

## Review

# Recent advances of genetic ancestry testing in biomedical research and direct to consumer! testing

Via M, Ziv B, Burchard EG. Recent advances of genetic ancestry testing in biomedical research and direct to consumer testing. (*Hum Genet* 2009; 76: 225-235. © John Wiley & Sons A/S, 2009)

In the post-Human Genome Project era, the debate on the concept of race/ethnicity and its implications for biomedical research are dependent on two critical issues: whether and how to classify individuals and whether biological factors play a role in health disparities. The advent of reliable estimates of genetic (or biogeographic) ancestry has provided this debate with a quantitative and more objective tool. The estimation of genetic ancestry allows investigators to control for population stratification in association studies and helps to detect biological causation behind population-specific differences in disease and drug response. New techniques such as admixture mapping can specifically detect population-specific risk alleles for a disease in admixed populations. However, researchers have to be mindful of the correlation between genetic ancestry and socioeconomic and environmental factors that could underlie these differences. More importantly, researchers must avoid the stigmatization of individuals based on perceived or real genetic risks. The latter point will become increasingly sensitive as several 'for profit companies' are offering ancestry and genetic testing directly to consumers and the consequences of the spread of the services of these companies are still unforeseeable.

*NB*

In the past 20 years, we have witnessed a genetic revolution. It began with the effort to sequence the entire genome, included stem cell cloning and now sequencing of the entire genomes of at least 1000 individual people (1, 2). Along the way, we have discovered how genetically similar we are to each other and to other mammalian species. At the same time, we discovered that major continental groups differ by as much as 4-9% of the total genetic variance (3, 4). Coincidentally, these major continental groups corresponded to what some have traditionally classified as 'racial' categories. Consequently, we have witnessed a re-emergence of an age-old debate on the concept of race and its implications for biomedical research and health outcomes. Human history is fraught with examples of our obsession with classifying people into racial categories. We have even obsessed over the potential biologic implications associated

with these categories. Often times, biologic differences were exploited and used to classify one group as superior to another. However, at no previous time in history has the point become more salient as we now enter the era of personalized medicine with widely publicized examples such as *LuiDil*® for African-Americans with heart failure and black box warnings against *Carbamazepine* for Asians (5, 6). These examples highlight a debate on the importance of race/ethnicity in biomedical research and clinical practice.

Recent scientific advances have given this debate new meaning. Through genetic testing, investigators can now measure genetic or biogeographic ancestry. In short, we can use genetic testing to determine and quantify an individual's ancestral background with statistical precision. The science behind this body of work serves as the basis for forensic investigation and television drama. Only

M Via, E Ziv, c,  
EG Burchard

Institute for Human Genetics,  
Department of Medicine, Helen Diller  
Family Comprehensive Cancer Center,  
and Department of Bioengineering &  
Therapeutic Sciences, University of  
California, San Francisco, CA, USA

Key words: admixture  
mapping – biomedical research – genetic  
ancestry – health disparities – population  
differences

Corresponding author: Marc Via,  
University of California, San Francisco,  
Box 2911, CA 94143-2911, USA  
Tel.: +1 415 5149931;  
fax: +1 415 5144365;  
e-mail: [marcviagarcia@ucsf.edu](mailto:marcviagarcia@ucsf.edu)

Received 21 April 2009, revised and  
accepted for publication 22 June 2009

**Via et al.**

recently have the applications of ancestry testing entered into biomedical research and clinical practice (7-9). In addition, as the costs associated with genetic testing have decreased, ancestry testing has taken on a new twist as 'direct to consumer' marketing has popularized this topic, which is now often referred to as 'recreational genetics'. Despite the social and fiscal implications associated with ancestry testing, a new question has emerged. For clinical outcomes, which is more precise, self-identified or genetically determined ancestry? herein, we will discuss this new technology and its potential clinical and social implications and pitfalls.

**Human population structure**

There is broad consensus that modern humans evolved from pre-existing populations in Eastern Africa, where the oldest fossils with modern physical characteristics have been found and dated to approximately 160,000-200,000 years ago (10, 11). Contemporary investigations of human genetic variation support not only a common African origin for modern humans but also a subsequent colonization of Eurasia, Oceania and the Americas around 60,000-100,000 years before present (3, 12). Recent reports using genome-wide polymorphisms confirmed that: (i) genetic variation seen outside of Africa is generally a subset of the total genetic variation that exists within Africa, (ii) genetic diversity decreases with increased geographic distance from Africa, and (iii) linkage disequilibrium (LD) patterns increase proportionally to the distance from Africa (3, 13, 14). Although all groups originated from sub-Saharan Africa, the subsequent migrations out of Africa resulted in a series of founder effects with the start of each new population.

The degree of genetic differentiation between human populations is relatively modest. Of the total genetic variation among humans in single nucleotide polymorphisms (SNPs), less than 2% (1.78%) is found only in one continent; 89-94% of genetic variance at the autosomal level (85% for the X-chromosome) is within populations and 4-9% is found between continental groups (3, 4, 13). Although there is a possibility that these estimates are skewed because of biases associated with the selection of polymorphisms included in genotyping platforms, analyses from non-biased SNPs identified using resequencing by hybridization reveal similar estimates (15).

