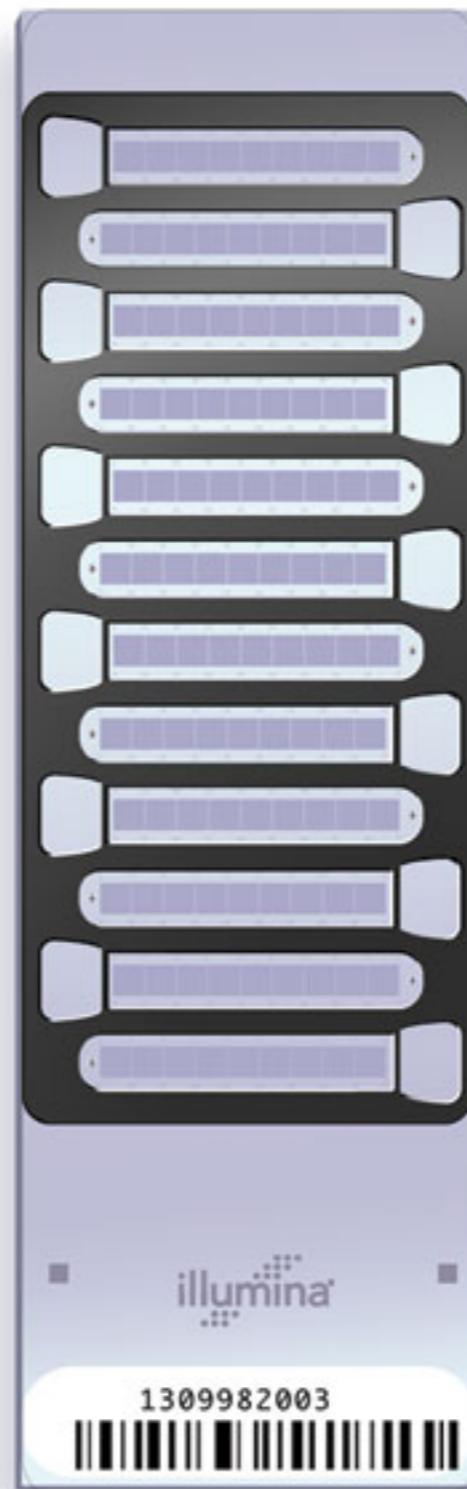


# Human variation and allele frequency spectrum

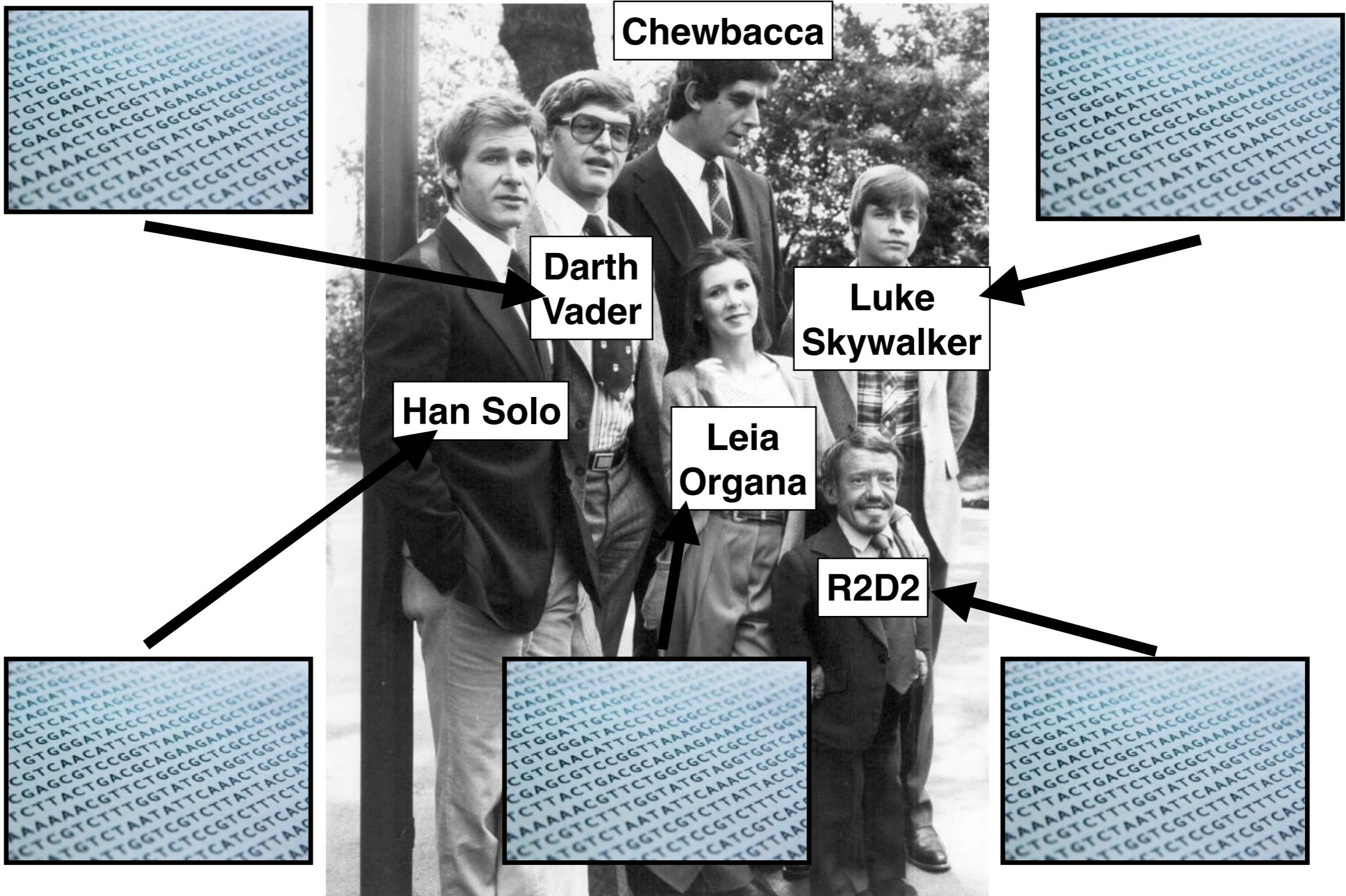


# An array to genotype at >4.3 million sites



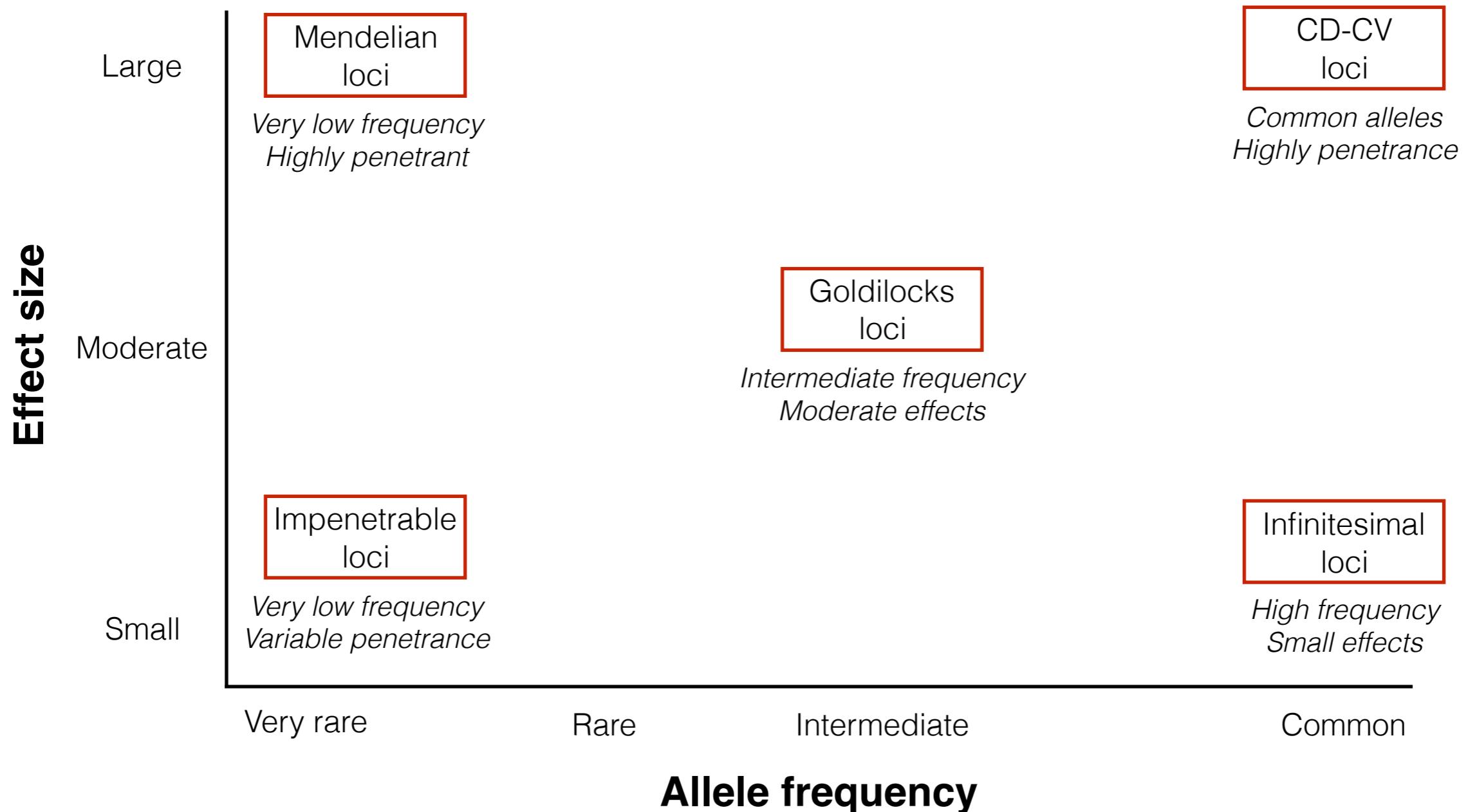
Tool to genotype intermediate and common variation

