Finding the global in the local: constructing population in the search for disease genes

Steve Sturdy

Introduction

Since the birth of genetics in the early twentieth century geneticists have been keen to understand what their science can say about how humans differ from one another, including how they differ from one part of the globe to another. Since the early 1990s there has been a notable surge in initiatives aiming to describe and analyse the genetic make-up of people living outside Europe and North America. Some of these initiatives profess a primarily anthropological orientation, aiming to capture and document as much as possible of the sheer diversity of human genetic constitutions: the Human Genome Diversity Project and National Geographic's Genographic Project are cases in point. Other initiatives – often framed in terms of documenting genetic 'variation' rather than 'diversity' – claim a more practical orientation towards identifying gene variants of medical significance: the International Haplotype Mapping (HapMap) Project and the African Genome Variation Project exemplify this latter approach to human genetics. In explaining the medical value of such work, advocates commonly state that it will generate genetic information that will help to combat illness and foster health in the countries under study. Thus the International HapMap Project is framed in terms of the 'need to be inclusive in the populations that we study to maximize the chance that all people will eventually benefit from this international research effort' (NIH News Advisory, 2002); while the African Genome Variation Project is

presented as a step towards 'provid[ing] a comprehensive resource for medical genomic studies in Africa' (Gurdasani et al., 2015: 331).

In this chapter I look beyond such professions of intent to examine a broader set of drivers that have structured the global expansion of genomic research and that skew how its findings are used. I show how recent initiatives to document the genetic constitution of people living in the world's poorer regions have actually grown out of, and serve the purposes of, efforts to identify genetic factors which influence the health of people living predominantly in the global North. This is perhaps unsurprising. For one thing, most research into the genetics of disease and ill health has been conducted by researchers based in the wealthiest countries of Europe and North America. For another, such research is to a significant extent driven by commercial interests, in the hope of profiting financially from any resulting medical innovations; consequently, research has tended to focus on understanding the ailments affecting those sections of the world's population that are most likely to be able to afford such innovations. In which case, why the recent turn to study the genomes of people living outside these rich regions? This chapter sets out to answer this question through a historical study of the changing aims and methodologies of medical genetics and human population genetics, and their recent convergence around new approaches to identifying genetic causes of ill health and disease.

Based on a close reading of medical genetics research literature, I show how, between the 1980s and the early 2000s, the focus of that research shifted from rare single-gene disorders to the genetics of common complex disorders such as heart disease and diabetes. I show how that shift in focus involved a shift in methodology: from studying the hereditary transmission of disease within families, to searching for correlations between genes and disease in populations. In particular, I show how a new approach to characterizing and analysing genetic populations came to be articulated in the course of this methodological turn. This new approach arose from the hybridization of two different and longer-established ways of thinking about populations. On the one hand, it drew on mainstream epidemiological ideas of populations as artificial constructs, created and defined for the purposes of research itself. And on the other hand, it drew on ideas inherited from older approaches to human population genetics which supposed that distinct genetic populations, far from being constructed by researchers, actually

exist in nature and are simply characterized and differentiated through the use of appropriate research methods. I argue that the adoption of this novel approach to populations as a basis for identifying genetic causes of common disorders was one of the principal drivers, in the 1990s, for the surge of interest in the genetics of people living outside the global North. Specifically, as we shall see, knowledge of particular genotypes that occur with different frequency in different parts of the world was deemed necessary so as to control for the confounding effects of so-called 'population structure' in the search for correlations between diseases and genes. In practice, this knowledge has overwhelmingly been used to facilitate research into disease–gene correlations among white Europeans and North Americans. As a result, the net effect of the globalization of genetic research has not been to reduce global inequalities in research into disease, its causes and treatment, but, if anything, to exacerbate them.

Populations and disease genes in the 1950s to 1970s

During the immediate post-war years research into human population genetics and research into genetic factors in human disease tended to diverge in their aims and methods. Consider first the research aimed at identifying genetic causes of disease. Such work focused primarily on single-gene disorders. These are mostly rare conditions, such as Huntington's disease or phenylketonuria, in which individuals possessing either one or more usually two variant copies (depending on the condition) of a particular gene will usually develop, sooner or later, the phenotypic signs of that condition. In such disorders researchers can trace transmission of the relevant genetic variants through successive generations of a family using the theoretical schema of Mendelian genetics. Family pedigrees accordingly served as the principal objects of research for medical geneticists interested in identifying and tracking such disorders, and remained so during the 1960s and 1970s, with researchers often devoting considerable effort to seeking out and cultivating affected families (see, inter alia, Comfort, 2012: 97–129; Gaudillière, 2000; Lindee, 2005; Nukaga, 2002).

As well as following the inheritance of genetic disorders from one generation to the next, medical geneticists were also keen to map the

location of the causal gene variants onto the human chromosomes. Gene-mapping techniques had been developed from the 1920s by geneticists working in the laboratory with experimental animals and plants. These techniques involved observing a phenomenon known as linkage. According to classical Mendelian genetics, in the course of sexual reproduction the alleles of any one gene will normally be distributed among members of the next generation quite independently of the alleles of any other gene. However, if two genes happen to be situated close together on the same chromosome their alleles will tend to be inherited together in a non-random fashion; in such cases, the genes are said to display linkage. The greater the tendency for alleles to be inherited together, the closer the linkage, and the closer together the relevant genes are assumed to be situated on the chromosome. By conducting controlled mutation and breeding experiments and observing multiple Mendelian traits, by the early post-war years geneticists had been able to construct genetic linkage maps of a range of organisms from the fruit fly *Drosophila* to maize and laboratory mice (Kohler, 1994; Rheinberger and Gaudillière, 2004).

Linkage mapping in experimental organisms played an important role in the development of genetics as a scientific discipline during the first half of the twentieth century. For medical geneticists it also offered a possible route to more practical benefits. If a genetic disorder could be shown to be linked to a more easily observable trait, the presence of that trait might serve as a diagnostic marker, enabling clinical geneticists to identify the disease before the onset of symptoms or to identify carriers who would not themselves develop the disease but could pass on the offending genetic variant to their offspring. Humans did not lend themselves well to available methods of genetic linkage mapping, however. For one thing, humans are not generally regarded as appropriate subjects for systematic breeding experiments. For another, humans are a young species in evolutionary terms and so are less genetically diverse than many other organisms. Consequently, they exhibit relatively few of the kinds of Mendelian traits that can be followed from one generation to the next. Given the paucity of such traits, such genetic linkage maps as were constructed up to the late 1970s contained relatively few genes, including no more than a handful of genetic disorders,

while the degrees of linkage remained too tenuous to be of diagnostic or predictive value (Harper, 2008: 194–212).

