

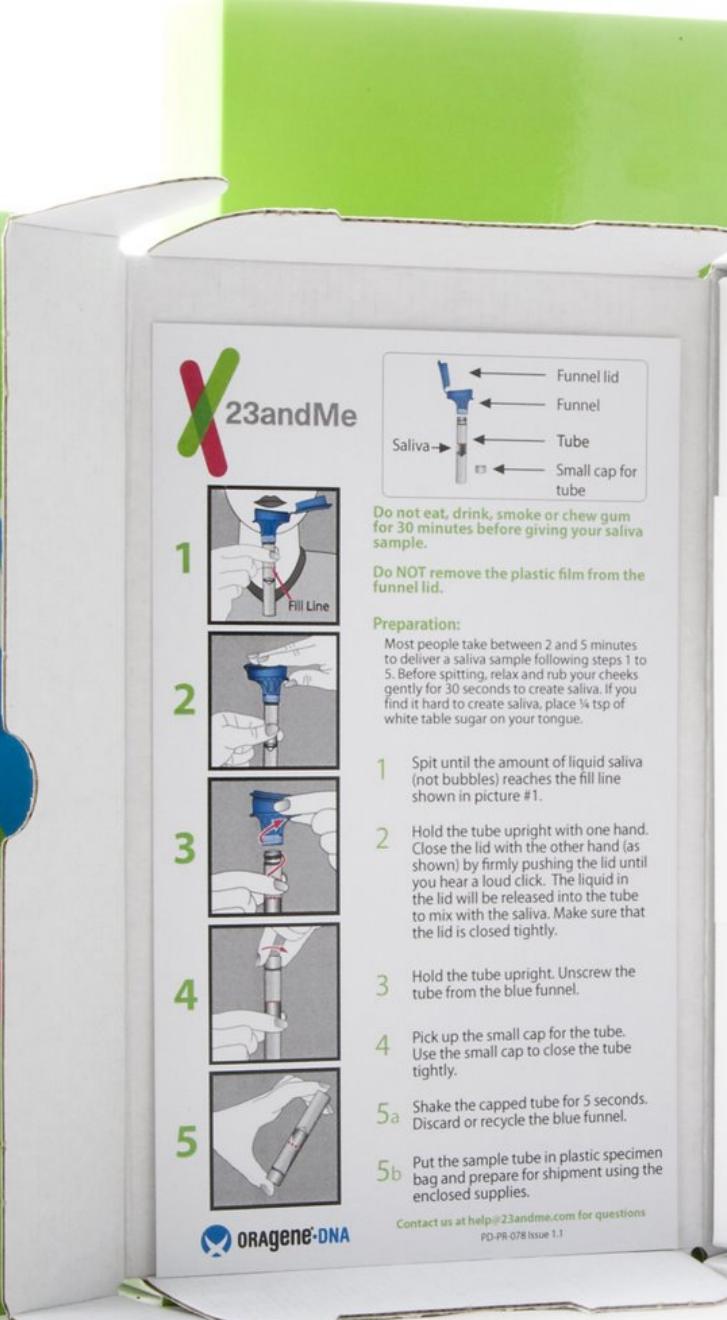


Data science and **23andMe**

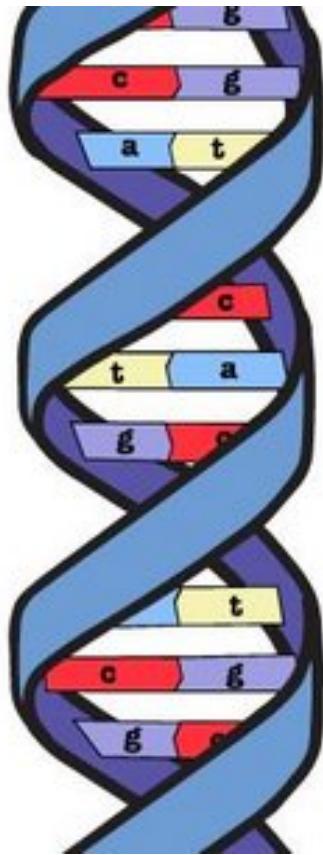
Alan Schoen

What is 23andme?

