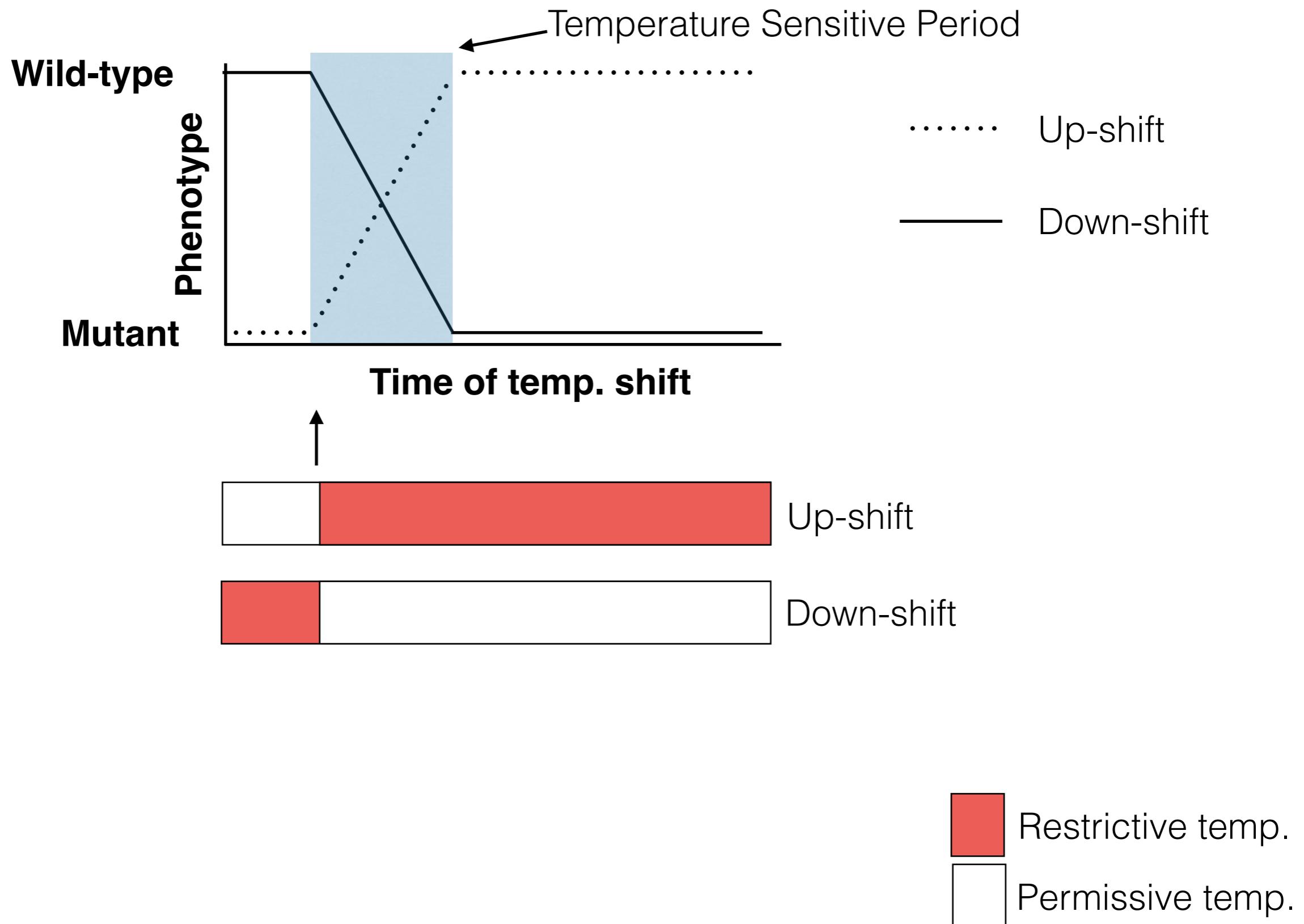


17. Determine time of gene action



Bio393: Genetic Analysis

Human variation and allele frequency spectrum



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



It took more than 10 years and \$3 billion

Who was sequenced?

We don't have one human genome



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

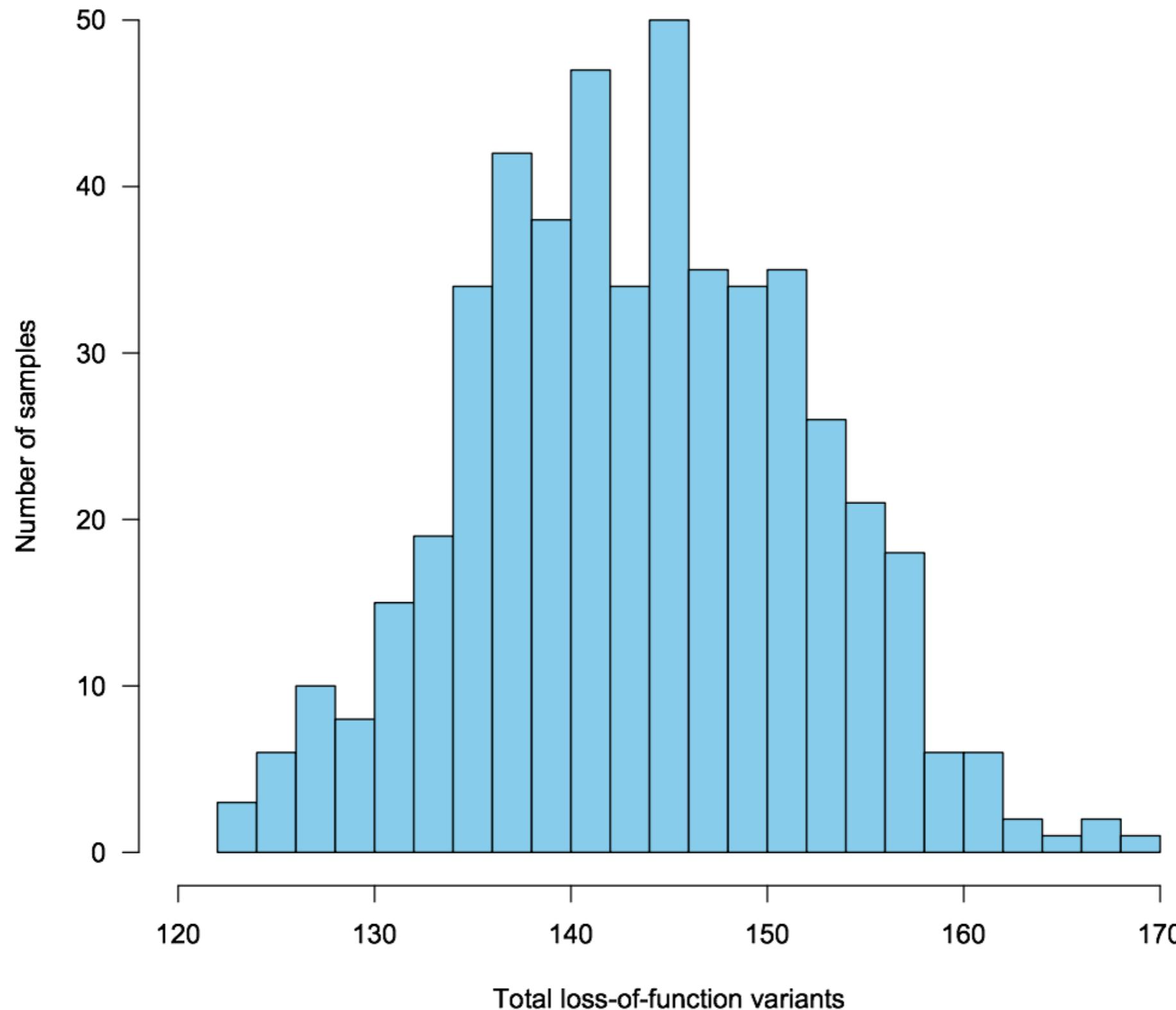
Types of variation

A large block of DNA sequence text, oriented vertically, consisting of multiple lines of sequence data. The sequence is composed of the four bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G).

TAGGC
CCATGGCTGA
TAGCGACGTGACTCTGAATG
CTTTCAAGATCATGCCGTAGGTC
CTACGCTGTACAGGTCCAGCTGATT
AATCGCAGCAGGTTTGAGCTTCAAC
ATGAGCAATGCTCTGAAAATCATAC
TGCTTGACTGATAAGGCTCGAAC
ATCTAGCAAACGTGCTCATGATGCC
AAGTGGCCGTACTTATGATGC
ACGTGTGATATCCAGA
GGACAGTCCAT
CATGC

Rare = variants found in less than 1% in population

We all have over 100 loss-of-function rare variants



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

A large block of DNA sequence text, oriented vertically, consisting of multiple lines of sequence data. The sequence is composed of the four bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G).

TAGGC
CCATGGCTGA
TAGCGACGTGACTCTGAATG
CTTTCAAGATCATGCCGTAGGTC
CTACGCTGTACAGGTCCAGCTGATTC
AATCGCAGCAGGTTTGAGCTTACAG
TGAGCAATGCTCTGAAAATCATAC
TGCTTGACTGATAAGGCTCGAAC
CTAGCAAACGTGCTCATGATGCC
AGTGGCCGTACTTATGATGC
ACGTGTGATATCCAGA
GACAGTCCATG
CATG

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?