While medical geneticists concentrated on families in their search for disease genes, human population geneticists pursued a rather different set of concerns during the post-war decades. Their principal aim was to elucidate the history and dynamics of human evolution – an enterprise which they conducted in sometimes fractious dialogue with physical anthropologists (Smocovitis, 2012). Beginning in the nineteenth century, studies of human evolution initially focused on efforts to characterize distinct races of humanity and to chart the hierarchical relationships between them. The anthropological and anthropometric techniques employed for this purpose were augmented during the first half of the twentieth century by new genetic methods of defining and differentiating populations, in particular by observing differences in the frequency of certain common genotypes. In particular, the realization that the ABO system of blood groups followed a simple pattern of genetic determination paved the way to large-scale population surveys of the distribution of different genotypes, made possible by the ease of collecting, storing, transporting and serotyping blood samples (Gannet and Griesemer, 2004; Marks, 2012). The post-war years saw further expansion in both the scale and geographical scope of such surveys, as well as the identification of additional serological markers of genetic difference (Bangham, 2014; Radin, 2014).

At the same time, dominant views about the nature of human evolutionary relationships were shifting as anthropologists and geneticists alike sought to cast off earlier, overtly racist assumptions about human difference. By the 1960s concerns to identify distinct racial types had largely given way to a more dynamic and relativistic understanding of human population. Earlier race theorists had posited that races constituted discrete geographically or reproductively separate populations. By contrast, post-war population geneticists were increasingly inclined to envisage a single, continuous global population connected by constant migration and interbreeding and structured not by discrete boundaries but by continuous variation in gene frequency from one part of the world to another. Consequently, they argued, biology could provide no basis for essentialist ideas of race. However, disagreements remained over whether this shift in thinking meant that it no longer made sense to talk about biological races at all. Some post-war human

geneticists argued that it was still possible to discern distinct human populations separated by cultural as well as geographical barriers. Even if the boundaries between those populations were blurred and porous, they still placed enough restrictions in the way of genetic exchange, and the populations they demarcated were still sufficiently different in their genetic make-up, to constitute distinct biological races. Others adopted a more pragmatic viewpoint. Discrete populations might not exist in reality, they argued. But for purposes of genetic research into human difference it was useful to behave as if they did, using geographical or other criteria to demarcate populations to sample and study. As a number of historians have observed, such research commonly drew on ethnic and other markers of population difference. In so doing it tended to reintroduce, albeit implicitly, older assumptions about race into the identification of populations for purposes of human genetic research. At the same time, that research itself tended to naturalize and reify those assumptions by showing that, in many cases, genetic differences could indeed be found between the populations so defined. While the precise relationship of these genetic differences to other forms of racial, ethnic or geographical difference continued to be debated, the circular reasoning underlying much of that research went largely unremarked, thereby helping to sediment assumptions that racial and ethnic differences are at least partly rooted in biology (see, inter alia, Bangham, 2015; Gannet, 2001, 2003; Gannet and Griesemer 2004; Gormley, 2009; Lipphardt, 2014; Marks, 2012; Reardon, 2004, 2005; Smocovitis, 2012).

On the whole, this line of research into population genetics offered little of practical interest to medical geneticists, at least at that time. Medical geneticists did sometimes take an ambivalent interest in the kinds of large-scale population surveys of variation in gene frequency that flourished in the post-war decades (de Chadarevian, 2014) – chiefly in the hope that they would provide insight into hereditary influences on susceptibility to infectious diseases and common disorders such as heart disease, as well as rare single-gene disorders. However, while such research certainly provided new epidemiological information about variability in disease incidence, it was less informative about supposed genetic determinants of that variation. Even where such research helped to throw new epidemiological light on the incidence of rare genetic disorders those findings were rarely of much practical benefit

to medical geneticists. The observation that gene variants associated with sickle cell disease appeared to confer a selective advantage on people living in malarial regions of the world, for instance, provided population geneticists with a neat example of how evolutionary theory could be applied to humans. For medical geneticists, meanwhile, the same findings helped to direct diagnostic attention in ways that often reflected racial assumptions about predisposition to disease (Wailoo and Pemberton, 2008). More generally, evidence of population-based variation in disease incidence proved largely unpersuasive as an argument for genetic causation: unless researchers could demonstrate clear Mendelian patterns of inheritance, they found it difficult to argue that increased incidence was due to genetic rather than environmental causes, as was evident for instance in the case of familial cancers (Cantor, 2006; Necochea, 2007). Consequently, despite the growth in research into population genetics during the 1960s and 1970s, medical geneticists continued to focus their attention on families rather than populations.

However, there was one important exception to this generalization. One key site where the interests of population geneticists and medical geneticists came into productive dialogue with one another was in relation to so-called 'population isolates'. Population geneticists had coined this term to denote specific populations which they considered to have been reproductively isolated, be it for geographical or cultural reasons, from the larger human gene pool. Groups identified as population isolates were typically quite small in size, although they included some larger groups such as the Finns and the Basques. They were also considered to be atypical in a world elsewhere dominated by movement and interbreeding. And that atypicality made them uniquely valuable. Human geneticists regarded population isolates as relics from older periods of population history, biologically untouched by more recent flows of people and genes that had shaped present-day populations. As such, they were seen to offer unique insight into human evolution. But they were also seen to be at risk. With the growing speed and reach of modern population movements, small population isolates were increasingly vulnerable to out-breeding, dilution and, ultimately, dissolution. Consequently, population geneticists were anxious to sample and characterize such isolates before they were lost to the homogenizing flood

of modernity, and the 1960s and 1970s saw a number of gene-hunting expeditions launched, particularly to sample such supposedly 'primitive' peoples as the Yanomami whose isolation was assumed to stretch back furthest in evolutionary time (Lindee, 2003; Lipphardt, 2012, 2014).

Medical geneticists, too, were keen to study population isolates – although for rather different reasons than population geneticists. For medical geneticists, population isolates represented unusually fertile ground on which to hunt for rare genetic diseases. Often descended from small groups of founders, with high rates of in-breeding and consanguinity, population isolates were typically judged to possess relatively low genetic diversity. Consequently, while such populations would likely harbour fewer genetic disorders, any disorders that did occur would do so with far higher frequency than in a larger, more heterogeneous population, recurring repeatedly in the pedigrees of densely interrelated families. Medical geneticists were therefore keen to work with population isolates as a means of identifying disorders which they would find much harder spot elsewhere. Unlike population geneticists, however, they tended to be less concerned with evolutionary 'primitiveness' than by accessibility and the availability of well-validated family histories. Consequently, medical geneticists were as likely to conduct their research among culturally defined 'isolates' living in Europe and North America, such as Ashkenazi Jews and the Pennsylvania Amish, as among more geographically remote populations (Lindee, 2005: 58–89, 156–186; Wailoo and Pemberton, 2008; Widmer, 2014).

This shared interest in population isolates was one of the few places where ideas from population genetics intersected directly with the medical geneticists' efforts to identify disease genes. Overall, work in human population genetics and research into genetic causes of human disease were notable as much for the extent to which they tended to diverge, both theoretically and methodologically, as for their occasional overlaps. Where population geneticists focused on statistical observations of the frequency of common variants within freely interbreeding populations, medical geneticists were concerned primarily with tracing individual occurrences of rare genetic variants within families. This divergence would become more marked during the 1980s as medical geneticists adopted new techniques from molecular genetics that

enabled them to focus ever more clearly on families rather than other kinds of human grouping.