# We want to be able to read genomes and make predictions



The cast of the original *Star Wars*

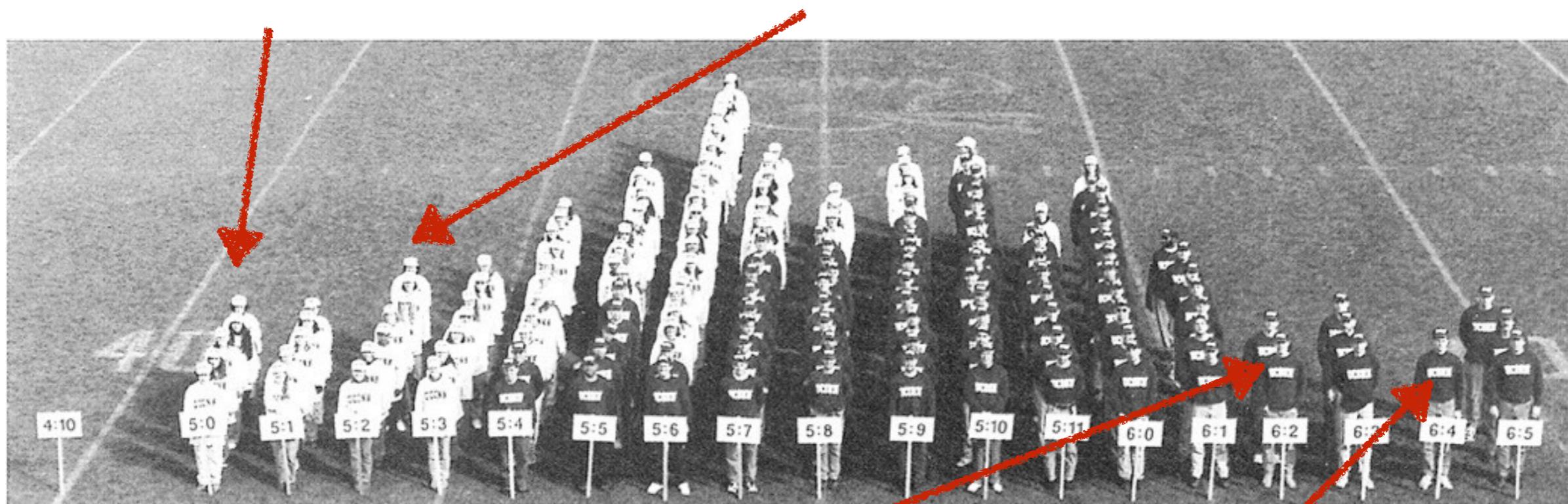
# The spectrum of how variation contributes to disease



How do we find the variants that cause common disease?

