Information Brief

The common thread: What genetics tells us about ancestry, health, and personal identity

Key points

- Advances in genetics have led to better understanding of how genetic variation and environmental factors contribute to health and disease, and provide a new lens to look at ethnicity and ancestry.
- Populations that are geographically close together tend to have more genetic variants in common than those that are far apart.
- Most genetic variants are shared across all populations.
- Not everyone in an ethnic group will have a particular genetic variant.
- An individual will not necessarily carry the genetic variants of all their ancestors.
- Some populations have a high prevalence for certain conditions or diseases. However, not everyone in the population have the genetic variants for the disease, and members of other groups may also have the conditions, caused by the same or different variants.
- Populations of non-European descent are often underrepresented in medical studies, so genetic variants important to their health may remain undiscovered.
- Non-genetic factors, e.g., diet, living conditions, and life experiences, also have important effects on health outcome.
- There is some evidence that environmental and historical factors may have health effects that could be passed on to subsequent generations.

Useful terms

Genome – An individual's full set of genetic information, including all genes as well as other sections of DNA that may regulate when genes are turned on or off.

Genetic variant – One of the possible variations in DNA sequence at a particular location in the genome.

Population – In the field of population genetics, a "population" refers to a group of individuals who are more likely to have children with others within the group than outside the group.

Admixture – Genetic mixing of two previously isolated populations.

(Genetic) Ancestry – An estimation of the ancestral population(s) that contributed to an individual's genome.

Ethnicity / ethnic group – A group of people who identify with each other based on shared ancestry, language, culture, etc.

Epigenetics – There are many definitions of this word, but one of the most popular is: Changes in gene function that do not involve changes in DNA sequence.

Introduction

Ancestry and ethnicity are at the heart of human identity – how people see themselves fit into cultural and social groups. Throughout history, people have sought explanations for physical differences that they observe between those from their own community and those from other communities. Advances in genetics in the past few decades have given researchers new tools to investigate and understand the ancestry of different human populations. At the same time, many researchers are studying how genetic variation may explain differences in health, disease susceptibility, and response to medications. The scientific community is trying to make sure that genetics-based, personalized medicine can benefit everyone, while at the same time identify and address important non-genetic causes of health differences.

A genetics perspective of ancestry

The genetic analysis of ethnicity and ancestry is based on the identification of genetic markers or *variants* in our DNA. From one generation to the next, our DNA often acquires new variants as a result of normal biological processes in our cells. The majority of these variants are biologically "neutral," meaning they are neither helpful nor harmful to the organism. The variants may not even be located in sections of our genome that code for genes or have any biological function. Because, throughout much of history, humans have tended to (though by no means exclusively) have children with others in close geographic proximity or within their own social or cultural communities, populations that are geographically close together tend to have more genetic variants in common than those that are far apart. Importantly, most genetic variants (90% or more) are shared across all populations.

DNA ancestry testing: Promises and limitations

The rise of affordable, consumer-oriented genetic tests means that more and more people are using these tests to discover their own personal ancestry and connect with different communities. For some, especially those who may otherwise have no access to their ancestral lineage or history, learning about their ancestry from these tests gives them a sense of belonging and reconnection with their culture. However, users are not always aware of the nuances and technical limitations of these tests. And sometimes, the test results may not align with an individual's or a community's personal and cultural identity.

The accuracy of DNA ancestry tests depends on many factors. These include the genetic markers used – different companies base their analyses on different, proprietary lists of markers that may have varying degrees of accuracy or relevance. A test's accuracy also depends on the sizes of the reference ancestral groups to which an individual's genetic information is compared. The fewer samples in a reference group, the less accurate is the analysis — and many ethnicities are currently underrepresented in the databases of genetic testing companies. Finally, it is important to remember that the tests provide estimates based on statistics: because an individual's inheritance of genetic variants is by-and-large random, they will not carry the variants of all their ancestors. It is not expected that everyone in an ethnic group will have a particular genetic variant, nor are these variants always unique or exclusive to particular populations.