Family studies and molecular mapping

As we have seen, the possibility of conducting genetic linkage studies in humans was severely constrained, up to the 1970s, by the paucity of common, phenotypically observable genetic variants. However, the development of new molecular biotechnologies during the 1970s made it possible to circumvent some of those constraints by identifying a growing number of variants which do not usually give rise to observable phenotypic differences but which can be identified in the laboratory using molecular biological techniques. These included so-called restriction fragment length polymorphisms (RFLPs) – specific combinations of nucleic acids that can occur at various points in the genome, which began to be identified in humans in significant numbers from the late 1970s; and later, from the early 1990s, microsatellite variants – repetitive sequences of nucleotides scattered through the genome, which vary in the number of repeats. Both RFLPs and microsatellite variants proved to be much commoner in humans than the kinds of phenotypically expressed gene variants that had previously been used for linkage studies. As such, they provided medical geneticists with a new and powerful set of techniques for studying linkage in humans, and ultimately for mapping conditions such as rare genetic disorders that could be shown to be linked to these genomic markers.

Medical geneticists first adopted the idea of using RFLP markers to study linkage to human disease genes in the late 1970s (Kan and Dozy, 1978; Solomon and Bodmer, 1979). The first steps in this direction were initiated by the Hereditary Disease Foundation (HDF), set up in 1968 by Dr Milton Wexler following his wife's diagnosis with Huntington's disease, with the aim of promoting research into Huntington's and other genetic conditions. In 1979 the HDF funded a group of geneticists to try to map the Huntington's disease gene using the new RFLP markers. The researcher worked first with an 'American family of reasonable size' identified through the National Research Roster for Huntington's Disease Patients and Families at Indiana University, and subsequently with what the researchers described as 'a unique community of interrelated Huntingdon's disease gene carriers living along

the shores of Lake Maracaibo, Venezuela'. The researchers struck lucky. Employing a panel of only a dozen RFLP markers, in 1983 they were able to report the mapping of Huntington's disease to chromosome 4, and the identification of an RFLP marker of potential value in identifying individuals who carried the fatal genetic variant (Gusella et al., 1983; Nukaga, 2002: 53–57).

Meanwhile, in 1980 a group of American molecular biologists had proposed that a concerted programme of identifying RFLPs should be undertaken with the aim of constructing a linkage map of the whole human genome, in order to facilitate more systematic mapping of genes associated with hereditary diseases (Botstein et al., 1980). In 1984, as the work on Huntington's disease made clear the potential of such an approach, a group of researchers from the US and France agreed to set up an international collaboration, sharing biological materials and analytical skills with the aim of creating the first complete human linkage map. The Centre d'Etude du Polymorphisme Humain (CEPH), as this collaboration was called, drew on three sets of three-generation families. In France, over 100 families had been recruited during the 1970s to provide a tissue-type reference group for the French tissue transplantation services; ten of these families were subsequently brought into the CEPH. In the US, researchers at the Howard Hughes Medical Institute in Salt Lake City, including two of the authors of the original 1980 paper proposing the construction of an RFLP linkage map, recruited a group of Mormon families. Two large families from the ongoing Huntington's disease study in Venezuela were also added to the CEPH's reference panel. Cell lines and pedigree data from these families were distributed to an increasingly large network of researchers and the results of linkage experiments were shared among the collaborators (Dausset et al., 1990; Rabinow, 1999). Using these resources, the first genetic linkage map of the whole human genome, involving over 400 genomic markers, was published in 1987 (Donis-Keller et al., 1987).

By that time, RFLP linkage data from the CEPH were already being successfully used to demonstrate linkage in a variety of hereditary diseases. In 1985 a collaborative study of forty-three Canadian families with children affected by cystic fibrosis made use of the accumulating body of RFLP markers to report a linked marker on chromosome 7 (Knowlton et al., 1985; Tsui et al., 1985). Other conditions quickly followed, including chronic granulomatous disease, Duchenne muscular

dystrophy and a rare genetic cancer called retinoblastoma. Clinical geneticists added the linked markers to the range of tools at their disposal for diagnosing and predicting these diseases. Detailed study of the inheritance of single-gene disorders through large pedigrees such as the Venezuelan Huntington's disease cohort in turn yielded additional markers that could be used to map other genes (Jones and Tansey, 2015, 33). By 1986 advocates of RFLP linkage mapping felt sufficiently confident to suggest that 'RFLPs can be found linked to any common human disease that shows simple Mendelian transmission and is caused by a single genetic locus' (Lander and Botstein, 1986: 49). Their confidence was further boosted during the early 1990s as researchers added microsatellite variants to their panels of RFLP markers, enabling them to produce even denser linkage maps of the human genome, while also streamlining the process of genotyping samples in the laboratory (Kaufmann, 2004; Kruglyak, 2008: 314; Weissenbach et al., 1992). By that time, molecular geneticists were not simply mapping the markers linked with hereditary diseases but also isolating, cloning and sequencing the genes themselves. Building on medical geneticists' well-established methods of studying inheritance within families, genomic linkage mapping had become a highly productive tool for researchers and clinicians working in the genetics of rare diseases.

Towards association studies

At the same time, researchers were becoming increasingly frustrated by the limitations of family studies. Such studies were effective for mapping the kinds of single-gene disorders which follow Mendelian patterns of transmission and segregation from one generation to the next. But Mendelian conditions are mostly rare, and attracted little medical interest beyond the specialism of medical genetics. With rapid developments in molecular biology yielding powerful new research tools, by the early 1990s scientists were growing increasingly ambitious to identify genetic determinants not just of Mendelian disorders but also of common disorders such as diabetes and heart disease.

It had long been known that elevated risk of developing some of the commonest health disorders often runs in families. Since the late 1950s a growing body of research using twin studies and other methods had

provided geneticists with what they regarded as compelling evidence that a significant proportion of that elevated risk was hereditary rather than environmental in origin (Lindee, 2005, ch. 5). However, such conditions rarely displayed anything that could be identified as Mendelian patterns of inheritance; they occurred sporadically, albeit more frequently in certain families than in others – a fact that geneticists attributed to the involvement of multiple genes, each of which contributed to the probability of disease occurring but was insufficient on its own to make that occurrence a certainty. Researchers agreed that, given the sporadic occurrence of such conditions within affected families, family linkage methods were largely useless for mapping the predisposing genes (e.g. Lander and Botstein, 1986).

Confronted with these limitations, researchers began to consider other methods that they believed would be better suited to identifying the genetic factors involved in complex disorders. In place of the kind of one-to-one correspondences between genotype and disease that were the mainstay of family studies, researchers now looked for ways of identifying statistical associations between genomic markers and the occurrence of particular diseases, not just within families but among larger groups of people. Association studies were well established in epidemiology, where they were used to identify associations between environmental factors and increased risk of developing certain diseases, for instance. However, adapting these methods to study the genetics of common disorders was not straightforward. In order to identify a statistical association between a disease and a particular genetic marker, that marker would need to be very tightly linked to the gene that actually predisposed to that disease. Even with the growing numbers of RFLP and microsatellite markers that were available by the early 1990s, researchers realized, those markers remained too sparsely scattered on the human genome to serve as effective indicators of the presence of predisposing genes. If association methods were to have any hope of identifying genetic risk factors for complex disorders, it would first be necessary to develop human genetic linkage maps to a level of detail and resolution that far exceeded what was possible using RFLP and microsatellite markers (Bodmer, 1986; Lander and Botstein, 1986).

