Bio393: Genetic Analysis

Human variation and allele frequency spectrum

Lecture 12

In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (haploid three gigabase pairs)

Good phenotyping!

Lots of $$$

How do we identify genes in humans?

Draft human genome announced in June 2000

ws
Print"

The New York Times

National Edition

Arizona and New Mexico: M
cloudy in New Mexico, thunderst
in the mountains. Partly sunny
where. Highs 80 mountains, ove
deserts. Weather map is on Page

. No. 51,432 Copyright © 2000 The New York Times TUESDAY, JUNE 27, 2000 Printed in Arizona ONE DOLL

tic Code of Human Life Is Cracked by Scientist

The Book of Life

The 3 billion base pairs ... ... of the intertwining double helix of DNA ... ... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS
Rungs between the strands of the double helix

BASES
A adenine
C cytosine
G guanine
T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions.

The New York Times

become part that Congress was entitled to the last word because Miranda's presump-

A SHARED SUCCE

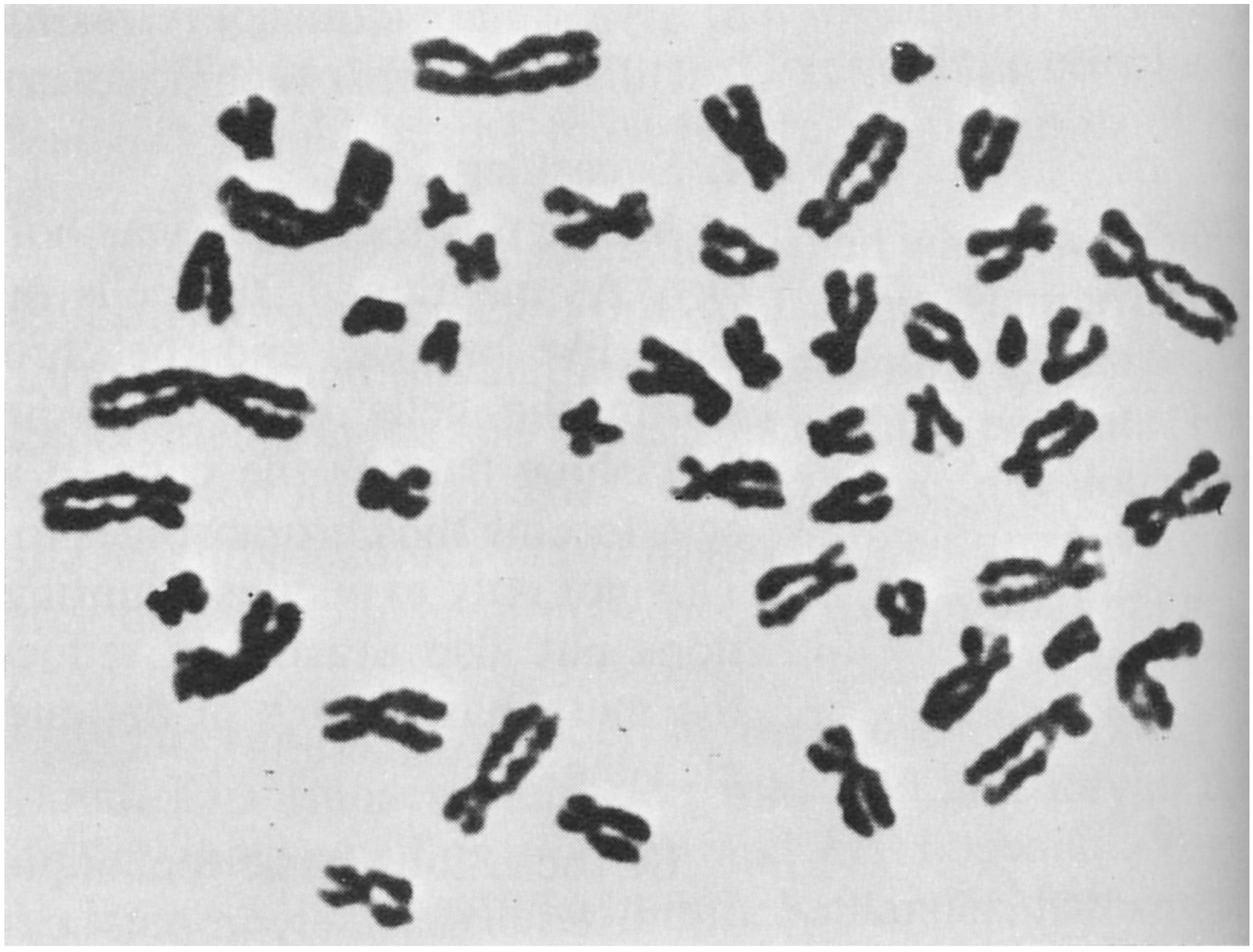
2 Rivals' Announcem
Marks New Medica
Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 —
achievement that represents a
nacle of human self-knowledge
rival groups of scientists said t
that they had deciphered the he
tary script, the set of instruc
that defines the human organis

It took more than 10 years and $3 billion

We need physical pieces of DNA to sequence

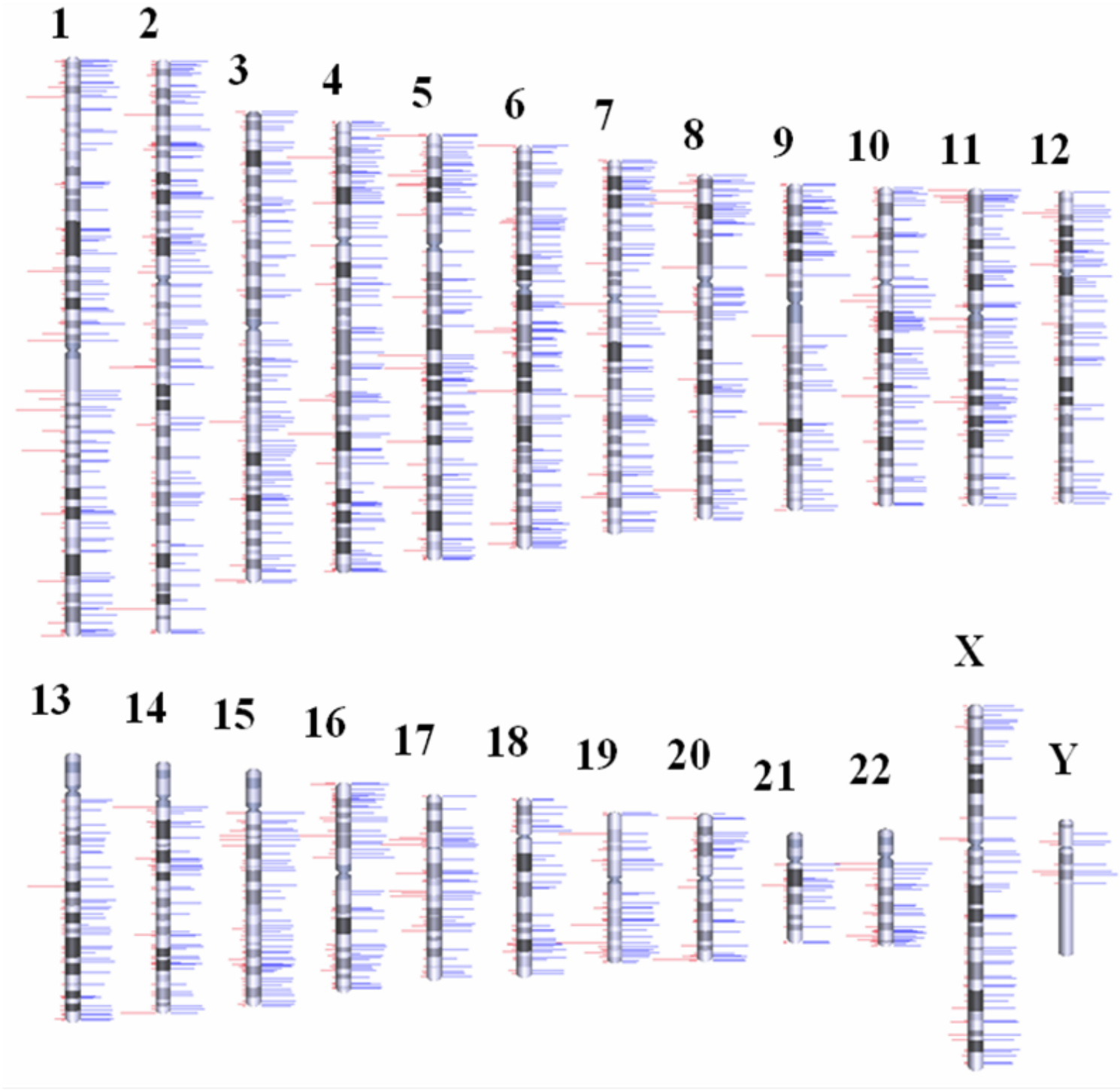


Who was sequenced?

We don’t have one human genome

Nine humans had parts of their genomes sequenced to make the first draft.

