# Bio393: Genetic Analysis

Human variation and allele frequency spectrum

**Lecture 14**

# In human genetics, experiments are all *post hoc*

# In human genetics, experiments are all *post hoc*

No controlled crosses

# In human genetics, experiments are all *post hoc*

No controlled crosses

# In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

# In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

## In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (Six gigabase pairs)

## In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (Six gigabase pairs)

# In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (Six gigabase pairs)

Good phenotyping!

# In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (Six gigabase pairs)

Good phenotyping!

## In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

Large genome (Six gigabase pairs)

Good phenotyping!

Lots of $$$

## In human genetics, experiments are all *post hoc*

No controlled crosses

No defined genetic backgrounds

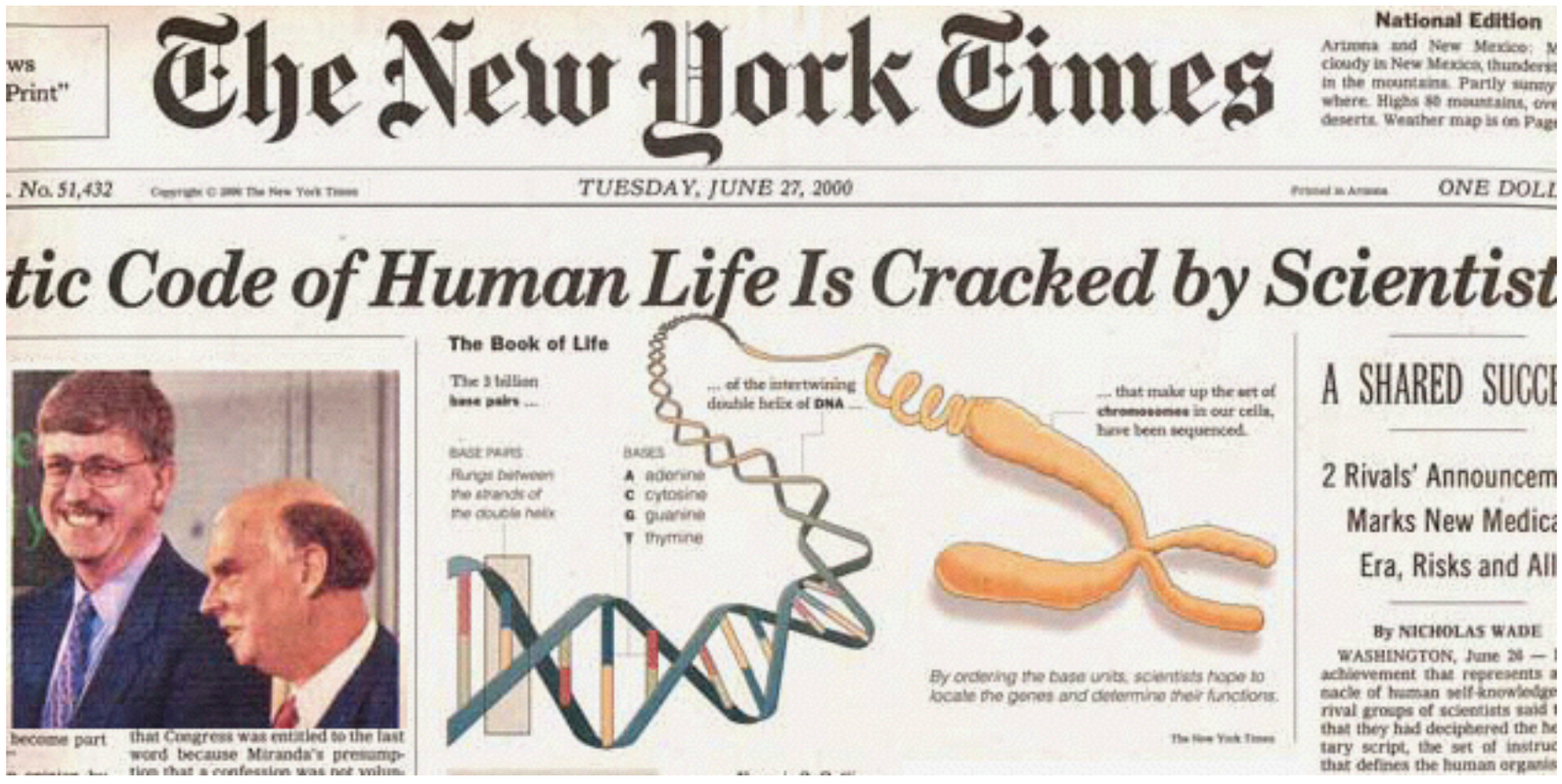
Large genome (Six gigabase pairs)

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 presumption that a confession was not volun-

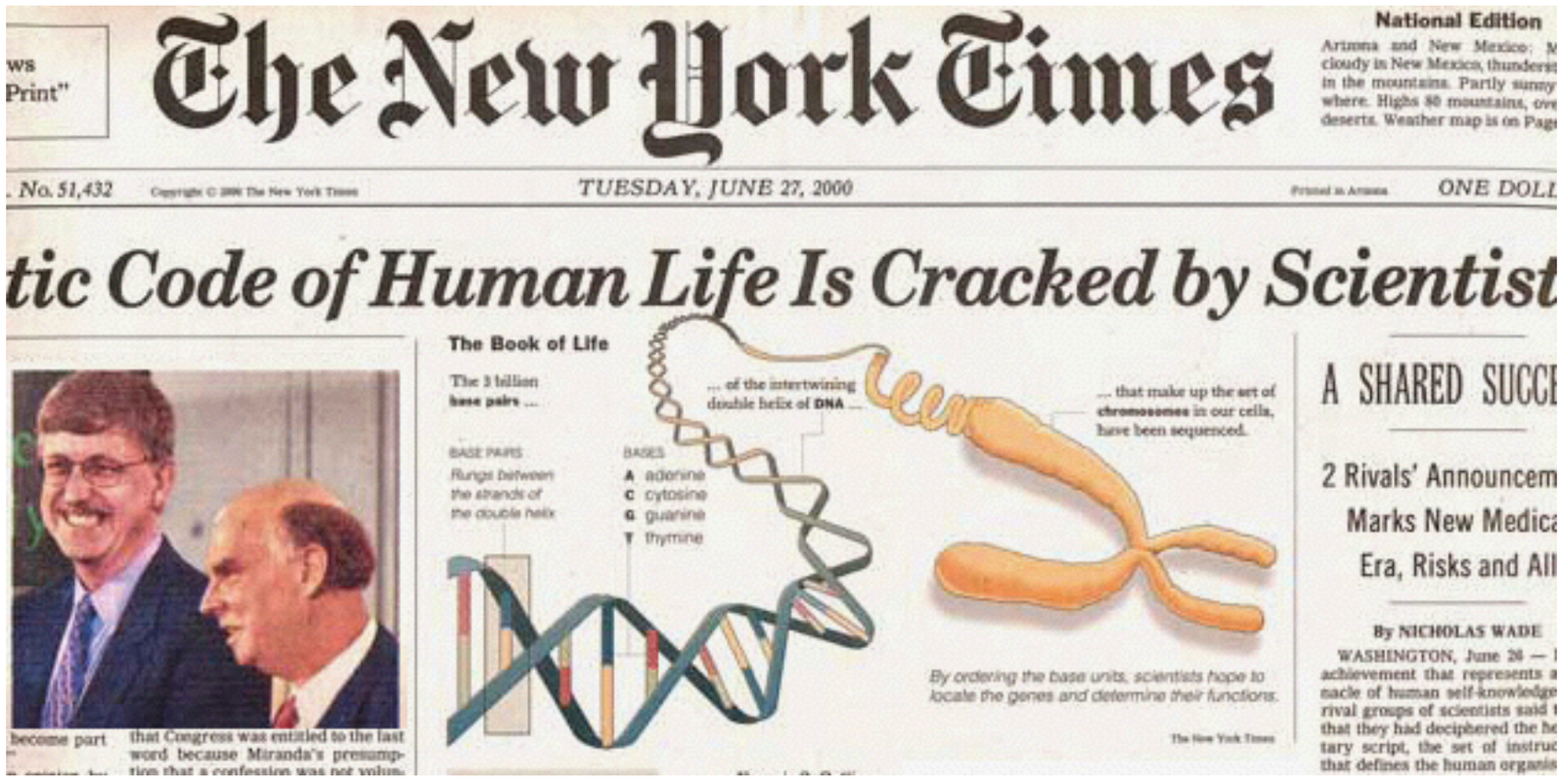
A SHARED SUCCE

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

# Draft human genome announced in June 2000



The New York Times

National Edition

TUESDAY, JUNE 27, 2000

tic Code of Human Life Is Cracked by Scientist

A SHARED SUCCE

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

By NICHOLAS WADE

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.