Consequently, for the time being, researchers sought instead to develop hybrid methods that in effect combined elements of family

studies with statistical methods for identifying associations. One such method, developed in the later 1980s, involved studying large numbers of pairs of siblings affected by particular conditions in order to identify statistical associations with genetic markers that they possessed in common (Bodmer, 1986; Kruglyak and Lander, 1995; Risch, 1989, 1990a, 1990b). This method proved effective in mapping gene variants associated with a number of common diseases, including demonstrating linkage between type 1 diabetes and a region of chromosome 6 associated with the body's immune response (Davies et al., 1994).

Another method, developed around the same time, involved studying population isolates where the high degree of in-breeding and interrelatedness meant that existing linkage maps were sufficiently fine-grained to detect disease associations (Lander and Botstein, 1986: 57–59; Lander and Schork, 1994). In the event, this method proved more successful in mapping a number of rare single-gene disorders (albeit without the need to reconstruct family pedigrees) than the common complex disorders that were increasingly preoccupying medical geneticists (Houwen et al., 1994; Puffenberger et al., 1994).

By the mid-1990s further rapid developments in the field of molecular biotechnology appeared to offer a way to develop more conventional forms of association studies. New DNA sequencing technologies connected with the Human Genome Project provided increasingly rapid means of identifying much larger numbers of genomic markers than had previously been possible. By mapping single nucleotide polymorphisms (SNPs) rather than RFLPs and microsatellite variants, much higher-resolution linkage maps now became a realistic prospect. This promised a step change in the rate at which researchers could identify and map human disease genes. Indeed, for a number of the most prominent scientific advocates of the Human Genome Project this was precisely the purpose of the whole enterprise (Fortun, 1999, 2008: 35–37).

Accordingly, as groups of researchers in Europe and North America began systematically collecting, cataloguing and mapping human SNPs they came increasingly to regard association methods as the most promising means of identifying and mapping gene variants implicated in common complex disorders (Collins, Guyer and Chakravarti, 1997; Lander, 1996; Risch and Merikangas, 1996). This commitment to creating the resources necessary to undertake genetic association studies would have profound consequences for how molecular and medical

geneticists thought about human genetic diversity and, ultimately, about human populations.

Capturing human genetic variation

Initially the work of identifying and cataloguing SNPs proceeded in a relatively uncoordinated fashion, with the establishment of local databases in a number of leading North American and European research centres. However, this work progressed against a backdrop of concern that researchers' access to large bodies of accumulated genomic data was threatened by moves to bring those data into private ownership. In the summer of 1997 it became apparent that a number of pharmaceutical companies were seeking to gain proprietary control of some of the leading collections of SNPs. In July of that year Abbot Laboratories announced a deal with Genset – a private company closely associated with the CEPH in Paris – to create two sets of SNPs, one for their own private research use and the other to market to other drug companies. At the same time Eric Lander, the founding director of the Massachusetts Institute of Technology's Whitehead Institute and one of the leading advocates of genetic association studies, was negotiating a deal with Bristol-Myers Squibb to create a similarly proprietary collection of SNPs to market for use in gene discovery (Marshall, 1997a).

Faced with this threat of privatization, which they feared would impede both publicly funded and commercial research into the genetics of disease and ill health, Francis Collins and other leading figures in the Human Genome Project were able to negotiate the creation in 1999 of a public-private partnership called the SNP Consortium, which included a number of leading pharmaceutical companies as well as the National Human Genome Research Institute (NHGRI) in the US and the Wellcome Trust in the UK (NHGRI, 2000). The SNP Consortium provided the organization and infrastructure to collate SNP discovery projects already under way, including a database called dbSNP which would serve as a public repository for the data generated.

More than simply collating existing SNP discovery projects, the creation of the SNP Consortium also provided an opportunity to impose a degree of order and shared purpose on those projects. In particular, it enabled the Consortium leadership to channel research in ways that were intended to promote the development of SNP maps that

would be optimally configured for use in genetic association studies. This led them to seek genomic data of a quite specific kind. SNPs are points of genetic variation between individuals. In order to facilitate the discovery of SNPs, Consortium members were keen to maximise the amount of variation present in the samples they analysed. They were helped in that aim by an initiative already under way at the NHGRI. As early as 1998 the NHGRI had announced the launch of a new research with the specific purpose of facilitating the identification of SNPs. The DNA Polymorphism Discovery Resource, as it was called, comprised a collection of DNA samples from '450 U.S. residents with ancestry from all the major regions of the world' (Collins, Brooks and Chakravarti, 1998; Marshall, 1997b). With the launch of the SNP consortium in July 2000, the NHGRI provided twenty-four samples from the DNA Polymorphism Discovery Resource, taken from donors 'with diverse geographic origins', to help with its work (NHGRI, 2000). This donation proved invaluable in facilitating the search for SNPs. Initially the Consortium had aimed to generate a map of 300,000 evenly spaced SNPs within three years. In the event, by the end of 2001 it had succeeded in compiling a map detailing over one million SNPs (International SNP Map Working Group, 2001).

The NHGRI's decision to collect and study genetic data from individuals 'with diverse geographic origins' bears careful analysis. The architects of the DNA Polymorphism Discovery Resource were at pains to declare that it was not intended to support research into the biology of racial or ethnic difference; indeed, it was deliberately designed in a way that rendered it useless for such research. 'No medical, phenotypic, or ethnicity information is included' with the samples, they stressed. 'The DNA Polymorphism Discovery Resource was designed to be used to discover variants in human DNA, not to assess the frequency of variants in particular groups. Thus, the DNA Polymorphism Discovery Resource is not useful for population-specific medical or anthropological studies' (Collins, Brooks and Chakravarti, 1998: 1229–1230).

The reasons for this were partly political. American genome researchers were acutely aware that any attempt to undertake genetic research that might be seen to impinge on matters of ethnic identity would be met with suspicion. A decade earlier, building on post-war surveys of human genetic variation as well as the availability of new molecular techniques to identify gene variants, population geneticists

and physical anthropologists had joined forces to launch the Human Genome Diversity Project. The aim was to sample indigenous populations around the world, particularly what geneticists regarded as endangered population isolates, in order to garner data on human origins and evolution (Reardon, 2005). The project had foundered amid charges of colonialism and racism. The architects of the DNA Polymorphism Discovery Resource therefore sought to distance themselves from that earlier debacle. They did so both by removing all racial, ethnic or geographical identifiers from the samples they collected and by presenting their work as an instance of how researchers were responding to complaints about Eurocentrism in biomedical research by deliberately including other ethnic groups (Bliss, 2012: 49–51).

