Personal Genomics

Wynn K Meyer, PhD

Bioscience in the 21st Century

Lehigh University

November 18, 2019

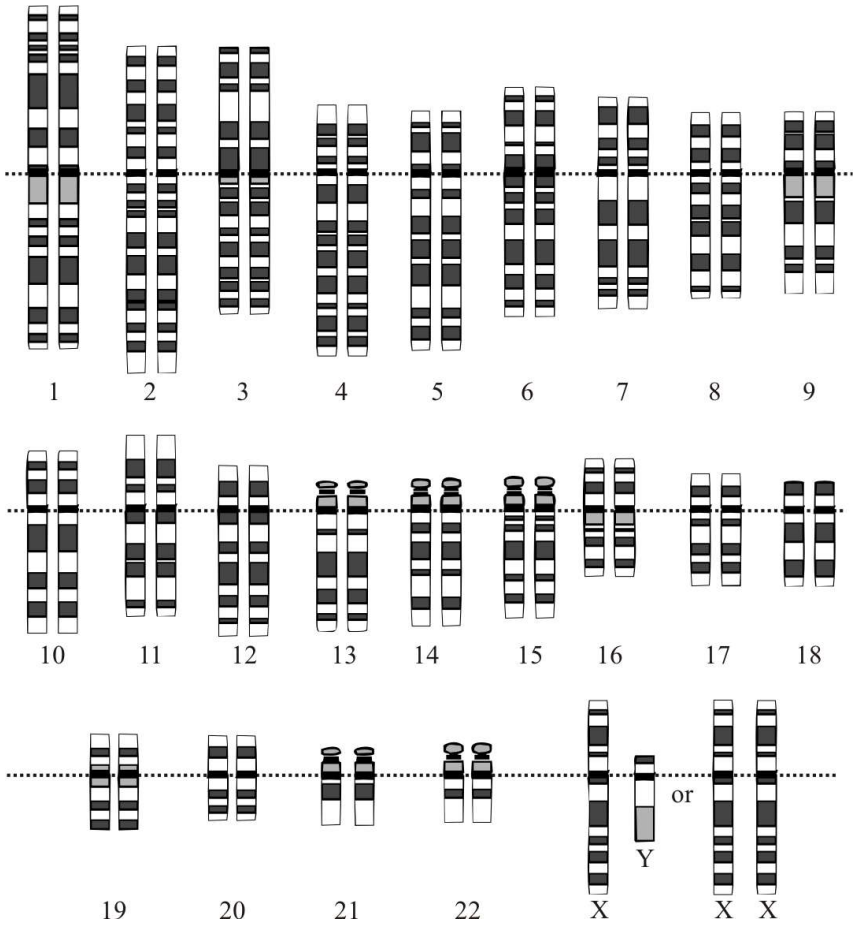
Draw your genome.

On your drawing, please label:

- The region(s) that can help you figure out where your ancestors lived.
- The region(s) that determine whether you will fall asleep in this class.
- The region(s) that determine your eye color.
- The region(s) that determine your skin color.
- The region(s) that contribute to your risk of cancer.

Guessing is fine!

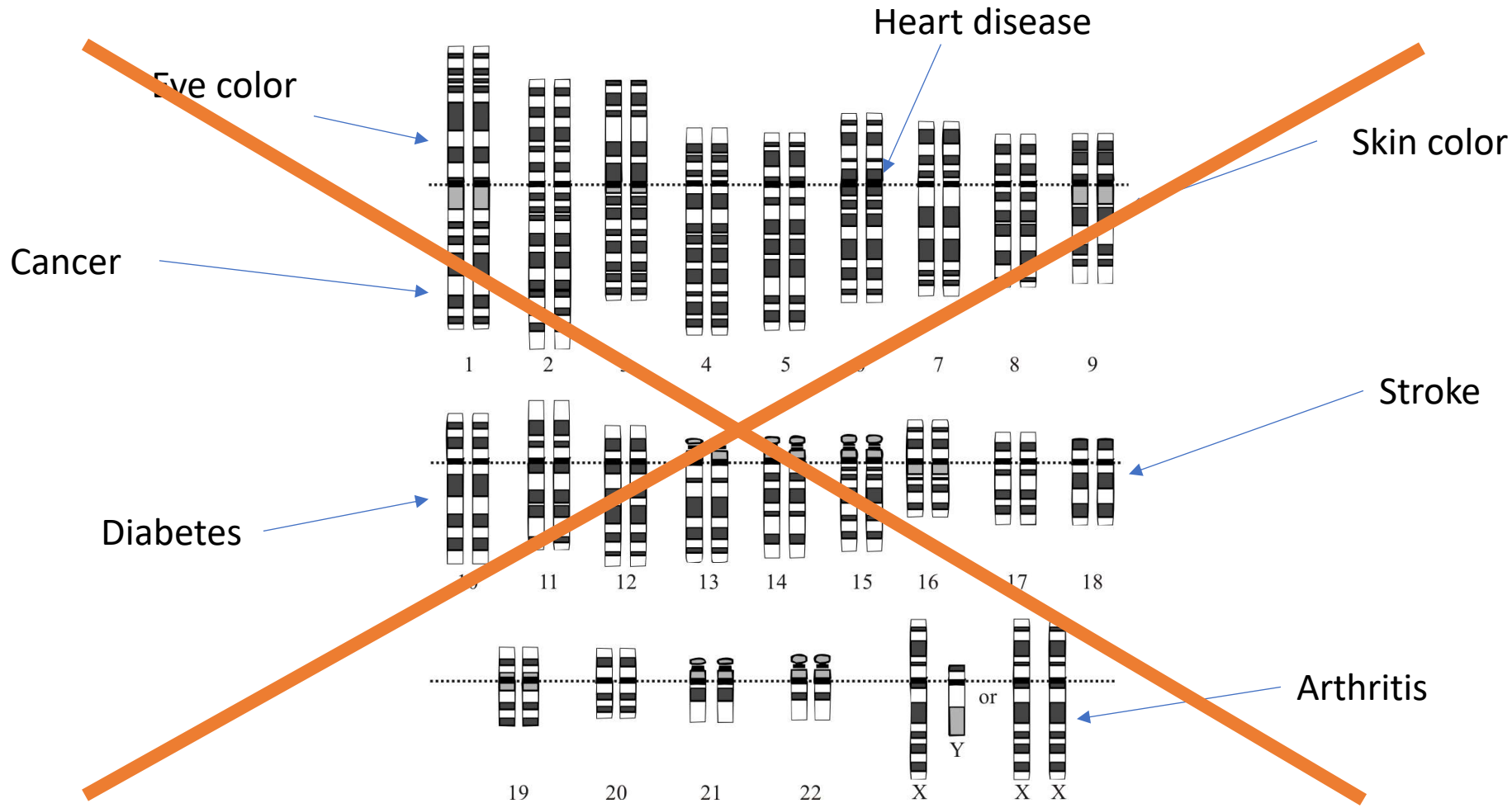
What does your drawing look like?