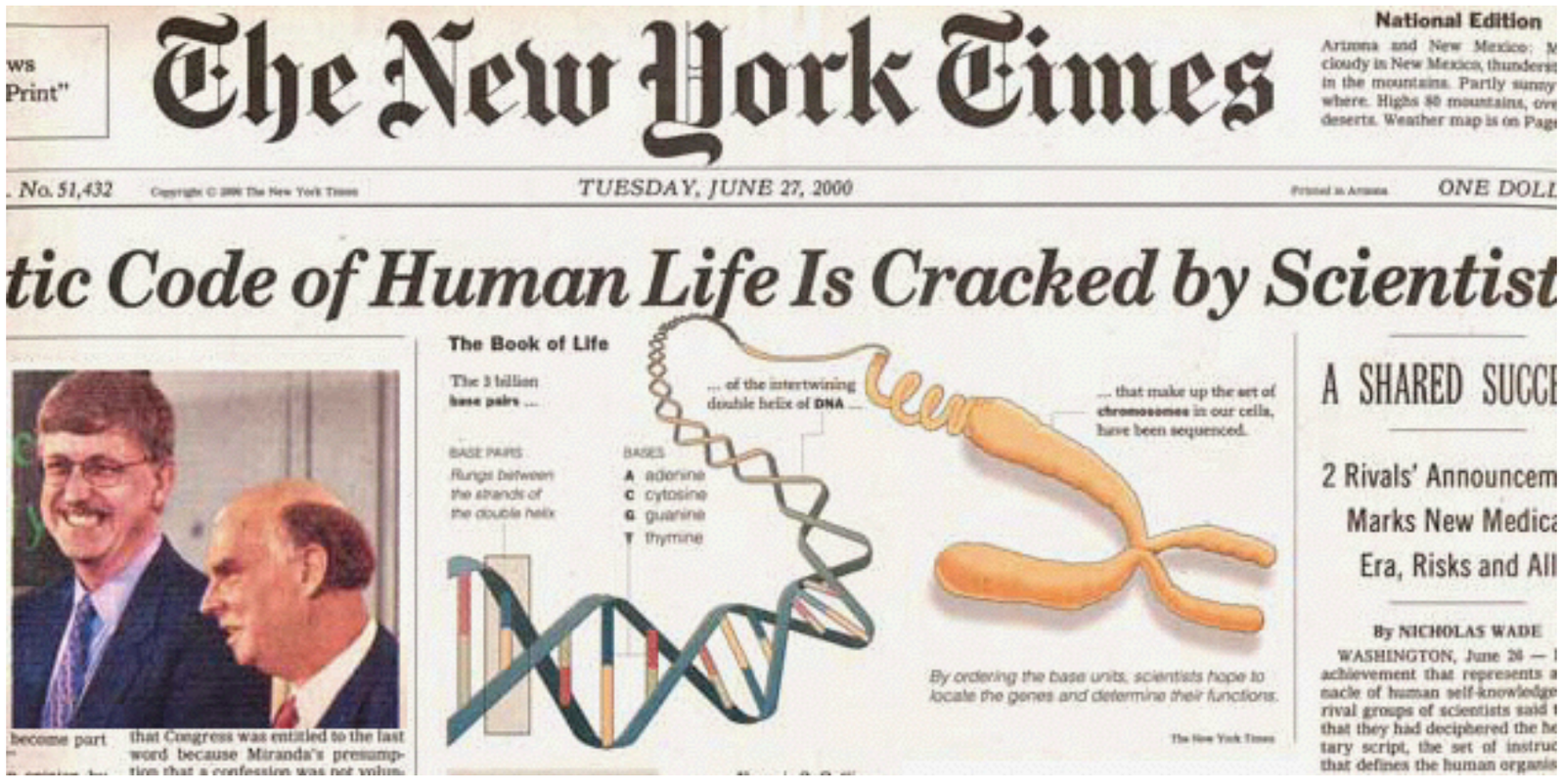
The New York Times

become part

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

It took more than 10 years and $3 billion

# 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”

# We don’t have one “human genome”

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

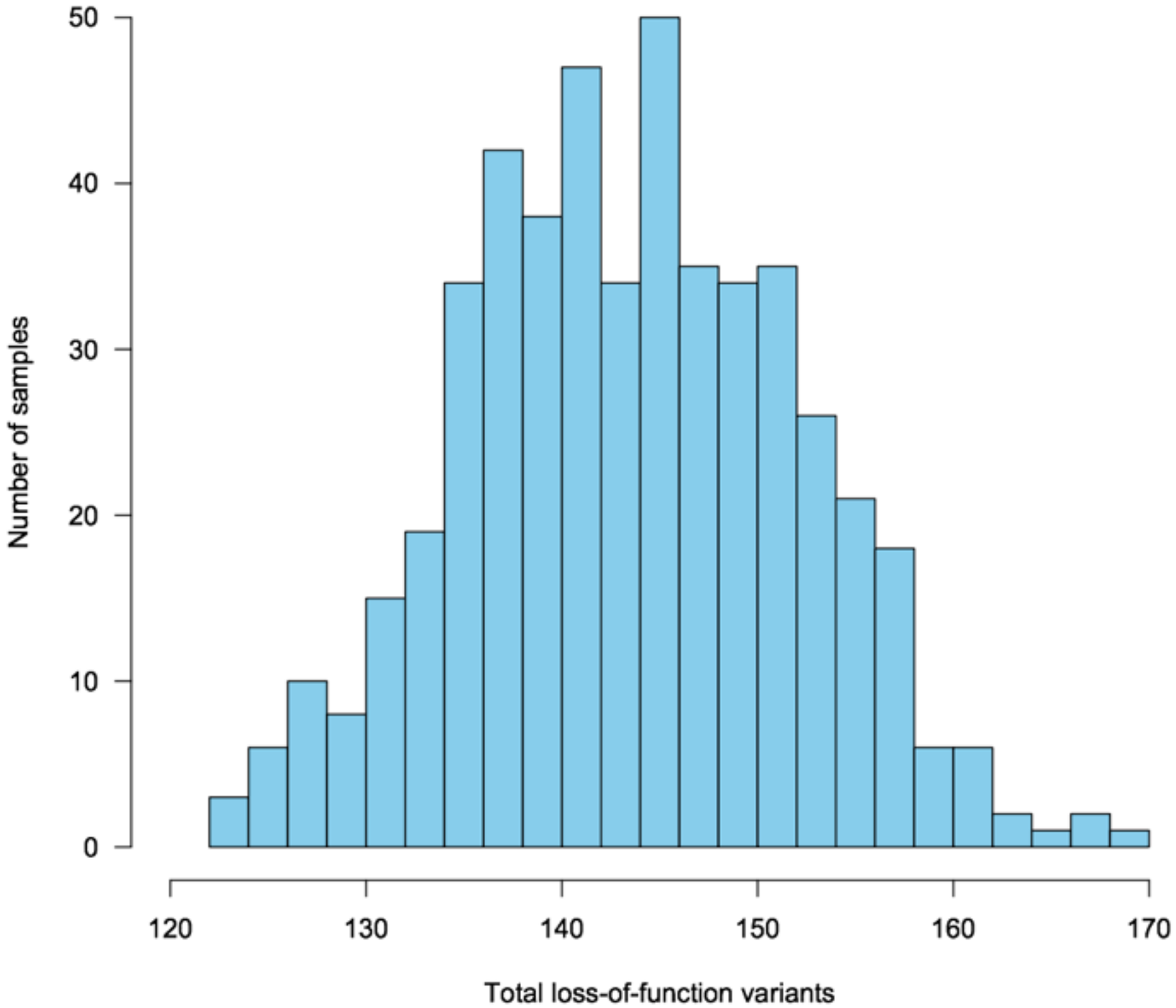
# Types of variation

Lecture 14

# Types of variation

Rare = variants found in less than 1% in population

# We all have over 100 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

Lecture 14

# Types of variation

Rare = variants found in less than 1% in population

# Types of variation

Rare = variants found in less than 1% in population

# Types of variation

Rare = variants found in less than 1% in population

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

# Types of variation

Rare = variants found in less than 1% in population

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

# 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?

# Where does variation come from?

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

# Where does variation come from?

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

# Where does variation come from?

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

Somatic or germline errors

# Where does variation come from?

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

Somatic or germline errors

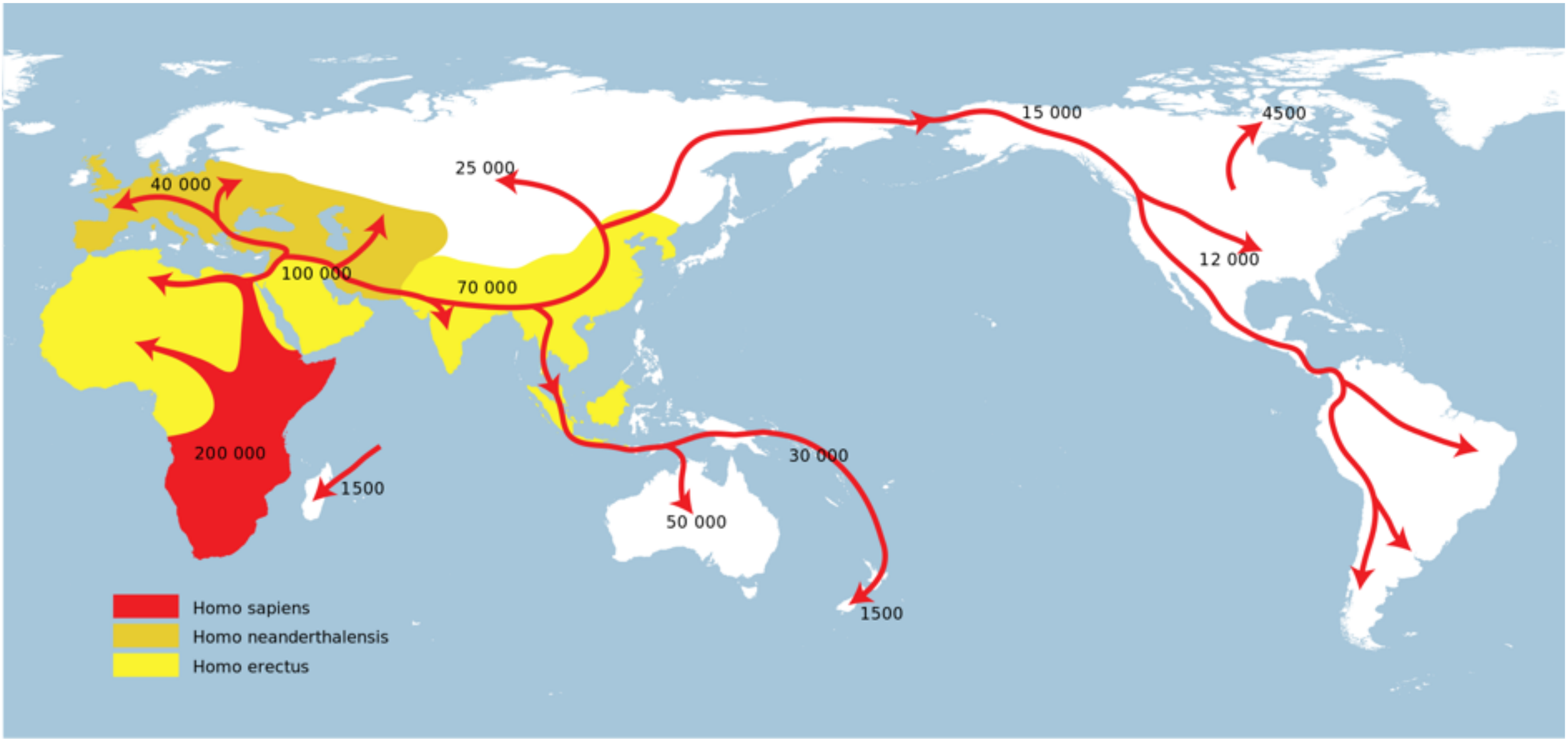
# Where does variation come from?