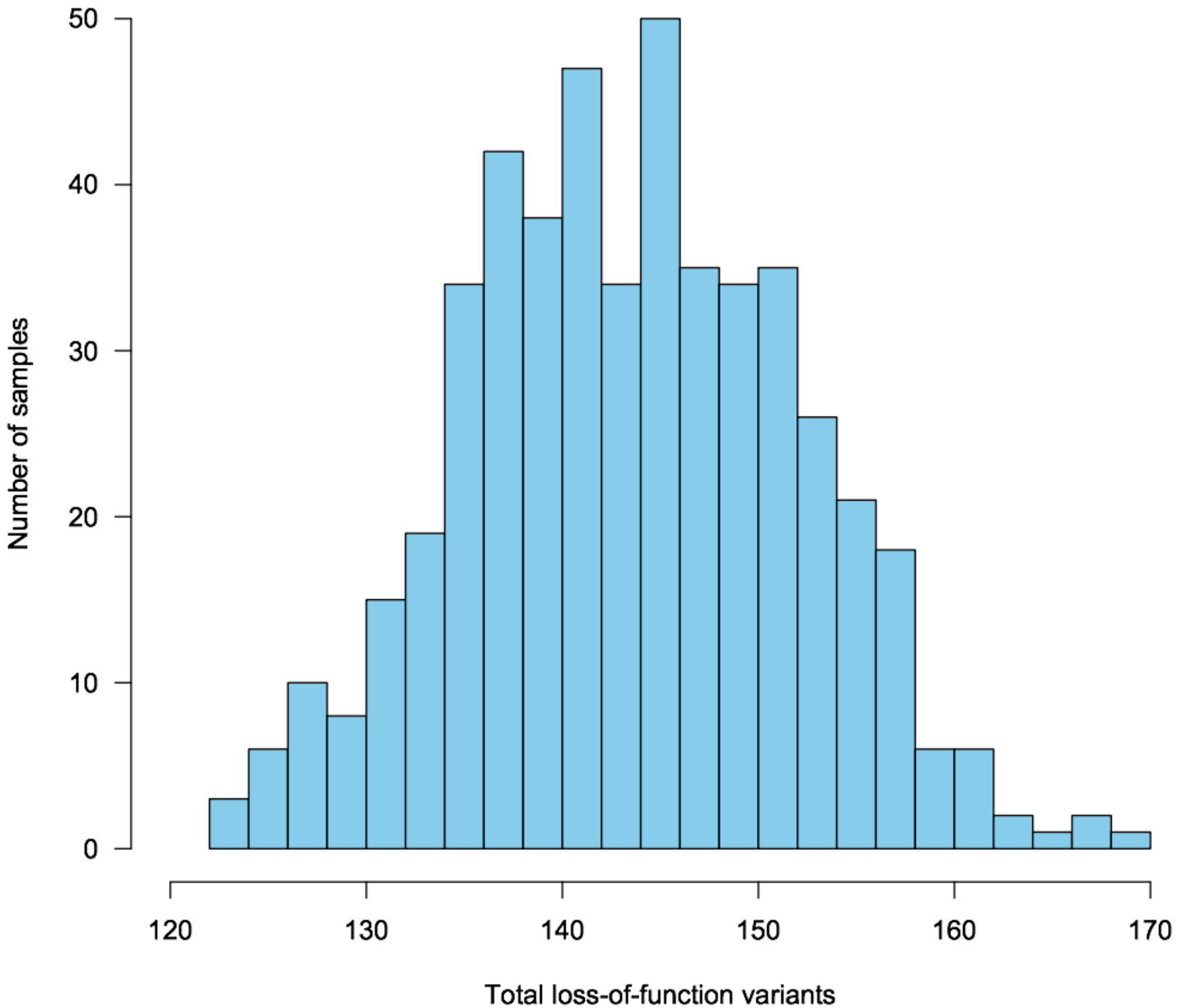
A genome sequence gives us the (incomplete) parts list



Types of variation

Rare = variants found in less than 1% in population

We each have over 100 unique loss-of-function rare variants



Francioli *et al.* 2014 Nature Genetics

Over 3,000 rare diseases have a known underlying genetic cause

One in twelve people have a rare disease

Compound heterozygosity underlies many diseases

Types of variation

Rare = variants found in less than 1% in population

Common = variants found in more than 5% of the population

Intermediate = variants found in 1-5% of the population

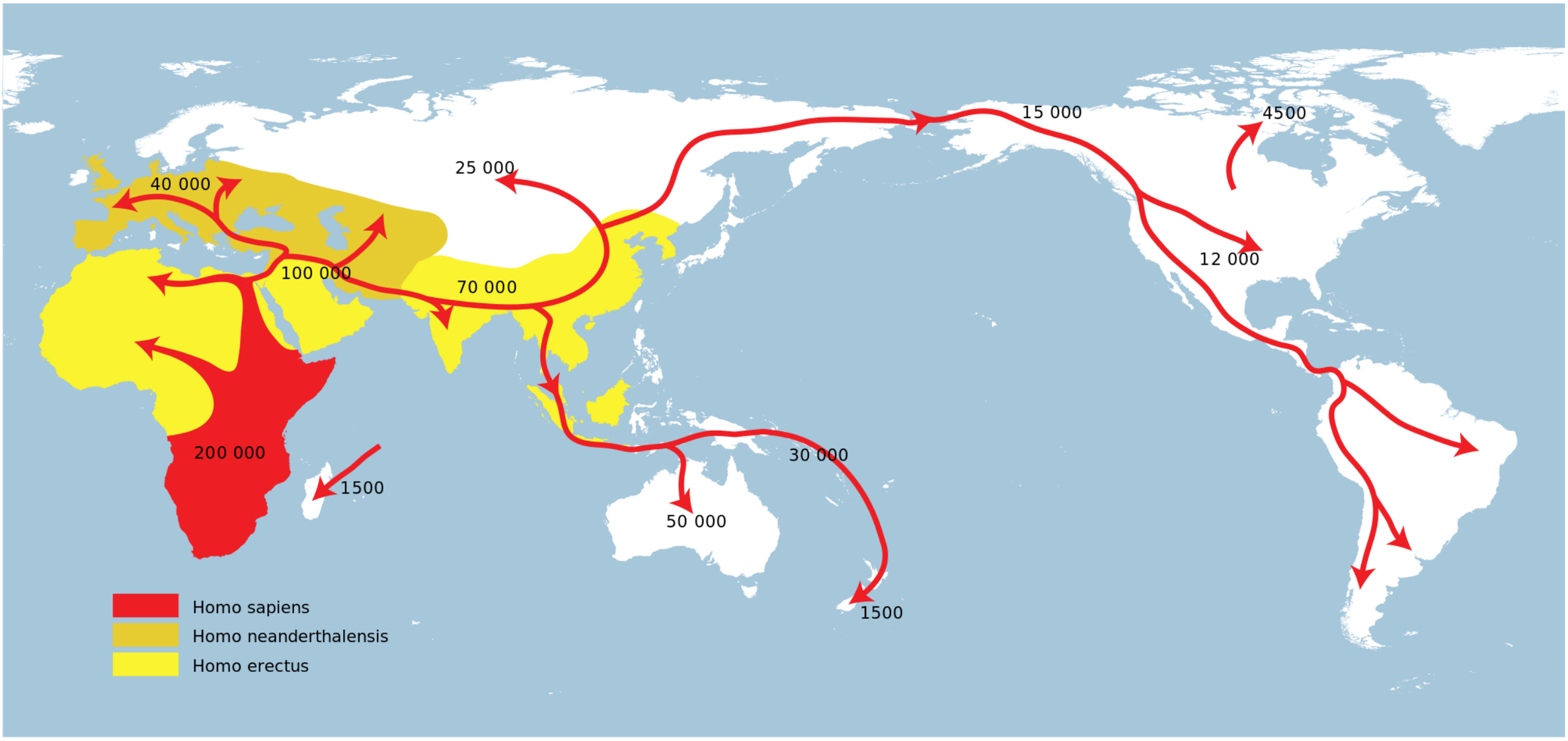
Where does variation come from?

Random errors in replication, transcription, DNA repair, etc.

