Individuals begin to interbreed (or ‘admix’) from populations with different frequencies of phenotypically relevant genetic variants, then these admixed individuals will exhibit the phenotype to varying degrees. The level of phenotypic expression will depend on the degree to which the admixed individuals have inherited causative variants that have descended from the ancestral population in which those variants were present (or, more likely, simply more frequent). We review studies that consider the association between the degree of admixture (or ancestry) and phenotypes of clinical relevance. We find a great deal of literature-based evidence for associations between the degree of admixture and phenotypic variation for a number of admixed populations and phenotypes, although not all this evidence is confirmatory. We also consider the implications of such associations for gene-mapping initiatives as well as general clinical epidemiology studies and medical practice. We end with some thoughts on the future of studies exploring phenotypic differences among admixed individuals as well as individuals with different ancestral backgrounds.

© 2014 S. Karger AG, Basel

Key Words
Admixture · Genetic ancestry · Principal components analysis · Metabolic disease · Cancer

Abstract
All human populations exhibit some level of genetic differentiation. This differentiation, or population stratification, has many interacting sources, including historical migrations, population isolation over time, genetic drift, and selection and adaptation. If differentiated populations remained isolated from each other over a long period of time such that there is no mating of individuals between those populations, then some level of global consanguinity within those populations will lead to the formation of gene pools that will become more and more distinct over time. Global genetic differentiation of this sort can lead to overt phenotypic differences between populations if phenotypically relevant variants either arise uniquely within those populations or begin to exhibit frequency differences across the populations. This can occur at the single variant level for monogenic phenotypes or at the level of aggregate variant frequency differences across the many loci that contribute to a phenotype with a multifactorial or polygenic basis. However, if individuals begin to interbreed (or ‘admix’) from populations with different frequencies of phenotypically relevant genetic variants, then these admixed individuals will exhibit the phenotype to varying degrees. The level of phenotypic expression will depend on the degree to which the admixed individuals have inherited causative variants that have descended from the ancestral population in which those variants were present (or, more likely, simply more frequent). We review studies that consider the association between the degree of admixture (or ancestry) and phenotypes of clinical relevance. We find a great deal of literature-based evidence for associations between the degree of admixture and phenotypic variation for a number of admixed populations and phenotypes, although not all this evidence is confirmatory. We also consider the implications of such associations for gene-mapping initiatives as well as general clinical epidemiology studies and medical practice. We end with some thoughts on the future of studies exploring phenotypic differences among admixed individuals as well as individuals with different ancestral backgrounds.

© 2014 S. Karger AG, Basel

Laura H. Goetz, MD, MPH
Department of Surgery
Scripps Clinic Medical Group
10666 North Torrey Pines Road, La Jolla, CA 92037 (USA)
E-Mail goetz.laura@scrippshealth.org

Prof. Nicholas J. Schork, PhD
Human Biology
J. Craig Venter Institute
4120 Torrey Pines Road, La Jolla, CA 92037 (USA)
E-Mail nschork@jcvi.org
Introduction

Human populations exhibit a significant degree of genetic and phenotypic variation. This variation occurs both among the individuals within a population and between individuals from different populations. The origins of the genetic and phenotypic differences between and within populations are complex and can be attributed to the combined effects of many phenomena, such as population isolation, genetic drift, selection and adaptation, and migration. Ultimately, and simply put, as populations become more and more isolated such that individuals do not mate across those populations, the resulting global consanguinity within those populations leads to an increased differentiation of the populations’ gene pools over time. Here we use the term ‘consanguinity’ in a general sense that, much like the term ‘inbreeding,’ encompasses the phenomenon of related individuals mating and producing offspring, but it is not confined to only matings involving very closely related individuals (e.g. siblings or cousins). This is consistent with general definitions of the term used in the population genetics literature [1]. In this light, although it has been shown that a greater proportion of global genetic variation can be attributed to differences among individuals within populations rather than to differences between individuals from separate populations [2–6], stratification or genetic differentiation between individuals from different populations is not negligible on a global scale and has important implications for gene mapping studies, public health campaigns, and clinical studies and practices [7–11].

One example of the problems created by the existence of between-population genetic and phenotypic differences involves genetic association studies. When association studies knowingly or unknowingly involve sampling individuals with a phenotype from one population and individuals without that phenotype from a different population, they are often plagued by false positive associations between the phenotype and any genetic variant that has frequency differences between those populations [11]. This ‘stratification’ issue can be overcome to some degree by assessing the ancestries or genetic backgrounds of the individuals in the study, although this strategy is not necessarily trivial or likely to work in all settings [11–14]. In addition, there are many phenotypes that have emerged in different populations due to different genetic variants or sets of causative variants arising in those populations (i.e. the genetic causes of these phenotypes manifest locus or allelic heterogeneity) [8]. Thus, it might be the case that two individuals, who present with the same condition but have different ancestral backgrounds, have different genetic bases for that condition.

These two problems, among others, are particularly vexing when attempts are made to characterize the genetic basis or architecture of a phenotype and then exploit the knowledge of that genetic basis in medical and health care settings. For example, a simple monogenic condition that is known to be higher in frequency in one population might still be found in individuals from another population, if the disorder results from allelic heterogeneity and certain variants at the offending locus are simply more frequent in one population rather than others. This is known to be the case for cystic fibrosis [15]. In addition, the identification of genetic variants that contribute to more complex multifactorial conditions with a polygenic genetic component may not be useful for predicting or diagnosing future disease incidence rates in another population because the variants identified in one population may not be as frequent in a different population. This has been shown to be the case for AIDS and type 2 diabetes (T2D) among other diseases [16, 17]. In this light, a true understanding, and subsequent accommodation, of the genetic and phenotypic differences among individuals from and between different populations can only occur if studies designed to specifically tease out those differences are conducted. Ultimate attention should be paid not only to whether those studies actually confirm a differential genetic basis for the phenotypes in question, but also whether they shed light on the impact of the differences between populations on clinical applications. This is particularly true for more complex multifactorial conditions with a polygenic component [16, 17].