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

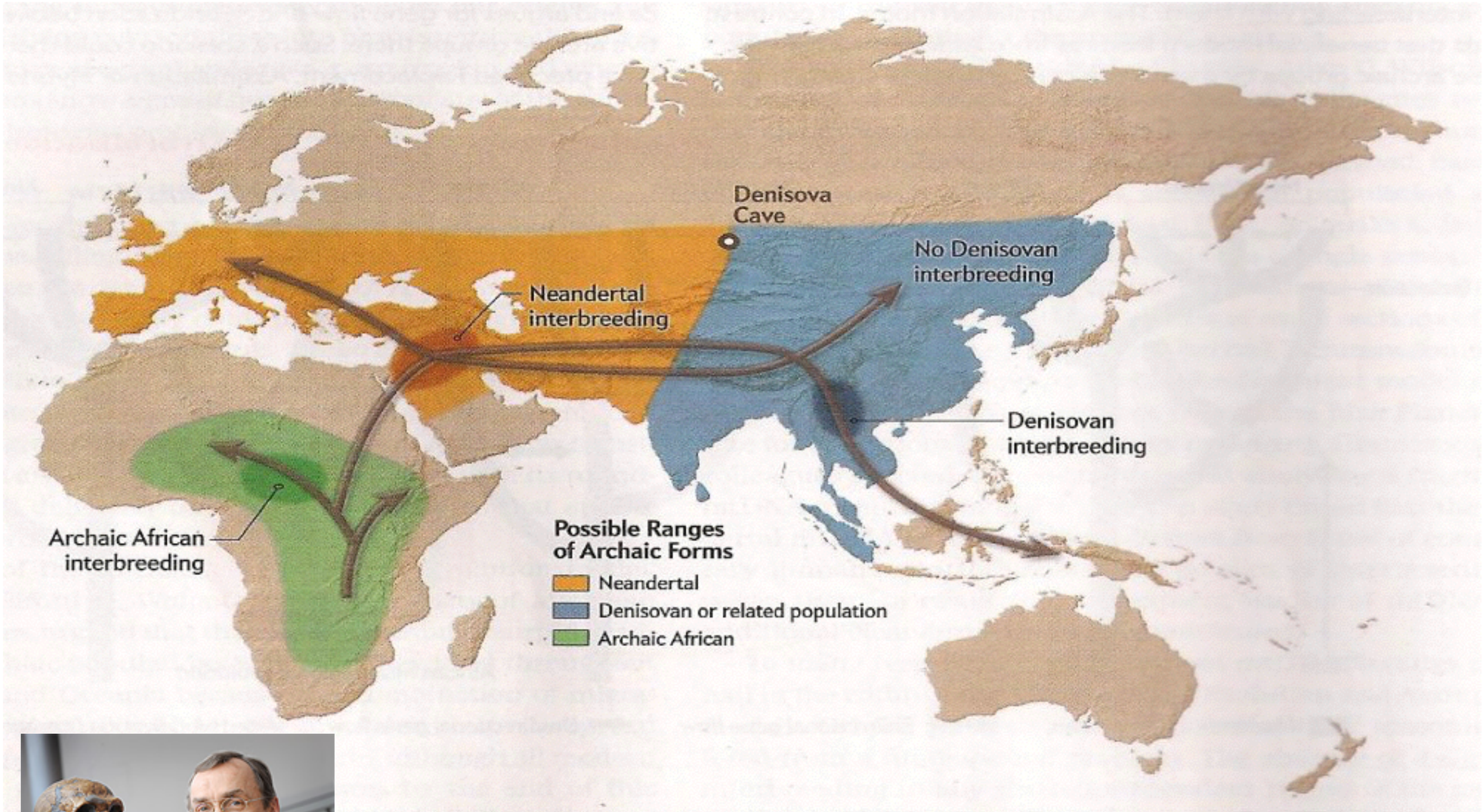
Somatic or germline errors

Once generated, germline variants are inherited

# Human history drives our genetics



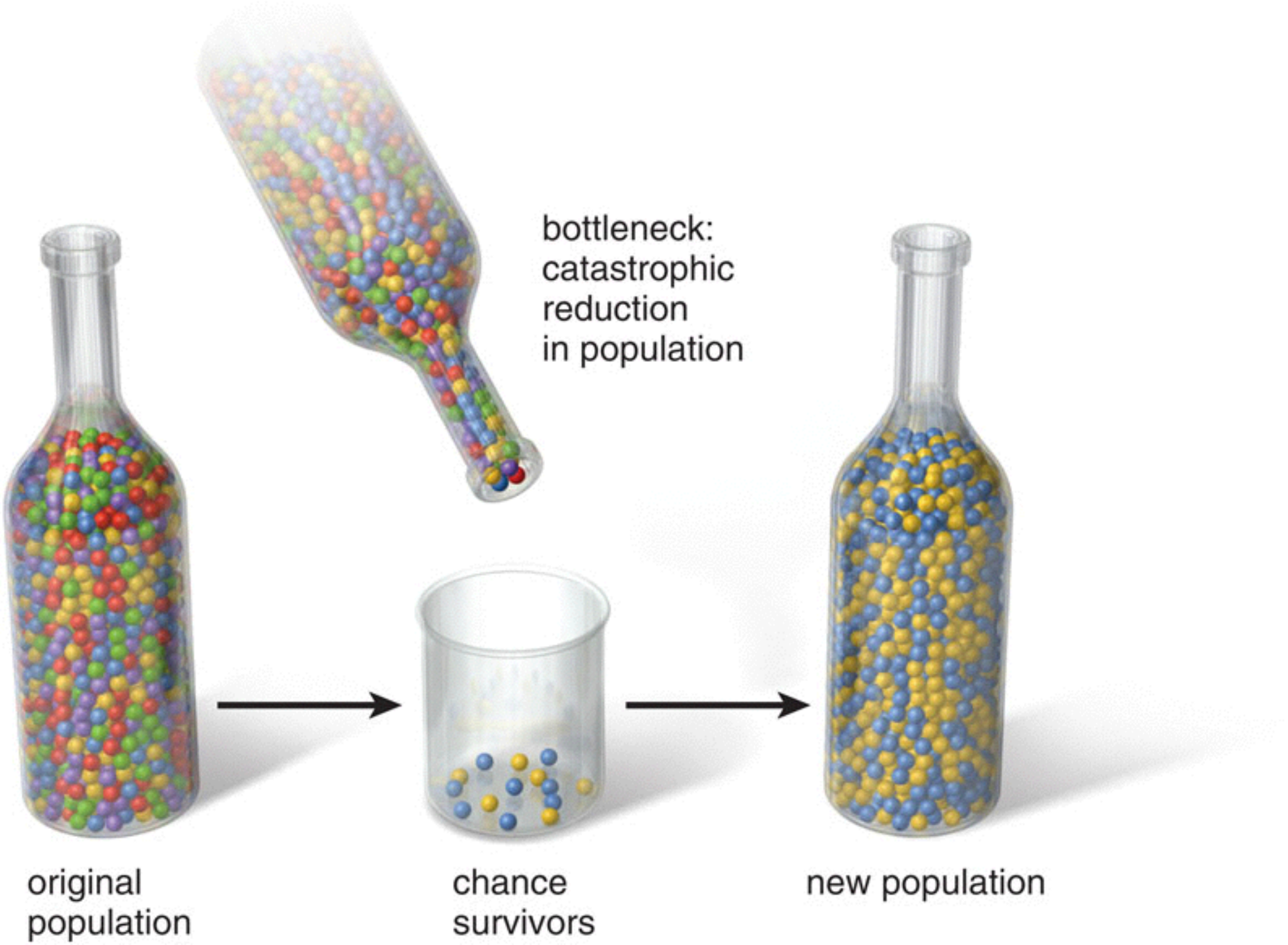
# Human history drives our genetics




**Svante Pääbo**

# Human history drives our genetics





# The common disease - common variant hypothesis

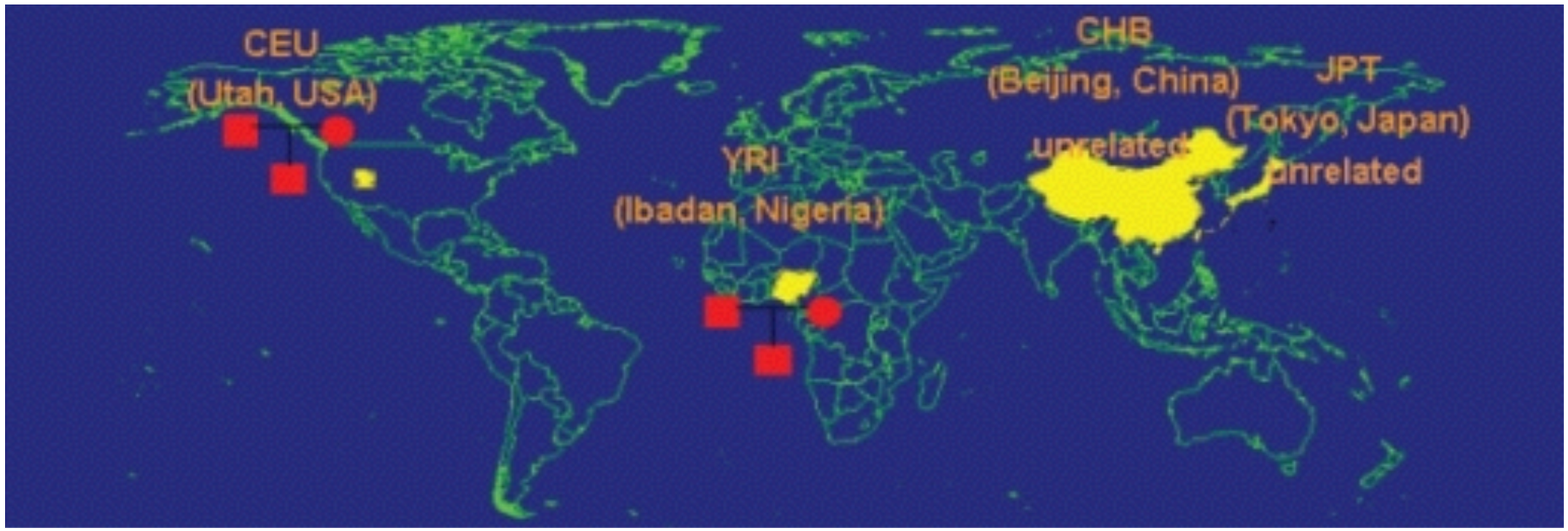
Diseases shared by lots of people
will be caused by variants shared by those same people

# 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

# All three types of variation can cause disease

Rare = variants found in less than 1% in population

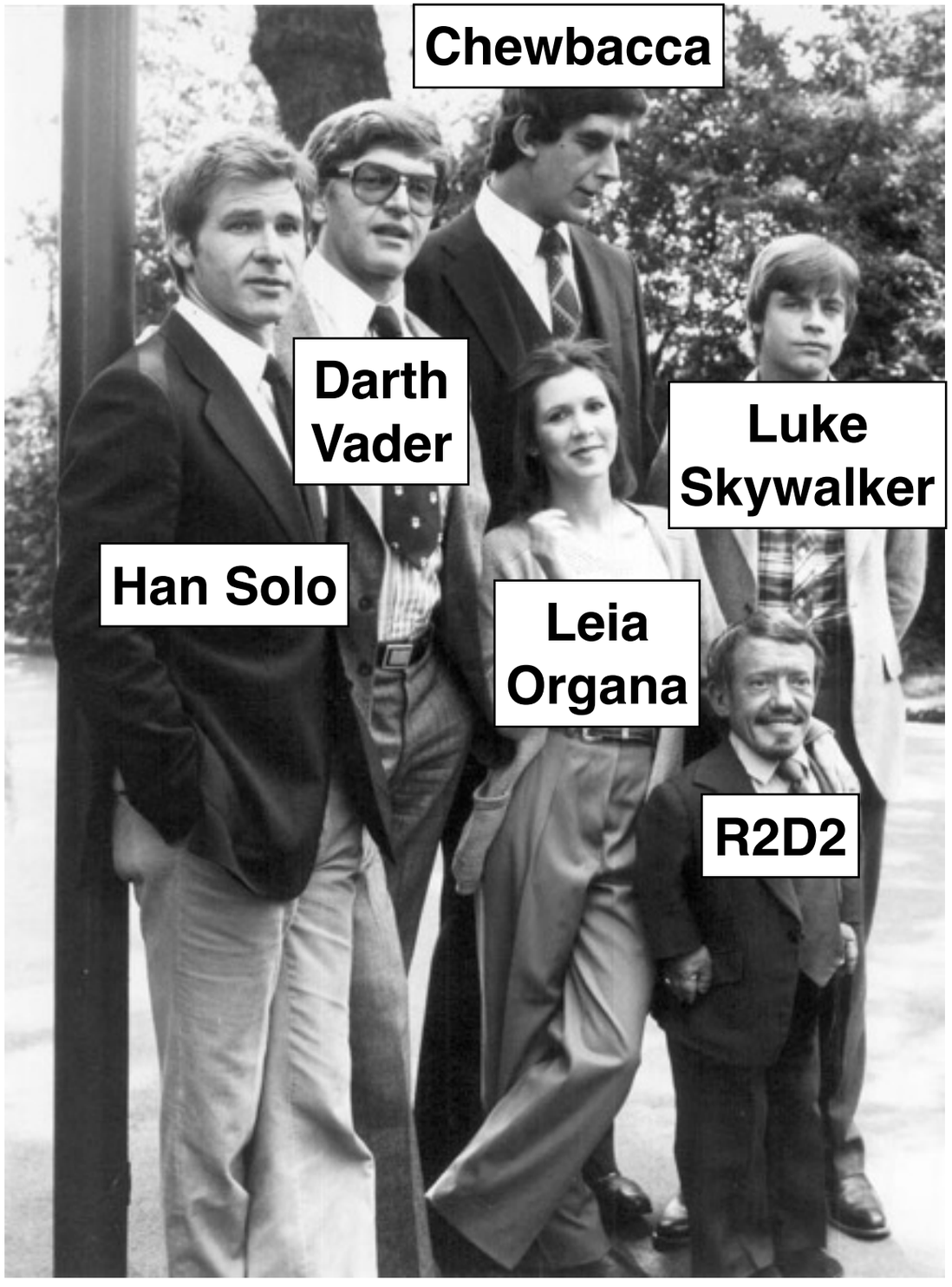
Common = variants found in more than 5% of the population

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

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

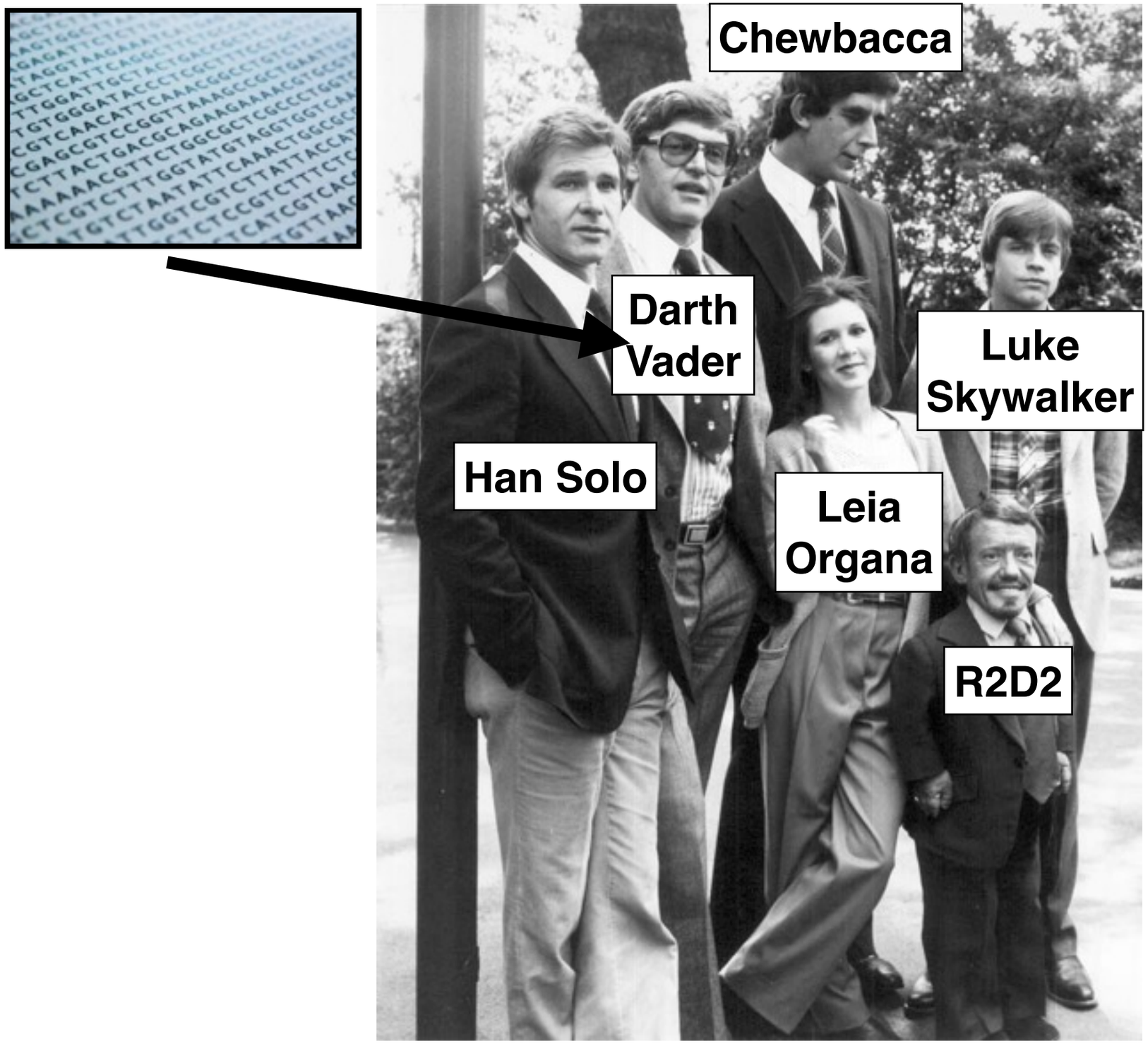
**The cast of the original *Star Wars***