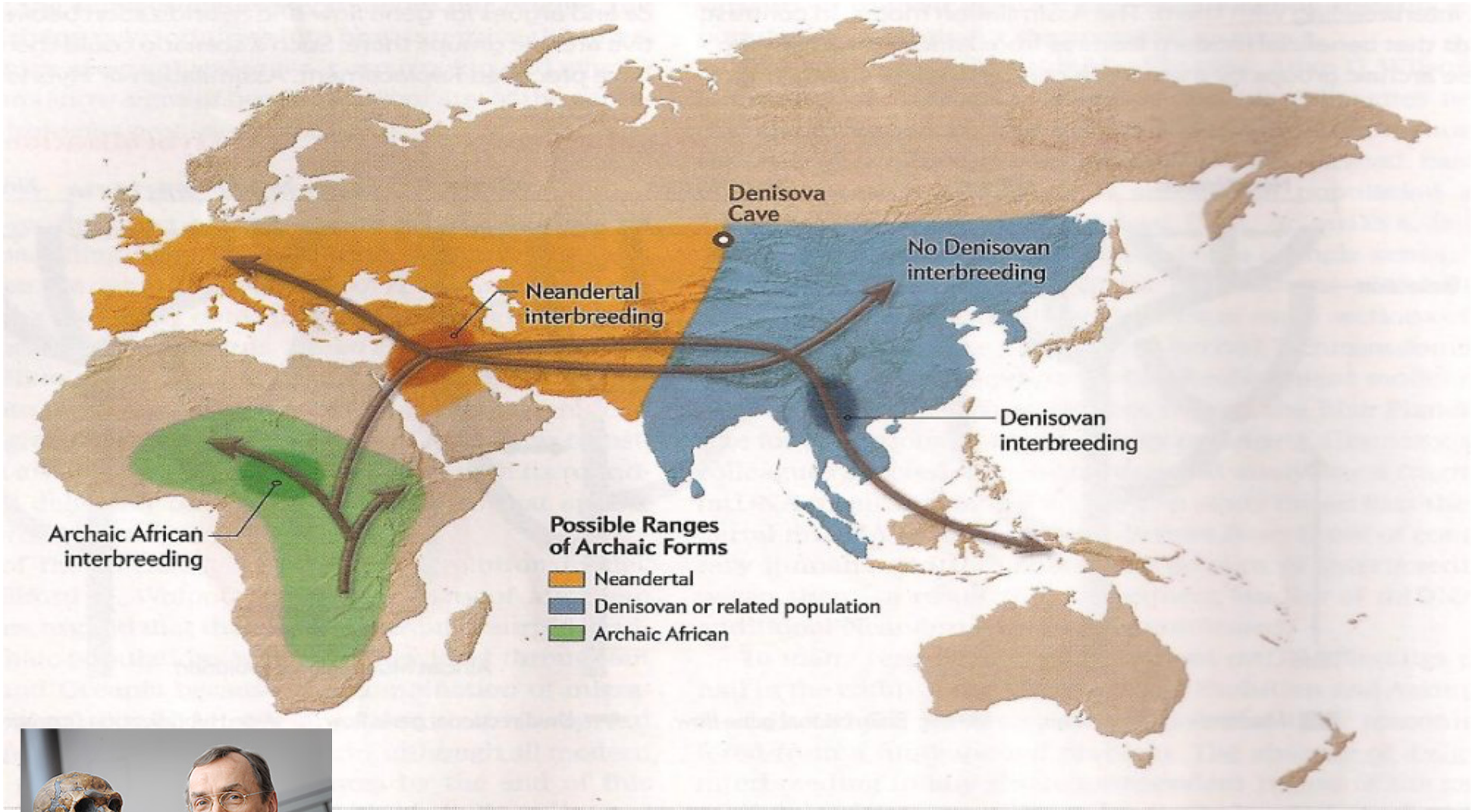
Somatic or germline errors

Once generated, germline variants are inherited

Human history drives our genetics

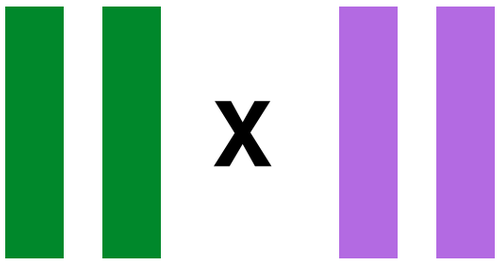


Human history drives our genetics



Svante Pääbo

H. neanderthalensis



H. sapiens

50%

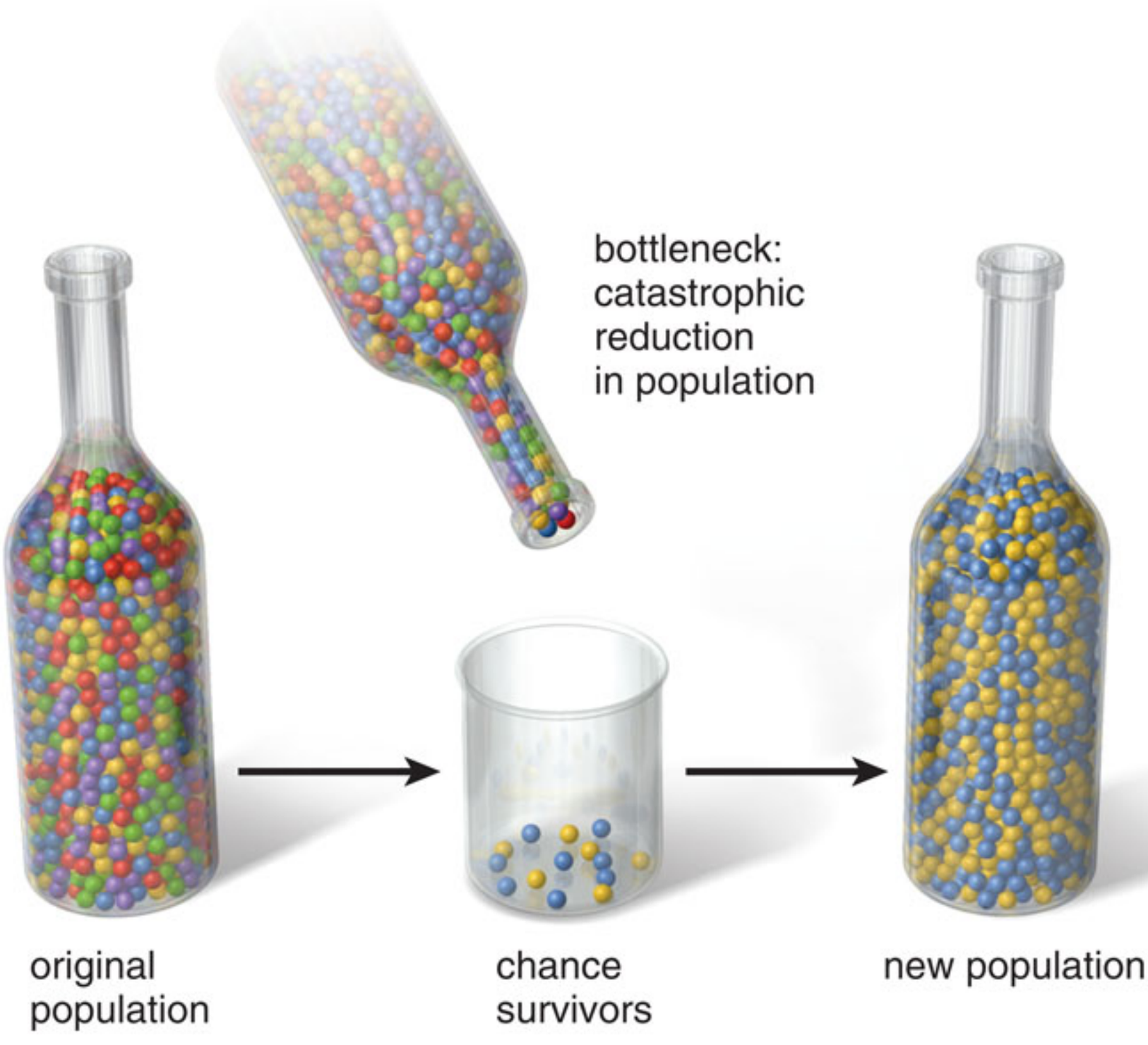
25%

12.5%

6.25%

3.125%

Human history drives our genetics



The common disease - common variant hypothesis

Diseases shared by lots of people
will be caused by variants shared by those same people

How do we find all these common variants?

To find common variants, we need markers shared by lots of people

Goal is to find all the common variants

After the HGP, the HapMap project was born.

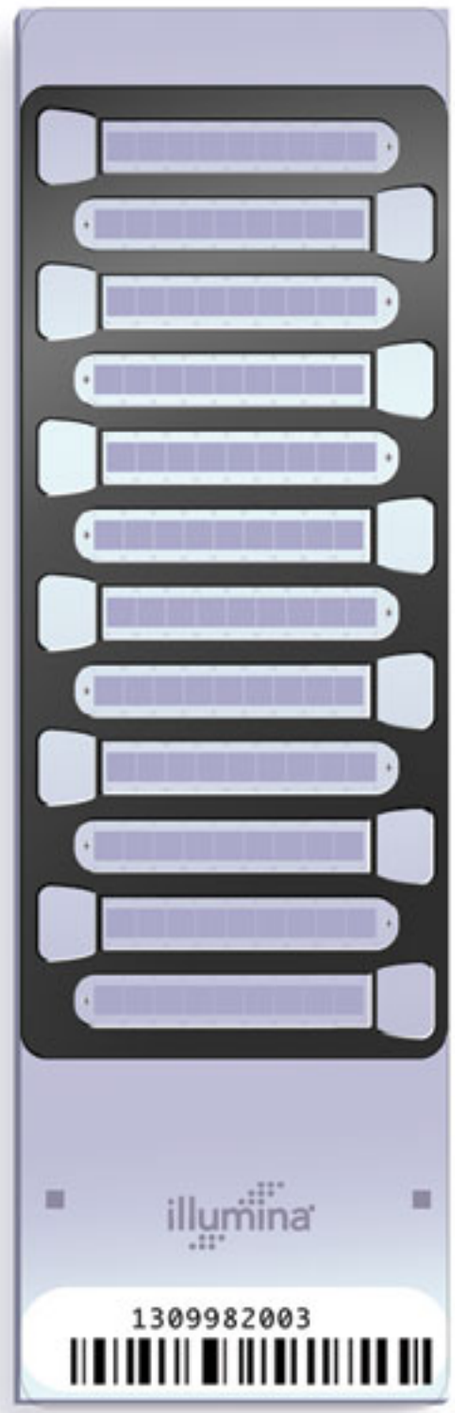
All three types of variation can cause disease

