admixture mapping approach, and sometimes as explanation for the observed.

Finally, we note that biomedical admixture maps include over a thousand AIMS, while commercial ancestry testing protocols use fewer than a hundred.

The AIMS in the biomedical admixture maps are similar to and may overlap with the first AIMS designed by Shriver et al. (1997; 2003). This first set of AIMS are being used in commercially marketed genetic ancestry tests sold by DNAPrint Genomics and other companies, to provide consumers with information about their percentages of ancestry from different continents. Some have pointed out that this use of AIMS to distinguish among continental populations over-simplifies and potentially geneticizes race, by inscribing sociopolitical conceptions of race onto test-takers’ DNA (Bohnick et al. 2007; Fullwiley 2008; TallBear 2008).

However, our respondents working in biomedical genomics, as opposed to ancestry testing, were careful in their publications to distinguish their marker sets from those of Shriver and others, indicating an unwillingness to use their markers and methods for identifying any specific individual’s continental ancestry.

The Practices of Admixture Mapping: How Specific Ancestries Get Linked to Specific Diseases

As discussed above, admixture mapping tools such as the admixture maps construct particular correspondences between groups and markers. Researchers believe a particular map or set of markers is useful only for the particular group it was designed for, such as African Americans, or Latin Americans. We will discuss how the marker maps are used to link groups (and particular “ancestries”) to diseases, referring to fieldwork done in a lab that narrowed down a large region of Chromosome 8 as being associated with prostate cancer in African Americans.

Once they have constructed an admixture map of AIMS for use with a particular population, researchers use the map to connect genetic markers to diseases. Researchers genotype DNA samples from disease-affected individuals, determining which variant each individual DNA sample has, at each of the SNP markers in the admixture map. They look for large chunks of the genomes of disease-affected individuals that appear to have an excess of the markers that they have assigned to one of the ancestries—for example, an excess of “African” SNP variants, or an excess of “European” SNP variants. Any regions that exhibit an excess of one of the two sets of “ancestral variants” is selected as a candidate for disease association. They then use statistical methods to assess candidate regions and identify those that appear to be most likely associated with the disease under study.

Put another way, in this step of admixture mapping researchers attempt (as they describe it) to infer continental ancestry at each point in the genome, for each of their samples. There is an analogous process in commercial genetic ancestry testing, whereby companies genotype the DNA of test-takers and generate a summary readout of the percentages of each of their ancestries across their genomes, calculated based on AIMS technology (Fullwiley 2008). However, in the case of admixture mapping in biomedical research, individuals are never told which chunks of their chromosomes are likely to be from which of their ancestries. But in positing that disease arises at least partly from a chunk of DNA inherited from one
ancestral group versus another, admixture mapping researchers explicitly connect not just the ‘race’ of the admixed group to the disease in question, but also indirectly designate a particular continental ancestry as being the “source” of the disease.

For example, our respondents reported that DNA inherited from African ancestors (or what they termed the “African chromosome”) was responsible for increasing the risk of prostate cancer in African Americans, compared to the risk in white Americans and other groups. However, one might surmise (according to the logic of admixture mapping) that if genetics plays a role, then West Africans should have a high incidence of prostate cancer as well, because of their continental ancestry. Yet, despite the finding of “risk alleles” for prostate cancer due to “West African” ancestry, it has never been established that peoples in West Africa have a high incidence of prostate cancer. It is likely that environmental factors play a considerable role, and it is possible that they interact with DNA (perhaps even DNA currently thought to be indicative of “non-African ancestry”), leading to variable outcomes in people who live in different parts of the world.

In this way, concepts of continental ancestry and related race categories are embedded within the continuum of practices constituting admixture mapping. They collide and tangle with notions of population (bounded by time and space) that are used to guide the map-making process, and they become central to the (re)constructed histories of various peoples and their diseases, which emerge from interpretations of disease-genetic marker associations. These concepts of continental ancestry are different from concepts of ancestry that are part of the scientific practices of, for example, genome-wide association studies (GWAS). Compared to GWAS, admixture mapping uses different theoretical frameworks and a different interpretation of ancestry, which can be described as “continental ancestry”—that is, they use a model in which ancestors of contemporary group study subjects are believed to have come from different continents. In other words, they try to identify genetic differences across continents. Thus in admixture mapping, the markers, chromosomes, risk alleles and “chunks of DNA” that are linked to disease are explicitly coded with continental and racial labels, and circulate in the literature and in further research with these labels firmly attached.

