Use of group labels in personal genomics

Joanna L. Mountain, PhD
Consulting Scientist, 23andMe

Population labels at 23andMe: Geopolitical labels

What best describes your White or European ancestry? Please check all that apply.

- [ ] Western European (e.g. German, French, Swiss, Belgian, Dutch, Austrian)
- [ ] Southern European (e.g. Italian, Spanish, Portuguese, Greek, Croatian, Slovenian)
- [ ] Eastern European (e.g. Polish, Czech, Ukrainian, Hungarian, Romanian, Russian)
- [x] Northern European (e.g. British, Irish, Swedish, Finnish, Norwegian, Danish, Icelandic, Latvian)
- [ ] Other
- [ ] I'm not sure

Population labels at 23andMe: Cultural and linguistic Groups

Do any of the following cultural group labels describe the ancestry of your father's father? Please check all that apply.

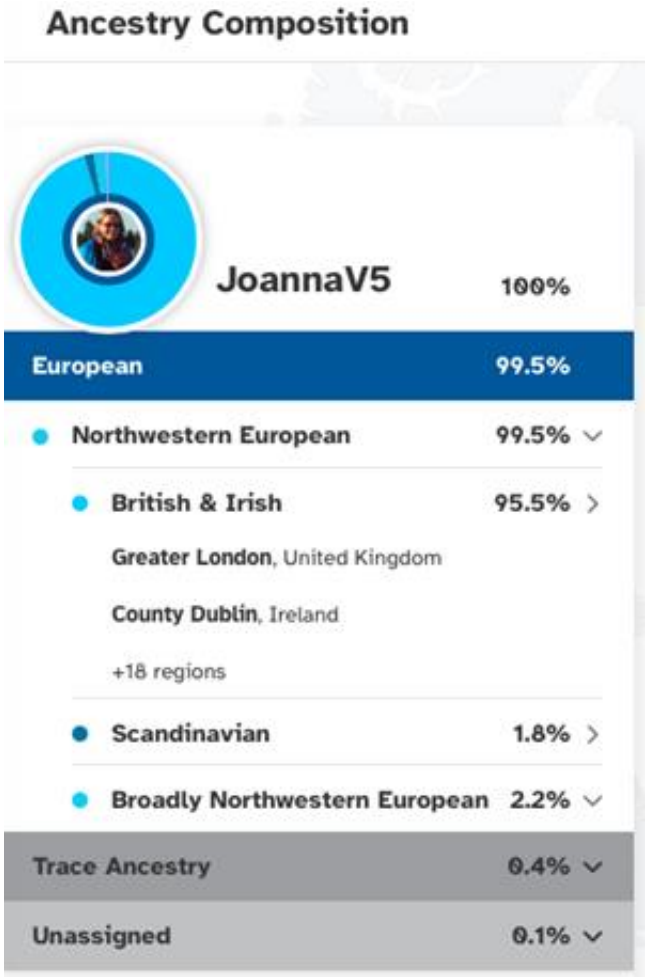
- [ ] Basque
- [ ] Jewish
- [ ] Romani
- [ ] Saami
- [x] None of the above
- [ ] I'm not sure

What was the language (sometimes referred to as the *mother tongue*) spoken in the childhood home of your mother's father? Please enter the name of the language.

If the language is not indigenous to the birth country of your mother's father or does not appear in the list, select 'other'.

