



Assessing genetic contributions to phenotypic differences among 'racial' and 'ethnic' groups

Joanna L Mountain^{1,2} & Neil Risch^{2,3}

Descriptions of human genetic variation given thirty years ago have held up well, considering the substantial accrual of DNA sequence data in the interim. Most importantly, estimates of between-group genetic variation have remained relatively low. Despite the low average level of between-group variation, clusters recently inferred from multilocus genetic data coincide closely with groups defined by self-identified race or continental ancestry. This correspondence implies that genetic factors might contribute to unexplained between-group phenotypic variation. Current understanding of the contribution of genes to variation in most complex traits is limited, however. Under these circumstances, assumptions about genetic contributions to group differences are unfounded. In the absence of detailed understanding, 'racial' and 'ethnic' categories will remain useful in biomedical research. Further, we suggest approaches and guidelines for assessing the contribution of genetic factors to between-group phenotypic differences, including studies of candidate genes and analyses of recently admixed populations.

Thirty years ago, a burning question for both scientists and the general public was the degree to which genes influence between-group variation in phenotypic traits, particularly behavioral traits such as IQ¹. The question of whether mean differences in IQ test scores had the same basis as differences in skin color, hair form and facial features (understood to be heavily influenced by genes) was at the forefront. At that time, scholars found much to agree on regarding genetic and phenotypic variation. Traits (*e.g.*, blood groups and enzyme polymorphisms) close to the DNA level indicated that there was a great deal of within-group variation, with much overlap between 'races' (as defined at the time)². Although studies had already shown the geographic pattern of human genetic variation³, the proportion of variation between groups was estimated to be low, between 0.06 and 0.15 depending on the groups considered⁴. For traits often associated with 'racial' or 'ethnic' categorization (*e.g.*, skin pigmentation, facial features and hair form), there was little information regarding the role of genes⁵. Although there were rough estimates of the number of genes underlying

variation in skin color (at least four; *ref.* 1), no specific genes were known. In contrast, geneticists were well aware of the genetic basis for a number of mendelian diseases, where between-group differences in incidence reflected differences in allele frequency⁶. They were also aware of between-group differences in incidence of more complex traits, traits without demonstrated genetic components even for the within-group variation⁶. In addition, there was agreement that for behavioral traits, little was known regarding a role for genes. Physical anthropologists and geneticists 30 years ago took the position that although we couldn't rule out a genetic basis for between-group variation in behavioral traits, we had no evidence indicating such a genetic basis but did have evidence for a role for nongenetic factors^{1,5}.

Today, the question of whether there is a biological or genetic basis to 'race' or 'racial' differences is once again in the popular and scientific press, indicative of public interest in the basis for between-group variation, especially for medical conditions and behavioral traits⁷⁻⁹. Two recent books have argued strongly for a genetic basis for what have been called 'race differences'^{10,11}. As writer Steve Olson notes, "Many people continue to believe that human groups have fundamental biological differences. They believe that outward variations in skin color, facial features, or body shape reflect more consequential differences of character, temperament, or intelligence."¹² Several authors, including Olson, have made the opposite argument, stating that there is no biological basis or meaning to race differences^{7,9,12,13}.

The controversy stems, at least in part, from the possibility that attitudes are influenced by whether people believe 'race' is a biological or social construct. Indeed, research has shown that attitudes differ depending on whether individuals conceive of 'race' as biological or social¹⁴. Many social scientists have suggested that by linking 'racial' or 'ethnic' categories to biology (especially genetics), one reinforces those categories and thereby influences attitudes and behavior¹⁵. For these reasons, authors have argued against assuming a genetic basis without evidence⁵.

In this discussion, we define a social category or group as one determined by social factors; an individual is associated with such a category (or categories) based on a set of socially negotiated criteria. Given this definition, 'race' and 'ethnicity' are social categories, even though some inclusion criteria may be biological. The extent to which 'racial' or 'ethnic' labels correlate with biological traits varies through time and across the labels themselves. Much of that correlation is the outcome of human population history, tied closely to geography¹⁶. Here, unless otherwise indicated, we use the term 'group' to refer to a group of individuals associated with a 'racial' or 'ethnic' label.

