

Are medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'?

Charles N Rotimi

"...with each birth and each death we alter the genetic attributes of human populations and drawing a line around an ephemeral entity like a human race is an exercise in futility and idiocy." —Pat Shipman, *The Evolution of Racism*

We now have the tools to describe the pattern of genetic variation^{1,2} across the whole genome and its relationship to the history of human origins and the differential distribution of diseases across populations and geography^{1–5}. We can begin to dissect common complex diseases^{1–5} and devise new therapeutic strategies to reduce adverse drug reactions, a key public health problem ranking between the fourth and sixth leading cause of death in the US^{6–8}. At the social level, the new genomic tools can help us to better appreciate the fluidity of social identity, including 'race', 'ethnicity' and the more complex notion of ancestry^{9–13}. Challenges surrounding the design of large-scale genotyping projects such as the international HapMap initiative and their future applications illustrate the complexities and ambiguities associated with the use of group labels in genomic research. Depending on how we use this information, the potential exists to describe simultaneously our similarities and differences without reaffirming old prejudices.

Genetic basis of common diseases and the HapMap project

Researchers have identified the genetic basis underlying several mendelian (single-gene) disorders using linkage studies in families with affected individuals. These success stories created an unrealistic expectation for the resolution power of linkage studies to unravel the genetic basis of common diseases in human populations³. In contrast to mendelian disorders, no single factor is either necessary or sufficient for describing the etiology of common diseases¹⁴. An individual's risk is the result of the complex interplay between an unknown constellation of genetic variants, environmental factors, lifestyle characteristics and some stochastic processes¹⁵. We are beginning to appreciate the complexity of common diseases, and our initial optimism, driven to some extent by candidate gene and genome-wide linkage approaches, has been tempered by the modest success rate. Notwithstanding this humbling reality, we continue to intensify our efforts, and as a result, we have achieved some success with asthma^{16–18}, cancer¹⁹, diabetes^{20–22}, Alzheimer disease²³, deep vein thrombosis²⁴, inflammatory bowel disease^{25,26}, schizophrenia²⁷ and stroke²⁸.

The international HapMap project explores patterns of DNA sequence variation in the human genome^{4,5}. Successful completion of the HapMap project will furnish scientists with powerful new tools for identifying genetic variants that contribute to common diseases and to differential drug response and for developing new diagnostic tools^{4,5,29,30}. The HapMap project is predicated on the common disease–common variant (CDCV) hypothesis, which assumes that complex diseases are influenced by genetic variants (single-nucleotide polymorphisms, SNPs) that are relatively common in human populations³⁰. If this hypothesis holds true, the HapMap project has the potential to advance the field of genetic epidemiology by facilitating association studies of candidate genes, chromosome regions or the whole genome without knowing the function of putative variants^{4,5}.

Association studies offer greater statistical power than linkage studies in examining the genetic basis of complex diseases when the risk variant is common in the population^{29,31}. The HapMap project will help develop an efficient and comprehensive catalog of common variants by contributing to our understanding of the patterns of linkage disequilibrium (LD) in multiple human populations⁴. LD is the tendency for alleles (in the case of the HapMap, SNPs) at separate sites in the genome to be found together more frequently than would be expected by chance³².

Some geneticists and statisticians have reservations regarding whether the HapMap will have sufficient resolution to be useful for understanding the genetics of common diseases across multiple populations^{14,15,33–38}. Reasons for the skepticism include what may be called a 'rush to judgment'. Some believe that the initial findings that seem to support the CDCV hypothesis may be the exception and not the rule^{14,15,33–38}. For example, some common diseases may result from the effects of multiple variants that are individually rare^{14,29,36–38}. This scenario, if true, has multiple implications for study design and analytic strategies for understanding common diseases. Because rare, as opposed to common, variants are probably differentially distributed in populations, a project like the HapMap will have to sample many more populations and develop much more detailed genetic maps than currently recommended. And some have argued that this problem may be magnified several times for African populations with more divergent patterns of genetic variation^{35,37,38}. Indeed, all bets will be off for the CDCV hypothesis and most other current analytic strategies if the manifestation of complex diseases is the result of the interacting effects of a set of rare variants (<1%) on other sets of rare (or even rarer) variants in the context of a changing environment³³.