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



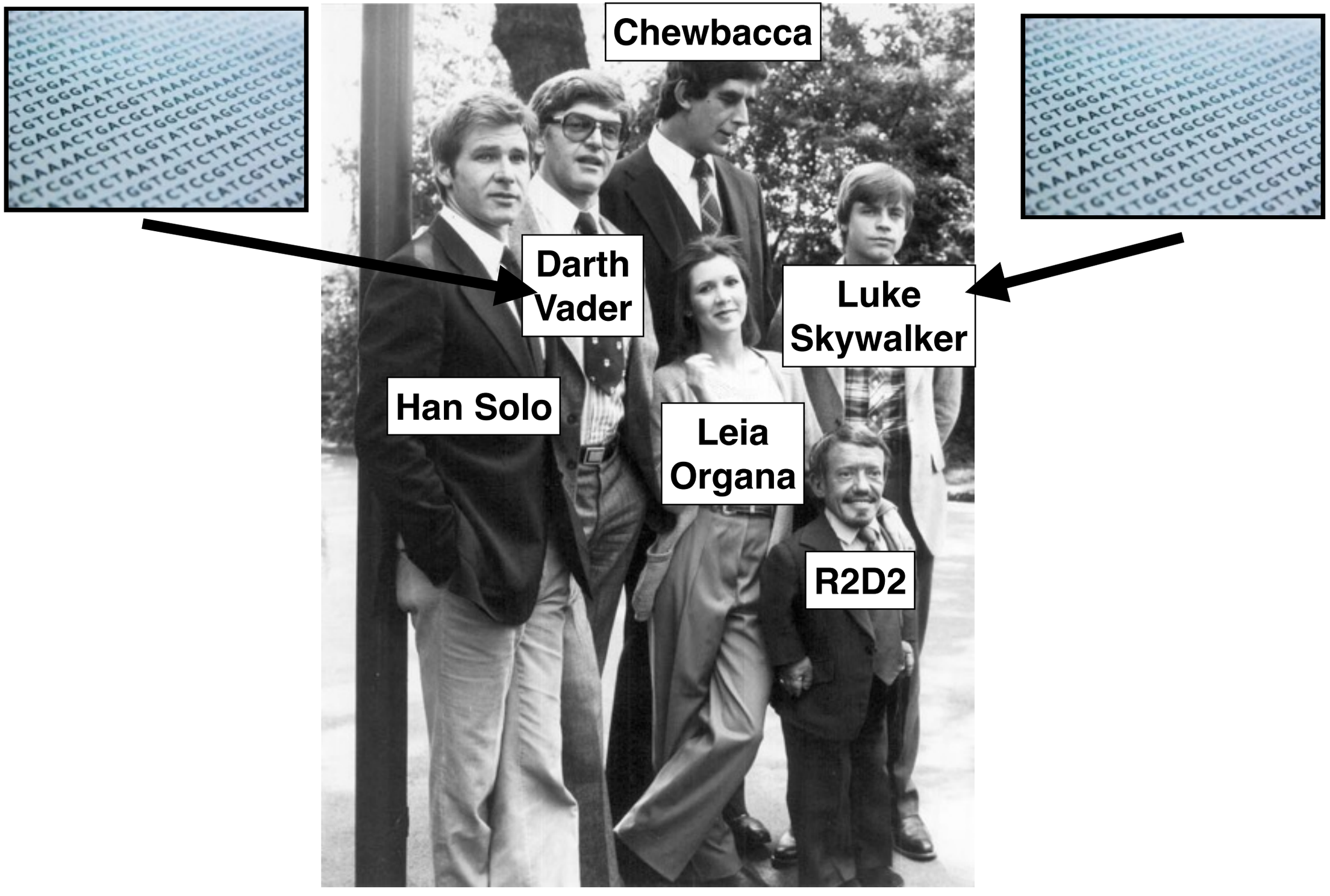

The cast of the original *Star Wars*

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



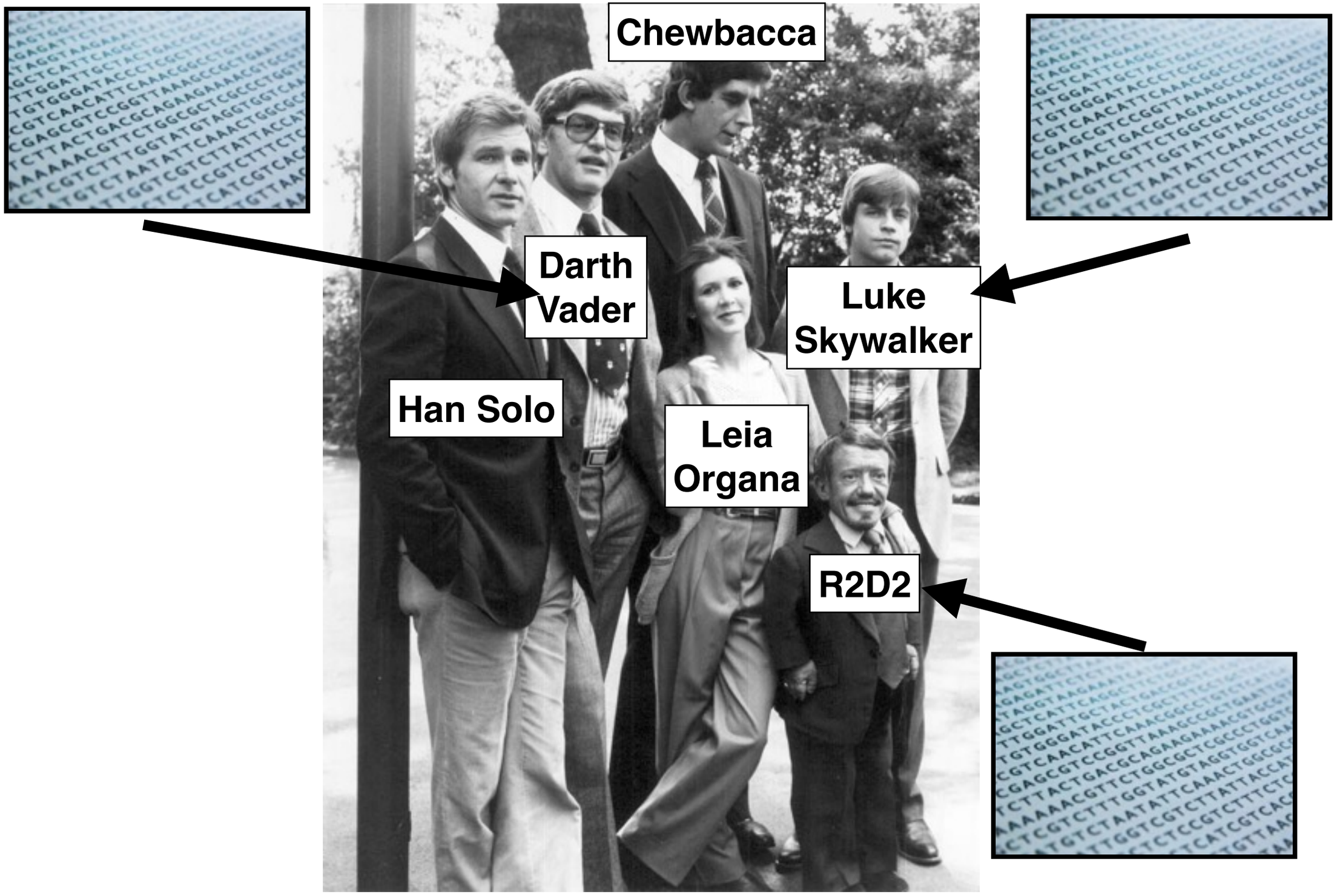

The cast of the original *Star Wars*

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



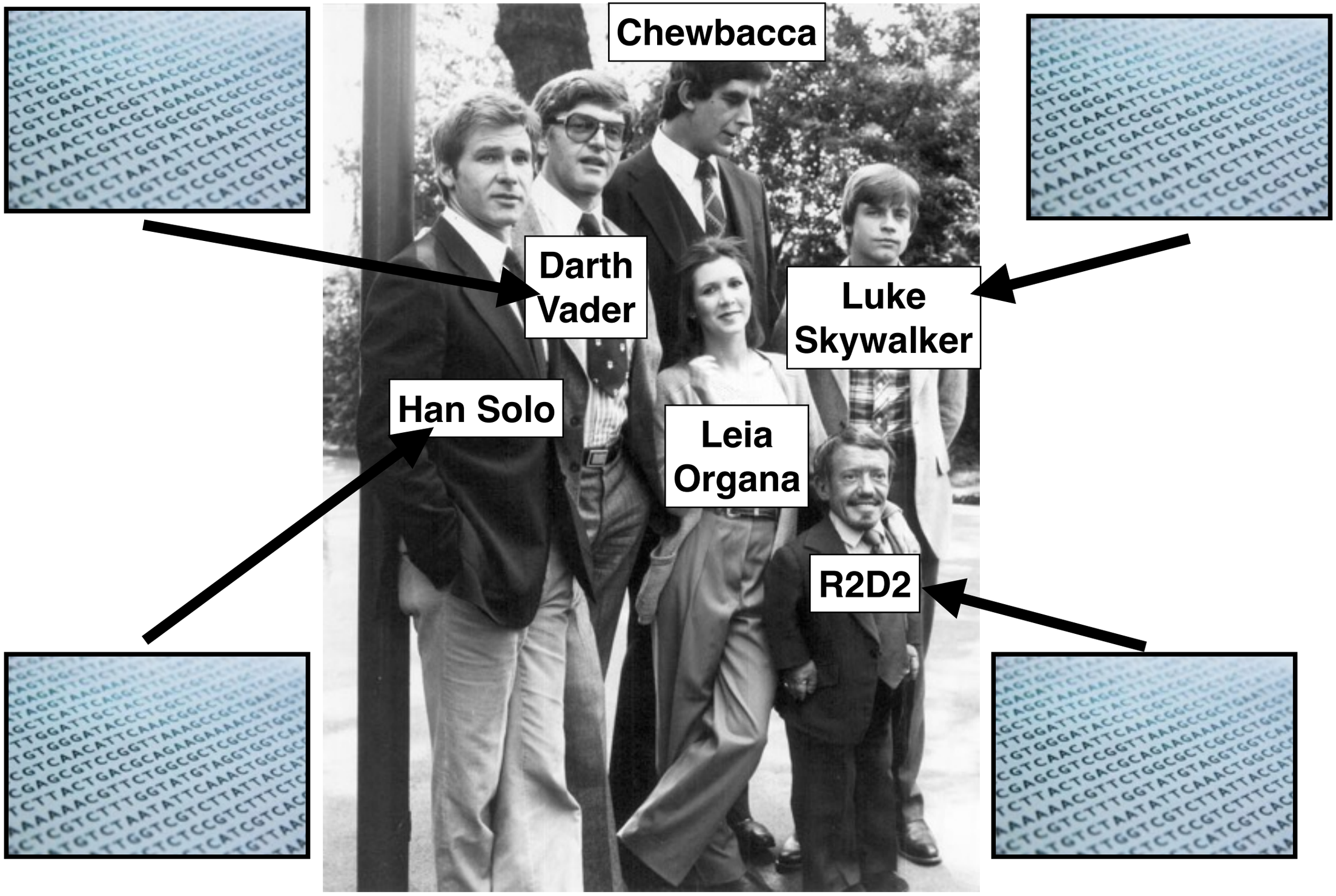

The cast of the original *Star Wars*

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



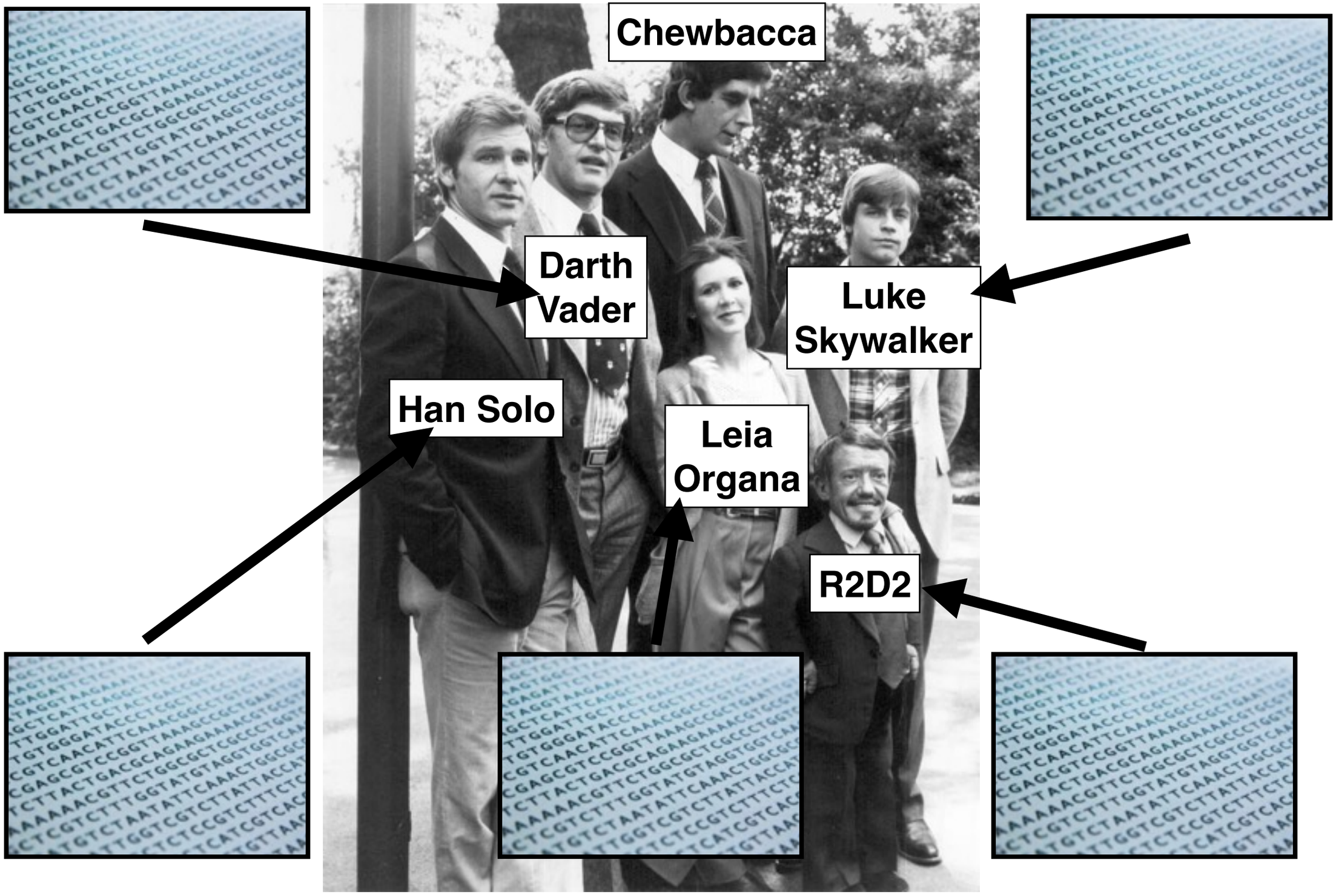

The cast of the original *Star Wars*

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



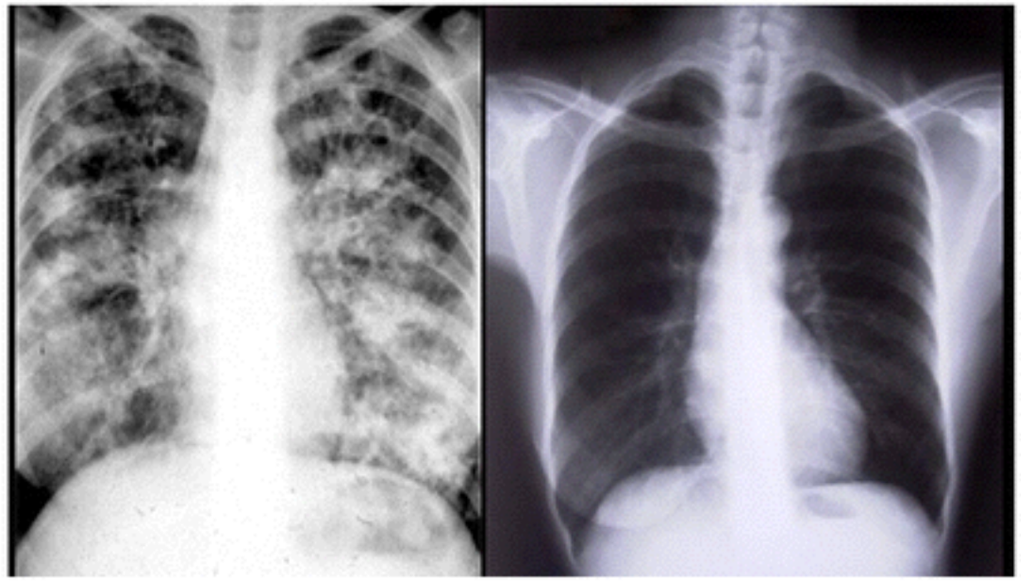

The cast of the original *Star Wars*

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




The cast of the original *Star Wars*

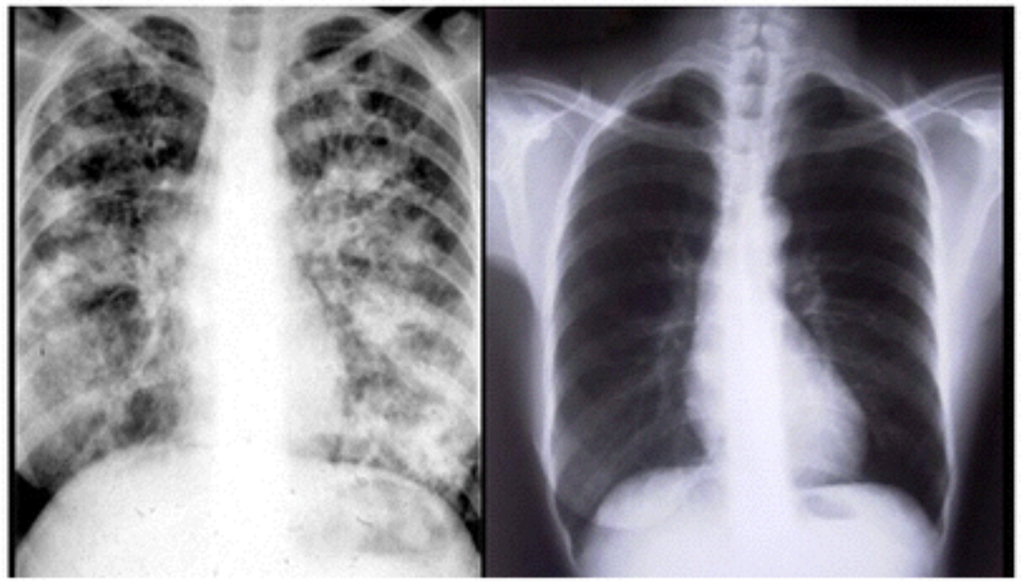
# Cystic fibrosis is caused by a mix of common and rare variants