¹Department of Anthropological Sciences, Stanford University, Stanford, California 94305-2117, USA. ²Department of Genetics, Stanford Medical School, Stanford University, Stanford, California 94305-5120, USA. ³Division of Research, Kaiser Permanente, Oakland, California 94611-5714. Correspondence should be addressed to J.L.M. (mountain@stanford.edu).

Published online 26 October 2004; doi:10.1038/ng1456

Has the scientific community made any progress since the early 1970s towards addressing the question of whether there is a genetic basis for between-group phenotypic differences? Some have argued that the sequencing of the human genome and related research have provided evidence that the notion of race is biologically meaningless and therefore useless^{17,18}. Others, however, have made the opposite argument, namely that recent studies show that genetic clusters correspond closely with groups defined by self-identified race or ethnicity or by continental ancestry¹⁹.

In light of these contradictory statements, we consider here current scientific understanding relevant to the question of the degree to which phenotypic variation has a genetic basis. We first note a distinction between the terms 'genetic' and 'biological'. Genetic (*i.e.*, DNA) factors can be considered a subset of biological factors, whereas biological, including physiological and morphological, factors reflect chance occurrences during development and the impact of environment (*e.g.*, diet or training) as well as DNA sequence. We next consider three approaches to the study of the basis for between-group phenotypic differences (Table 1).

Inference from familial correlations

Study designs from genetic epidemiology, including family, twin and adoption studies, have been used extensively to estimate the proportion of trait variance that is due to genetic factors, commonly known as heritability. Heritability estimates, however, are far from absolute and depend not only on the population studied but also on the time and location of study. Any differences in allele frequency between populations or differences in relevant environmental contributions will alter heritability estimates. In addition, simplified models often lead to overestimates of heritability²⁰. Nonetheless, an estimate can provide some insight into genetic contributions to trait variation within groups.

Does a heritability estimate inform us of the extent to which a genetic explanation underlies a group difference in a trait? The answer depends on the nature of that trait. For most complex disease and behavioral traits of interest, heritability is uninformative regarding group differences. Such traits typically result from the interaction between genetic and environmental contributors. A trait that seems to be purely genetic (heritability of 1.0) in one group may also have a group difference that is entirely environmental because an environmental factor differentiates the two groups. For example, studies have shown that without musical training before age six, few individuals will develop absolute pitch (also called perfect pitch). With musical training before age six, some individuals will develop absolute pitch, and whether they do so seems to be strongly familial and possibly

genetic^{21,22}. Thus, if two groups (*e.g.*, racial or ethnic groups) have, on average, different experiences with early musical training, a large difference in prevalence of absolute pitch will occur (Fig. 1a). Trait variation will seem to have a genetic contribution (at least in a group with early musical training), even when the group difference is largely environmental. Such a pattern has been observed in one study where individuals of East Asian ancestry were three to four times more likely to have absolute pitch than individuals of European ancestry²².

The opposite pattern is also possible. When the allele that influences the trait segregates in only one group, and the variation in the other group reflects only environmental variation, the trait will appear to have 0 heritability (in the latter group) even though the group difference is explained fully by genetic variation. We illustrate this pattern with the case of taste sensitivity to phenylthiocarbamide (Fig. 1b). A low-sensitivity haplotype was observed at frequency 0.50 in a European sample but was not observed at all in a Native American sample. Genes are responsible for the difference in the distribution of the trait, even though heritability is low in one of the populations.

