Human genetic variation and health: new assessment approaches based on ethnogenetic layering

Fatimah LC Jackson

University of Maryland, MD, USA

Introduction

Human genetic variation is often biologically relevant, particularly when it influences (or is influenced by) health outcomes. For example, human genetic variation can modulate disease aetiology as in the case of homozygous beta sickle gene (βS/βS or sickle cell) pathology. Conversely, health outcomes, such as the frequency and duration of homozygous sickle cell pathology, can change affected group gene frequencies by selectively targeting and culling specific genotypes in a group, such as clinically more severe βS^Bantu/βS^Bantu versions of the βS gene, thereby changing future patterns of genetic variation in this gene.

Whereas the above case of the βS gene is a classic example, identifying the actual role of human genetic variation in health is, in many other cases, often problematic. This is because, in many health-related conditions, it is difficult to discern the precise contribution of genetics to general human biological variability. Genetic variability is only one component of human biodiversity and its relationship to human biological variability is non-linear. Genes interact with each other and with the environment and the products of these interactions may vary throughout the lifetime of the affected human. Discriminating between the various contributors to an interactive and dynamic condition (human biodiversity) and then correlating this with health remains an important contemporary challenge. For example, efforts to identify key genes influencing multifactoral complex phenotypes such as many of the psychiatric disorders continue to be less than satisfactory[1,2]. Most major diseases and traits are polygenic; the results of multiple genes with small additive effects. Scientists are just beginning to understand how these genes interact with each other[3] and with environmental factors in ways that impact on health.

The most important factors modulating human biological variability include genetic factors with biological effects, non-genetic factors with biological effects, sociocultural/environmental factors with biological ramifications, and/or idiosyncratic factors with biologically evident phenotypic

Correspondence to:
Fatimah LC Jackson,
Ph.D., University of
Maryland, MD, USA

effects. Genetic factors ultimately must pass through several filters before they become part of the expressed genotype as depicted in Figure 1. Identifying and isolating the genetic component from the other factors can be complicated, particularly when the health phenotypes under question are highly nuanced as, for example, in such expressed genotypes as nicotine resistance, alcohol tolerance and emotional balance. Identifying the 'moving target' of health is complicated by the fact that we often lack objective diagnostic tests, that human behaviour is culturally and ecologically complex, and that the genetic traits we suspect as contributing to health status often have variable and incomplete penetrance.

Furthermore, insights gained through using specific family studies, animal (usually mouse) models and retrospective assessments of affected individuals may not be broadly applicable to large groups of humans across geographical space or through generational time. At the individual level, human biological variation is presumably more strongly linked to one's health status than when larger ethnic, regional, or geographical group patterns of biological variation are used to predict group or individual health. This is because genetic and non-genetic contributions to human biological variation may produce substructured diversity in health as assessed by phenotype (expressed genotype) and/or actual genotype.
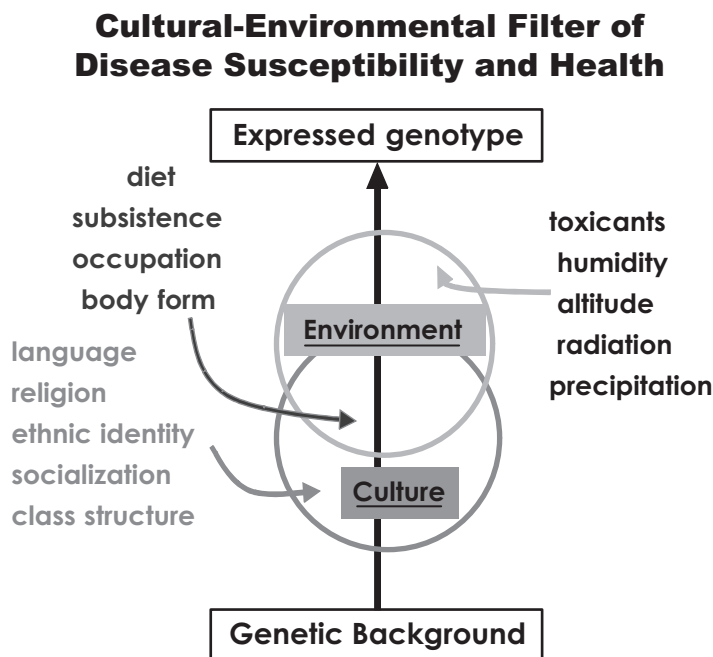
Cultural-Environmental Filter of Disease Susceptibility and Health

Fig. 1 Important filters of influence from genotype to expressed genotype (phenotype). These filters may interact with specific aspects of the genome, influencing functional gene expression as well as gene–gene interaction thresholds.

Presumed biomarkers for many important environmental health disorders often vary by ethnicity, such as in the case of asthma. This inflammatory airways disorder is phenotypically heterogeneous and appears to have an important genetic component in its expression. When linkage studies were done on an ethnically diverse cross-section of affected families with the disease, linkage to six novel genome regions were detected. However, the significant genome regions differed among African Americans, European Americans and US Latinos (Hispanic Americans). Ethnicity-specific analyses, revealed different frequencies of asthma-susceptibility genes in each ethnic group, suggesting linkage at 6p21 in the European American population, at 11q21 in the African American population, and at 1p32 in the US Latino population[3]. This diversity was revealed at a crude level of substructuring. However, had this ethnic substructuring not been a part of the study's research design, broad inter-ethnic differences in the specific genome regions linked to asthma-associated phenotypes would not have been detected. Imagine the diversity that could have been revealed had the researchers examined substructuring *within* each of these macro-ethnic groups.

It has been reported that there is a growing sense in genetic epidemiology that many findings are failing to replicate, because many of the claimed associations are false positive. These false positives are seen because of our inability to study many genetic variants in relation to many disease outcomes[4,5] without knowing the precise biocultural background of the groups being studied. Perhaps this lack of replication is a reflection of the over-reliance in public health, epidemiology, and biomedicine in particular, on poorly specified, sociologically defined 'racial' groups as a corner-stone of biological analysis. This chapter presents an alternative to the classic macroethnic, racial approach by proposing more carefully defined categories and smaller units of assessment in determining the reciprocal influences of human genetic variation on health, and of health outcomes on population human genetic diversity.

Chapter overview

This chapter considers those aspects of human biodiversity (specifically genetic and nongenetic variability with biological consequences) that appear to vary within and between geographical groups and examines their correlation (or lack of association) with various parameters of health. Case studies are presented to illustrate the complex interrelationships of human genetic variability and health. The chapter then proposes an alternative bioanthropological strategy for identifying genetic and nongenetic substructuring within and between geographical groups, called ethnogenetic layering. This technique, when applied in multiethnic settings, may

facilitate the identification and testing of smaller and presumably more genetically homogenous and socioculturally uniform groups, thereby providing an alternative, more nuanced, and anthropologically precise strategy for assessing the interrelationships of human genetic variation and health.