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

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

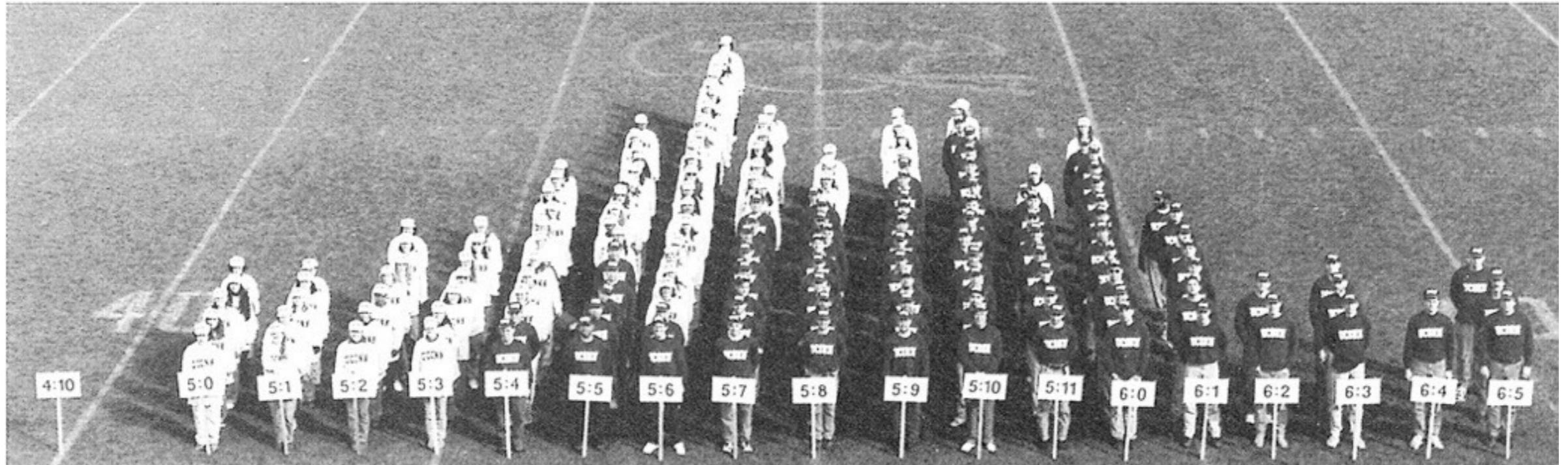


University of Connecticut, 1997

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

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

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

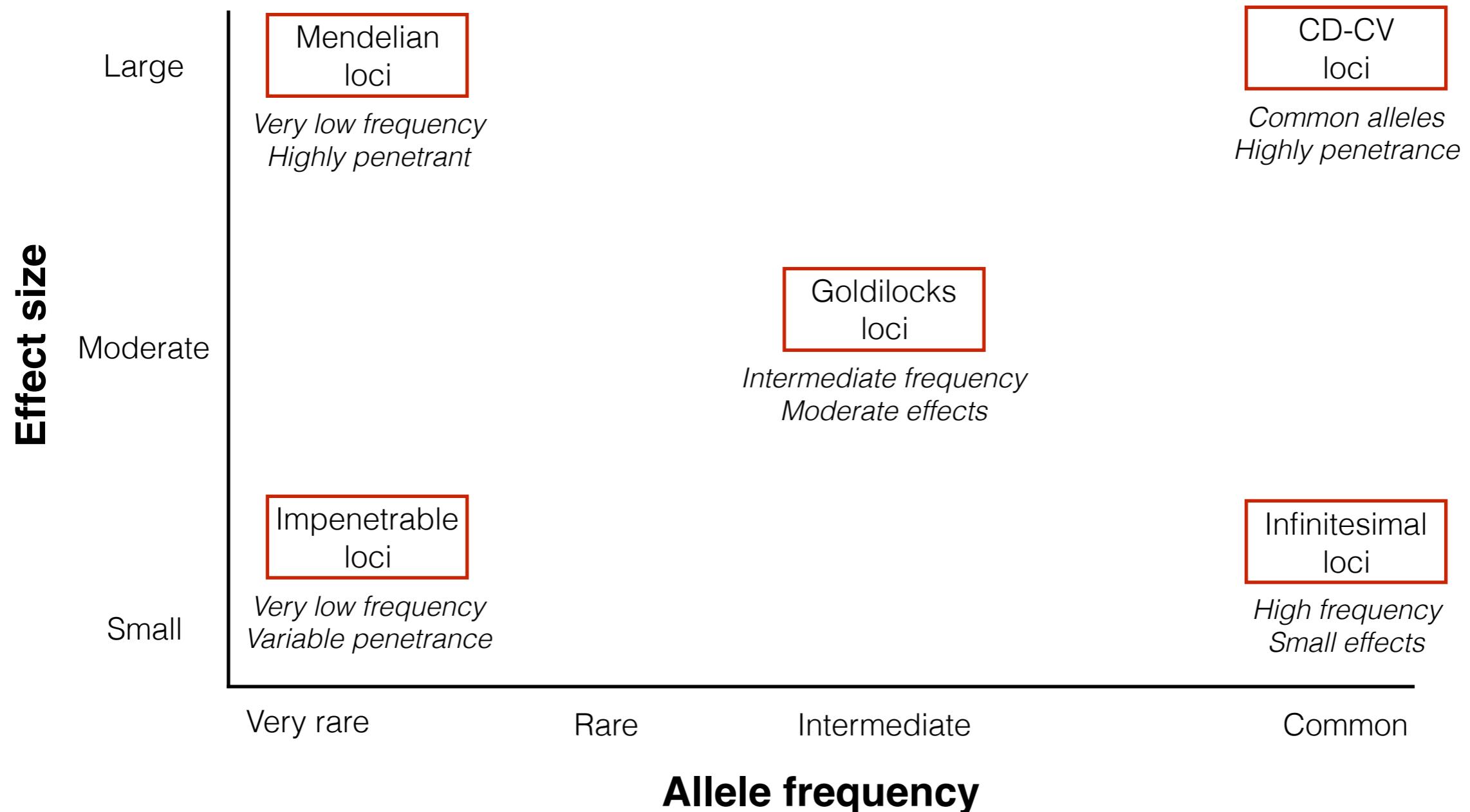


University of Connecticut, 1997

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

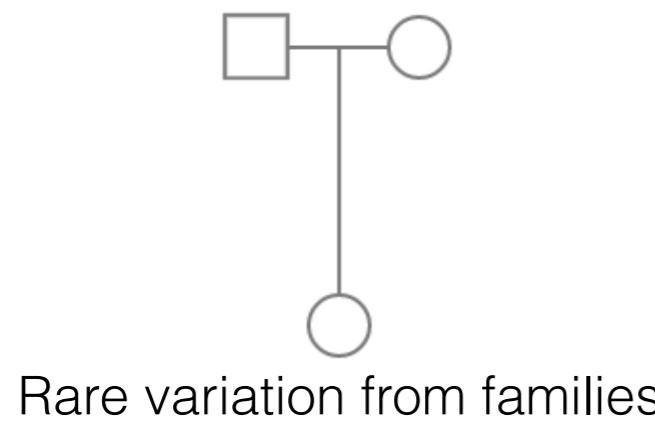
~700,000 people genotyped led to 60%  
of height differences explained