Inference from DNA analysis

Inferences of the role of genes from familial correlations, even using unusual and informative relationships such as twins and adoptees, are indirect because they are not based on the identification of specific DNA regions that contribute to trait variation. Identification of genes influencing trait variation can be more informative. Many genetic studies today are carried out with reference to a particular population to minimize the influence of genetic or environmental heterogeneity. When a gene (or genes) that influences a trait has been identified, it can then be studied in a broader set of populations, including racially or ethnically defined groups. One might even estimate the between-group difference in trait frequency (in the case of a categorical trait) or in average trait value (otherwise) due to that gene. Such an estimate will be inaccurate, however, unless the effect of the alleles at this locus is the same in the two groups being compared. Because of gene-environment or gene-gene interactions, differences between the two groups in environmental conditions or in genetic background can lead to different impacts of the locus under consideration (*e.g.*, the *APOE* ϵ 4 allele and Alzheimer disease²³).

Inference from admixture analysis

One approach that has the potential to further our understanding of group differences in the distribution of a trait is admixture analysis. When individuals with ancestry from geographically disparate regions form a new population (*e.g.*, African Americans, Mexican Americans), it may be possible to correlate trait values with degree of admixture

Table 1 Three methods for assessing the genetic component to within- and between-group variation

Method	Approach	Strengths	Weaknesses
Familial correlations	Examine similarity of trait value for sets of family members, typically including twins, half siblings or adoptees.	Can provide estimates of heritability within groups.	Provides limited information regarding any genetic contribution to between-group differences.
DNA analysis	Examine specific DNA variations and associated risks both within and between groups.	Can provide estimate of genetic contribution to between-group difference due to specific genes.	Depends on accurate within-group estimates of risk; does not provide a global genomic estimate.
Admixture analysis	Regress trait value (or incidence) on individual admixture estimates in a recently admixed population.	Can provide global genomic estimate of genetic contribution to between-group differences.	Highly susceptible to confounding due to correlation of individual admixture with nongenetic factors. Inference strengthened when individual admixture is cryptic or specific genomic regions are identified.

