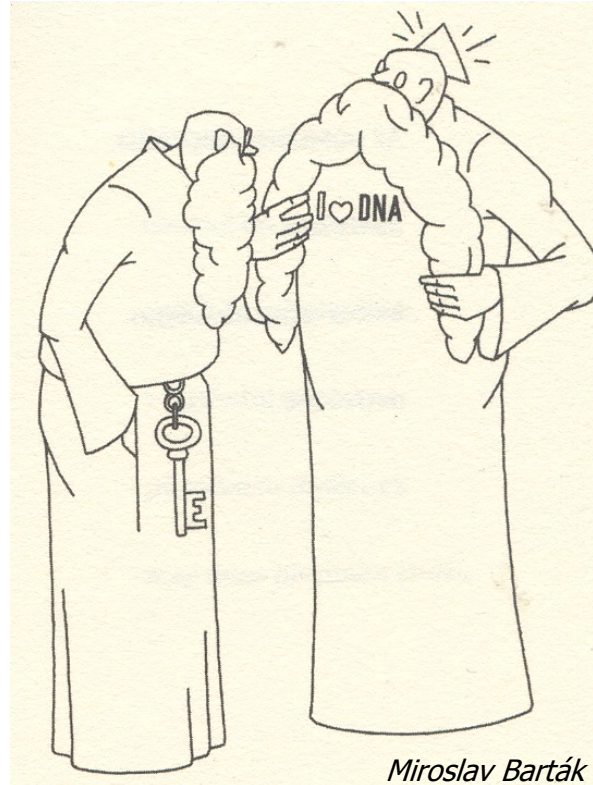


Medical Genetics



Kateřina Staňo Kozubík, Michael Doubek

Why medical genetics

➤ **Genome role in diagnostics, therapy and prevention**

= application in medical practice

It is possible to implement into practice only what I know and what I have in mind

What you should already know

➤ **What is a gene**

genes: structural
for functional RNAs

- housekeeping genes
- gene expression
- exons, introns, non-transcribed regions, promoters

➤ **Informational macromolecules**

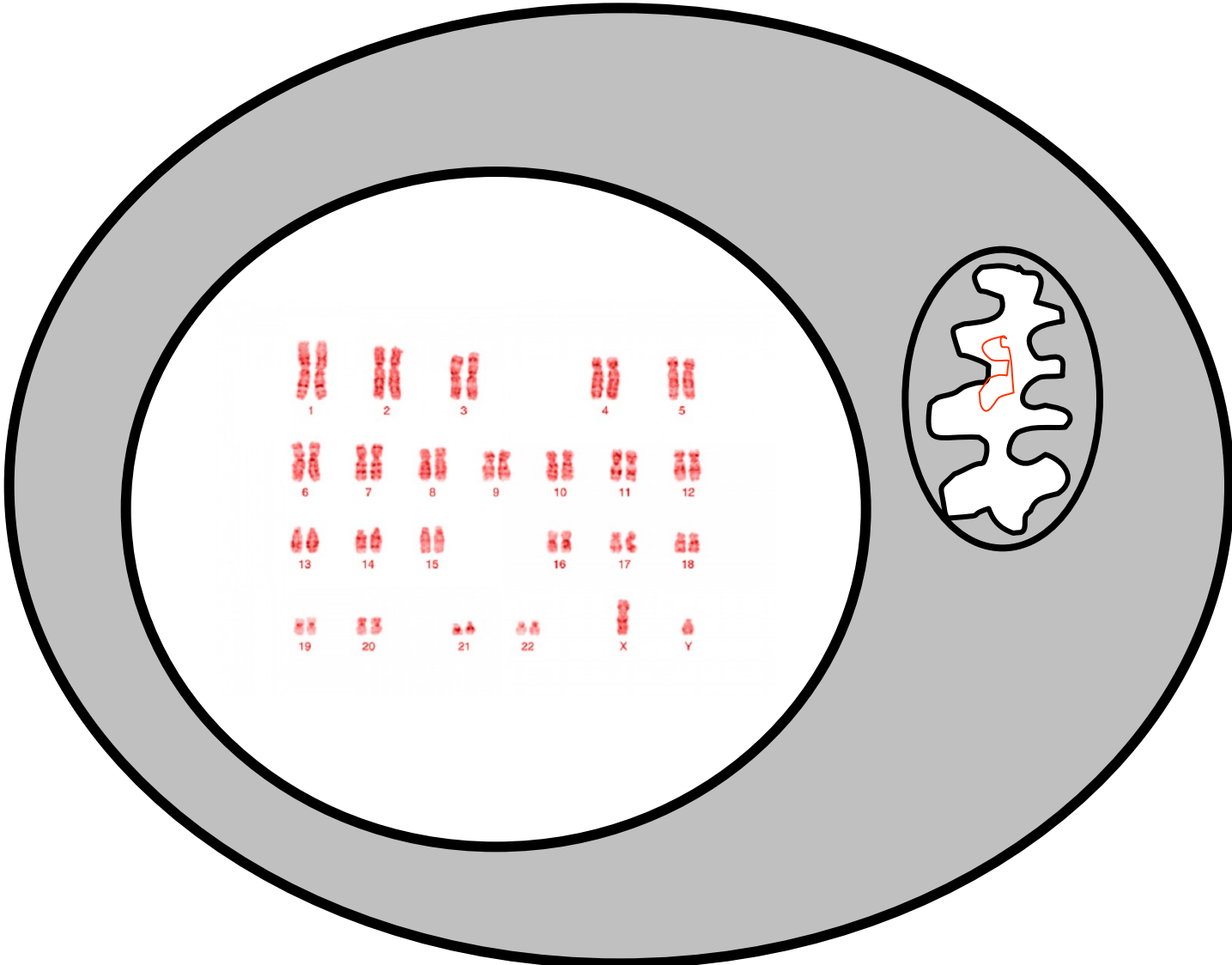
➤ **Transcription, alternative splicing, translation**