# The spectrum of how variation contributes to disease

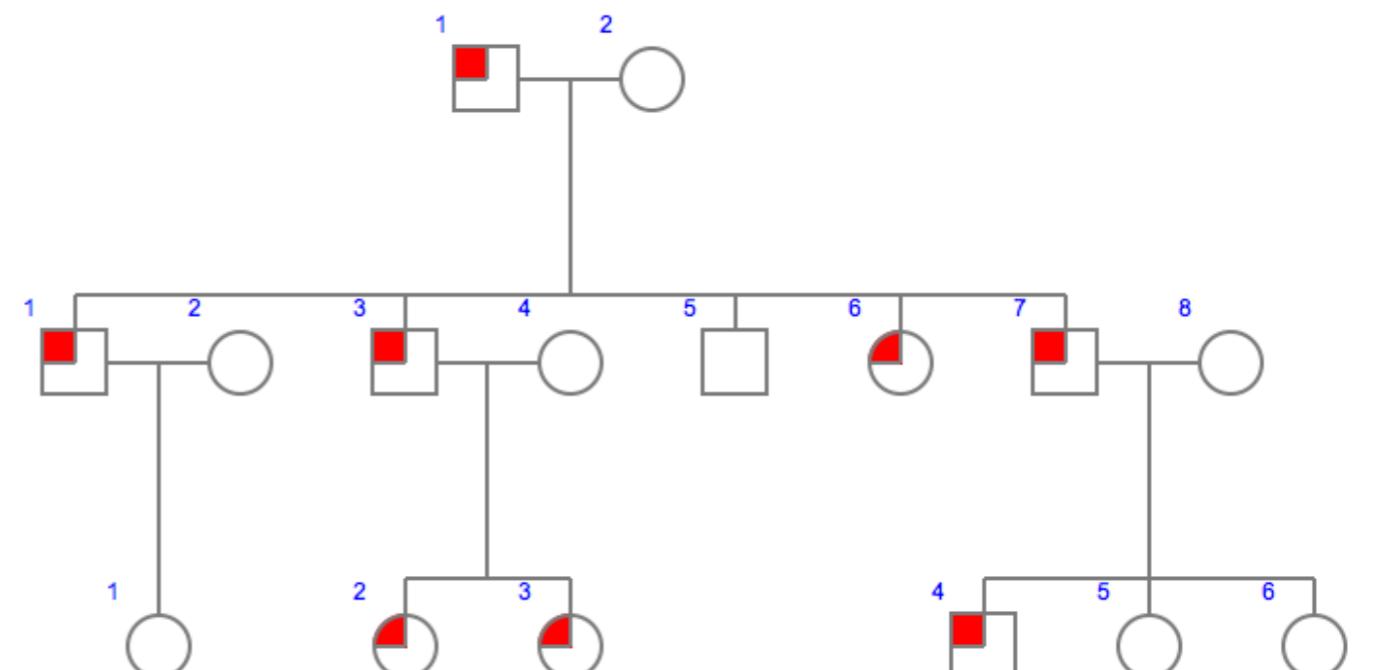


How do we find the variants that cause rare disease?

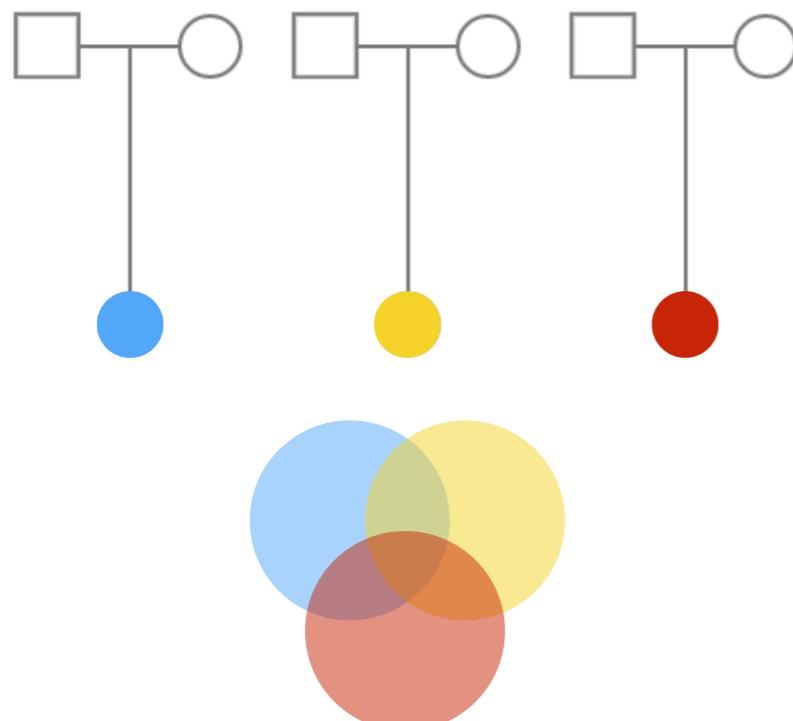
# Strategies to identify disease-causing rare variants



