Latino Populations: A Unique Opportunity for the Study of Race, Genetics, and Social Environment in Epidemiological Research

Latinos are the largest minority population in the United States. Although usually classified as a single ethnic group by researchers, Latinos are heterogeneous from cultural, socioeconomic, and genetic perspectives.

From a cultural and social perspective, Latinos represent a wide variety of national origins and ethnic and cultural groups, with a full spectrum of social class. From a genetic perspective, Latinos are descended from indigenous American, European, and African populations.

We review the historical events that led to the formation of contemporary Latino populations and use results from recent genetic and clinical studies to illustrate the unique opportunity Latino groups offer for studying the interaction between racial, genetic, and environmental contributions to disease occurrence and drug response. (Am J Public Health. 2005;95:2161-2168. doi:10.2105/AJPH.2005.068668)

Latinos are the largest, youngest, and fastest-growing minority in the United States, accounting for 14% of the nation’s total population. Currently, 42.6 million Latinos reside in the United States, not including residents of Puerto Rico. By 2050, 25% of the US population is predicted to be Latino.1

The term “Hispanic” or “Latino” describes a population with a common cultural heritage and most often a common language, but it does not refer to race or a common ancestry. Although Latinos have been considered to be first and foremost an ethnic group, they represent a heterogeneous mix of Native American, European, and African ancestries.2 Therefore, they can self-identify as any race or mixed race as defined by the 2000 US Census.

In the context of the US Census, race is usually considered a fixed characteristic of the individual, linked to his or her genetic makeup, while ethnicity represents a broader construct based on cultural tradition, common history, language, religion, and often a shared genetic heritage. The 2000 US Census classification scheme, often used in biomedical research, includes 5 major racial groups: Black or African American, White, Asian, Native Hawaiian or Other Pacific Islander, and American Indian or Alaska Native. In general, this classification scheme emphasizes a person’s ancestry according to geographic region(s) of origin.3

Although the relative ancestral contributions to the contemporary Latino gene pool make each Latino national group unique, there is substantial overlap in ancestry. Individuals with similar proportions of ancestral admixture may therefore still belong to different Latino subgroups. Furthermore, most Latinos are uncertain of their precise ancestry and base their reported ancestry on the national origin of their family and on their physical appearance. The unavailability of self-report of ancestry and the genetic complexity among Latinos may complicate biomedical research studies in this population.

On the other hand, precisely because of this complexity, Latinos also present a unique opportunity to disentangle the clinical, social, environmental, and genetic underpinnings of population differences in health outcomes. We describe the historical events that led to the formation of contemporary Latino populations. We also use Latino-based examples of clinical, social, and genetic research to highlight the importance and potential benefits, from the perspective of epidemiological research, of studying Latino populations.

Latin American founding populations

To understand the genetics of Latin American populations and relationship of that genetics to health, we must understand the diversity of this population’s 3 primary founding populations: indigenous Americans, Europeans, and Africans. Genetic and archeological data suggest that the first Americans arrived from Asia by way of the Bering Strait sometime between 30,000 and 100,000 years ago and that there were at least 3 major migrations to North, Central, and South America.4-6 It has been estimated that there may have been up to 75 million people living in North and South America prior to the arrival of Columbus.4 After 1492, Latin America and the Caribbean underwent an unprecedented mixing of the 3 populations.

European ancestry in Latin America originated in the Iberian Peninsula, where, before 1492, the population was extremely diverse, consisting of Iberians, Celts, Greeks, Romans, Sephardic Jews, Arabs, Gypsies, and other groups. The Spanish Inquisition resulted in the religious oppression of Jews and Muslims.7 In 1492, Spain’s Edict of Expulsion forced more than 200,000 Jews to choose between conversion to Christianity, death, or exile from...
Spanish territory. Jewish converts have been called Conversos (converts), Marranos (pigs), or Crypto-Jews. Immediately after the Edict of Expulsion, Columbus set sail on his first voyage. The early European founders of the New World reflected the diversity of the Iberian Peninsula. Although Jews were explicitly prohibited from emigrating to the New World, many Crypto-Jews fled with the colonizers, who were recruited by the Spanish Crown. Columbus’s accidental arrival in the Caribbean in 1492 was a defining event in the formation of Latino populations in the region. His arrival opened up the New World (Western Hemisphere) to Europe for trade and immigration, and subjected the indigenous peoples of the Americas to use as slave laborers.