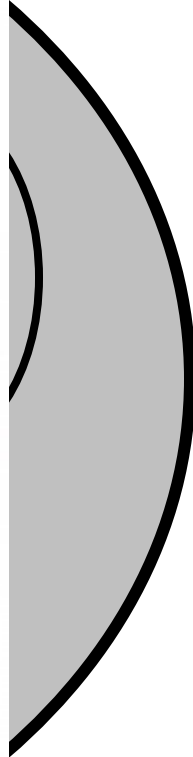
➤ **Chromosomes**

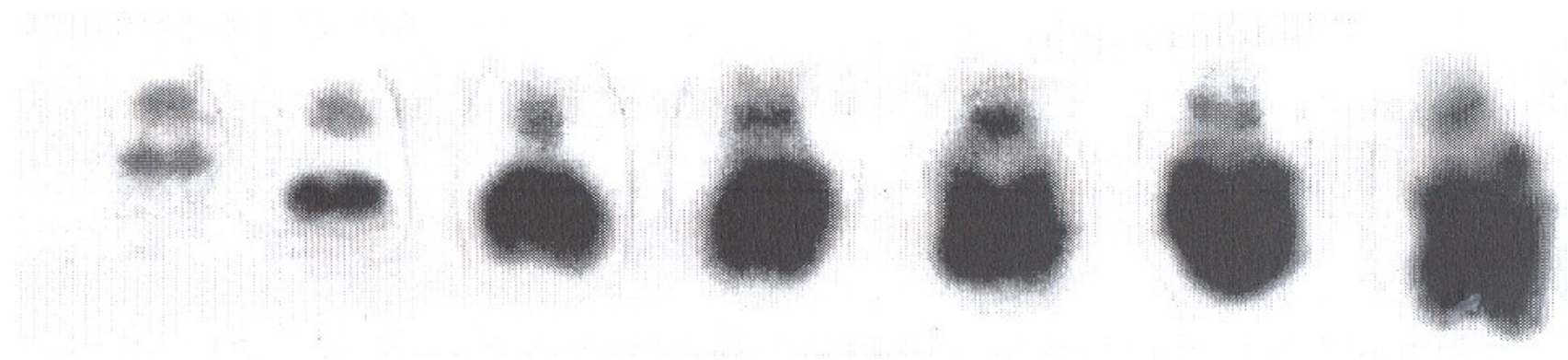
What is a DNA?



**cold ethanol + salt + detergent
> 1 m DNA**









Cavendish Laboratory and The Eagle Pub
Watson + Crick + Wilkins + Franklin



?

Using results without
permit
Eugnenics
Black men's IQ



Cavendish Laboratory and The Eagle Pub
Watson + Crick + Wilkins + Franklin

Terminology

- **Genetics:** study of genes, genetic variation, and heredity in living organism
- **Genome:** complete set of DNA within a single cell of an organism
- **Genomics:** focuses on the structure, function, evolution, and mapping of genomes

Terminology

➤ **Genetics**

➤ **Genome**

➤ **Genomics**



Genome is more than
just a sum of genes

Terminology

➤ **Genetics**

➤ **Genome**

➤ **Genomics**

- Structural
(DNA, chromosomes)
- Functional
(RNA, gene expression)
- Comparative

Terminology

- **Genetics**
- **Genome**
- **Genomics**
- **Microbiome**
- **Transcriptome**
- **Epigenetics**