A logical way to establish that genetic differentiation, or stratification, between populations influences a phenotype with a polygenic basis is to study individuals who are admixed between those populations (i.e. individuals who are the offspring or descendants of individuals from each of those populations) [18–20]. If the admixed individuals who exhibit the phenotype known to be more frequent in one of the populations harbor relevant variants also known to emanate from that population (i.e. are admixed and have inherited phenotypically relevant variants from that population), then the phenotype is likely attributable to multiple variations with differential frequencies between the populations. For quantitative phenotypes, such as blood pressure or body mass index (BMI), the assumption is that if two parental populations exhibit different average levels of a trait (i.e. one population tends to have higher blood pressure readings than the other), then admixed individuals deriving from mat-
ings involving ancestors from those populations will tend to have blood pressure readings closer in value to the average of one of the parental populations as a function of the percentage of the genome they have inherited from their ancestors that were from that population.

An important implication of the degree of admixture/phenotypic correlations is that admixed individuals are likely to complicate genetic and clinical assessments of their conditions, as noted above, since it might not be possible to exploit simple patient categorizations on the basis of their ancestry given that admixed individuals have, in fact, multiple ancestries. Rather, accurate assessments of the genetic causes of an admixed individual's disease will require the knowledge of both the potentially unique sets of variants that contribute to the disease in both populations and also of how those sets of variants might be functional in an admixed individual [21].

In addition, admixture studies are particularly important and relevant for contemporary society, since any historical population genetic differentiation and stratification that may have existed at one time is being broken down more and more over time due to greater access to transportation, a breaking down of social barriers that might have limited interbreeding and admixture in the past and globalization generally, all of which contribute to greater admixture on a global scale. We review the literature investigating the degree to which individuals that are admixed between two or more ancestral populations exhibit disease-related phenotypic features in a way that suggests the existence of both a polygenic basis for that phenotypic feature as well as population-level genetic stratification surrounding the genetic determinants of that feature. Such studies not only shed light on the degree to which genetic differences between populations explain disease frequency differences, but also help put into perspective modern conceptions about race and human genetic diversity and its impact on clinical care and public health practices. Ultimately, we find a great deal of evidence for associations between admixture and phenotypic expression in a variety of disease-specific settings, but not all this evidence is confirmatory across different studies.

The format for the rest of the paper is as follows. We first give a brief overview of how we approached the literature survey. We then discuss methodological issues surrounding the assessment of admixture and, importantly, the degree of admixture exhibited by admixed individuals. We then consider the results of studies relating the degree of admixture to different categories of disease phenotypes. We close with a brief discussion of the limitations and clinical implications of studies published to date.

Criteria for the Consideration of Publications

We performed database searches on MEDLINE and ‘Web of Science’ using both MeSH and keyword search terms that included ‘genetic admixture’, ‘genetic ancestry’, ‘disease’, ‘race’, ‘cancer’, ‘diabetes’, ‘cardiovascular disease’, ‘pulmonary disease’, and ‘obesity’. The initial review of articles was done independently by two of the authors. From relevant titles, article abstracts were reviewed to ensure that the assessment of individual genetic ancestry was included as a predictor (independent) variable and phenotypes associated with complex chronic diseases were taken as the primary outcome or dependent variables. We included both articles that adjusted for social determinants of health and those that did not. We excluded articles that reflected only genome-wide association studies and did not consider admixture, those studying admixture only to control for population stratification in association studies, and those using ancestry primarily for admixture mapping and identification of ancestry informative markers and not to directly explore the correlation between the overall degree of admixture and clinical phenotypes. We included studies investigating individual genetic ancestry as a predictor of complex chronic disease phenotypes. Based on these criteria, we ultimately identified a total of 50 articles for review. We then separated these 50 studies into the following broad disease categories: cardiovascular, pulmonary, cancer, and metabolic. Those articles not clearly falling into one of these categories were included under a ‘miscellaneous’ category. The articles were evaluated for the strength of association between the degree of ancestry and a disease phenotype as well as for their use of specific methodologies to infer genetic ancestry.

Statistical Methods for Ancestry Analysis

Although many methods for determining individuals’ geothnic ancestry or their degree of admixture have been devised to either reduce genomic confounding caused by population substructure in large genome-wide association studies [12, 22] or help identify cryptic or unexpected relatedness among individuals in an association study [23], a number of methods have been devised to estimate population diversity for its own sake or to help identify relationships between the degree of admixture and phenotypic variation [24–28]. Virtually all methods designed to estimate an individual’s geothnic ancestry involve interrogating genetic variants or ‘marker loci’ – often single nucleotide polymorphic loci – that either span the genome or are chosen based on their ability to distinguish populations. Since the majority of polymorphic markers vary in frequency (albeit possibly slightly) among major global populations, an individual’s genetic ancestry can be inferred with a combination of a sufficiently dense set (e.g., tens of thousands) of common polymorphic markers randomly positioned throughout the genome and appropriate statistical methods that can determine the population(s) from which an individual has most likely inherited those marker alleles [12, 22, 24–26]. To make ancestry inference more efficient and less reliant on costly genome-wide marker panels, many studies use a limited set of ancestry informative markers whose frequencies are known to differ appreciably among the source populations used to infer an individual’s admixture proportions (see table 4 in Winkler et al. [21]).
From a statistical inference perspective, two main approaches have been used to assess an individual’s ancestry. First, there are approaches that implement a ‘supervised’ clustering algorithm to define an a priori statistical model that prespecifies ancestral populations to be considered in the admixture analysis as well as marker allele frequencies in those populations. These methods thus assume some ‘model’ for the admixture observed in a population and then leverage likelihood and/or related techniques to compute an estimate of an individual’s admixture proportions based on the marker allele frequencies in those ancestral populations. These analyses typically maintain certain constraints, such as Hardy-Weinberg equilibrium of allele frequencies within the populations, when performing the calculations. Model-based approaches of this type are implemented in the popular programs Structure [29], Admixture [30], MALDsoft [31], and FRAPPE [18]. Since this approach relies on a priori information as to the number of source populations to be used in the admixture assessment, it might not result in reliable estimates of admixture and ancestry for an individual whose ancestry involves populations not prespecified in the model. Supervised model-based methods often also incur a high computational cost, but they are likely to become a standard, as more and more global populations are assessed for allele frequencies [11, 27, 28].