Although the genetic differences between human populations are relatively small, the subtle differences between populations are nonrandom and can

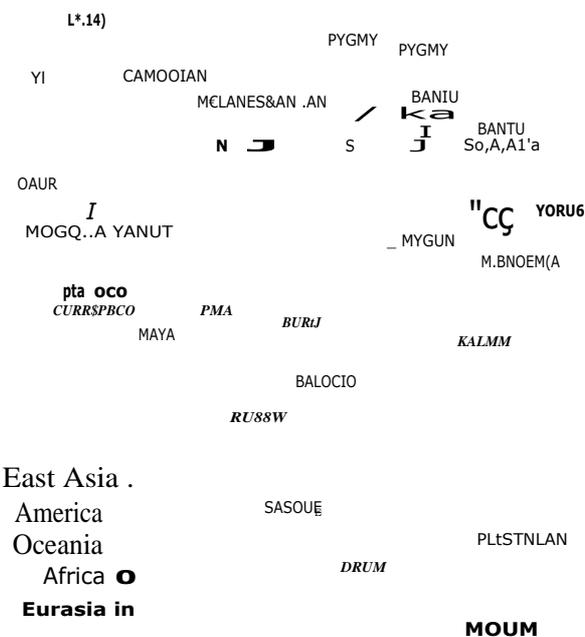


Fig. 1. Neighbor-joining tree of world population relationships ascertained from > 500,000 genome-wide SNPs. Based on the data presented by Jakobsson et al. (13) and available from <http://Aieurogenetics.nia.nih.gov/paperdata/public/> and <http://www.cephb.fr/hgdp-cephdb/>.

accumulate over a large number of loci. Widely accepted statistical methods can then use the information generated from the aggregation of these genetic differences to classify populations into broad continental groups (Fig. 1). Furthermore, these statistical methods can also be applied to assign the biogeographical origin of individuals within major groups (16).

**Classification of humans: race/ethnicity/ancestry**

Genetic differences between human populations have reified an old and widespread debate on the concept of 'race' and its validity as a taxonomic classification. Although many investigators have argued that this construct offers an opportunity to study the interaction between genetic, environmental, and social contributions to disease occurrence and drug response (17-19), several investigators disagree and see racial identity primarily as a social construct that can misdirect the categorization of participants in research projects (20-23). To complicate the debate further, it is not uncommon for the same individual to report their racial identity differently in different contexts and at different points in their lives (24, 25). Thus, many scholars view 'racial identity' as a dynamic and complex

*But Human Groups*

*NB*

construct, which represents an amalgam of biological and social factors. Consequently, some have promoted the term 'ethnicity' to characterize people simply to avoid the controversy and strong emotions associated with racial categorization. Ethnicity gives special emphasis to the cultural, socioeconomic, religious, and political dimensions of human populations rather than their genetic background, even though both terms, race and ethnicity, are related (26). However, the term 'ethnicity' also suffers from some flaws such as assuming a higher uniformity into a group ascribed with an ethnic identity (i.e., combining Cubans, Mexicans, Brazilians, Peruvians, and Argentinians, among others, in the ethnic Hispanic/Latino group) and the possibility that ethnicity can also change in different times or circumstances (27).

A separate question about the classification of humans into subpopulations is whether a model of discrete subgroups vs a model of continuous variation over geographic distance is most appropriate. Serre and Paabo argued that human genetic diversity is best represented in discrete subgroups (28). Using the data analyzed by Rosenberg et al. (4), they suggested that in many cases discontinuities may arise from insufficient sampling of the geographical area. However, Rosenberg et al. expanded their data analysis and concluded that a model that includes discrete clustering is appropriate and unlikely to be an artifact of sampling (29). Instead, it appeared to be correlated with geographic boundaries across continents.

In the United States, the Office of Management and Budget (OMB) Directive 15 sets the standard classifications of racial and ethnic data for federal statistics and administrative reporting. Since 2001, federally funded researchers have been required to categorize study participants into the OMB categories. Researchers rely on the study participants' self-description of their race and ethnicity and, in many cases, there is the risk that researchers could assume that self-identified race/ethnicity is a reasonable proxy for genetic homogeneity and use these same variables in the subsequent analysis and theoretical framing of the research (30).

### Genetically determined ancestry

As an alternative to the use of self-identified racial and ethnic categorization, it has been proposed that genetically determined ancestry may be more accurate when considering biomedical and/or clinical research (27). However, we must be clear about our definition of ancestry as it can be defined

on several levels: biogeographical (i.e., African vs Asian), geographical (i.e., southeast Asian vs northern European), geopolitical (i.e., Cambodian vs Swedish), or cultural terms (i.e., Jewish vs Berber). Furthermore, the description of ancestry can be self-identified, identified by an observer, or estimated from genetic data. In addition, ancestry can be defined by one or multiple sources.

Some reports have suggested that self-identified race/ethnicity correlates with clustering of genetic ancestral groups (31). However, self-identified estimates of genetic ancestry are less accurate than genetic testing and likely vary by population (27, 32, 33). Genetic estimates of ancestry become especially relevant for populations that have undergone recent admixture, where distortions in the relationship between genetic and self-assessed ancestry have been described. As an example, a recent study performed in the Southwest United States reported that 85% of Hispanics underestimate their Native American admixture proportions, whereas most Native Americans systematically underestimate their European ancestry (34).

Genetic ancestry (admixture) of a given individual or population can be estimated by using most genetic polymorphisms. However, ancestry informative markers (AIMs) are routinely used because the number of markers required to estimate ancestry is inversely proportional to the informativeness of the marker. AIMs are those genetic markers, usually SNPs, which exhibit high allele frequency differences between parental populations, i.e., African vs European. Using highly informative AIMs means that fewer markers are required to obtain robust ancestry estimations, which also means lower genotyping costs. One measure of ancestral informativeness of a specific polymorphism is  $\Delta$  ( $\delta$ ), the absolute difference in allele frequency between two ancestral populations. A  $\delta$  value of 1 implies complete ancestry informativeness and a  $\delta$  value of 0 implies no informativeness for ancestry. Most markers are only informative for one pair of ancestral populations, whereas some are informative for more than one pair and, in general, a  $\delta$  value of  $> 0.5$  is considered as highly informative for ancestry (35). Even though  $\delta$  is the most obvious measure for ancestral informativeness, other measures such as  $1(n)$ , and Fisher's information content have been used (36-38). In general,  $F_r$  and  $1(n)$  are slightly more accurate methods of ranking markers than  $\delta$  (37).