Terminology

➤ **Genetics**

➤ **Genome**

➤ **Genomics**

➤ **Microbiome**

➤ **Transcriptome**

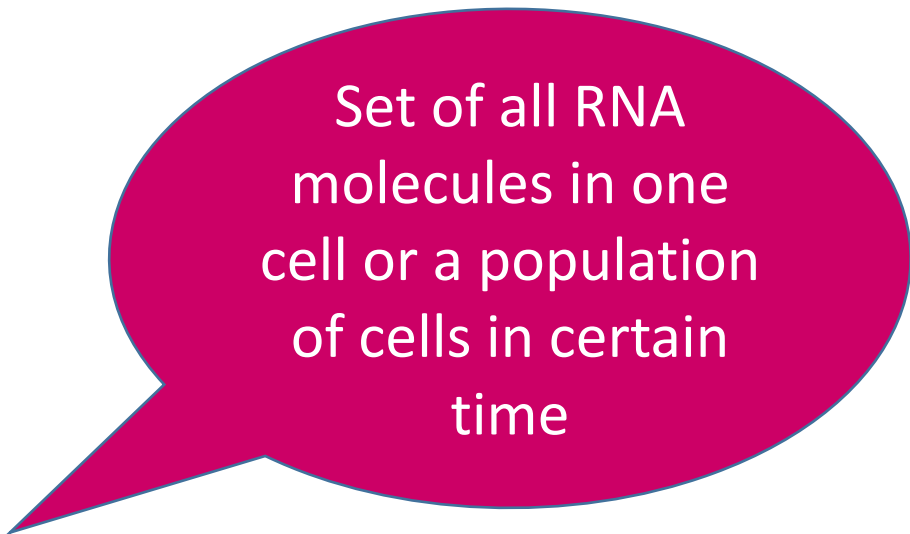
➤ **Epigenetics**



Community
of microorganisms
inhabiting
a particular
environment

Terminology


- **Genetics**
- **Genome**
- **Genomics**
- **Microbiome**
- **Transcriptome**
- **Epigenetics**



Set of all RNA molecules in one cell or a population of cells in certain time

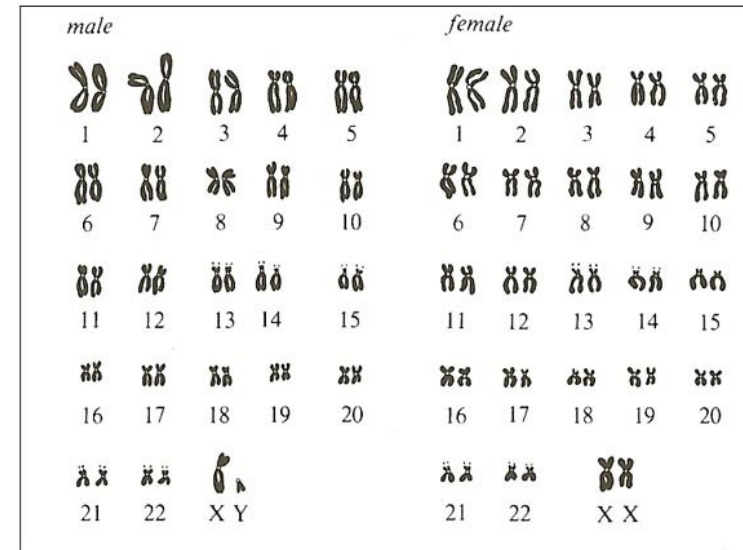
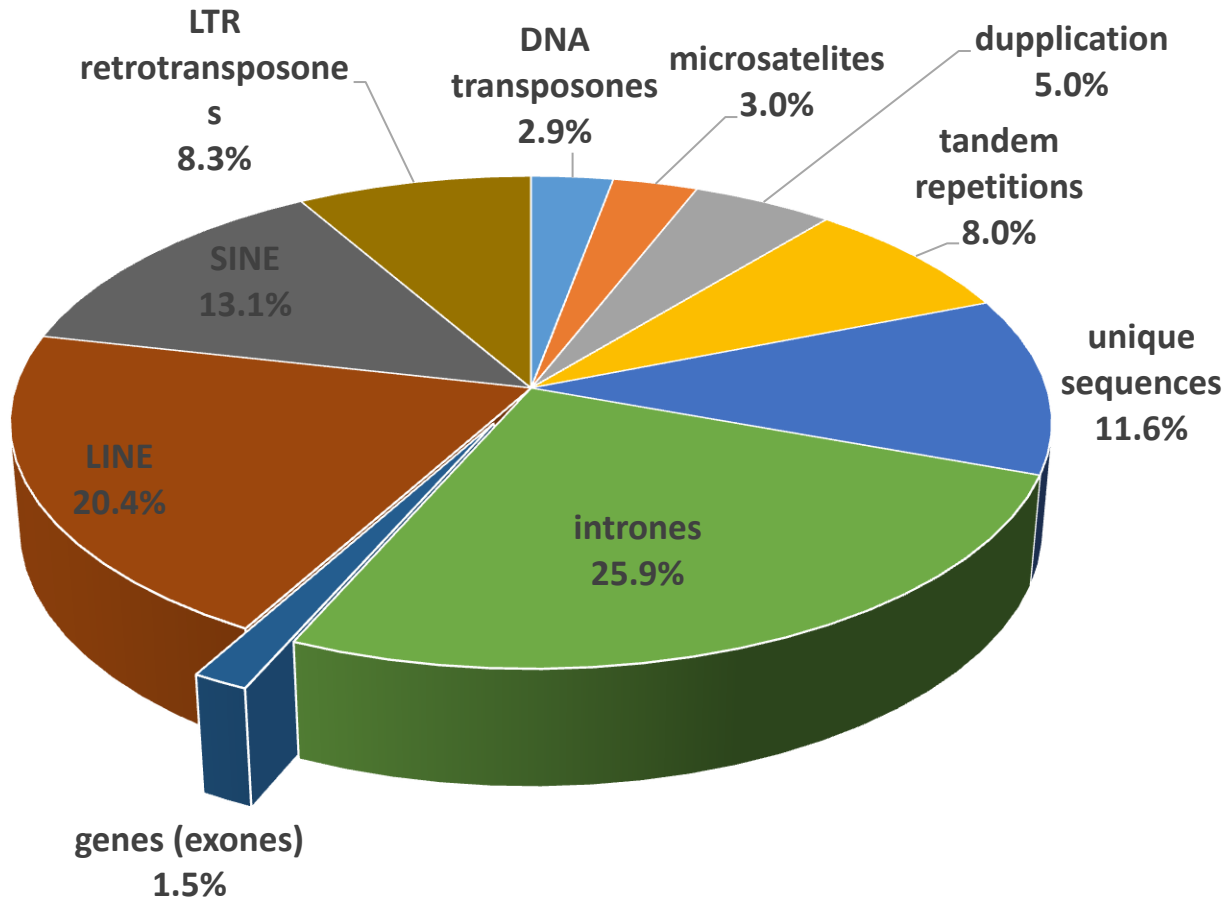
Terminology