Although the first Africans brought to the New World were Spanish-born African servants (Ladinos) who arrived with Columbus, the stimulus for the large-scale slave trade between Africa and the Americas evolved with the establishment of large plantation complexes and mines in the 16th century. Enslaved Africans came from 7 coastal regions. Africans were transported from interior regions (the hinterland) and brought to slave ports along the coast. Although the vast majority of African slaves were transported to what are now Colombia, Brazil, and the Caribbean, over 230,000 were also imported to Mexico, Chile, Argentina, and Bolivia between 1519 and 1650. However, the number of enslaved Africans transported by the Spanish are low compared with the tens of millions shipped to the New World by British, French, Dutch, and Portuguese contractors in subsequent years.

### THE NEW MESTIZO RACE

Although Spain tolerated intermarriage between Spanish Christians and Native Americans, the progeny of which were called Mestizos, there were separate laws for Spaniards, Native Americans, and Africans in the New World. Nevertheless, continued racial intermixing led to an increasingly stratified society and the creation of a caste system, called the Society of Castes (Sociedad de Castas), which was initially based on racial ancestry and physical appearance. The relegation of Africans and Native Americans to the lower social rungs of society led to African–Native American unions, the progeny of which were called Zambos.

The top tier of the “pigmentocracy” was reserved for Whites born in Europe. However, it became impossible to apply any universally valid criteria for classifying admixed individuals. A large number of people could therefore “pass” as belonging to a higher status by simply assuming a new status and moving to a new region. Finally, the caste system disintegrated with the wars of independence, the introduction of democratic principles, and the abolition of slavery in all countries in Spanish Latin America except Cuba.

Modern genetic studies have revealed complex patterns of ancestry in the former Spanish colonies. Analyses of Y chromosomes and mitochondrial DNA demonstrate that Native Americans are closely related to people from the border area between Mongolia and Siberian Russia, and these analyses confirm the common ancestral origin of contemporary indigenous populations in Latin America and the Caribbean. The precise European ethnicity of explorers to South America is unknown, as Spain consisted of an amalgam of ethnicities. However, some of the European ancestry may be traced back to unions between Crypto-Jews (Conversos) and Native American women. Studies of populations in Costa Rica and Colombia have demonstrated that the majority of Y chromosomes are consistent with Iberian or Sephardic origin, while most mitochondrial DNA is Native American. In contrast, the precise genetic origin of the enslaved Africans is obscured by the fact that many were classified according to their port of departure and not their true geographic origin.

The process of genetic mixing continues in contemporary Latin populations. Recently, there have been dramatic shifts in the “source country” of US immigrants, with more than half coming from Latin America. The top 10 Latino sources of immigrants to the United States are Mexico, El Salvador, Dominican Republic, Colombia, Guatemala, Peru, Cuba, Ecuador, Brazil, and Honduras. Puerto Ricans are US citizens by birth and not considered immigrants. Although, in the 2000 US Census, 97.9% of the non-Hispanic US population self-identified as one of the 5 major racial categories, 48% of Hispanics self-identified as White, 2% as African/African American, 1% as American Indian, and 42% as “some other race.” This demonstrates the complexity of self-identification in this group for epidemiological studies.

Such complexity does not mean that genetic analysis of this population is impossible. While self-report among Latinos is generally non-specific for determination of ancestry, genetic markers that provide information on ancestry (called ancestry informative markers, AIMs) and newly developed statistical methods are making genetic estimation of ancestry increasingly more accurate.

