

Slicing soup

The end of May was the deadline for applications to join an initiative among several institutes and centers at the US National Institutes of Health to develop a haplotype map of the human genome. Once developed, this map should prove a valuable resource for association studies aimed at finding genes affecting health and responses to drugs and other environmental factors. A particularly thorny issue surrounding construction of the map is whether data should be collected from populations defined by race and ethnicity, which are at best imperfect proxies for biological relatedness. Tracking disease down such ethnic pathways not only is scientifically questionable, but also raises concern over a new potential for racial profiling in medicine.

The literature abounds with research efforts that attempt (and fail) to find significant correlations between race and molecularly characterized disease. A sample of papers from the past few months illustrates the point: DNA damage in morphologically normal breast tissue correlates with the smoking status of patients but not with, among other things, race (*Carcinogenesis* 23, 301–306, 2002); likewise, the distribution of low-renin hypertension, which represents about a quarter of all essential hypertension, is the same among different racially defined populations (*Hypertension* 39, 914, 2002); and there is no evidence for a link between ethnic group and such gross parameters as the thickness of the carotid artery intima-media wall, which is a useful measure of atherosclerosis (*Stroke* 33, 1420, 2002).

What these studies demonstrate all too clearly, however, is that biomedical investigators are still in the habit of looking for racial correlations of molecular or physiological differences. The ostensible justification is that race is an important parameter in stratifying human disease phenotypes. This is dangerous territory, not because it is likely to inflame cultural sensitivities (although it certainly will), but because the genetic foundation of race-based studies is tenuous.

There are, of course, many simple inherited conditions that occur more frequently in some human groupings than in others. Members of certain Jewish groups, for example, often seek testing for genes that signal a risk for breast cancer or Tay–Sachs disease. Complex conditions such as diabetes, hypertension, coronary heart disease, and obesity also occur with varying frequencies in populations with particular geo-social histories. Recent studies in South Africa and Antioquia, Colombia suggest that genetic predispositions to conditions such as Parkinson's disease, hypertension, or hypercholesterolemia occur in the genomes of admixed populations as a result of gene flow from European settlers. Such disease-related linkage disequilibrium can seduce researchers into believing that focusing studies by ethnic grouping is, in general, a good way to discover genetic associations. The ease of racial pigeonholing may be another attraction.

However, to paint genetic gloss onto current ethnotyping, or even to redefine racial groups in terms of haplotypes, is a fundamental error. In applying the lancet of polymorphisms and haplotypes to dissect possible genetic contributions to complex conditions, the conditions must be well defined and their environmental influences documented. Race studies cannot meet this requirement, except perhaps in cases where race has a basis in geographically based reproductive isolation—

among such peoples as the Khoi, the San, Amerind groups, and other isolated populations. In these cases, geography provides an independent variable for defining the group.

But in most Western nations, especially perhaps in the United States, this makes little sense. On average, humans are remarkably (~99.99%) alike genetically, regardless of skin color. The human genome sequence reveals, for example, that many people of black African descent are closer genetically to whites than they are to other black Africans. The entire human population differs from its closest relatives (chimps) by only 1% of its genome sequence (roughly 1 base pair in 100) and has less genetic variation than a single tribe of east African chimpanzees.

In addition, ethnic origins are poorly characterized. In the United States, “whites” have origins in Europe and the Middle East; “blacks” are largely of African stock, admixed liberally with “whiteness”; “Asian Americans” have hugely diverse geographical origins; and “Hispanic-Latinos” have mixed European, Amerind, and African roots.

The current predilection for the use of ethnicity as a variable in the evaluation of clinical studies and the increasing willingness to separate populations into distinct ethnic or racial groupings for targeting by specific treatments (a kind of poor man's pharmacogenetics) is also worrisome. Already, companies are attempting to obtain approval for drugs that have failed in the general population but may work for a particular ethnic subgroup. In March of last year, for example, the US Food and Drug Administration issued an approvable letter for NitroMed's heart-failure drug BiDil, pending the results of a late-stage trial in “African Americans.” As yet, no genes have been pinpointed to explain why BiDil might work better for this group; this could be attributable to diet, access to health care, or any of a multitude of other factors.

Last year, a study reported in *Nature Genetics* (29, 265–269, 2001) definitively showed that genetics is a far better predictor of drug response than clustering by race; in fact, grouping by race can actually mask important differences in drug response. Race is thus a crude measure of whether a drug will work for a given patient. But the acceptance of racial profiling to subdivide populations in drug studies raises the much more serious concern that pharmaceutical companies may target products particularly toward whites—the largest and most affluent market—while ignoring poorer minority groups.

For researchers, it is tautological to seek genetic definitions of race or ethnicity in order to simplify or streamline subsequent genetic studies. To explore genetic associations with race, one must first find a substantive nongenetic basis for describing and separating racial groups. Most “races” have no basis in physical, physiological, or biochemical criteria, at least none that has been convincingly established.

Race, in countries like the United States at least, is a fuzzy social construct by which people with one or two superficial similarities are often clumped together. It reflects simplistic cultural habits, reinforced by the questionable practices of government statisticians and medical researchers, among others. Ethnic binning may simplify thought processes and, in some cases, negate them altogether. But using genetics to define race is like slicing soup. You can cut wherever you want, but the soup stays mixed.