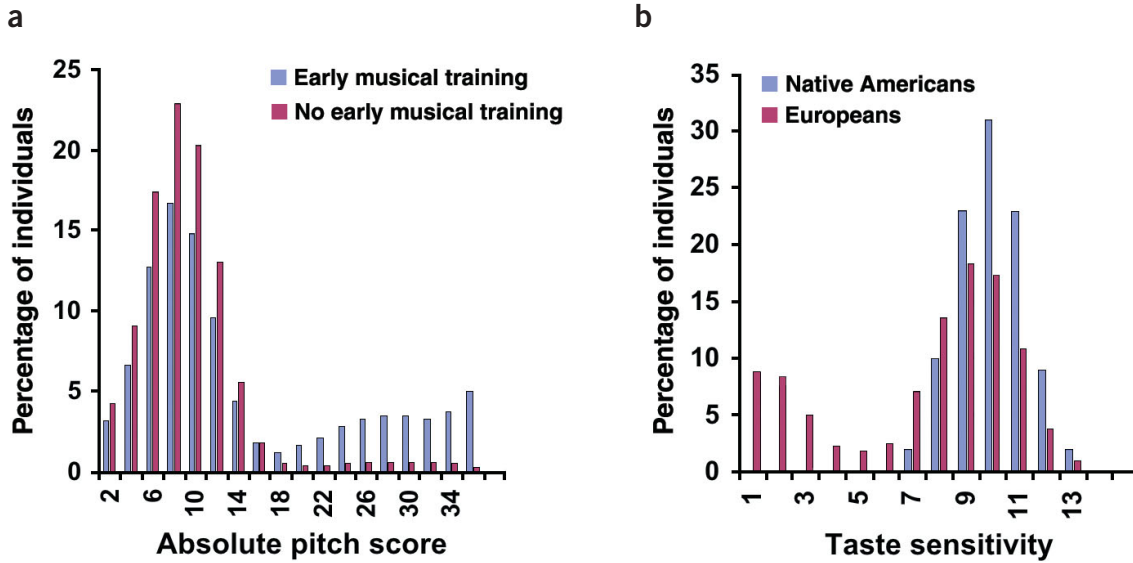


Figure 1 Two examples of group differences influenced by environmental or genetic factors. (a) Absolute pitch in individuals with early musical training (before age 6) versus those without early musical training. Absolute pitch manifests primarily in the group that received early musical training, is familial and may have a genetic component in the presence of this environmental factor²². This potential genetic influence does not manifest in the absence of early musical training, and therefore the heritability of absolute pitch in a group with no early musical training would presumably be low. The difference in distribution is primarily environmental. Data derived from ref. 21. (b) Taste sensitivity to phenylthiocarbamide in Europeans versus Native Americans. The phenylthiocarbamide tasting gene (*TAS2R38*, also called *PTC*) on chromosome 7 is polymorphic in a European population sample, in which the low-sensitivity haplotype has a frequency of 0.50 and homozygotes are at frequency of 0.25. This haplotype is absent in a Native American population sample, in whom the heritability is presumably low. In this case, the difference in distribution is genetic. Data derived from ref. 56.

(i.e., individual admixture) from the contributing ancestral populations. A positive correlation in the direction of the original group difference provides support for some genetic contribution to the between-group variation. If individual admixture correlates with other, unmeasured (or omitted) environmental, social, cultural or behavioral traits, however, as is often the case, then the observed trait correlation with individual admixture will be confounded, and a genetic interpretation of the correlation will be unreliable. Although inclusion of relevant environmental factors in the regression analysis is helpful, it may still be impossible to know that all relevant environmental factors have been considered.

There are two ways a genetic conclusion from admixture analysis can be more robust. First, if the individual admixture of individuals in

the population is cryptic (not discernible by themselves or others), then residual correlation of individual admixture with social or other environmental agents is less likely. Second, correlation of individual admixture at specific locations in the genome with the trait is strong evidence for a genetic role. This method is the basis of what has been called mapping by admixture disequilibrium. Environmental confounding is unlikely to contribute to a positive regression of the trait at a single or small number of genomic locations.

Inference depends on the nature of the trait

Assessment of genetic contributions to between-group differences depends heavily on the nature of the trait under consideration. Traits that are closely related to DNA sequence variation may show different

Table 2 Current understanding of genetic contribution to group differences for five different categories of traits

Trait category	Current understanding
DNA sequence variation, including SNPs and STRs	Very detailed. Large-scale studies of DNA variants reveal average F_{ST} values of 0.05–0.15 and average allele frequency difference of 0.12–0.15 between continental groups (global comparison) and racial groups (US comparison). Differences in allele frequency vary in that some differences are close to zero while others are much higher than 0.15.
Mendelian diseases	Detailed. Most mutations are rare and are specific to regions (e.g., continents) or groups (e.g., races or ethnicities), reflecting the timing of mutations and human population history, including founder effects, genetic drift more generally and, in some cases, natural selection.
Physical traits (skin, hair and eye color, hair form, body shape)	Modest. Presumed polygenic with few genes (e.g., <i>MC1R</i>) identified.
Complex diseases	Limited. Partial evidence for a few diseases (e.g., HLA-associated diseases, Crohn disease, Alzheimer disease) where some variants are group-specific and others are more broadly distributed, possibly at different frequencies in different groups.
Behavior	Very limited. Based only on within-group heritability studies.

patterns, and group differences may be more easily resolved, than is true for those traits more remotely related to sequence variation. In a classic paper, Penrose gave the example of phenylketonuria where traits closely related to the actual genetic variant (such as the DNA sequence itself, electrophoretic mobility and serum phenylalanine levels) showed large differences between homozygotes and nonhomozygotes, whereas other traits (e.g., hair color) had much smaller differences²⁴. For the latter traits, many factors other than the gene underlying phenylketonuria also contribute to variation. In discussing how traits have been studied, and the inferences about group differences that have resulted, we therefore consider five sets of traits: (i) DNA-level traits, where genotype is equivalent to phenotype; (ii) mendelian traits; (iii) physical traits often used to categorize people in terms of race or ethnicity; (iv) complex diseases; and (v) other complex phenotypes including behavior, cognitive abilities and temperament (Table 2).