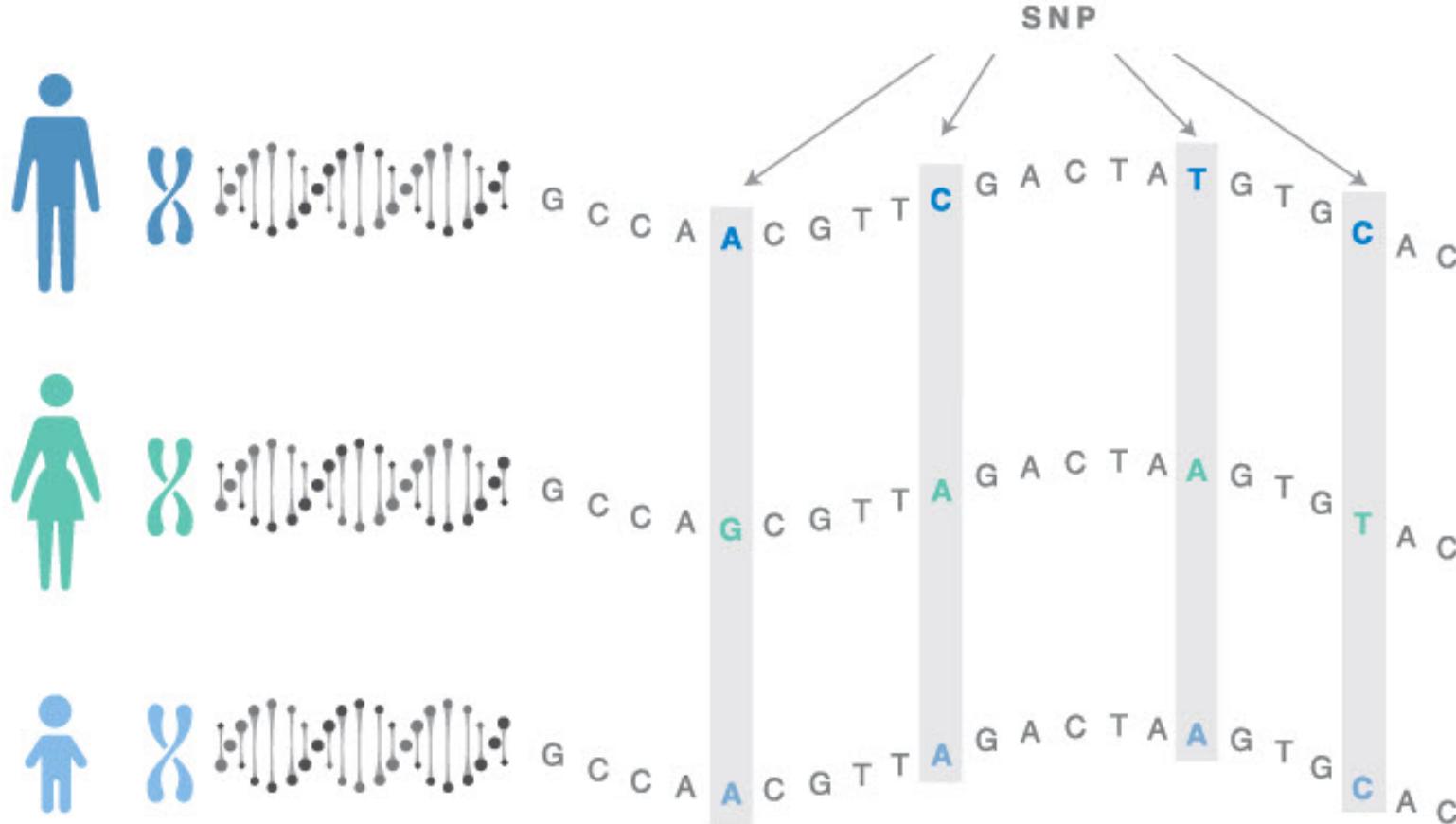
- Direct-to-consumer genetic testing service
- Generates ML-based findings about ancestry
- Links results to published science about genetic traits and medical risk
- Conducts studies by sending questionnaires to users
- Sells data for research



SNP arrays

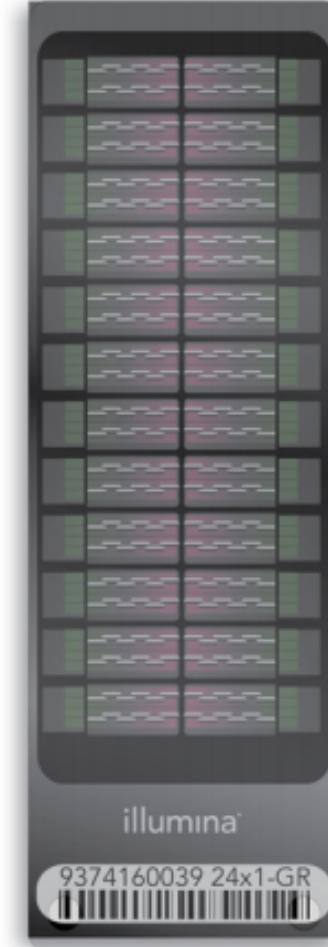


Single Nucleotide Polymorphism (SNPs)