Second, as an alternative to supervised or model-based methods for inferring ancestry, there are methods that make no assumptions about the populations that may contribute to the admixture observed in a subpopulation. Many manifestations and extensions of principal components analysis (PCA) fall into this category [32, 33]. PCA is a statistical technique that seeks to identify independent axes or sources of variation across genotype data possessed by a group of individuals (i.e. principal components, PCs) whose origins are likely attributable to different ancestral populations. The relative amount of variation each component explains is used to guide the attribution of ancestral sources to any one individual’s genome based on his or her genotype data. In essence, this assumes that a PC reflects the variation in the individuals’ genetic profiles that is attributable mainly to (the degree of) the individuals’ origins with respect to two ancestral populations. Where each individual sits along the (PC) coordinates reflects the contribution of the two ancestral populations to the genotype data, and this can be used to estimate the degree to which each individual’s unique genetic profile reflects ancestry shared with those populations [13]. In actual datasets containing sufficiently dense genotype data collected on a relatively large number (e.g. hundreds) of unrelated individuals of varying ancestry, this approach is valid for at least a few PCs that explain the most variance [13]. However, the primary PCA component may actually reflect phenomena, such as genealogical or familial relatedness, linkage disequilibrium or stratification by a genotyping method, which often yield very similar results [22, 39]. However, based on extensive, comparisons, Lawson and Falush [20] concluded that model-based clustering approaches perform better on real data. As noted, however, model-based approaches also tend to be more computationally intensive than PCAs due to the need to estimate and optimize many parameters simultaneously for a given data set and hence could be supplanted with PCA methods. In addition, PCAs admit easy and intuitive graphical displays which often makes the assessment of ancestry more appealing when using them.

**Ancestry-Phenotype Associations**

In the following, we discuss the highlights abstracted from the 50 articles we identified that investigated individual genetic ancestry as a predictor of a disease or clinically relevant phenotype. The references reviewed are listed in tables for 5 categories: complex metabolic disease (table 1); cardiovascular disease (table 2); pulmonary disease (table 3); cancer (table 4), and miscellaneous phenotypes (table 5). These summary tables are divided into columns for types of ancestry, study population, specific phenotype, the number of study subjects, the method for determining ancestry, the number of markers used in the ancestry analysis, and the main findings of the paper. For types of ancestry, the nomenclature used in the tables corresponds to that used in the article, thus, for example, ‘Native American’ and ‘Amerindian’ both appear in the tables since different research groups have used these terms.

**Complex Metabolic Disease**

Over one third (17/50) of the studies reviewed explore the relationship between ancestry and complex metabolic diseases, such as T2D, dyslipidemia and obesity. This reflects the steadily increasing incidence rates of these illnesses in general, and their disproportionately higher rates in admixed individuals in the United States [40]. In 2011, the incidence of T2D was 12.4:1,000 in African Americans, 11.1:1,000 in Hispanics and 7.0:1,000 in Whites [41]. Obesity rates were 49.5% in African Americans, 39.1% in Hispanics and 34.3% in Whites [42]. Thus, elucidating potential genetically determined factors for these differences is of critical importance in predicting
# Published studies considering the relationship between degrees of admixture and complex metabolic disease phenotypes