Courtesy: National Human Genome Research Institute [Public domain]

Public domain photo / NASA]

If things were simple:



Courtesy: National Human Genome Research Institute [Public domain]

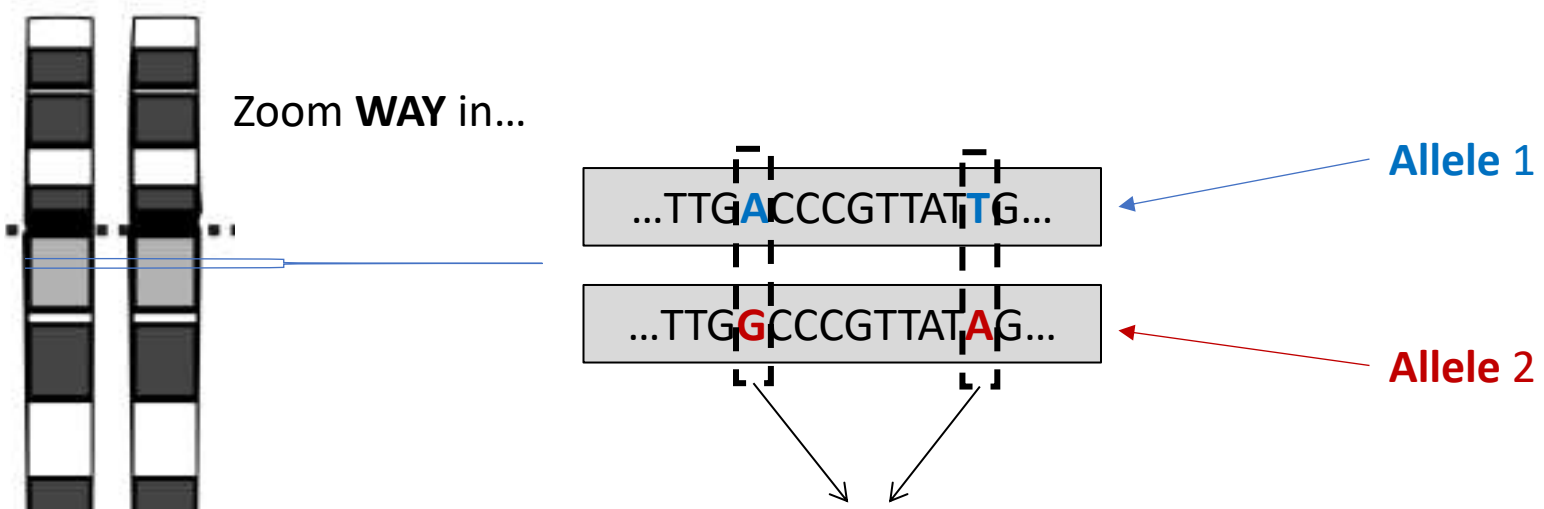
Why not?

How do we link **phenotypes** with the **genetic variants** that underlie them?

Arthritis
Cancer
Diabetes
Eye color

Heart disease
Skin color
Stroke

What are **genetic variants**?



What process generates SNPs?

The term "**allele"** can refer either to the nucleotide at the **SNP** (A/G or T/A) …
or to the combination across both SNPs (AT/GA).

Either way, it represents one "version" of the sequence.

The combination of **alleles** that a person has at a **SNP** is called a **genotype**.
E.g., for SNP1: AA, AG, and GG.

What do genetic variants **do**?

Mostly **nothing!**

Why?

In order to impact a **phenotype**, a variant would have to influence a **gene**...

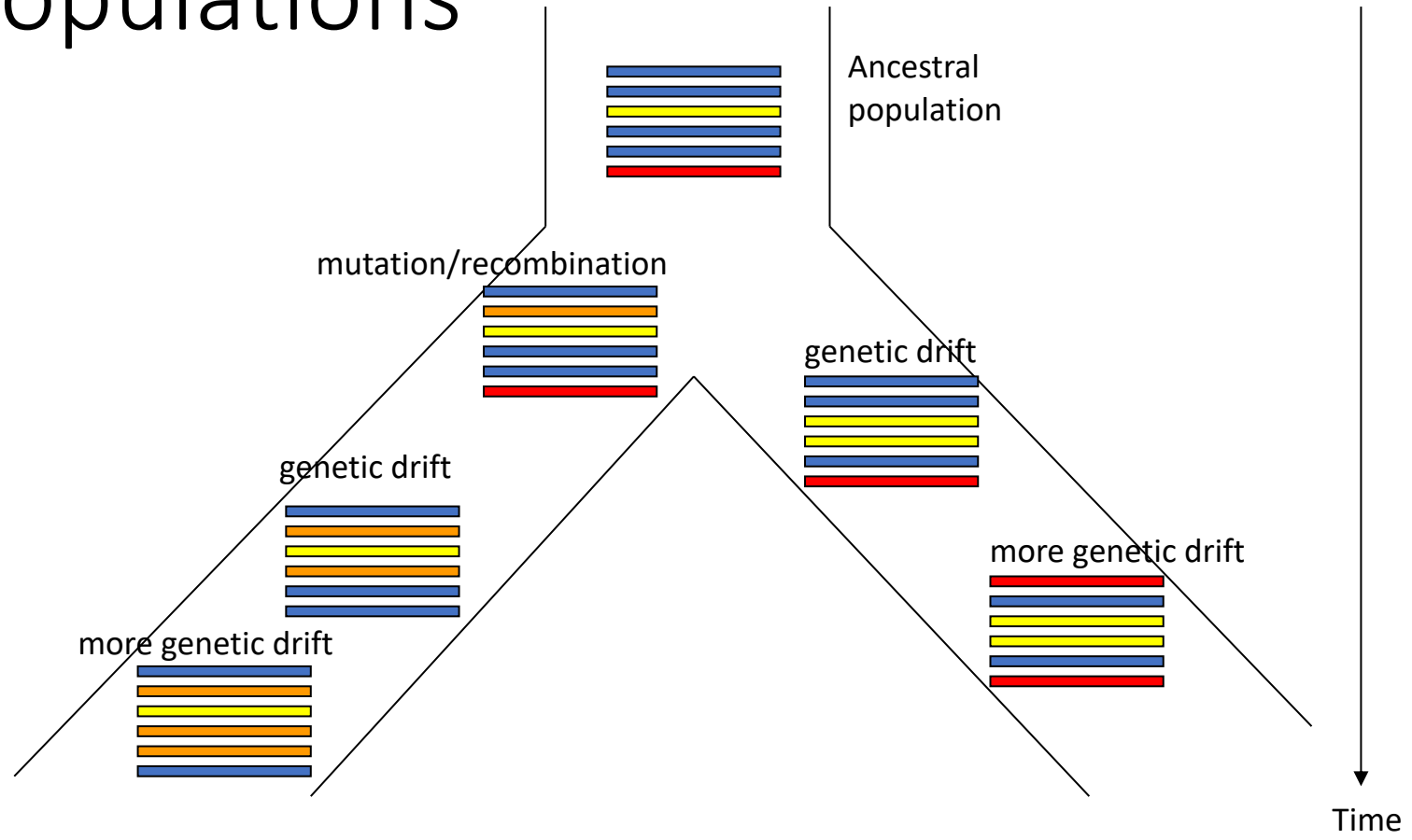
Gene*: one or more exons (and introns) that can be transcribed into RNA and translated into protein

