Studies in History and Philosophy of Biological and Biomedical Sciences 47 (2014) 173–184

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Biogeographical ancestry and race

Lisa Gannett

Department of Philosophy, Saint Mary's University, 923 Robie Street, Halifax B3H 3C3, Canada

ARTICLE INFO

Article history:
Available online 29 June 2014

Keywords:
Ancestry-informative marker (AIM)
Biogeographical ancestry (BGA)
Genetics
Genomics
Population structure
Race

ABSTRACT

The use of racial and ethnic categories in biological and biomedical research is controversial—for example, in the comparison of disease risk in different groups or as a means of making use of or controlling for population structure in the mapping of genes to chromosomes. Biogeographical ancestry (BGA) has been recommended as a more accurate and appropriate category. BGA is a product of the collaboration between biological anthropologist Mark Shriver from Pennsylvania State University and molecular biologist Tony Frudakis from the now-defunct biotechnology start-up company DNAPrint genomics, Inc. Shriver and Frudakis portray BGA as a measure of the 'biological', 'genetic', 'natural', and 'objective' components of race and ethnicity, what philosophers of science would call a natural kind. This paper argues that BGA is not a natural kind that escapes social and political connotations of race and ethnicity, as Shriver and Frudakis and other proponents believe, but a construction that is built upon race—as race has been socially constructed in the European scientific and philosophical traditions. More specifically, BGA is not a global category of biological and anthropological classification but a local category shaped by the U.S. context of its production, especially the forensic aim of being able to predict the race or ethnicity of an unknown suspect based on DNA found at the crime scene. Therefore, caution needs to be exercised in the embrace of BGA as an alternative to the use of racial and ethnic categories in biological and biomedical research.

1. Introduction: categories of race and ethnicity in biological and biomedical research

Completion of the Human Genome Project (HGP), and the consequent shift of resources to investigating patterns of variability across genomes that are of possible biological and biomedical interest, have contributed to an increased use of categories of race and ethnicity to study group-based genetic and genomic differences.

'Single-gene' diseases that appear at different frequencies in different racial and ethnic groups are familiar to us: for example, in the United States, sickle-cell anaemia is more common in African Americans and cystic fibrosis is more common in European Americans. Mutations implicated in these relatively rare diseases have been mapped to locations on the chromosomes. The focus is now on mapping genes that are involved in more common 'complex' diseases such as hypertension, asthma, cancer, and dementia, but the task has proved more challenging than the rhetoric selling the benefits of the HGP promised, and than geneticists themselves envisioned. In complex diseases, multiple genetic, epigenetic, and environmental causal factors need to be identified and their separate roles and interaction effects unravelled. Some complex diseases occur at increased frequencies in particular racial and ethnic groups: for example, hypertension in African Americans and diabetes in Native Americans. Although they recognize the importance of epigenetic and environmental factors and discount genetic determinism, many biomedical geneticists believe that genetic variants found at different frequencies in different racial and ethnic groups are implicated in complex diseases, just as in single-gene diseases.

The use of racial and ethnic categories in biological and biomedical research is controversial, however. In the United States, for example, criticisms are directed against the double duty that is

E-mail address: lisa.gannett@smu.ca.

http://dx.doi.org/10.1016/j.shpsc.2014.05.017
1369-8486/

asked of the Office of Management and Budget (OMB) categories of race and ethnicity. The OMB regulates the collection of racial and ethnic data by federal agencies. When OMB Directive 15 was issued in 1977, its aim was to standardize the racial and ethnic categories used to collect data so that compliance with recently passed civil rights legislation that prohibited discrimination in areas like housing, education, and employment could be monitored. The OMB system of classification, as it was updated most recently in 1997, asks individuals to self-identify, first, by 'ethnicity' (their choices are 'Hispanic or Latino' and 'Not Hispanic or Latino') and, second, by 'race' (their choices are 'American Indian or Alaska Native', 'Asian', 'Black or African American', 'Native Hawaiian or Other Pacific Islander', and 'White', and more than one racial class can be chosen). According to the OMB (1997), 'The categories represent a social-political construct designed for collecting data on the race and ethnicity of broad population groups in this country, and are not anthropologically or scientifically based' ('Supplementary Information', Sec. A). Consistent with this mission is use of the OMB categories in clinical research where the variables of interest are social and related to the history of racial and ethnic discrimination in the United States: for example, access to health care, exposure to racism, income level, educational attainment, proximity to environmental toxicity, etc. But, as critics point out, use of the OMB categories in biologically-based clinical research goes beyond this, and risks suggesting that 'there are fundamental biological and behavioral differences among racial groups' (e.g., Institute of Medicine, 1999, p. 82).