DNA-level traits

Overall, our understanding of the pattern of human genetic variation is quite consistent with that of authors three decades ago. We now know details of the distribution of F_{ST} (the proportion of genetic variation between groups) across thousands of DNA markers. Recent estimates of average F_{ST} are similar to early estimates²⁵, with exceptions such as the human Y chromosome and mitochondrial DNA (somewhat higher F_{ST} values) and sets of STR loci (relatively low F_{ST} values)^{26–28}. A low average F_{ST} for neutral genetic markers suggests that the power to define phenotypic differences in racial and ethnic categories is not typical of a single neutrally evolving locus. On the other hand, an F_{ST} estimate of 0.10–0.15 does not rule out a genetic basis for phenotypic differences between groups. In fact, a low F_{ST} estimate implies little about the degree to which genes contribute to between-group differences. Small differences in allele frequency indicated by a low average F_{ST} are still sufficient for generating clusters of individuals that correspond closely with groups defined by continental ancestry or self-identified race or ethnicity^{19,27,29–31}. The joint effect of small differences in allele frequencies at multiple neutral genetic markers provides the power to cluster individuals^{19,32}.

Mendelian traits

The scientific community knows a great deal more about the role of genes in disease, especially mendelian traits, today than it did in the early 1970s. More than 1,000 traits have been mapped to specific chromosomal regions in the last 20 years or so³³. Among these traits are a number of drug-metabolizing enzyme activities^{34,35}. These enzyme activities have relatively high differentiation among populations, possibly because of the role of the enzymes in detoxifying compounds in the diet^{36,37}. Also, the extent to which genes underlying ‘simple’ traits have variable penetrance continues to be apparent; the correspondence between genotype and phenotype is far from simple.

As was true thirty years ago, loci with unusually high or low F_{ST} values receive a great deal of attention because such loci may be under selection^{38–40}. Some of those markers are known to be associated with differences in disease incidence between groups (e.g., Duffy locus⁴¹). In general, however, estimates of F_{ST} are of little clinical utility in characterizing racial or ethnic differences. For example, numerous disease or resistance mutations (many associated with mendelian diseases and others not) are found almost exclusively in a single population (e.g., in individuals with European ancestry: hemochromatosis, *HFE* C282Y; HIV, *CCR5* Δ 32; and Crohn disease, three *CARD15* mutations)^{42–44}. Yet these variants have low F_{ST} values because there is little absolute variation in frequency across racial groups (F_{ST} typically less than 0.05 in these cases)¹⁹.

Physical traits

Physical traits (e.g., skin pigmentation, hair form and facial features) are sometimes used to categorize people according to race or ethnicity: We know a little more regarding the genetic basis underlying these traits now than we did in 1970, but not much in the case of humans. The gene *MC1R* accounts for a small proportion of the variation in human pigmentation⁴⁵. Much more has been learned about mouse coat pigmentation. As Jackson⁴⁶ notes, “Despite decades of research on coat-color genes, mutagenic screens continue to identify new dominant mutations that affect pigmentation.” Given the mouse data, it seems likely that more than four genes (a minimum estimate 30 years ago) contribute to variation in human pigmentation. Even though these physical traits may have large racial differences, we cannot at this time conclude that such differences reflect radically different allele frequencies at one or a few genes. It remains plausible that many genes contribute to these traits and to the differences among populations.

Complex diseases

Some progress has been made with respect to a number of diseases, including inflammatory bowel disease and neurodegenerative diseases^{47,48}, but inference of genes influencing complex traits such as diabetes, blood pressure, heart disease, psychiatric disorders and glaucoma remains a challenge⁴⁹. Reports of linkage of these complex diseases to specific genetic regions are rarely replicated. Therefore, the genetic basis for any between-group variation remains undetermined in most cases. The greater the environmental contribution, the more genetic factors and the more complex the interactions, the more difficult it is to characterize group differences in terms of genetic and environmental contributions.