…and **genes** are rare!

From the 1000 Genomes Project Consortium (Nature 2015):

- We find that a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites (out of 3.2 billion).
  - *SNPs are rare! Only ~1 in 1000 sites is a SNP.*
- [A] typical genome contained 149–182 sites with **protein truncating variants**, 10,000 to 12,000 sites with **peptide-sequence-altering variants**, and 459,000 to 565,000 variant sites overlapping known **regulatory regions**.
  - These are the variants that could **possibly influence genes**.
  - *Genes are also rare! Likely ~1 in 300 SNPs could affect a gene.*

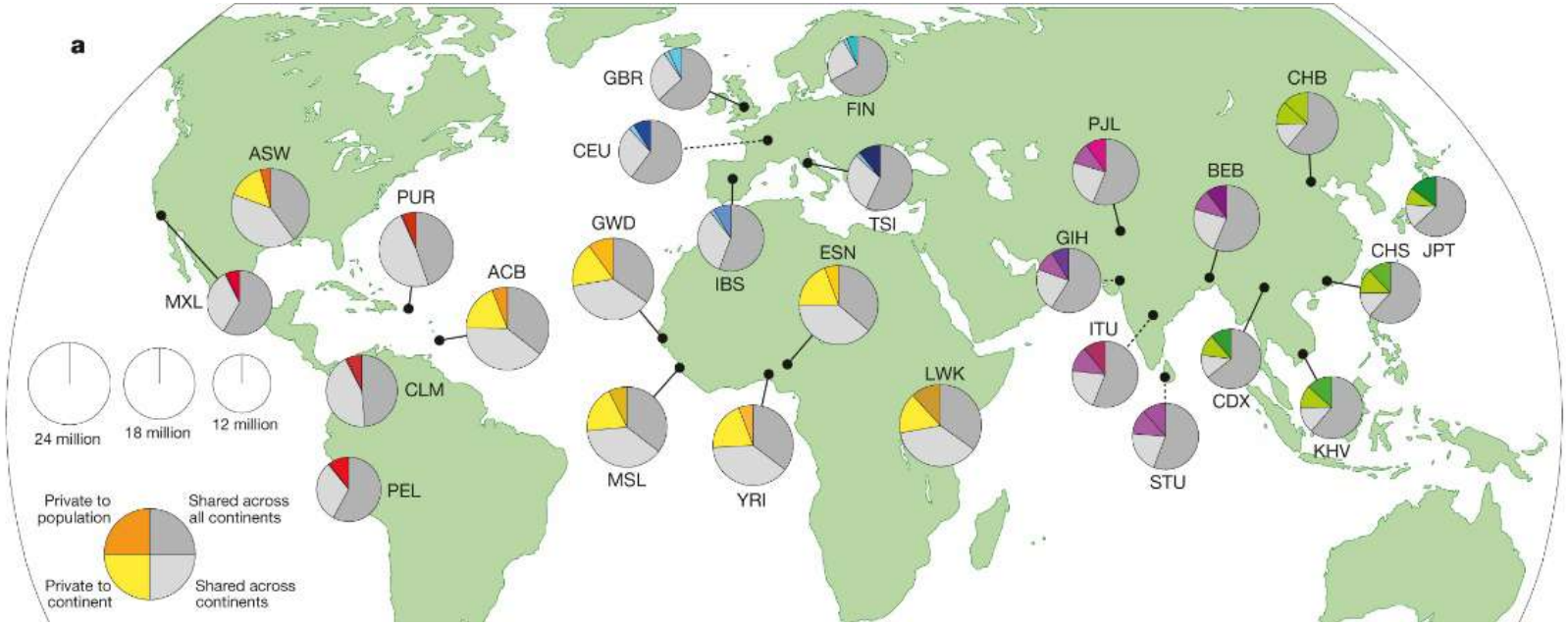
Genetic variants arise and spread within populations



What is the current **frequency** of the **alleles** present in the two populations?