Ancestry report populations:

my own results



Ancestry report populations at 23andMe: Beyond European

- Central & South Asian
- East Asian & Indigenous American
- Melanesian
- Sub-Saharan African
- Western Asian & North African

Indigenous American

Alaska

Columbia River Basin

Great Basin & Lower Colorado Basin

Great Lakes and Canada

Northeast

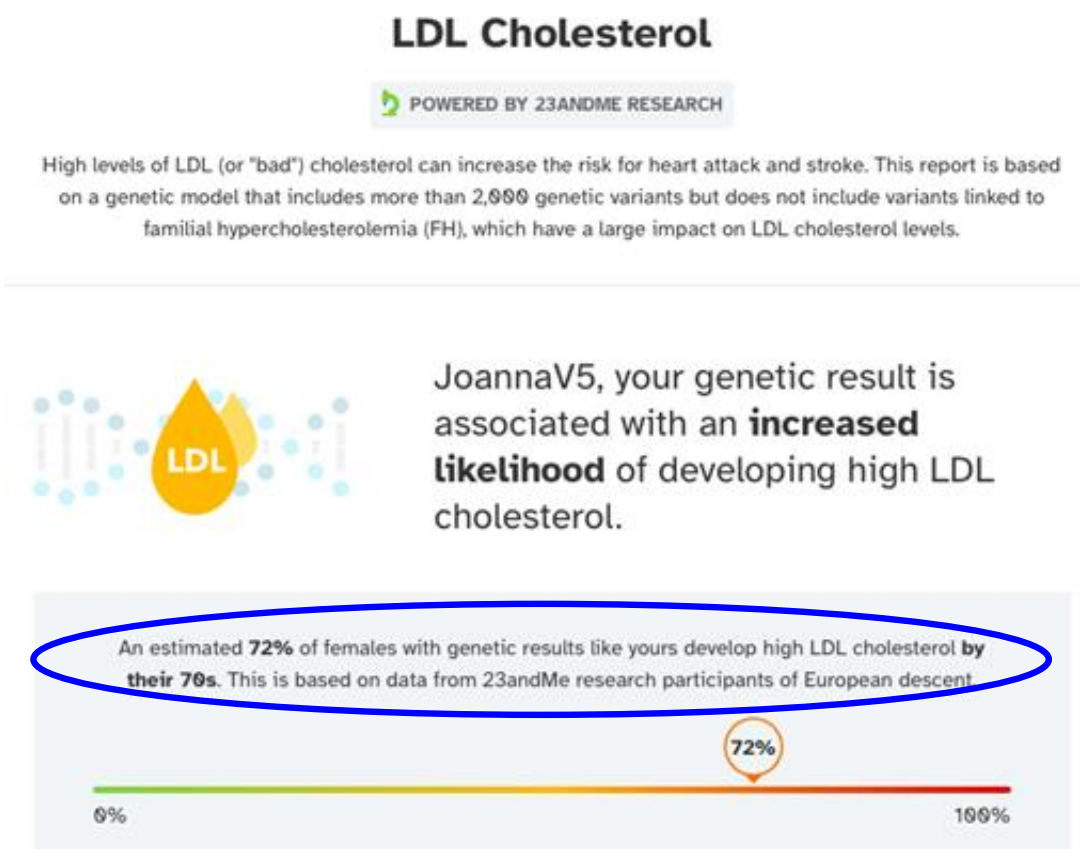
Plains

South Central

Southwest

Polygenic Risk Scores at 23andMe: my result

Methods
This report is based on a statistical model that takes into account your genetic results at **2,950** genetic markers, along with the **ethnicity** and **sex** you reported in your account settings, to estimate the likelihood of developing high LDL cholesterol.



Polygenic Risk Scores at 23andMe:"ancestries"

Methods for non-European ancestries

One of the biggest challenges for PGS today is transferability between ancestries (Martin et al., 2017). Individuals of European descent make up the overwhelming majority of genetics research participants even though they represent a minority of global genetic diversity (Popejoy & Fullerton, 2016). As a result, PGS trained with data from individuals of European descent typically perform worse among individuals of other ancestries.

https://customercare.23andme.com/hc/en-us/articles/115006037188-Navigating-and-Understanding-Health-Predisposition-Reports

Polygenic Risk Scores at 23andMe: “ethnicities”

Scientific validity across ethnicities

We verified that the model meets our scientific standards for individuals of European, Hispanic/Latino, East/Southeast Asian, South Asian, Sub-Saharan African/African American, and Northern African/Central & Western Asian descent.

