

A genetic melting-pot

Marcus W. Feldman, Richard C. Lewontin and Mary-Claire King

Race as a biological concept has had a variety of meanings. In the taxonomic literature, a race is any distinguishable type within a species, such as dark-bellied and light-bellied variants of small mammals. In 1937, Theodosius Dobzhansky introduced the idea of geographical races — populations of species that differ in the frequencies of one or more genetic variants. But as no two populations have identical gene frequencies at variable (polymorphic) loci, Dobzhansky's definition of race becomes synonymous with that of population.

The classical definition of race, as applied to our species, is based on phenotypes such as skin colour, facial features and hair form that clearly differ between native inhabitants of different regions of the world. An underlying assumption is that all of these defining features (all largely genetic traits, although few of their genes have been identified) are characteristic of the genome in general. In other words, just as there are large differences between races in genes for skin colour, so there should be large genetic differences between races in general. In the previous absence of data to confirm or deny this assumption, it was not an unreasonable one to make.

But recent studies of genetic diversity indicate that the genes underlying the phenotypic differences used to assign race categories are atypical, in that they vary between races much more than genes in general. Together, the iconic features of race correlate well with continent of origin but do not reflect genome-wide differences between groups.

Discussion has arisen over the implications of these findings for the utility of racial classification in medical practice. The issue of whether race is a biologically useful or even meaningful concept when applied to humans in a medical context is controversial — holders of opposing views each claim to have evidence to support them. But there is no contradiction between these two well-substantiated bodies of data, as they actually

deal with two different questions that have become confused with one another.

The first question is: "Is it possible to find DNA sequences that differ sufficiently between populations to allow correct assignment of major geographical origin with high probability?" The answer to this question is yes, as shown by studies of genetic polymorphisms and by universal personal experience.

The second question is: "What fraction of human genetic variation, whether based on protein-coding genes or other sequences, falls within geographically separated populations, and what fraction occurs between these populations?" The answer to this question is that most genetic diversity occurs within groups, and that very little is found between them.

Why this apparent paradox? The answer is that genes that are geographically distinctive in their frequencies are not typical of the human genome in general.

It has been suggested that racial categorization has a valid role in good medical practice because many medically important genes vary between populations from different regions. But although knowing a patient's ancestry is often extremely useful in diagnosis and treatment, race is both too broad and too narrow a definition of ancestry to be biologically useful.

For any species, definitions of race can lose their discriminating power when individuals migrate to different regions and mate with their counterparts there. Among humans, large-scale migrations between continents — particularly through European colonial expansion and the commercial slave trade — has resulted in matings of individuals from different continents and the creation of new populations, especially in the Western Hemisphere and Oceania. Many people thus have ancestry from more than one major geographical region, meaning that the association of phenotype and geography breaks down.

For example, sickle-cell disease, which is often thought to be an African trait, is instead characteristic of ancient ancestry in a geographic region where malaria was endemic. Africa is one such region, but so are the Mediterranean and southern India. If sickle-cell disease is suspected, then the correct diagnostic approach is not simply to determine the patient's race, but to ask whether they have African, Mediterranean or South Indian ancestry. To use genotype effectively in making diagnostic and therapeutic decisions, it is not race that is relevant, but both intra- and trans-continental contributions to a person's ancestry.

Race and ancestry are confounded both by genetic heterogeneity within groups and by the widespread mixing of previously iso-

Race

Ancestral genetic data are far more useful for medical purposes than are racial categories, which may be correlated with disease for social or economic rather than biological reasons.

lated populations. The assignment of a racial classification to an individual hides the biological information that is needed for intelligent therapeutic and diagnostic decisions. A person classified as 'black' or 'Hispanic' by social convention could have any mixture of ancestries, as defined by continent of origin. Confusing race and ancestry could be potentially devastating for medical practice.

Other attempts to classify people into broad genetic groups based on the frequency of specific genes for, say, drug-metabolizing enzymes, are also likely to be poor predictors of medical outcome. As with racial groupings, the overall variation in the frequencies of such genes between groups is likely to be less than that within each group.

The conventional, social definition of race is useful in a medical context as it provides information about the social circumstances and lifestyle of patients. But this is a consequence of social history, so any variation is (at least in principle) transitory. By contrast, information on the likelihood that a person carries specific disease-related or treatment-response genes is grounded in their ancestry in far more complex ways. We suggest that identifying all contributions to a patient's ancestry can be useful in diagnosing and treating diseases with genetic influences. Eventually, for both diagnosis and treatment, specific genetic variants will provide concrete, useful information.

Marcus W. Feldman is in the Department of Biological Sciences, Stanford University, California 94305, USA.

Richard C. Lewontin is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Mary-Claire King is in the Departments of Genome Science and Medicine, University of Washington, Seattle, Washington 98195, USA.

FURTHER READING

- Rosenberg, N. A. *et al. Science* **298**, 2381–2385 (2002).
- González Burchard, E. *et al. N. Engl. J. Med.* **348**, 1170–1175 (2003).
- Cooper, R. S., Kaufman, J. S. & Ward, R. *N. Engl. J. Med.* **348**, 1166–1170 (2003).
- Lewontin, R. C. *Evol. Biol.* **6**, 381–398 (1972).
- Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L. L. *Proc. Natl Acad. Sci. USA* **94**, 4516–4519 (1997).



Human migration, as depicted in Charles Hunt's *Landing at Madras*, blurs the boundaries of race.