Most human genetic variants are present in multiple populations

SNPs discovered from sequencing the whole genomes of 2,504 humans from 26 populations



A Auton *et al*. *Nature* **526**, 68-74 (2015) doi:10.1038/nature15393

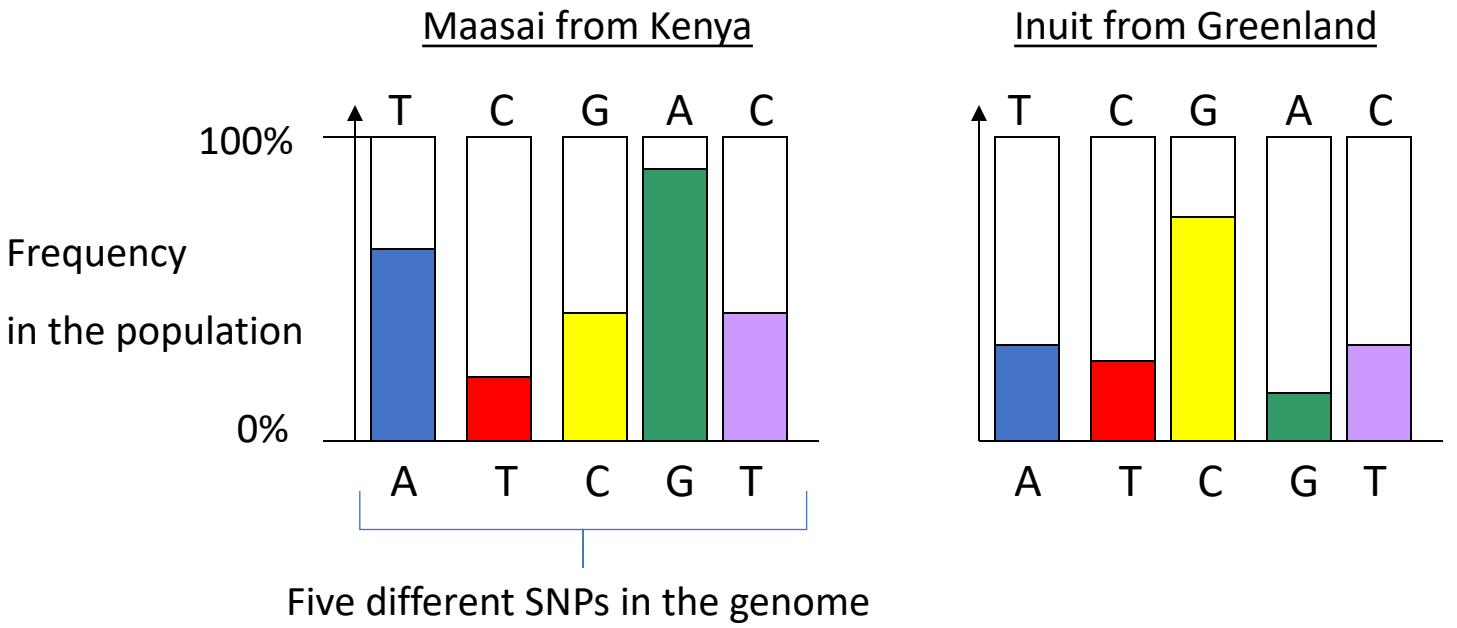
Why are "private" SNPs so rare?

- Human migrations are very recent.
- Human populations experience frequent gene flow.

Once discovered, SNPs are genotyped using a **SNP chip (microarray)**.

nature

Some variants can help us trace our history



Which of these SNPs would be most useful in determining whether an individual's ancestors came from Kenya or from Greenland?

Where do you think these are in the genome?

- The region(s) that can help you figure out where your ancestors lived.

…Now back to phenotypes!

Is it **heritable?**
(If so, how heritable?)

Informally: runs in families

https://harrypotter.fandom.com/wiki/Weasley_family?file=Weasleys.png

Formally: proportion of variance due to genetic contributions

Variation due to genetics

$$h^2 = \frac{V_G}{V_G + V_E} = \frac{V_G}{V_P}$$

Total variation

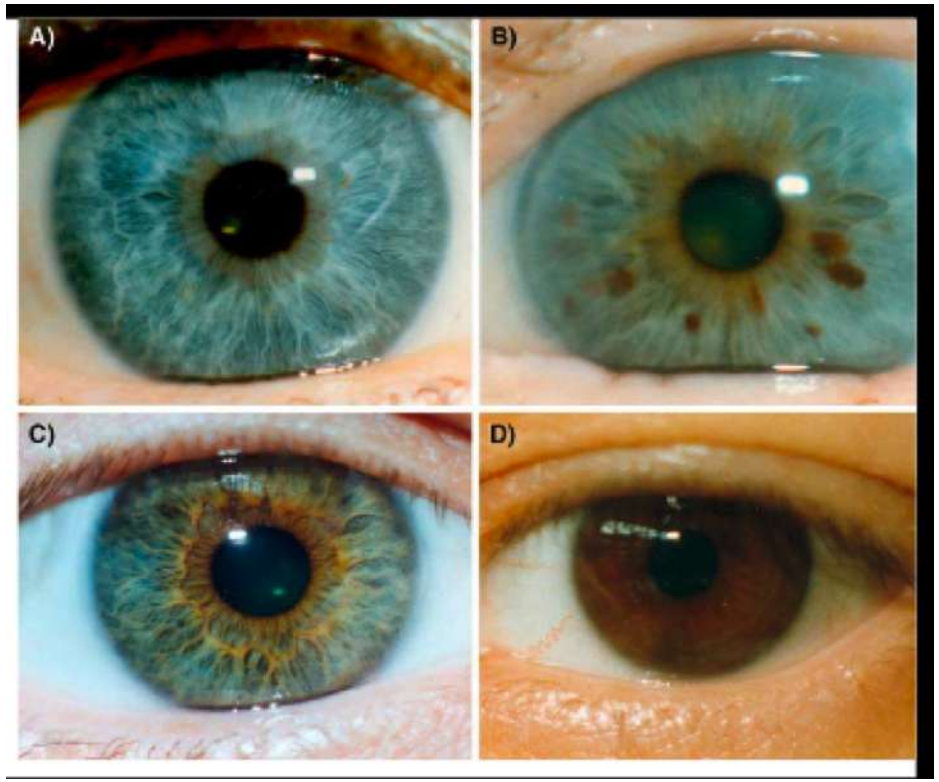
Heritability of height in US: ~80%
Heritability of height in West Africa: ~65%

What is the **heritability** of your propensity to fall asleep in this class?

Is it **simple** or **complex?**
(one gene or several/many?)