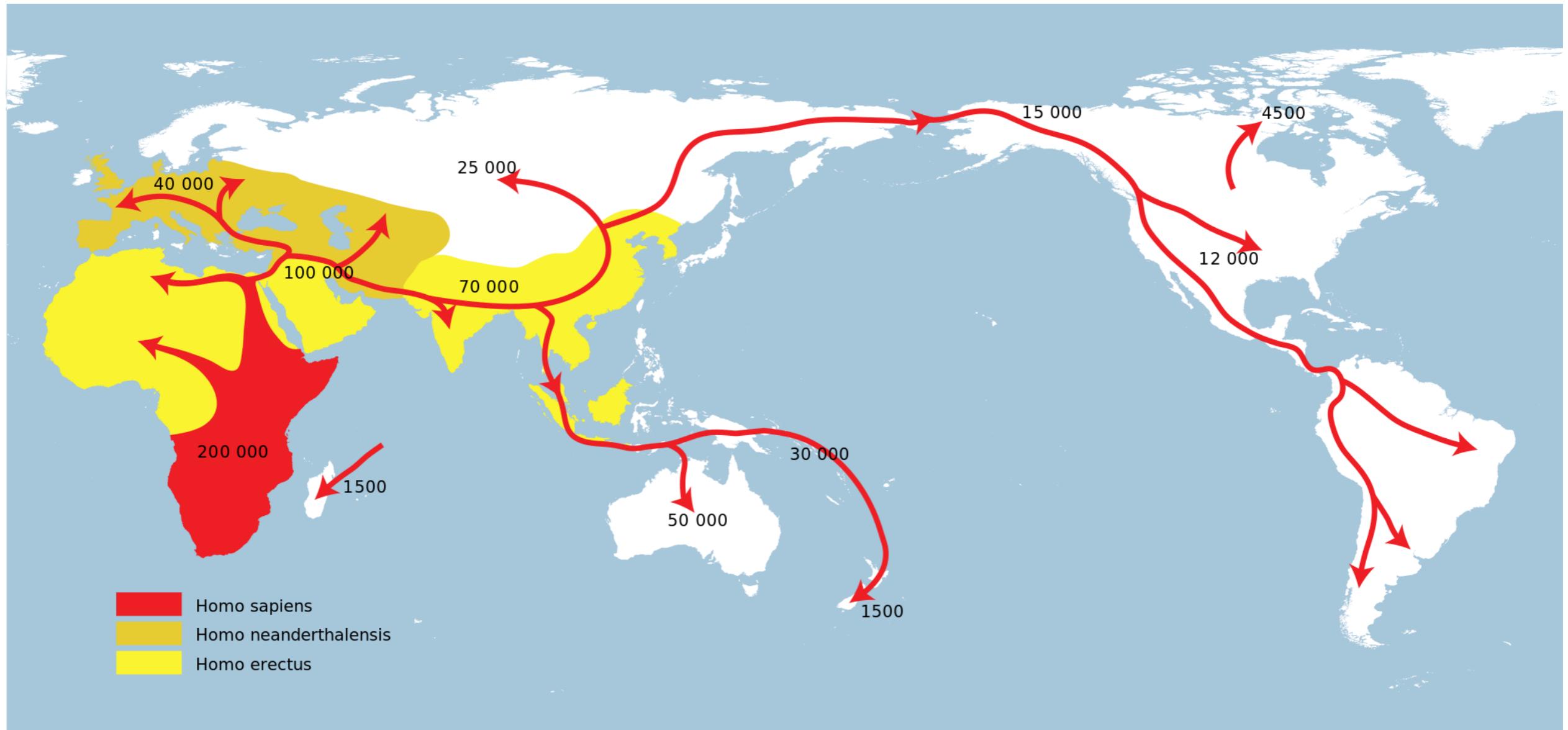
A large block of DNA sequence text, oriented vertically, representing a genome. It consists of four lines of sequence, each starting with a capital letter and followed by lowercase letters representing the four nucleotides: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). The sequence is: T A G G C C A T G G C T G A A G T A T C T G A A T G C T G C A C G T G A C T C T G A A T G C T T C A A G A T C A T G G C C G T A G G T C G A T T N C T A C G C T G T C A C A G G T C C A G G T C A G G T T T C G A G C T T C A C G A A T G C T G C A A A A T C A T A C G A T G A G C A A T G C T G C T G C A A G G C T C G A A C T G C T T G A C T G A T C A A G G C T C A T G C C A T C T A G C A A A C G T G C T C A C T T A T G A T G C A A G T G G C C G T C A C T T A T G A T C A G T G T G A T A T C C A G A C A G T C C A T G C A T G C