National Human Genome Center, College of Medicine, Howard University, Washington, DC, USA. Correspondence should be addressed to C.N.R. (crotimi@howard.edu).

Published online 26 October 2004; doi:10.1038/ng1439

Ethical issues and nonmedical uses of the HapMap project

Although the HapMap project is designed to answer medical questions, it would be naive to think that results of future studies based on the HapMap data will not be applied to the emotional and volatile issues surrounding group identity and their subsequent correlation with health and social outcomes^{5,39,40}. As discussed by Morris Foster (<http://www.genome.gov/10001683>), “the proposed haplotype map project cannot be considered in isolation from the more general, ongoing discussion of the implications of using socially constructed identities in genetic research. Nor can it be considered apart from prior efforts to catalogue human genetic diversity and the controversies that surrounded them.” This is particularly relevant because some may attempt to use the HapMap data to validate old notions of ‘race’ and its correlation with multiple phenotypes (e.g., behavioral characteristics), making use of population identifiers recorded in the HapMap project.

The inclusion of population and ethnic labels in the HapMap project (Yoruba, Han Chinese, Japanese and Americans of northern and western European descent)^{4,5} is source of considerable debate⁵. Although the project used population samples rather than racial or ethnic groupings, by concentrating on common variants, the project ran the risk that this first approximation of human population structure might be subsequently used to reinforce existing racial or ethnic categories or even taken as evidence for a new categorization of human stereotypes. Why did the designers of the HapMap take this risk? Were guidelines put in place to reduce potential harm to individual participants and their communities? Could the HapMap project have been implemented without this population and ethnic information? How would this have affected the team’s goal of including the broadest spectrum of common human genetic variations?

Several rationales supported the decision to include population and ethnic labels. First, available data show differences in haplotype structure and frequency across populations (e.g., African populations tend to have shorter haplotype lengths than non-African populations do)^{35,41}. Second, having the population information will make it possible to choose the most efficient sets of SNPs for association studies. Third, removing population labels may create a false sense of protection from collective risks (e.g., stereotyping) because this information can easily be reconstructed given publicly available information including the names of the researchers and institutions involved in the project. Also, it would not be difficult to discern the identity of participating populations from previously collected data sets. Fourth, identifying the populations will allow HapMap researchers and ethicists to provide better context for interpreting the biological importance of genetic findings that are associated with particular population identities⁵.

I argue that the HapMap team of investigators selected the most appropriate design for the specific hypothesis of CDCV and have gone to great lengths to address most of these issues. For example, extensive strategies for community engagement were used to discuss potential harms and benefits. Scientists have been given descriptive guidelines for interpreting group findings and are advised to present their data in ways that avoid stigmatizing groups, conveying an impression of genetic determinism or attaching inappropriate levels of biological importance to largely social constructs such as race⁵. But the challenges associated with correct and ethical use of the HapMap results are ongoing and will probably manifest themselves in unanticipated ways.

Genetic variation and social identity

To reap the full benefits of the Human Genome Project and spin-offs like the HapMap project, we must be willing to move beyond old and simplistic interpretations of differential frequencies of disease variants

by poorly defined social proxies of genetic relatedness like ‘race’. We should allow the genome to teach us the extent of our evolutionary history without abbreviating it with preconceived notions of population boundaries and social identities. We must recognize that social identities are formed in various ways—ancestry, ethnic and tribal background, geopolitical boundaries, language, and other social and behavioral activities^{42,43}. Identities change over time and from one context to another. Their use as markers of ‘relatedness’ in genetic research without appreciation for how they were formed is likely to produce misleading information concerning the distribution of genetic variation⁷.