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheng et al. [73], 2012</td>
<td>Af</td>
<td>US AA</td>
<td>T2D</td>
<td>7,021</td>
<td>Locus-specific-based method (AncestryMAP)</td>
<td>2,189</td>
<td>Increasing Af ancestry associated with higher T2D risk</td>
</tr>
<tr>
<td>Campbell et al. [46], 2012</td>
<td>NA</td>
<td>Colombian Mestizos</td>
<td>T2D</td>
<td>1,275</td>
<td>PCA (EIGENSTRAT, Lavase)</td>
<td>1,536</td>
<td>Increasing NA ancestry associated with T2D, but association disappears when accounting for socioeconomic status</td>
</tr>
<tr>
<td>Qu et al. [74], 2012</td>
<td>Am</td>
<td>US Mexicans</td>
<td>Metabolic traits</td>
<td>1,551</td>
<td>Model-based clustering method (Admixture)</td>
<td>103</td>
<td>Increasing Am ancestry associated with higher triglycerides in men, higher non-high-density lipoproteins in men, lower high-density lipoproteins in both genders, higher alanine aminotransferase in both genders and lower systolic blood pressure in females</td>
</tr>
<tr>
<td>Nassir et al. [69], 2012</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>Adiposity and adipose distribution</td>
<td>16,800</td>
<td>Model-based clustering method (Structure)</td>
<td>92</td>
<td>Increasing Af ancestry is associated with increasing BMI in AA Increasing Am ancestry is associated with increasing waist-to-height ratio but not BMI in H</td>
</tr>
<tr>
<td>Qi et al. [75], 2012</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>T2D risk</td>
<td>16,476</td>
<td>Model-based clustering method (Structure)</td>
<td>92</td>
<td>Increasing Af ancestry associated with increasing risks of T2D in AA Increasing Am ancestry associated with increasing risk of T2D in H</td>
</tr>
<tr>
<td>Lins et al. [72], 2012</td>
<td>Af, Na, Eu</td>
<td>Brazil</td>
<td>BMI and serum lipids</td>
<td>176</td>
<td>Maximum likelihood (Tsai et al., 2005) [101]</td>
<td>23</td>
<td>Increasing Eu ancestry associated with higher BMI and obesity Increasing Af ancestry associated with higher low-density lipoproteins, triglycerides and lipids, and lower very low-density lipoproteins Increasing NA ancestry associated with high triglycerides</td>
</tr>
<tr>
<td>Willig et al. [76], 2011</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>Aerobic fitness and BMI</td>
<td>232</td>
<td>Maximum likelihood (Hanis et al., 1996) [102]</td>
<td>140</td>
<td>Higher Af ancestry associated with lower aerobic fitness Higher Am ancestry associated with higher body fat</td>
</tr>
<tr>
<td>Signorello et al. [77], 2010</td>
<td>Af, Eu</td>
<td>US AA</td>
<td>Blood vitamin D levels</td>
<td>758</td>
<td>Model-based clustering method (Structure)</td>
<td>276</td>
<td>Increasing Af ancestry associated with decreasing blood vitamin D levels</td>
</tr>
<tr>
<td>Cheng et al. [78], 2010</td>
<td>Af, Eu</td>
<td>US AA</td>
<td>Obesity-related traits</td>
<td>3,531</td>
<td>Locus-specific-based method (AncestryMAP)</td>
<td>1,350</td>
<td>Increasing Eu ancestry associated with decreasing BMI</td>
</tr>
<tr>
<td>Lai et al. [50], 2009</td>
<td>Af, Am, Eu</td>
<td>Puerto Ricans</td>
<td>T2D and obesity</td>
<td>1,129</td>
<td>Model-based clustering method (Structure, IAE3CI)</td>
<td>100</td>
<td>Higher Af ancestry correlated with lower risk of T2D Obesity not correlated with Af or Am ancestry</td>
</tr>
<tr>
<td>Cheng et al. [79], 2009</td>
<td>Af, Eu</td>
<td>US AA</td>
<td>BMI</td>
<td>15,280</td>
<td>Locus-specific-based method (AncestryMAP)</td>
<td>1,411</td>
<td>Increasing Eu ancestry associated with decreasing BMI</td>
</tr>
<tr>
<td>Casazza et al. [80], 2009</td>
<td>Af</td>
<td>US AA</td>
<td>Insulin secretion</td>
<td>100</td>
<td>Maximum likelihood (Hanis et al., 1996) [102]</td>
<td>22</td>
<td>Increasing Af ancestry associated with increased first-phase insulin secretion in peripubertal children, association attenuated with higher adiposity</td>
</tr>
<tr>
<td>Florez et al. [45], 2009</td>
<td>Eu</td>
<td>Mexicans and Columbians</td>
<td>T2D</td>
<td>931</td>
<td>Model-based clustering method (mixture of binomials model) [12]</td>
<td>67</td>
<td>Higher E ancestry associated with lower risk of T2D, not significant after adjusting for socioeconomic status</td>
</tr>
<tr>
<td>Wassel et al. [81], 2007</td>
<td>Eu</td>
<td>US AA</td>
<td>Adipocytokines</td>
<td>1,241</td>
<td>Model-based clustering method (Structure)</td>
<td>35</td>
<td>Increasing Eu ancestry associated with higher levels of adipocytokines and with lower C-reactive protein</td>
</tr>
</tbody>
</table>
risk, designing better prevention initiatives and individualizing treatment modalities for these diseases.

While most studies did find an association between the degree of ancestry and the complex metabolic disease being investigated, there was no consensus as to the direction of the correlation. Increasing African-American, Amerindian/Hispanic or European ancestry was found to be either positively or negatively correlated with an increased risk for T2D and greater BMI. One of the factors that may explain these differences is that population admixture patterns differ among the studies, with some including groups with two prevalent genetic ancestries (e.g. African and European) and others with three components of genetic admixture (e.g. African, European and Amerindian). Also, the groups studied live in different areas of the USA and the world.

Given these discrepancies, it would be hasty to assign complete biological significance to associations between genetic ancestry and admixture and disease pathogenesis, since social determinants of health and gene-environment interactions cannot be overlooked. The importance of sociocultural factors that mediate disease have been highlighted in studies investigating the association between Native American (Amerindian) ancestry and the risk for T2D, a genetically complex and common disease. The prevalence of T2D is known to vary by socioeconomic status and is also affected by environmental factors, such as diet and lifestyle [43]. Any significant association found between, for example, greater levels of Native American ancestry and an increased risk for T2D is greatly attenuated after controlling for socioeconomic status [44–46]. Thus, studying the influence of genetically determined admixture and ancestry on disease without adjusting for social determinants of health, such as education level, household income, physical activity, and environmental factors, could potentially lead to erroneous conclusions.

Cardiovascular Disease

In the United States, African Americans have the highest incidence rates of hypertension, coronary artery disease and stroke, whereas Hispanics have lower rates than both African Americans and Whites [40]. This could be reflected in associations between the degree of admixture and these conditions. Table 2 provides the detailed results of our assessment of ancestry and cardiovascular disease. Greater levels of African ancestry were consistently shown to be associated with an increased risk of hypertension [47–50], which is in accordance with clinical epidemiological data, whereas an inverse association was found between Amerindian ancestry and the risk for hypertension [49]. Given the higher rate of coronary artery disease in African Americans, it is noteworthy that two papers [51, 52] identified an association between higher European, not African-American, ancestry and coronary artery calcifications in African-American individuals. Similarly, one paper identified an association between higher rates of African ancestry and a lower incidence of coronary vascular disease in US Puerto Ricans [50]. This discrepancy between the higher prevalence of coronary artery disease in African Americans but the negative association between higher African ancestry and coronary