In the United States, there are significant disparities in socioeconomic status between Latinos and
Whites. However, of all major racial/ethnic groups, Latinos use fewer health care services than White non-Latino Americans and are less likely to have entered the health care system for any type of care. Moreover, access to health care is strongly influenced by the options available, and Latinos often work for employers who do not provide health insurance.

Nevertheless, Latino health profiles contradict many assumptions made about poor and underserved minority groups. Despite higher poverty rates, less formal education, and reduced access to health care, health outcomes of Mexicans living in the United States today are generally equal to, or better than, those of Whites and other minorities. In the United States, Latinos have lower cancer rates and reduced access to health care. By contrast, many Mexicans are recent or undocumented immigrants and therefore do not have access to these services, but they nevertheless still tend to have better health outcomes than Whites and other minorities. In addition, as globalization expands and US companies increase their presence and market share in Latin America, it will be imperative that researchers account for the adaptation of mainstream American diet, lifestyle, and culture in Latin American countries.

In the United States, Latinos have lower incidence and mortality rates from most common types of cancer (breast, prostate, lung and bronchus, colon and rectum) than non-Hispanic Whites. In contrast, rates are higher among Latinos for cancers of the stomach, liver, uterus, cervix, and gallbladder. Of the 3 major Latino groups, Puerto Ricans appear to have the highest age-adjusted mortality rates and Cuban Americans the lowest. The apparent Hispanic advantage in mortality holds regardless of gender and age. There is also considerable variation in cancer mortality rates across Latin American countries (Table 1) for as yet undetermined reasons.

Using 44 AIMs, Choudhry et al. found that among Puerto Ricans, ancestry is associated with socioeconomic status. Healthy Puerto Rican volunteers reporting “upper” socioeconomic status had 9.1% lower African ancestry and 9.2% higher European ancestry than healthy volunteers reporting “moderate” and “middle” socioeconomic status (P=.004 and .008, respectively). Other studies have shown that Puerto Ricans who self-identify as Black have lower mean household income and are more likely to live below the poverty level than those who self-identify as White. Moreover, racial reporting was a significant predictor of hourly wages for Puerto Rican men in New York City, even after those elements that might be interacting with race reporting (i.e., language, disability, work experience, inner-city residence, the presence of children, and industrial and occupational location) were taken into account. Similarly, among Mexican Americans, those with dark skin/American Indian appearance are more likely to be discriminated against, receive less education, and hold occupations with lower prestige than their counterparts with light skin/European appearance. This relationship also was observed for earnings. Thus, future research among Latino populations should complement genetic data with measures of racial identity, ancestry, acculturation, and socioeconomic status.

The primary bottleneck in dissecting the etiology of differences in health and disease experiences among racial/ethnic groups is that caused by confounding. Racial and ethnic groups differ from one another in terms of culture, socioeconomic status, levels of discrimination in work and housing, and genetic ancestry; many of these characteristics are interrelated, but not necessarily in a causal manner. Because Latino populations represent different admixtures of 3 major racial groups, it may be possible to begin to unravel some of the differences in

![Individual ancestry (IA) estimates for 96 healthy Puerto Ricans, clustered by admixture levels.](Image 1)
disease incidence and outcomes through modern genetic techniques and a variety of epidemiological study designs.

**Ecological Studies**

The first and simplest level of analysis involves ecological studies. Here, rates of disease and phenotypic characteristics are contrasted among a variety of Latino ethnic groups and assessed for correlation with group admixture levels. When a high level of a specific ancestry correlates with a particular disease, differences in ancestral group genetics could be responsible. One example is provided by systemic lupus erythematosus, a complex autoimmune disease that arises from genetic, hormonal, and environmental factors. Investigators from the Grupo Latinoamericano de Estudio del Lupus study conducted a prospective multinational cohort study of 1214 patients with systemic lupus erythematosus recruited from 9 Latin American countries. Racial groups were determined by self-report and were classified as White, Mestizo, African-Latin Americans, or Pure Amerindians.