Rare = variants found in less than 1% in population

Common = variants found in more than 5% of the population

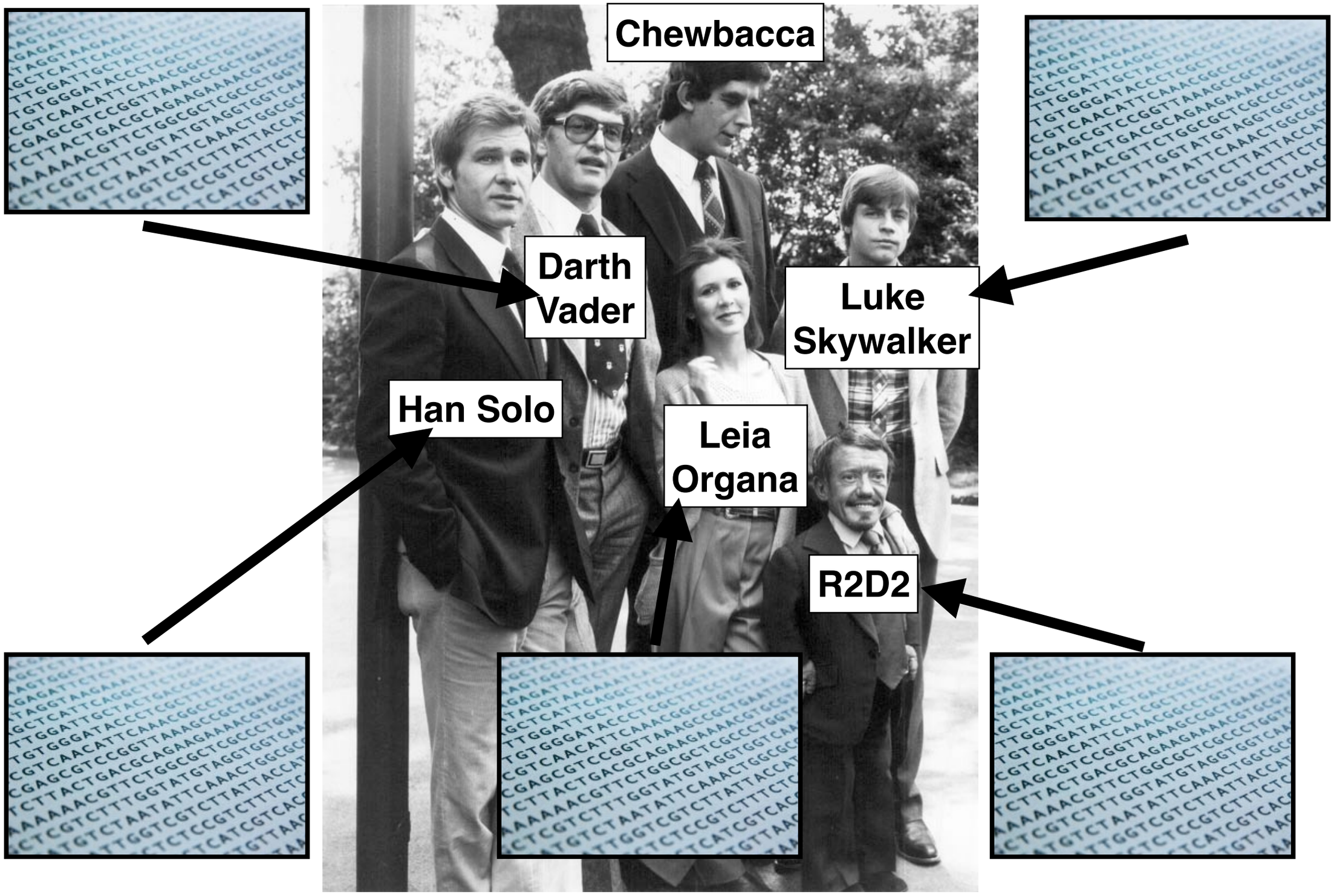
Intermediate = variants found in 1-5% of the population

An array to genotype at >4.3 million sites



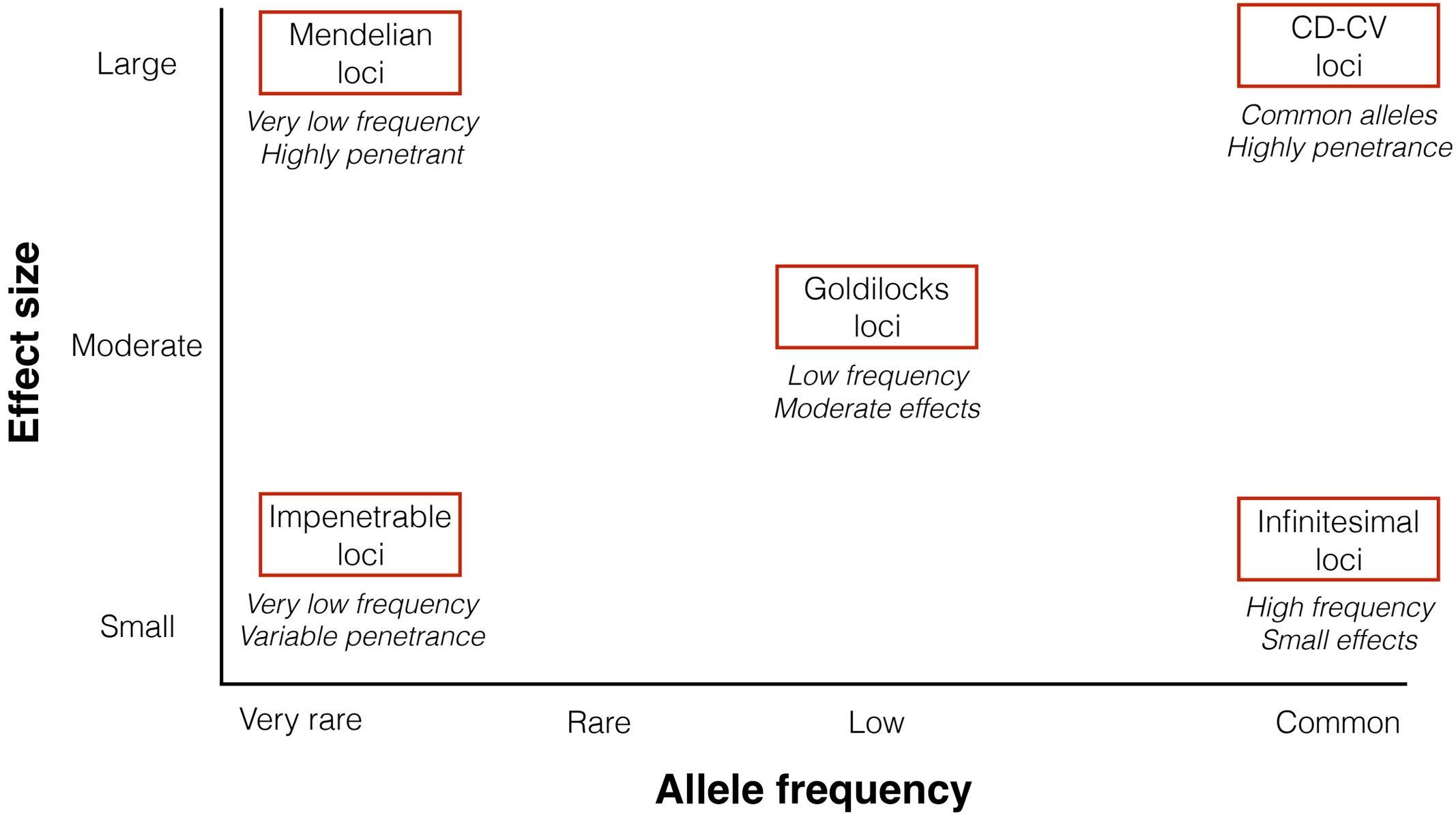
Tool to genotype intermediate and common variation

We want to be able to read genomes and make predictions



The cast of the original *Star Wars*

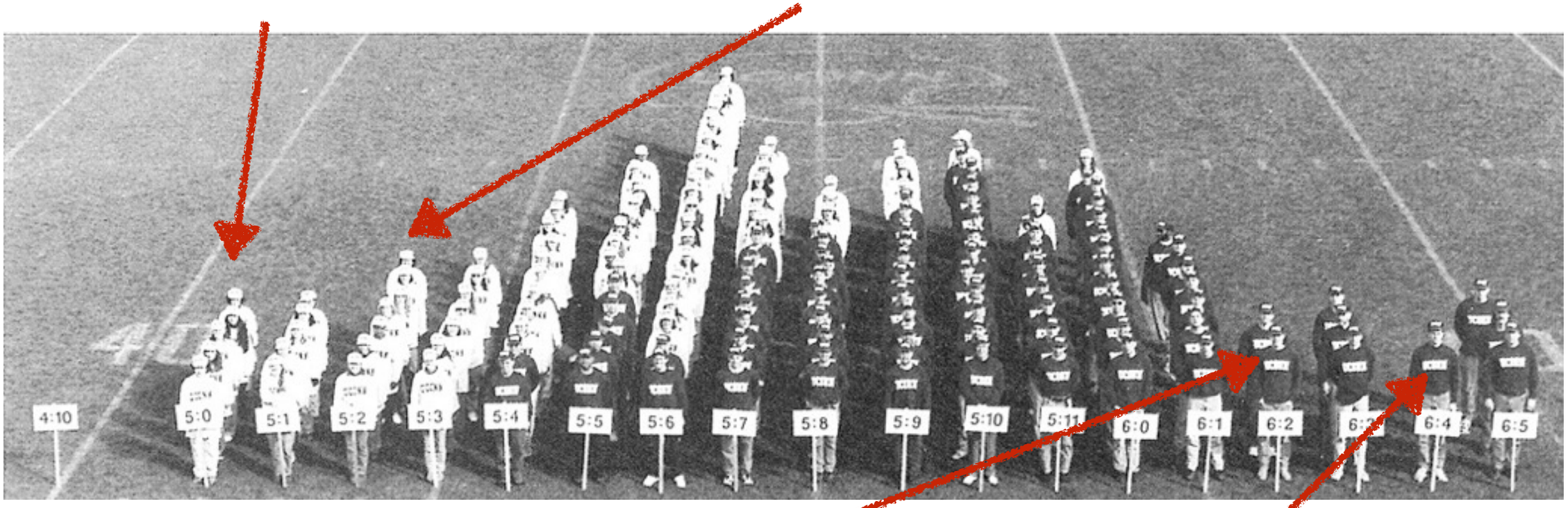
The spectrum of how variation contributes to disease



How do we find the variants that cause common disease?