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tang et al. [47], 2006</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>BMI</td>
<td>4,711</td>
<td>Model-based clustering method (Structure)</td>
<td>284</td>
<td>Higher Af ancestry correlated with higher BMI in AA Higher Am ancestry correlated with lower BMI in H</td>
</tr>
<tr>
<td>Parra et al. [44], 2004</td>
<td>Am</td>
<td>US H</td>
<td>T2D</td>
<td>446</td>
<td>Locus-specific-based method (AdmixMAP)</td>
<td>21</td>
<td>Increasing Am ancestry associated with higher risk for T2D, but attributed to confounding from low socioeconomic status</td>
</tr>
<tr>
<td>Fernandez et al. [82], 2003</td>
<td>Af</td>
<td>US AA</td>
<td>Obesity and resting metabolic rate</td>
<td>145</td>
<td>Maximum likelihood</td>
<td>18</td>
<td>Increasing Af ancestry associated with BMI, fat mass, fat-free mass, and bone mineral density, but not with resting metabolic rate</td>
</tr>
</tbody>
</table>

Af = African; Eu = European; Am = Amerindian; NA = Native American; US = United States; AA = African American; H = Hispanic.

A reference population was included when assessing ancestry.
Table 2. Published studies considering the relationship between degrees of admixture and cardiovascular conditions

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shahabi et al.</td>
<td>Af, Eu, NA</td>
<td>US Latinas</td>
<td>Hypertensive disorders of pregnancy</td>
<td>286</td>
<td>Model-based clustering method (Structure)</td>
<td>98</td>
<td>Increased odds of hypertensive disorders for increasing percentiles of Af ancestry</td>
</tr>
<tr>
<td>Divers et al.</td>
<td>Af, Eu</td>
<td>US AA</td>
<td>Coronary artery calcification plaques</td>
<td>1,040</td>
<td>Locus-specific-based method (regional admixture mapping [83])</td>
<td>–</td>
<td>Increased coronary artery calcification plaques in individuals with increased Eu ancestry</td>
</tr>
<tr>
<td>Smith et al.</td>
<td>Af, Eu</td>
<td>US AA</td>
<td>Prolonged QT* Interval</td>
<td>12,097</td>
<td></td>
<td>3,192</td>
<td>No association of ancestry with QT interval</td>
</tr>
<tr>
<td>Lutsey et al.</td>
<td>Af, Eu, NA</td>
<td>US AA, H</td>
<td>Plasma hemostatic factor</td>
<td>1,413</td>
<td>Model-based clustering method (Structure)</td>
<td>199</td>
<td>In AA, higher Af ancestry associated with higher fibrinogen and lower plasmin-antiplasmin levels In H, higher Af ancestry associated with higher fibrinogen and factor VIII levels</td>
</tr>
<tr>
<td>Wassel et al.</td>
<td>Af, NA</td>
<td>US AA, H</td>
<td>Large and small artery elasticity, pulse pressure</td>
<td>1,295</td>
<td>Model-based clustering method (Structure)</td>
<td>199</td>
<td>In AA, higher Eu ancestry associated with lower pulse pressure and marginally associated with higher large artery elasticity In H, higher NA ancestry associated with higher small artery elasticity</td>
</tr>
<tr>
<td>Kosoy et al.</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>Hypertension</td>
<td>17,118</td>
<td>Model-based clustering method (Structure)</td>
<td>92</td>
<td>Increasing Af ancestry associated with increasing blood pressure Increasing Am ancestry associated with lower blood pressure</td>
</tr>
<tr>
<td>Allison et al.</td>
<td>Af, Eu, NA</td>
<td>US AA, H</td>
<td>Lower extremity peripheral artery disease</td>
<td>1,417</td>
<td>Model-based clustering method (Structure)</td>
<td>199</td>
<td>Increasing NA ancestry associated with lower risk of lower extremity peripheral artery disease No association between increasing Eu ancestry and increased risk for lower extremity peripheral artery disease in AA or H</td>
</tr>
<tr>
<td>Marcus et al.</td>
<td>Af, Eu</td>
<td>US AA, H</td>
<td>Atrial fibrillation</td>
<td>19,784</td>
<td>PCA (Eigensoft, SmartPCA, AncestryMAP)</td>
<td>1,747</td>
<td>Increasing Eu ancestry associated with increasing risk for atrial fibrillation</td>
</tr>
<tr>
<td>Wassel et al.</td>
<td>Eu</td>
<td>US AA, H</td>
<td>Subclinical cardiovascular disease (coronary artery calcification and carotid intima media thickness)</td>
<td>1,417</td>
<td>Model-based clustering method (Structure)</td>
<td>199</td>
<td>Increasing Eu ancestry associated with higher coronary artery calcification in AA and H and with lower carotid intima media thickness in AA</td>
</tr>
<tr>
<td>Lai et al.</td>
<td>Af</td>
<td>US Puerto Ricans</td>
<td>Hypertension and cardiovascular disease</td>
<td>1,129</td>
<td>Model-based clustering method (Structure, IAE3CI)</td>
<td>100</td>
<td>Increasing Af ancestry associated with higher risk of hypertension and lower risk of cardiovascular disease</td>
</tr>
<tr>
<td>Tang et al.</td>
<td>Af, Am, Eu</td>
<td>US AA, H</td>
<td>Blood pressure</td>
<td>4,711</td>
<td>Model-based clustering method (Structure)</td>
<td>284</td>
<td>Increasing Af ancestry correlates with hypertension in both AA and H</td>
</tr>
</tbody>
</table>

Af = African; Am = Amerindian; Eu = European; NA = Native American; US = United States; AA = African American; H = Hispanic.

* The QT interval is the distance between the Q wave and the T wave on an electrocardiogram and can be an indicator of susceptibility for ventricular arrhythmia.
artery disease illustrates the most compelling need for a better understanding of the roles both the degree of admixture and socio-environmental factors play in the etiology of disease processes.