Origins and maintenance of human biodiversity

Recent work in anthropological genetics suggests that the traditional, historic and socially-constructed 'racial' aggregates that have permeated the Western biomedical literature since the 18th century are largely genetic illusions. Important human biological variation exists, but classical races, as the term is used systematically and taxonomically in the natural sciences, appears inapplicable to modern humans. Traditionally, Western biomedicine and public health have embraced four or five continent-based major races of humans, often using local US or European representatives as proxies for these continental groups. This blunt, often ahistorical comparative strategy continues to dominate the biomedical literature, even though its deficiencies are well noted. Underpinning the medical acceptance of biological race has been the assumption that substantial human genetic variability is at the core of racial group-level human differences. In the United States, the groups of choice, for comparative health status studies are usually identified by such terms as 'Black', 'White', 'Latino' or 'Hispanic', and 'Asian'. Significant within-group variation is often ignored and this inherent variability is now returning to haunt researchers searching for broad racial generalizations. The key questions in using these macroethnic 'racial' groups have been (and continue to be): (1) Do these groups represent statistically valid biological categories? And (2) can they be used as reliable shortcuts to making predictions (probability statements) about group disease susceptibilities and health status? The answers to both of these questions are a resounding No.

Modern human origins are in continental Africa where our collective ancestors spent most of human evolutionary history. Subsets of modern humans appear to have migrated out of (and back into) Africa as early as 100,000 years ago, eventually spreading to encompass the world. Our species collective origins are too recent, the extent of gene flow between us is too great, and our current diversity is too evolutionarily superficial to warrant the racial or subspecies level of differentiation among contemporary humans. Human variability does not neatly package itself into separate and discrete categories, as the term race would indicate. In fact, from a scientific point of view, we humans are a single, highly variable, polytypic race—*Homo sapiens sapiens*. The second 'sapiens' is actually the subspecies or race category. What biodiversity exists among modern humans exists taxonomically below the subspecies level.
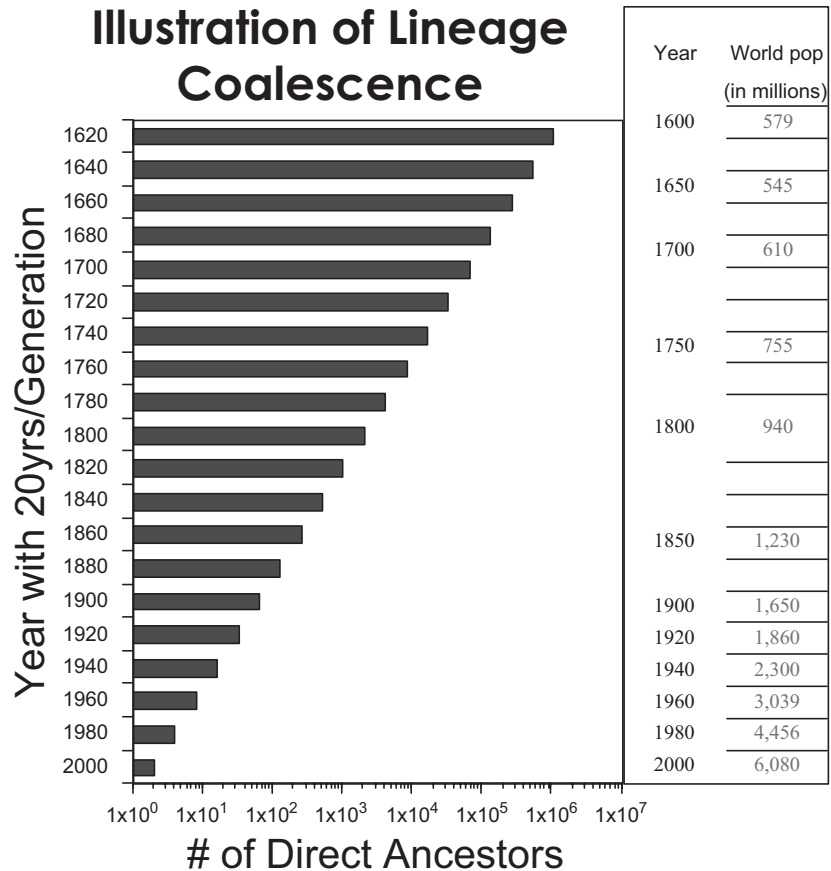
Illustration of Lineage Coalescence

Year	World pop (in millions)
1600	579
1650	545
1700	610
1750	755
1800	940
1850	1,230
1900	1,650
1920	1,860
1940	2,300
1960	3,039
1980	4,456
2000	6,080

Fig. 2 Biological lineage coalescence. Note that the number of ancestors increases exponentially going back in time but that the number of actual humans on the planet decreases. Therefore, it is obvious that biological lineages must converge around shared ancestors, thus increasing the potential for genetic similarity among all modern humans.

Since patterns of genetic variation at the molecular level do not always faithfully correspond to the phenotypically expressed individual and group patterns of biodiversity, this presents an initial dilemma for researchers and policy makers. The old race/ethnic paradigm mentioned earlier that still reigns in much of public health, epidemiology and biomedicine would have us believe that there is no lineage coalescence or shared ancestries between members of different racial or macroethnic groups. It presumes that each 'human race' has its own constellation of group-specific genes and that most, if not all, of the phenotypic variation evident in comparisons between these groups is because of differential genetics. This old paradigm is based on the presumption that real human biological races exist, that they can be easily delineated, that they represent long-standing patterns of reproductive isolation, and that they have perpetuated

with significant consistency through time. This is referred to as the 'fallacy of racial thinking' that continues to pervade much of the human sciences[6].

In fact, there is tremendous biological lineage overlap in modern humans. We all share many ancestors in common and the farther we go back in time, the more common ancestors are found. In the year 1620 C.E., each of us alive today had, on average, 20 generations of direct ancestors or 1,048,576 individuals. However if we contrast the graph of direct ancestors with the estimated world population figures, as depicted in Figure 2, it is clear that there must have been as much significant biological lineage redundancy among humans 380 years ago as there is today. In a biological sense, we remain 'subracially labile'. Our diversity is at the subspecific (subracial) level, well below the second '*sapiens*' in our taxonomic classification as *Homo sapiens sapiens*. The more we understand about the fine tuning of human biology and culture, the more difficult it is to match what we now know about human biological diversity with the pervasive, traditional 19th and 20th century sociologically-based categories of human biological variation.