Genetic test

- Illumina HumanOmniExpress-24 format chip
- High-throughput test using pre-selected SNPs
- Tests 713,599 SNPs across the genome
- Fewer than 1% of the nucleotides in human genome, selected for relevance to scientific findings
- 3 minutes to do measurements, but computing results takes a lot of time and computing power

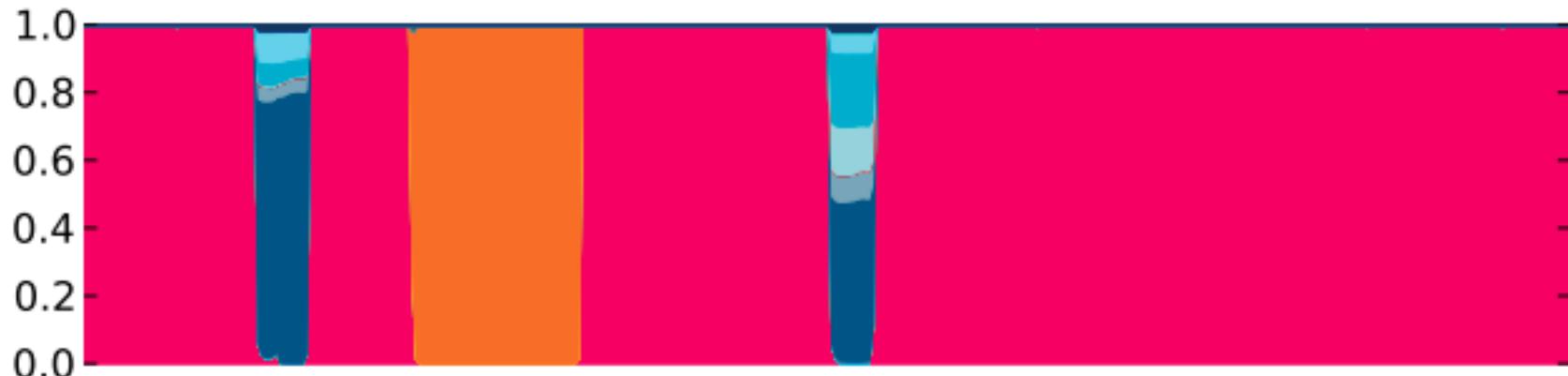


Ancestry

ML application: ancestry

- Start with database of individuals with known ancestry (public + own data)
- Divide genome into windows of about 100 markers
- Classify each window
- Smooth across adjacent windows to handle borderline windows

Ancestry of an African American user





An estimated **2.6%** of your DNA is from Neanderthals.

Jean-Francois Foufant (you)

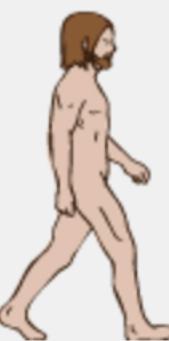


Average European user



MODERN HUMANS

- Higher brow
- Narrower shoulders
- Slightly taller



NEANDERTHALS

- Heavy eyebrow ridge
- Long, low, bigger skull
- Prominent nose with developed nasal chambers for cold-air protection



Freckling

Show results for all profiles

Journal	PLoS Genet
Study Size	3 people
Replications	None
Contrary Studies	None
Applicable Ethnicities	European
Marker	rs2153271

A study of 4,405 individuals of European ancestry who participated in 23andMe research surveys identified a novel genetic variant associated with freckling. People with the C version of rs2153271 (located in the BNC2 gene) exhibited less freckling than people with the T version.

Citations

Eriksson N et al. (2010). "Web-based, Participant-driven Studies Yield Novel Genetic Associations for Common Traits" *PLoS Genet* 6(6):e1000993.