**Pulmonary Disease**

Pulmonary illnesses, such as asthma and chronic obstructive pulmonary disease, are largely attributable to environmental exposures, such as smoking cigarettes and poor air quality [53], which are in turn associated with social determinants of health, e.g., race/ethnicity and socioeconomic status. Cigarette smoking is more prevalent among African-American males than any other group and least prevalent among Hispanic women [40]. However, in addition to the socioeconomic factors, the degree of admixture may contribute to an individual’s susceptibility to ill effects of the environment and pulmonary disease progression or severity. Table 3 provides a listing of studies investigating the relationship between ancestry and pulmonary disease. Only 6 studies were identified. There is no consensus regarding the association between the degree of ancestry and incidence of pulmonary disease, with different studies reporting both positive and negative associations.

The study by Salari et al. [54] included Native-American ancestry in two populations of patients with asthma, and the associations were different for both populations: in Mexican Americans there was an inverse correlation between Amerindian ancestry and the severity of asthma, but in Puerto Ricans there was no association found with ancestry proportions. Interestingly, the prevalence of asthma in US Puerto Ricans, as reported by the CDC, is higher than in any other racial/ethnic group: 16.1%, as compared with 11.2% in African Americans, 7.7% in Whites and 5.4% in Mexican Americans [55]. Given these statistics, many factors other than ancestry are influencing the development of asthma as well as chronic obstructive pulmonary disease. A more careful assessment of the degree of ancestry and asthma in larger cohorts while controlling for environmental factors and other social determinants of health will further our understanding.

**Cancer**

The prevalence of different cancers can vary widely by race/ethnicity. Even different subtypes of the same cancer, such as certain very aggressive breast cancers, can have a higher prevalence in admixed individuals [56]. Sorting out what causative factors may be associated with an individual’s specific ancestral background could lead to the development of tailored treatment protocols based on ancestry. Table 4 lists studies exploring the relationship between cancer

---

Table 3. Published studies considering the relationship between degrees of admixture and pulmonary disease

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powell et al. [89], 2013</td>
<td>Af, As, E, NA</td>
<td>US AA, As, H</td>
<td>Lung function, chronic obstructive pulmonary disease</td>
<td>7,746</td>
<td>PCA, model-based clustering method (Admixture)</td>
<td>Not stated</td>
<td>No association between genetic ancestry components and lung function or severity of chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>Brehm et al. [90], 2012</td>
<td>Af</td>
<td>Puerto Ricans</td>
<td>Lung function</td>
<td>416</td>
<td>Locus-specific-based method (LAMP)</td>
<td>85,059</td>
<td>Increasing Af ancestry associated with decreased FEV₁ and FVC pre- and post-bronchodilator</td>
</tr>
<tr>
<td>Kumar et al. [91], 2010</td>
<td>Af</td>
<td>US AA</td>
<td>Lung function</td>
<td>2,169</td>
<td>Model-based clustering method (Structure, Admixture, Maximum likelihood)</td>
<td>Variable</td>
<td>Increasing Af ancestry associated with lower FEV₁ and lower FVC</td>
</tr>
<tr>
<td>Rumpel et al. [92], 2012</td>
<td>Af</td>
<td>US AA</td>
<td>Asthma, exacerbation</td>
<td>392</td>
<td>Model-based clustering method (Structure)</td>
<td>59</td>
<td>Increasing Af ancestry associated with increasingly severe asthma exacerbation in males but not females</td>
</tr>
<tr>
<td>Aldrich et al. [93], 2012</td>
<td>Af</td>
<td>US AA</td>
<td>Smoking/lung function interaction</td>
<td>1,281</td>
<td>Model-based clustering method (Structure)</td>
<td>1,332</td>
<td>Increasing Af ancestry associated with lower FEV₁ per pack-year of smoking</td>
</tr>
<tr>
<td>Salari et al. [54], 2005</td>
<td>Eu, NA</td>
<td>US H</td>
<td>Asthma severity</td>
<td>362</td>
<td>Maximum likelihood (IBGA)</td>
<td>44</td>
<td>Increasing NA ancestry associated with less severe asthma</td>
</tr>
</tbody>
</table>

Af = African; Am = Amerindian; As = Asian; E = European; NA = Native American; US = United States; AA = African American; H = Hispanic. FEV₁ = Forced expiratory volume in 1 s; FVC = forced vital capacity.
and admixture. We identified one study that did tailor treatment based on ancestry. This was done with Hispanic American children with acute lymphoblastic leukemia [57]. Essentially, greater levels of Native-American ancestry in the children were associated with an increased risk of relapse after treatment. By changing the standard treatment and adding additional chemotherapy for these children, this risk of relapse decreased. Other notable findings were a consistent association between greater levels of European ancestry and larger risk for breast cancer [58–60] and a protective effect against breast cancer with greater levels of Native-American ancestry [61]. Palmer et al. [62] reported an association between greater levels of African ancestry and certain breast cancer subtypes. Other cancer types studied, Table 4.