Race, human genetic variation, and health

If we humans are a single biological race—sapiens—then this implies that the biological variation that exists among us is at a lower level than the subspecies or racial category. Genetically, only a very small, low level of sequence diversity is evident between 'races'. Less than 0.01% overall base pair differences separate one geographical group of humans from another. This 0.01 difference represents approximately 300,000 bases. Since the draft human genome sequence (about 3 billion base pairs) was completed in 2001, appropriately issuing in the 21st century as the century of the study of human genetic variability, we have come to realize that humans have fewer protein-coding genes than expected, and that most of these are highly conserved. Much of the variation between individual humans, including that which may affect our predispositions to common diseases, is probably the result of differences in the non-coding regions of the genome (*i.e.* the control architecture of the system). Humans and other complex organisms produce massive amounts of non-coding RNAs, which may function as another level of genetic output that controls phenotypic differentiation and development. Classical monogenic diseases and other differences caused by mutations and polymorphisms still seem to be caused by variations in protein-coding genes[7].

The differences between geographical groups of modern humans are less than the differences between any two unrelated humans. Overall human heterogeneity is less than that observed among existing populations

of contemporary chimpanzees. Modern chimpanzees (*Pan troglodytes*) are more biologically diverse than are modern humans. Yet how do we reconcile the lack of remarkable inter-group distinctions at the evolutionary level with the fact that human biological differences are often clustered and often have important clinical relevance?

Intra-group genetic variability and health outcomes

Genetic variability within a group (particularly among macroethnic groups aggregated on the basis of cultural, geographical, or linguistic criteria) may complicate detecting and correlating specific, broadly applicable candidate genes with specific health outcomes. Most large and socially constructed groups evidence a high degree of intra-group genetic variability at many of the loci of interest. For example, marked interindividual variability in genetic and non-genetic factors can influence the disposition of many endo- and xenobiotics, including the metabolism of environmental toxins[8] affecting health. An unusual genetic background or localized behaviour trait may place an exposed individual at a higher risk for adverse health when in contact with particular constellations of environmental toxins. Even dormant genetic alleles that are a part of normal variation may become activated in specific environmental contexts, for example, when coupled with certain environmental toxins. Long and colleagues[9] have discussed the interaction of arylsulphatase-A (ASA) allelic variation, environmental lead exposure, and an increased risk for neurodevelopmental damage in urban children.

As another example of human genetic variability and health, consider the case of the enzyme methyltetrahydrofolate reductase or MTHFR; the gene product is a cofactor for folic acid metabolism. Some researchers have suggested that there is a relationship between MTHFR polymorphism and human neoplasia. Carriers of genotypes containing the methylenetetrahydrofolate reductase 677T allele show constitutive low levels of 5-methylcytosine in their genomes, and tumours in these patients do not achieve severe degrees of global hypomethylation[10]. The methyltetrahydrofolate reductase (MTHFR) 677C–>T polymorphism is associated with a reduced risk of some forms of cancer, however, the protective effect of this folate-related polymorphism is dependent on adequate folate status. In a folate deficiency state, genetic polymorphisms of methyltetrahydrofolate reductase polymorphisms produce megaloblastic anaemia, classic neuropathy of the spinal cord, and an increase in homocysteine in the blood (a major risk factor for cardiovascular disease). So, in this case, the health effects of genetic variation in MTHFR are modulated by folate status. Together, these gene×environment interactions can influence important aspects of disease diathesis.

US Latinos, one of the fastest growing multiethnic groups in the United States, are a perfect example of imbedded heterogeneity within a highly diverse, socially constructed group. Rather than being considered genetically 'unusual' among modern humans, their magnitude of internal variability is, in fact, similar to other large, multi-ethnic aggregates. All US Latinos are basically either di- or trihybrid, their ancestral populations being a combination of European, African and Native American Indian biological and cultural lineages. However, the proportion of biologically important genes and cultural factors US Latinos received from these ancestral populations varies greatly. In the Western part of the United States, Latinos are mostly of Mexican origin, and in the East, they are predominantly of Cuban and Puerto Rican origin[11]. Using six autosomal DNA markers (LDLR, GYPA, HBGG, D7S8, GC and HLA-DQA), Bertoni and colleagues[11] identified, by US region of sampling, the different ancestral contributions to the US Latino population. Genetic diversity ranged from a trihybrid structure with European, Native American Indian and African contributions in the states of California, Nevada, Florida, New Jersey, and Virginia to a dihybrid structure with European and Native American contributions among the US Southwest population. However, in the state of Pennsylvania and among the US Southeast population, European and African ancestral contributions are more important. In another study of Y-STR haplotypes among US Latinos[12], the population exhibited significant geographic heterogeneity. Since the genetic propensity for developing a number of chronic diseases in the United States is being addressed in more sophisticated and comprehensive ways with the new genetic technology, it is imperative that the technology be applied and interpreted in a culturally and historically informed manner. This is the only way we will be able to truly assess the relationship of human genetic variability and health. US Latino populations are of particular interest because they seem to show different disease susceptibilities depending on their point of geographical origin[13].

Additionally, recent research suggests that US Latinos exhibit clear variabilities in response to the traditional US health care system[14], in part because of inherent within-group biological and cultural diversity. Health is not simply the absence of disease. More meaningfully, health is a biocultural state of relative equilibrium and normal function maintained within specific temporal, social, cultural and ecological contexts. The relationship of health with inherent human genetic variability is rarely predictable since most human genetic variability is not linked to obvious pathologies. For highly diverse groups such as US Latinos and others, who are usually analysed at the macroethnic (*i.e.* 'racial' or demic) levels of assessment, we are only in the earliest stages of decoding to what extent group biodiversity dictates health outcomes.

HLA variability in populations illustrates the complexity of health

Numerous studies have clearly indicated a role for the major histocompatibility complex (MHC) in susceptibility to autoimmune diseases. Susceptibility to coeliac disease (discussed later in this chapter) shows such a relationship with HLA variability. Most of the studies of HLA background genetics and health status have focused on the genetic variation of a small number of classical human-leukocyte-antigen (HLA) genes in the autoimmune disease gene region. By using linkage disequilibria to study the relationship between human genetic variation and health status and a high-density map of single-nucleotide polymorphisms (SNPs), researchers are increasingly able to propose potentially good candidate genes. However, although these immunologically-associated genes represent potentially good candidates within well defined groups, linkage disequilibria (LD) surrounding these genes has made it difficult to rule out neighbouring genes, many also with important immune functions, which may also influence disease susceptibility and health in more broadly distributed human groups. Perhaps using a high-density map of single-nucleotide polymorphisms (SNPs) would begin to facilitate a better understanding of the nature of the observed associations in diverse groups, as well as lead to the identification of causal variation. A comprehensive analysis of the patterns of linkage disequilibria and human variation was recently done using 201 SNPs, nine classical HLA loci, two TAP genes and 18 microsatellites. From these results, researchers were able to propose that the MHC has patterns of linkage disequilibria and variation that are essentially no different from those in the rest of the human genome. The exception seems to be the classical HLA markers that behave in a more straightforward, Mendelian way with respect to human genetic variation and health.