**The Triangulations of Geography, Ancestry, Genetic Histories, and Disease**

Within admixture studies, practitioners mix geographical descriptors with their concepts of race and continental ancestry to describe individuals and groups. Where groups (actually present today, or imagined as ancestors) are or were located comes to matter crucially for the links to fit together. But people move, and these histories are rarely straightforward or traceable. Still, genetic history is constructed alongside geographical history, and the two are woven to arrive at present-day “race” groups and their diseases.

This leads to the question, how do notions of continental ancestry differ from notions of race? We argue that the two are different, but often superimposed. Notions of ancestry and stories of geographical and continental origins are frequently interlaced, and even talked of interchangeably. But researchers also relate stories of continental origins to contemporary American race categories. We interviewed members of a leading lab in this area of study who have been using the admixture map described above to identify genetic variants for prostate cancer in African American men. One scientist respondent in our study described what they are doing as “tracking continental origin of lines of descent.” This population geneticist asserted that the ancestral populations of African Americans were “very different . . . at a population level, genetically.” The use of a whole set of tools and stories (as described in this paper) create the knowledge that leads to such assertions, and such assertions justify the tools and stories, and validate the knowledge. This includes theories about the relative incidence of disease, the contributions of genetics to disease risk, the relationship between continental ancestry and genetics, and the relationship between continental ancestry and race. In triangulating all of these, this population geneticist and colleagues in his discipline have constructed and cast admixture mapping as a reasonable approach to test their belief that genetics might partially explain why prostate cancer appears to be more prevalent in African Americans than in other groups.

In explicitly connecting continental ancestry to race, scientists encode geographical descriptors in the code of the genome. Alleles and even chromosomes are labeled with continental descriptors. For example, in publications scientists have referred to “the African chromosome” to indicate that a particular chunk of MALD markers in the genome appears to be more closely related to the “ancestral African” sequence of SNPs in that region than to the “ancestral European.” “African allele” is used to refer to a particular SNP variant that is estimated to be more frequent in African ancestors than European ancestors. Thus, “chunks of DNA” in the genome acquire geographical qualifiers, via a vernacular that pins down their origins to (and collapses the genetic histories and diversity of) entire continents.

As scientists deploy particular genetic, historical, and geographical discourses that cross and intersect, genomic stretches of SNP markers acquire genealogies that are rooted in place. However, in another move, scientists extend the explanatory power of the risk SNPs to people in other groups who have the disease. Even if African American or other supposedly “admixed” samples are used to identify genetic associations for the diseases under study, such associations are considered to be potentially translatable across populations. For example, SNPs in a genomic region of chromosome 8 that were believed to confer elevated risk for prostate cancer in African Americans were subsequently tested in other “ethnic” groups, such as Japanese Americans, Native Hawaiians, Latin Americans, and European Americans, to see if the findings could be replicated in other groups. All of these groups displayed evidence of statistically significant association of variants at this locus to prostate cancer risk (Freedman et al.
Complexity in Genomic Studies of Common Complex Diseases: Clinical Implications

The diseases under study in these projects are complex and thought to be the result of multiple genetic and environmental factors acting in combination. Reinforcing this idea, the first paper to report the 8q24 region linked to prostate cancer described above did not use admixture mapping; its authors found this region using family linkage analysis in several pedigrees of affected Icelandic males (Amundadottir et al. 2006). Others have identified the same region as associated with increased risk of prostate cancer, in various populations, using genomic-wide association studies.