To find genes in humans, we must correlate genotype with phenotype

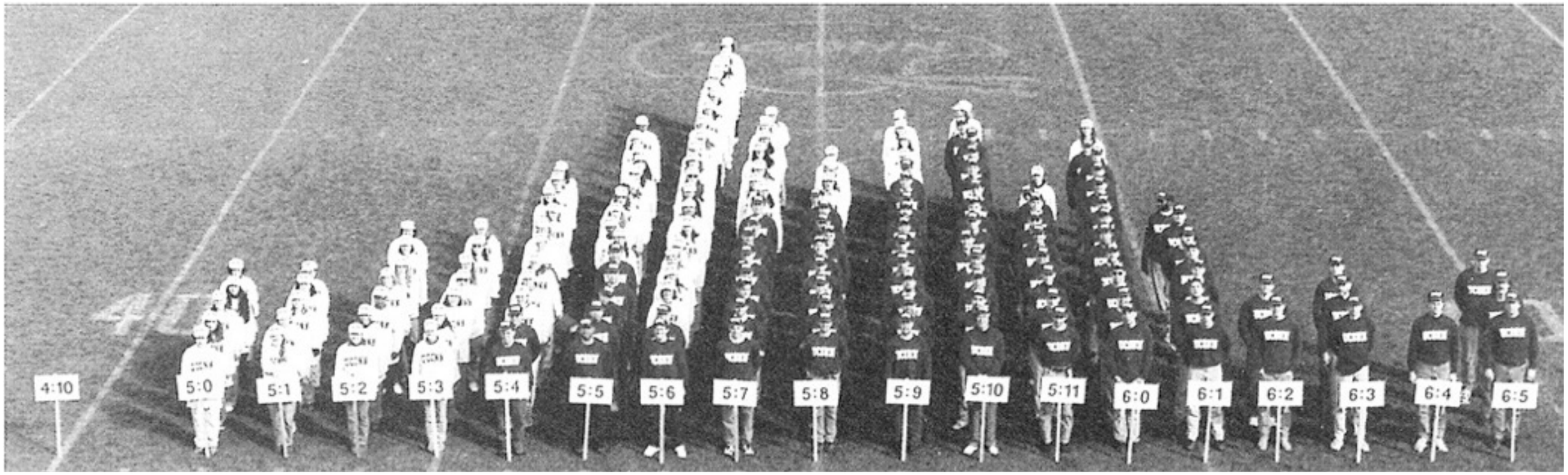
CAGCGATAGGCTTTAATGTT CAGCGATAGGCTTTAATGTT
AGCCCGTTTTATGACCAACG AGCCCGTTTTATGACCAACG
GGGTTCACAGTGAGCTGTGT GGGTTCACAGTGAGCTGTGT



University of Connecticut, 1997

CAGCGATAGGCTTTAATGTT CAGCGATAGGCTTTAATGTT
AGCCCGTTTGATGACCAACG AGCCCGTTTGATGACCAACG
GGGTTCACAGTGAGCTGTGT GGGTTCACAGTGAGCTGTGT

For traits controlled by many genes, we need many, many people

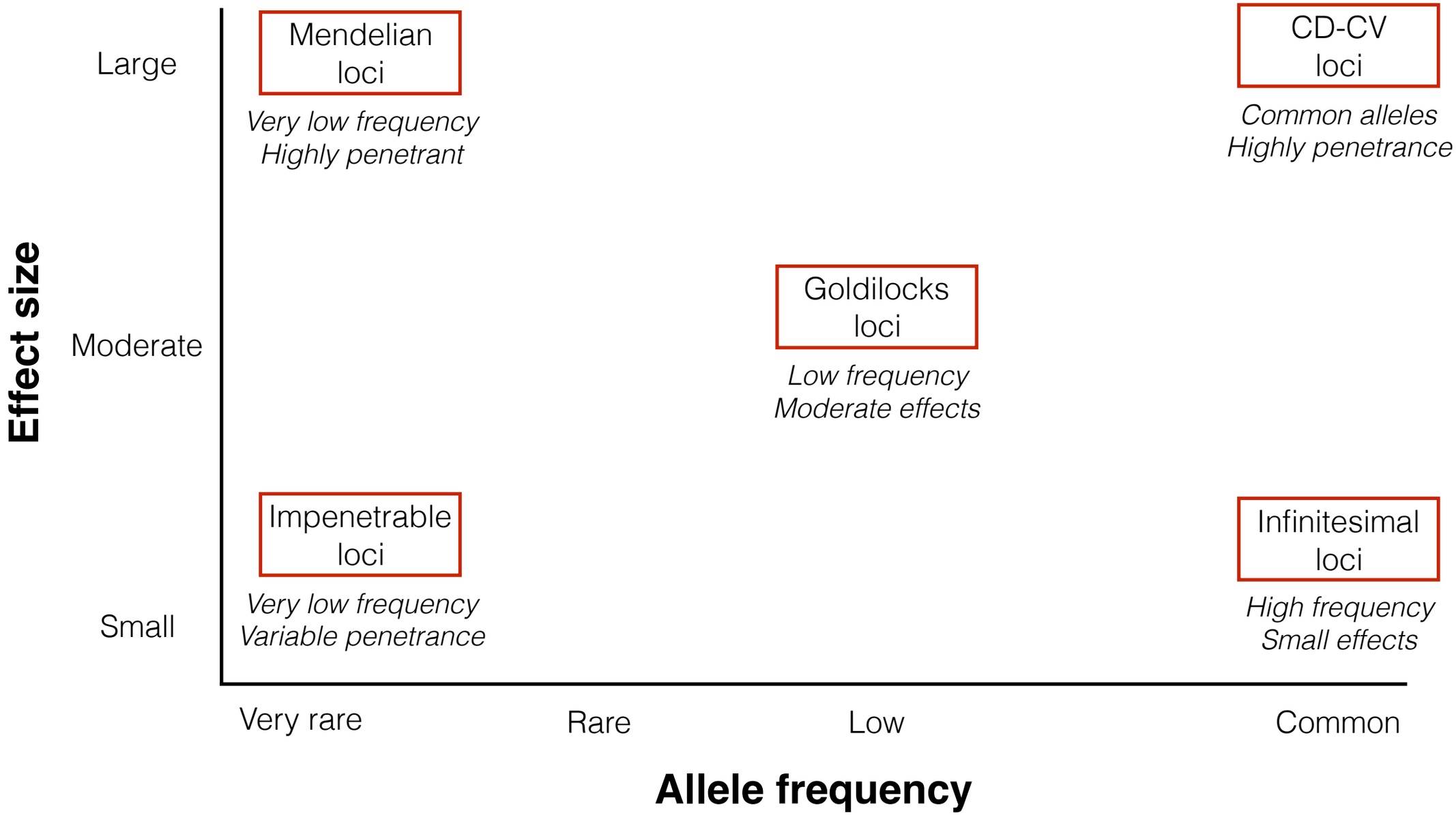


University of Connecticut, 1997

Variation shared by lots of tall people
and not shared by lots of short people

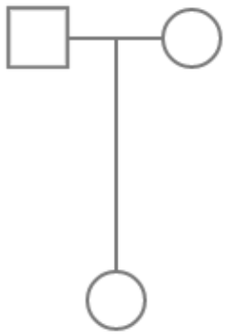
~250,000 people genotyped led to 20%
of height differences explained

The spectrum of how variation contributes to disease

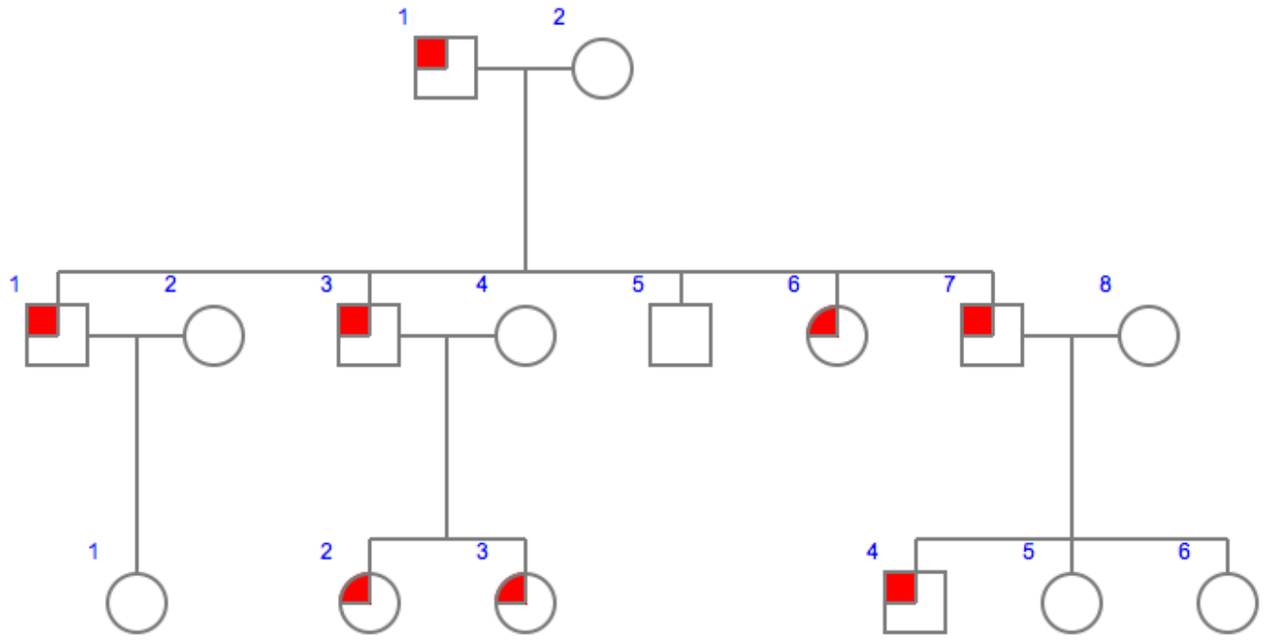


How do we find the variants that cause rare disease?