Regional differences in HLA genetics may define, to some extent, subgroup susceptibilities to harmful environmental agents and even the identities of such harmful agents. Compounds that may be highly toxic for a significant segment of one local regional group, may be essentially nontoxic for another (otherwise similar) regional group. For example, North Americans with full or partial ancestry in the northern and highland part of Atlantic Europe (the British Isles, Norway and much of Sweden and Denmark) have an increased frequency of HLA*DQ2+* phenotypes. The genotypes underlying these phenotypes, for example HLA*DQ beta 1*0201*, demonstrate clear sensitivity to wheat gliadin and susceptibility to the cell-mediated immunity disorder, coeliac disease[15,16]. For these individuals, wheat gliadins and as yet unknown compounds in rye, barley, oats and triticale[17] can provoke often fatal sensitivities and are,

for genetically susceptible individuals, clearly environmental toxins. In these cases, the maintenance of health requires avoidance of wheat gliadins. Without this environmental trigger, the responsible human genetic variants are unable to initiate disease (and impair health).

Case studies of specific human genetic variants and health

hNP and hGSTO1-1 genes and arsenic metabolism

Human genetic differences are known to modulate toxicant metabolism, and in so doing influence health status. An example of such a toxicant with differential metabolism based on human genetic variability is arsenic. Millions of persons worldwide are exposed to arsenic, primarily through natural enrichment of drinking water drawn from deep wells. When humans come in contact with inorganic arsenic, a known cause of skin cancer[18], this toxin is methylated (primarily in the liver but in other organs as well) through a detoxification process to methylarsonic acid (MMA) and dimethylarsinic acid (DMA). Variations in arsenic metabolism may affect individual risks of toxicity and carcinogenesis. In fact, a study in southwest Taiwan[18] recently concluded that arsenic biotransformation including methylation capacity is likely to have a role in the development of arsenic-induced skin disorders, particularly skin cancer.

Marked differences in arsenic metabolism have been observed in humans at the individual and group levels. Some authors have suggested that individuals with low MMA in their urine have faster elimination of ingested arsenic, compared to those with more MMA in urine[19]. Of the arsenic in urine, on average, human urine contains 10–30% inorganic arsenic, 10–20% MMA and 60–80% DMA[20]. MMA and DMA are less reactive with tissue constituents, less toxic, and more readily excreted in the urine than is inorganic arsenic. Recent studies[21] have identified ethnic and regional groups with unusually low or high urinary excretion of MMA and there seems to be functional genetic polymorphism in the biomethylation of arsenic and the potential for resulting toxicity. Using data from three populations, from Mexico, China and Chile, Loffredo and colleagues[21] analysed the distribution in urine of total arsenic and arsenic species, inorganic arsenic (InAs), monomethyl arsenic (MMA), and dimethyl arsenic (DMA). In all persons, most urinary arsenic was present as DMA. Male to female differences were discernible in both high- and low-exposure groups from all populations, but the gender differences varied by populations. In 1995, Vahter and colleagues[22] reported an unusual pattern of arsenic metabolism in indigenous Andean women from four northwestern Argentinean Andean villages with elevated levels of As in the drinking water (2.5, 14, 31, and 200 μg/l, respectively). Andean group median

concentrations of metabolites of inorganic As, methylarsonic acid (MMA), and dimethylarsinic acid (DMA) in the urine varied between 14 and 256 µg/l. Urinary concentrations of total As were only slightly higher (18–258 µg/l), indicating that inorganic As was the main form of As ingested. In contrast to all other world populations studied so far, arsenic was excreted in the urine mainly as inorganic As and DMA. There was very little MMA in the urine. Furthermore, studies among Andean women demonstrate that they are individually stable in their arsenic metabolism patterns[23], again suggesting that the variability is likely the result of genetic rather than nongenetic influences.

Another indication of there being a strong genetic component to arsenic metabolism is the finding that methylation patterns aggregate in families and are correlated in siblings[24]. Yu and colleagues[25] screened two genes responsible for arsenic metabolism, human purine nucleoside phosphorylase, *hNP*, which functions as an arsenate reductase converting arsenate to arsenite, and human glutathione S-transferase omega 1–1, *hGSTO1-1*, which functions as a monomethylarsonic acid (MMA) reductase enzyme, converting MMA(V) to MMA(III). Their goal was to develop a comprehensive catalogue of commonly occurring genetic polymorphisms in these important arsenic detoxification genes. This screening allowed them to generate a catalogue of DNA sequences from 22 individuals of European ancestry and 24 individuals of indigenous Native American Indian ancestry. In the *hNP* gene, 48 polymorphic sites were observed, including six that occurred in exons, of which one was nonsynonymous (G51S). One intronic polymorphism occurred in a known enhancer region. In the *hGSTO1-1* gene, 33 polymorphisms were observed. Six polymorphisms occurred in exons, of which four were nonsynonymous. In contrast to the *hNP* gene, in which the Native American Indian group was more polymorphic than the European group, in the *hGSTO1-1* gene the European group was more polymorphic than the Native American Indian group, which had only one polymorphism with a frequency >10%. These macroethnic group differences may be potentially important in explaining geographical-group level functional differences in arsenic detoxification patterns and the resulting health consequences. Furthermore, they suggest that the *hNP* and *hGSTO1-1* genes need to be evaluated as potential susceptibility genes in human arsenicism in diverse microethnic groups and the results of this screening be correlated with the functional health consequences of this variability.

GST polymorphisms, and lung and squamous cell cancers

Polymorphisms in glutathione S-transferase genes (*e.g.* GSTM1, GSTT1, GSTP1) have variable ethnic distributions and are associated with the

detoxification of many carcinogens, including polycyclic aromatic hydrocarbons such as those from tobacco cigarette smoke. The enzymes produced by these genes detoxify reactive epoxides, including carcinogens produced by tobacco smoke[26]. It is suspected that the null polymorphisms in the GSTM1, GSTT1 and GSTP1 genes that code for glutathione S-transferase may differentially influence susceptibility to smoking-related lung cancer in various groups of modern humans.