- **Genetics**
- **Genome**
- **Genomics**
- **Microbiome**
- **Transcriptome**
- **Epigenetics**



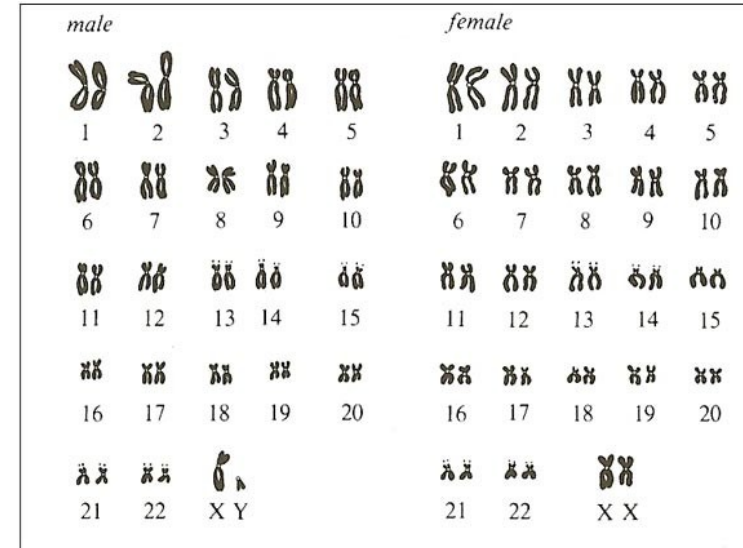
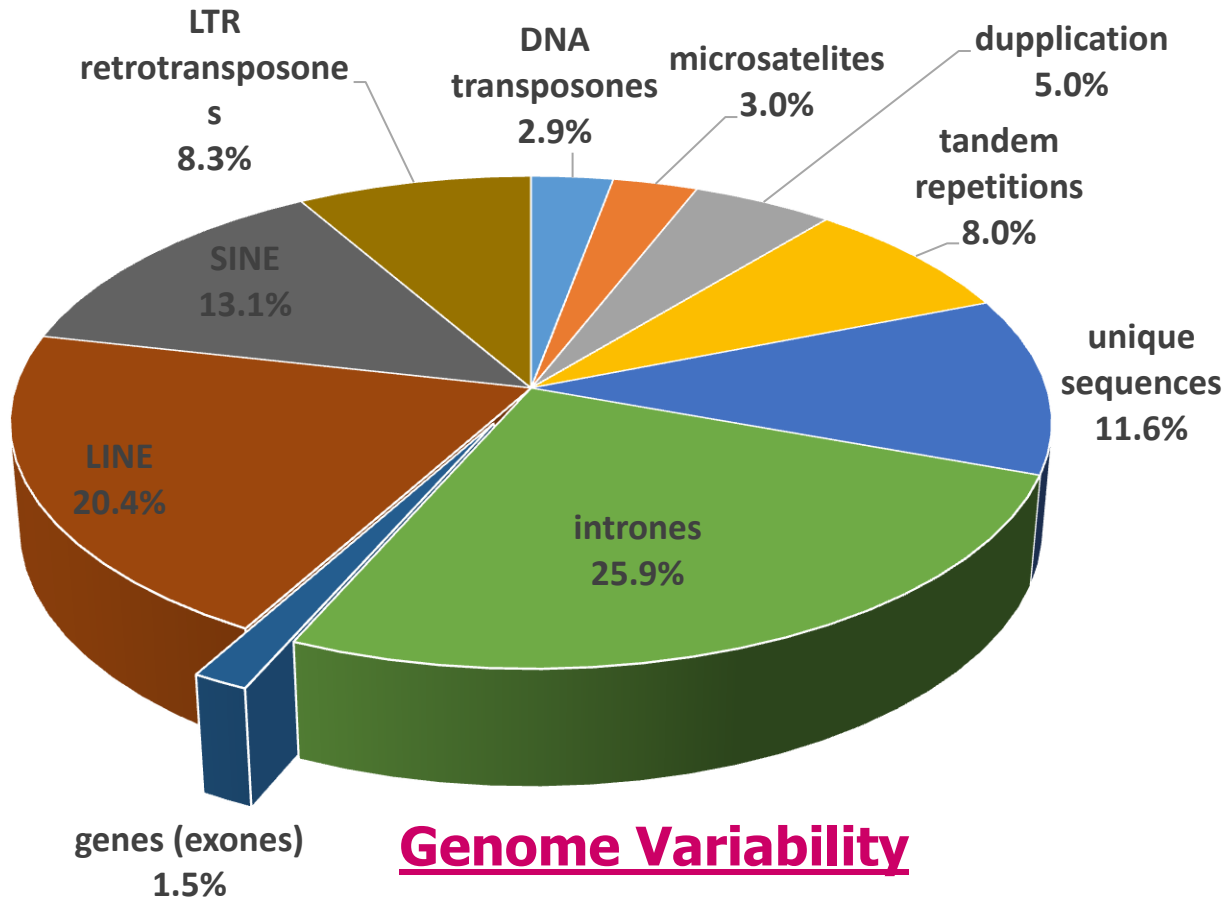