Such studies have been done in African American men and non-Hispanic white or European men. Each of these studies has reported different SNPs, and even differing variants of SNPs, within the large 8q24 genomic region as being most significantly associated with prostate cancer risk. Therefore different SNPs, SNP variants, genes, or even alleles may be involved in different individuals. Other studies have found SNP alleles in this region to also be involved in colorectal cancer (Haiman et al. 2007b; Tomlinson et al. 2007; Zanke et al. 2007), and urinary bladder cancer (Kiemenei et al. 2008), suggesting that some of these risk variants may be pleiotropic, that is, they have multiple effects on the etiologies of several kinds of cancers, or perhaps even other diseases.

With the level of complexity involved in prostate cancer and other common complex diseases, what are the implications for association studies like admixture mapping to identify causative genetic elements? Because researchers use the technique to focus their attention on very large regions of the genome, and not on genes, or even specific genetic markers, further experimental analysis is necessary to determine if and how 8q24 is involved in prostate cancer, either directly or indirectly.

Given the variety of reports of genomic regions, disease associations and populations studied, it remains unclear how this information can and will be used in clinical settings to shed light on individual or group risk. As the American Society of Human Genetics has stated, “Numerous studies using MALD are underway, but even at its best, MALD is likely to be an effective strategy for only a small fraction of health-related traits, since genetic differences may not be the major cause of observed population differences in disease incidence. These limitations justify caution in the interpretation of data from these studies and in the clinical application of results from the related DTC genetic tests” (ASHG 2008).

Indeed, one respondent admitted that in terms of ancestry and conceiving of admixed populations, “what they are generating is a model not the truth” and that some details may be incorrect. In addition certain outliers may not fit the model. One respondent stressed that in practice “it’s very important clinically that individuals should be treated as individuals.” The genetic findings of admixture studies, however, are based on risk probabilities for groups, not individuals. Indeed, the scientists we studied were quick to point out that they cannot make any claims about individual cases, either in terms of their geographic or ethnic ancestry, or in terms of their disease risk status.

There are many examples of diseases for which genetic explanations emerging from laboratory research became associated with more frequent occurrence in certain socio-historical groups. For example, several diseases are now thought to be most prevalent in Jewish, African, or Caucasian genetic backgrounds, such as Tay-Sachs, sickle-cell anemia, and cystic fibrosis, respectively. Thus some diseases and associated genes come to be described with ethnic or racial qualifiers. Indeed, as part of the standard of care at many clinics, healthcare providers test for each of these diseases specifically in individuals who self-identify with the relevant race or ethnic category. In similar fashion, when regions of the genome that appear to derive from one of two presumed “ancestral populations” harbor SNP variants that become associated with disease in a particular population through admixture mapping studies, they acquire names such as “African risk allele” in the biomedical literature. Such interpretations reinforce the imputed group specificity that motivates these studies in the first place, even though, as described above, the risk alleles are often found in other groups as well.

This is not simply a labeling problem, since the practices used to arrive at such labels shape and reinforce how biomedical scientists think about the relationships between genes, diseases, and certain populations. These practices construct diseases and population groups as interconnected in particular ways. For example, one such assumption about self-identified African Americans is that they are so recently admixed that their ancestry is predominantly African, and only in small fraction European. However, many people who self-identify as African American may have had a “more European” ancestor as recently as one generation back, for example, Barack Obama. Another assumption is that those who identify with a particular race or ethnic group are more genetically related to each other than to individuals who do not identify with that group. Relatedly, another assumption is that genetics may be involved in diseases that are reported to occur more often in individuals who identify with the same socio-cultural race or ethnic group. Together, these assumptions stack and imbibe in ways that connect estimates of disease incidence to estimates of African ancestry in an appeal to “common sense” interpretations of which groups are “more” or “less” African, or more or less prone to disease.

In a more applied example, companies offering direct-to-consumer genetic risk tests are channelling the results of admixture mapping studies to individual diagnosis, regardless of scientists’ claims that these findings are too premature. Within the field of admixture mapping, a company called Proactive Genomics is
developing a test to predict risk for prostate cancer, which is expected to cost about $300 (Kolata 2008). After genotyping DNA from blood or saliva, they will analyze the SNP variants at five different regions of the genome thought to affect susceptibility, each conferring an independent (and therefore additive) risk (Haiman et al. 2007a; Zheng et al. 2008). The more susceptibility variants a person has, the higher that person’s estimated disease risk. The availability of Proactive Genomics’ test will, according to some clinicians, lead to earlier screening of higher risk groups, including African American males (Kolata 2008).