A number of studies suggest that GSTM1 and GSTT1 polymorphisms play an important role in the development of lung cancer and modify the risk for smoking-related lung cancer in the macroethnic group known as African Americans. A number of studies have been published about the association between GSTM1 and GSTT1 polymorphisms and lung cancer, including a recent case–control study[27]. In this study, samples of DNA from 117 lung cancer cases and 120 controls were assayed to detect glutathione S-transferase polymorphisms. The authors estimated the odds ratios (ORs) and 95% confidence intervals (CIs) for lung cancer associated with homozygous deletion of the GSTM1 gene and other risk factors using logistic regression. In 37 of the 117 cases (31.6%) and 24 of the 120 controls (20.0%), the GSTM1 null genotype was observed. The OR was 2.10 (95% CI 1.07–4.11) after adjustment for age, gender and smoking. The association was higher for squamous cell carcinoma (OR 2.98, 95% CI 1.09–8.19) than for adenocarcinoma (OR 1.95, 95% CI 0.81–4.66). Ford and colleagues[27] observed a stronger association between the GSTM1 null genotype and lung cancer among heavy smokers exposed to 30 or more pack-years (OR 4.35, 95% CI 1.16–16.23). A similar association was also found in squamous cell carcinoma (OR 6.26, 95% CI 1.31–29.91). When GSTM1 polymorphism was combined with cigarette smoking, smokers with the null genotype had a high risk (OR 8.19, 95% CI 2.35–28.62) compared with non-smokers with the wild-type genotype, and the risk significantly increases as the number of smoking cigarette pack-years increase.

A second case–control study[26] investigated the association of the GSTT1 and GSTM1 polymorphisms with lung cancer and compared a second group of 108 African Americans and 60 US Latinos (Mexican Americans) with lung cancer and a group of 132 African Americans and 146 US Latinos (Mexican Americans) as controls. In the unadjusted data, there was a borderline significant association of the GSTM1 null polymorphism with lung cancer in these US Latinos (OR 1.8, 95% CI 1.0–3.3) that was not observed in this second group of African Americans. The GSTT1 null polymorphism also had a higher (but not statistically significant) prevalence in cases than controls in both ethnic groups. Using logistic regression (controlling for age, gender, ethnicity and smoking), no significant association of either genetic trait with lung cancer was observed, with ORs for both traits of approximately 1.3. However, when the researchers compared the prevalence of individuals who were

null for both polymorphisms on a case by case basis, a significant interaction was observed. Logistic regression models showed the OR for the association of lung cancer and the presence of both null polymorphisms compared with one (either GSTT1 or GSTM1) or no null genotype to be 2.9, suggesting that there may be carcinogenic intermediates in cigarette smoke that are substrates for both the GSTT1 and GSTM1 enzymes. This would suggest that, in these groups, the risk for lung cancer is increased in a greater than additive fashion when certain African American and certain US Latino individuals have both the GSTT1 and GSTM1 null polymorphisms.

The role of these same genetic polymorphisms among a group of French Europeans (Caucasians) suggests that GSTM1 null genotypes pose a moderate risk factor for lung cancer[28]. In this group, the GSTT1 genotypes had no significant effects on their lung cancer risk. However, a third class of glutathione S-transferases coded for by the GSTP1*B/*B genotype posed a two-fold risk (OR 2.0, 95% CI 1.0–4.1) of developing small cell lung cancer in this group when compared with genotypes containing the GSTP1*A allele (another variant within the P class of glutathione S-transferase). Among this group of French Europeans, the most remarkable risk for small cell carcinoma was seen among subjects with the GSTP1*B/*B genotype and concurrent lack of the GSTM1 gene (OR 6.9, 95% CI 1.6–30.2). In this group, the deficient genotypes for GSTM1 and GSTP1 seem to be important risk modifiers for lung cancer, especially when observed in combination.

GSTM1 has also been studied among Chinese populations and evaluated as to its impact on the metabolism of tobacco-related carcinogens. Using allele specific PCR and multiplex PCR techniques to identity the genotypes of GSTM1 in a case–control study, Chen and colleagues[29] evaluated 106 lung carcinoma patients with histopathological diagnosis and 106 matched controls free of malignancy in Jiangsu Province, China. Logistic regression analysis was carried out to calculate the OR and 95% CI. The results showed that in this group of Chinese, individuals with GSTM1 null had an elevated risk of lung cancer. Light smokers (<30 packs per year) with the GSTM1 null genotype were shown to have the increased risk to lung carcinoma (OR 3.47; CI, 1.13–7.57), suggesting that the null GSTM1 genotype might affect the genetic susceptibility for lung carcinoma in these particular Chinese people.

The take-home message in considering these seemingly contradictory results on the impact of GST polymorphisms on cancer morbidity and mortality is that overall human biodiversity ultimately modulates many of the health outcomes of the expression of particular genes and gene complexes. Human biodiversity will reflect genetic variation that is filtered by social, economic, cultural, historical, geographical and other non-genetic sources with biological effects to produce individual and microethnic specific 'expressed genotypes'.

CASR-BsaHI, AHSG-SacI, ESR1-PvuII, ESR1-XbaI, VDR-ApaI and PTH-BstBI polymorphisms, BMD and osteoporosis

Bone mineral density (BMD) is an important risk factor for osteoporosis and has a strong genetic component. Osteoporosis is an important health problem in the world. Whereas average BMD differs among macroethnic groups, several important candidate genes have been shown to underlie intra-macroethnic group BMD variation. Dvornyk and colleagues[30] investigated whether important candidate genes contributed to macroethnic differences in BMD by evaluating the degree of genetic differentiation among five important candidate genes observed in European Caucasians and Han Chinese. The genetic variability of these two highly diverse groups was assessed using 1131 randomly selected individuals evaluated at six polymorphic restriction sites for five important candidate genes for BMD. Specifically, Dvornyk and colleagues[30] focused on the BsaHI polymorphism of the calcium-sensing receptor (CASR) gene, the SacI polymorphism of the alpha2HS-glycoprotein (AHSG) gene, the PvuII and XbaI polymorphisms of the oestrogen receptor alpha (ESR1) gene, the ApaI polymorphism of the vitamin D receptor (VDR) gene, and the BstBI polymorphism of the parathyroid hormone (PTH) gene.

Among the Chinese and the Europeans studied, significant allelic and genotypic variability was observed in each of the polymorphisms assessed. The mean F_{ST}, a test of correlation of the genetic distances, was 0.103, which significantly differed from zero. The Chinese people had lower mean heterozygosity (0.331) compared to the Europeans (0.444) with genetic diversity in the CASR-BsaHI and PTH-BstBI polymorphisms contributing most significantly to this difference. Another study of the relationship of the AHSG gene to bone formation, metabolism, BMD and the development of osteoporosis evaluated 1260 individuals from 401 Chinese nuclear families[31]. In this study, subjects were genotyped using PCR-RFLP at polymorphic Sac I site inside the exon 7 of the AHSG gene and BMD was measured at the lumbar spine and hip region by dual-energy X-ray absorptiometry (DXA). Using the QTDT (quantitative trait transmission disequilibrium test), Liu and colleagues[31] found no association or linkage between the AHSG-Sac1 gene and BMD variation at the spine or hip.