Study of heritable changes in gene function that do not involve changes in the DNA sequence

What is a genome?



**Human genome:
3.2 x 10⁹ bp,
~ 20,000 genes**

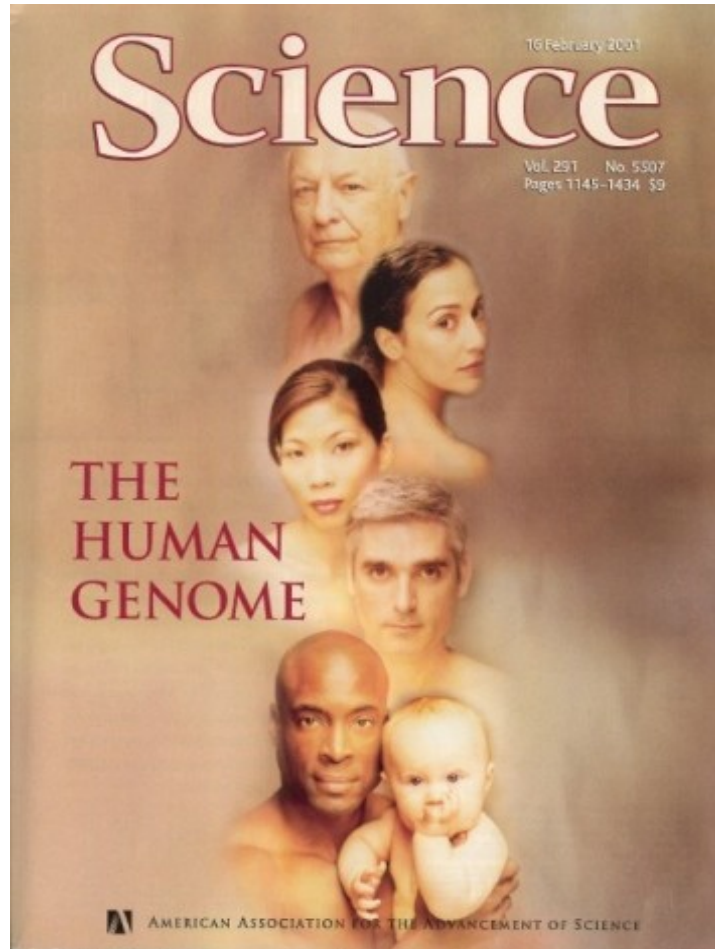
What is a genome?