The removal of ethnic and other identifiers from the DNA Polymorphism Discovery Resource was not merely a political gesture, however. It was also consistent with the purpose which the Resource was designed to serve – namely, to identify SNPs. For that purpose, there was no need to know anything about the ethnic or geographical origins of the genomes under study; it was sufficient merely to compare them and to identify the differences between them. In that respect, the Resource marked a significant break with earlier approaches to human genetic diversity. Previous anthropologically informed studies of genetic diversity, culminating in the Human Genome Diversity Project, had focused on identifying and characterizing the genetic differences between what they took to be different populations living or originating in different parts of the world – differences that were most starkly exemplified in so-called population isolates. In such studies the search for 'diversity' meant documenting how the human species had become subdivided into a number of more or less distinct evolutionary branches or backwaters. The DNA Polymorphism Discovery Resource certainly drew on such assumptions when deciding to recruit individuals whose ancestry was seen to lie in different parts of the world. But the aim in doing so was markedly different from that of earlier studies. The Resource did not seek to identify or describe genetic differences between the ancestral populations from which those individuals were supposedly drawn. On the contrary, it sought simply to maximise the number and range of genetic variants available for mapping. Beyond seeking to recruit as diverse a range of individuals as possible, the origins and ancestry of the individuals sampled were irrelevant to the aims of these studies

– hence the decision to remove geographical identifiers from the collected data. For the purposes of the DNA Polymorphism Discovery Resource, the genetic diversity of the population of the US was a useful resource, but it was not a matter for analysis.

Reconstituting populations

If the DNA Polymorphism Discovery Resource marked a step away from earlier efforts to use genetics to differentiate human populations, subsequent research initiatives decisively reasserted the old direction of travel. Following the success of the SNP Consortium in cataloguing and mapping unexpectedly large numbers of SNPs, scientists at the NHGRI proposed an even more ambitious project to identify the kinds of genomic variants that would help them to identify genes associated with common diseases. In autumn 2002 the US National Institutes of Health announced the launch of the International HapMap Project (NIH News Advisory, 2002). The aims and methods of the HapMap Project differed significantly from those adopted by the DNA Polymorphism Discovery Resource.

For one thing, where the SNP Consortium focused solely on sampling American citizens, the HapMap Project looked abroad to sample 'several populations from different ancestral geographic locations': initially Han Chinese living in Beijing, Japanese living in Tokyo, Yoruba from Ibadan in Nigeria and selected members of the Mormon families originally collected in 1980 for the CEPH project and classified within the HapMap Project as from Northern and Western Europe (International HapMap Consortium, 2003: 791). The reasoning behind this decision was again in part political. In the wake of the announcement of the first draft of the human genome in February 2001 and the growing public prestige that now attached to genome research, many feared that confining the research to the US would be seen as exclusionary. At the same time, by sampling large, culturally dominant groups in African and Asian countries it would be possible to avoid the charges of racism and colonialism that had attended the Human Genome Diversity Project. While neither of these strategies entirely avoided controversy and contestation, they were sufficient to secure participation by members of the four communities listed in the press release (Reardon, 2017: 70–93).

For another thing, the kind of genetic variation that the HapMap project sought to document differed in important ways from anything that had gone before. By the time that planning for the HapMap Project got under way, research using increasingly detailed SNP maps was revealing new details about how the human genome is structured. Among other things, it showed that DNA is organized not just into chromosomes but, at a finer level of organization, into haplotypes (Daly et al., 2001; Gabriel et al., 2002). A haplotype is a specific combination of SNPs that are not only situated close together on the genome but also tend to be inherited together across many generations – they exhibit particularly close genetic linkage, in other words. For molecular geneticists interested in mapping SNPs, and ultimately in identifying genetic variants associated with common diseases, the existence of haplotypes provided a welcome methodological shortcut: if researchers identified the presence of one or more SNPs peculiar to a particular haplotype, then they could infer with a high degree of probability what other SNPs markers were likely to be located in the immediate vicinity. Consequently, the HapMap project was organized with the express purpose of collecting not just SNPs but haplotypes as the preferred markers of human genetic variation.

However, the turn to haplotypes also opened the door to other kinds of genetic analysis. Since haplotypes are groups of genetic markers that tend to be inherited together, geneticists were able to read them not just as units of genetic variation but as indicators of common descent: if individuals share a haplotype, they must also have a common ancestor. In the context of the sampling strategy adopted by the HapMap Project this aspect of haplotypes quickly acquired a set of meanings that went well beyond the Project's professed claim to be concerned solely with variation. HapMap researchers did not just collect DNA samples from individuals; they sampled what they saw as specific populations, defined by ethnic identity and geographical location. As a result, the particular patterns of haplotypes identified in each of those populations were strongly associated from the start with particular ethnic groups and their supposedly disparate ancestral origins. As a number of commentators have observed, the particular choice of populations to study, in Africa, Asia and white North America, effectively served to reinscribe long-standing ideas about race into the findings of the HapMap Project, including the idea that different racial types could be mapped onto

particular continental locations (e.g. Duster, 2015; Hamilton, 2008). More generally, the very act of assembling different groups of people to sample, then characterising the differences between those groups in terms of distinctive hereditary patterns of haplotypes, served in effect to constitute the very populations which those haplotypes were supposed to represent (Reardon, 2017: 80–82).

This concern with sampling disparate populations, and the idea that those populations were genetically different from one another in important ways, in turn resonated with another, rather different understanding of populations that was becoming increasingly salient in debates about the feasibility of association studies as a means of identifying disease genes. As we have seen, earlier family linkage methods for identifying disease genes had not involved any explicit conceptualization of populations, since such studies focused on families as the object and means of investigation. However, once geneticists began considering the possibility of conducting association studies, the language of 'populations', and particular technical ideas about those populations, became central to their work.

For at least three decades before medical geneticists began to consider adopting association methods to elucidate genetic factors in disease, epidemiologists had been refining those methods for use in identifying environmental and other causes of ill health. Central to their methodological armamentarium were so-called case-control studies. In order to identify possible causes of illness, epidemiological researchers typically compare a group of affected cases with a group of non-affected controls and seek to identify statistical associations between the occurrence of the disease and specific environmental or other factors. In the course of developing such methods epidemiologists quickly realized that false positive results can occur if the cases and controls are not sufficiently similar to one another in relevant respects. Systematic differences between cases and controls, for instance in potentially confounding factors such as age or socio-economic status, could lead to misleading statistical associations between disease and environmental or other circumstances. For this reason, as early as the 1950s epidemiologists developed tools and methodologies designed to ensure as far as possible that cases and controls embodied the same 'population structure'.