How we may use ethnicity and sex to customize this result

- If you indicated in your account settings that you are of European, Hispanic/Latino, East/Southeast Asian, South Asian, Sub-Saharan African/African American, or Northern African/Central & Western Asian (Middle Eastern) descent, your result is tailored based on data from individuals of that ancestry.
- Otherwise, your result may be based on data from individuals of European descent because there is not enough data from individuals of your ancestry at this time. Data from individuals of European descent is used because the most data is available for this population.
- Your LDL Cholesterol result also takes into account the sex you indicated in your account settings.

See our white paper to learn more about the science behind this report.

Table 1: High LDL-C participant cohort descriptives

Platform	Ancestry Group	Sample Use	N	Age mean (SD)	Sex (% female)	High LDL-C Prevalence (%)
V1 to V5	European	GWAS	617,165	56.2 (13.8)	54.60%	41.99%
V5	European	Training the European Model	511,469	55.0 (13.9)	55.50%	40.60%
		Testing	56,749	55.1 (14.0)	55.24%	40.94%
	Sub-Saharan African/African American	Testing	18,710	50.1 (13.5)	59.02%	40.94%
	East/Southeast	Testing	18,357	44.7 (14.2)	57.51%	27.07%
	Hispanic/Latino	Testing	72,806	47.8 (14.0)	56.46%	33.86%
	South Asian	Testing	6,128	44.3 (13.0)	37.73%	34.48%
	Northern African/ Western Asian	Testing	5,267	49.4 (14.7)	40.38%	38.47%

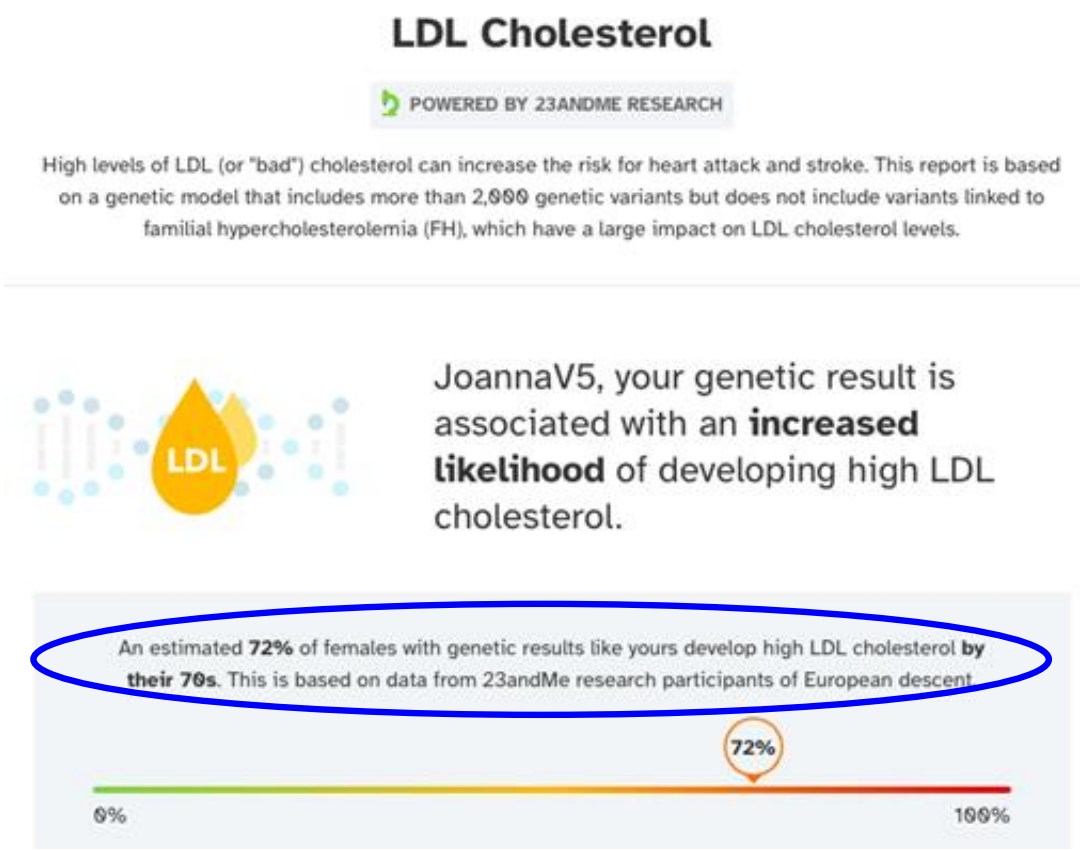
Table 2: High LDL-C PGS performance characteristics

