



What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era

Francis S Collins

A true understanding of disease risk requires a thorough examination of root causes. 'Race' and 'ethnicity' are poorly defined terms that serve as flawed surrogates for multiple environmental and genetic factors in disease causation, including ancestral geographic origins, socioeconomic status, education and access to health care. Research must move beyond these weak and imperfect proxy relationships to define the more proximate factors that influence health.

A small meeting convened at the National Human Genome Center at Howard University in Washington, D.C., on 15 May 2003, titled "Human Genome Variation and 'Race': The State of the Science," marked an important, positive milestone in the turbulent history of genetics, race and ethnicity. Experts in sociology, anthropology, history and genetics gathered together to discuss, in an honest and unemotional way, the substance of what we know and what we don't know about the connections between genetics and race. The few meetings held in the past decade to discuss this highly charged topic have often been unsatisfactory, either because participants with strong opinions tended to talk past each other or, more commonly, because heightened sensitivity to the possibility of giving inadvertent offense caused those present to speak only in politically correct generalities. As a historically black university, Howard University served science and society by sponsoring this frank discussion, and the National Human Genome Center's leaders are to be congratulated for their vision in putting together such a thought-provoking agenda at a time when large amounts of new

information about human genetic variation are coming to light. Many of the salient points made by participants in this meeting were captured in the preceding articles.

The meeting at Howard University focused on exactly the right questions. What does the current body of scientific information say about the connections among race, ethnicity, genetics and health? What remains unknown? What additional research is needed? How can this information be applied to benefit human health? How might this information be applied in nonmedical settings? How can we adopt policies that will achieve beneficial societal outcomes?

Is race biologically meaningless?

First, it is essential to point out that 'race' and 'ethnicity' are terms without generally agreed-upon definitions. Both terms carry complex connotations that reflect culture, history, socioeconomic and political status, as well as a variably important connection to ancestral geographic origins. Well-intentioned statements over the past few years, some coming from geneticists, might lead one to believe there is no connection whatsoever between self-identified race or ethnicity and the frequency of particular genetic variants^{1,2}. Increasing scientific evidence, however, indicates that genetic variation can be used to make a reasonably accurate prediction of geographic origins of an individual, at least if that individual's grandparents all came from the

same part of the world³. As those ancestral origins in many cases have a correlation, albeit often imprecise, with self-identified race or ethnicity, it is not strictly true that race or ethnicity has no biological connection. It must be emphasized, however, that the connection is generally quite blurry because of multiple other nongenetic connotations of race, the lack of defined boundaries between populations and the fact that many individuals have ancestors from multiple regions of the world.

Race and health disparities

What about health disparities? Are genetic differences between populations likely to have a role in health status, both in the US and around the world? In many instances, the causes of health disparities will have little to do with genetics, but rather derive from differences in culture, diet, socioeconomic status, access to health care, education, environmental exposures, social marginalization, discrimination, stress and other factors⁴. Yet it would be incorrect to say that genetics never has a role in health disparities. This is most obvious in the unequal distribution of disease-associated alleles for certain recessive disorders, such as sickle cell disease or Tay-Sachs disease, but has also been noted recently for certain nonmendelian disorders, such as Crohn disease⁵.

The question of whether genetics will explain a substantial proportion of health disparities for most common diseases is largely

National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. Correspondence should be addressed to F.S.C. (fc23a@nih.gov).