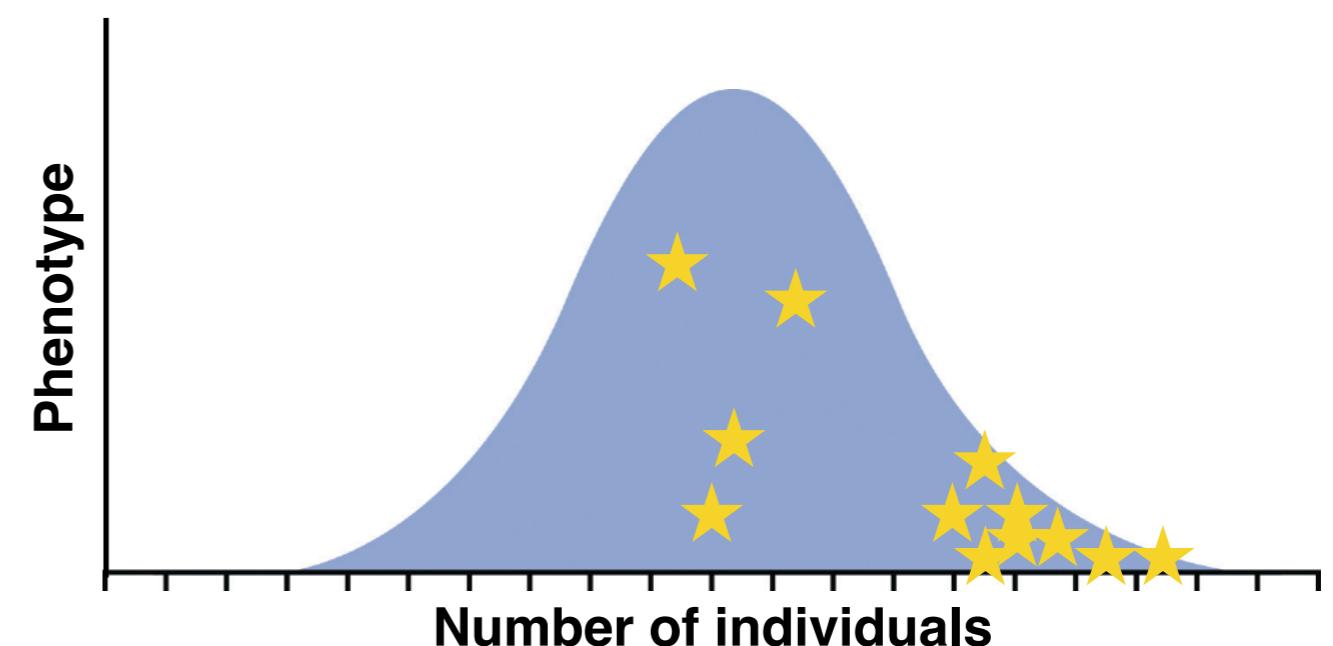
Rare variation from families



Shared variants from affected individuals in large families

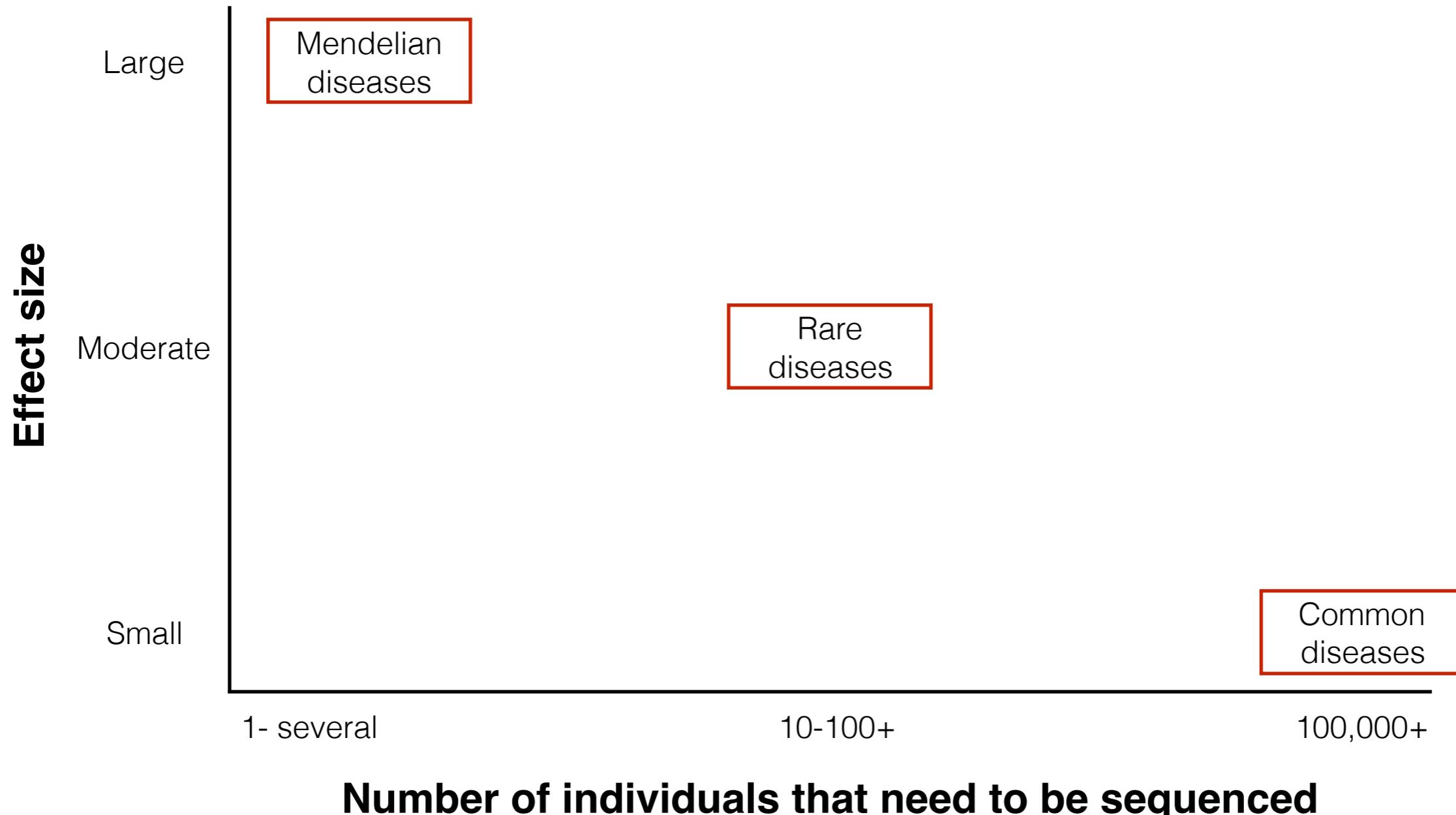


Shared variation from trios



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.

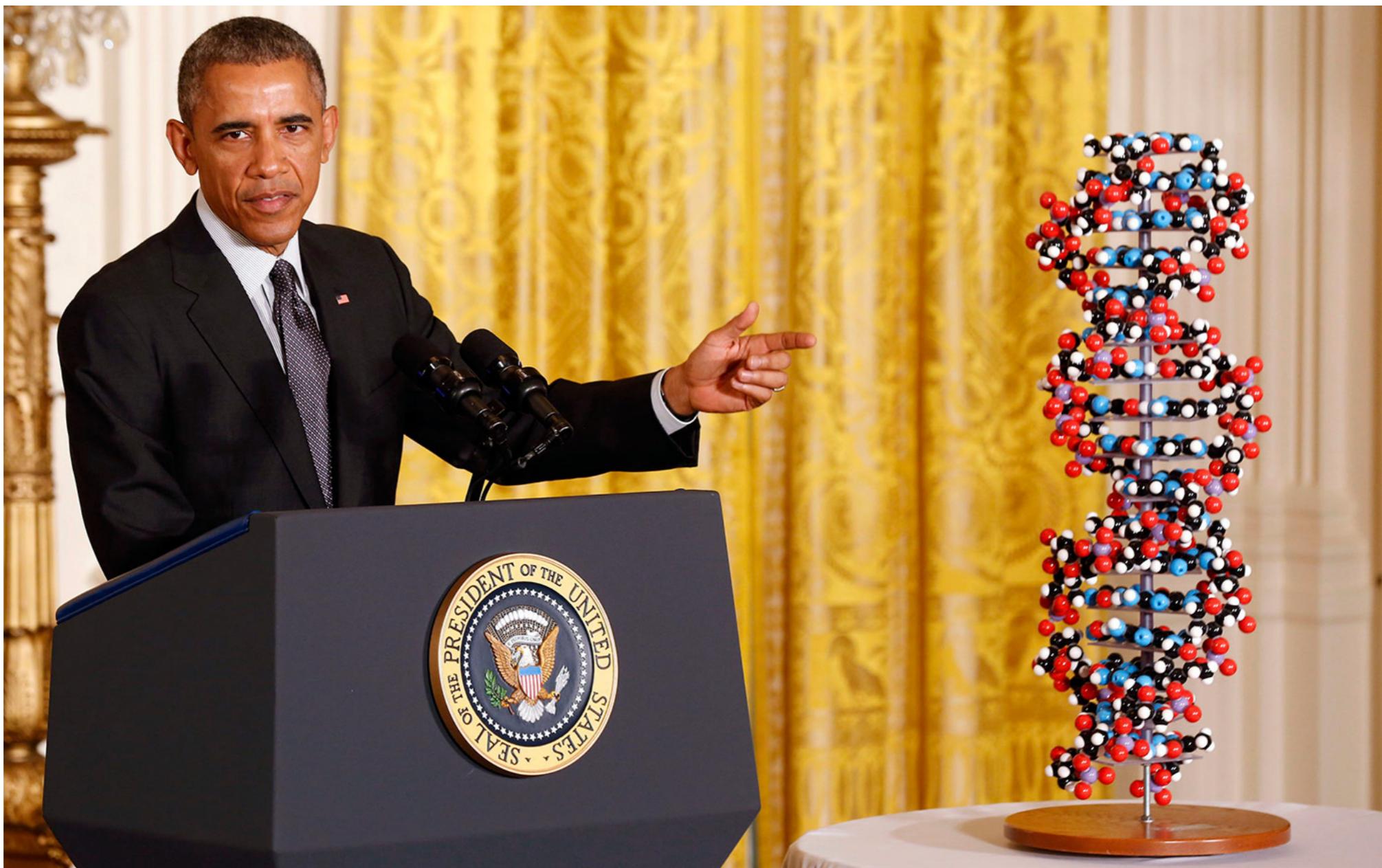
We don't know which ones affect phenotype.