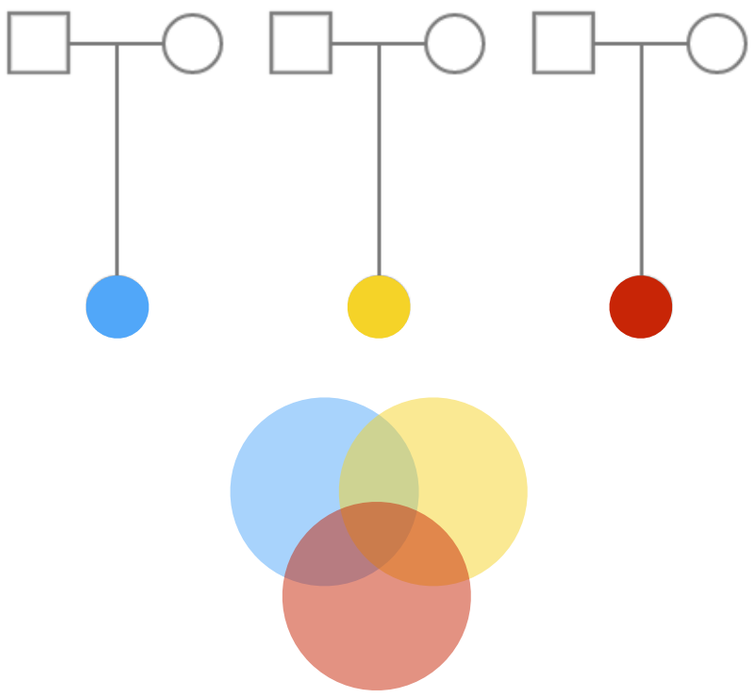
Strategies to identify disease-causing rare variants



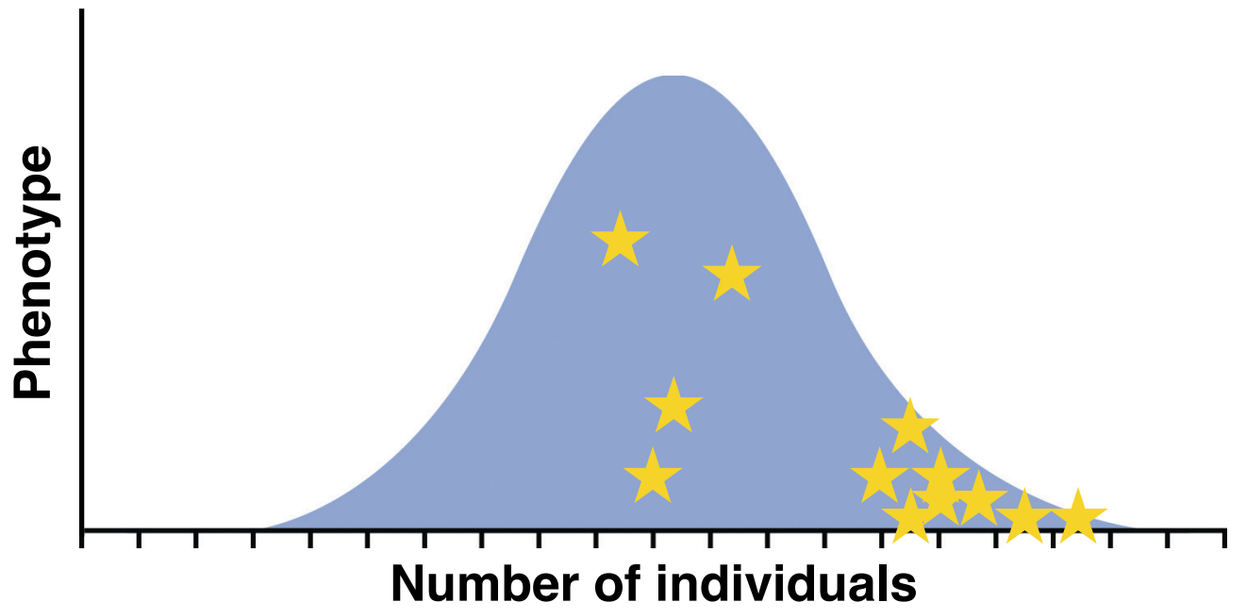
Rare variation from families



Shared variants from affected individuals in large families

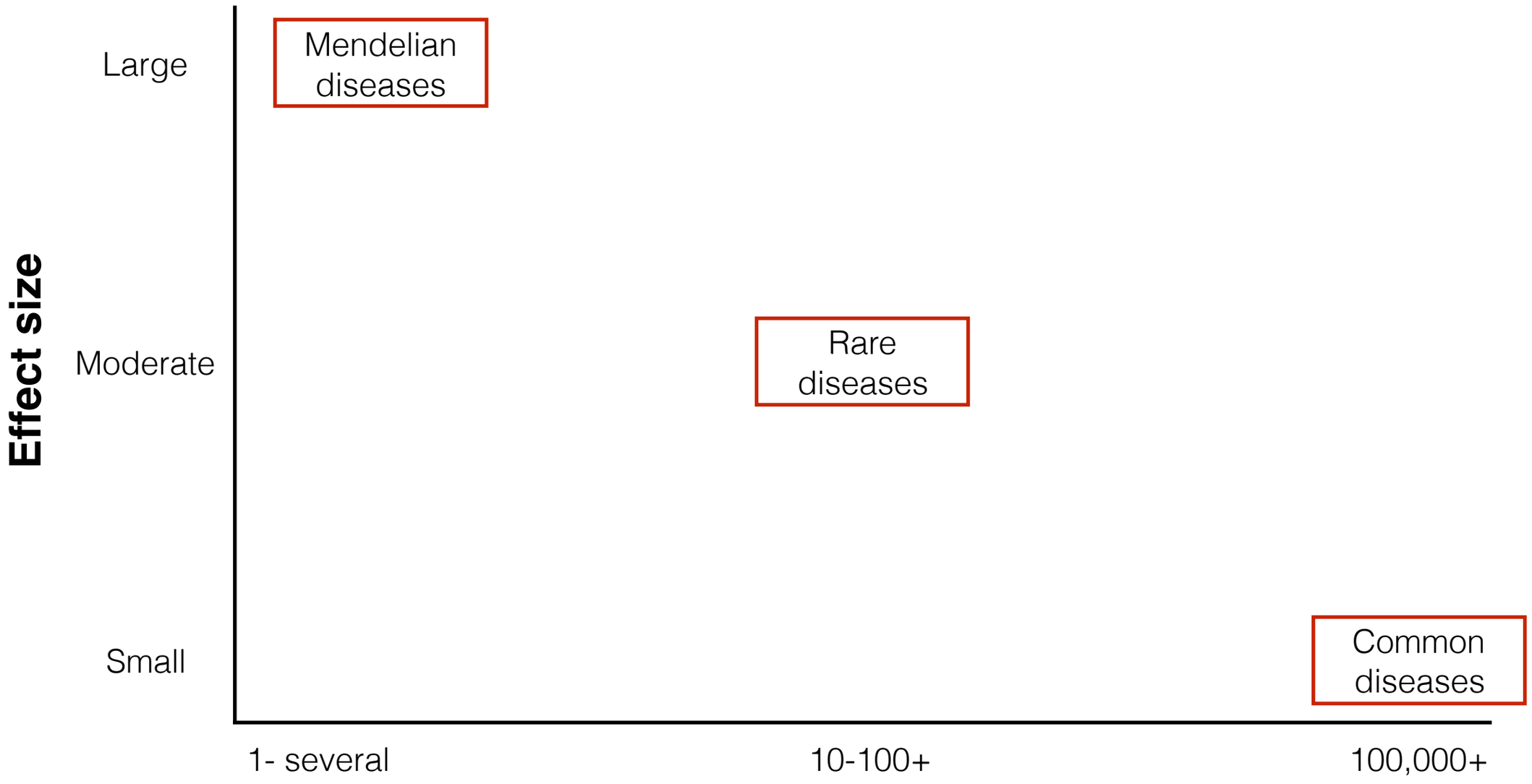


Shared variation from trios



Shared variants from many people

How can sequencing help us to identify these variants?



Why can’t we read the genome?

We don’t know all the variants.