We all have a common birthplace somewhere in Africa^{44,45}, and this common origin is the reason why we share most of our genetic information^{46,47}. Our common history also explains why contemporary African populations have more genetic variation than younger human populations that migrated out of Africa ~100,000 years ago to populate other parts of the world, carrying with them a subset of the existing genetic information^{44,45}.

Given this shared history, why do we interpret human genetic variation data as though our differences rise to the level of subspecies? Two facts are relevant: (i) as a result of different evolutionary forces, including natural selection, there are geographical patterns of genetic variations that correspond, for the most part, to continental origin^{46–48}; and (ii) observed patterns of geographical differences in genetic information do not correspond to our notion of social identities, including ‘race’ and ‘ethnicity’^{46–48}. In this regard, no matter what categorical framework is applied, we cannot consistently use genetics to define racial groups without classifying some human populations as exceptions¹⁰. Our evolutionary history is a continuous process of combining the new with the old, and the end result is a mosaic that is modified with each birth and death. This is why the process of using genetics to define ‘race’ is like slicing soup: “You can cut wherever you want, but the soup stays mixed”⁴⁹.

How can we grasp the population structure of our species? I believe this requires universal awareness that genomic information cannot be used either to confirm or to refine old social, political and economic classifications such as ‘race’. In particular, we should understand the following points: (i) individuals in genetics studies may have membership in more than one biogeographical clusters; (ii) the borders of these clusters are not distinct; and (iii) population clusters are influenced by sampling strategies^{47,48}. For example, the inference drawn from a study with one or two African populations will probably be very different from that drawn from a study with 100 African populations sampled from north, east, west, central and south Africa. As Steve Olson observed, “Not only do all people have the same set of genes, but all groups of people also share the major variants of those genes. Geneticists have never found a genetic marker that is of one type in all the members of one large group and of a different type in all the members of another large group”⁵⁰. Furthermore, because most alleles are widespread, genetic differences among human populations are the result of gradations in allele frequencies rather than distinctive diagnostic genotypes^{46,48}.

Differential distribution of disease genes across populations

Genetic variations in human populations are distributed in a nonrandom manner¹. For example, a greater degree of genetic variation is seen in present-day Africa populations, resulting in more haplotypes, lower levels of LD, more divergent patterns of LD and more complex patterns of population substructure³⁵. As observed by Reich and collaborators, LD in a sample of Yoruba individuals from Nigeria extends only to an average distance of 5 kb, compared with 60 kb in a Mormon

population of European descent⁵¹. Similarly, Gabriel *et al.* report an average haplotype block size of 11 kb in their Yoruba and African American samples, compared with 22 kb in European and Asian samples⁵². The nonrandom pattern of genetic variation by populations has implications for mapping disease genes and for understanding how population and genomic diversity have influenced evolution, differentiation and adaptation of humans^{35,41}.

The impact of the forces of evolution, including adaptation (natural selection), on the differential distribution of disease genes is currently better understood in the context of the worldwide distribution of monogenic traits⁴¹. Gene variants that cause monogenic diseases are more common in some populations than others⁵³. But the worldwide distribution of these genetic variants does not always follow our usual definition of continental populations or social groups. For example, Tay-Sachs disease is more frequent among individuals of Ashkenazi Jewish ancestry⁵³, but people without known Jewish ancestry also have mutations in the gene responsible for the disease⁵³. Cystic fibrosis, though more common in people of European ancestry, is found in other groups, including those of African and Arab ancestry⁵³. A notable example of the signature of selection on the genome is provided by the human need to survive the mosquito-borne disease malaria⁵³. Because gene variants of hemoglobin (Hb^S and Hb^E) and glucose-6-phosphate dehydrogenase provide survival advantages for populations who lived in the area where malaria was endemic, they have been maintained at high frequency, despite the fact that they cause multiple hemopathologies^{41,53-55}.