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

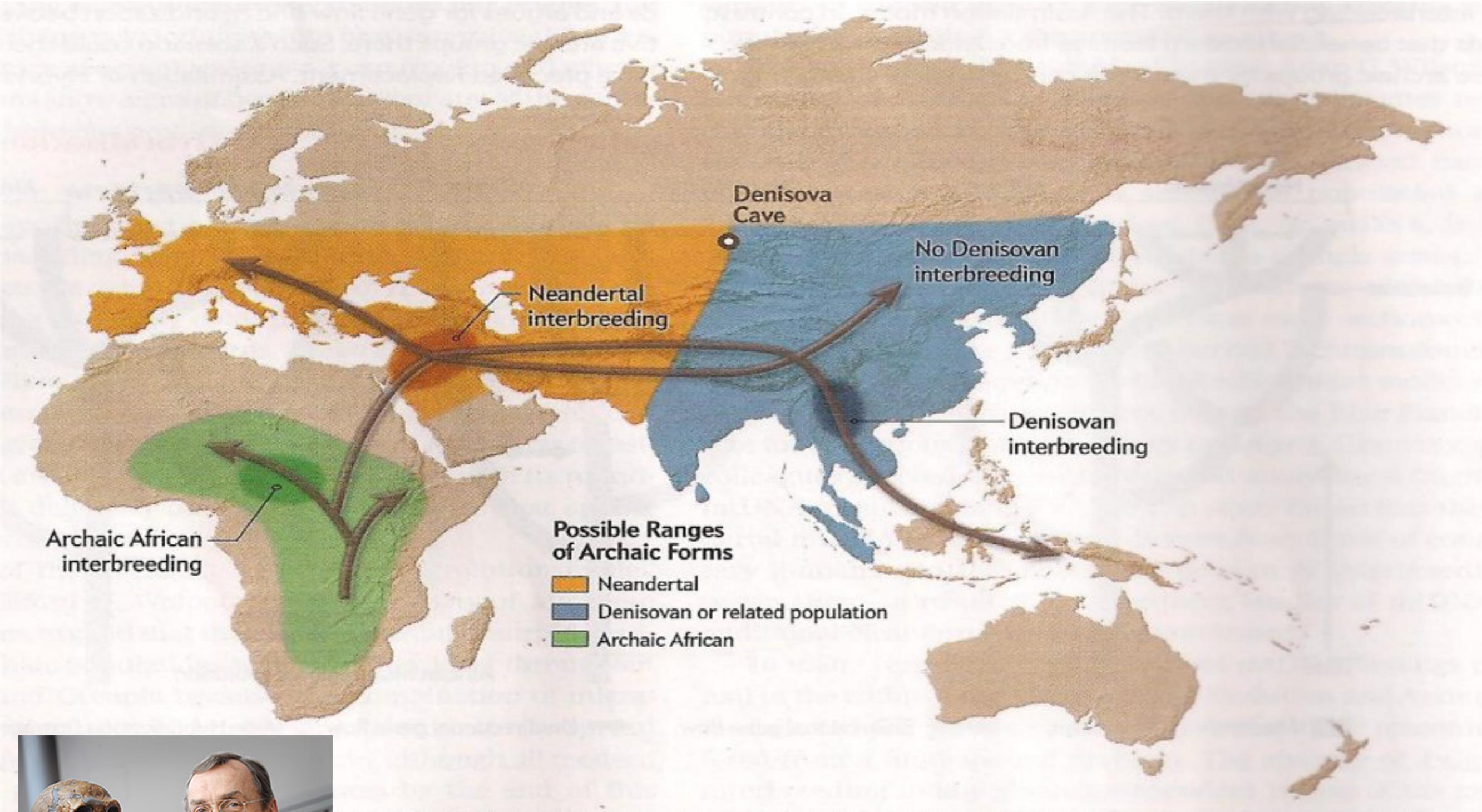
Somatic or germline errors

Once generated, germline variants are inherited

Human history drives our genetics



Human history drives our genetics

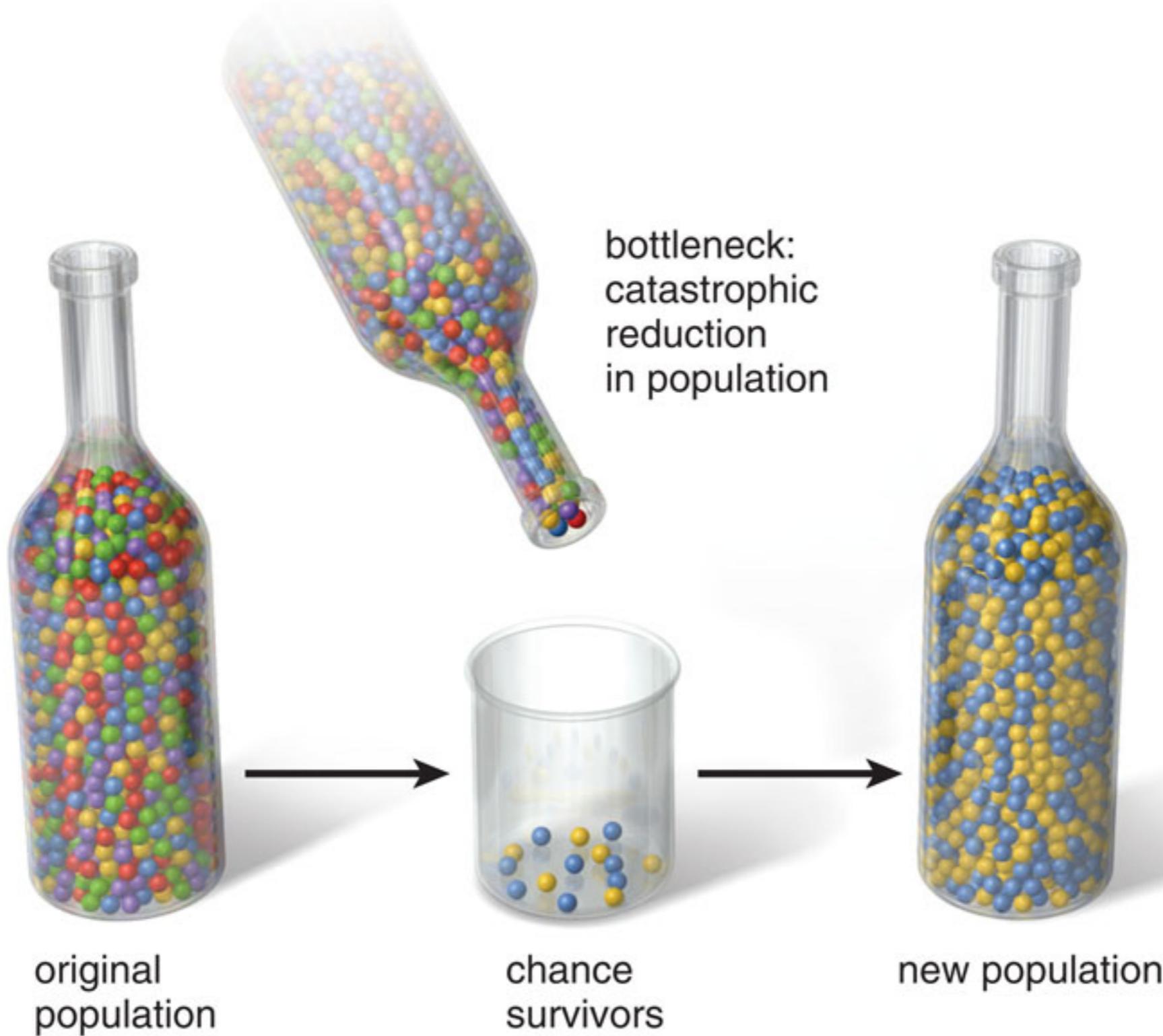


Svante Pääbo

Lecture 13

Human history drives our genetics

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



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

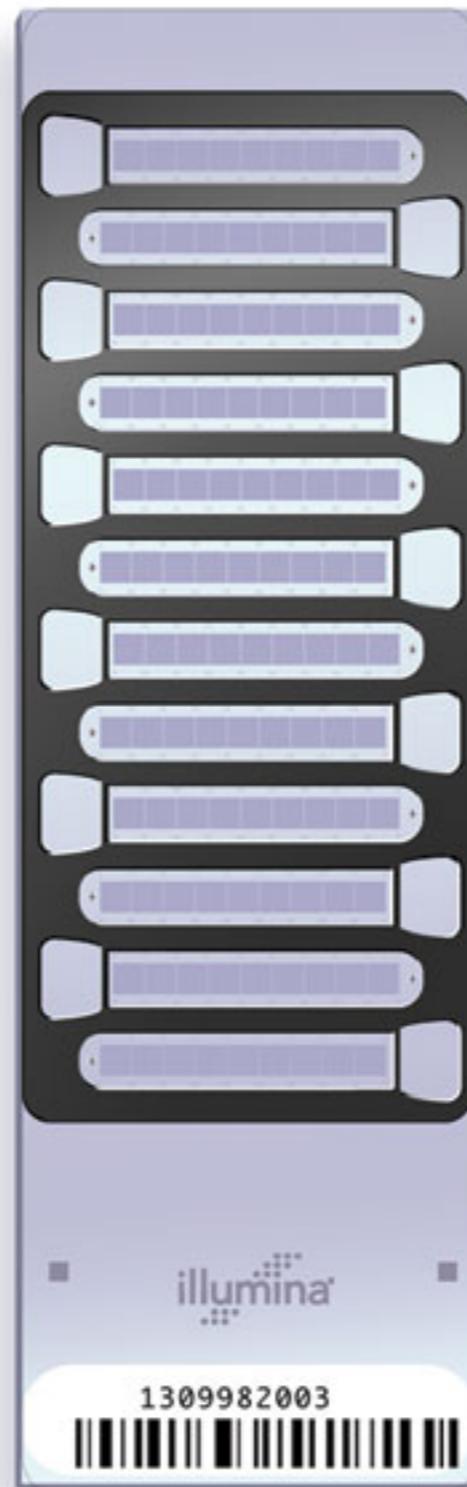


Rare = variants found in less than 1% in population

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

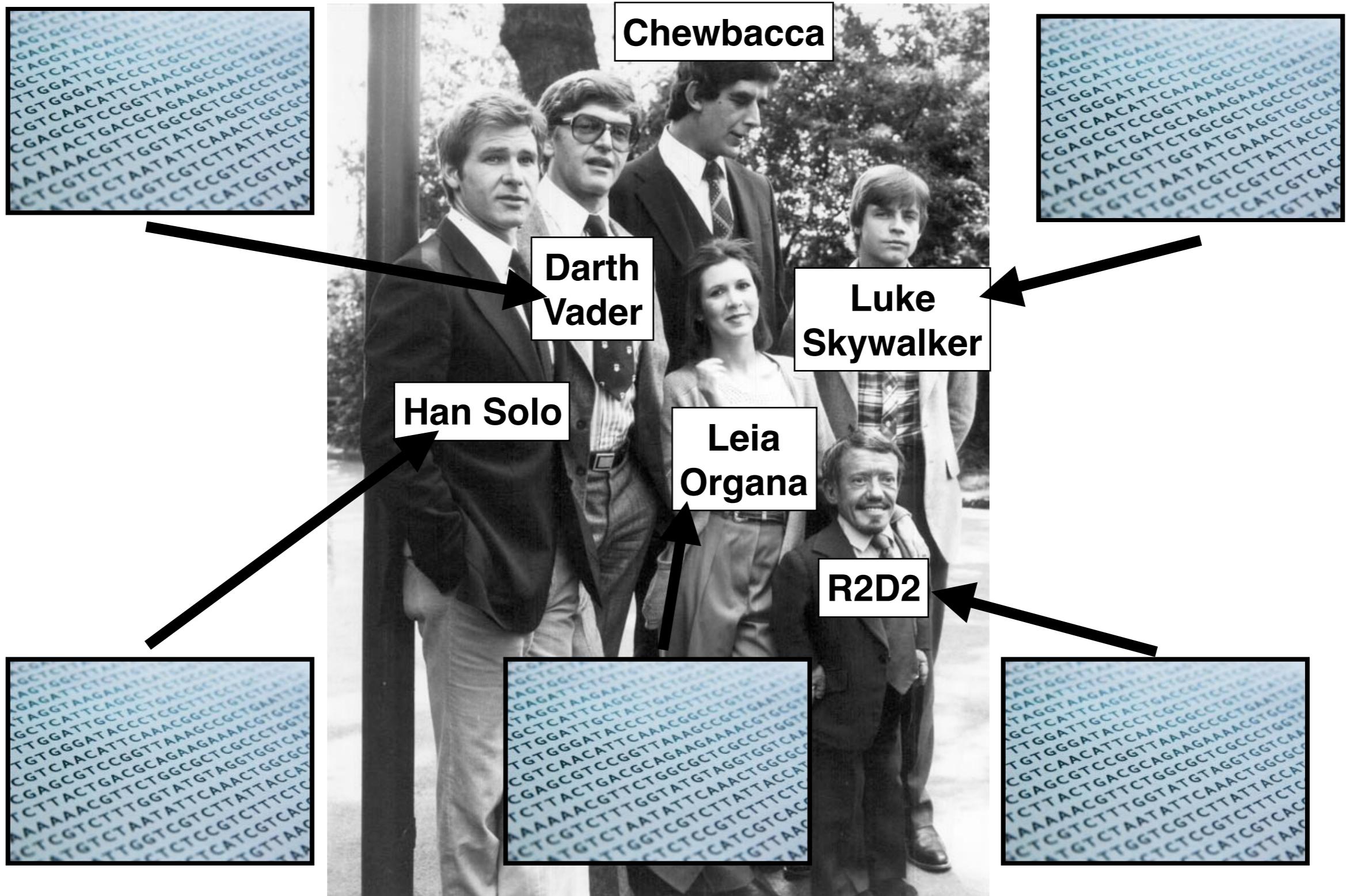
Intermediate = variants found in 1-5% of the population