Ancestry Group (test sets)	Full Model AUROC	Genetics Only AUROC	Odds Ratio top 5% versus average (95%CIs)	Odds Ratio top 5% versus bottom 5% (95%CIs)
European	0.7770	0.6456	2.81 (2.58 to 3.07)	10.24 (9.02 to 11.63)
Sub-Saharan African/African American	0.7312	0.5985	1.91 (1.67 to 2.23)	4.10 (3.34 to 5.05)
East/Southeast Asian	0.7635	0.5888	1.91 (1.64 to 2.22)	4.30 (3.43 to 5.39)
South Asian	0.7828	0.6222	2.69 (2.08 to 3.47)	7.75 (5.29 to 11.37)
Northern African/Western Asian	0.7776	0.6188	2.81 (2.13 to 3.72)	7.49 (5.04 to 11.14)

Polygenic Risk Scores at 23andMe

Methods

This report is based on a statistical model that takes into account your genetic results at **2,950** genetic markers, along with the **ethnicity** and **sex** you reported in your account settings, to estimate the likelihood of developing high LDL cholesterol.



Health Report Configuration at 23andMe: race/ethnicity

Report Configuration

COMPLETED

Genetic conditions can affect certain ethnicities more than others. Provide some information about who you are to help us personalize some of your reports.

What is your race / ethnicity? Please check all that apply.

- [ ] Native Hawaiian or Other Pacific Islander
- [x] White or European
- [ ] Middle Eastern
- [ ] Black, African, or African American
- [ ] Asian
- [x] American Indian or Alaska Native
- [ ] Other

Enter text

- [ ] I'm not sure

Are you Hispanic or Latino?

- ( ) Yes
- (•) No
- ( ) I'm not sure

Health Report Configuration at 23andMe: Cultural groups & geopolitical ancestry

Group labels at 23andMe: Indigenous Americans

What steps did our team take to mitigate potential harm to Indigenous sovereignties while providing additional geographic detail?

1. With guidance from experts including Indigenous scientists, we created an educational page with answers to common questions about this update, and information about how 23andMe defines Indigenous American genetic ancestry. This page is easily discoverable to customers through their results, and it is also viewable by the public.
2. In providing these new regions, and following recommendations by advisors, we did not use information about Tribe or Indigenous Nation affiliation to curate the genetic groups. Furthermore, because cultural, Tribe, Nation, or ethnic affiliation was not used to curate these groups — and because Indigenous peoples are not discrete genetic groups — each of the eight regions likely reflect ancestry from many Tribes or First Nations.
3. To create the region maps, we used only the customers' reported grandparent birth locations, minimizing biased data curation.
4. We are transparent about the methods that were used to determine these regions as well as the methods used to provide customers with matches to these regions.
5. We have published educational material available to the public about how genetic ancestry testing can impact Indigenous sovereignties and peoples, including this recent blog post.

We hope that this update will inspire many to learn more about the complex histories and diversity of Indigenous Peoples in North America in a manner that respects the ways in which Indigenous Peoples define themselves.

https://blog.23andme.com/ancestry-reports/indigenous-genetic-ancestry/