Sickle cell anemia is a good example by which we can evaluate some consequences of ethnic labeling of genetic traits. Though more frequently observed in populations of African descent, it is found in a wide range of people including Hispanic people and inhabitants of northwestern India and areas around the Mediterranean⁵⁶. The label 'black disease', however, rendered the distribution of sickle cell anemia invisible in other populations⁵⁶, leading to erroneous understanding of the geographical distribution of the underlying genetic variants. This is one reason why many people, including physicians, are unaware that the town of Orchomenos in central Greece has a rate of sickle cell anemia that is twice that of African Americans and that black South Africans do not carry the sickle-cell trait^{56,57}.

As the following case illustrates, labeling this disease on the basis of phenotype (skin color) resulted in serious health consequences to individuals who are not phenotypically 'black' but have the relevant genetic variants. An 8-year-old boy, phenotypically European, presented with acute abdominal pain and anemia (hematocrit 0.21). Although his body temperature was only 37.9 °C, surgery was considered. A technician found red corpuscles with hemolytic characteristics on a smear. Surgery was canceled after the results of a subsequent sickle preparation were found to be positive, and the child was treated for previously undiagnosed sickle cell anemia. His parents were from Grenada and were of Indian, northern European and Mediterranean ancestry⁵⁸. This case highlights the idea that ancestry is a better indicator than 'race' or 'ethnicity' of whether one carries the markers for sickle cell anemia, Tay Sachs disease, cystic fibrosis or other genetic diseases.

The geographical distribution of genes associated with common diseases is less skewed by populations and by geographic origin than that of monogenic diseases⁵⁹. For example, the $\epsilon 4$ allele of *APOE* is found in all populations, albeit at varying frequencies^{60,61}. Those carrying a variant of the $\epsilon 4$ allele have a greater risk of developing Alzheimer disease. The frequency of the $\epsilon 4$ allele ranges from 9% in Japanese individuals to 14% in populations of European descent to 19% in African Americans⁶⁰. There is no evidence supporting the view that common functional variants are organized in discrete racial or ethnic categories. In contrast, available data show that coding sequences are conserved across popula-

tions and common polymorphisms are usually old and are therefore shared⁶². The take-home message is that variation is continuous, it is discordant with race, and the future categorization of groups for drug development and treatment will probably not correspond to our current sociopolitical group definitions.

Genetic variation: implications for drug development

Studies of genetic variation among population groups have implications for development of drugs aimed to reduce adverse reactions. As high-throughput genotyping methodologies are applied to large populations, the opportunity exists to develop genetic tests that will allow scientists and physicians to tailor medicine to individuals and to groups defined by a collection of specific genetic variants⁶³. Unfortunately, the new genomic information is being interpreted along old familiar social labels such as 'race' and 'ethnicity'. An important side effect of this phenomenon is the promotion of 'ethnicity'- or 'race'-based medicine. In recent years, there have been a flurry of newspaper articles with titles like "Shouldn't a Pill Be Color Blind?" and "Are 'racialized drugs' a marketing ploy or part of medical progress?" (ref. 64 and <http://www.bio-itworld.com/archive/files/043003.html>).

This heated debate reached a boil with the announcement of the first and only trial to test the efficacy of a drug, BiDil, in treating heart failure in a sample consisting only of African Americans (ref. 64 and <http://www.bio-itworld.com/archive/files/043003.html>). BiDil is a combination drug (two vasodilators, hydralazine and isosorbide dinitrate) designed to restore low or depleted nitric oxide levels in the blood to treat or prevent cases of congestive heart failure⁶⁵. The trial, which was cosponsored by the Association of Black Cardiologists and supported by the National Medical Association and members of the Congressional Black Caucus, was recently halted because the drug was so effective⁶⁶. The seeming success of an 'ethnic drug' and the support for this trial by leading African Americans highlights the confusion surrounding race and biology. It also illustrates the potential of genomic research to contribute inadvertently to the harmful effects of using 'race' as a variable in research. Similar thinking on the part of some physicians led to the uncritical acceptance of the decision not to treat African Americans suffering from chronic heart failure with inhibitors of angiotensin-converting enzyme⁶⁷. Despite the favorable preliminary results of the BiDil trial in African Americans, earlier concerns about the development of 'ethnicity'- or 'race'-based drugs, including the potential to exacerbate health disparity, remain valid⁶⁸⁻⁷⁰. 'Race' or 'ethnicity' is an inadequate proxy for the subset of human populations that are likely to benefit from a certain drug.