Given existing controversy concerning the use of racial and ethnic categories in biological and biomedical research, their cooption in service of the genetic mapping of complex traits has been a topic of debate. Mapping genes implicated in complex traits is a more complicated endeavour than mapping genes implicated in single-gene or Mendelian traits. An early strategy extended the linkage mapping used for Mendelian diseases that segregate in families to populations supposed to be relatively homogeneous, such as the Icelandic population. In more heterogeneous populations, different strategies are required, several of which have introduced the expertise of population geneticists, evolutionary geneticists, and anthropological geneticists into biomedical genetics. One such strategy is admixture mapping. Admixture mapping (AM) makes use of linkage in a different way: specifically, the linkage disequilibrium that arises when historically separated populations combine. Correlations are expected in the 'admixed population' between incidence of the disease and regions of the genome derived from the ancestral population in which the disease is most prevalent. Another strategy is made possible by the ability to screen for many mutations at the same time. Genome-wide association studies (GWAS) use large numbers of individuals divided into cases and controls to seek correlations between disease incidence and SNP (single nucleotide polymorphism) markers or haplotypes (characteristic patterns of neighbouring markers) across the genome.

Progress has been slow using AM and GWAS to map genes associated with complex traits, and 'population structure' has both helped and hindered. 'Population structure' refers to the pattern by which genetic variability is distributed across a species or subpopulation of a species. Genetic variability is patterned in ways that reflect the past and present operation of factors of evolutionary importance such as selection, drift, migration, and mating structure (random, assortative, etc.). Population genetic structure is a help in mapping genes associated with complex traits insofar as it is the very basis of the linkage disequilibrium used in AM. Population genetic structure is a hindrance in both AM and GWAS insofar as it functions as a confounding variable when 'false positives' are yielded by genetic variants that are more frequent in the population in which the trait of interest is more frequent but play no causal role. Racial and ethnic categories are employed in the delineation of ancestral populations in AM and in order to control for population structure in AM and GWAS. Scientists recognize that doing so is problematic. They realize that it is difficult to extricate race and ethnicity from their social and political connotations when these categories are used in biological and biomedical research. They are also concerned about whether racial and ethnic self-identification by subjects is appropriate when the goals are scientific. Consequently, there have been efforts to come up with alternative group concepts that gain in scientific objectivity by ostensibly foregoing the social and political connotations of race and ethnicity.

In this paper, I focus on the newly invented concept 'biogeographical ancestry'. Biogeographical ancestry has met with a positive reception by knowledgeable critics and policy-makers. A critical examination of the use of race as a category of classification in the mapping of complex traits by a multidisciplinary group of U.S. biological and social scientists accepts that biogeographical ancestry provides an objective method of controlling for population structure (Shields et al., 2005). The authors consider self-identified race (i.e., the OMB classification) to be the appropriate category to use for monitoring health disparities and an acceptable category to use for recruitment for genetic studies. But they recommend that genetic studies themselves assess population structure empirically by genotyping individuals to determine their 'continental ancestry' proportions. Because of the social harms that attach to using race as a variable in genetic research, they urge the National Institutes of Health (NIH) to provide financial support for developing and facilitating the use of such tools. Citing this work, a 2005 review article by the National Human Genome Research Institute's (NHGRI's) Race, Ethnicity, and Genetics Working Group, titled 'The Use of Racial, Ethnic, and Ancestral Categories in Human Genetics Research', recognizes that the U.S. census categories were not designed for genetic research and suggests that instead, 'labels based on biogeographical ancestry may be suited for many genetics studies' (p. 526), thus providing a means of moving beyond self-identified race, ethnicity, and ancestry as proxies.

I am more sceptical than these critics and policy-makers about biogeographical ancestry as an alternative to the use of categories of race and ethnicity in biological and biomedical research. Biogeographical ancestry emerged as a product of the collaboration between biological anthropologist Mark Shriver and members of his Pennsylvania State University laboratory and molecular biologist Tony Frudakis's now-defunct biotechnology start-up company DNAPrint genomics. We will see that the research context surrounding the emergence of the concepts and techniques associated with biogeographical ancestry was shaped by diverse interests—social and commercial as well as scientific—in DNA forensics, gene mapping, pharmaceutical development, and direct-to-consumer genealogy testing. Close attention to this research context reveals that BGA is itself a construction built upon race as race has been socially constructed in the European scientific and philosophical traditions, but especially in the United States. As such, BGA does not provide a means of moving beyond 'proxy' social categories like race, ethnicity, and ancestry, as the NHGRI's Race, Ethnicity, and Genetic Working Group assumes, and BGA does not provide a wholly empirical method of assessing population structure, as is the hope of the multidisciplinary group of critics.

In the next section, I outline the steps leading to the invention of biogeographical ancestry through the collaborative efforts of Shriver and Frudakis. To say that BGA is an invention is to emphasize not only its invention as a concept but as a technology for which patent protection was sought.