Several statistical methods have been proposed to estimate individual admixture proportions using different maximum likelihood, Bayesian, and principal component (PC) approaches. According to

some studies, the method selected has a relatively small impact on the accuracy of individual admixture estimates (39, 40). By far, the most important factor in determining accuracy of the admixture estimate appears to be the number of markers used to estimate admixture and their informativeness. Even though they have been used in population genetics for decades, PC-based estimations are more widely applied recently to many large-scale dense genotype datasets, especially ones in which the variation in ancestry may be difficult to ascertain through other methods (41-43).

In the United States, a significant proportion of the population consists of admixed populations. Therefore, categorical classifications are likely to misrepresent the rich genetic variation that exists within these populations. For example, in the 2000 U.S. Census, 48% of Hispanics self-identified as White, 2% as African/African-American, 1% as American Indian, and 42% as 'Some Other Race' (44). As illustrated in Fig. 2(a), we can see an example of genetic ancestry estimated in Puerto Ricans, who are considered to be a Latino ethnic group. In this case, individuals self-identified themselves and their four grandparents as being 'pure' Puerto Ricans. In contrast with the homogeneity in self-identification, there is a remarkable genetic heterogeneity between individuals in the contributions of the different ancestral groups.

Given the continuum of African ancestry in African-Americans (Fig. 2(b)), it is surprising that remnants of the 'One-Drop Rule' still persist in the eyes of most Americans. The 'One-Drop Rule' defines a person as African-American with as little as a single drop of 'African blood', regardless of the origin of his or her other ancestors (45). This rule was historically implemented as a way to enlarge the slave population with the children of slave holders and it was maintained in the Jim Crow era to keep the *status quo* of social groups. From a social perspective, this 'One-Drop Rule' has encouraged racism but has also brought together the African-American community. Recently, Barack Obama was elected as president of the United States, a historic event. Although half of his ancestry is of European descent, media and general public opinion have 'unambiguously' classified him as the first African-American president. From a genetic point of view, there is no scientific justification to classify such a diverse population as a single and homogenous group. From a social point of view, there is likely a 'threshold' of ancestry in which all members of the population are class i within the category (i.e., President Obama). This social classification is likely to be contextual and specific to population and time/era. Measurement of genetic ancestry is the only method available so far to estimate the degree of African ancestry among

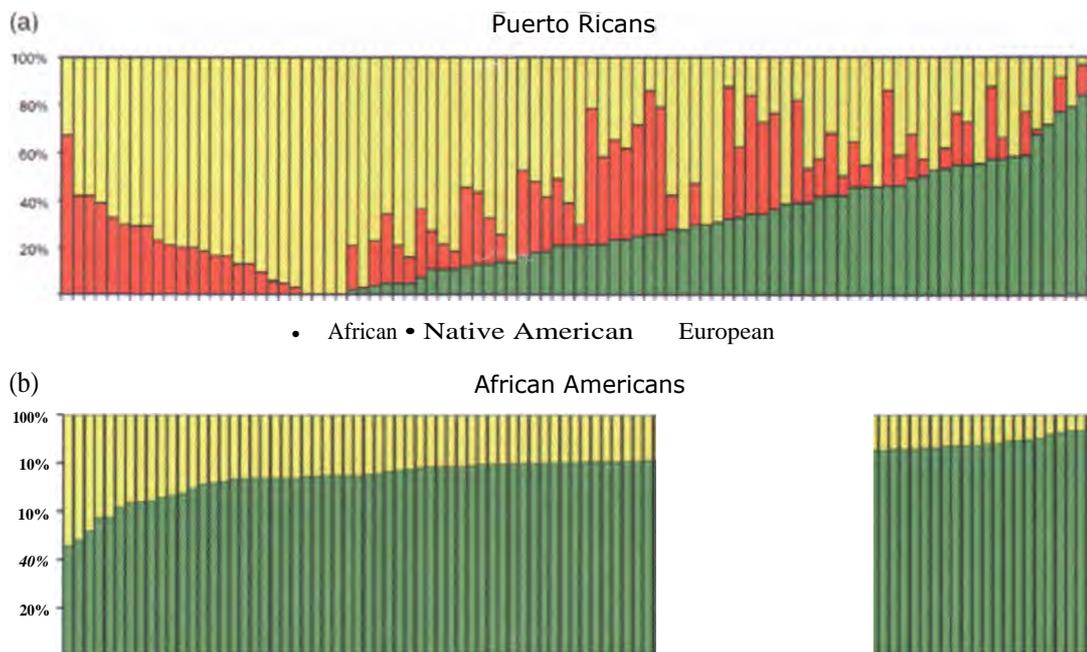


Fig. 2. Individual ancestry estimates for (a) 90 healthy Puerto Ricans and (b) 100 healthy African-Americans, clustered by admixture levels. Each individual is shown as a thin vertical line partitioned into different colored components representing inferred membership in the ancestral groups. Results are fully discussed by Choudhry et al. and Tsai et al. (61, 87).

African-Americans, as family genealogy and questionnaire data are not reliable predictors (33). We (also must be mindful of the fact that gene icily determined ancestry does not capture the social and cultural determinants that contribute to an individuals' affiliation with a particular racial or ethnic group.

### Ancestry and population differences In disease

It is well known that specific monogenic diseases (sickle cell anemia, Tay-Sachs, cystic fibrosis, etc.) differ between populations. It is also well known that many common and complex diseases, such as asthma, show significant disparities in prevalence, mortality and drug response among different populations. In the United States, asthma prevalence, morbidity and mortality are the highest in Puerto Ricans, African-Americans, Filipinos and Native Hawaiians, whereas it is the lowest in Mexicans and Koreans (46). Similar differences can also be observed for breast cancer incidence and severity, heart disease, or diabetes (47-49).