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

The human genome is big.

Phenotypes are highly variable.

# What is precision medicine?

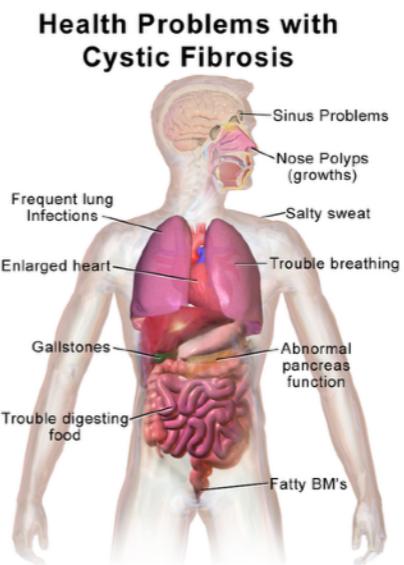
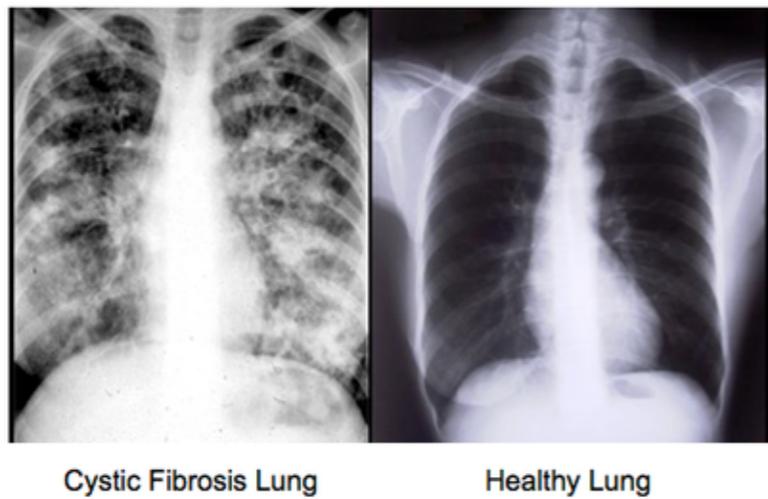


# We are living in the human genetics renaissance



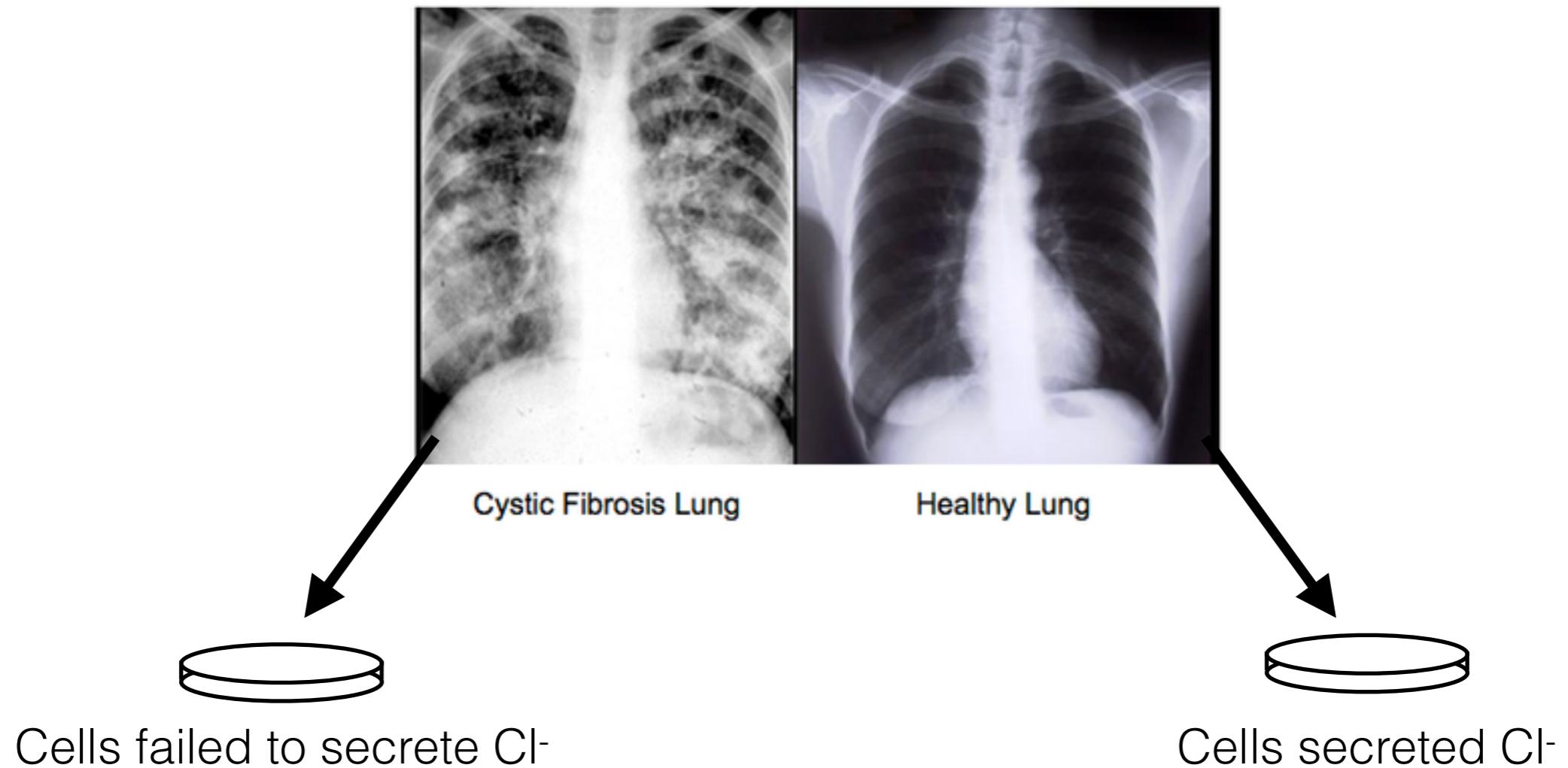
Under \$1000 genome  
Rare disease sequencing for Mendelian disorders  
Family genetics  
Fetal testing from sequence  
Disease outbreaks and diagnosis  
Drug response prediction  
Cancer genome sequencing