Behavioral traits

Little progress has been made towards elucidating a genetic basis for any variation in traits such as cognitive ability, temperament and athletic ability. Many studies have made claims regarding the role of specific genes, but few of these have been replicated. The *DRD4* polymorphism is a good example. Although initial reports claimed that the polymorphism had a role in variation of ‘novelty-seeking’^{50,51}, a recent meta-analysis failed to find consistent support⁵². This category is a more extreme version of the complex-disease category, possibly having even more important social implications.

Why so little progress?

For complex diseases, and particularly for behavioral traits, little progress has been made towards understanding any genetic basis for within- or between-group variation. Does the lack of progress reflect a lack of research? Complex diseases have certainly received a great deal of attention and funding. Authors of the early 1970s felt that there were good reasons not to spend money on research into the genetic basis of between-group variation of cognitive traits¹. Meta-analyses of studies of traits such as novelty-seeking, however, suggest that the lack of progress is not simply a matter of a lack of research.

Under relatively simple models, small differences in allele frequency can lead to relatively large differences in trait frequencies. A trait influenced by several genes that interact with one another may be rare in one group and far less rare in another. Similar models involving interactions between multiple nongenetic factors might also be explanatory. Such models might account for different frequencies of individuals at the extremes of phenotypic distributions (e.g., marathon success of Kenyan runners⁸). If a model of interactions among many genes and nongenetic factors best fits the data, however, detecting any specific genetic contribution to trait variation will be

challenging. Furthermore, our ability to predict phenotype accurately (even given genotypes for a subset of the genes) is much lower than for more simply inherited traits. Much larger studies are needed to measure the effects of genes and the interactions among them.

The relationship between differences in allele frequency and group-level phenotypic differences need not be consistent. We anticipate that in the future all possible relationships will be represented. There may even be cases where there are no reported between-group phenotypic differences but contributing alleles differ in frequency between groups.

Conclusions

During the last three decades, the scientific community has generated a large volume of data regarding human genetic and phenotypic variation. Recent analyses of multilocus genetic data have generated clusters that correspond closely with socially or geographically defined groups. There is the potential, therefore, for genetic factors to contribute to phenotypic differences between groups. But the skepticism of some scientists of the early 1970s regarding our ability to find genes (if such genes exist) underlying group differences in behavioral traits such as intelligence seems warranted. We are far from characterizing the contribution of genes to between-group variation of any complex trait and are likely to continue struggling in the future. The current lack of detailed understanding has two implications.

First, there is a need for stringent criteria for drawing conclusions regarding the contribution of genes to between-group differences. Generalizations and assumptions are unwarranted and may exacerbate group disparities. We therefore advocate standards for statements regarding genetic contributions to between-group differences. With that goal in mind, we discussed promising approaches that take advantage of identified candidate genes or recent population admixture.

Second, our lack of understanding of the etiology of many complex traits means that racial and ethnic labels remain useful in epidemiological and clinical settings. Given the potential for furthering racism by discussing race and genetics together in a scientific context, we might seek to eliminate the use of racial categories in these contexts. Given current health disparities, however, and assuming that our society values the goal of understanding the underlying basis of those disparities, the continued use of labels in epidemiological research and clinical practice seems justified^{19,53}. Racial and ethnic categories are proxies for a wide range of factors, potentially genetic and nongenetic. Many nongenetic correlates of race and ethnicity have been identified. Other factors will probably be elucidated through future epidemiological and genetic research. Once these factors are known, we may no longer need to take those categories into consideration. Until then, however, an individual's racial or ethnic affiliation will continue to be valuable information in the context of epidemiological research⁵⁴.

The key to successful use of information about group differences in disease incidence (health disparities) is accurate incidence statistics with as much ancestry detail as possible. For now, racial or ethnic categories may be the best we can do in some cases. Not all racial or ethnic labels are equally useful in an epidemiological or clinical context. The labels are useful to the extent that they predict phenotypic differences and are particularly useful if they are predictive over and above any potentially correlated factors, such as education level or health care access. Self-identified race and continental ancestry often have relatively high predictive value. Self-identified ethnicity typically has somewhat less predictive power.