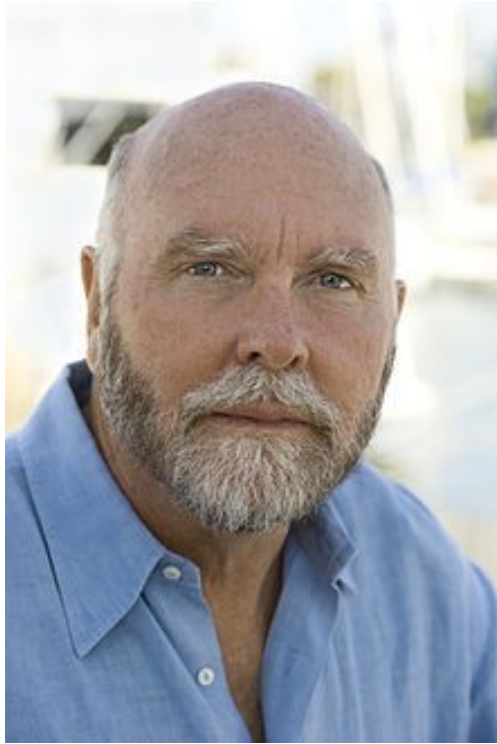
Genome Variability

- Nucleotide polymorphism
 - Single Nucleotide Polymorphisms - SNP
- Structural variations
 - Copy Number Variations – CNV
 - Short Tandem Repeats – STR (2-5)

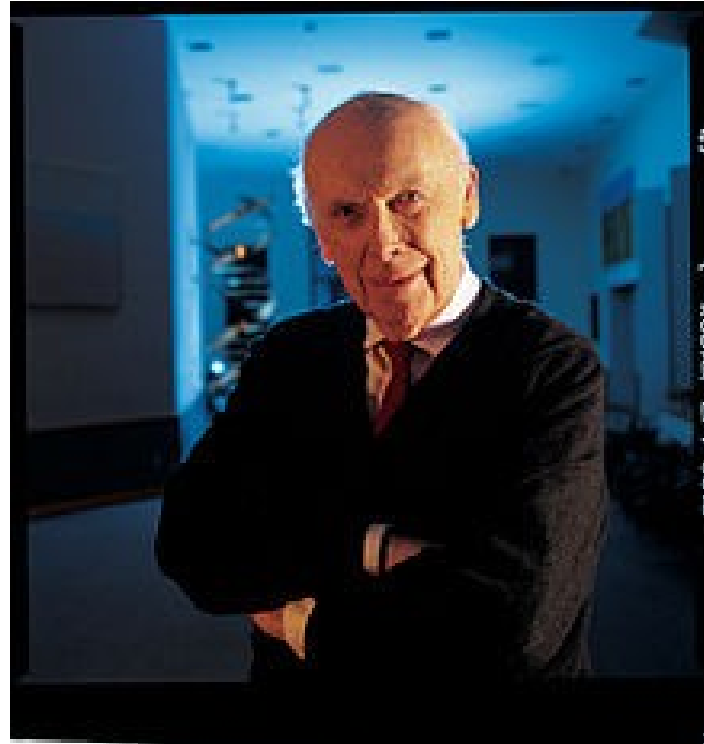
**Human genome:
3.2 x 10⁹ bp,
~ 20,000 genes**