We have come full circle in biomedical research. Due to social, political and economic forces, biomedical research was almost exclusively conducted in people, especially men, of European descent. The results of such studies were then extrapolated to other groups. In the end, the BiDil story will have similar outcome; if the drug continues to be effective in the treatment of heart failure, the subset of individuals with heart failure who will benefit from the drug will not be accurately described by the label 'African American'. It is reasonable to assume that the distribution of genetic variants underlying the effectiveness of this drug will not be limited to African Americans. Moreover, it is important to note that the label used to designate the African American population in studies like the clinical trial for BiDil is too imprecise to be relevant for individual therapy. Some members of this population 'supergroup' with heart failure will benefit from this drug, and others will not. More importantly, some members of other ethnic groups will probably benefit from this drug as well. In this regard, and because of ethnic binning, a drug like BiDil, which may help other ethnic groups, could never achieve its full potential.

'Race'-based hypotheses in biomedical research sell. Reporting the nuances underlying group differences does not and, more importantly, will probably not receive the same attention in the popular press. Unfortunately, instant notoriety can be attained by reporting genetic explanations for 'racial' differences in disease, at least in North America. For example, the College of Medicine at Howard University⁷¹ made the front page of the *New York Times* and received considerable amount of air time in other venues when it was erroneously reported that the university was developing a 'black' genetic biobank to understand the genetic basis of health disparity. Although it is the intention of Howard University to facilitate the development of a biobank to house demographic, epidemiologic, clinical and genetic materials in populations of the African diaspora to study the complex interplay between environmental and genetic factors in the etiology of diseases, only the genetic component of this huge infrastructural development was reported. Given the polarized atmosphere of race relations in the US, it is easy to see why the Howard University story gained that much currency. The story was too good to resist: "Howard University, Race and Genetics"; a readymade controversy.

Conclusion

Future clinical trials may be driven by the delineation of subpopulations using DNA polymorphisms as opposed to current imprecise classifications such as 'race', 'ethnicity' or skin color⁶³. Polymorphism-based stratification of populations is expected to reduce adverse reaction to drugs and facilitate the identification of genetic variants that confer resistance or predispositions to many diseases⁶³. In this regard, and if successful, genomic data in the context of drug development may contribute to the deconstruction of 'race' and other imprecise group definitions as currently applied. Until we achieve the ultimate goal of genotyped-based medicine, however, drug developers and health-care providers will struggle with how to interpret differential drug response by groups when group definition is imprecise, fluid and time-dependent^{2,56,63}. Similarly, they will struggle with whether an individual's response to a drug or other medical interventions can be inferred from group data.

ACKNOWLEDGMENTS

I thank P. Marshall for comments and suggestions. C.N.R. is supported by grants from the National Institutes of Health (National Human Genome Research Institute; National Heart, Lung and Blood Institute; National Institute of Diabetes and Digestive and Kidney Diseases; National Center on Minority Health and Health Disparity; and National Center for Research Resource).

COMPETING INTERESTS STATEMENT

The author declares competing financial interests (see the Nature Genetics website for details).