Inter- and intra-group variability in these candidate genes implies that, at some loci, various types of natural selection may have influenced the observed patterns of variation. Since each of these candidate genes presumably contributes to the observed variation in BMD in some groups of humans worldwide, it is possible that a broad range of genetic polymorphisms may underpin some of the existing differences in bone mineral densities and osteoporosis. The key point may well be to be able to match the correct human microethnic group with the most informative candidate

genes for the specific health concern being addressed. To do this, however, we can no longer rely on macroethnic level assessments. We must assess groups at a much lower level, at a microethnic level so as to be able to capture the most salient genetic and nongenetic factors affecting, in this case, osteological status.

Rationale for new approach

Major technological innovations in molecular biology have accelerated the need for more sophisticated anthropological models to guide the assessment of human biodiversity and its relationship to health. Recent advances in genome science indicate that the human genome is comprised of only approximately 30,000 genes and that only a mere 300 of these genes are uniquely human. The other 29,700 are shared with other species. Humanity is genetically linked to the rest of life on this planet and all humans are intimately bound to each other. There are just not enough genes to explain, by themselves, the tremendous variation that exists among humans as it relates to health or other aspects of our existence. Since we are more complex than our genes, to understand the relationship of human genetic variability to health, we must integrate detailed knowledge of our environment, including data on geographical and cultural stratification, with the newly discovered genetic information.

A cogent and consistent message from the new genetic knowledge of human biodiversity is that what differences exist among modern humans need to be examined under highly detailed scrutiny. Our strategy for viewing human variation now needs to be as sophisticated as the emerging laboratory and computer technology for determining human molecular variation. Paying careful attention to historical nuance, social context and demographic detail, for example, can initiate a better understanding of the interactions of genes and non-genetic factors in gene expression and subsequent health impact. The development of bioanthropologically-rich, regional population models that are amenable to scientific manipulation and hypothesis testing can lead to better and more appropriate use of health-related resources. A technique developed by the Genomic Models Research Group at the University of Maryland, ethnogenetic layering, can accelerate candidate gene (and candidate cultural behaviour) discovery process by providing detailed, site-specific approaches as a precursor to the search for disease-susceptibility biomarkers and clusters of microethnic group marker genes. The ability to subdivide various and diverse human groups who had otherwise been lumped as macroethnic groups is important because it can help researchers correlate microethnicity, genetics, cultural behaviour and disease susceptibility more precisely, particularly when the phenotypes produced are clinically similar.

Background on the concept of ethnogenetic layering

Over the last 12 years, I have led a group, the Genomic Models Research Group, that has been developing the concept of ethnogenetic layering in response to the reality of human genetic variability and its differential (and often unexpected) effects on human health over geographical space and through generational time. Ethnogenetic layering is a conceptual innovation that recognizes that in any given area of the world, many different constellations of ethnic groups have settled, often sequentially rather than simultaneously, and they have interacted over time. The intensity and duration of these interactions have varied in each area, as have the actual ethnic compositions of the resident groups. By identifying the population substructuring present in most diverse human groups, we have developed a tool for stratifying groups, based upon biologically important social, cultural and historical criteria, before their molecular genetic and clinical assessments[32].

The term, 'ethnogenetics' is at the core of ethnogenetic layering and is based on the concept that population genetics and ethnicity are often intertwined but not necessarily overlapping identities[33]. The term was originally used by scientists in the former Soviet Union who focused on the genetic variations and biochemical polymorphisms observed within the many social and cultural groups dispersed across the expanse of these republics. Ethnogenetic studies have been conducted among the Yakuts of the Republic of Sakha[34] and the Komi-Zyrians in Kormis[35]. Ethnogenetics has also been applied to the discussion of hypotheses about human genetic adaptation since the Paleolithic[36]. However, ethnogenetics in both of these contexts did not seek to integrate data on gene–environment interactions or health, rather it was more narrowly focused on genetic variation and the substructuring of this diversity by ethnicity.

We have taken the term, expanded its original meaning, broadened its application, and modified it to fit the 21st century landscape of human biological variability. We have found that when macroethnic groups (such as European Americans) are regionally subdivided, and when genetic and cultural studies emphasize the assessment of those traits and clusters of traits that geographically distinguish regional microethnic groups within macroethnic groups, the correlations between the incidence of specific expressed genotypes (=phenotypes) and regional microethnic groups is stronger. This is because most microethnic groups are actually temporal-based constellations of specific, loosely affiliated biological lineages. Biological lineage affiliation alone (with its assumptions of shared socialization templates) will account for a higher probability of biocultural and genetic redundancy among microethnic group members and more likely shared health outcomes. Ethnogenetic layering maps have been developed to

provide spatial depictions of the interface of genetics, ethnic identity and health status, holding geography constant.

Ethnogenetic layering (Table 1) involves identifying important historical and ethnographic (cultural/historical/behavioural) detail over geographic space and then superimposing upon these depictions the geographic distributions of specific genes, gene clusters and health outcomes. A Geographic Information System (GIS) is usually applied to our database to generate raster and vector maps of the variables deemed interesting. Raster maps are used for continuous numeric values, using state counties as our cell size, to grid data for reclassification, interpolation and creation of surfaces. Vector maps were produced for county demarcated locations of microethnic groups where possible by defining each feature by an *x,y* location in space. We then connected the dots to draw lines and area outlines. Image data was added to our vector data to provide general geographical points of reference. The analysis of vector data involved summarizing the attributes in the layers data tables.

In multiethnic settings such as the United States, ethnogenetic layering often includes an array of local microethnic groups represented at any one geographic site. For example, in the Mississippi Delta region (depicted in Figure 3), ethnogenetic layering might include such microethnic groups as the Cajun [Acadia French] (as a subset of European Americans), the Creole and Black groups [with African origins in Senegambia, Central Africa and Bight of Benin] (as a subset of African Americans), and Choctaw, Houmas, Chickasaw, Coushatta, Caddo, Atakapa, Karankawa and Chitimacha peoples (as subsets of Native American Indians). When researchers interested in specific genes, gene clusters and/or health outcomes sample vertically through the layered groups, they are able to more easily identify shared potential candidate genes as well as shared cultural behaviours of biological importance.