Categorical classifications missed the complexity behind these differences. For example, in the United States, Puerto Ricans and Mexicans have the highest and the lowest asthma prevalence, morbidity and mortality, respectively. This is paradoxical because both groups are classified as 'hispanic or Latino'. Differences in the contribution of the ancestral populations to the contemporary populations may in part underlie the ethnic-specific differences observed in the epidemiology of asthma. For example, increasing proportions of Native American ancestry have been associated with milder asthma among Mexican Americans (8). In a similar way, associations between several other diseases and genetic ancestry have been described. Among U.S. Latinas with breast cancer, higher European ancestry was significantly associated with increased breast cancer risk after adjustment for known risk factors and place of birth (50). In a very recent study among Puerto Ricans, African ancestry was negatively associated with type 2 diabetes and cardiovascular disease and positively correlated with hypertension (51).

However, such studies need to be mindful of the historical association between socioeconomic status and ancestry and the influence that this may still have on disease associations today. Neglecting to collect information on and to control for socioeconomic factors in studies of admixture may lead to associations between ancestry and disease phenotypes that are confounded by non-genetic factors (23). Even after adjustment for all

known confounders, it is important to interpret associations between ancestry and health-related outcomes with caution, because unmeasured environmental confounders may still explain the effect. Studies based on phenotype have shown that Puerto Ricans who self-identify as black have lower mean household income and were more likely to live below the poverty level when compared with those who self-identified as white (52). Moreover, racial reporting was a significant predictor of hourly wages for Puerto Rican men in New York City, even after controlling for language, disability, work experience, inner-city residence, presence of children, and industrial and occupational location (53). Similarly, among Mexican Americans those with dark skin/American Indian physical appearance are more likely to be discriminated against, receive less education and hold occupations with lower prestige than their light skin/European-appearance counterparts (54). This relationship also was observed with respect to earnings (55). Finally, it has been demonstrated that genetic ancestry interacts with socioeconomic status to confer differential risks for asthma among Puerto Ricans (56). Social factors such as discrimination, which may result in increased allostatic load, have been shown to be associated with worse physical and mental outcomes (57). Therefore, any association between disease phenotypes and ancestry may not be causal but rather a proxy for increased discrimination. Ultimately, if a difference in disease or health-related outcome is suspected to be at least partially because of genetic causes, it is important to consider interactions between genetic, social, and environmental factors.

### Genetic ancestry to identify genetic risk factors for disease

Case-control genetic association studies suffer from the potential problem of genetic confounding. Adjusting for population genetic structure in association testing is particularly important, because differences in population genetic structure between cases and controls can confound SNP-disease associations, leading to false-positive or false-negative findings (58-60). Population stratification is the result of admixture but refers to a phenomenon in which genetic confounding occurs as a result of this process, although some authors have used the term with slightly different meanings. Some reports have empirically assessed the effects of stratification in case-control studies in admixed populations (61, 62). However, while this issue

ight seem of remarkable importance only in recently admixed populations, population stratification has been identified in supposedly more homogenous populations such as Europeans, European Americans or even in Icelandics (43, 63-65).

Another method that implements the information provided by the estimates of genetic ancestry is admixture mapping. Admixture mapping is a genetic method to perform genome-wide association analysis to identify regions harboring population-specific risk alleles for a disease in admixed populations. The phenotypes that are of interest for admixture mapping are those which demonstrate differences among continental groups and which may not be explained by environmental factors, access to care, and other nongenetic factor differences between populations. Admixture mapping capitalizes on the fact that recently admixed populations are known to have large regions of LD (or genetic blocks) across genetic markers that are informative for ancestry (66, 67). Admixture mapping uses this increased LD to identify loci associated with complex disease phenotypes. The underlying premise behind admixture mapping is that if a marker increases the risk of disease and is found at a much higher frequency in one population (the high risk population), then that marker will also be found more frequently among cases. Furthermore, that marker will be in LD with other AIMs, which are specific to the high risk population and that this LD will be spread across large regions of the genome. By genotyping thousands of AIMs across the genome, one may be able to identify genomic regions in which the cases share ancestry from the high risk population more commonly than expected. Such loci presumably harbor disease-causing variants.

The ideal period of admixture for admixture mapping is approximately 5-20 generations (68). More remote admixture would mean that LD would have decayed and therefore would require any more markers. Conversely, more recent admixture (one to three generations ago) would mean that LD would extend too far to accurately localize a genomic region. Another limitation is the availability of genome-wide panels of markers specific of the population to be analyzed.

Admixture mapping is especially relevant in Latino and African-American populations because their admixture is relatively recent and this results in long-range LD (69, 70). A recent admixture mapping approach has estimated ancestry across the entire genome among African-American subjects with hypertension and healthy controls. These investigators identified two novel loci associated with hypertension on chromosomes 6 and 20,

which partially explained the excess African ancestry in subjects with hypertension compared with healthy controls (71). In another recent high-powered admixture scan, using 605 African-American cases and 1043 controls, revealed a locus on chromosome 1 that is significantly associated with multiple sclerosis (72). More recently, another study performed an admixture mapping study to characterize the genetic factors associated with the lower white blood cell count that African-Americans characteristically show (Fig. 3). In the figure, we could observe a significant increase in the African ancestry proportions in a region of chromosome 1 in two independent African-American cohorts (the Health ABC and the Jackson Heart Studies). The SNP with the strongest association was in the Duffy blood group antigen gene and it is known to eliminate expression of the Duffy blood group antigen (7). Admixture mapping has also identified a gene responsible for end stage renal disease in African-Americans (9, 73). In addition, admixture mapping helped to identify novel genetic variants for prostate cancer risk in African-Americans in a locus that had already been associated with prostate cancer in Caucasians (74). These promising results indicate a strong possibility for success in well-designed admixture mapping studies to identify genetic risk factors for complex traits.