Received 11 August; accepted 10 September 2004

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

- Bamshad, M *et al.* Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–608 (2004).
- Shriver, M.D. & Kittles, R.A. Genetic ancestry and the search for personalized genetic histories. *Nat. Rev. Genet.* **5**, 611–618 (2004).
- Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex. *Nat. Genet.* **33** Suppl, 228–237 (2003).
- The International HapMap Project. *Nature* **426**, 789–796 (2003).
- International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nat. Rev. Genet.* **5**, 467–475 (2004).
- Lazarou, J., Pomeranz, B.H. & Corey, P.N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* **15**, 200–205 (1998).
- Nebert, D.W. & Menon, A.G. Pharmacogenomics, ethnicity, and susceptibility genes. *Pharmacogenomics J.* **1**, 19–22 (2001).
- Deloukas P, Bentley D. The HapMap project and its application to genetic studies of drug response. *Pharmacogenomics J.* **4**, 88–90 (2004).
- Sankar, P. & Cho, M.K. Genetics. Toward a new vocabulary of human genetic variation. *Science* **298**, 1337–1338 (2002).
- Risch, N., Burchard, E., Ziv, E. & Tang, H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* **1**, 3 (2002).
- Ellison, G.T.H. & De Wet, T. 'Race', ethnicity and the psychopathology of social identity. In *Psychopathology and Social Prejudice* (eds. Hook, D. & Eagle, G.) 130–149 (Cape Town, Juta, 2002).
- Ellison, G.T.H. & Jones, I.R. Social identities and the 'new genetics': Scientific and social consequences. *Critical Public Health* **12**, 265–282 (2002).
- Foster, M.W. & Sharp, R.R. Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity. *Genome Res.* **12**, 44–50 (2002).
- Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
- Pritchard, J.K. & Cox, N.J. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum. Mol. Genet.* **11**, 2417–2423 (2002).
- Bell, J. Predicting disease using genomics. *Nature USA* **429**, 453–456 (2004).
- Van Eerdedewegh, P. *et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* **25**, 426–430 (2002).
- Allen, M. *et al.* Positional cloning of a novel gene influencing asthma from chromosome 2q14. *Nat. Genet.* **35**, 258–263 (2003).
- van de Vijver, M.J. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **25**, 347 (2002).
- Dorman, J.S., LaPorte, R.E., Stone, R.A. & Trucco, M. Worldwide differences in the incidence of type I diabetes are associated with amino acid variation at position 57 of the HLA-DQ beta chain. *Proc. Natl. Acad. Sci. USA* **87**, 7370–7374 (1990).
- Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
- Bell, G.I., Horita, S. & Karam, J.H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33**, 176–183 (1984).
- Strittmatter, W.J. & Roses, A.D. Apolipoprotein E and Alzheimer's disease. *Annu. Rev. Neurosci.* **19**, 53–77 (1996).
- Dahlback, B. Resistance to activated protein C due to factor V R506Q mutation as a cause of venous thrombosis. *Rev. Invest. Clin.* **49** Suppl, 3–5 (1997).
- Hugot, J.P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
- Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
- Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* **71**, 877–892 (2002).
- Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
- Carlson, C.S., Eberle, M.A., Kruglyak, L. & Nickerson, D.A. Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446–452 (2004).
- Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**, 29–32 (2001).
- Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).
- Goldstein, D.B. & Weale, M.E. Population genomics: linkage disequilibrium holds the key. *Curr. Biol.* **11**, 576–579 (2001).
- Clark, A.G. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. *Curr. Opin. Genet. Dev.* **13**, 296–302 (2003).
- Cardon, L.R. & Abecasis, G.R. Using haplotype blocks to map human complex trait loci. *Trends Genet.* **19**, 135–140 (2003).
- Tishkoff, S.A. & Williams, S.M. Genetic analysis of African populations: human evolution and complex disease. *Nat. Rev. Genet.* **3**, 611–621 (2002).
- Weiss, K.M. & Clarke, A.G. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* **18**, 19–24 (2002).
- Tishkoff, S.A. & Verrelli, B.C. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293–340 (2003).
- Kittles, R.A. & Weiss, K.M. Race, ancestry, and genes: implications for defining disease risk. *Annu. Rev. Genomics* **4**, 33–67 (2003).
- King, P.A. *Gene Mapping: Using Law and Ethics As Guides* (eds. Annas, G.J. & Elias, S.) 94–111 (Oxford University Press, Oxford, 1992).
- Clayton, E.W. The complex relationship of genetics, groups, and health: what it means for public health. *J. Law Med. Ethics* **30**, 290–297 (2002).
- Tishkoff, S.A. *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
- Smedley, A. *Race in North America: Origin and Evolution of a Worldview* 2nd edn (Westview, Boulder, 1999).
- Juengst, E. Group identity and human diversity: Keeping biology straight from culture. *Am. J. Hum. Genet.* **63**, 673–677 (1998).
- Yu, N., Fu, Y.X. & Li, W.H. DNA polymorphism in a worldwide sample of human X chromosomes. *Mol. Biol. Evol.* **19**, 2131–2141 (2002).
- Cann, R.L., Stoneking, M. & Wilson, A.C. Mitochondrial DNA and human evolution. *Nature* **3**, 31–36 (1987).
- Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519 (1997).
- Rosenberg, N.A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
- Bamshad, M.J. *et al.* Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**, 578–589 (2003).