# What about cystic fibrosis?

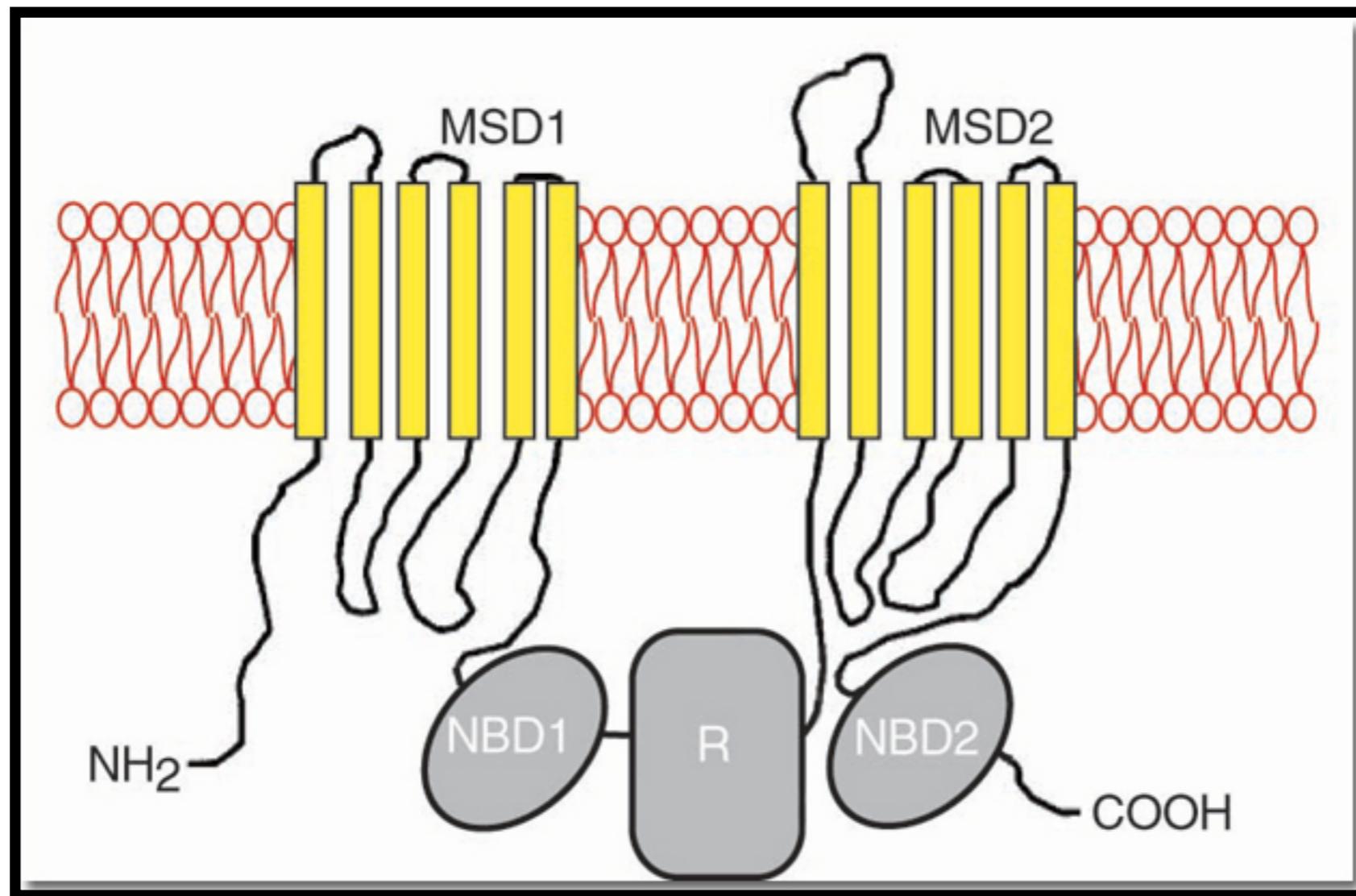


1. Autosomal recessive disorder
2. Not caused by chromosomal aberrations or meiotic NDJ
3. Mapped to chromosome 7
4. Mutations in CF gene are null or hypomorphs
5. Compound heterozygosity (failure to complement) is common
6. No known epistatic genes to CF gene
7. Genetic enhancers are known (immune modulatory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

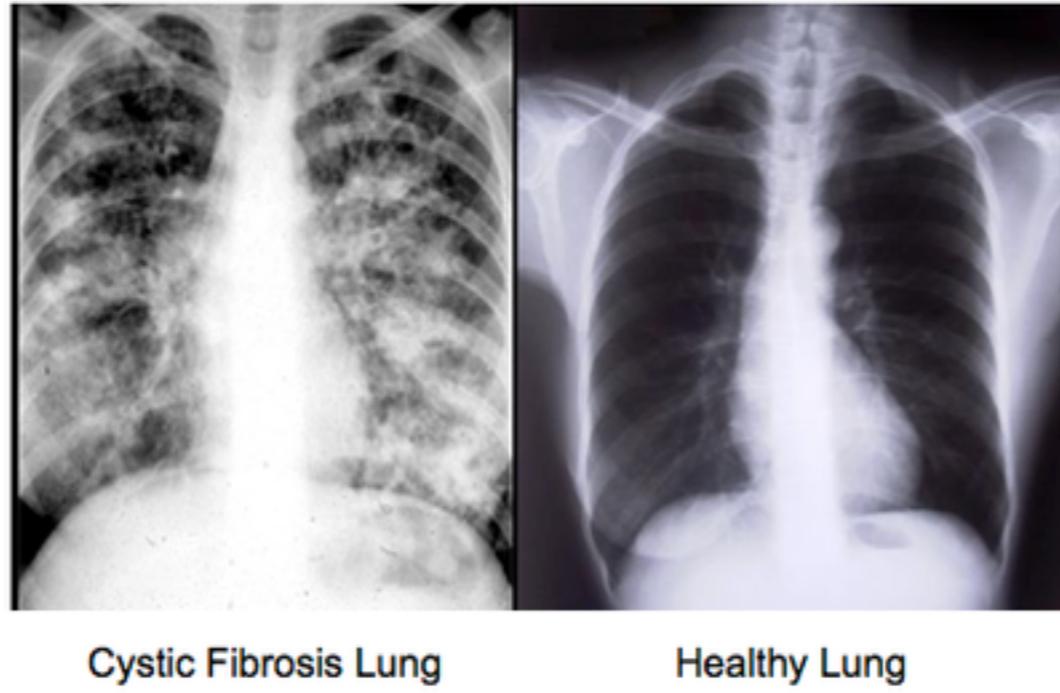
# Cell autonomy of CF mutation was shown in the 1960's