Ancestry and health

Do different human populations have genetic variations that lead to differences in health? As a result of the historical patterns of reproduction and migration, some populations have a particularly high prevalence for certain conditions or diseases. For example, conditions such as Tay-Sachs disease, hemophilia C, and Bloom syndrome are highly prevalent (at least historically) among Ashkenazi Jews. Thus, before members of these ethnic groups plan to have biological children, they frequently use genetic tests to check if both parents are "carriers" of the same genetic variants, which could result in a child born with such genetic conditions. However, as with any other trait that correlates with ancestry or ethnicity, it is important to note that these conditions do not define the ethnic group, i.e., not everyone in the group will have the genetic variants for the disease, and members of other ethnic groups may also have these conditions, caused by the same or different genetic variants. For example, Tay-Sachs is also prevalent among French Canadians, but due to genetic variants that are different from those in Ashkenazi Jews and not commonly found in European French people.

Another disease that is commonly seen to correlate with ancestry is sickle cell disease (SCD). Often considered a "black" disease, prevalence of SCD is high among people of west and central African descent, but actually not high among many east or southern African ethnicities. At the same time, SCD prevalence is also high in groups from around the Mediterranean, Middle East, and South Asia. (The reason for this distribution, which closely parallels that of malaria, may be because of the protective effect that sickling red blood cells provide against the malaria parasite.) Similarly, cystic fibrosis (CF) is often considered a "white" disease, given its prevalence among people of Northern European descent, but it does also occur in people of other ethnicities. Non-Europeans may thus suffer from delayed diagnosis and treatment if their physicians do not consider the possibility of CF based simply on the patient's ancestry. Because people of non-European descent are often underrepresented in medical studies, genetic variants important to their health may remain undiscovered.

The important role of the environment

Non-genetic factors in one's environment also have biological effects on health outcome. Differential access to healthcare (preventative care, diagnosis or treatment options), nutrition, life experience, etc., influence the rate and manifestation of health conditions, and how well individuals do when living with such conditions. Social and environmental factors also affect which and how genes are *expressed* (turned on and off) in our cells, exerting direct biological and health effects. Moreover, there is some evidence that environmental and historical factors, such as chronic stress or episodes of famine, potentially have health effects that could be passed on to subsequent generations, even if those descendants no longer experience the same stressing factors. The presence or causes of such "transgenerational inheritance" are still under active research, but experiments in animal models such as mice point to possible *epigenetic* mechanisms – the inheritance of traits beyond ("epi-") the level of changes in DNA sequence – that may explain these biological effects.

References and further reading

Rosenberg. (2011) A Population-Genetic Perspective on the Similarities and Differences among Worldwide Human Populations. *Human Biology* 83:659. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531797/

"Understanding genetic ancestry testing" https://www.ucl.ac.uk/mace-lab/debunking/understanding-testing

"Siblings Can Have Surprisingly Different DNA Ancestry. Here's Why" by Nicole Wetsman, *National Geographic* (2018). https://news.nationalgeographic.com/2018/03/dna-ancestry-test-siblings-different-results-genetics-science/

"How African Americans Use DNA Testing to Connect With Their Past" by Ed Yong, *The Atlantic* (2017). https://www.theatlantic.com/science/archive/2017/06/how-african-americans-use-dna-testing-to-connect-with-their-past/531834/

Martin *et al*. (2017) Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics* 100:635. https://www.sciencedirect.com/science/article/pii/S0002929717301076

Rotimi *et al*. (2010) Ancestry and Disease in the Age of Genomic Medicine. *New England Journal of Medicine* 363:1551. https://www.nejm.org/doi/full/10.1056/NEJMra0911564

Piel *et al*. (2010) Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications* 1:104. https://www.nature.com/articles/ncomms1104

"Ethnic Differences in Cystic Fibrosis Genetic Coding Not Addressed in Screening Tests for Nonwhite Patients" (2015) https://www.elsevier.com/about/press-releases/research-and-journals/ethnic-differences-in-cystic-fibrosis-genetic-coding-not-addressed-in-screening-tests-for-nonwhite-patients

"Epigenetics: The Controversial Science Behind Racial and Ethnic Health Disparities" by Janelle Ross, *The Atlantic* (2014). https://www.theatlantic.com/politics/archive/2014/03/epigenetics-the-controversial-science-behind-racial-and-ethnic-health-disparities/430749/

Sharma *et al*. (2016) Gene × Environment Determinants of Stress- and Anxiety-Related Disorders. *Annual Review of Psychology* 67:239. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5739029/