In the future we can strive for more detailed databases, acknowledging missing data regarding the ancestry of many individuals, the fluidity of racial and ethnic categories and increasing rates of intermarriage⁵⁵. As always, the most accurate diagnoses and predictions

(including recommendations for treatment) will come from an individualized analysis of genetic, nongenetic and additional unspecified (in terms of the role of genes) factors. Racial or ethnic categorization will continue to be useful as long as such categorization 'explains' variation unexplained by other factors.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 3 September; accepted 23 September 2004

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

1. Bodmer, W.F. & Cavalli-Sforza, L.L. Intelligence and race. *Sci. Am.* **223**, 19–29 (1970).
2. Dobzhansky, T. Race equality. in *The Biological and Social Meaning of Race* (ed. Osborne, R.) 14–24 (Freeman, San Francisco, 1971).
3. Cavalli-Sforza, L.L. & Bodmer, W.F. *The Genetics of Human Populations* (Freeman, San Francisco, 1971).
4. Lewontin, R. The apportionment of human diversity. in *Evolutionary Biology* vol. 6 (eds. Hecht, M.K. & Steere, W.S.) (Plenum, New York, 1972).
5. Howells, W.W. The meaning of race. in *The Biological and Social Meaning of Race* (ed. Osborne, R.) (Freeman, San Francisco, 1971).
6. McKusick, V.A. The ethnic distribution of disease in the United States. *J. Chronic Dis.* **20**, 115–118 (1967).
7. Graves, J. *Race Myth: Why We Pretend Race Exists in America* (Dutton, New York, 2004).
8. Holden, C. Peering under the hood of Africa's runners. *Science* **305**, 637–639 (2004).
9. Schwartz, R.S. Racial profiling in medical research. *N. Engl. J. Med.* **344**, 1392–1393 (2001).
10. Entine, J. & Smith, E. *Taboo: Why Black Athletes Dominate Sports and Why We're Afraid to Talk About It* (Public Affairs, New York, 1999).
11. Sarich, V. & Miele, F. *Race: The Reality of Human Differences* (Westview, Boulder, 2004).
12. Olson, S. *Mapping Human History* (Mariner Books, Boston, 2002).
13. Long, J.C. & Kittles, R.A. Human genetic diversity and the nonexistence of biological races. *Hum. Biol.* **75**, 449–471 (2003).
14. Parrott, R.L., Silk, K.J. & Condit, C. Diversity in lay perceptions of the sources of human traits: genes, environments, and personal behaviors. *Soc. Sci. Med.* **56**, 1099–1109 (2003).
15. Sankar, P. *et al.* Genetic research and health disparities. *JAMA* **291**, 2985–2989 (2004).
16. Bamshad, M., Wooding, S., Salisbury, B.A. & Stephens, J.C. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–609 (2004).
17. Lander, E.S. Cracking the code of life. (WGBH, Boston, 2001).
18. Haga, S.B. & Venter, J.C. Genetics. FDA races in the wrong direction. *Science* **301**, 466 (2003).
19. Risch, N., Burchard, E., Ziv, E. & Tang, H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* **3**, comment2007 (2002).
20. Feldman, M.W. & Lewontin, R.C. The heritability hang-up. *Science* **190**, 1163–1168 (1975).
21. Baharloo, S., Johnston, P.A., Service, S.K., Gitschier, J. & Freimer, N.B. Absolute pitch: an approach for identification of genetic and nongenetic components. *Am. J. Hum. Genet.* **62**, 224–231 (1998).
22. Gregersen, P.K., Kowalsky, E., Kohn, N. & Marvin, E.W. Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component. *Am. J. Hum. Genet.* **65**, 911–913 (1999).
23. 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).
24. Penrose, L.S. Measurement of pleiotropic effects in phenylketonuria. *Ann. Eugen.* **16**, 134–141 (1951).
25. Barbujani, G., Magagnoli, A., Minch, E. & Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519 (1997).
26. Seielstad, M.T., Minch, E. & Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280 (1998).
27. Rosenberg, N.A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
28. Jorde, L.B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
29. Bowcock, A.M. *et al.* High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457 (1994).
30. Mountain, J.L. & Cavalli-Sforza, L.L. Multilocus genotypes, a tree of individuals, and human evolutionary history. *Am. J. Hum. Genet.* **61**, 705–718 (1997).
31. Bamshad, M.J. *et al.* Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**, 578–589 (2003).
32. Edwards, A.W. Human genetic diversity: Lewontin's fallacy. *Bioessays* **25**, 798–801 (2003).
33. McKusick-Nathans. Online Mendelian Inheritance in Man, OMIM (TM). vol. 2004 (Johns Hopkins University, Baltimore, 2000).