In such studies epidemiologists use the language of populations pragmatically, to refer simply to the groups of cases and controls involved in the study. 'Population' in this sense implies nothing about the background of those who take part in a study, while 'population structure' is a consequence of the way that cases and controls are selected. That usage, and its connotations, changed markedly as medical geneticists began to adopt case-control methods, and epidemiological ideas about populations ran up against ideas drawn from population genetics. Problems of confounding due to unrecognized differences between cases and controls quickly became apparent as attempts to find associations between disease and specific genetic markers began to gain momentum. For instance, a 1980s study conducted among the Pima people of Arizona initially appeared to show an association between type 2 diabetes and a particular genetic marker. However, on further analysis the association was instead judged to be 'an artifact of population admixture'. Pima people had been selected for study because they experience a much higher incidence of diabetes than white North Americans – a fact that researchers hoped would facilitate their search for predisposing genes. However, on re-examining their findings the researchers found that the non-affected controls recruited into the study reported having more white Americans among their forebears than did the affected cases, who mostly claimed to have solely Pima ancestors. The researchers concluded that 'the association was apparently because tribe members have different degrees of Caucasian ancestry'; they had been misled by their failure to select 'a control group that is perfectly matched for ethnic ancestry' (Lander and Schork, 1994: 2041–2042).

It is worth pausing to reflect on the language adopted here. It reveals an important slippage: from thinking about populations and population structure in the instrumental language of epidemiology – referring simply to those individuals who together make up a study population – to thinking about populations in terms of population genetics – referring to the larger groups of people *from which* those individuals are judged to have been drawn. It also reveals a tendency for geneticists to think about 'population structure' not simply as an artefact of the selection of cases and controls but as something that already exists in the world from which the cases and controls are selected. This is particularly clear in the way that 'population structure' was equated with

'population admixture' in the Pima diabetes study. The very notion of 'population admixture' presumed not only that the Pima people from whom the research participants were drawn represented an uneven mix of two previously distinct genetic populations – the original Pima and 'Caucasians' – but also that the cases and controls had in effect been drawn from different sub-populations with 'different degrees of Caucasian ancestry'. In genetic case-control studies, in other words, the 'populations' which non-genetic epidemiologists would in principle have understood to have been constituted through the act of selecting cases and controls now came, in practice, to be seen as representing genetic populations that existed independently of the study methodology.

Such thinking persisted as high-density SNP maps and, subsequently, haplotype maps became available and researchers began conducting much more high-powered association studies using much larger populations of cases and controls. In such large-scale studies the potentially confounding effects of population structure (in the narrow epidemiological sense) would present a constant risk of spurious association. Consequently, researchers began developing increasingly powerful statistical techniques for analysing the distribution of SNPs within study populations, in order to discern any systematic differences between cases and controls. Their arguments were marked by constant slippage between instrumental talk of study populations and realist talk of populations of origin, and between population structure and population admixture (e.g. Devlin and Roeder, 1999; Marchini, Cardon, Phillips and Donnelly, 2004; Pritchard and Donnelly, 2001). This slippage was reinforced by the rolling-out of the HapMap Project and the growing use of haplotypes to identify population structure in association studies of disease-linked genetic markers.

Since the early 2000s association studies have proliferated and expanded, attracting large-scale research funding to study genetic factors in an increasingly wide range of medical and other conditions. Analysis of haplotypes is now routinely used in such studies as a means of controlling for population structure and ensuring that cases and controls are properly matched. In principle, this need not involve making inferences about what external populations might be represented in a study. It is possible, for instance, to use haplotype analysis to ensure simply that cases are compared with haplotypically similar controls within the study population. In practice, however, haplotype

analysis commonly draws on assumptions about the ancestry of study participants, and about the genetic make-up of geographically and ethnically defined populations from around the globe. The method of admixture mapping, for instance, relies on researchers not only identifying different genetic sub-populations within the study population but also attributing common ancestry and geographical origins to those sub-populations (Fujimura, Rajagopalan, Ossorio and Doksum, 2010; Fullwiley, 2008). By contrast, attributions of ancestry are not a necessary step in large-scale genome-wide association studies, which use specialized software to conduct purely statistical analyses of population structure. Even here, however, the haplotypes used to conduct such analyses typically derive from initiatives such as the HapMap Project, and hence ultimately refer back to assumptions about the differences between geographically and ethnically defined populations; while researchers often make their own assumptions about what ancestral populations they might expect to be represented in their study sample when deciding how to interpret and classify the sub-populations identified by their software. As a result, ideas of race, ethnicity and the genetic differences between populations are constantly being reinscribed in research into the genetic determinants of common disorders (Fujimura and Rajagopalan 2011; Gannett, 2014).

Conclusion

During the 1980s and 1990s efforts to identify and map genetic variants of possible significance for disease aetiology focused primarily on families. More recently, such research has shifted to include large-scale association studies in populations rather than in families. This has in turn led to the development of new methods to determine and control for population structure, which, in the case of genomic research, has come to mean the presence of sub-populations of different biological ancestry. Researchers are accordingly anxious to know about the genomic constitution not just of those populations that are the principle focus of their research but of other populations that might in effect intrude into their study samples. The implications of this have been twofold.

First, it has entailed a shift in medical thinking about human populations. Not only has it fostered a new reification of the idea of a population as something that is defined by common biological descent; it has

also led to a renewed interest in finding molecular techniques for dif-
ferentiating between such populations. As Joan Fujimura and Ramya
Rajagopalan put it, 'contrary to emphasizing the notion that humans
are all related', studies of population structure in the context of genomic
association studies are 'buttressed by a logic of difference' (Fujimura
and Rajagopalan, 2011, 21). This logic of difference provides a vehicle
by which old and supposedly discredited biological notions of race find
their way back into human genetics.

Second, this new thinking about genetic populations, and the desire
to differentiate between them, has led to the rolling-out of genetic sam-
pling on an increasingly global scale. In order to know what genes might
be involved in the incidence of heart disease among the inhabitants of
America or Britain, researchers now need to know about the genetic
constitution of populations from Mexico to Kenya to Japan. Local
studies must routinely take into account the global distribution of genes
and genotypes. In this respect, research into the genetic causes of
disease, wherever it is conducted, is increasingly global in its purview,
even when it is local and parochial in its concerns. This has prompted
a proliferation of studies, from the International HapMap Project to the
Human Heredity and Health in Africa (H3Africa) Initiative (launched
in 2010), designed to reveal in ever more detail the genetic make-up of
populations around the world.

Advocates of such initiatives declare that they are expected to benefit
the populations being studied. But, insofar as the data they produce are
used in the search for disease-causing gene variants, the vast majority of
that work has been oriented towards elucidating and ultimately reliev-
ing the health problems of people – especially white people – living
in North America and Europe (Need and Goldstein, 2009; Popejoy
and Fullerton, 2016). By comparison, the flow of medical knowledge,
and of such health interventions as result from these studies, back to
the world's poorer regions has been tiny. For one thing, the extent to
which knowledge about genetic causes of disease among Europeans
and North Americans is applicable to people with different genetic
constitutions is often unclear. For another, impoverished patients
and health systems simply do not have the resources to enable them
to make use of the often expensive and complex interventions that
modern biomedicine affords. The outcome of human genetic variation
research in Africa, Asia and South America has been overwhelmingly to

facilitate the development of health-related investigations and interventions among white Europeans and Americans. To the extent that this is the case, the globalization of genomic research has tended simply to reproduce the extractive relationships of neo-colonialism by extracting biological resources from the global South and realizing the value of those resources predominantly in the global North.