Cystic Fibrosis Lung

Healthy Lung

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



Cystic Fibrosis Lung  Healthy Lung

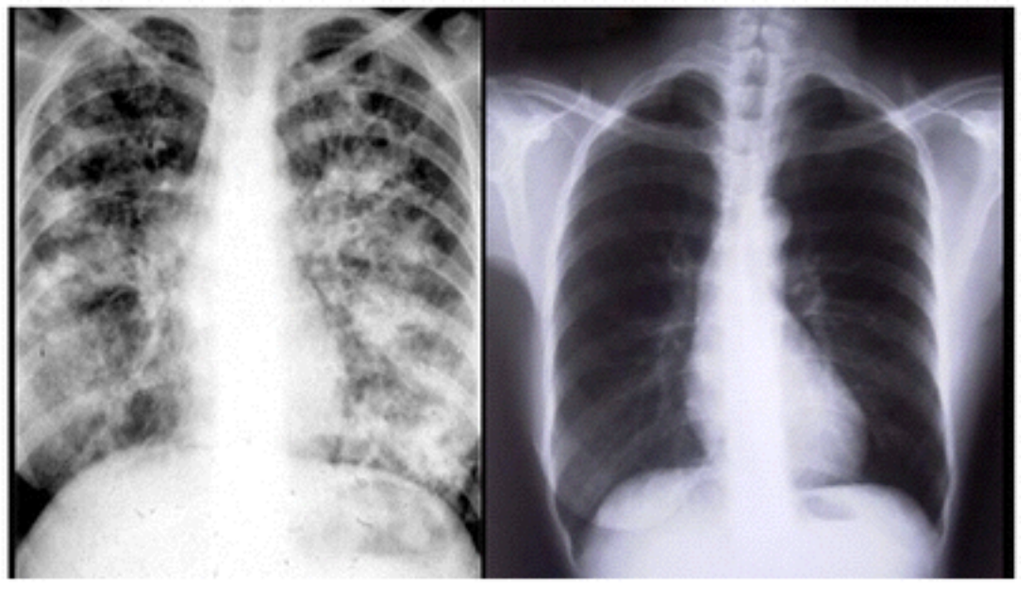
Rare disease affects 1/10,000 live births

Caused by mutations in the CFTR gene

Selection removes homozygotes from population

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

# 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

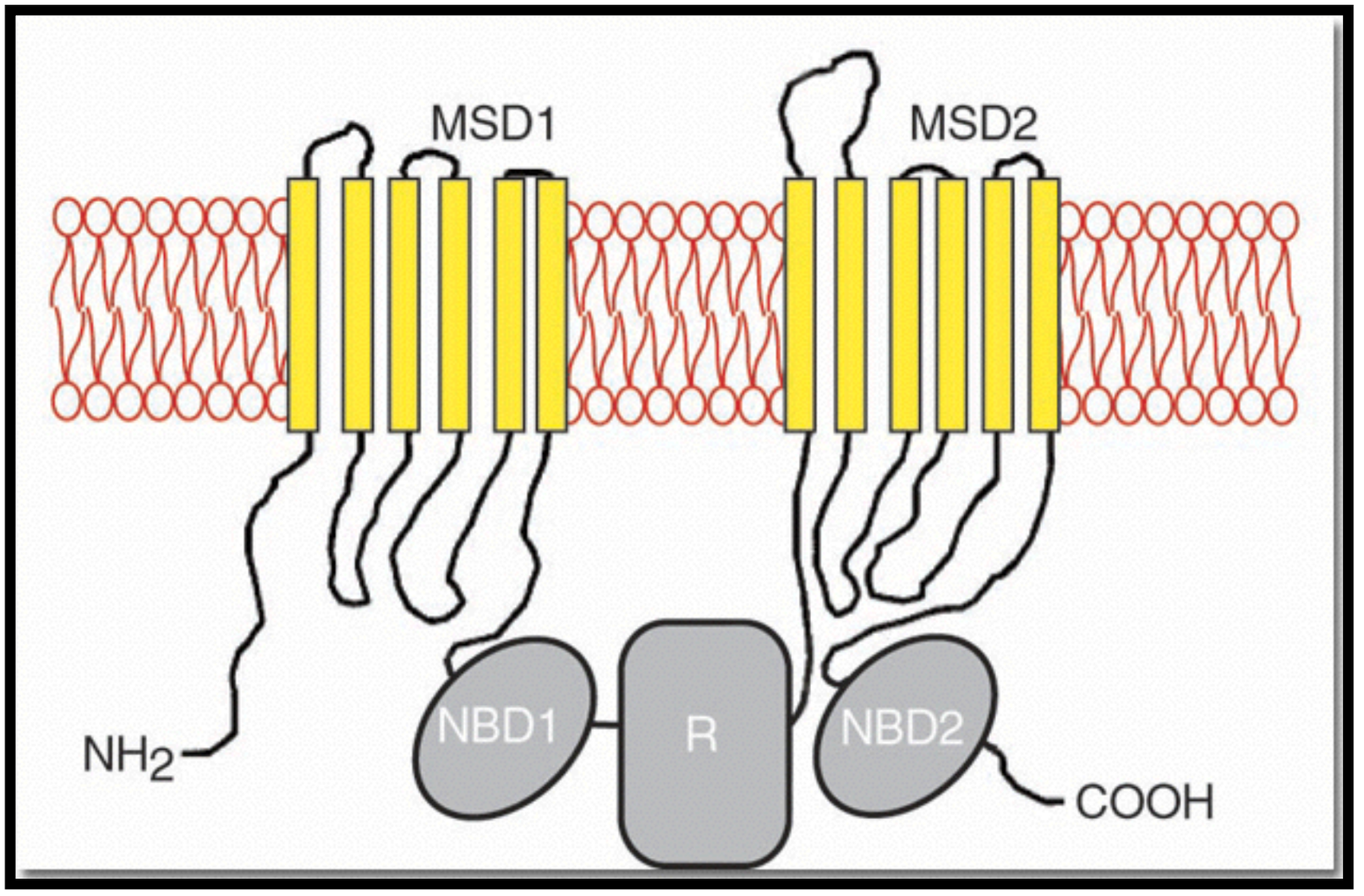
Caused by mutations in the CFTR gene

Selection removes homozygotes from population

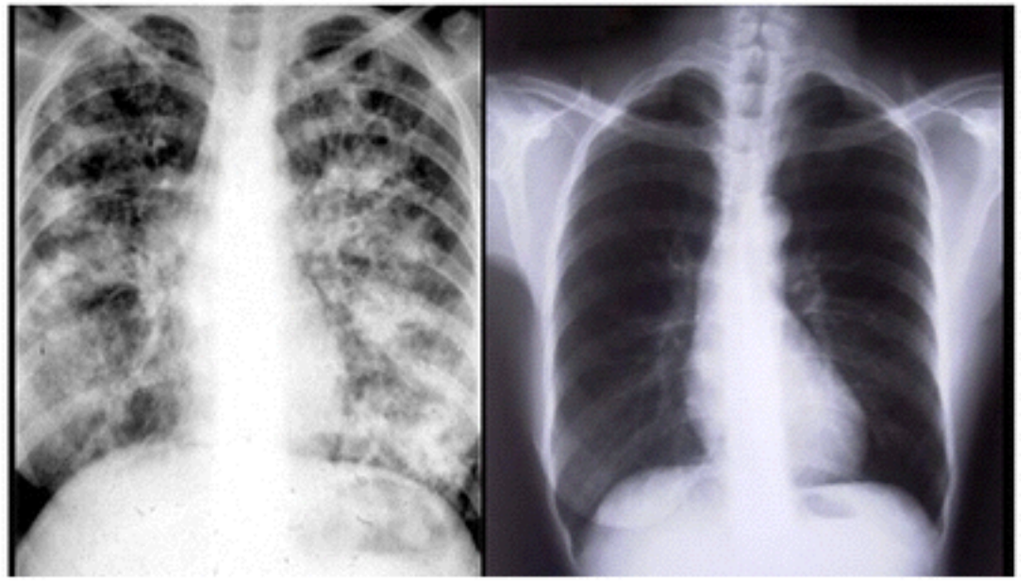
H-W equilibrium tell us that 1/50 people are carriers