**Table 4. Published studies considering the relationship between degrees of admixture and cancers**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Types of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldrich et al. [94], 2013</td>
<td>A, Am</td>
<td>US AA, H</td>
<td>Lung cancer</td>
<td>368</td>
<td>Maximum likelihood</td>
<td>184</td>
<td>Increased Eu ancestry in H associated with increased risk of lung cancer  No association in AA</td>
</tr>
<tr>
<td>Gamazon et al. [95], 2013</td>
<td>A, As, E, NA</td>
<td>US AA, H, As</td>
<td>Neuroblastoma</td>
<td>2,709</td>
<td>PCA, locus-specific-based method (Eigenstrat, Admixture)</td>
<td>Not stated</td>
<td>Increasing Af ancestry associated with high-risk neuroblastoma and with lower survival</td>
</tr>
<tr>
<td>Palmer et al. [62], 2013</td>
<td>A, E</td>
<td>US AA</td>
<td>Breast cancer subtypes</td>
<td>646</td>
<td>Model-based clustering method (AdminMAP)</td>
<td>30</td>
<td>Higher degrees of Af ancestry associated with higher risk of ER-/PR- and triple-negative breast cancer</td>
</tr>
<tr>
<td>Fejerman et al. [58], 2012</td>
<td>E, NA</td>
<td>Mexicans</td>
<td>Breast cancer</td>
<td>1,881</td>
<td>LS (Hapmix, LAMP)</td>
<td>106</td>
<td>20% increase in risk for breast cancer per 25% increase in Eu ancestry</td>
</tr>
<tr>
<td>Slattery et al. [61], 2012</td>
<td>E, NA</td>
<td>US H</td>
<td>Breast cancer risk factors</td>
<td>7,775</td>
<td>Model-based clustering method (Structure)</td>
<td>104</td>
<td>Breast cancer risk decreased with increasing NA ancestry</td>
</tr>
<tr>
<td>Pereira et al. [96], 2012</td>
<td>A, E, NA</td>
<td>Peruvians</td>
<td>Gastric cancer</td>
<td>541</td>
<td>Model-based clustering method-ref (Structure)</td>
<td>103</td>
<td>Higher NA ancestry associated with higher risk of gastric cancer, but not after adjusting for socioeconomic status</td>
</tr>
<tr>
<td>Erdei et al. [97], 2011</td>
<td>A, E, NA</td>
<td>Puerto Ricans</td>
<td>Oral cancer</td>
<td>310</td>
<td>Model-based clustering method (Leadmix)</td>
<td>12</td>
<td>No association found</td>
</tr>
<tr>
<td>Yang et al. [57], 2011</td>
<td>A, As, E, NA</td>
<td>US AA, H, As</td>
<td>Acute lymphocytic leukemia</td>
<td>2,534</td>
<td>Model-based clustering method (Structure)</td>
<td>444,034</td>
<td>Higher NA ancestry associated with higher risk of relapse after treatment</td>
</tr>
<tr>
<td>Fejerman et al. [60], 2010</td>
<td>E</td>
<td>Mexicans</td>
<td>Breast cancer</td>
<td>1,938</td>
<td>Maximum likelihooda</td>
<td>106</td>
<td>Increasing Eu ancestry associated with increasing risk for breast cancer</td>
</tr>
<tr>
<td>Giri et al. [98], 2009</td>
<td>A, E</td>
<td>US AA</td>
<td>Prostate-specific antigen and prostate cancer</td>
<td>646</td>
<td>Model-based clustering method (Structure)</td>
<td>100</td>
<td>No association between prostate-specific antigen level and Af ancestry</td>
</tr>
<tr>
<td>Fejerman et al. [59], 2008</td>
<td>E, NA</td>
<td>US H</td>
<td>Breast cancer risk factors</td>
<td>1,013</td>
<td>Maximum likelihood</td>
<td>106</td>
<td>Increasing Eu ancestry associated with increasing risk of breast cancer</td>
</tr>
<tr>
<td>Ziv et al. [99], 2006</td>
<td>E, NA</td>
<td>US H</td>
<td>Breast cancer</td>
<td>563</td>
<td>Maximum likelihood</td>
<td>44</td>
<td>Higher NA ancestry associated with overweight, obesity and low hormone therapy use</td>
</tr>
</tbody>
</table>

A = African; Am = Amerindian; As = Asian; E = European; NA = Native American; US = United States; AA = African American; H = Hispanic.  
ER = Estrogen receptor; PR = progesterone receptor.  
a Reference population was included when assessing ancestry.
including prostate cancer, gastric cancer and oral cancer, were not found to be associated with genetic ancestry, despite these cancers having vastly different incidence rates in different racial/ethnic groups. Prostate cancer, for example, has an incidence of 192.9:100,000 in African Americans compared with 115.6:100,000 in Whites and 104.8:100,000 in Hispanics [63]. Thus, as with the other diseases discussed above, the role an individual’s genetic ancestry plays in carcinogenesis remains to be elucidated fully.

Discussion

We reviewed the literature describing the results of studies assessing the correlation between the degree of admixture exhibited by individuals from different global populations and clinically relevant phenotypes. We ultimately identified 50 articles that varied widely with respect to study design, study population(s), total number of participants or individuals assessed, genotyping methods used, analytic techniques used to determine the degree of admixture, and interpretation of results. Thus, although a number of associations have been reported in the literature, as our review makes clear, few truly definitive conclusions about the precise relationships between admixture and, for example, disease susceptibility can be drawn for any one disease. Despite this fact, it is important to recognize that our results are consistent with the notion that genetic background does indeed impact clinically relevant phenotypes, and this should motivate further research.

### Table 5. Published studies considering the relationship between degrees of admixture and miscellaneous clinical phenotypes