Simple*: Eye color

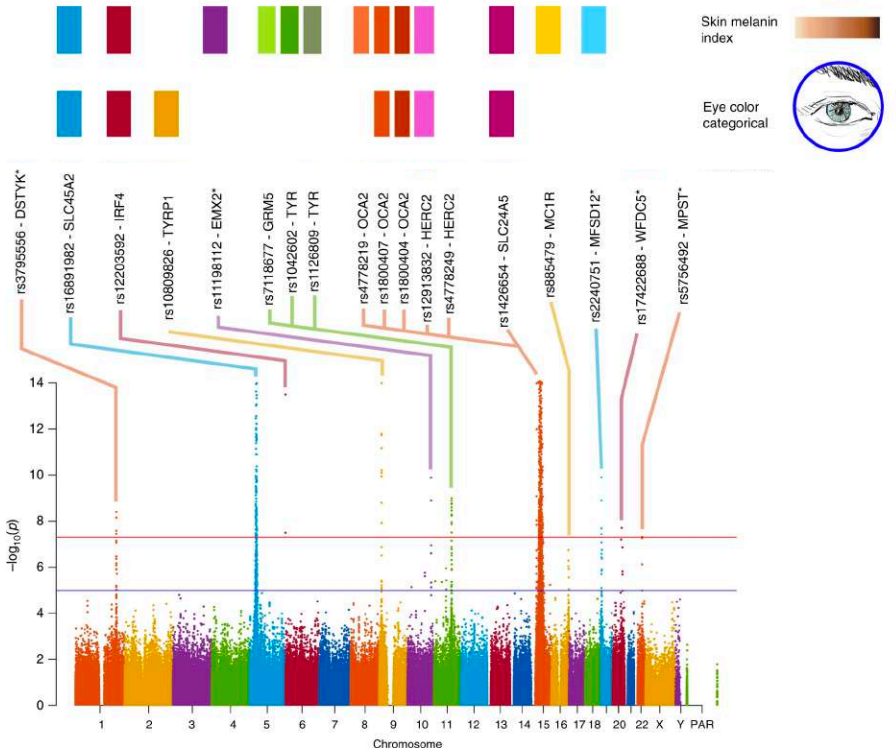
(can predict with ~90% accuracy from genotype at 1 site)



Eiberg et al. Hum Genet. 2008

Complex: Skin color

(many regions throughout the genome contribute a small amount)



Adhikari et al. Nat. Comms. 2019

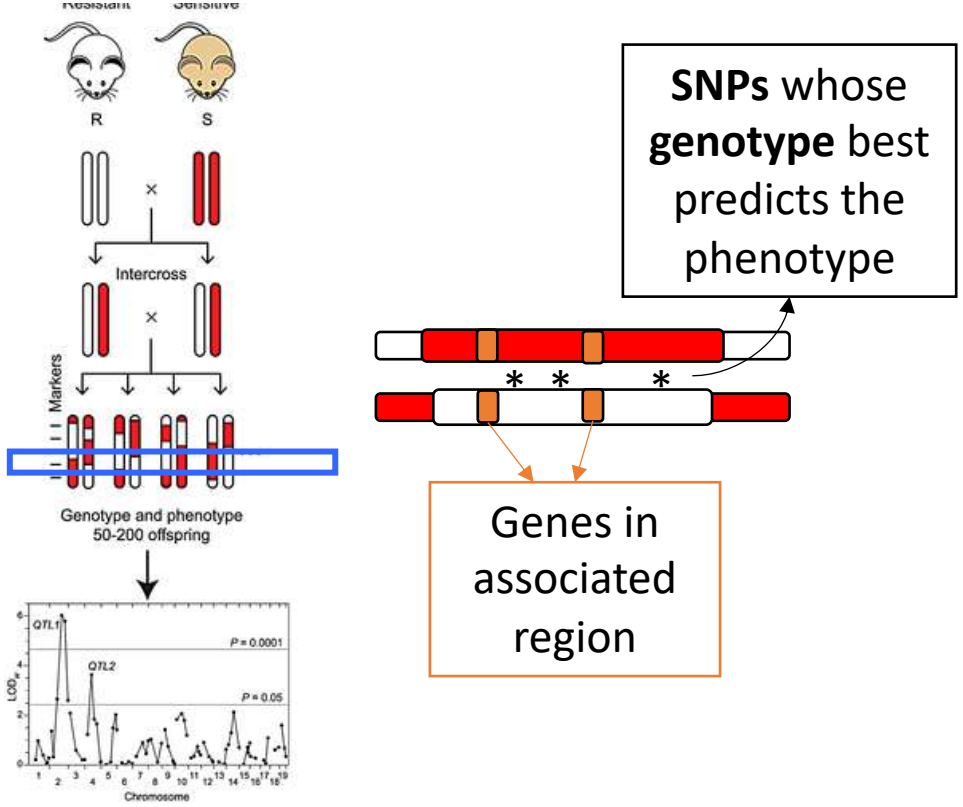
Which **genes** are involved?

Gene*: one or more exons (and introns) that can be transcribed into RNA and translated into protein

How have we identified genes associated with phenotypes?

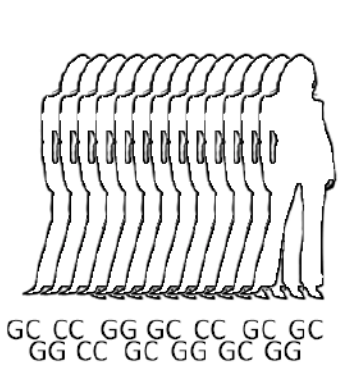
Try to isolate a single gene that is inherited with a phenotype…

- In model organisms (can induce mutations and use controlled breeding)
- In human families (often follow **simple** phenotypes)



Drinkwater and Gould PLOS Genet. 2012

Genome-wide association studies (GWAS)



	SNP1	*SNP2*	*SNP…*
	Cases	**Cases**	*Repeat for all SNPs*
	Count of G: 2104 of 4000	Count of G: 1648 of 4000	
	Frequency of G: 52.6%	Frequency of G: 41.2%	
	Controls	**Controls**	
	Count of G: 2676 of 6000	Count of G: 2532 of 6000	
	Frequency of G: 44.6%	Frequency of G: 42.2%	

Figure by Lasse Folkersen [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)]

Cases = people with the phenotype of interest

Controls = people without the phenotype of interest

Which of these two **SNPs** is strongly **associated** with the phenotype?

- Which **allele** is associated with higher risk of the phenotype?