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

# Cystic fibrosis is caused by a mix of common and rare variants in the chloride ion channel CFTR



# 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

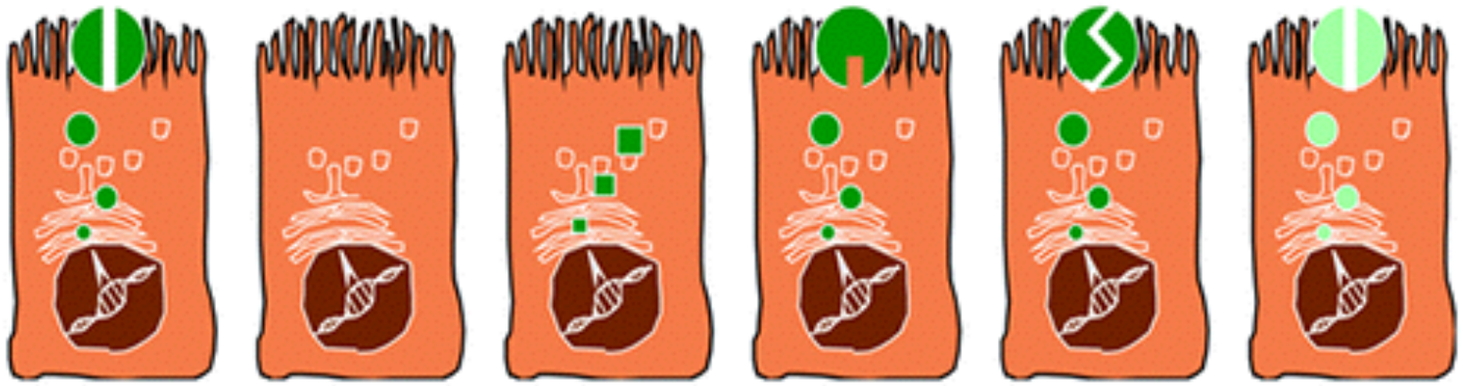
Over 1000 other mutations are known

Compound heterozygotes found often

Genetic heterogeneity

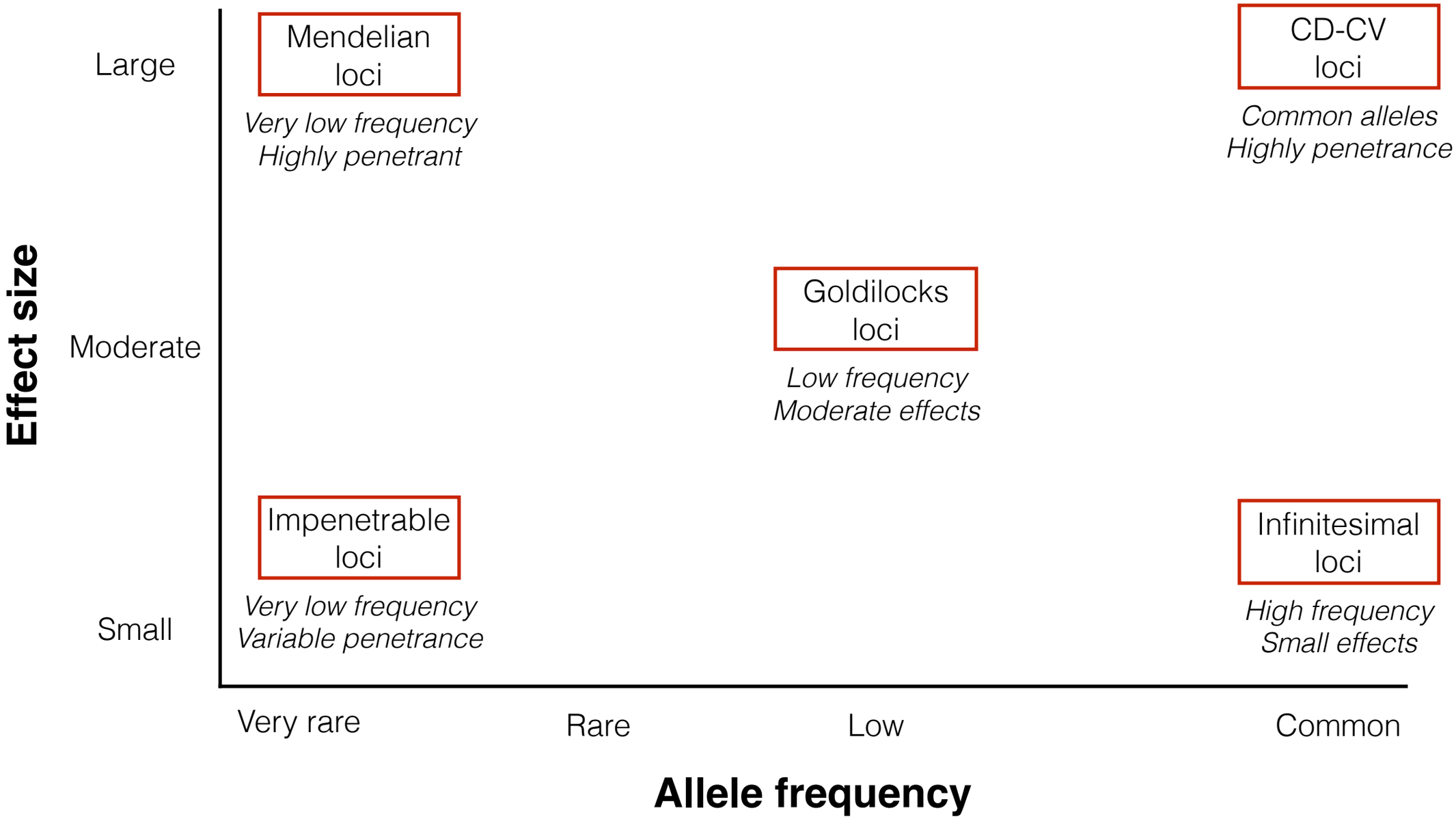
# CFTR

## *Classes of Mutations*

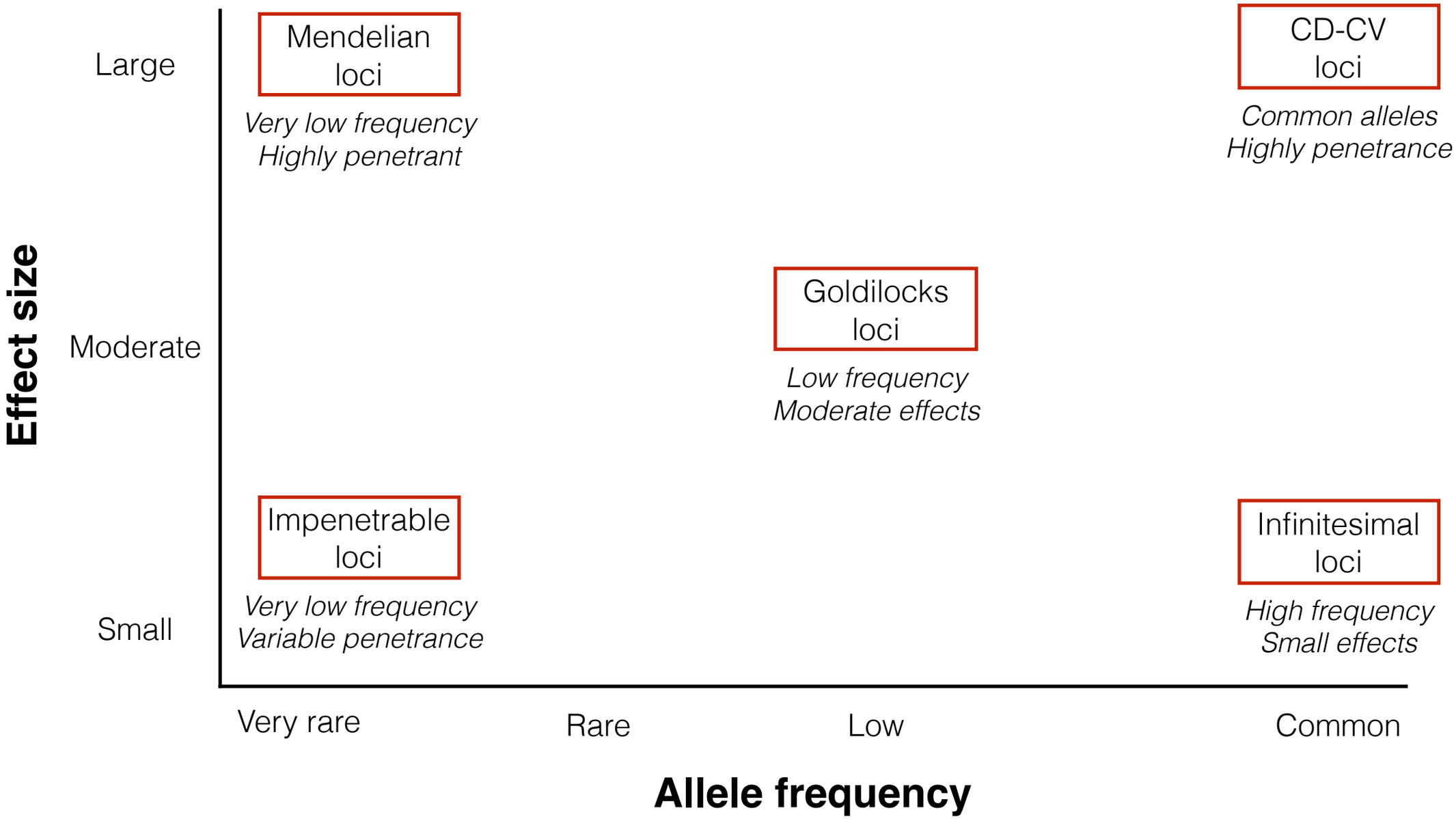



| | No synthesis | Block in processing | Block in regulation | Altered conductance | Reduced synthesis |
|---|---|---|---|---|---|
| | G542X | F508del | G551D | R117H<br>D1152H | 3849+10kbC→T<br>5T<br><br>A455E |

# The spectrum of how variation contributes to disease

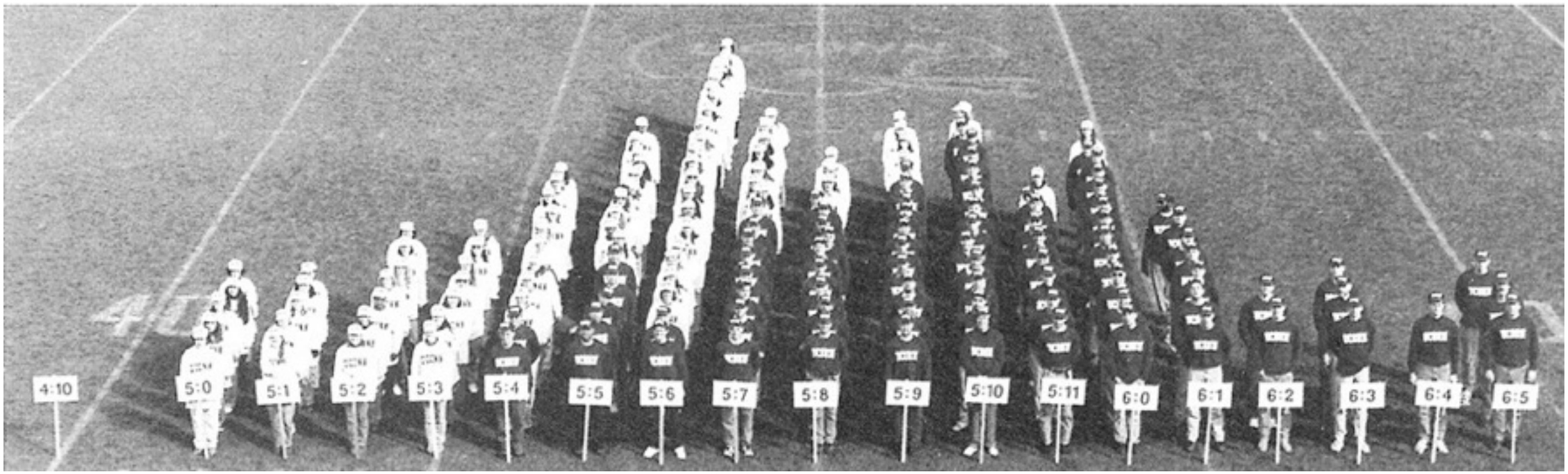


# The spectrum of how variation contributes to disease




How do we find the variants that cause common disease?

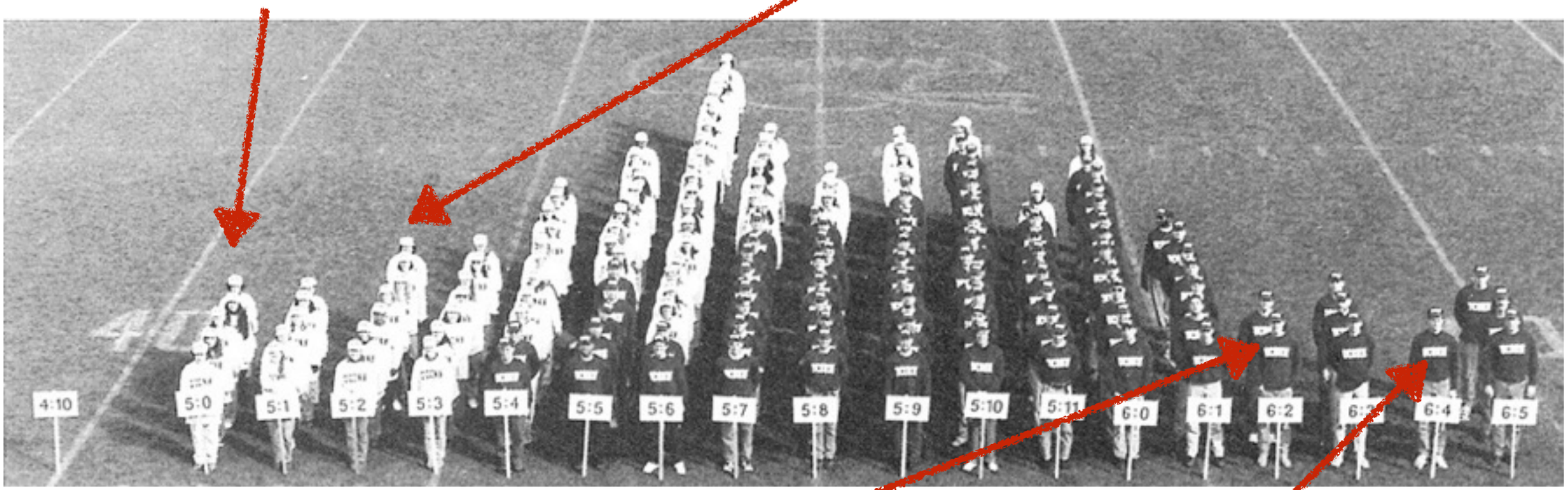
# To find genes in humans, we must correlate genotype with phenotype




**University of Connecticut, 1997**

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

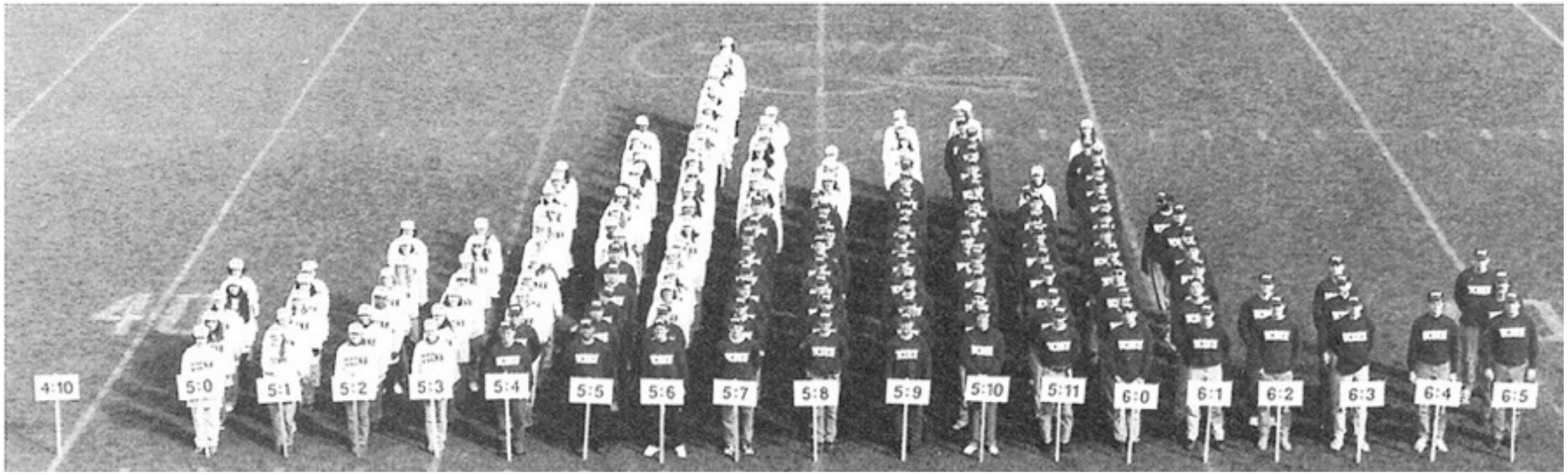
CAGCGATAGGCTTTAATGTT  CAGCGATAGGCTTTAATGTT
AGCCCGTTTTATGACCAACG  AGCCCGTTTTATGACCAACG
GGGTTCACAGTGAGCTGTGT  GGGTTCACAGTGAGCTGTGT




University of Connecticut, 1997

CAGCGATAGGCTTTAATGTT  CAGCGATAGGCTTTAATGTT
AGCCCGTTTGATGACCAACG  AGCCCGTTTGATGACCAACG
GGGTTCACAGTGAGCTGTGT  GGGTTCACAGTGAGCTGTGT