Published online 26 October 2004;
doi:10.1034/ng1436

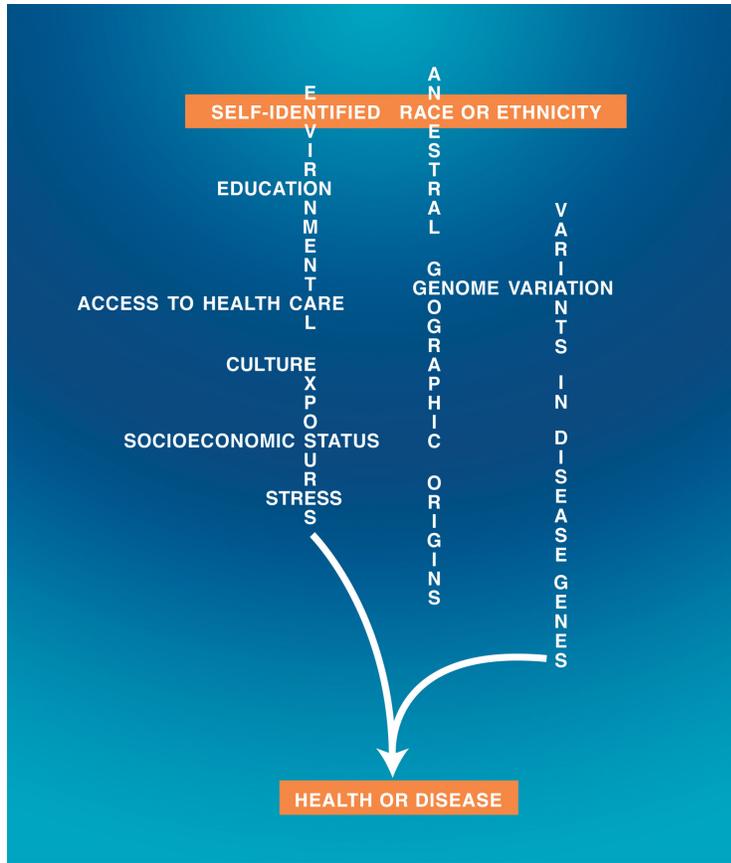


Figure 1 Interconnections between self-identified race or ethnicity and health status. The undeniable existence of health disparities indicates that there is a correlation between self-identified race or ethnicity and health or disease in some cases. But this is a complex and poorly understood relationship. On the left side of the diagram, multiple environmental factors that are influenced by race and ethnicity, and that potentially contribute to health disparities, are depicted. On the right side, the potential genetic contribution to health disparities, which operates through a series of proxy relationships, is depicted. To unravel the real causes, research into health disparities must move beyond weakly correlated variables, such as self-identified race or ethnicity, towards an understanding of the more proximate environmental and genetic factors.

unanswered and will be clarified only by further research studies of many populations. Given that the frequency of many genetic variants is not equal in all parts of the world⁶, however, genetic variations conferring disease susceptibility are expected to be unequally distributed, at least in some cases.

Finding common ground

A vigorous debate has raged in the scientific and medical literature over the last few years about whether there is any value in using self-identified race or ethnicity to identify factors that contribute to health or disease^{7,8}. Proponents of maintaining such identifiers argue that even if the genetic component of health disparities is small, self-identified race or ethnicity is also a useful proxy for other correlated nongenetic variables, and to lose the opportunity to explore these would be

doing a disservice to the public. Detractors argue that race and ethnicity are such flawed concepts that the persistent use of such descriptors prolongs the delay in seeking real causes and lends more scientific validity to the race-health connection than it deserves.

After reviewing these arguments and listening to the debate during the meeting at Howard University, one could conclude that both points are correct. The relationship between self-identified race or ethnicity and disease risk can be depicted as a series of surrogate relationships (Fig. 1). On the nongenetic side of this diagram, race carries with it certain social, cultural, educational and economic variables, all of which can influence disease risk. On the genetic side of the diagram, race is an imperfect surrogate for ancestral geographic origin, which in turn is a surrogate for genetic variation across an indi-

vidual's genome. Likewise, genome-wide variation correlates, albeit with far-from-perfect accuracy, with variation at specific loci associated with disease. Those variants interact with multiple environmental variables, with the ultimate outcome being health or disease.

Considered in this context, it is apparent why self-identified race or ethnicity might be correlated with health status, through genetic or nongenetic surrogate relationships or a combination of the two. It is also evident that a true understanding of disease risk requires us to go well beyond these weak and imperfect proxy relationships. And if we are not satisfied with the use of imperfect surrogates in trying to understand hereditary causes, then we should not be satisfied with them as measures of environmental causation either.

What additional research is needed?