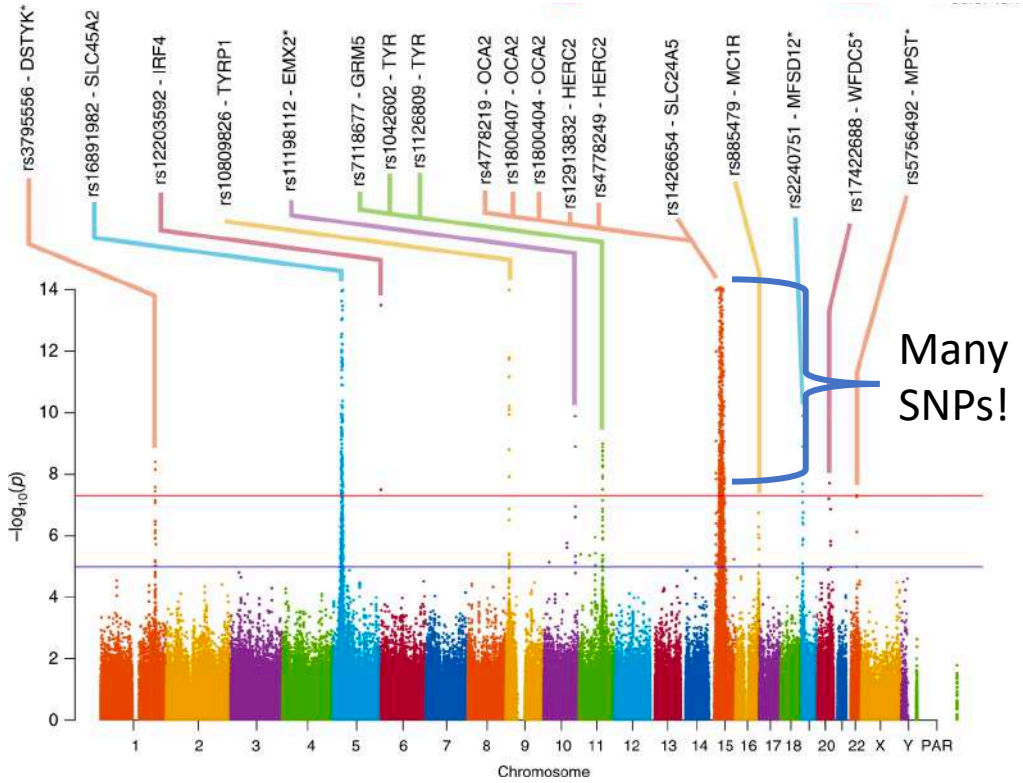
Frequently SNPs that are near each other in the genome show similar association patterns. Why might this be?

…TTG**A**CCCGTTAT**T**G…

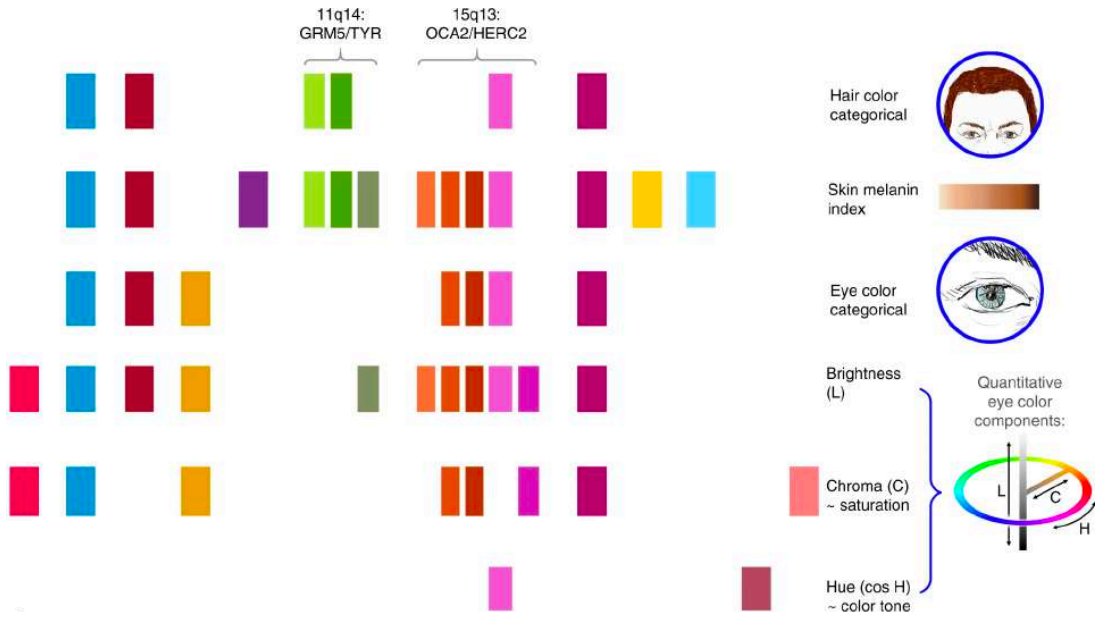
…TTG**G**CCCGTTAT**A**G…

How GWAS results are typically shown

“Manhattan plot”



Several phenotypes “map” to the same regions of the genome



Adhikari et al. Nat. Comms. 2019

Will this experiment work?

- A researcher wants to determine the genetic basis of speaking French. She collects samples from 500 French people and 500 people from elsewhere in Europe and **genotypes** them at 100,000 **SNPs**. She analyzes the data to find **SNPs** that have different **allele frequencies** in the two samples.
- What will she find?

When designing association studies, researchers take care to match cases and controls as much as possible, and to correct for any remaining **population structure** (ancestry differences unrelated to the phenotype) using statistical tools.