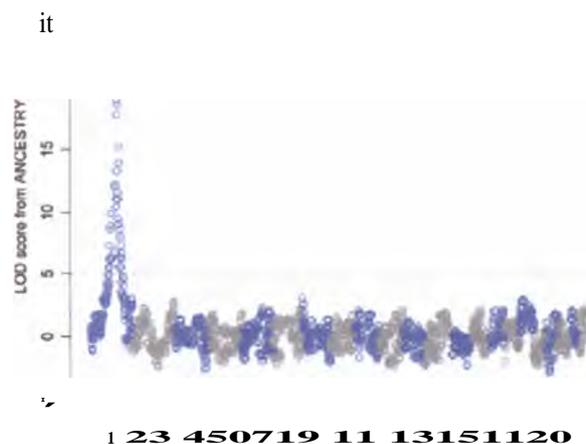


Fig. 3. Admixture mapping results (LOD Scores) of a case-control analysis of white blood cell count in African-Americans. Ancestry results calculated using ANCESTRYMAP (88) for initial genome-wide scan are plot for the Health ABC Study (diamonds) and the Jackson heart Study (circles) cohorts. Broken red line shows significance threshold. Results were presented by Nails et al. (7).

### Clinical applications of genetic ancestry

Despite the promising role that genetic ancestry plays in biomedical research, the debate still remains open about whether the profiling of individuals based on their population of origin has to be included in biomedical research or not. For example, should prescriptions take into account the 'racial' profile of patients? After all, BiDil<sup>®</sup> was the first Food and Drug Administration approved race-specific drug to be solely marketed to self-identified African-American patients with heart failure (5, 75). However, acting on rapid 'racial' assessments can lead to inappropriate treatments (30).

We speculate that, with the advent of high-throughput genotyping, we will bypass the need for self-identified categorical definitions of individuals recruited for genetic and pharmacogenetic testing. However, we believe that there are several milestones that we must achieve before we arrive at this point. First, genetic and pharmacogenetic studies will need better representation of diverse populations. Despite government initiatives to include different populations, most research is still done on populations of European origin. For example, all the recombinant factor VIII products available to treat hemophilia A correspond to the amino acid sequences present in Europeans, but ignore the ones in other populations and are responsible for a great proportion of alloantibody production among African-American patients (76). Second, future studies will be required to determine whether ancestry modifies genetic and pharmacogenetic effects. To date, we are only aware of one investigation that has demonstrated that ancestry modifies pharmacogenetic associations (77). Third, we are in the discovery phase of identifying novel genetic risk factors for disease and this has raised expectations for predicting risk (78). However, the statistics used for this discovery phase (odds ratios and p-values) are not appropriate for evaluating the predictive value of the genetic profiles of individual patients. Rather, we will need statistics that are more commonly used in clinical practice such as sensitivity, specificity, positive and negative predictive value (79, 80). Fourth, to make genetic and **pharmacogenetic** testing, a practical reality for everyone, the costs of genetic testing will need to be reduced. Otherwise, from a public health perspective, it can be easily argued that resources would be better spent on prevention at the population level rather than at the individual level. Finally, we will need to have in place appropriate ethical, legal and social guidelines to avoid stigmatization of specific members of a given

population. For example, studying health disparities and identifying genetic variation responsible for disease using population groups can lead to the 'racialization' of disease, irrevocably linking a disease state to a particular group (81). This would ultimately lead to the stigmatization of the members of that group and decreased information, surveillance, and access to treatment to other groups (81, 82). This all can lead to the overemphasis of the magnitude of genetic differences between populations and the overemphasis of the role of genetics as the basis for health disparities.

### Personal applications of genetic ancestry

Several for-profit companies are now providing ancestry testing directly to consumers. To the general lay public, these companies may *prima facie* be offering the same service. However, depending on the genetic testing performed, they may be offering completely different services. Some companies will only analyze mitochondrial DNA (mtDNA) or Y-chromosomal DNA, which represent the maternal lineage or the paternal lineage, respectively, while others will analyze autosomal DNA that provides an average estimate of the ancestry of all the lineages (83). Each method provides unique data and each has its own limitations. For example, the use of only mitochondrial or Y-chromosomal markers will only provide information about one lineage when in reality there are thousands of lineages that contribute to contemporary populations (84). Over the last 10 generations (dating back 200-250 years in the past), any individual has a total of 1024 different ancestors. Uniparental tests (Y or mitochondrial) provide a customer with a description of a geographical (or even ethnic or tribal) group that can be different depending on the samples in a company's reference database. For autosomal markers, the other source of variation in the results, besides the information in the company's database, is the number of markers tested which positively correlates with the statistical accuracy of ancestry estimates. In the best case scenario, genetic markers can only quantify the different continental contributions to a given individual's genome. These limitations are often unknown to the lay public and can give highly skewed or misleading results. For instance, these tests are not able to trace the ancestry of an individual to a single village in northern Europe or to prove kinship with Genghis Khan or the High Kings of Ireland just from the mitochondrial haplogroup. The sociological implications of such ancestry testing are complex and far beyond the

scope of this article. On the one hand, they may overstate, and even misinform, the precision of the estimates. On the other hand, they may provide some information to individuals whose ancestors are members of a diaspora and seek to learn more about their ancestors.