Acknowledgements

References

Bangham, Jenny (2014) 'Blood Groups and Human Groups: Collecting and Calibrating Genetic Data after World War Two', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (A), 74–86.

Bangham, Jenny (2015) 'What is Race? UNESCO, Mass Communication and Human Genetics in the Early 1950s', *History of the Human Sciences* 28 (5), 80–107.

Bliss, Catherine (2012) *Race Decoded: The Genomic Fight for Social Justice*. Stanford, CA: Stanford University Press.

Bodmer, Walter F. (1986) 'Human Genetics: The Molecular Challenge', *Cold Spring Harbor Symposia on Quantitative Biology* 51, 1–13.

Botstein, David, Raymond L. White, Mark Skolnick and Ronald W. Davis (1980) 'Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms', *American Journal of Human Genetics* 32, 314–331.

Cantor, David (2006) 'The Frustrations of Families: Henry Lynch, Heredity, and Cancer Control, 1962–1975', *Medical History* 50 (3), 279–302.

de Chadarevian, Soraya (2014) 'Chromosome Surveys of Human Populations: Between Epidemiology and Anthropology', *Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (A), 87–96.

Collins, Francis S., Lisa D. Brooks and Aravinda Chakravarti (1998) 'A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation', *Genome Research* 8, 1229–1231.

Collins, Francis S., Mark S. Guyer and Aravinda Chakravarti (1997) 'Variations on a Theme: Cataloguing Human DNA Sequence Variation', *Science* 278 (5343), 1580–1581

Comfort, Nathaniel (2012) *The Science of Human Perfection: How Genes Became the Heart of American Medicine*. New Haven, CT: Yale University Press.

Daly, Mark J. et al. (2001) 'High-Resolution Haplotype Structure in the Human Genome', *Nature Genetics* 29, 229–232.

Dausset, Jean et al. (1990) 'Centre d'Etude Du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome', *Genomics* 6 (3), 575–577.

Davies, June L. et al. (1994) 'A Genome-Wide Search for Human Type 1 Diabetes Susceptibility Genes', *Nature* 371, 130–136.

Devlin, Bernard J. and Kathryn Roeder (1999) 'Genomic Control for Association Studies', *Biometrics* 55, 997–1004.

Donis-Keller, Helen et al. (1987) 'A Genetic Linkage Map of the Human Genome', *Cell* 51, 319–337.

Duster, Troy (2015) 'A Post-Genomic Surprise: The Molecular Reinscription of Race in Science, Law and Medicine', *British Journal of Sociology* 66 (1), 1–27.

Fortun, Mike (1999) 'Projecting Speed Genomics', in *The Practices of Human Genetics: International and Interdisciplinary Perspectives*, ed. Mike Fortun and Everett Mendelsohn (pp. 25–48). Dordrecht: Kluwer.

Fortun, Mike (2008) *Promising Genomics: Iceland and DeCODE Genetics in a World of Speculation*. Berkeley; Los Angeles; London: University of California Press.

Fujimura, Joan H., Ramya Rajagopalan, Pilar N. Ossorio and Kjell A. Doksum (2010) 'Race and Ancestry: Operationalizing Populations in Human Genetic Variation Studies', in *What's the Use of Race? Modern*

Governance and the Biology of Difference, ed. Ian Whitmarsh and David S. Jones (pp. 169–186). Cambridge, MA: MIT Press.

Fujimura, Joan H. and Ramya Rajagopalan (2011) 'Different Differences: The Use of 'Genetic Ancestry' Versus Race in Biomedical Human Genetic Research', *Social Studies of Science* 41 (1), 5–30.

Fullwiley, Duana (2008) 'The Biologistical Construction of Race: 'Admixture' Technology and the New Genetic Medicine', *Social Studies of Science* 38 (5), 695–735.

Gabriel, Stacey B. et al. (2002) 'The structure of haplotype blocks in the human genome', *Science* 296, 2225–2229.

Gannett, Lisa (2001) 'Racism and Human Genome Diversity Research: The Ethical Limits of "Population Thinking"', *Philosophy of Science* 68 (3), Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association: S479–S492.

Gannett, Lisa (2003) 'Making Populations: Bounding Genes in Space and Time', *Philosophy of Science* 70 (5), 989–1001.

Gannett, Lisa (2014) 'Biogeographical Ancestry and Race', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (A), 173–184.

Gannett, Lisa and James A. Griesemer (2004) 'The ABO Blood Groups: Mapping the History and Geography of Genes in *Homo Sapiens*', in *Classical Genetic Research and Its Legacy: The Mapping Cultures of Twentieth-Century Genetics*, ed. Hans-Jörg Rheinberger and Jean-Paul Gaudillière (pp. 119–172). Abingdon: Routledge.

Gaudillière, Jean-Paul (2000) 'Mendelism and Medicine: Controlling Human Inheritance in Local Contexts, 1920–1960', *Comptes Rendus de l'Académie Des Sciences.* Série III: *Sciences de la Vie* 323, 1117–1126.

Gormley, Melinda (2009) 'Scientific Discrimination and the Activist Scientist: L. C. Dunn and the Professionalization of Genetics and Human Genetics in the United States', *Journal of the History of Biology* 42, 33–72.

Gurdasani, Deepti et al. (2015) 'The African Genome Variation Project Shapes Medical Genetics in Africa', *Nature* 517 (7534), 327–332.

Gusella, James F. et al. (1983) 'A Polymorphic DNA Marker Genetically Linked to Huntington's Disease', *Nature* 306 (5940), 234–238.

Hamilton, Jennifer A. (2008) 'Revitalizing Difference in the HapMap: Race and Contemporary Human Genetic Variation Research', *Journal of Law and Medical Ethic* 36 (3), 371–477.

Harper, Peter S. (2008) *A Short History of Medical Genetics*. Oxford: Oxford University Press.

Houwen, R. H. J. et al. (1994) 'Genome Scanning by Searching for Shared Segments: Mapping a Gene for Benign Recurrent Intrahepatic Cholestasis', *Nature Genetics* 8, 380–386.

International HapMap Consortium (2003) 'The International HapMap Project', *Nature* 426, 789–796.

International SNP Map Working Group (2001) 'A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms', *Nature* 409, 928–933.

Jones, Emma M. and E. M. Tansey (eds) (2015) *Human Gene Mapping Workshops c.1973–c.1991*. Wellcome Witnesses to Contemporary Medicine, vol. 54. London: Queen Mary University of London.

Kan, Yuet Wai and Andreé M. Dozy (1978) 'Polymorphism of DNA Sequence Adjacent to Human β-Globin Structural Gene: Relationship to Sickle Mutation', *Proceedings of the National Academy of Sciences* 75, 5631–5635.

Kaufmann, Alain (2004) 'Mapping the Human Genome at Généthon Laboratory: The French Muscular Dystrophy Association and the Politics of the Gene', in *From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth-Century Genetics*, ed. Jean-Paul Gaudillière and Hans-Jörg Rheinberger (pp. 129–157). Abingdon and New York: Routledge.