An array to genotype at >4.3 million sites



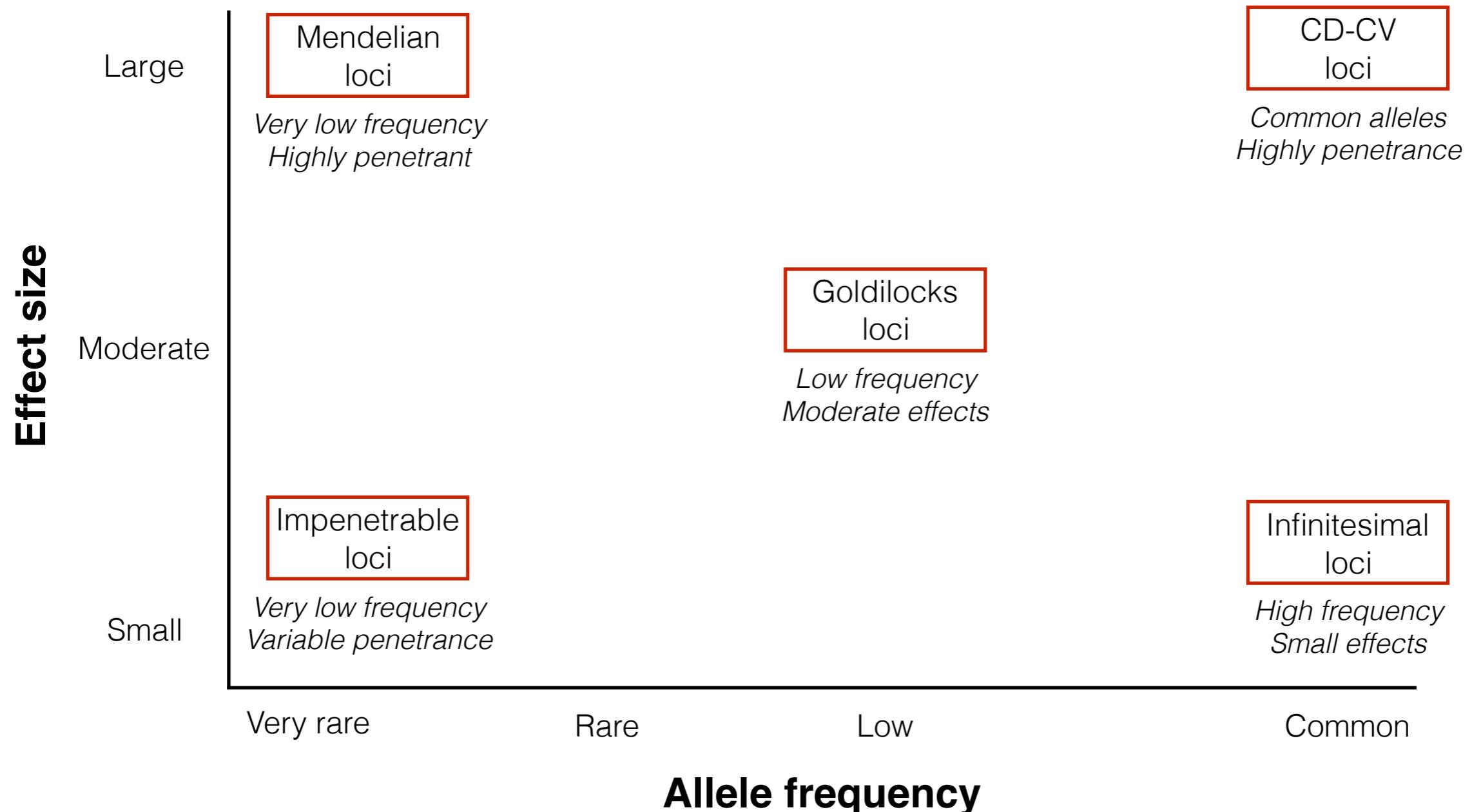
Tool to genotype intermediate and common variation

We want to be able to read genomes and make predictions



The cast of the original *Star Wars*

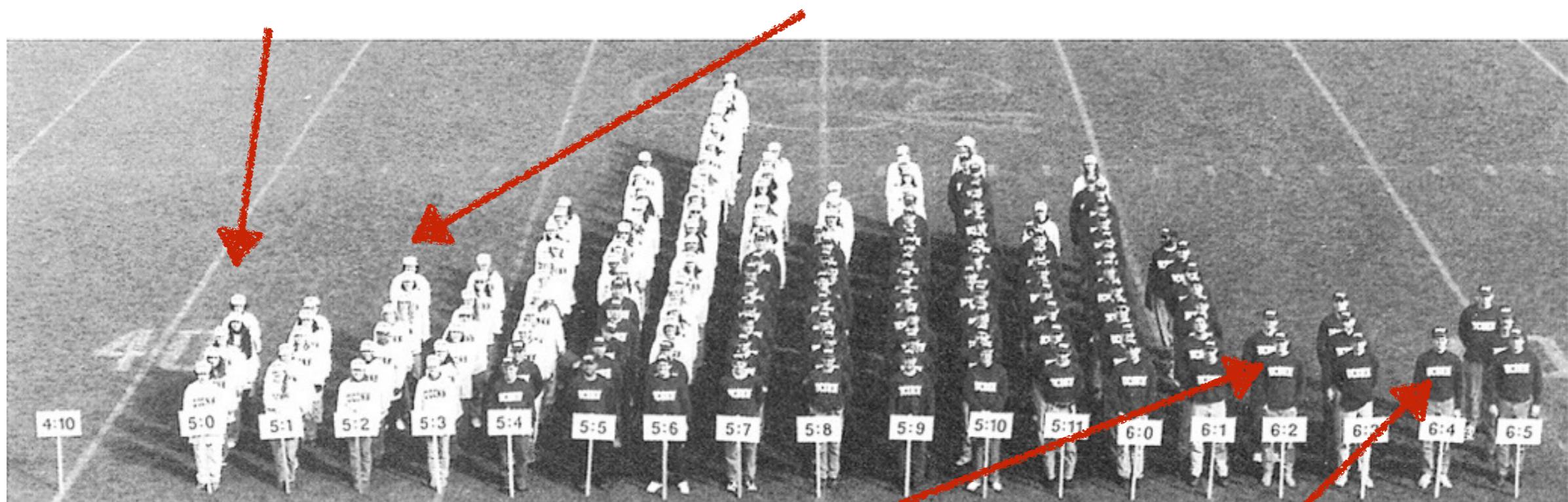
The spectrum of how variation contributes to disease



How do we find the variants that cause common disease?

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

CAGCGATAGGCTTAATGTT	CAGCGATAGGCTTAATGTT
AGCCC G T T T <u>T</u> ATGACCAACG	AGCCC G T T T <u>T</u> ATGACCAACG
GGGTTCACAGTGAGCTGTGT	GGGTTCACAGTGAGCTGTGT