What we have learned from GWAS

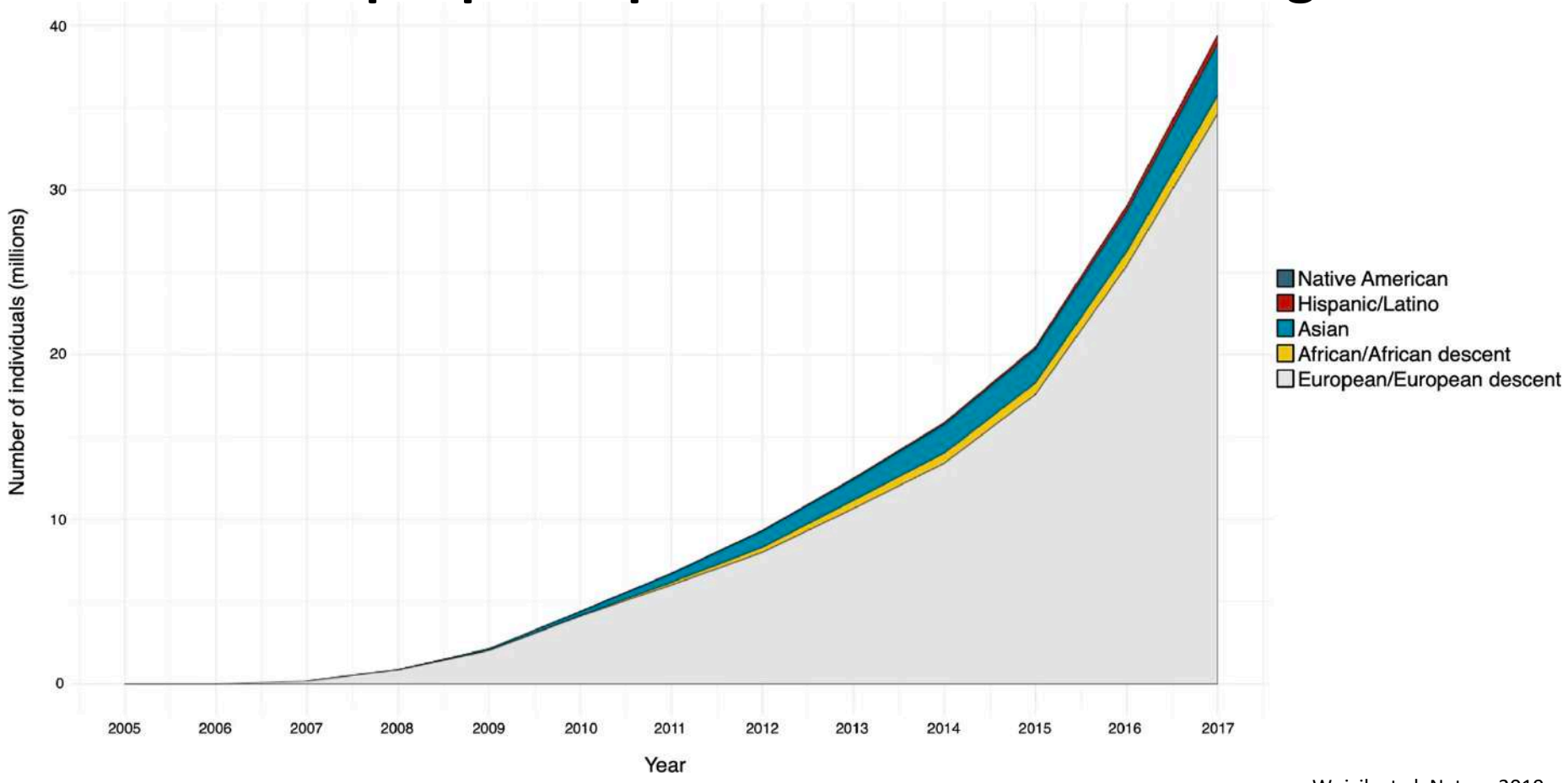
NIH National Human Genome Research Institute

EMBL-EBI

NHGRI-EBI GWAS Catalog
www.ebi.ac.uk/gwas

But we still have a long way to go

Number of unique participants in the GWAS catalog



Wojcik et al. Nature 2019

What we've seen so far

- **Genetic variants** arise by **mutation** and spread through populations mostly by **drift**. This creates **SNPs** (variants within populations) that vary in **allele frequency** between populations.
- Combinations of **SNPs** can be used to infer **ancestry** because of these **allele frequency** differences.
- Phenotypes can be non-**heritable** (usually caused by the environment), **simple** (mostly predicted by variation at one **gene**), or **complex** (determined by multiple genes).
- **Genome-wide association studies** have helped us to understand the links between genetic variants and phenotypes. These studies have to correct for **population structure** and can sometimes find different associations in different populations.

What does this mean for personal genomics?

A few direct-to-consumer (DTC) genomics companies

***Note: This is not an endorsement of any of these companies!**

What do these companies do?

What you provide

- A vial of spit

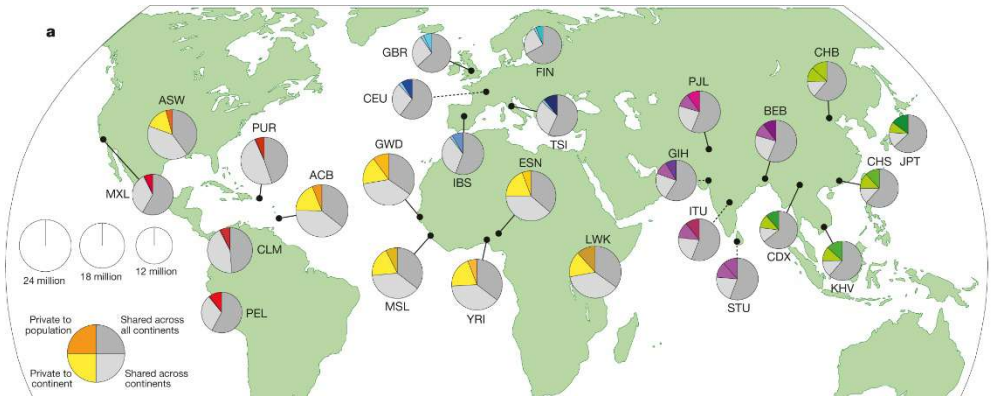
What they promise

- “Understand the past, present, and future of your family’s health.”
- “Get genetic insights into your ancestry, traits and health so you can learn more about your past and make more informed decisions about your future.”

What's really going on?

The accuracy and informativeness of any test depends upon the panel of SNPs used to infer ancestry and disease risk.