Who	Genotype	Genetic Result
	TT	More freckling than average
Jean-Francois Foufant	CT	Typical amount of freckling
	CC	Less freckling than average

Medical results

FDA problems with medical results

- FDA sent a warning letter in 2013
- 23andme suspended medical results in USA
- Resumed in 2015 with FDA approval

FDA warning letter

Some of the uses for which PGS is intended are particularly concerning, such as assessments for BRCA-related genetic risk and drug responses (e.g., warfarin sensitivity, clopidogrel response, and 5-fluorouracil toxicity) because of the potential health consequences that could result from false positive or false negative assessments for high-risk indications such as these. For instance, if the BRCA-related risk assessment for breast or ovarian cancer reports a false positive, it could lead a patient to undergo prophylactic surgery, chemoprevention, intensive screening, or other morbidity-inducing actions, while a false negative could result in a failure to recognize an actual risk that may exist. Assessments for drug responses carry the risks that patients relying on such tests may begin to self-manage their treatments through dose changes or even abandon certain therapies depending on the outcome of the assessment. For example, false genotype results for your warfarin drug response test could have significant unreasonable risk of illness, injury, or death to the patient due to thrombosis or bleeding events that occur from treatment with a drug at a dose that does not provide the appropriately calibrated anticoagulant effect. These risks are typically mitigated by International Normalized Ratio (INR) management under a physician's care. The risk of serious injury or death is known to be high when patients are either non-compliant or not properly dosed; combined with the risk that a direct-to-consumer test result may be used by a patient to self-manage, serious concerns are raised if test results are not adequately understood by patients or if incorrect test results are reported.

Locked Reports

NAME	CONFIDENCE	STATUS
Hereditary Breast and Ovarian Cancer Syndrome (BRCA1- and BRCA2-Related, Selected Mutations)	★★★★	

NAME	CONFIDENCE	STATUS
Celiac Disease (HLA-DQ2-Related)	★★★★	Variant Present; Higher Risk
Parkinson's Disease (LRRK2- and GBA-Related)	★★★★	Variant Present; Higher Risk
Alpha-1 Antitrypsin Deficiency	★★★★	Variant Absent; Typical Risk
Alzheimer's Disease (APOE Variants)	★★★★	ε4 Variant Absent
Early-Onset Primary Dystonia (DYT1-TOR1A-Related)	★★★★	Variant Absent; Typical Risk
Factor XI Deficiency	★★★★	Variant Absent; Typical Risk
Familial Hypercholesterolemia Type B (APOB-Related)	★★★★	Variant Absent; Typical Risk
Familial Transthyretin (TTR) Amyloidosis	★★★★	Variant Absent; Typical Risk
Hereditary Hemochromatosis (HFE-Related)	★★★★	Variant Present; Typical Risk
Hypertrophic Cardiomyopathy (MYBPC3 25-base-pair Deletion)	★★★★	Variant Absent; Typical Risk
Inherited Thrombophilia (Factor V Leiden- and Prothrombin-Related)	★★★★	Variant Absent; Typical Risk

23andme research

23andme conducts survey-based studies



You answer questions.



Other 23andMe members
answer questions.



23andMe scientists
work their magic.



23andMe
discoveries



And make discoveries!

If forced to choose, would you say you generally prefer to talk or to listen?

Talk

33%

Listen

56%

I'm not sure

11%

[Edit Answer](#)

Is your desk typically...