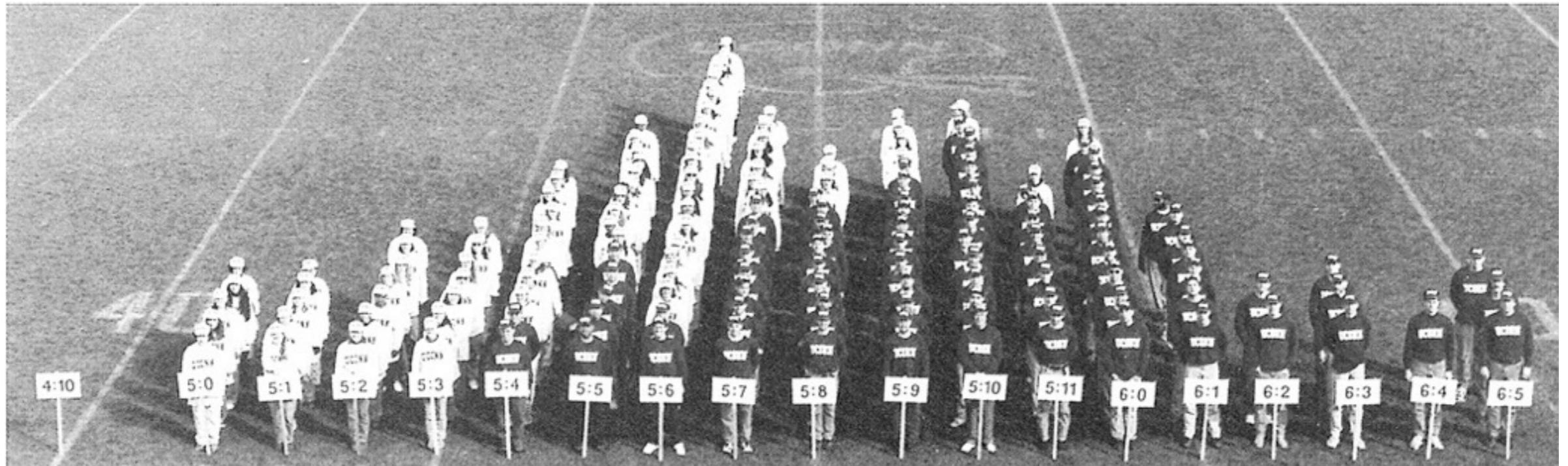


University of Connecticut, 1997

CAGCGATAGGCTTAATGTT
AGCCC G T T T <u>G</u> ATGACCAACG
GGGTTCACAGTGAGCTGTGT

CAGCGATAGGCTTAATGTT
AGCCC G T T T <u>G</u> ATGACCAACG
GGGTTCACAGTGAGCTGTGT

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

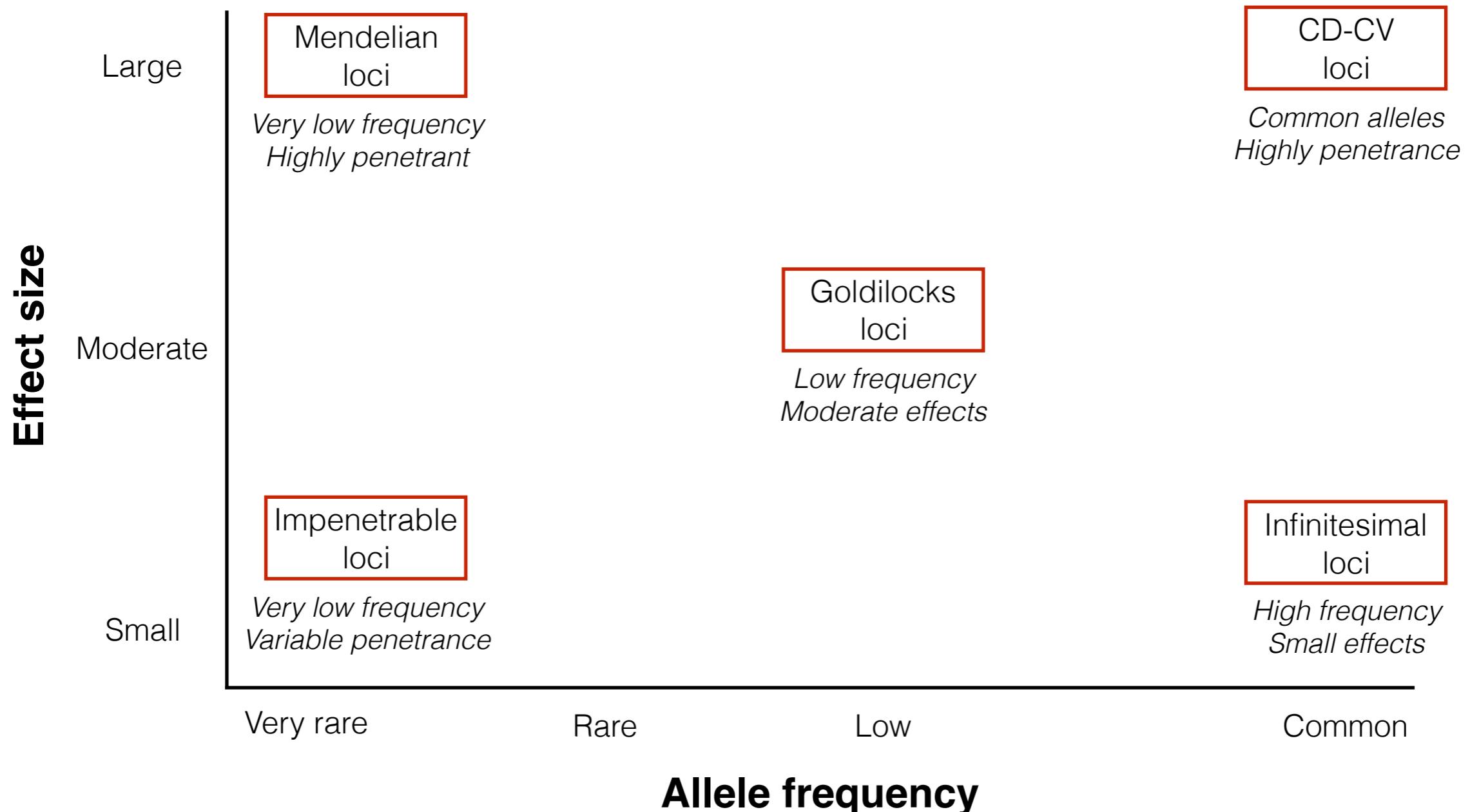


University of Connecticut, 1997

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

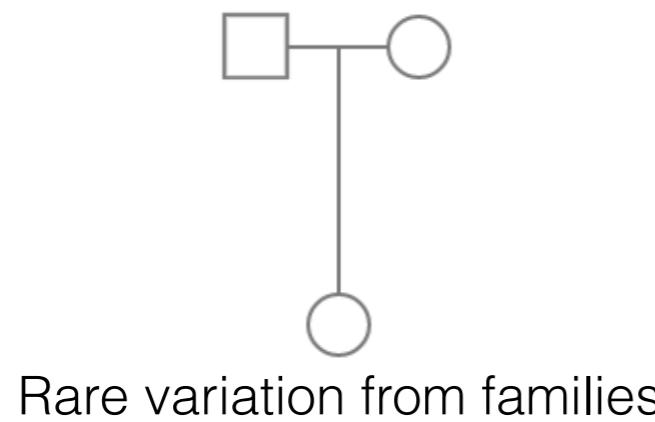
~250,000 people genotyped led to 20%
of height differences explained

The spectrum of how variation contributes to disease

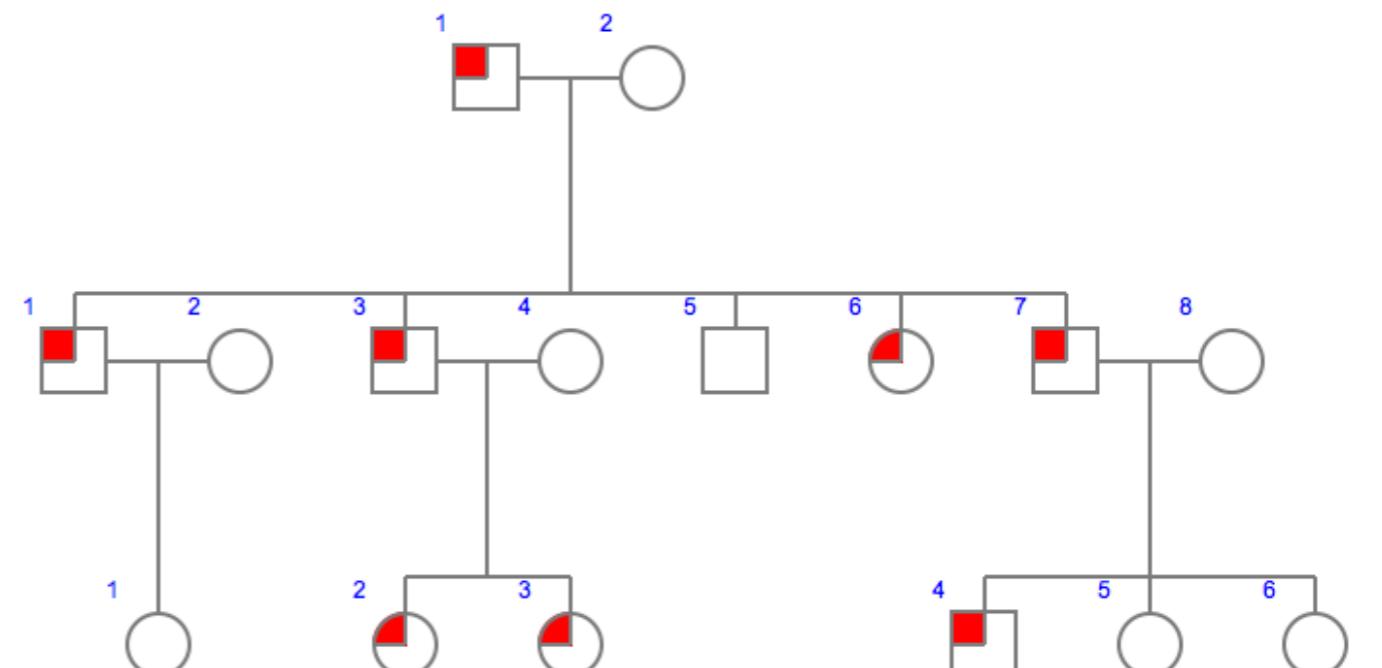


How do we find the variants that cause rare disease?

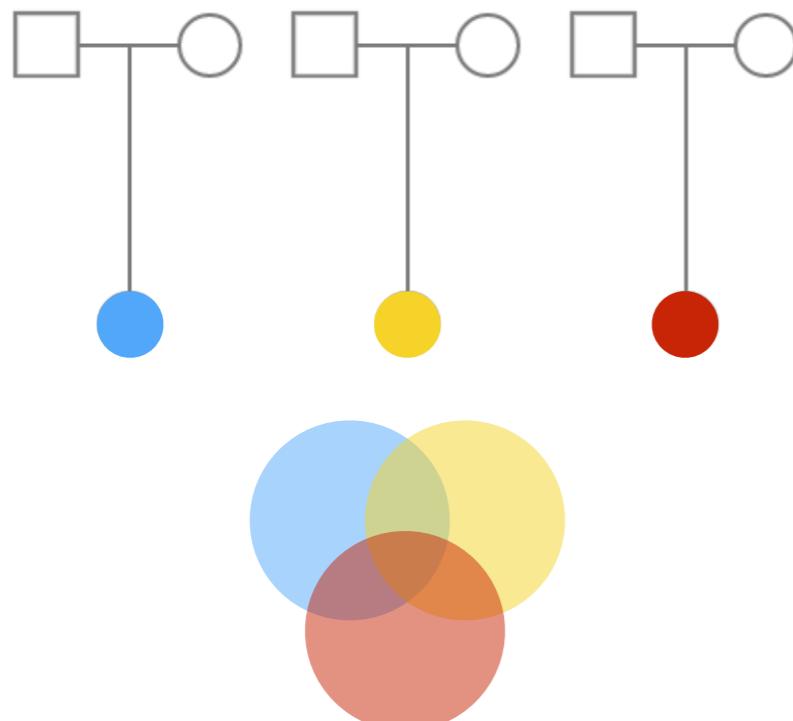
Strategies to identify disease-causing rare variants