As such, ethnogenetic layering offers far more within-group differentiation and nuanced detail than the classic macroethnic (=racial) approach affords. Furthermore, the attention to genetic and cultural regional variance in ethnogenetic layering offers a non-racial model for identifying regional genetic variation and gene–environment interactions that may significantly predict disease susceptibility and health status. For example, among the Mississippi Delta microethnic groups noted above, all are well known for

Table 1 Key data sources for ethnogenetic layering

Historical assessments (Archived US Customs records, Census Bureau data, US and British Naval records, Plantation diaries, Slave narratives, Post-emancipation letters, *etc.*)

Geographical appraisals (Geographic Information System facilitated reconstructions of population densities)

Cultural reconstructions (regional ethnographies, historical records)

Genetic evaluations (published literature, origin country studies)

Health risk factors (WHO database on toxicity, American Cancer Society reports, CDC reports, *etc.*)
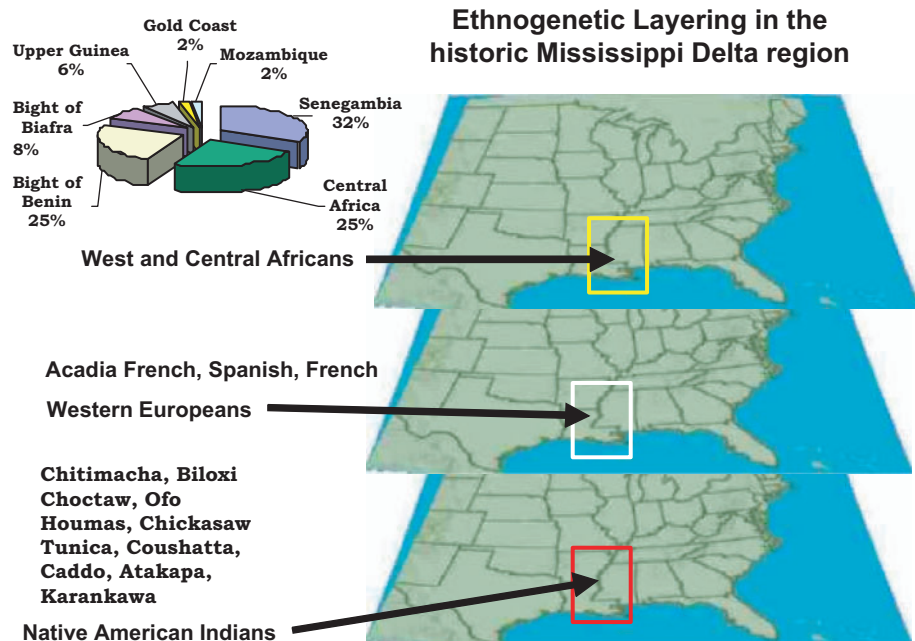
Ethnogenetic Layering in the historic Mississippi Delta region

Gold Coast 2%
Upper Guinea 6%
Mozambique 2%
Bight of Biafra 8%
Senegambia 32%
Bight of Benin 25%
Central Africa 25%

West and Central Africans

Acadia French, Spanish, French
Western Europeans

Chitimacha, Biloxi Choctaw, Ofo Houmas, Chickasaw Tunica, Coushatta, Caddo, Atakapa, Karankawa
Native American Indians

Fig. 3 Ethnogenetic layering map of the historic Mississippi Delta region. The Native American Indian groups represent the baseline microethnic groups during the 18th and 19th centuries, followed by the Western Europeans, followed by the West and Central Africans. Estimated proportions of West and Central Africans from various exit ports are presented in lieu of specific microethnic affiliations.

an important shared cultural dietary practice: the extensive use of sassafras in their traditional cuisine. Whereas sassafras (*Sassafras albidum*), the prime ingredient in gumbo filé, is originally a Native American Indian domesticate, its current broad multiethnic regional use is of potential biomedical importance since this plant contains safrol[37], a potent phytochemical associated with increasing susceptibility to pancreatic cancer[38]. Hence, in spite of the genetic variability distinguishing Cajun, Creole and Choctaw peoples, for example, their similar non-genetic but biologically important dietary practices have an important shared impact on their collective health.

Each day, our list lengthens of human genetic variants that affect health and that display variability both between and within macroethnic groups. In fact, as the list lengthens, the onslaught of genetic variation and health data often appear chaotic because the variations are clinal but non-uniform. As has been noted, irreproducible results have begun to appear increasingly in the literature when studies done on one segment of a macroethnic ('racial') group fail to be evident in subsequent studies of another segment of the same macroethnic group. Furthermore, patterns of variation in one set of genetic polymorphism and health outcome dyads do not correspond ethnically or geographically to patterns of variation observed in a second

Table 2 Relevant applications of ethnogenetic layering to various health-related research

- Pharmacogenetics
- Evaluation of environmental toxicant exposure
- Risk assessment for prematurity and low birth weight
- Susceptibility to nutritional toxicants
- Chronic disease expression
- Organ-tissue transplant compatibility
- Cancer risk

set. The results are confusing and at times present seemingly contradictory profiles of health-impacting human genetic variation. With so much of the previous scientific and biomedical literature based on race-level assessments, researchers have been limited in their abilities to easily substratify diverse macroethnic groups, with the result that much of the nuance of disease susceptibility is lost. Some have proposed that we move away from group assessments altogether, relying instead on individual assessments[39]. However, since many biologically important non-genetic processes affecting health are expressed in concert with others, in a group setting, only assessing individuals would reduce our access to these data, rendering our evaluations incomplete.

In this chapter, I have provided examples of the wealth of emerging information on the interface of specific human genetic variants and health, and have identified a model, ethnogenetic layering, that can be used to tease out underlying genetic and cultural complexity in disease susceptibility affecting health status. Table 2 identifies the broad range of application of ethnogenetic layering that, when applied to specific and well-defined human groups, genes and gene clusters, and disease entities, should greatly accelerate our understanding of the subtle yet dynamic nature of their health-influencing interactions.

References

1 Inoue K, Lupski JR. Genetics and genomics of behavioral and psychiatric disorders. *Curr Opin Genet Dev* 2003; **13**: 303–9

2 Van de Bree MB, Owen MJ. The future of psychiatric genetics. *Ann Med* 2003; **35**: 122–34

3 Xu J, Meyers DA, Ober C, Blumenthal MN, Mellen B, Barnes KC, King RA, Lester LA, Howard TD, Solway J, Langefeld CD, Beaty TH, Rich SS, Bleecker ER, Cox NJ; Collaborative Study on the Genetics of Asthma. Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. *Am J Hum Genet* 2001; **68**: 1437–46. Epub 2001; May 10

4 Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; **2**: 91–9

5 Colhoun HM *et al*. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; **361**: 865–72

6 Keita SO, Boyce AJ. 2001 'Race': confusion about zoological and social taxonomies, and their places in science. *Am J Human Biol* 2001; **13**: 569–75

7   Mattick JS. The human genome and the future of medicine. *Med J Aust* 2003; **179**: 212–6

8   Lamba JK, Lin YS, Thummel K, Daly A, Watkins PB, Strom S, Zhang J, Schuetz EG. Common allelic variants of cytochrome P4503A4 and their prevalence in different populations. *Pharmacogenetics* 2002; **12**: 121–32