What is clinically important here is that most prostate cancer cases advance so slowly that they are generally benign if left alone, but unnecessary treatment can result in unwanted side effects like impotence or incontinence. Because doctors cannot predict in advance which tumors will become aggressive, increased screening is likely to result in more decisions to prematurely and unnecessarily treat patients, hastening side effects and worsening the overall condition of patients. Many prostate cancer specialists worry about these implications. Like the Prostate-Specific Antigen screening test, which has never been shown to actually reduce the risk of prostate cancer mortality and for which false positives are common, it is unclear if the risk estimates generated by Proactive Genomics’ test will actually lead to better health outcomes for individuals or unnecessary invasive surgery and treatment.

Finally, AIMs have also been tailored to the goal of trying to determine individuals’ ancestry proportions. At least one company, DNAPrint Genomics, uses AIMs, for example, in consumer tests that purport to be able to distinguish a person’s various continental “biogeographical” ancestries, which they claim is the heritable portion of “race” (Bliss 2008; Bolnick et al. 2007; Fullwiley 2008; TallBear 2008). The cautions and limitations of ancestry testing are summarized in Bolnick et al. (2008).

It is not possible for scientists or regulatory bodies to fully control the commercial applications of their findings, regardless of their intended usage. Though our respondents helped to find an association to prostate cancer at one region of the genome, much more work will be required to determine whether, if any, of the DNA within that stretch is causative or even indirectly correlated to prostate cancer. Even if genes are found to be statistically associated with prostate cancer, researchers would need to do more analysis to determine their function. The interpretation and application of admixture mapping findings, once disseminated into public domains (including to other practitioners in the sciences, who may or may not be aware of these caveats), becomes difficult to manage. With this degree of uncertainty, it is difficult to imagine what information the commercial test for prostate cancer can actually offer patients, or how they might act on it. Many of our respondents worry about the ramifications of untrained commercial ventures using genetic findings to make broad claims about an individual’s risk for certain diseases, particularly because the FDA has not implemented any regulatory oversight or approval process for “homebrew”

tests offered directly from a company, rather than through licensed clinical or medical practitioners.


Given that other methods exist, such as GWAS, which can potentially avoid the imprecision of race categories in genomics research (Fujimura and Rajagopalan, 2011), why do some researchers choose admixture mapping over GWAS? One rationale our respondents have provided for this choice is that admixture mapping is much smaller in scale and thus cheaper in terms of the technologies. Population geneticists have designed the science of admixture mapping to be more affordable, while still generating statistically significant results. Further, admixture mapping studies are less time-consuming and technology-intensive to conduct than GWAS studies. Admixture mapping studies use fewer DNA samples from disease-affected individuals than GWAS, and only genotype about three thousand AIMs in each DNA sample, as opposed to the hundreds of thousands of markers genotyped per DNA sample in GWAS. Thus some researchers believe that certain diseases may be studied more inexpensively using an admixture mapping approach. However, as one respondent noted, GWAS can be used to generate the same findings as admixture mapping without being restricted to admixed populations, and the admixture mapping approach is thus falling out of favor at some labs, even as it remains in use at others.

Understanding how admixture mapping works, and how it makes connections between populations and diseases, is critical for understanding the limitations of commercial ancestry testing and disease risk testing, whose methods are based in genomic assumptions like those we have described. Many companies have begun to offer both types of tests directly to consumers, for a fee, including DNAPrint for the former and 23AndMe, Navigenics, and deCODEme for the latter. These tests have received much attention from consumers and the media. They are not simply recreational or for entertainment—many consumers place some degree of trust in the results they receive (Bolnick et al. 2007; Nelson 2008). Significantly, the databases and methods that these tests use (derived from the findings of genomics research like that which we describe) are specifically not designed for individual ancestry or disease risk assessments, according to our respondents. As genomics becomes increasingly integrated into health care in the United States, perhaps even under the Obama administration, recognizing the limitations of such approaches as admixture mapping will be important for informing policy around the potential clinical applications of such methods and their findings.