# 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

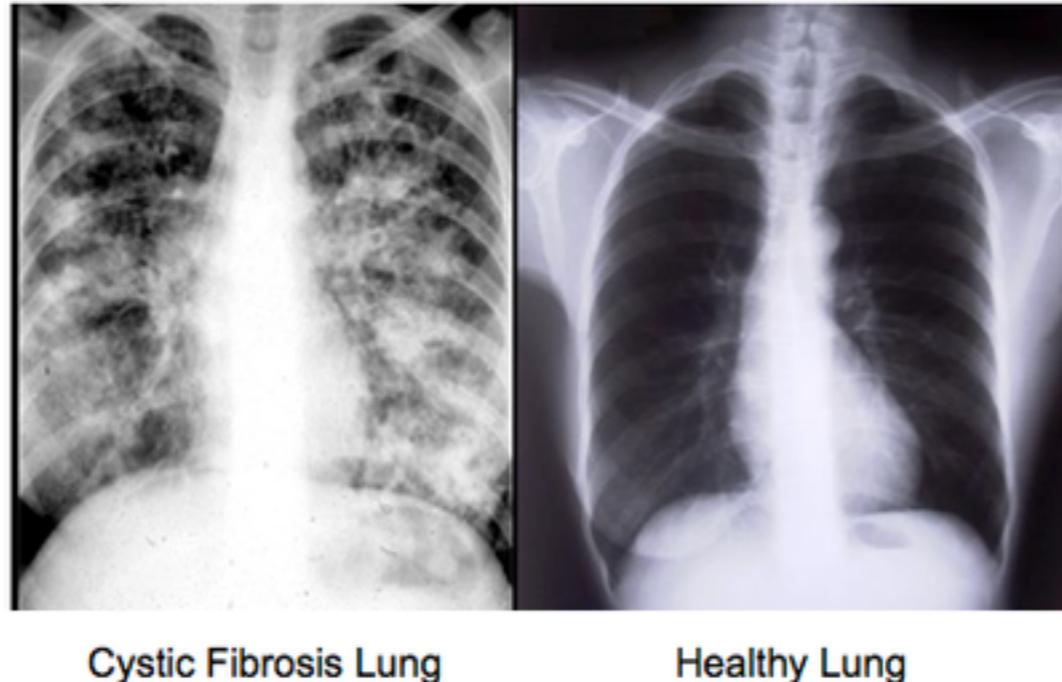
Caused by mutations in the CFTR gene

Selection removes homozygotes from population

Hardy-Weinberg 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  $\Delta F508$

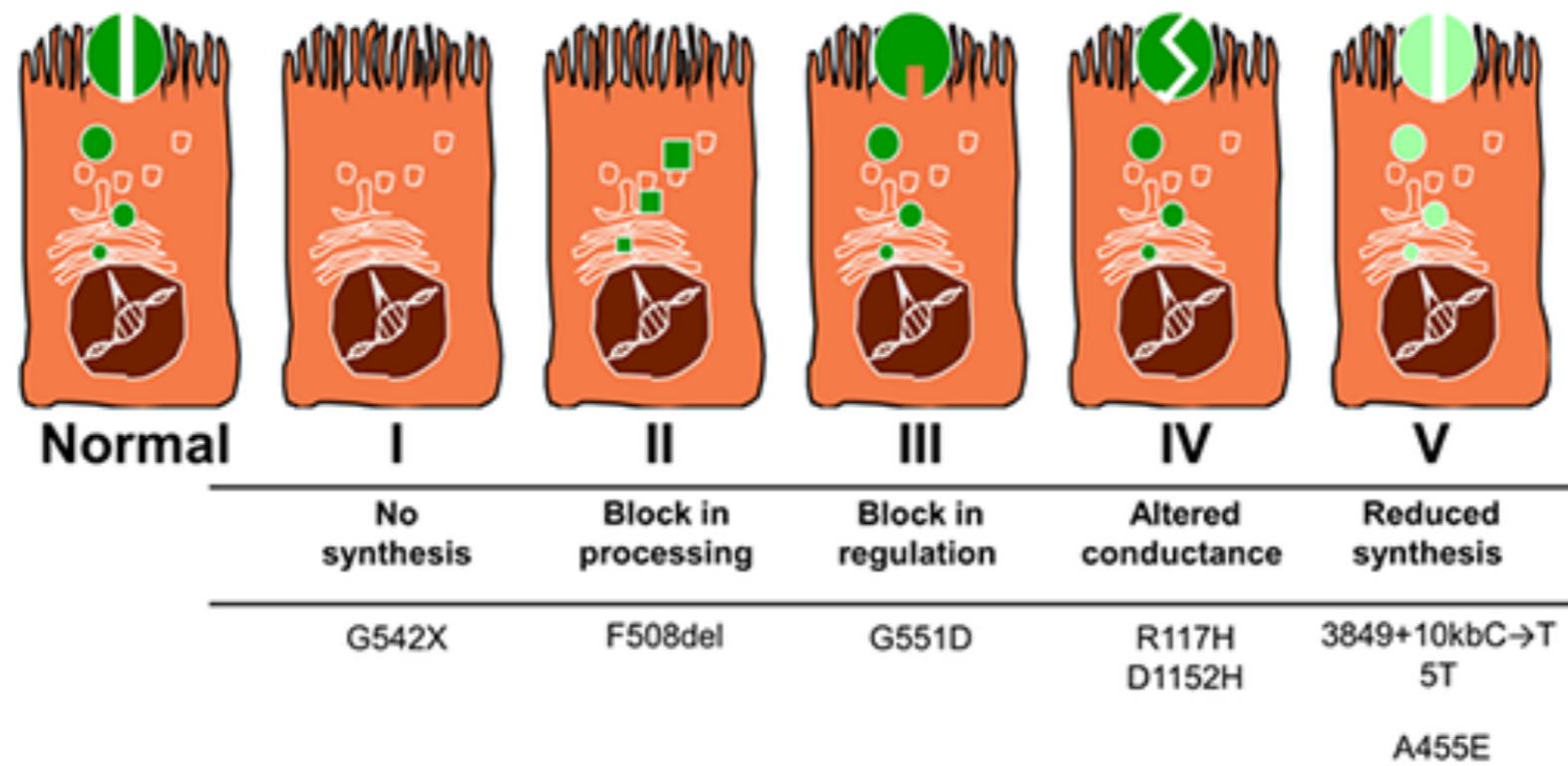
Over 1000 other mutations are known

Compound heterozygotes found often

Genetic heterogeneity

## CFTR

### *Classes of Mutations*



What do you think the phenotypes of these mutations are?