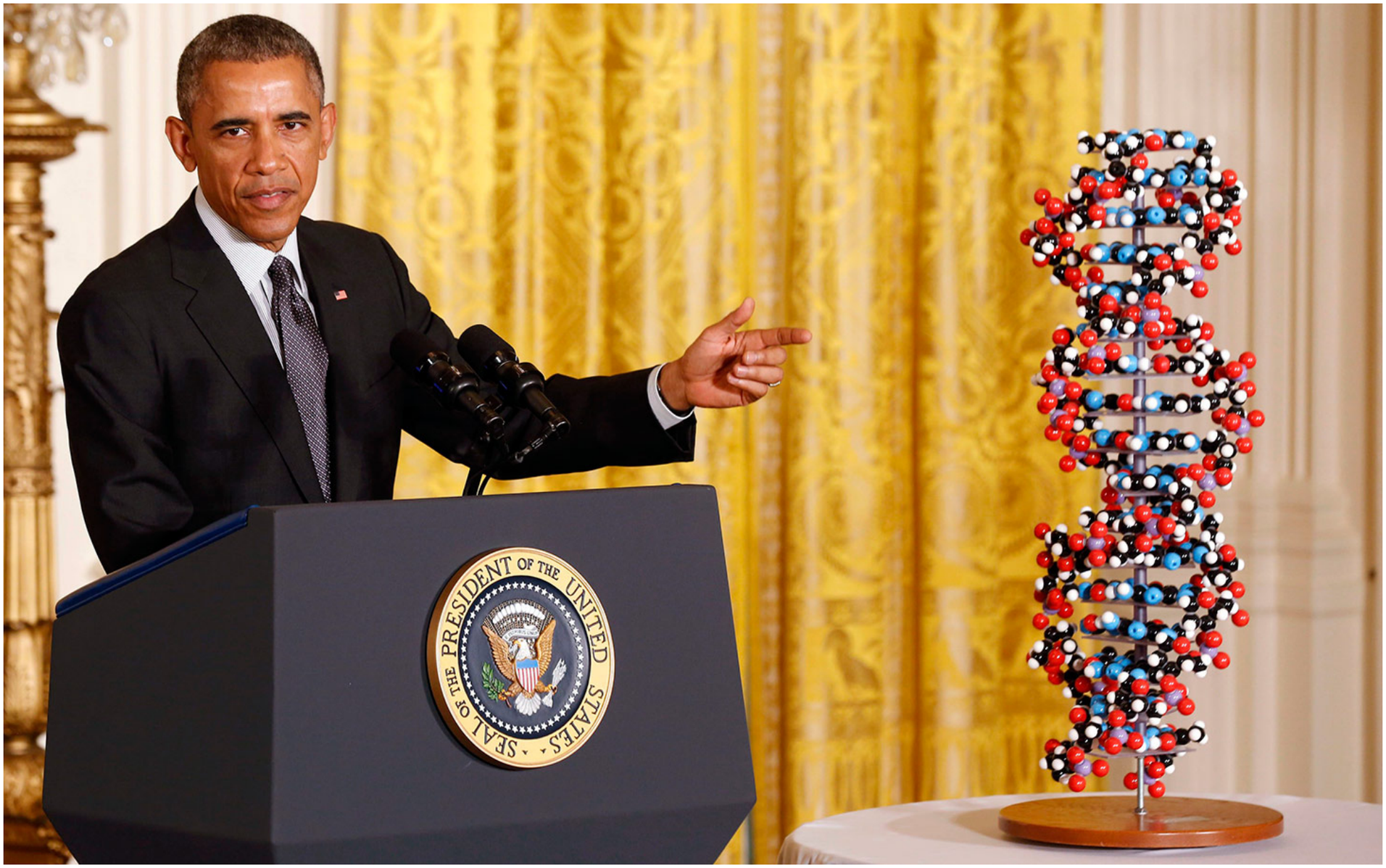
We don’t know which ones affect phenotype.

Single genes don’t cause most disease
or control most traits

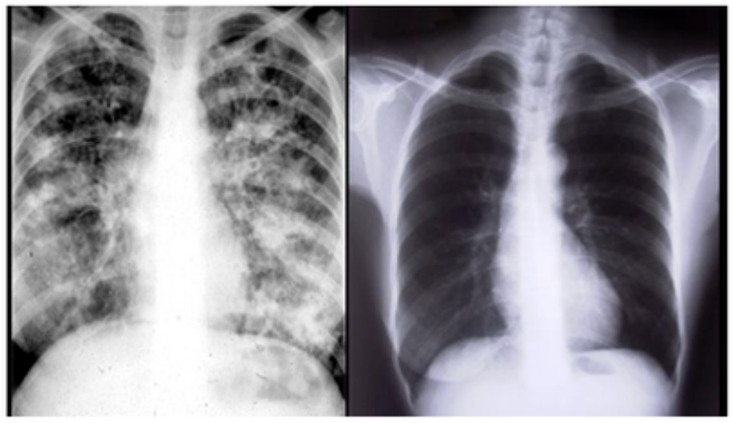
The human genome is big.

Phenotypes are highly variable.

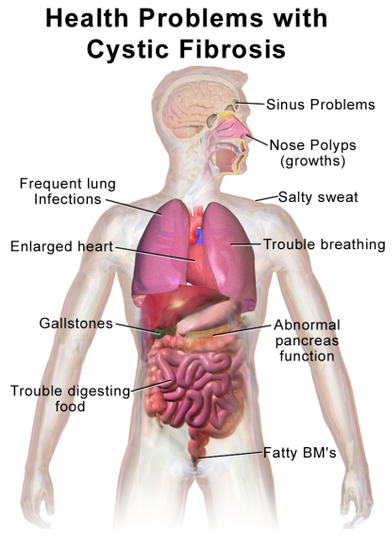
What is precision medicine?



What about cystic fibrosis?



Cystic Fibrosis Lung Healthy Lung



1. Autosomal recessive disorder

2. Not caused by chromosomal aberrations or meiotic NDJ

3. Mapped to chromosome 7

4. Mutations in CF gene are null or hypomorphs

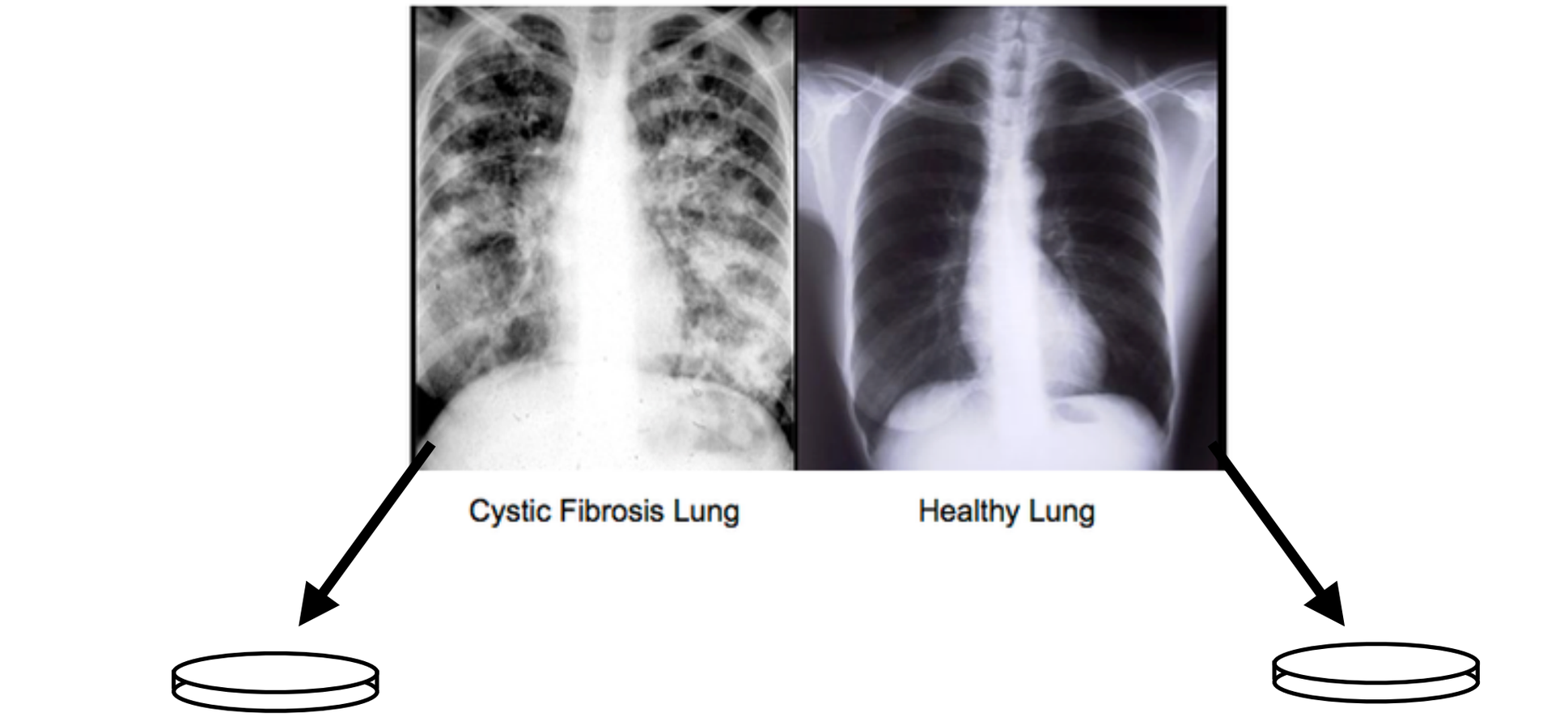
5. Compound heterozygosity (failure to complement) is common