In addition to ancestry testing, many for profit companies are providing customers with genetic testing of potentially clinically important genetic risk factors and there are even 'risk tests' for specific diseases such as breast cancer. Although there is increasing popularity for commercial genomic profiles, recent evaluations found the scientific evidence to be insufficient to support their usefulness to measure genetic risk for diseases or for disease prevention (85, 86). These tests are also lacking the inclusion of environmental and dietary information to provide a good estimate of the real risk for complex diseases of their costumers and there are also concerns about the stability of the current risk estimates (78, 79). Furthermore, many of these risk factors are described on a population level and have been most often examined in populations of European decent and less so in African-Americans or Asians. I however, to our knowledge, no one has determined whether or not the strength of the risk associated with a given genotype varies with the percentage of individual ancestry of a given population. For example, among African-Americans, does percentage of African ancestry of a given individual modify the genotypic relative risk associated with the apolipoprotein E (APOE)

4 allele for Alzheimer's disease? We believe that this sort of question will become more common with increased globalization and inter-racial mixing of populations. This area of inquiry may be a new frontier for assessing genotypic relative risk profiles.

## Conclusions

The use of genetic ancestry estimations is a topic of growing importance in biomedical research. The awareness that population stratification is a potential confounding factor has been increasing among researchers. Currently, the inclusion of ancestry estimates has become essential when performing genetic association studies. In the era of high-throughput genotyping, genetic ancestry also has the potential to leave us with the dilemma of how best, if at all, to categorize individuals in biomedical research and clinical practice. Its more objective nature has made genetic ancestry a less polemic tool to approach the biological heterogeneity between populations. Consequently, there

is now a convergence emerging in the opinion of interdisciplinary groups and workshops about how to relate genetic variation to population-level differences in complex traits (27, 81). Epidemiologic research must broaden its reach to include all aspects of genetic, environmental and sociologic risk factors. Beyond genetics, research must include traditional environmental risk factors associated with disease (i.e., diet, age, gender, environmental exposures, family history) and the social dimension of individuals also has to be properly taken into account. As presented, factors such as socioeconomic status, educational level, racism and discrimination, access to healthcare, religion, language of use, immigration history, etc. have to be integrated in research. This paradigm shift in research will require the engagement of a broad range of specialists from different disciplines not only in the study design but also in the interpretation and discussion of results.

With the availability of genetic ancestry estimates, admixed populations represent a valuable opportunity to study complex diseases and drug response. Admixed groups, such as Latinos, African-Americans, or Cape Coloreds from South Africa, share varying proportions of different ancestral populations and their genetic complexity can potentially complicate biomedical research studies. On the other hand, precisely because of this complexity, admixed populations can also provide a unique opportunity to disentangle the clinical, social, environmental, and genetic underpinnings of population differences in health outcomes. Specifically, their mixed ancestry provides the intrinsic variability needed to untangle complex gene environment interactions that may help to explain the population differences in the epidemiology of complex diseases. A good example could be the striking disparities in asthma that are seen among different Latino groups and other populations.

The opportunities that genetic ancestry is offering to biomedical and clinical research are also opening new challenges to protect individuals from unethical uses of these technologies. The first indications for drugs based on population background are seen by some as the advent of the golden age of personalized medicine. In contrast, others will view this as opening Pandora's box to new forms of racism, discrimination, and population stigmatization based on pseudoscientific evidence. At the same time, the accessibility to inexpensive genotyping tools has allowed the emergence of a whole array of 'for profit companies' offering personalized genetic services to customers (79). The

products offered by these companies have limitations that are not well understood by most consumers. Again, a multidisciplinary approach has to be taken to set in place appropriate ethical, legal and social guidelines. Market forces can exert pressure to get additional customers, but strong science and reliable information are essential, especially when it is the consumers' health that is at stake.

The routine application of genetic ancestry estimation is a milestone to the future that we envision for biomedical research: the sequencing of the whole genome of individuals, adding social and environmental factors and being able to make predictions on overall (genetic and environmental) risk for common diseases, personalized diets, lifestyle recommendations, and drug treatments. It is important to note, however, that as we resolve some debates, new questions will emerge. In our view, the most interesting question to emerge from the development of ancestry testing is which is better for clinical outcomes, self-identified vs genetic ancestry?

#### Acknowledgements

E. G. B. has been supported by the National Institutes of Health (111 A)78885. A1077439.111,088133). the flight Attendant Medical Research Institute (FAMRI) and RWJF Amos Medical Faculty Development Award; E. Z. by the National Institutes of Health (CAI 20120. A623122) and Susan (3 Kometi Foundation for Breast Cancer Research (K(i080165); and M. V. by Beairiu de Pinos Postdoctoral Grant (2006 HP-A 10144). We would like to thank the Sandier Center for Basic Research in Asthma and the Sandler Family Supporting Foundation. We also thank Mattias Jakobsson, Michael A. Nails, and Noah A. Rosenberg for their support to illustrate this article.