34. Evans, W.E. & Relling, M.V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487–491 (1999).
35. Evans, W.E. & Relling, M.V. Moving towards individualized medicine with pharmacogenomics. *Nature* **429**, 464–468 (2004).
36. Wilson, J.F. *et al.* Population genetic structure of variable drug response. *Nat. Genet.* **29**, 265–269 (2001).
37. Goldstein, D.B. & Chikhi, L. Human migrations and population structure: what we know and why it matters. *Annu. Rev. Genomics Hum. Genet.* **3**, 129–152 (2002).
38. Beaumont, M.A. & Nichols, R.A. Evaluating loci for use in the genetic analysis of population structure. *Proc. R Soc. Lond. B Biol. Sci.* **263**, 1619–1626 (1996).
39. Akey, J.M., Zhang, G., Zhang, K., Jin, L. & Shriver, M.D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
40. Bowcock, A.M. *et al.* Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* **88**, 839–843 (1991).
41. Hamblin, M.T., Thompson, E.E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70**, 369–383 (2002).
42. Croucher, P.J. *et al.* Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *Eur. J. Hum. Genet.* **11**, 6–16 (2003).
43. Limdi, J.K. & Crampton, J.R. Hereditary haemochromatosis. *Q. J. Med.* **97**, 315–324 (2004).
44. Su, B. *et al.* Distribution of three HIV-1 resistance-conferring polymorphisms (SDF1-3'A, CCR2-641, and CCR5-delta32) in global populations. *Eur. J. Hum. Genet.* **8**, 975–979 (2000).
45. Harding, R.M. *et al.* Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351–1361 (2000).
46. Fitch, K.R. *et al.* Genetics of dark skin in mice. *Genes Dev.* **17**, 214–228 (2003).
47. Kennedy, J.L., Farrer, L.A., Andreasen, N.C., Mayeux, R. & St George-Hyslop, P. The genetics of adult-onset neuropsychiatric disease: complexities and conundra? *Science* **302**, 822–826 (2003).
48. Brant, S.R. & Shugart, Y.Y. Inflammatory bowel disease gene hunting by linkage analysis: rationale, methodology, and present status of the field. *Inflamm. Bowel Dis.* **10**, 300–311 (2004).
49. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33** Suppl, 228–237 (2003).
50. Benjamin, J. *et al.* Population and familial association between the D4 dopamine receptor gene and measures of Novelty Seeking. *Nat. Genet.* **12**, 81–4 (1996).
51. Ebstein, R.P. *et al.* Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of Novelty Seeking. *Nat. Genet.* **12**, 78–80 (1996).
52. Kluger, A.N., Siegfried, Z. & Ebstein, R.P. A meta-analysis of the association between DRD4 polymorphism and novelty seeking. *Mol. Psychiatry* **7**, 712–717 (2002).
53. Burchard, E.G. *et al.* The importance of race and ethnic background in biomedical research and clinical practice. *N. Engl. J. Med.* **348**, 1170–1175 (2003).
54. Karter, A.J. Commentary: Race, genetics, and disease—in search of a middle ground. *Int. J. Epidemiol.* **32**, 26–28 (2003).
55. Lin, S.S. & Kelsey, J.L. Use of race and ethnicity in epidemiologic research: concepts, methodological issues, and suggestions for research. *Epidemiol. Rev.* **22**, 187–202 (2000).
56. Kim, U.K. *et al.* Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* **299**, 1221–1225 (2003).