49. Anonymous. Slicing soup. *Nat. Biotechnol.* **20**, 637 (2002).
50. Olson, S. The genetic archaeology of race. *The Atlantic Monthly* April (2001).
51. Reich, D.E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 99–204 (2001).
52. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **21**, 2225–2229 (2002).
53. Molnar, S. *Human Variation: Races, Types and Ethnic Groups* 5th edn. (Prentice Hall, New Jersey, 2001).
54. Luzzatto, L. & Metha, A. Glucose-6-phosphate dehydrogenase deficiency. in *The Metabolic Basis of Inherited Disease* 5th edn. (eds. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle D.) 3367–3398 (McGraw-Hill, New York, 1995).
55. Beutler, E., Vulliamy, T. & Luzzatto, L. Hematologically important mutations: glucose-6-phosphate dehydrogenase. *Blood Cells Mol. Dis.* **22**, 49–56 (1996).
56. Braun, L. Race, ethnicity, and health: can genetics explain disparities? *Perspect. Biol. Med.* **45**, 159–174 (2002).
57. Kevles, D.J. In the Name of Eugenics: Genetics and the Uses of Human Heredity (Harvard University Press, Cambridge, 1995).
58. Witzig, R. The medicalization of race: scientific legitimization of a flawed social construct. *Ann. Intern. Med.* **125**, 675–679 (1996).
59. Cooper, R.S., Kaufman, J.S. & Ward, R. Race and genomics. *N. Engl. J. Med.* **348**, 1166–1170 (2003).
60. Farrer, L.A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
61. Burchard, E.G. *et al.* The importance of race and ethnic background in biochemical research and clinical practice. *N. Engl. J. Med.* **348**, 1170–1175 (2003).
62. Halusha, M.K. *et al.* Patterns of single nucleotide polymorphisms in candidate genes for blood pressure homeostasis. *Nat. Genet.* **22**, 239–247 (1999).
63. Wilson, J.F. *et al.* Population genetic structure of variable drug response. *Nat. Genet.* **29**, 265–269 (2001).
64. Duster, T. Unlikely mix — Race, biology and drugs. *The San Francisco Chronicle* (2003).
65. Taylor, A.L., Cohn, J.N., Worcel, M., Franciosa, J.A. & A-HeFT Investigators. The African-American Heart Failure Trial: background, rationale and significance. *J. Natl. Med. Assoc.* **94**, 762–769 (2002).
66. Pollack, A. Drug approved for heart failure in black patients. *New York Times* (2004).
67. Anonymous. Genes, drugs and race. *Nat. Genet.* **29**, 239–240 (2003).
68. Cooper, R.S. & Psaty, B.M. Genomics and medicine: distraction, incremental progress, or the dawn of a new age? *Ann. Intern. Med.* **138**, 576–580 (2003).
69. Kahn, J. Getting the numbers right: statistical mischief and racial profiling in heart failure research. *Perspect. Biol. Med.* **46**, 473–483 (2003).
70. Kahn, J. How a drug becomes “ethnic”: Law, Commerce and the production of racial categories in medicine. *Yale J. Health Policy Law Ethics* **VI**:1, 1–46 (2004).
71. Pollack, A. DNA of blacks to be gathered to fight illness. *New York Times* (2003).