After the researchers controlled for clinical and sociodemographic variables, including medical coverage, education, socioeconomic status, and country of origin, both Mestizo and African-Latin American ethnicities were significantly associated with higher probability of lymphopenia and renal disease. In addition, Mestizos and African-Latin Americans had more severe disease than Whites, as evidenced by higher frequencies of renal disease, pericarditis, and polyadenopathy. Similar associations were found in Mestizos from Mexico and Argentina, indicating that such associations may be independent of the country of origin. The similarities between these 2 Mestizo groups may be the result of ethnic-specific genetic risk factors that originated from common ancestral populations. An obvious candidate is African ancestry, which should be largest in the African-Latin American group, followed by Mestizos, and lowest in Whites and Native Americans. These results also highlight the importance of studying the association between ancestry and the endophenotypes of a complex disease.

Asthma is another complex disorder with well-documented environmental, socioeconomic, and genetic contributions. Among all US populations, asthma prevalence, morbidity, and mortality are highest and lowest among Puerto Ricans and Mexicans, respectively. In this case, the prevalence pattern is not consistent with a genetic hypothesis of increasing admixture from any one ancestral population.

**Migrant Studies**

Large migrations of various Latino groups to the United States over the last several generations facilitate migrant studies, which may provide important clues about disease etiology. For example, if a group migrating into a new country with a different rate of disease takes on the same rate of disease as the

---

**TABLE 1—Age-Adjusted Cancer Mortality Rates in Spanish- or Portuguese-Speaking Latin American and Caribbean Countries, by Site and Gender**

<table>
<thead>
<tr>
<th>Human Development Index</th>
<th>Deaths per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gastric Cancer</td>
</tr>
<tr>
<td>Guatemala</td>
<td>0.649</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>0.667</td>
</tr>
<tr>
<td>Honduras</td>
<td>0.672</td>
</tr>
<tr>
<td>Bolivia</td>
<td>0.681</td>
</tr>
<tr>
<td>El Salvador</td>
<td>0.72</td>
</tr>
<tr>
<td>Ecuador</td>
<td>0.735</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>0.738</td>
</tr>
<tr>
<td>Paraguay</td>
<td>0.751</td>
</tr>
<tr>
<td>Peru</td>
<td>0.752</td>
</tr>
<tr>
<td>Colombia</td>
<td>0.773</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.775</td>
</tr>
<tr>
<td>Venezuela</td>
<td>0.778</td>
</tr>
<tr>
<td>Panama</td>
<td>0.791</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.802</td>
</tr>
<tr>
<td>Cuba</td>
<td>0.809</td>
</tr>
<tr>
<td>Uruguay</td>
<td>0.833</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>0.834</td>
</tr>
<tr>
<td>Chile</td>
<td>0.839</td>
</tr>
<tr>
<td>Argentina</td>
<td>0.853</td>
</tr>
</tbody>
</table>

*The UN Human Development Index (HDI) is a composite index of life expectancy, literacy, and per capita gross domestic product that measures the socioeconomic development of a country. HDI values range from 0 (lowest living standards) to 1 (highest living standards).*
resident population in succeeding generations, it strongly suggests that environmental factors are responsible for the initial difference. Conversely, if the migrant group retains the same rate of disease as the country from which they migrated over several generations, a genetic difference may be responsible. Contrasting the migration effects of different Latino populations may also provide additional clues, especially regarding issues of genetic, socioeconomic, and cultural factors and their interactions. However, migrant studies need to be interpreted with caution, particularly among Latino populations, where the populations in the countries of origin are diverse and may have unique genetic and socioeconomic characteristics. Migrant groups may not reflect the general population of the country of origin, which results in a selection bias.