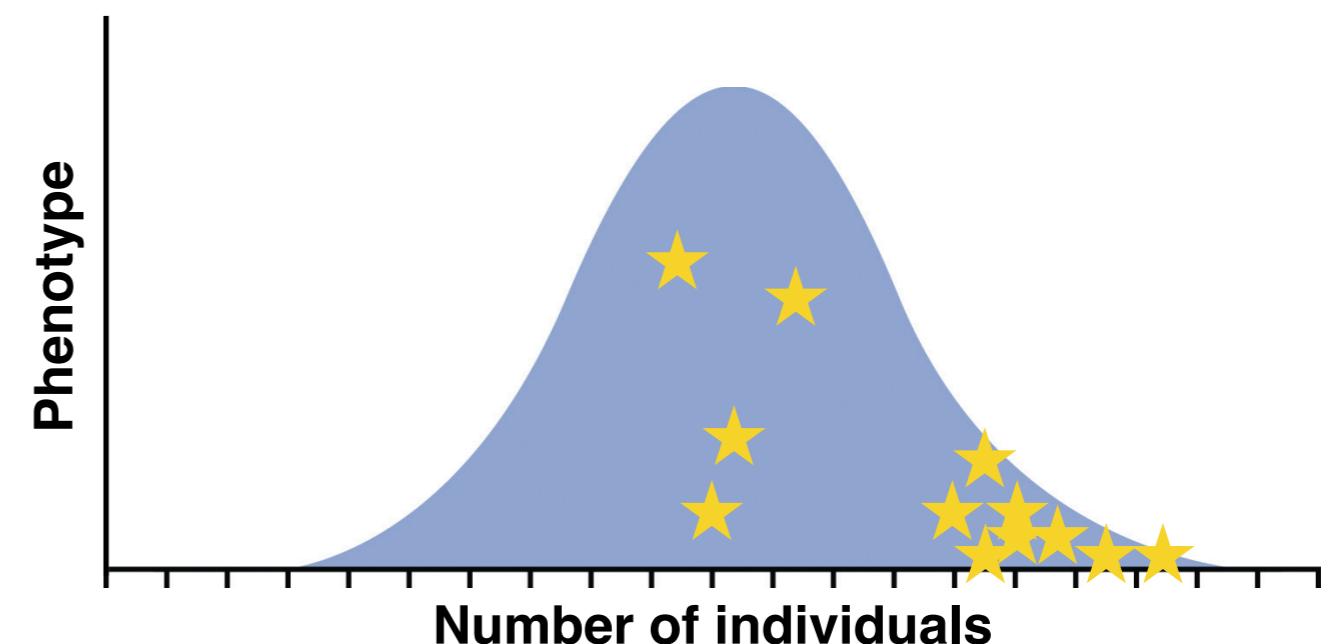
Rare variation from families



Shared variants from affected individuals in large families

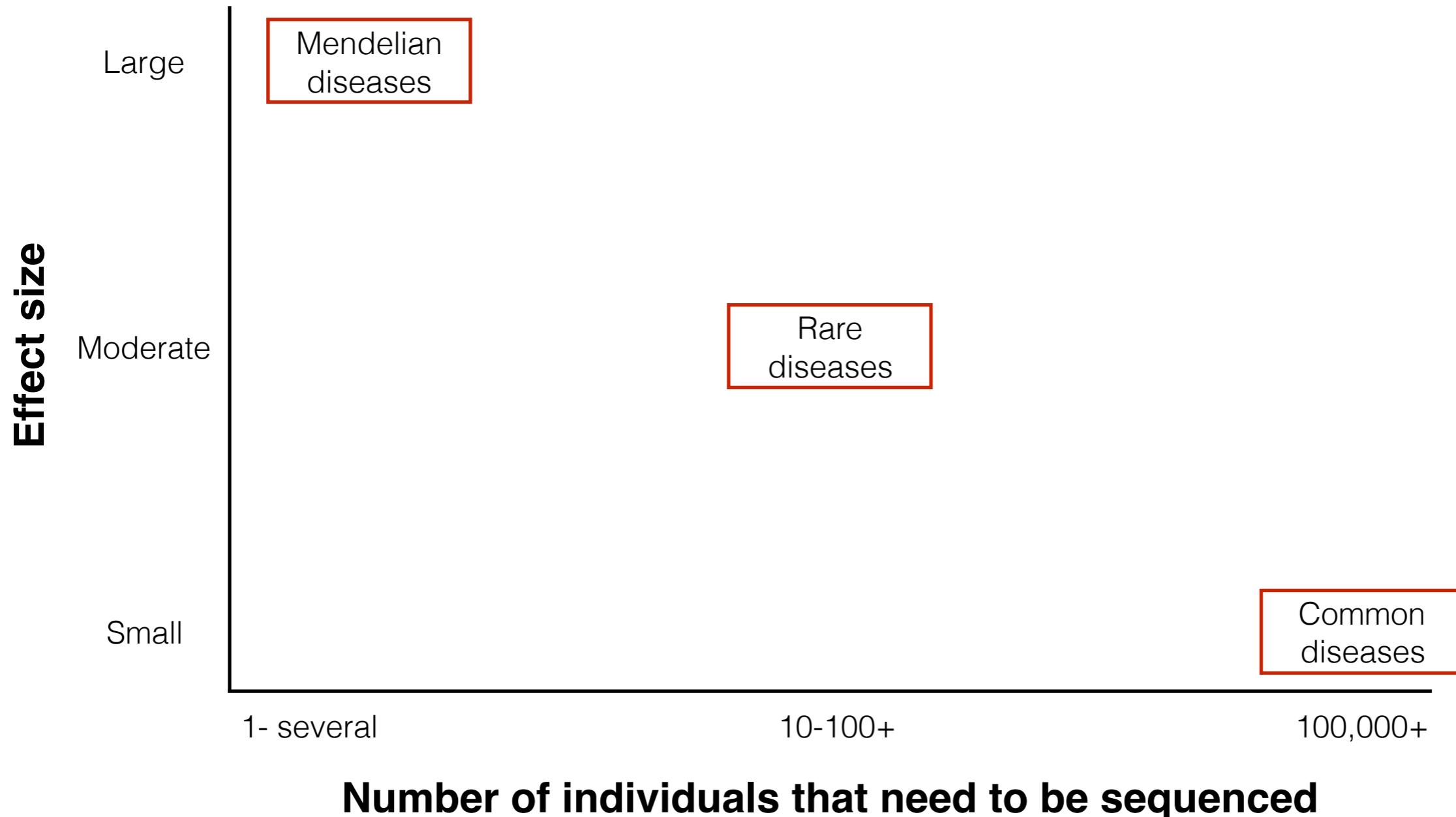


Shared variation from trios



Shared variants from many people

How can sequencing help us to identify these variants?



Why can't we read the genome?



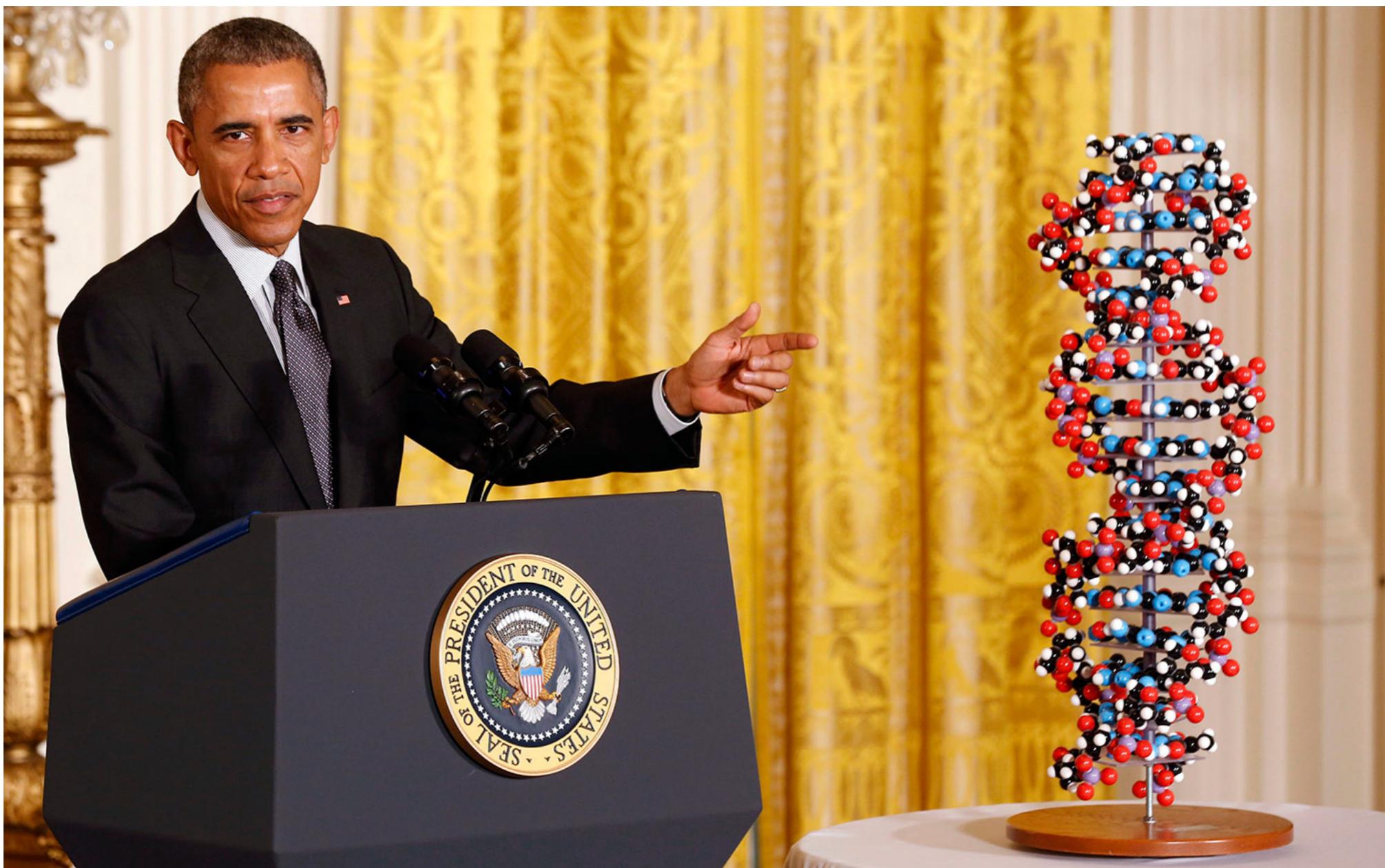
We don't see all the variants.

