

The American Society of Human Genetics
Ancestry Testing Statement
November 13, 2008

Scope

Ancestry testing and ancestry estimation are utilized in a variety of settings. Ancestry testing is done on an individual basis, in an attempt to determine the ancestral origins or population(s) of origin for a person or family. Ancestry estimation is performed to infer biogeographical origins or admixtures of populations for research purposes. This document from the human genetics community focuses on issues pertaining to the assessment of genetic ancestry in both research and individual testing situations, the latter usually being performed in a commercial environment. We acknowledge that, in addition to these uses, genetic ancestry data are being utilized for other purposes. The forensic applications have drawn much attention, and along with other possible uses of these data, foster questions about privacy and the security of ancestry-related databases. It is yet unknown what the full potential of the applications and implications of genetic ancestry information might be, but The American Society of Human Genetics (ASHG) will continue to take a leadership role in discussions about the issues.

Ancestry Testing

Public interest in ancestry and genealogical research is increasing, and there has been growth in the number of direct-to-consumer (DTC) companies offering genetic ancestry analysis as a means of supplementing traditional genealogical research methods. This new wave of genetic services – currently provided by approximately 30 companies – raises a range of unique, as well as familiar issues related to the interpretation, application, and impact of genetic information.

A recent ASHG statement on DTC genetic testing acknowledged the prominence of commercial ancestry testing, but focused explicitly on tests that make health-related claims or that directly affect health care decision making (see [Hudson et al, 2007](#)). However, the Society believes that ancestry testing generally warrants independent consideration for the following reasons: 1) an increasing number of DTC genetic testing companies offer both ancestry and health-related genetic information; 2) the impact of ancestry testing on people, families, communities and societies traverses a wide range of psychosocial, ethical, legal, political and health-related issues;

and 3) many scientific and non-scientific challenges and implications of DTC ancestry testing are also present – and are not being adequately addressed – in the genetic and genomic research arenas from which it originated.

Ancestry Estimation

Ancestry can be assessed at a number of different levels. The concept of "ancestry" is least ambiguous when we speak of our closest ancestors such as our parents or grandparents, or when we speak of our most distant ancestors, such as the earliest hominids or the first modern *Homo sapiens*. Ancestry estimation has enormous value in human genetics research, illuminating patterns of past human migration and providing a background pattern of human genetic variation that is essential for inferences about the past action of natural selection and genetic disease association. Genetic ancestry assessment often addresses the intermediate levels of ancestry that are usually imprecisely defined and identified. It is exactly this intermediate level of ancestry, however, that may be especially informative for identification of the genetic basis for complex disease, as it provides a combination of advantages of pedigree analysis and association testing.

Many people pursue genetic ancestry testing because they wish to find out more information about either the local populations or broad geographical regions in which their ancestors lived. However, the power of commercial genetic tests to answer such questions is limited, and the precision of the answer is often limited by the imprecision of the question. The limitations arise from the fact that every person has hundreds of ancestors going back even a few centuries and thousands of ancestors in just a millennium. There is thus enormous non-deterministic variation to the portion of the genome retained in a descendant from a given ancestor, with a rough expectation that it halves every generation. Consequently, genetic tests can access only a fraction of these ancestral contributions. The genomic segments contributed by a particular ancestor are far from all being uniquely identifiable, so even if one's genome has those specific genome contributions, identification of particular ancestry is always uncertain and statistical. It is also unclear how well-inferred ancestry serves to predict the tested individual's genotypes at untested loci.

Subjectivity arises from the fact that geneticists make specific choices about which levels of ancestry to examine. For example, many estimations of genetic ancestry are designed to distinguish contributions from geographic regions which were prominent in colonial era population movements, especially as they affected the New World (e.g., West Africa, Europe, East Asia, and the Americas). This creates a bias that may lead us to define ancestry in reference to particular sociopolitical groups, rather than the wider range of demographic influences on our genome architecture or diversity.

Motivations for Assessing Ancestry

Consumers and scientists have different reasons for pursuing assessment of genetic ancestry, and these rationales, in turn, tend to influence how the genetic information is interpreted and applied.

Most consumers are interested in using genetic ancestry testing to confirm or extend their knowledge of family genealogy. Scientists offering these commercial services use Ancestry Informative Markers (AIMs, which are defined as showing higher than average allele frequency differences between particular human populations that are judged as appropriate ancestral populations in some specific setting), mitochondrial DNA (mtDNA, which is passed from mother to all children) markers, Y-chromosome (which is passed from father to son) markers, or increasingly, genome-wide single nucleotide polymorphisms (SNPs) to provide information on personal biogeographical ancestry, or maternal or paternal lineage.