6. No known epistatic genes to CF gene

7. Genetic enhancers are known (immune modulatory genes)

8. No genetic suppressors are known yet.

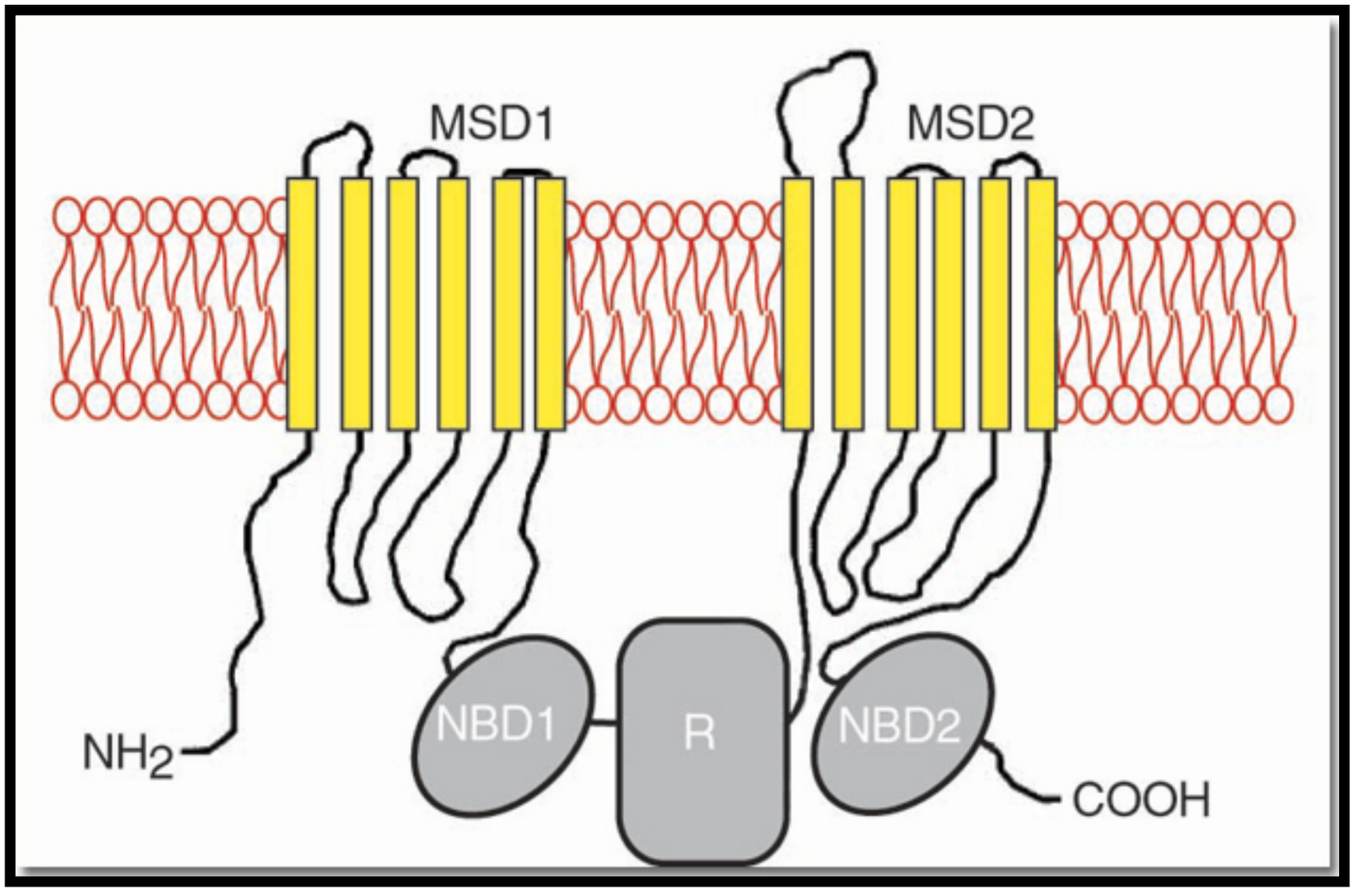
Cell autonomy of CF mutation was shown in the 1960's



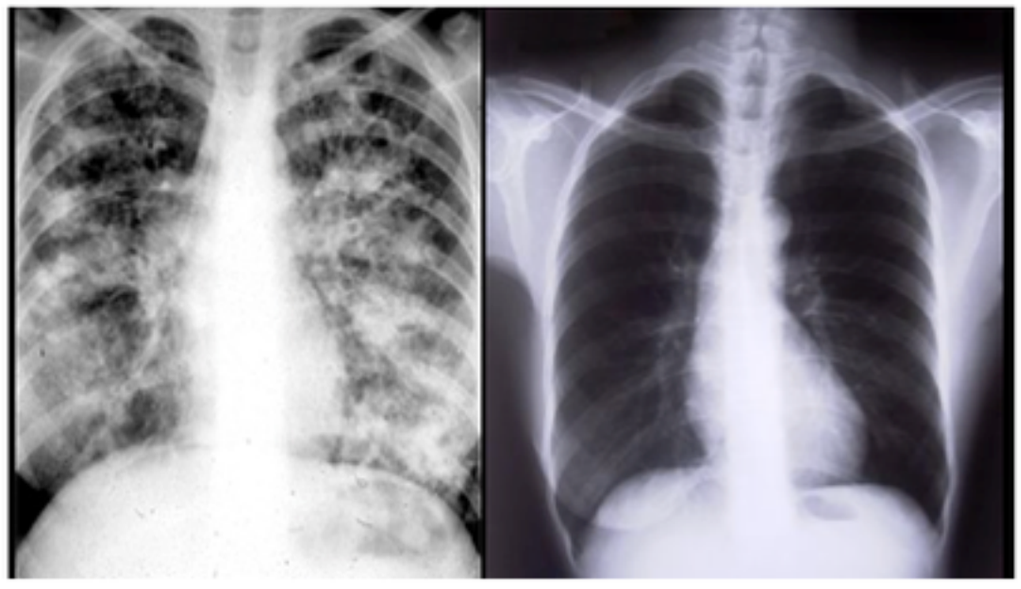
Cells failed to secrete Cl^-

Cells secreted Cl^-

Cystic fibrosis was mapped to the chloride ion channel CFTR



Cystic fibrosis is caused by a mix of common and rare variants



Cystic Fibrosis Lung Healthy Lung

Rare disease affects 1/10,000 live births

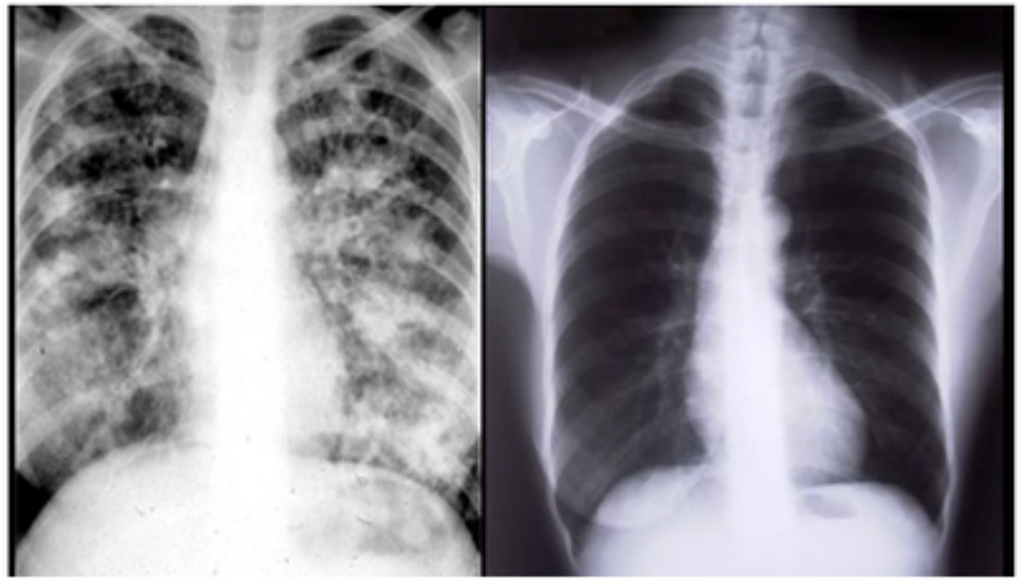
Caused by mutations in the CFTR gene

Selection removes homozygotes from population

H-W equilibrium tell us that 1/50 people are carriers

Why is eugenics (or genome editing) next to impossible?

Cystic fibrosis is caused by a mix of common and rare variants



Cystic Fibrosis Lung Healthy Lung

50% of all cases have the same allele ΔF508

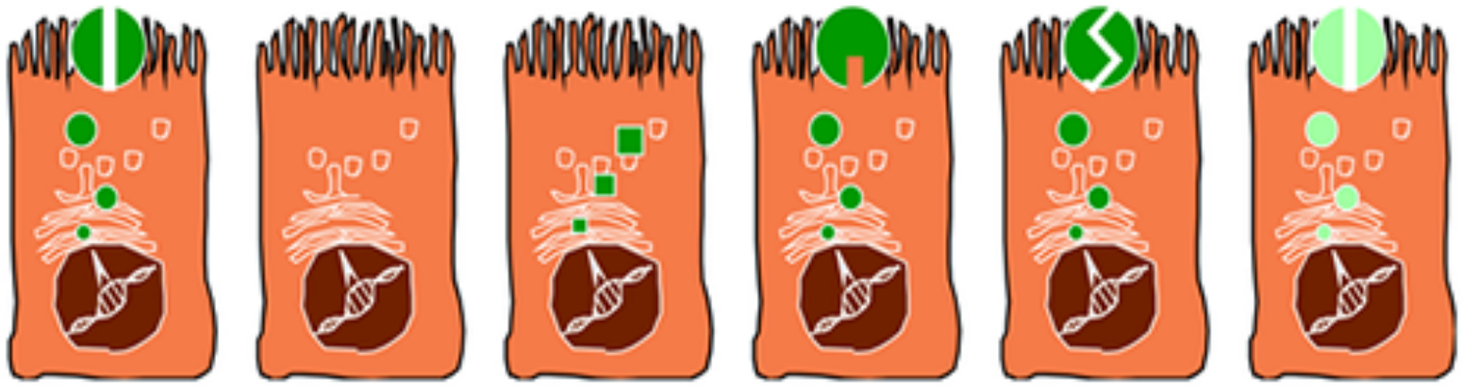
Over 1000 other mutations are known

Compound heterozygotes found often

Genetic heterogeneity

CFTR

Classes of Mutations



	No synthesis	Block in processing	Block in regulation	Altered conductance	Reduced synthesis
	G542X	F508del	G551D	R117H D1152H	3849+10kbC→T 5T A455E

What do you think the phenotypes of these mutations are?

We are living in the human genetics renaissance

Under $1000 genome

Rare disease sequencing for Mendelian disorders

Family genetics

Fetal testing from sequence

Disease outbreaks and diagnosis

Drug response prediction

Cancer genome sequencing