#### References

1. Lander ES, Linton I.M, Birren B et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409 (6822): 860-921.
2. Siva N. 1000 Genomes project. *Nat Biotechnol* 2008; 26(3): 256.
3. Li J, Absher DM, Tang H et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; 319 (5866): 1100-1104.
4. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002; 298 (5602): 2381-2385.
5. Berdmid KC. Fixed-dose isosorbide dinitrate-hydroxyethylamine: race-based cardiovascular medicine benefit or mirage? *J Law Med Ethics* 2008; 36 (3): 458-463.
6. Ferrell PH Jr, McLeod HI, Caibamazepine. H1A-131502 and risk of Stevens-Johnson syndrome and toxic epidermal necrolysis: FDA recommendations. *Pharmacogenomics* 2008; 9 (10): 1543-1546.
7. Nalls MA, Wilson JF, Patterson NJ et al. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am J Hum Genet* 2008; 82 (1): 81-87.
8. Salad K, Choudhry S, Tang H et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genet Epidemiol* 2005; 29 (1): 76-86.
9. Kao WL, Kao M, Meoni LA et al. MY119 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet* 2008; 40 (10): 1185-1192.
10. Balaesque PL, Ballereau S, Jobling MA. Challenges in human genetic diversity: demographic history and adaptation. *Hum Mol Genet* 2007; 16 (Spec No. 2): R134-R139.
11. Relothford MI. Genetic evidence and the modern human origins debate. *Heredity* 2008; 100 (6): 555-563.
12. Jobling M, Hurles ME, Tyler-Smith C. *Human evolutionary genetics: origins, peoples and disease*. New York: Garland Science, 2004.
13. Jakobsson M, Scholz SW, Scheel PA et al. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 2008; 451 (7181): 998-1003.
14. Tishkoff SA, Reed PA, Friedlaender FR et al. The genetic structure and history of Africans and African Americans. *Science* 2009; 324 (5930): 1035-1044.
15. Clark AG, Hubisz MJ, Bustamante CD et al. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res* 2005; 15 (11): 1496-1502.
16. Bamshad M, Wooding S, Watkins WS et al. Human population genetic structure and inference of group membership. *Am J Hum Genet* 2003; 72 (3): 578-589.
17. Risch N, Burchard E, Ziv E et al. Categorization of humans in biomedical research: genes, race and disease [comment]. *Genome Biol* 2002; 3 (7): 2007.
18. Burchard EG, Ziv E, Coyle N et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003; 348 (12): 1170-1175.
19. Burchard EG, Borrell LN, Choudhry S et al. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am J Public Health* 2005; 95 (12): 2161-2168.
20. Anonymous. AAA statement on race. *Am Anthropol* 1998; 100: 712-713.
21. Anonymous. Genes, drugs and race. *Nat Genet* 2001; 29 (3): 239-240.
22. Schwartz RS. Racial profiling in medical research. *N Engl J Med* 2001; 344: 1392-1393.
23. Cooper RS, Kaufman IS, Ward R. Race and genomics. *N Engl J Med* 2003; 348 (12): 1166-1170.
24. Eschbach K, Snipp CM. Changes in racial identification and the educational attainment of American Indians. *Demography* 1998; 35: 3543.
25. Iltis S, Scott J, Elder JH. Racial self-categorization in adolescence: multiracial development and social pathways. *Child Dev* 2006; 77: 1298-1308.
26. Huxley I, Haddon AC. *We Europeans: a survey of racial problems*. New York: Harper, 1936.
27. Race, Ethnicity, and Genetics Working Group. The use of racial, ethnic, and ancestral categories in human genetics research. *Am Hum Genet* 2005; 77 (4): 519-532.
28. Sere D, Paabo S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res* 2004; 14 (9): 1679-1685.

29. Rosenberg NA, Mahajan S, Ramachandran S et al. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet* 2005; 1 (6): e70.
30. Braun L, Fausto-Sterling A, Fullwiley D et al. Racial categories in medical practice: how useful are they? *PLoS Med* 2007; 4 (9): e271.
31. Tang H, Quillen T, Rodriguez B et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 2005; 76 (2): 268-275.
32. Goldstein DR, Hirschhorn IN. In genetic control of disease, does race matter? *Nat Genet* 2004; 36 (12): 1243-1244.
33. Yaeger R, Avila-Bront A, Abdul K et al. Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiol Biomarkers Prev* 2008; 17 (6): 1329-1338.
34. Klimentidis YC, Miller GF, Shtiver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am J Phys Anthropol* 2008; 138 (4): 375-383.
35. Choudhry S, Taub M, McRae C et al. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. *Hum Genet* 2008; 123 (5): 455-468.
36. Weir B, Cockerham C. Estimating F-statistics for the analysis of population structure. *Evolution* 1984; 38: 1358-1370.
37. Rosenberg NA, Li LM, Ward R et al. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 2003; 73: 1402-1422.
38. Pfaff CL, Barnholz-Sloan I, Wagner JK, et al. Information on ancestry from genetic markers. *Genet Epidemiol* 2004; 26 (4): 305-315.
39. Tsai III, Choudhry S, Nasvi M et al. Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. *Hum Genet* 2005; 118 (3-4): 424-433.
40. Aldrich MC, Selvin S, Hansen HM et al. Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. *Am J Epidemiol* 2008; 168 (9): 1035-1046.
41. Price AL, Patterson NJ, Plenge RM et al. Principal component analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38 (8): 904-909.
42. Paschou P, Fejerman L, Burchard E et al. PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 2007; 3 (9): 1672-1686.
43. Novembre J, Johnson T, Bryc K et al. Genes mirror geography within Europe. *Nature* 2008; 456 (7218): 98-101.
44. United States Census 2000. Washington, DC: US Census Bureau. 2000.
45. Wright S. One drop of blood. New York: The New Yorker, 1994. Retrieved March 30, 2009: from <http://www.lawrencewright.com/laxt-dmp.html>.
46. Drake KA, Galanter JM, Burchard F.G. Race, ethnicity and social class and the complex etiologies of asthma. *Pharmacogenomics* 2008; 9 (4): 453-462.
47. Fejerman L, Fejerman E. Population differences in breast cancer severity. *Pharmacogenomics* 2008; 9 (3): 323-333.
48. Iltis DR, Benjamin MR, Rogers RG. Racial and ethnic disparities in health and mortality among the U.S. elderly population. In: Anderson NB, Iltis DR, eds. *Critical perspectives on racial and ethnic differences in health in later life*. Washington, DC: National Academy Press. 2004: 53-94.
49. Davis TM. Ethnic diversity in type 2 diabetes. *Diabet Med* 2008; 25 (Suppl. 2): 52-56.
50. Fejerman L, John EM, Huntsman S et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. *Cancer Res* 2008; 68 (23): 9723-9728.
51. Lai CQ, Tucker KL, Choudhry S et al. Population admixture associated with disease prevalence in the Boston Puerto Rican health study. *Hum Genet* 2009; 125 (2): 199-209.
52. Rodriguez CE. Racial Classification among Puerto Rican men and women in New York. *Hisp J Behav Sci* 1990; 12: 366-379.
53. Rodriguez C. The effect of race on Puerto Rican wages. In: Melendez E, Rodriguez C, Figueroa JR. *Hispanics in the labor force: issues and policies*. New York: Plenum Press, 1991: 77-98.
54. Arce CIT, M.E., Frisbie WP. Phenotype and life chances among Chicanos. *Hisp J Behav Sci* 1987; 9: 19-32.
55. Telles F. Phenotypic Discrimination and income differences among Mexican Americans. *Soc Sci Med* 1990; 71: 682-693.
56. Choudhry S, Burchard EG, Borrell LN et al. Ancestry-environment interactions and asthma risk among Puerto Ricans. *Am J Respir Crit Care Med* 2006; 174 (10): 1088-1093.
57. Bonnell I.N., Kiefe C.I., Williams DR et al. Self-reported health, perceived racial discrimination, and skin color in African Americans in the CARDIA study. *Soc Sci Med* 2006; 63 (6): 1415-1427.
58. Clayton DG, Walker NM, Smyth DJ et al. Population stratification, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005; 37: 1243-1246.
59. Freedman ML, Reich D, Penney KL et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004; 36: 388-393.
60. Marchini J, Cardon LR, Phillips MS et al. The effects of human population structure on large genetic association studies. *Nat Genet* 2004; 36: 512-517.
61. Choudhry S, Coyle NE, Tang H et al. Population stratification confounds genetic association studies among Latinos. *Hum Genet* 2006; 118 (5): 652-664.
62. Kim H, Hysi PG, Pawlikowska I et al. Population stratification in a case-control study of brain aneurysm in Latinos. *Neuroepidemiology* 2008; 31(4): 224-228.
63. Campbell CD, Ogburn EL, Lunetta KL et al. Demonstrating stratification in a European American population. *Nat Genet* 2005; 37: 868-872.
64. Uelgason A, Yngvadottir B, Lirafnelsson B et al. An Icelandic example of the impact of population structure on association studies. *Nat Genet* 2005; 37: 90-95.
65. Paschou P, Drincas P, Lewis I et al. Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genet* 2008; 4 (7): e1000114.
66. Chakraborty R, Kainbo M, Ferrell RE. 'Unique' alleles in admixed populations: a strategy for determining 'hereditary' population differences of disease frequencies. *Hum Hered* 1991; 1 (3): 245-256.
67. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations. by conditioning on parental admixture. *Am J Hum Genet* 1998; 63 (1): 241-251.
68. Smith MW, O'Brien SI. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet* 2005; 6 (8): 623-632.