Very clean

13%

Fairly clean

38%

Fairly messy

36%

Very messy

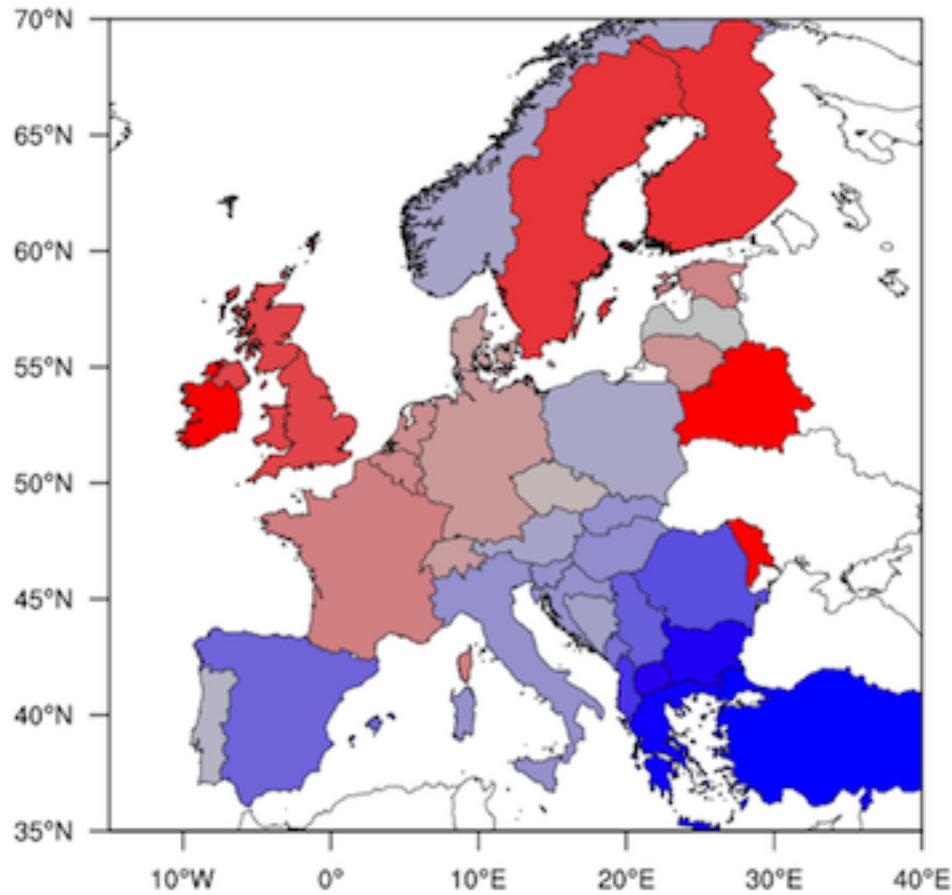
12%

I'm not sure

1%

Sociocultural study

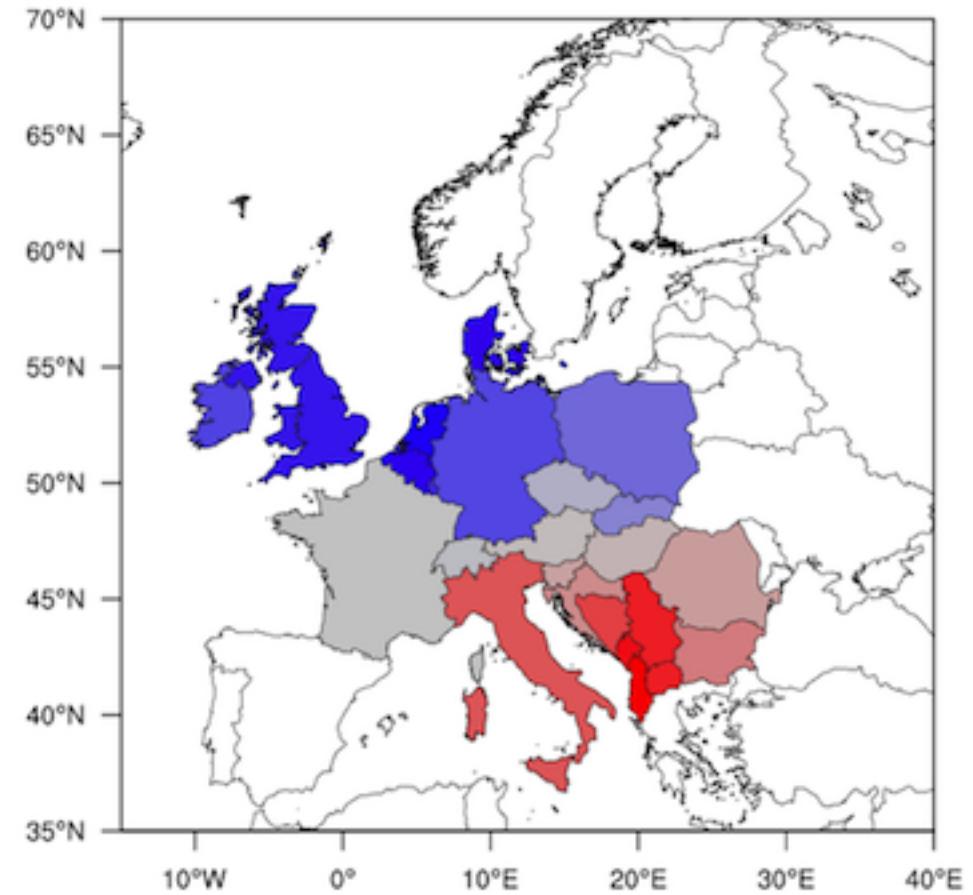
Diagnosed with alcoholism



Less likely

More likely

Extraversion



Less

More

Allergies



https://www.23andme.com/you/23andwe/discoveries/allergy_correlations/