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of ancestry</th>
<th>Study population</th>
<th>Specific phenotype</th>
<th>Study subjects, n</th>
<th>Method for determining ancestry</th>
<th>Markers, n</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qi et al. [75], 2012</td>
<td>A, E</td>
<td>US AA</td>
<td>Risk of hysterectomy</td>
<td>10,439</td>
<td>Model-based clustering method (Structure)</td>
<td>92</td>
<td>Increased Af ancestry associated with increased risk of hysterectomy</td>
</tr>
<tr>
<td>Nassir et al. [100], 2012</td>
<td>A, Am, E</td>
<td>US AA, H</td>
<td>Risk of gallbladder surgery</td>
<td>15,461</td>
<td>Model-based clustering method (Structure)</td>
<td>92</td>
<td>For H, increasing Am ancestry associated with decreasing risk of gallbladder surgery; For AA no significant association between Af ancestry and risk of gallbladder surgery</td>
</tr>
<tr>
<td>Wise et al. [68], 2012</td>
<td>A, E</td>
<td>US AA</td>
<td>Uterine leiomyomata (fibroids)</td>
<td>2,453</td>
<td>Locus-specific-based method (AncestryMAP, AdmixMAP)</td>
<td>1,430</td>
<td>Increased Af ancestry associated with increased risk of fibroids</td>
</tr>
<tr>
<td>Eppelein et al. [66], 2011</td>
<td>A, E</td>
<td>US AA</td>
<td>Helicobacter pylori infection</td>
<td>686</td>
<td>Model-based clustering method (Structure)</td>
<td>276</td>
<td>Increasing Af ancestry associated with higher odds of Helicobacter pylori infection</td>
</tr>
<tr>
<td>Peralta et al. [65], 2010</td>
<td>A, E</td>
<td>US AA</td>
<td>Creatinine levels</td>
<td>1,623</td>
<td>Model-based clustering method (Structure, Frappe)</td>
<td>42</td>
<td>Increasing Af ancestry associated with increasing creatinine</td>
</tr>
<tr>
<td>Tsai et al. [67], 2009</td>
<td>A, E</td>
<td>US AA</td>
<td>Preterm birth</td>
<td>812</td>
<td>Model-based clustering method (Structure)</td>
<td>57</td>
<td>Increasing Af ancestry associated with increased risk of preterm birth</td>
</tr>
</tbody>
</table>

Af = African; Am = Amerindian; Eu = European; NA = Native American; US = United States; AA = African American; H = Hispanic.
The reasons for this are obvious and were alluded to in the Introduction section: the world simply has not reached equilibrium on a genetic basis. That is, despite increasing globalization efforts, more affordable and efficient methods of transportation and less stigma associated with marriages and matings between individuals on opposite sides of geographic and cultural borders, global populations remain genetically differentiated and isolated to some degree. This differentiation and isolation leads to global levels of consanguinity that create and preserve genetically mediated phenotypic differences between populations. These differences will then manifest themselves in unique ways in admixed individuals that are the offspring of matings involving individuals of different genetically distinct populations. In fact, the identification of relationships between the degree of admixture exhibited by individuals and the presence of a phenotype confirms the existence of genetic differentiation and global levels of consanguinity.

The studies we have reviewed also have important implications for not only general epidemiological studies, but also clinical research and practice. For example, by investigating individual ancestry and admixture using genetic markers rather than traditional constructs built off the concepts of race and ethnicity, the authors of the 50 studies we reviewed were able to focus on potentially biologically mediated risk factors of disease. Thus, they were able to avoid, to some degree, a reliance on the often ill-defined and more nuanced social and cultural constructs of race/ethnicity [71] that are used in historical studies of admixture and population-level differences in disease prevalence. This suggests that more certain diseases observed in individuals who are admixed or from different populations may have a more organic and less sociocultural component. In this light, the identification of biological causes of a disease that are consistent with an individual’s genetic ancestry may elucidate the unique, possibly subtly idiosyncratic, pathogenesis of that individual’s disease. We note that several studies listed in tables 1 and 2 looked at the relationship between ancestry and intermediate phenotypes or surrogate endpoints associated with disease, such as serum lipids or insulin secretion. Significant associations between ancestry and many of these phenotypes were identified and may shed light on connections between ancestry and biologically mediated risk factors for diseases with which they are associated. More research in this area needs to be pursued.

Not all of the 50 articles we reviewed controlled for socioeconomic status. In fact, it may be the case that the lack of control for socioeconomic and environmental variables could account, at least in part, for some of the discordant study results regarding associations of genetically determined ancestry proportions and complex phenotypes such as BMI and T2D risk in admixed populations [47, 72]. In the context of the potential confounding effects of socioeconomic status on the relationship between the degree of genetically determined admixture and disease, it is important to keep one very important phenomenon in mind: many studies found a graded response between the degree of genetically determined admixture and a disease phenotype (e.g. as between the degree of African ancestry and BMI). It is hard to argue that socioeconomic factors, or even diet and lifestyle, track along perfectly with the degree of genetically determined admixture. This suggests that although socioeconomic factors are likely contributors to disease phenotypes among admixed individuals, and hence need to be taken into consideration, they are likely to contribute to disease in tandem with genetic factors and not necessarily explain the entire association between the degree of admixture and disease.

There are other issues that need to be considered in studies investigating genetic admixture and disease. For example, many of the published studies we reviewed included a relatively small sample size and considered only a handful of globally admixed populations. The inclusion of more admixed groups with similar socioeconomic, cultural and environmental characteristics may allow for more accurate studies of genetic ancestry and clinical phenotypes. Thus, some of the apparent discrepancies found in the studies reviewed here could be plagued by these issues. Also, the exploration of what are thought to be similarly admixed populations might be problematic if the populations actually have different average levels of admixture. For example, this phenomenon might be the case in the analysis of T2D and Amerindian ancestry in Mexican Americans from Starr County, Texas, who predominantly share European and Amerindian ancestries [47], if compared to studies of populations with a higher admixture level, such as studies of Brazilians who predominantly share European, Amerindian and African ancestries [72].

Ultimately, our review of the literature is indeed consistent with the notion that admixture impacts clinically relevant phenotypes, but it may also raise more questions than it answers with respect to the best methodology to use, the best populations and phenotypes to study and the need for a better control of confounding factors. It is hoped that our efforts will motivate further research and help move the biomedical research community towards greater sensitivity to global issues in population health, as these issues undoubtedly impact health care in most urbanized centers in first-world countries.
References


Admixture and Clinical Phenotypic Variation

DOI: 10.1159/000362233

85


106 Willig AL, Hunter GR, Casazza K, Heimbuch DC, Beasley TM, Fernandez JR: Body fat and racial genetic admixture are associated with aerobic fitness levels in a multiethnic pediatric population. Obesity (Silver Spring) 2011;19:2222–2227.