## Genetic ancestry in biomedical research

69. Bonilla C, Pares RI, Pfaff CL, et al. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 2004; 68 (Pt 2): 139-153.
70. Smith MW, Patterson N, Lautimberger JA, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 2004; 74 (5): 1001-1013.
71. Zhu X, Luke A, Cooper RS, et al. Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 2005; 37 (2): 177-181.
72. Reich U, Patterson N, DeLager PL, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 2005; 37(10): 1113-1118.
73. Kopp IB, Smith MW, Nelson RW, et al. MY119 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* 2008; 40(10): 1175-1184.
74. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 2006; 103 (38): 14068-14073.
75. Tutton R, Smart A, Martin PA, et al. (Genotyping the future: scientists' expectations about race/ethnicity after Bill Clinton). *J Law Med Ethics* 2008; 36 (3): 464-470.
76. Vigneron KR, Amen A, Ahshiri TC, et al. Inhibitors of factor VIII in black patients with hemophilia. *N Engl J Med* 2009; 360 (16): 1618-1627.
77. Corvol H, DeGiacomno A, Eng C, et al. Genetic ancestry modifies pharmacogenetic gene-gene interaction for asthma. *Pharmacogenet Genomics* 2009; 19 (7): 489-496.
78. Kraft P, Hunter DJ. Genetic risk prediction - Are we there yet? *N Engl J Med* 2009; 360 (17): 1701-1701.
79. Hunter DJ, Khoury MJ, Eberly IM. Identifying the genome out of the bottle - will we get our wish? *N Engl J Med* 2008; 358 (2): 105-107.
80. Kraft P, Wacholder S, Cornelis MC, et al. Beyond odds ratios - communicating disease risk based on genetic profiles. *Nat Rev Genet* 2009; 10 (4): 264-269.
81. Caulfield T, Fullerton SM, Ali-Khan SE, et al. Race and ancestry in biomedical research: exploring the challenges. *Genome Med* 2009; 1: 8.
82. Brandt-Rauf SI, Raveis VII, Drummond NF, et al. Ashkenazi Jews and breast cancer: the consequences of linking ethnic identity to genetic disease. *Am J Public Health* 2006; 96(11): 1979-1988.
83. Bolnick DA, Fullwiley D, Duster T, et al. Genetics. The science and business of genetic ancestry testing. *Science* 2007; 318 (5849): 399-400.
84. Bandelt H-J, Yao YG, Richards MB, et al. The brave new era of human genetic testing. *Bioessays* 2008; 30 (11-12): 1246-1251.
85. Janasens AC, Gwinn M, Bradley LA, et al. A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 2008; 82 (3): 593-599.
86. Couzin J. Genetics. DNA test for breast cancer risk draws criticism. *Science* 2008; 322 (5900): 357.
87. Tsai H, Shaikh N, Kho JY, et al. Beta 2-adrenergic receptor polymorphisms: pharmacogenetic response to bronchodilator among African American asthmatics. *Hum Genet* 2006; 119 (5): 547-557.
88. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; 74 (5): 979-1000.