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

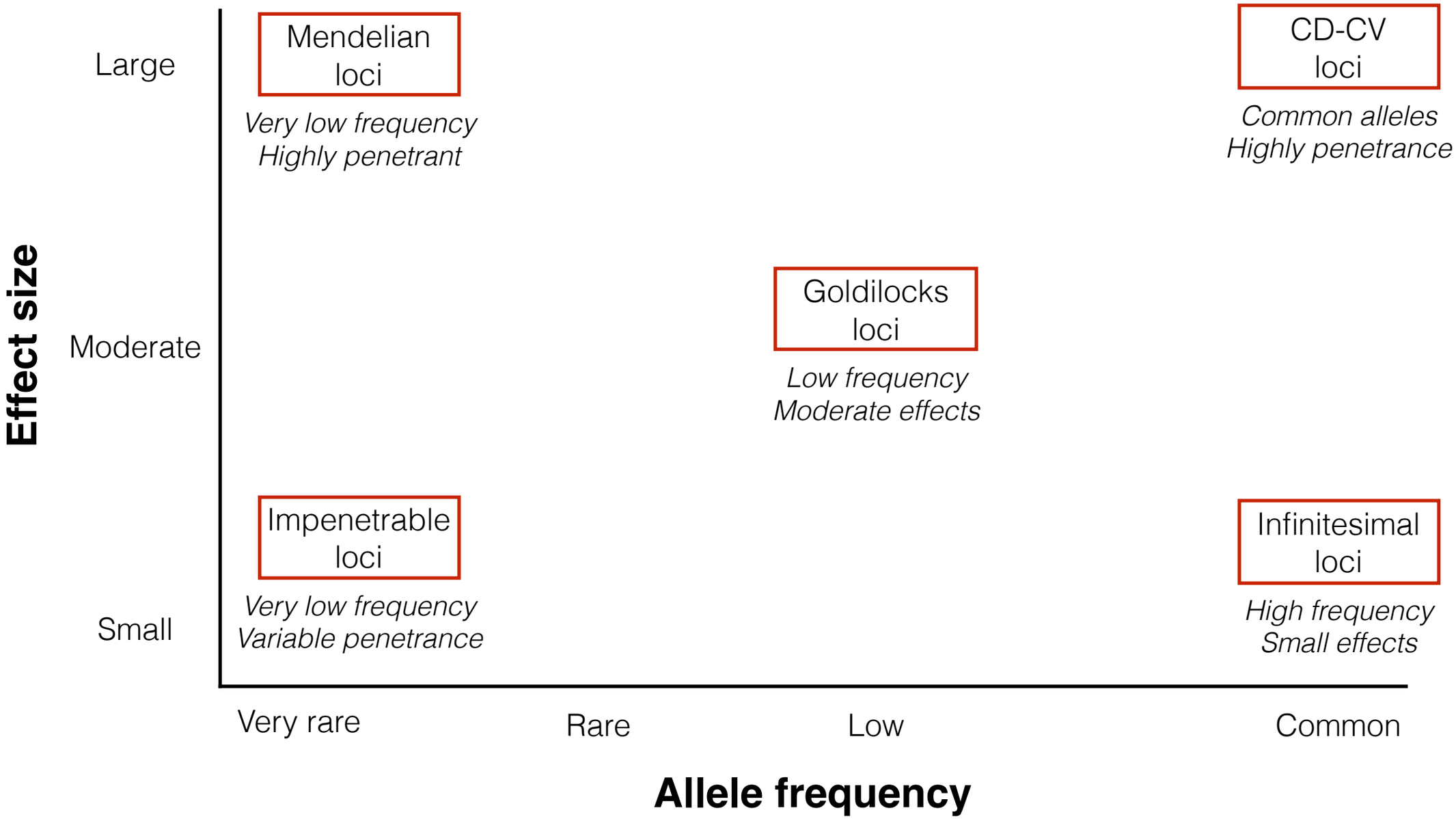



**University of Connecticut, 1997**

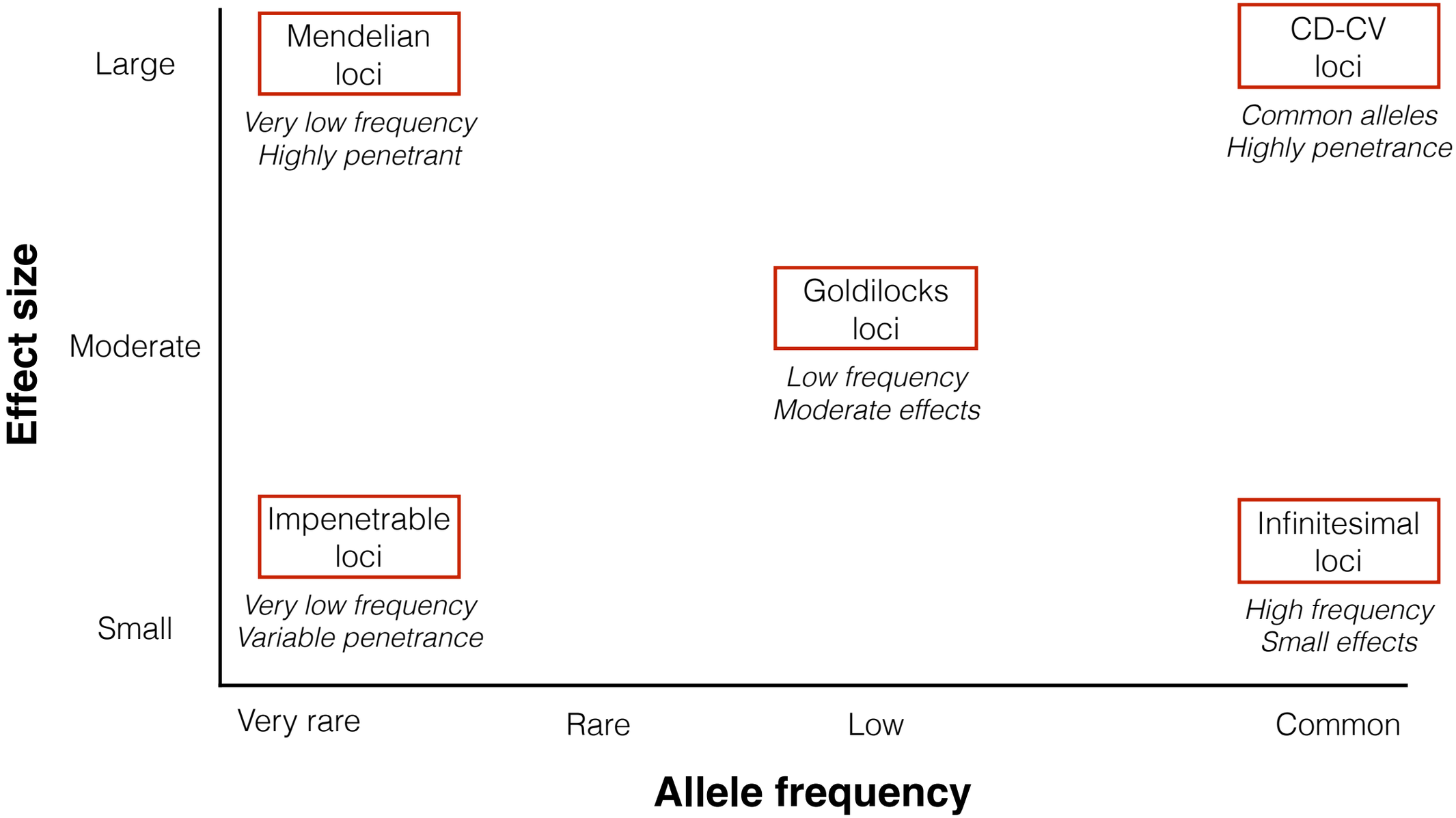
Variation shared by lots of tall people
and not shared by lots of short people

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

# The spectrum of how variation contributes to disease

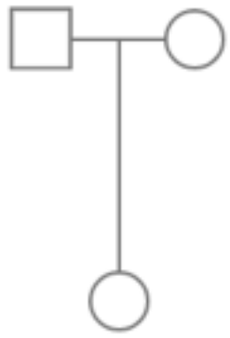


# The spectrum of how variation contributes to disease



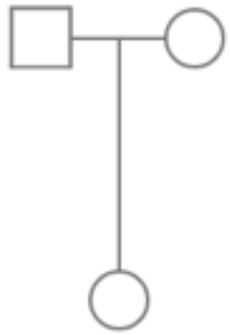

How do we find the variants that cause rare disease?

# Strategies to identify disease-causing rare variants

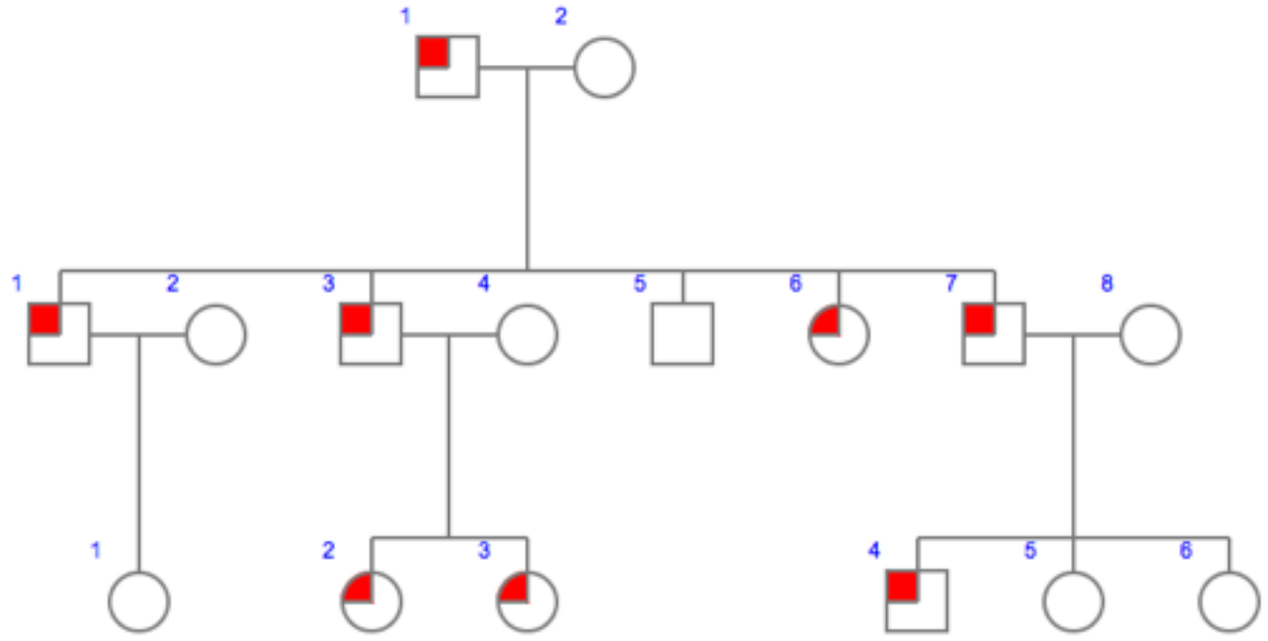


Rare variation from families

# Strategies to identify disease-causing rare variants

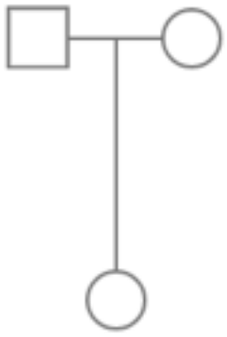


Rare variation from families

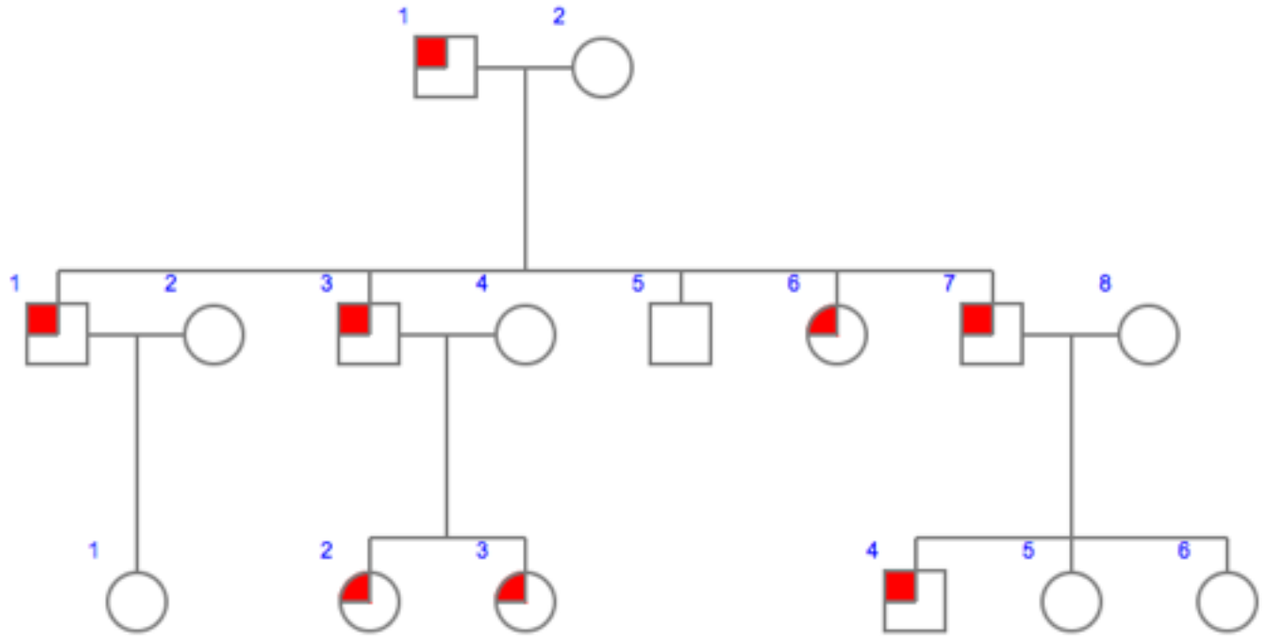



Shared variants from affected individuals in large families

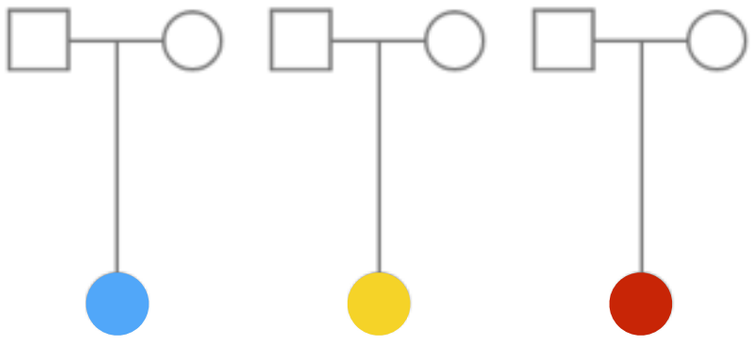
# Strategies to identify disease-causing rare variants