9   Long J, Covington C, Delaney-Black V, Nordstrom B. Allelic variation and environmental lead exposure in urban children. *AACN Clin Issues* 2002; **13**: 550–6

10 Paz MF, Avila S, Fraga MF, Pollan M, Capella G, Peinado MA, Sanchez-Cespedes M, Herman JG, Esteller M. Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors. *Cancer Res* 2002; **62**: 4519–24

11 Bertoni B, Budowle B, Sans M, Barton SA, Chakraborty R. Admixture in Hispanics: distribution of ancestral population contributions in the Continental United States. *Hum Biol* 2003; **75**: 1–11

12 Kayser M, Brauer S, Schadlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res* 2003; **13**: 624–34

13 Gopal KS, Siahpush M. Ethnic-Immigrant differentials in health behaviors, morbidity, and cause-specific mortality in the United States: An analysis of two national data bases. *Hum Biol* 2003; **75**: 1–11

14 Bastida E. Hispanic health: persistent disparities and promising developments. *Ethn Dis* 2002; **12**: 456–8

15 Rocher A, Calero M, Soriano F, Mendez E. Identification of major rye secalins as coeliac immunoreactive proteins. *Biochim Biophys Acta* 1996; **1295**: 13–22

16 Sturgess R, Day P, Ellis HJ, Lundin KE, Gjertsen HA, Kontakou M, Ciclitira PJ. Wheat peptide challenge in coeliac disease. *Lancet* 1994; **343**: 758–61

17 Campbell JA. Foods for patients with celiac disease. *Can Med Assoc J* 1982; **127**: 963–5

18 Yu RC, Hsu KH, Chen CJ, Froines JR. Arsenic methylation capacity and skin cancer. *Cancer Epidemiol Biomarkers Prev* 2000; **9**: 1259–62

19 Vahter M. Mechanisms of arsenic biotransformation. *Toxicology* 2002; **181–182**: 211–7

20 Vahter M, Concha G. Role of metabolism in arsenic toxicity. *Pharmacol Toxicol* 2001; **89**: 1–5

21 Loffredo CA, Aposhian HV, Cebrian ME, Yamauchi H, Silbergeld EK. Variability in human metabolism of arsenic. *Environ Res* 2003; **92**: 85–91

22 Vahter M, Concha G, Nermell B, Nilsson R, Dulout F, Natarajan AT. A unique metabolism of inorganic arsenic in native Andean women. *Eur J Pharmacol* 1995; **293**: 455–62

23 Concha G, Vogler G, Nermell B, Vahter M. Intra-individual variation in the metabolism of inorganic arsenic. *Int Arch Occup Environ Health* 2002; **75**: 576–80. Epub 2002; Jul 9

24 Chung JS, Kalman DA, Moore LE, Kosnett MJ, Arroyo AP, Beeris M, Mazumder DN, Hernandez AL, Smith AH. Family correlations of arsenic methylation patterns in children and parents exposed to high concentrations of arsenic in drinking water. *Environ Health Perspect* 2002; **110**: 729–33

25 Yu L, Kalla K, Guthrie E, Vidrine A, Klimecki WT. Genetic variation in genes associated with arsenic metabolism: glutathione s-transferase omega 1–1 and purine nucleoside phosphorylase polymorphisms in European and indigenous Americans. *Environ Health Perspect* 2003; **111**: 1421–7

26 Kelsey KT, Spitz MR, Zuo ZF, Wiencke JK. Polymorphisms in the glutathione S-transferase class mu and theta genes interact and increase susceptibility to lung cancer in minority populations (Texas, United States). *Cancer Causes Control* 1997; **8**: 554–9

27 Ford JG, Li Y, O'Sullivan MM, Demopoulos R, Garte S, Taioli E, Brandt-Rauf PW. Glutathione S-transferase M1 polymorphism and lung cancer risk in African-Americans. *Carcinogenesis* 2000; **21**: 1971–5

28 Stucker I, Hirvonen A, de Waziers I, Cabelguenne A, Mitrunen K, Cenee S, Koum-Besson E, Hemon D, Beaune P, Loriot MA. Genetic polymorphisms of glutathione S-transferases as modulators of lung cancer susceptibility. *Carcinogenesis* 2002; **23**: 1475–81

29 Chen S, Xue K, Xu L, Ma G, Wu J. Polymorphisms of the CYP1A1 and GSTM1 genes in relation to individual susceptibility to lung carcinoma in Chinese population. *Mutat Res* 2001; **458**: 41–7

30 Dvornyk V, Liu XH, Shen H, Lei SF, Zhao LJ, Huang QR, Qin YJ, Jiang DK, Long JR, Zhang YY, Gong G, Recker RR, Deng HW. Differentiation of Caucasians and Chinese at bone mass candidate genes: implication for ethnic difference of bone mass. *Ann Hum Genet* 2003; **67**: 216–27

31 Liu XH, Liu YJ, Jiang DK, Li YM, Li MX, Qin YJ, Jian WX, Zhou Q, Deng HW. No evidence for linkage and/or association of human alpha2-HS glycoprotein gene with bone mineral density variation in Chinese nuclear families. *Calcif Tissue Int* 2003 Jul 24 [Epub ahead of print]

32 Jackson FLC. Concerns and priorities in genetic studies: Insights from recent African American biohistory. *Seton Hall Law Rev* 1997; **27**: 951–70

33 Jackson FLC. Race and ethnicity as biological constructs. *Ethn Dis* 1992; **2**: 120–5

34 Tarskaia LA, Makarov SV, Bychkovskaia LS, Pai GV, Pakendorf B, El'chinova GI, Deriabin VE, Spitsyn VA. Ethnogenetics of yakuts from three regions of Republic of Sakha (Yakutia) inferred from the frequencies of biochemical gene markers. *Genetika* 2002; **38**: 1282–91

35 Shubin PN, Morokov VA, Efimtseva EA, Chelpanova TI. Ethnogenetics of the Komi-Zyrians according to data on the distribution of gene frequencies of erythrocyte and serum systems in the blood. *Genetika* 1997; **33**: 235–42

36 Balanovskaia EV, Rychkov IG. Ethnogenetics: adaptive structure of the gene pool of the mankind from the data on human polymorphic genetic markers. *Genetika* 1990; **26**: 739–48

37 Carlson M, Thompson RD. Liquid chromatographic determination of safrole in sassafras-derived herbal products. *J AOAC Int* 1997; **80**: 1023–8

38 Ireland CM, Cooper CS, Marshall CJ, Hebert E, Phillips DH. Activating mutations in human c-Ha-ras-1 gene induced by reactive derivatives of safrole and the glutamic pyrolysis product, Glu-P-3. *Mutagenesis* 1988; **3**: 429–35

39 Haga SB, Venter JC. Genetics. FDA races in wrong direction. *Science* 2003; **301**: 466