The recent National Human Genome Research Institute's "Vision for the Future of Genomics Research"⁹ outlined a bold agenda for the future, including a number of compelling research opportunities. The meeting at Howard University underscored the importance of additional research in certain crucial areas:

(i) Without discounting self-identified race or ethnicity as a variable correlated with health, we must strive to move beyond these weak surrogate relationships and get to the root causes of health and disease, be they genetic, environmental or both.

(ii) To determine accurate risk factors for disease, we need to carry out well-designed, large-scale studies in multiple populations. Such studies must be equally rigorous in their collection of genetic and environmental data. If only genetic factors are considered, only genetic factors will be discovered.

(iii) To validate quantitative conclusions about genes, environment and their interactions in health and disease for multiple groups, long-term, longitudinal prospective cohort studies, as well as carefully designed case-control studies, will be needed¹⁰.

(iv) We must continue to support efforts to define the nature of human variation across the world, focused primarily on medical goals. The International Human Haplotype Map Project¹¹ will open a new window into human variation and generate a powerful tool for discovering disease associations, but the project will provide a resource, not all of the answers.

(v) We need more anthropological, sociological and psychological research into how individuals and cultures conceive and internalize concepts of race and ethnicity.

(vi) We must assess how the scientific community uses the concepts of race and ethnicity and attempt to remedy situations in which the use of such concepts is misleading or counterproductive.

(vii) We need to formulate clear, scientifically accurate messages to educate researchers, health-care professionals and the general public on the connections among race, ethnicity, genetics and health.

Conclusion

The individuals attending the meeting at Howard University represented a group of highly informed and sophisticated thinkers. Many participants had spent more than a decade trying to untangle these complicated concepts. A substantial degree of consensus was achieved regarding what we currently know, but it was impossible to escape the fact that substantial gaps in our current knowledge remain. Therefore, the research and the conversation must continue.

In that vein, the National Human Genome Research Institute convened a Roundtable on Race, Ethnicity, and Genetics on 8–10 March 2004, which was attended by a wide range of

thought leaders in genetics, anthropology, sociology, history, law and medicine. A report of that meeting is being prepared for publication. The National Human Genome Research Institute is also sponsoring a consortium of funded investigators, known as the Genetic Variation Consortium (<http://www.genome.gov/10001551>), which is striving to address many of these unanswered questions.

Much remains to be done, but the meeting at Howard University set the stage for a new era of interdisciplinary inquiry into the challenging topic of race and genetics, an era characterized by openness, freedom of scientific inquiry, an appreciation of history and a respect for differing points of views. It would be naive to portray these early steps as a breakthrough, but the committed efforts of the band of scholars and thinkers involved in these discussions are a good start in that direction. ■

ACKNOWLEDGMENTS

I thank V. Bonham for assistance in preparing this commentary.

COMPETING INTERESTS STATEMENT

The author declares that he has no competing financial interests.

Received 9 August; accepted 3 September 2004

Published online at <http://www.nature.com/naturegenetics/>

1. Angier, N. Do races differ? Not really, genes show. *The New York Times* Aug. 22, F1 (2000).
2. Marshall, E. DNA studies challenge the meaning of race. *Science* **282**, 654–655 (1998).
3. Rosenberg, N.A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
4. Sankar, P. *et al.* Genetic research and health disparities. *JAMA* **291**, 2985–2989 (2004).
5. Inoue, N. *et al.* Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* **123**, 86–91 (2002).
6. Bamshad, M., Wooding, S., Salisbury, B.A. & Stephens, J.C. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–608 (2004).
7. Cooper, R.S., Kaufman, J.S. & Ward, R. Race and genomics. *N. Engl. J. Med.* **348**, 1166–1170 (2003).
8. Burchard, E.G. *et al.* The importance of race and ethnic background in biomedical research and practice. *N. Engl. J. Med.* **348**, 1170–1175 (2003).
9. Collins, F.S., Green, E.D., Guttmacher, A.E. & Guyer, M.S. A vision for the future of genomics research: A blueprint for the genomic era. *Nature* **422**, 835–847 (2003).
10. Collins, F.S. The case for a U.S. prospective cohort study of genes and environment. *Nature* **429**, 475–477 (2004).
11. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).