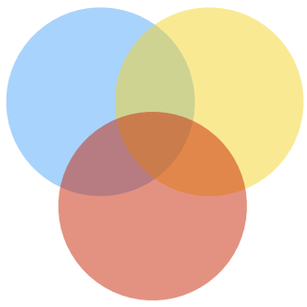


Rare variation from families



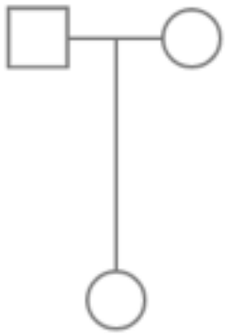

Shared variants from affected individuals in large families



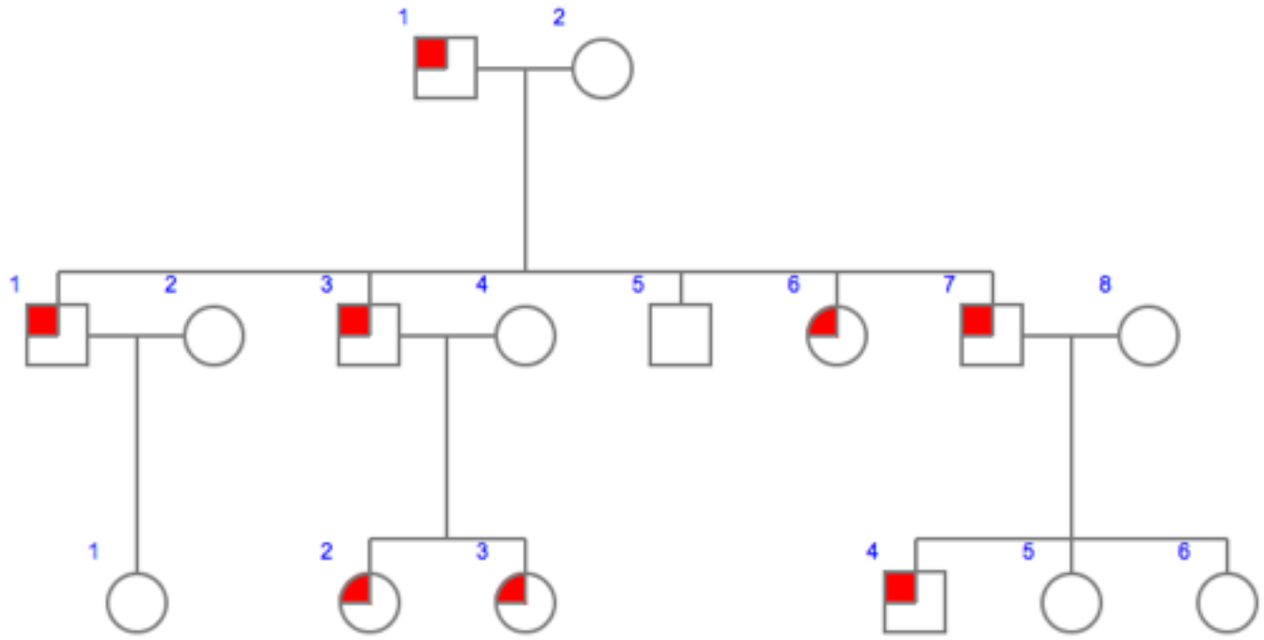


Shared variation from trios

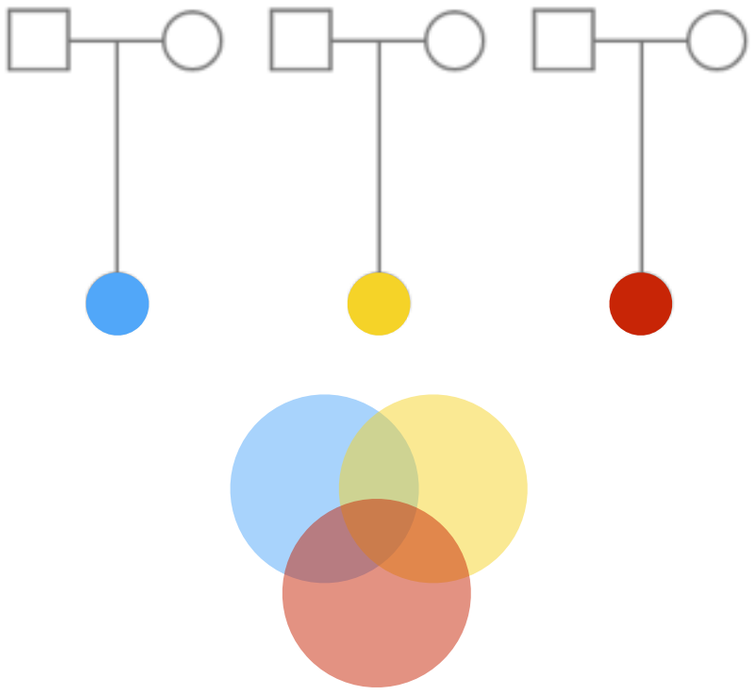
# Strategies to identify disease-causing rare variants



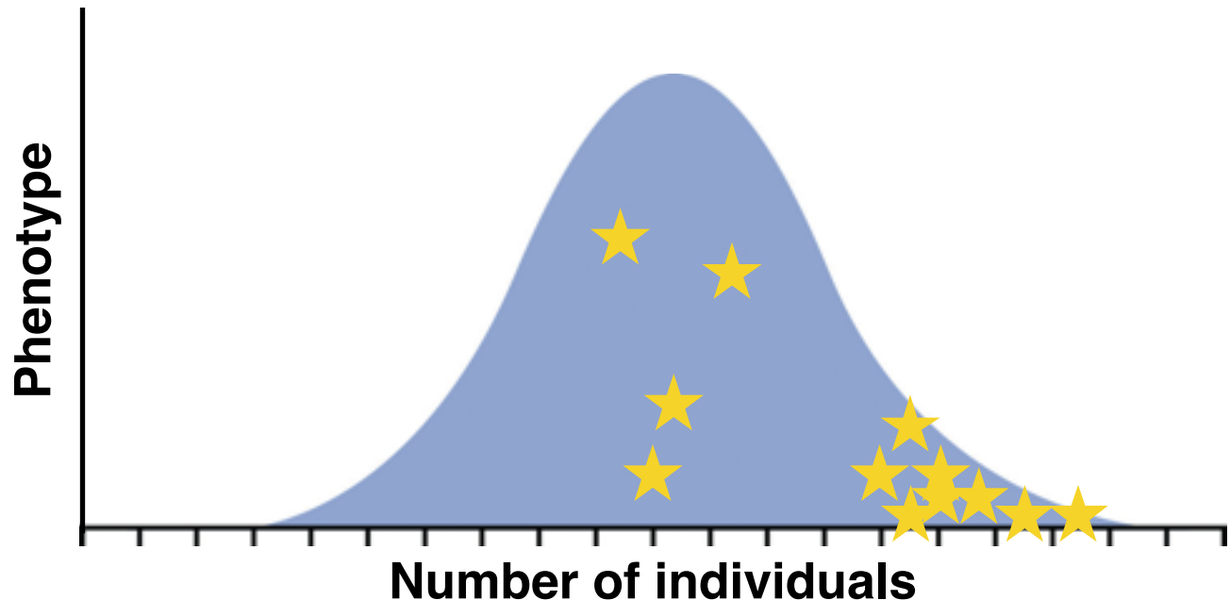
Rare variation from families




Shared variants from affected individuals in large families

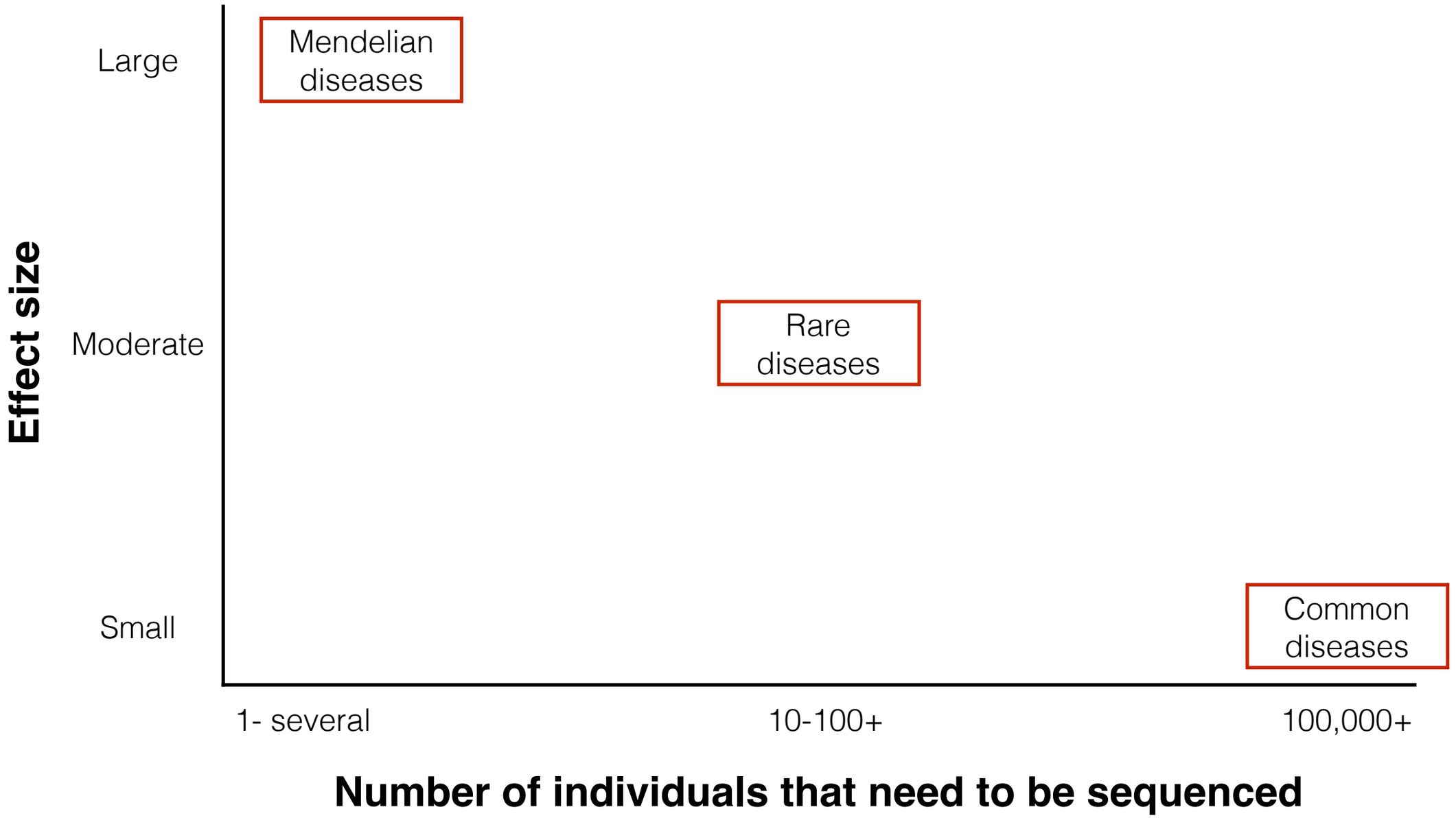


Shared variation from trios




Shared variants from many people

# How can sequencing help us to identify these variants?



# Why can’t we read the genome?

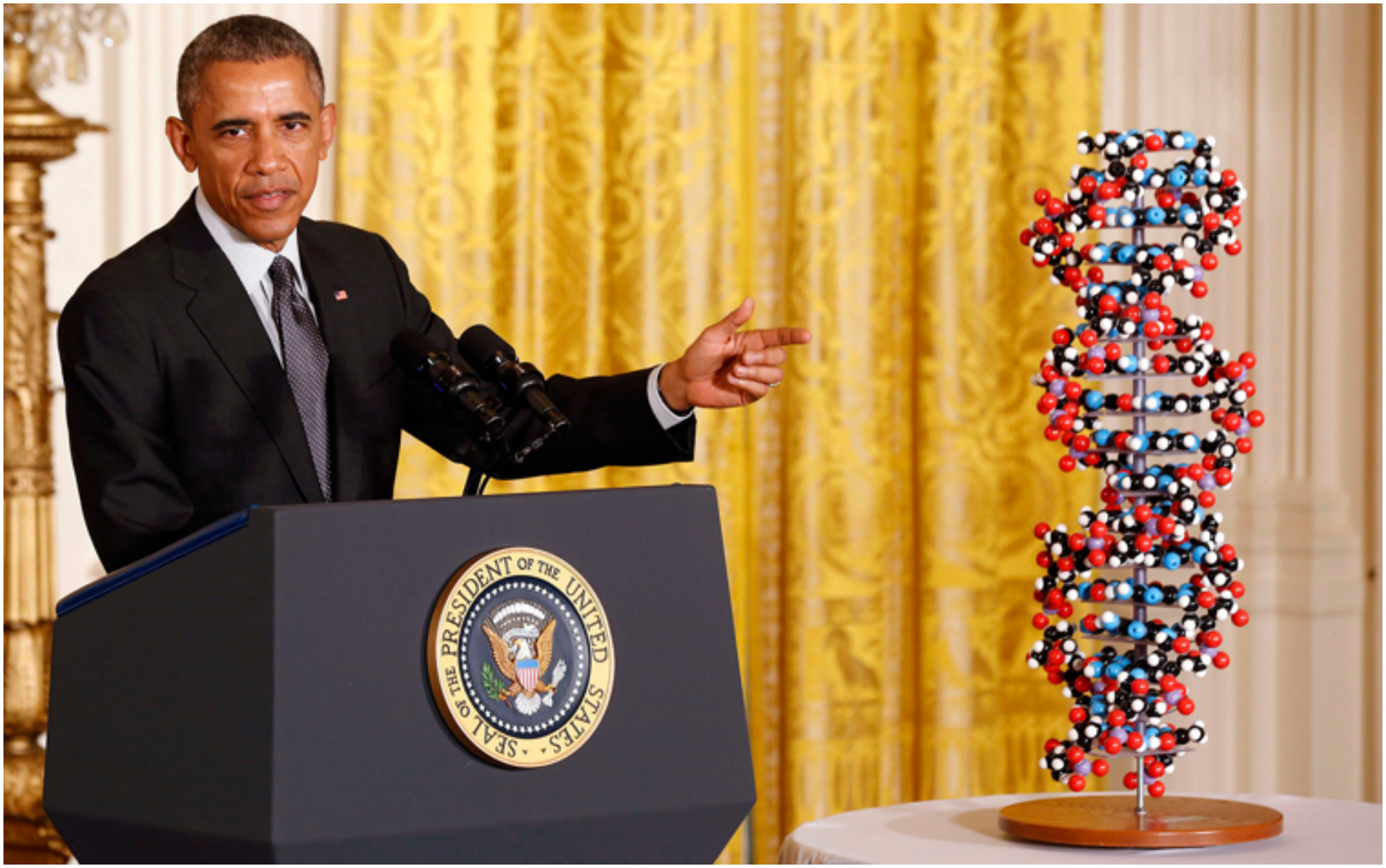
We don’t all the variants.

We don’t know which ones affect phenotype.

The human genome is big.

Phenotypes are highly variable.

# What is precision medicine?



# We are living in the human genetics renaissance

Lecture 14

# We are living in the human genetics renaissance

Under $1000 genome

# We are living in the human genetics renaissance

Under $1000 genome
Rare disease sequencing for Mendelian disorders

# We are living in the human genetics renaissance

Under $1000 genome
Rare disease sequencing for Mendelian disorders
Family genetics

# We are living in the human genetics renaissance

Under $1000 genome
Rare disease sequencing for Mendelian disorders
Family genetics
Fetal testing from sequence

# 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

# 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

# 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