We don't know which ones affect phenotype.

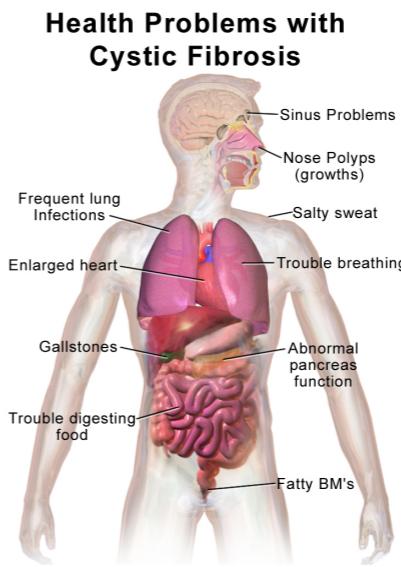
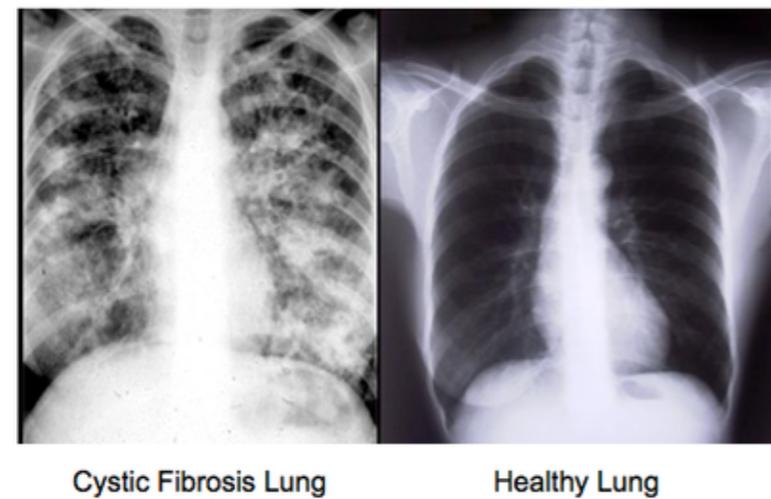
The human genome is big.

Phenotypes are highly variable.

What is precision medicine?

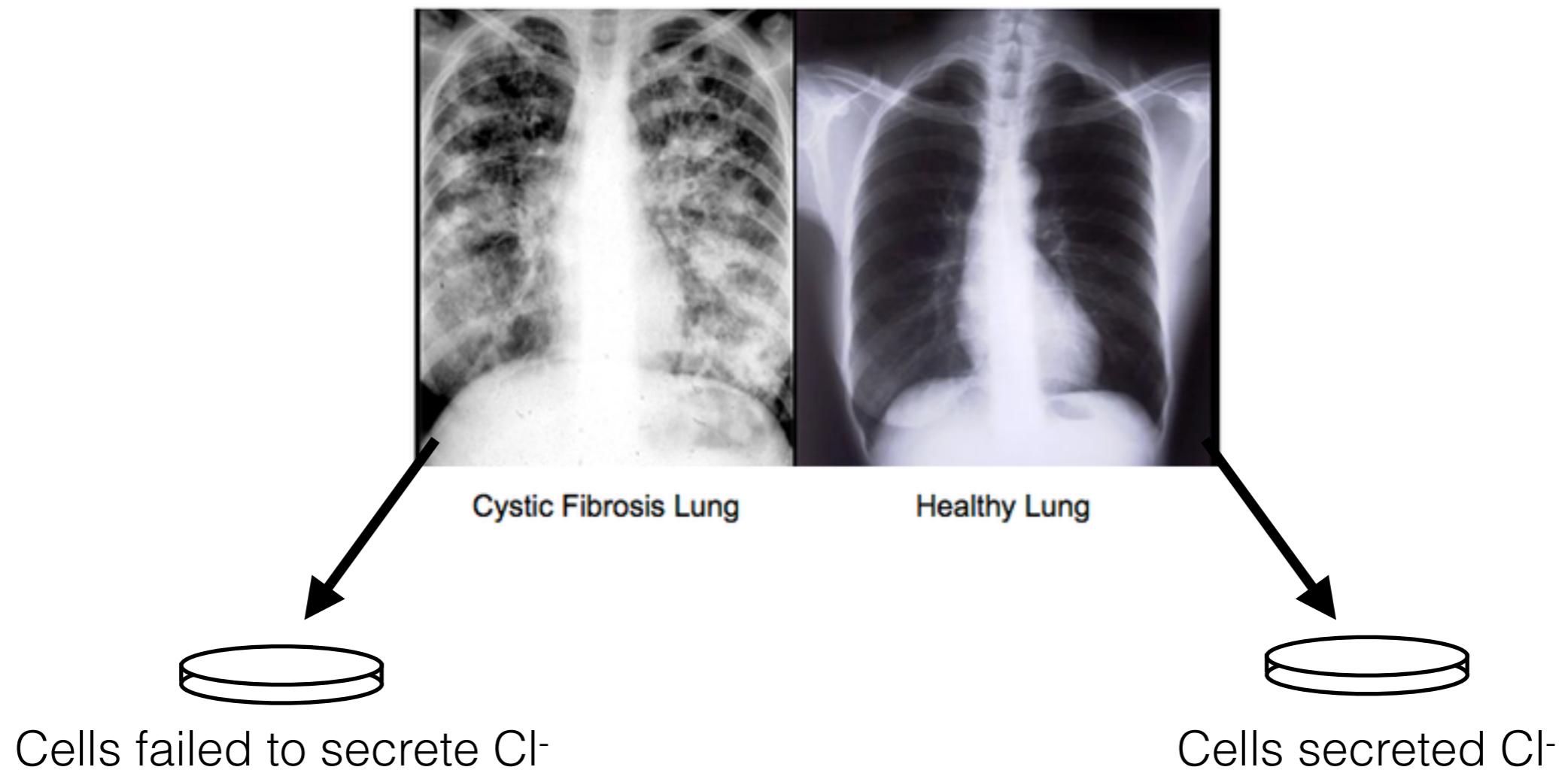


What about cystic fibrosis?

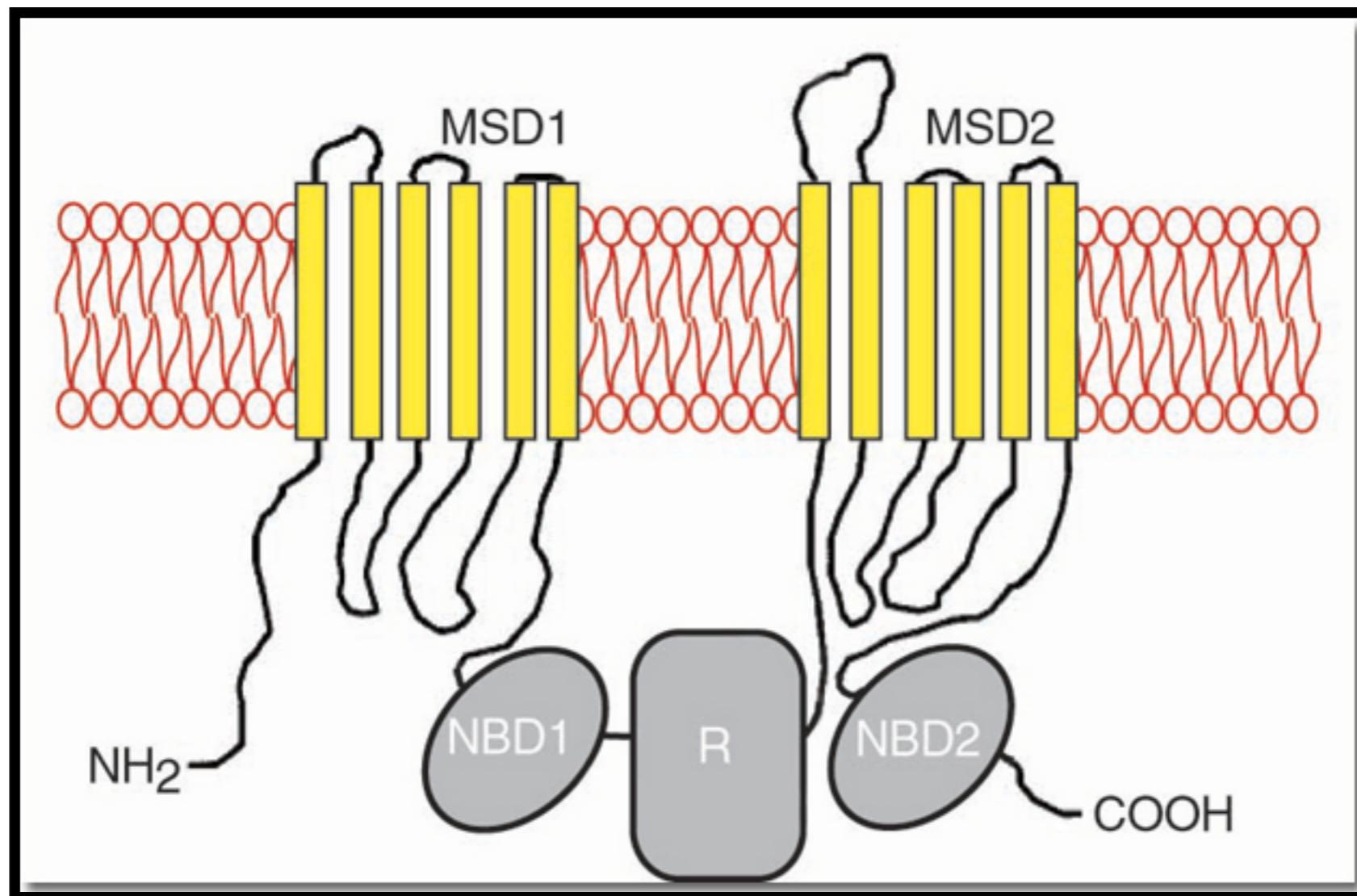


1. Autosomal recessive disorder
2. Not caused by chromosomal aberrations or meiotic NDJ
3. Mapped to chromosome 7
4. Mutations in CF gene are null or hypomorphs
5. Compound heterozygosity (failure to complement) is common
6. No known epistatic genes to CF gene
7. Genetic enhancers are known (immune modulatory genes)
8. No genetic suppressors are known yet.