The choice of groups that scientists study as admixed in this vein of genomic research are particularly indicative of the tenor of debates around race-based medicine in the United States. As Steve Epstein has noted (2008), there has been
a regulatory push at the level of the federal funding agencies for the inclusion of minorities in American biomedical research. The rise of admixture mapping research speaks to both this regulatory context and to the desire on the part of some scientists to include their own minority communities in their research programs (Fullwiley 2008).

Indeed, thus far in our reading, individuals who identify as European American or non-Hispanic white have not been the focus of an admixture mapping search for associations between disease and genetics. Why not? Are they considered too admixed, and therefore not amenable because the ancestral populations are too difficult to define? Here, the operative history may be one of the multiple waves of migration from various parts of Europe, a Europe that is differentiated in ways that “west Africa” is not, into national, geographic, linguistic, cultural, and possibly reproductively distinct groups. Alternatively, such groups may be viewed by researchers as not admixed, or not admixed recently enough. The idea that European American groups are not admixed in the way that other groups who encountered Europeans are suggests an “unmarked category” status for European American categories within this science.

Other groups, such as those identifying as African American, become, in a sense, Europeanized by this work, in that European ancestry is imputed regardless of individuals’ own genealogical self-understandings. These assumptions gain the authority of science, and stories of geography, history, and genetics stabilize each other in mutually reinforcing ways. Just as socio-cultural histories and human geographies are used to legitimate the inquiry into genetics, the genetic findings and the presumed stability of DNA reaffirm and stabilize historical stories.

Conclusion

In this essay we have begun to untangle the relationships between continental ancestry, geography, history, population, and race in the practices, technologies, research designs, and research analyses/results of some admixture mapping studies in U.S. biomedical research. In admixture mapping, researchers differentiate DNA into chunks by deploying geographical, continental and ancestral labels that buttress certain stories and make them robust. During the research processes, social histories of population origins and migrations thus become embedded in the research philosophy. Geography plays a major role, as these researchers trace groups of peoples across the globe to separate “ancestral” homes delineated by continent, thus shaping concepts of “continental ancestry.” Discourses of race and ancestry are deeply entangled in the constitution of genetic histories in contemporary biomedicine, in ways that are contingent and mutually reinforcing. These discourses, as embedded in the technologies of admixture mapping, have consequences for how disease studies, medical practices, public health policies, and popular culture use and interpret genetics to construct categories of difference.

NOTES

1. The quotations indicate that these terms are used by the population geneticists who use admixture mapping.

2. A phenotype is an observable characteristic, attribute, or trait of an organism.

3. Although some of our respondents argue that methods used in GWAS to account for ancestry will supersede/displace/supplant admixture mapping technologies, the latter are still very much in use today.

4. We note here that genetic studies are different from clinical practices, in which self-reported race is a very common heuristic used by clinicians in diagnosis and treatment decisions. Some argue that race is a useful proxy for other kinds of information when dealing with patients, especially in budget-restricted or time-sensitive medical situations, where obtaining genetic or other information is too expensive or could delay treatment and compromise the patient’s health. Still, there are debates about the use of race in clinical practice.

5. American research groups have also used this same approach to study disease in other geographical contexts, including Argentina and Mexico. However, our discussion is limited to the American biomedical context and study subjects in the United States.

6. Population geneticists have formulated the theory of genetic drift to describe processes that may cause two groups of people to become more genetically different from each other with each generation. These can include: geographic isolation from each other, different selection pressures, mutation, and bottlenecks (when only a small subset of the group reproduces and contributes genetic material to the next generation).

7. A population geneticist used this nontechnical term with us to help make the point clear. These “chunks of DNA” are more technically described as portions of the genome with significant linkage disequilibrium, or regions that are believed to be relatively unchanged over recent generations. These portions of the genome are assumed to have “travelled” intact through generations with minimal recombination.

8. This is the case at one of our field sites.

WORKS CITED