Ancestry testing



A Auton *et al*. *Nature* **526**, 68-74 (2015) doi:10.1038/nature15393

Health testing

NIH National Human Genome Research Institute

EMBL-EBI

NHGRI-EBI GWAS Catalog
www.ebi.ac.uk/gwas

What DTC genomic testing can and cannot tell us

We can learn…

- Proportions of ancestry from populations well represented in the companies' databases
- Risk for diseases that are highly heritable and have strong genetic associations
  - (in the populations well represented in the companies' databases)

We cannot learn…

- Proportions of ancestry from other populations (may be mistakenly assigned to known populations)
- Risk for diseases that do not have clear associations or have not been studied in the relevant populations
- Our identity

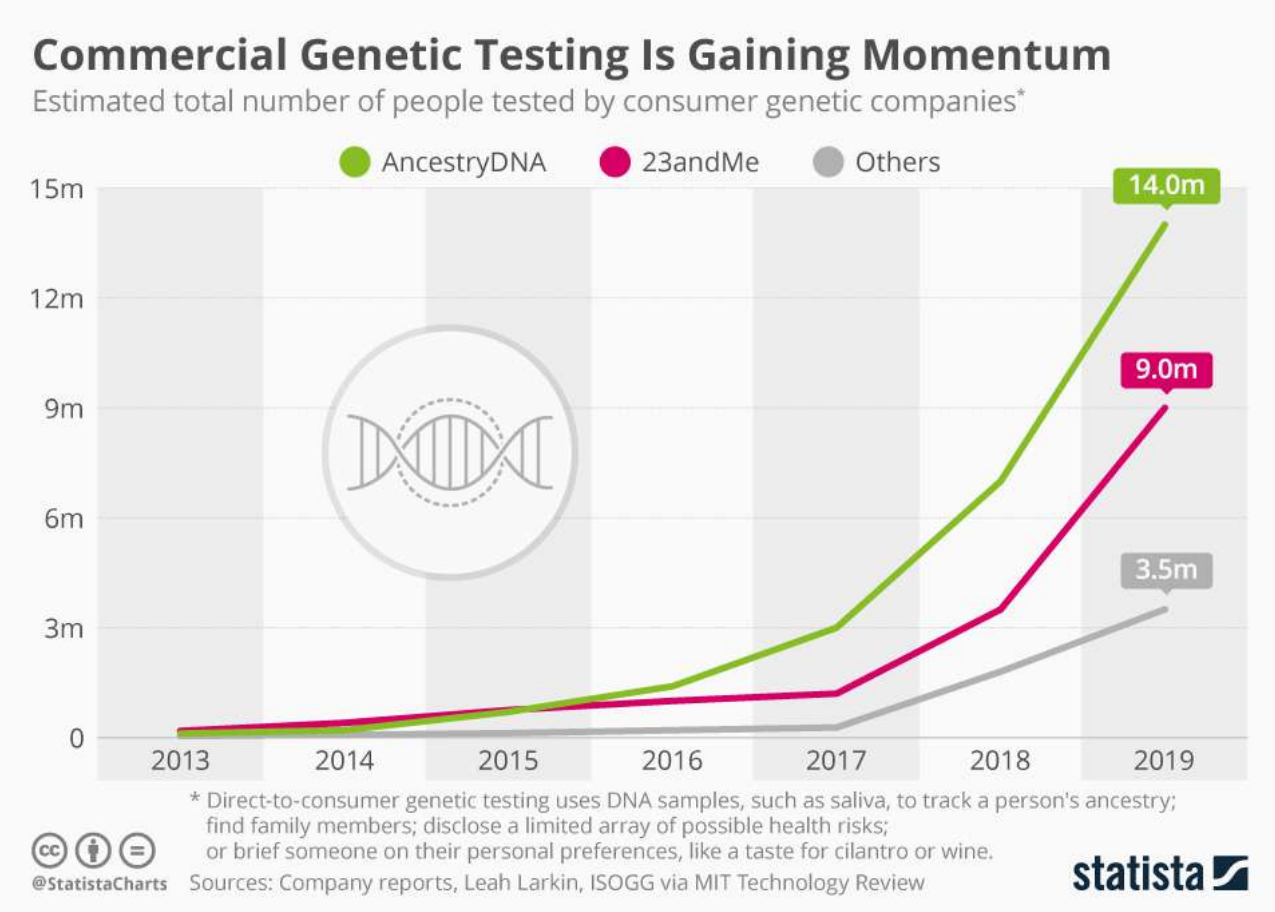
Making decisions based on personal genomic test results

Imagine: You have sent your DNA to a DTC genomics company to learn more about your ancestry. One day you receive an email with the header “New results about your genetic testing with company X.” You open the email and read the following:

- *A new association study has identified a strong association between breast cancer and one of the markers that was included in your ancestry testing panel. Click the link to find out your genotype at this marker.*

Do you click the link? What additional information would you want to know?

What the future holds



Who may need personal genomics expertise?

- Researchers at DTC genomics companies
- Genetic counselors
- Health professionals
- Patients
- Forensic scientists
- Legal professionals
- Policy-makers

Up next!

- How can your liver cells and your heart cells have the same genome but be so different? (Epigenetics in Health and Disease)
- How can you determine the sequence of a genome? (Bioinformatics: Genes and Genomes)
- Could you use technology to change your genome to prevent disease? (CRISPR and Genome Editing)

Thank you!

- Questions? Please email wynn.meyer@gmail.com.

References

1. Auton A, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.
2. Eiberg H, et al. (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 123(2):177–87.
3. Adhikari K, et al. (2019) A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nat Commun* 10(1):358.
4. Drinkwater NR, Gould MN (2012) The long path from QTL to gene. *PLoS Genet* 8(9):e1002975.
5. Buniello A, et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47(D1):D1005–D1012.
6. Wojcik GL, et al. (2019) Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570(7762):514–518.

Additional resources

More information about GWAS:

From the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI): https://www.ebi.ac.uk/training/online/course/gwas-catalog-exploring-snp-trait-associations-2019/what-gwas-catalog/what-are-genome-wide

From the National Human Genome Research Institute (NHGRI): https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet

Visualize results of GWAS by phenotype (NHGRI-EBI GWAS catalog):

https://www.ebi.ac.uk/gwas/diagram

Popular press about trends in DTC genomic testing:

Regalado, Antonio. "More than 26 million people have taken an at-home ancestry test." *MIT Technology Review*, Feb 11, 2019, https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/

Personal stories from two women who found out they were BRCA+ through DTC testing:

Altschule, Sara. "I did a random DNA test to learn about my ancestry. Then I found out I was BRCA positive." *Bustle*, Jul 31, 2018, https://www.bustle.com/p/i-did-a-random-dna-test-to-learn-about-my-ancestry-then-i-found-out-i-was-brca-positive-9723470

Pomerantz, Dorothy. "23AndMe had devastating news about my health. I wish a person had delivered it." *STAT*, August 8, 2019, https://www.statnews.com/2019/08/08/23andme-genetic-test-revealed-high-cancer-risk/

Limitations to DTC genomic test results:

Tandy-Connor S, et al. (2018) False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. *Genet Med* 20(12):1515–1521. https://www.nature.com/articles/gim201838