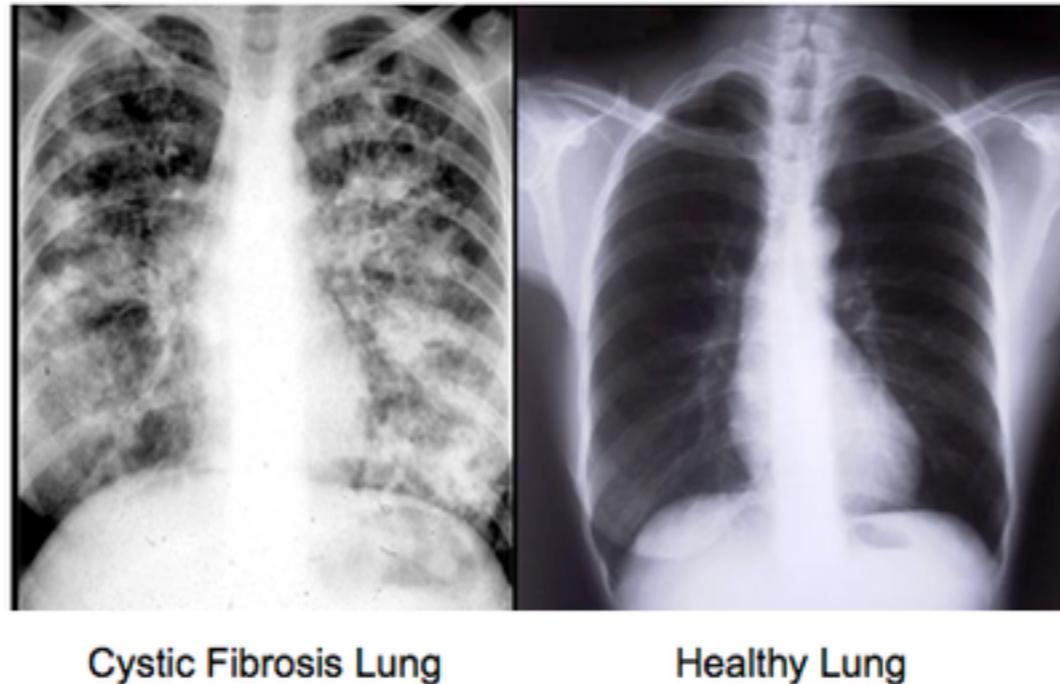
Cell autonomy of CF mutation was shown in the 1960's



Cystic fibrosis was mapped to the chloride ion channel CFTR



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



Rare disease affects 1/10,000 live births

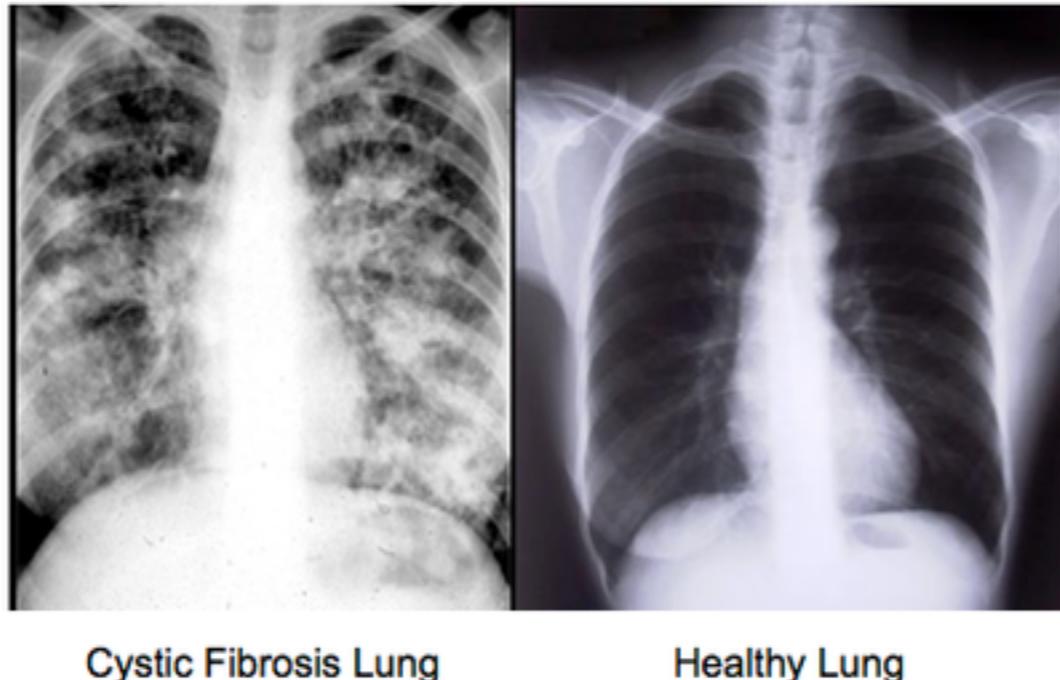
Caused by mutations in the CFTR gene

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



50% of all cases have the same allele $\Delta F508$

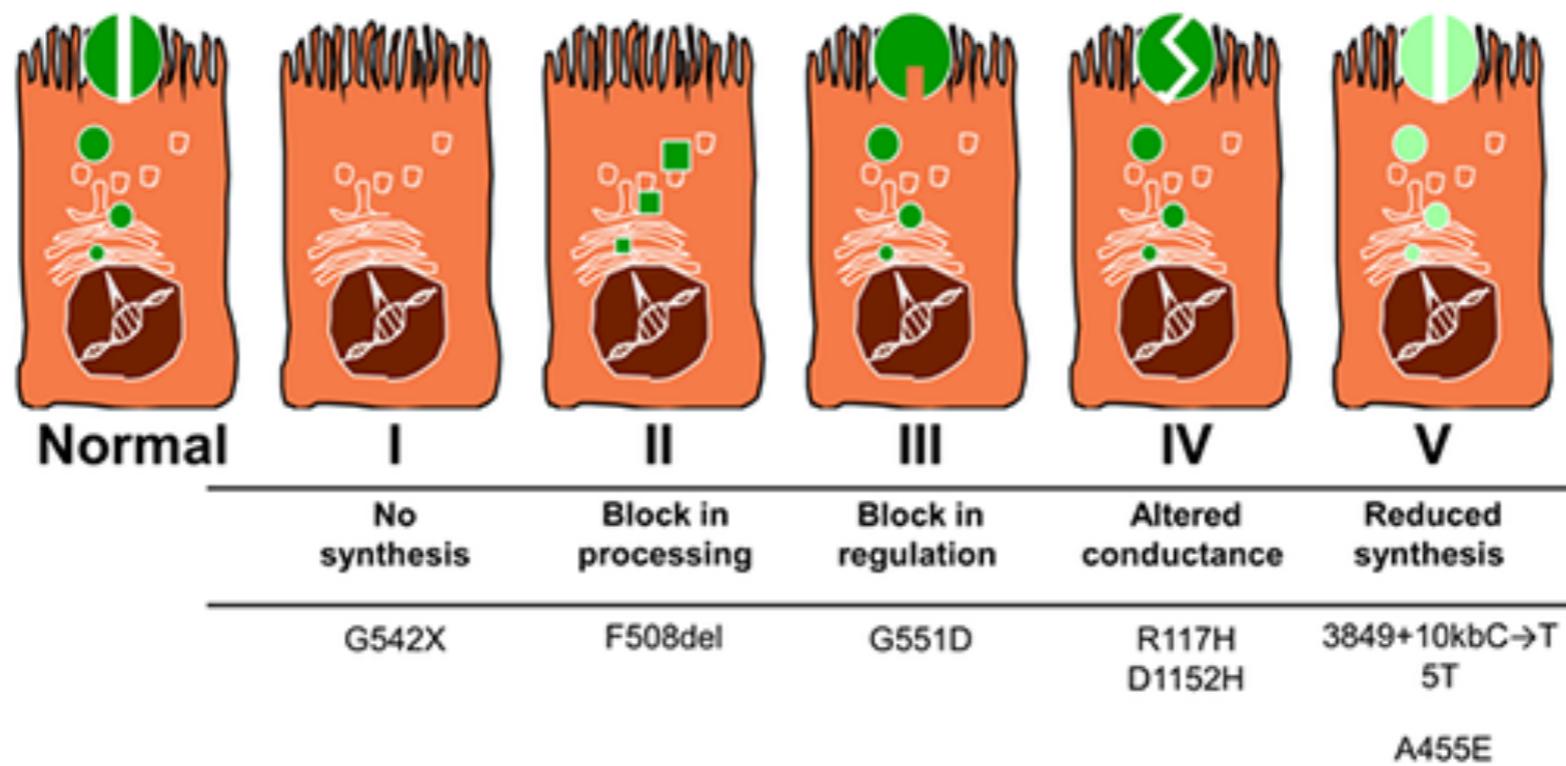
Over 1000 other mutations are known

Compound heterozygotes found often

Genetic heterogeneity

CFTR

Classes of Mutations



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