For example, Holguín and colleagues analyzed 2 independent national databases and confirmed that asthma prevalence is higher among US-born Mexicans than Mexican-born Mexicans. In this case, it might be concluded that such differences in disease prevalence arise purely from environmental factors. However, the racial ancestry of Mexican subjects participating in these studies was based on self-report, not genetic ancestry. Using 44 AIMs, we have demonstrated that in a population of self-identified Mexicans with asthma recruited from California and Mexico City, admixture proportions differ among US-born Mexicans, Mexican-born Mexicans currently living in California, and Mexicans currently living in Mexico City (Figure 2). Specifically, among our study participants, the proportion of Native American ancestry is higher among Mexican asthmatics living in Mexico City than among those living in California. Furthermore, Mexican asthmatics born in Mexico who then immigrate to the United States also have less Native American ancestry than Mexicans in Mexico City. Finally, the migrating populations may change over time because of socioeconomic and political forces.

Estimation of Individual Admixture and Regression Analysis

While ecological studies focus on group admixture levels, with the advent of ancestrally informative genetic markers, it is now possible to perform regression analyses of disease or trait frequencies on individual admixture estimates. For example, in a previous study, we used 44 AIMs to demonstrate that among Mexican asthmatics, Native American ancestry is associated with mild asthma whereas European ancestry is associated with severe asthma. In a case–control study, individual admixture levels can be contrasted between the cases and controls. 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 nongenetic factors. However, even after adjustment for all known confounders, it is important to interpret associations between ancestry and health-related outcome with caution, because unmeasured environmental confounders may still explain the effect. 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 gene–environment interactions.

There may also be significant differences among subgroups of the same population residing in different environments and geographic locations. For example, in a recent study of Puerto Rican asthmatics that used 44 AIMs, Choudhry et al. demonstrated that ancestry varied by clinic site even thought recruited subjects were homogenous based on self-reported ethnicity (Figure 3).

Genetically Similar Individuals of Different Ethnicities

From what has been discussed thus far, it is clear that much is to be lost if we classify Latinos only as a single ethnic group and that much is to be gained by further investigation of the genetic ancestry of groups of Latino national origin in different environments. Latino groups vary tremendously in terms of cultural influences according to national origin and level of acculturation. These cultural factors are in turn important in shaping attitudes toward health care. In addition, language, legal status, and socioeconomic status may be strong factors in determining access to care and disease-related outcomes. An important opportunity arises when ancestral admixture levels among different
try could indicate interaction with variant that coincide with ancestral risks associated with a genetic gene–environment and gene–opportunities for examining association, as well as provide studies could also strengthen the relations in such gene association Analysis of multiple Latino populations overlap. Comparisons of disease history among individuals with comparable admixture levels but membership in different national or cultural groups may also provide information on the relative importance of genetic versus sociological and cultural factors; migrant studies also provide such information, but they generally focus only on a single national group.

**Candidate Gene Studies and Admixture Mapping**

The strongest evidence for a genetic contribution to group differences would be the identification of a specific genetic variant associated with a disease outcome that varies in frequency across racial/ethnic groups.\(^3\) Analysis of multiple Latino populations in such gene association studies could also strengthen the association, as well as provide opportunities for examining gene–environment and gene–gene interactions. For example, risks associated with a genetic variant that coincide with ancestry could indicate interaction with genetic background. Latino populations may also offer a unique opportunity to dissect the genetic basis of complex traits. Because Latinos are known to be an admixed population, they may be an ideal population for “admixture mapping,” an approach that can efficiently identify genomic regions that underlie racial differences in disease.\(^3,4\) This approach uses the fact that admixed populations are known to have large regions of linkage disequilibrium (genetic blocks) across AIMS. If there are particular genetic variants that account for racial differences in disease susceptibility, then AIMS can be used to identify regions of the genome associated with a given trait. Admixture mapping may be particularly relevant in Latino subpopulations because their admixture is relatively recent, which results in very long-range linkage disequilibrium. Furthermore, compared with regression analyses using genome-wide estimates of individual admixture, estimates based on specific chromosome locations are more resistant to residual confounding.\(^5,5\)

**CONCLUSION**

From the perspective of clinical and genetic epidemiology, Latinos are a complex and potentially challenging population to study. The recent formation of this population through a complex admixing of ancestral populations has also been shaped by socioeconomic, sociopolitical, and geographic factors. Great demographic shifts of Latino population throughout the United States and Latin America have compounded the issues of mixed race, shared culture, unique environments, significant migrations, and continued socioeconomic and ethnic discrimination.