In the research arena, population geneticists and anthropologists use these same technologies as used in DTC ancestry testing, but more often summarized on a population scale, to make inferences about demographic history and population relationship on the basis of genetic identity of groups.

Epidemiologists with an interest in identifying genetic associations with disease, in contrast, employ methods of ancestry inference either to control for complexities due to population stratification among cases and controls, or as an explicit strategy to map susceptibility variants that might be differentially distributed with respect to ancestry in recently admixed groups (such as African Americans or Hispanic Americans) through mapping by admixture

linkage disequilibrium (MALD). Epidemiological estimations of ancestry are typically subsequently applied to individuals and nearly always based on the analysis of genome-wide single nucleotide polymorphisms (SNPs) or AIMs. For epidemiological purposes, inference of cohesiveness of ancestral history is more relevant than is the specification of particular populations of ancestral origin.

Accuracy

Ideally, any quantitative claims about ancestry should have an easily interpreted assessment of confidence or accuracy associated with them. The accuracy of ancestry inference methods is a function of: 1) how underlying patterns of human genetic diversity are distributed among populations; 2) how that diversity is surveyed (i.e., which genetic markers are used and how many); 3) which populations are used as references; and 4) the statistical methods used to interpret patterns of variation. Perhaps the most important aspect of reporting confidence in ancestry determinations is to accurately convey the level of uncertainty in the interpretations and to convey the real meaning of that uncertainty.

There are already large and growing data sets describing the geographic pattern of variation of related lineages of the Y chromosome and of mitochondrial DNA. While it is now possible to identify related groups of Y chromosome and mtDNA lineages with very high accuracy, population-level inferences that have been made from these uniparental systems are substantially less accurate. Ancestry inferences made from multi-locus data (e.g., autosomal AIMs) provide a far more accurate estimate of total ancestry than uniparental systems, but even the best methods have limitations that are important to consider.

The underlying patterns of human genetic diversity determine how well ancestry inference could potentially perform. Accordingly, the accuracy of ancestry inference greatly depends on the reference database of populations available. Commercial scientists and private groups often have their own unpublished databases with the potential to provide more refined information than that available from publicly available resources. Yet, even the best databases reflect a woefully incomplete sampling of human genetic diversity, and this has important consequences for ancestry inference.

One problem is that the "ancestral populations" assumed by some methods are not explicitly represented in these databases – and indeed cannot be represented, because we do not have the ability to sample true ancestral populations. Instead, samples from a related population are used as a proxy. For example, present-day West Africans are the most frequently used proxy for inferring African American ancestry even though the African origins of African Americans are quite heterogeneous. A second problem is that oftentimes populations that are mixtures of the typical reference populations (e.g., Africans, Asians, and Europeans) are under-represented in most ancestry testing databases.

The accuracy of ancestry estimation also depends on the nature of the markers that are used and the statistical methods used to perform ancestry inference. Markers vary in terms of their power and informativeness, and methods vary with regards to the assumptions they make, how much of the information available in the genetic data is extracted, and how their statements about inference are summarized for the consumer or researcher receiving the information. A major concern about the DTC ancestry testing business is that there is no quality assurance guarantee, and there is not even a mechanism to couple market performance with anything relating to accuracy. Cost pressures and market competition will likely drive costs down, and lower costs for ancestry testing services will probably be tolerated in this environment even if the accuracy suffers.

Population genetic inference is ultimately a statistical exercise, and rarely can definitive conclusions about ancestry be made beyond the assessment of whether putative close relatives are or are not related. As a result, whenever ancestry inference moves beyond such simple questions it must rely on complex inference procedures that necessitate a fairly sophisticated understanding of probability to fully understand the level of uncertainty.

Health Implications

The relationship of genetic ancestry to individual and population health is still poorly understood by researchers, but an important emergent idea with social and political consequences. In the U.S. and elsewhere, “racial” and/or “ethnic” identity is often considered a key determinant of health. Yet the features of racial/ethnic identity that contribute to differential health outcomes

are frequently unclear and widely debated. “Race” might co-vary or correlate with different environmental or genetic risk factors, different interactions between genetic and environmental factors, or different combinations thereof. Therefore, differences in disease prevalence among racial groups may be weak predictors of the genetic differences that may be found in a particular person or group. Conversely, similar prevalence rates of disease among so called racial groups do not imply that genetic risk factors will be shared or are equivalent (identical) among people or groups.