Knowlton, Robert G. et al. (1985) 'A Polymorphic DNA Marker Linked to Cystic Fibrosis is Located on Chromosome 7', *Nature* 318, 380–382.

Kohler, Robert E. (1994) *Lords of the Fly: Drosophila Genetics and the Experimental Life*. Chicago: University of Chicago Press.

Kruglyak, Leonid (2008) 'The Road to Genome-Wide Association Studies', *Nature Reviews Genetics* 9 (4), 314–318.

Kruglyak, Leonid and Eric S. Lander (1995) 'Complete Multipoint Sib-Pair Analysis of Qualitative and Quantitative Traits', *American Journal of Human Genetics* 57, 439–454.

Lander, Eric S. (1996) 'The New Genomics: Global Views of Biology', *Science* 274, 536–539.

Lander, Eric S. and David Botstein (1986) 'Mapping Complex Genetic Traits in Humans: New Methods Using a Complete RFLP Linkage Map', *Cold Spring Harbor Symposia on Quantitative Biology* 51, 49–62.

Lander, Eric S. and N. J. Schork (1994) 'Genetic Dissection of Complex Traits', *Science* 265 (5181), 2037–2048.

Lindee, M. Susan (2003) 'Voices of the Dead: James Neel's Amerindian Studies', in *Lost Paradises and the Ethics of Research and Publication*, ed. Francisco Salzano and Magedelene Hurtado (pp. 40–73). New York: Oxford University Press.

Lindee, M. Susan (2005) *Moments of Truth in Genetic Medicine*. Baltimore, MD: Johns Hopkins University Press.

Lipphardt, Veronika (2012) 'Isolates and Crosses in Human Population Genetics; or, a Contextualization of German Race Science', *Current Anthropology* 53 (Supplement 5), S69–S82.

Lipphardt, Veronika (2014) '"Geographical Distribution Patterns of Various Genes": Genetic Studies of Human Variation After 1945', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (A), 50–61.

Marchini, Jonathan, Lon R. Cardon, Michael S. Phillips and Peter Donnelly (2004) 'The Effects of Human Population Structure on Large Genetic Association Studies', *Nature Genetics* 36 (5), 512–517.

Marks, Jonathan (2012) 'The Origins of Anthropological Genetics', *Current Anthropology* 53 (Supplement 5), S161–S172.

Marshall, Eliot (1997a) 'Snipping Away at Genome Patenting', *Science* 277 (5333), 1752–1753.

Marshall, Eliot (1997b) '"Playing Chicken" Over Gene Markers', *Science* 278 (5346), 2046–2048.

Necochea, Raul (2007) 'From Cancer Families to HNPCC: Henry Lynch and the Transformations of Hereditary Cancer, 1975–1999', *Bulletin of the History of Medicine* 81 (1), 267–285.

Need, Anna C. and David B. Goldstein (2009) 'Next Generation Disparities in Human Genomics: Concerns and Remedies', *Trends in Genetics* 25, 489–494.

NHGRI (2000) 'Human Genome Project and SNP Consortium Announce Collaboration to Identify New Genetic Markers for Disease', July, www.genome.gov/10001456 (accessed 10 January 2015).

NIH News Advisory (2002) 'International Consortium Launches Genetic Variation Mapping Project', October, www.genome.gov/10005336 (accessed 10 January 2015).

Nukaga, Yoshio (2002) 'Between Tradition and Innovation in New Genetics: The Continuity of Medical Pedigrees and the Development of Combination Work in the Case of Huntington's Disease', *New Genetics and Society* 21 (1), 39–64.

Popejoy, Alice B. and Stephanie M. Fullerton (2016) 'Genomics is Failing on Diversity', *Nature* 538 (7624), 161–164.

Pritchard, Jonathan K. and Peter Donnelly (2001) 'Case-Control Studies of Association in Structured or Admixed Populations', *Theoretical Population Biology* 60, 227–237.

Puffenberger, Erik G. et al. (1994) 'Identity-by-Descent and Association Mapping of a Recessive Gene for Hirschsprung Disease on Human Chromosome 13q22', *Human Molecular Genetics* 3, 1217–1225.

Rabinow, Paul (1999) *French DNA: Trouble in Paradise*. Chicago: University of Chicago Press.

Radin, Joanna (2014) 'Unfolding Epidemiological Stories: How the WHO Made Frozen Blood into a Flexible Resource for the Future', *Studies in*

History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 47 (A), 62–73.

Reardon, Jenny (2004) 'Decoding Race and Human Difference in a Genomic Age', *Differences: A Journal of Feminist Cultural Studies* 15 (3), 38–65.

Reardon, Jenny (2005) *Race to the Finish: Identity and Governance in an Age of Genomics*. Princeton; Oxford: Princeton University Press.

Reardon, Jenny (2017) *The Postgenomic Condition: Ethics, Justice, and Knowledge After the Genome*. Chicago and London: Chicago University Press.

Rheinberger, Hans-Jörg and Jean-Paul Gaudillière (eds) (2004) *Classical Genetic Research and Its Legacy: The Mapping Cultures of Twentieth-Century Genetics*. London: Routledge.

Risch, Neil (1989) 'Genetics of IDDM: Evidence for Complex Inheritance with HLA', *Genetic Epidemiology* 6, 143–148.

Risch, Neil (1990a) 'Linkage Strategies for Genetically Complex Traits. II. The Power of Affected Relative Pairs', *American Journal of Human Genetics* 46, 229–241.

Risch, Neil (1990b) 'Linkage Strategies for Genetically Complex Traits. III. The Effect of Marker Polymorphism on Analysis of Affected Relative Pairs', *American Journal of Human Genetics* 36, 242–253.

Risch, Neil and Kathleen R. Merikangas (1996) 'The Future of Genetic Studies of Complex Human Diseases', *Science* 273, 1516–1517.

Smocovitis, Vassiliki Betty (2012) 'Humanizing Evolution: Anthropology, the Evolutionary Synthesis, and the Prehistory of Biological Anthropology, 1927–1962', *Current Anthropology* 53 (Supplement 5), S108–S125.

Solomon, Ellen and Walter F. Bodmer (1979) 'Evolution of Sickle Variant Gene', *Lancet* 313, 923.

Tsui, Lap-Chee et al. (1985) 'Cystic Fibrosis Locus Defined by a Genetically Linked Polymorphic DNA Marker', *Science* 230 (4729), 1054–1057.

Wailoo, Keith and Stephen Pemberton (2008) *The Troubled Dream of Genetic Medicine: Ethnicity and Innovation in Tay-Sachs, Cystic Fibrosis, and Sickle Cell Disease*. Baltimore: Johns Hopkins University Press.

Weissenbach, Jean et al. (1992) 'A Second-Generation Linkage Map of the Human Genome', *Nature* 359 (6398), 794–801.

Widmer, Alexandra (2014) 'Making Blood "Melanesian": Fieldwork and Isolating Techniques in Genetic Epidemiology (1963–1976)', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (A), 118–129.