Latinos are not a homogeneous ethnic group, as there is great genetic diversity and socioeconomic, educational, and demographic variation both between and within Latino ethnic groups. However, Latinos living in the United States will often share a common language, immigration experience, and a culture with attitudes and values that often differ from those of the mainstream English-language culture of White non-Latinos. This diversity and similarity among Latinos provide a valuable opportunity to study the interactions of race, genetics, culture, and environment. By taking advantage of such diversity, we may gain a much more thorough understanding of disease, its causes, and its distribution that will benefit all.

**About the Authors**

Esteban González Burchard is with the Division of Pulmonary and Critical Care and Clinical Pharmacology, Department of Medicine, and the Division of Pharmaceutical Sciences and Pharmacogenetics, Department of Biopharmaceutical Sciences, University of California, San Francisco. González Burchard, Shueta Choudhry, Mariam Napi, Hue Ju Tsai, Elad Ziv, and Neil Risch are with the Center for Human Genetics, University of California, San Francisco. González Burchard, Eliseo Perez-Stable, and Ziv are also with the Medical Effectiveness Research Center for Diverse Populations, University of California, San Francisco. Luisa N. Borrell is with the Department of Epidemiology, Mailman School of Public Health and School of Dental and Oral Surgery, Columbia University, New York, NY. Jose R. Rodriguez-Santana is with the Pediatric Pulmonary Program of San Juan, San Juan, Puerto Rico. Rocio Chapela is with the Instituto Nacional de Enfermedades Respiratorias, Mexico City, Mexico. Scott D. Rogers and Jose F. Arena are with the Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Md. Bui M. is with Affymetrix, Inc, Santa Clara, Calif. William Rodriguez-Citron is with the Veterans Affairs Medical Center, University of Puerto Rico School of Medicine, San Juan. Rich Kittles is with Human Cancer Genetics, Comprehensive Cancer Center, Ohio State University, Columbus. Neil Risch is with the Department of Epidemiology and Biostatistics, University of California, San Francisco, and the Division of Research, Kaiser Permanente, Oakland, Calif.

Requests for reprints should be sent to Esteban Gonzalez Burchard, MD, University of California, San Francisco, CA 94143-0833 (e-mail: esteban@sfgh.ucsf.edu).

This article was accepted July 29, 2005.

**Contributors**

E. Gonzalez Burchard, J.R. Rodriguez-Santana, R. Chapela, and W. Rodriguez-Citron recruited and characterized subjects whose data were analyzed as part of this study. S. Choudhry performed all analyses related to ancestry. S.D. Rogers
and J. F. Arena provided additional epidemiological data on cancer in Latin America. R. Mei performed genotyping of samples in which GeneChip was used. R. Kittles provided background information on Africans in Latin America. E. G. Burchard, L. N. Borrell, N. Risch, M. Naqvi, S. Choudhury, E. Ziv, J. F. Arena, R. Kittles, S. D. Rogers, H.-J. Tsai, E. J. Perez-Stable, R. Mei, J. R. Rodriguez-Santamaria, R. Chapela, and W. Rodriguez-Clinton contributed to the writing of the article.

Acknowledgments

This work was supported by the National Institutes of Health grants K23 HL04464, HL07185, and GM63430, the American Lung Association of California, the Robert Wood Johnson Foundation, the National Center for Minority Health Disparities, the Extramural Clinical Research Loan Repayment Program for Individuals From Disadvantaged Backgrounds, 2001–2003 (to E.G.B), the National Cancer Institute (through Redes En Acción, grant U01 CA63117), the Sandler Center for Basic Research in Asthma, and the Sandler Family Supporting Foundation.

We dedicate this report to improvement in health care for all populations but especially those that heretofore have been underserved and understudied. The first author thanks his mother, wife, and children for their unwavering support of this and other projects.

References


41. Abrardo-Lanza AF, Chao MT,