There are circumstances in which the genetic factors influencing health-related traits are associated with specific genetic variations that tend to be more prevalent in a particular racial group, compared to the rest of the population. In this scenario, disease risk or treatment response is often purported to be associated with and, in some situations, influenced by genetic factors that vary among racial groups. Yet, it is unclear whether, or to what extent, such genetic risk factors explain variation in the prevalence of these diseases among these groups. Indeed, many racial/ethnic health disparities probably are only modestly affected by genetics, influenced more strongly instead by environmental factors such as differences in diet, education, and socioeconomic class, and inequities in access to and the provision of health care services.

Admixture mapping methods including MALD have been used successfully to identify some genomic regions associated with several health-related traits including prostate cancer, hypertension and white blood cell count. To date, however, inferences about ancestral populations have been extrapolated from a relatively small number of the world’s populations and sampled from a limited number of geographic regions, therefore the extent to which MALD will be useful for identifying population based genetic variants underlying health-related traits is not fully known. Numerous studies using MALD are underway, but even at its best, MALD is likely to be an effective strategy for only a small fraction of health-related traits, since genetic differences may not be the major cause of observed population differences in disease incidence. These limitations justify caution in the interpretation of data from these studies and in the clinical application of results from the related DTC genetic tests.

Personal and Societal Implications

Ancestry assessment – in both its research and personal applications – poses a host of political, legal, psychological, social and ethical issues. Anthropological and population genetics research that postulate or cast doubt on ancestral relationships has historically incited varying degrees of conflict.

For some groups (some Native American tribes, for example), a major concern about scientific efforts to explain origins is the apparent diminished regard for important cultural, religious, social, historical and political processes that also inform group origin, membership, and identity, and access to group rights. Some related issues include the use of genetic ancestry information as the basis for: changing one's identity on various government forms; making claims to certain group rights or benefits; and immigration purposes, such as seeking dual citizenship. These issues are of increasing practical concern and likely to become more so in the future.

Knowledge about genetic ancestry – if undesirable and unexpected – can elicit a range of psychological responses including shock, disbelief, denial, anxiety, anger, fear and other well-known reactions to unwanted news. It can also lead to the reshaping of individual or group identity. The occurrence of or potential for emotional distress in people and groups following receipt of conflicting information about their ancestry has been documented, but still needs more research.

The use of AIMS and admixture mapping techniques, in general, has brought about anxiety with regard to its apparent reification of race. Similarly, commercially available lineage tests and research on lineages often imply clear-cut connections between DNA and specific regions or ethnic groups. The treatment of ancestral groups as bounded biological entities increases the potential for stigmatization and/or discrimination of the groups and the people within them on the basis of traits, behaviors, diseases or other attributes.

Consideration of the ethical implications of ancestry estimation calls for an evaluation of scientific integrity, obligations and accountability, and benefits versus harms. The ever-present challenge and obligation of the scientific community engaged in this work is to refine existing

methodologies while effectively utilizing and communicating knowledge about the inherent uncertainties. In the commercial setting, accountability in this regard might be further compromised by various market pressures.

Recommendations

1. Because the science of ancestry determination has limitations, greater efforts are needed on the part of both industry and academia to make the limitations of ancestry estimation clearer to consumers, the scientific community, and the public at large. In turn, the public has the responsibility to avail themselves of information regarding ancestry testing and strive to better understand the implications and limitations of these assessments.
2. Additional research is required to further understand the extent to which the accuracy of genetic ancestry estimation is influenced by the individuals represented in existing databases, geographical patterns of human diversity, marker selection and statistical methods.
3. The complex consequences of ancestry estimation for people, families, and populations need to be assessed and guidelines should be developed to facilitate explanation and/or counseling about ancestry estimation in research, DTC and health care settings.
4. Scientists inferring genetic ancestry should consult or collaborate with scholars who have expertise in the historical, sociopolitical and cultural contexts needed to inform the processes and outcomes of their research and commercial efforts.
5. Mechanisms for greater accountability of the DTC ancestry testing industry should be explored.

Implementation of these recommendations is likely to have many benefits, including an improved understanding of human evolution and demographic history (an important story applicable to all humans), more accurate ancestry testing with quantifiable limits, better informed users of ancestry information, and the establishment of a framework for interpreting ancestry information in a culturally appropriate and socially sensitive manner.

Reference:

Hudson, K, Javitt, G, Burke, W, Byers, P with ASHG Comm. *Am. J. Hum. Gen.*, 2007;81:635-637. (Web URL: http://www.ashg.org/pdf/dtc_statement.pdf)