Human genome was published in 2001

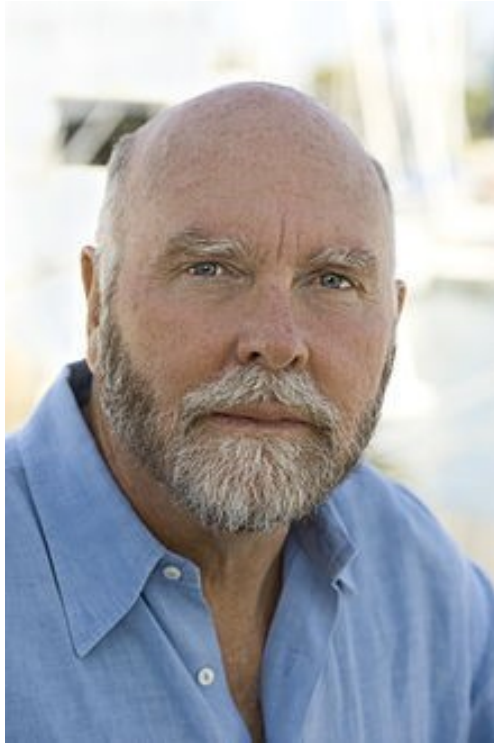


C. Venter

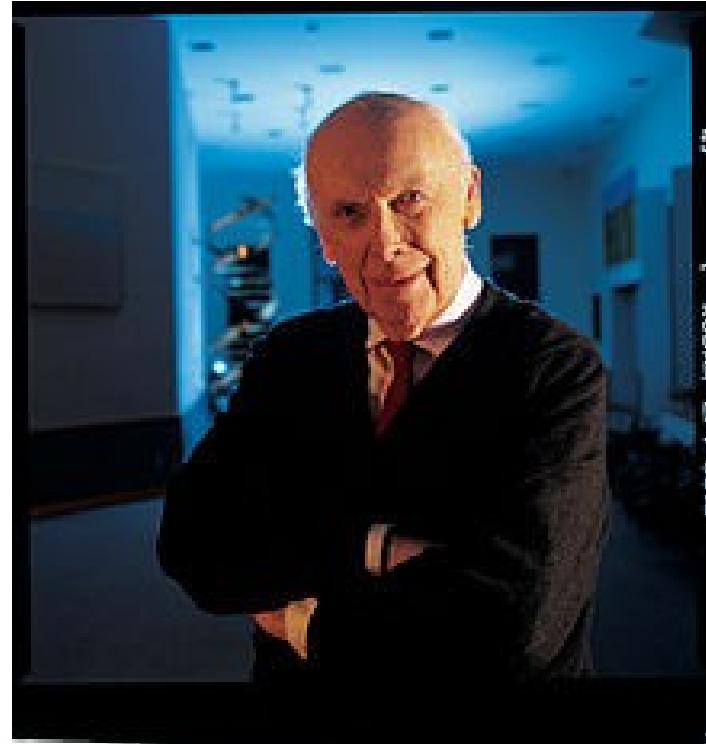


J. D. Watson

Individual sequences of human genomes were published in 2007 and 2008



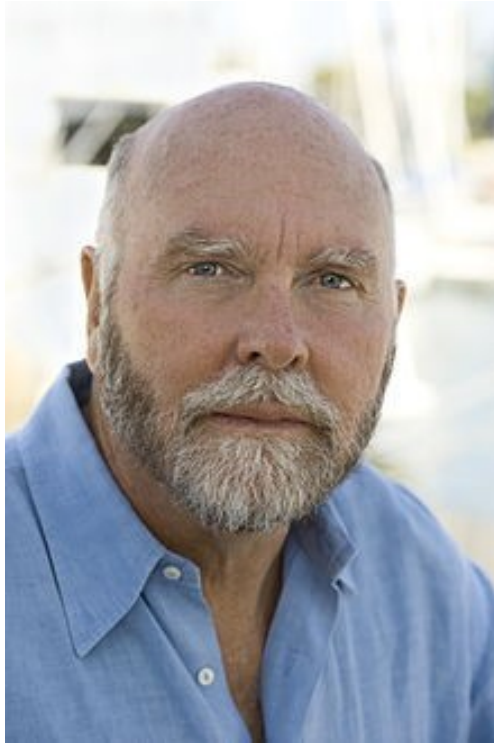
C. Venter



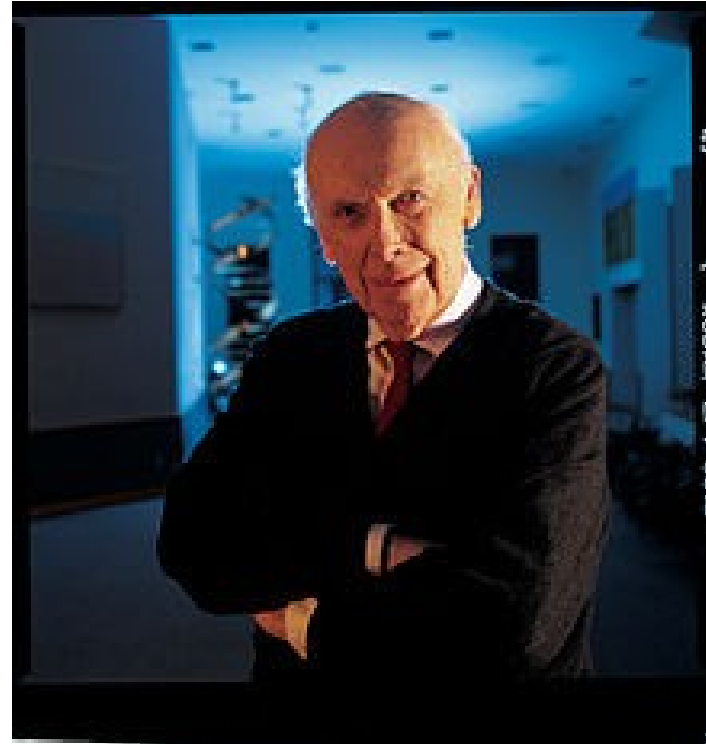
J. D. Watson

Individual sequences of human genomes were published in 2007 and 2008

Difference in 7648 amino acid substitutions



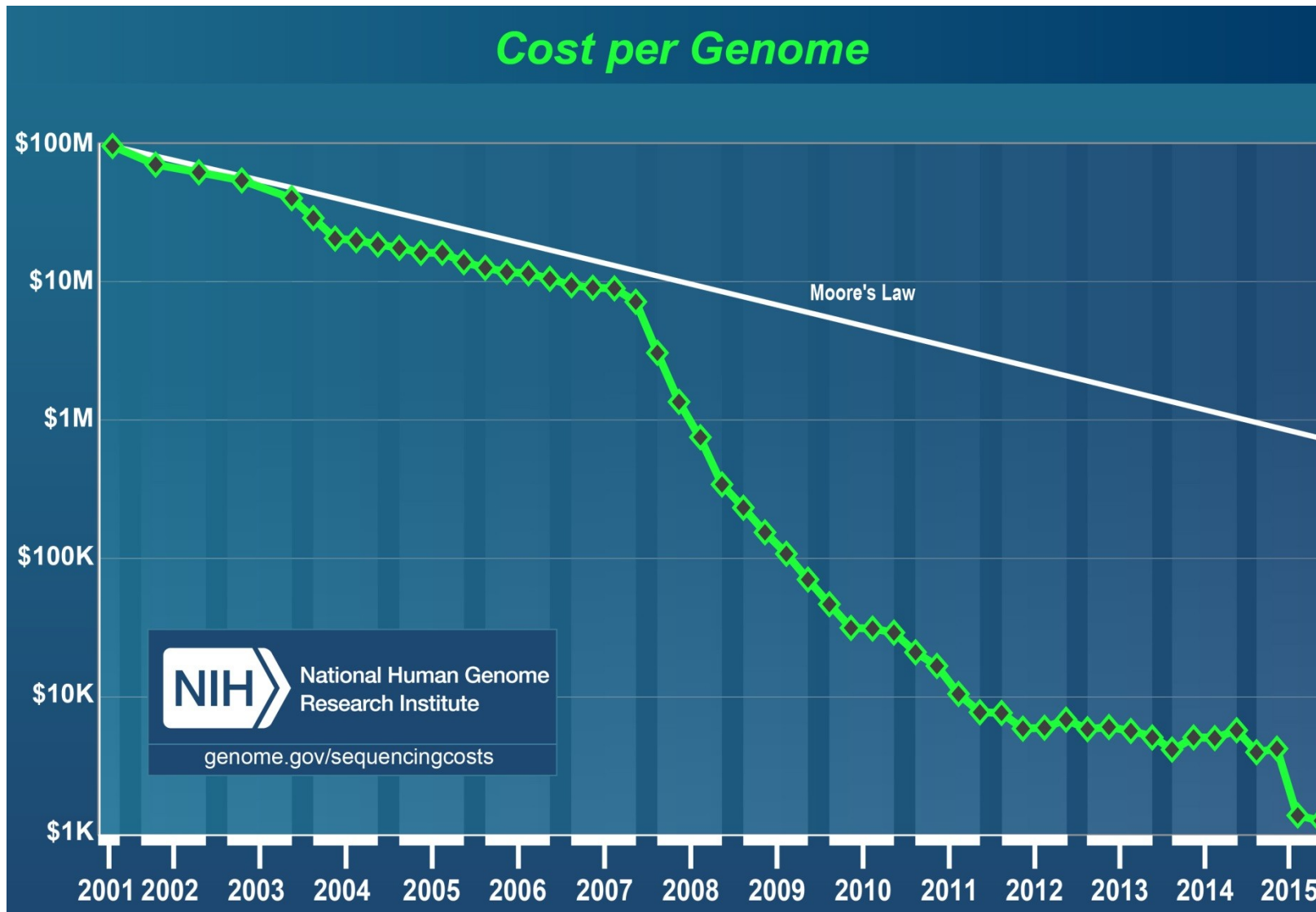
C. Venter



J. D. Watson

Individual sequences of human genomes were published in 2007 and 2008

The 1000 genome project published in 2010



Moore's law (1965): „The number of transistors (hence the processing power) that can be squeezed onto a silicon chip of a given size will double every 18 months”.

Postgenomic era

- **Genomes were described**
- **Ongoing genomes annotations**



Genetics today

from phenotype to genotype

DNA → RNA → protein → phenotype



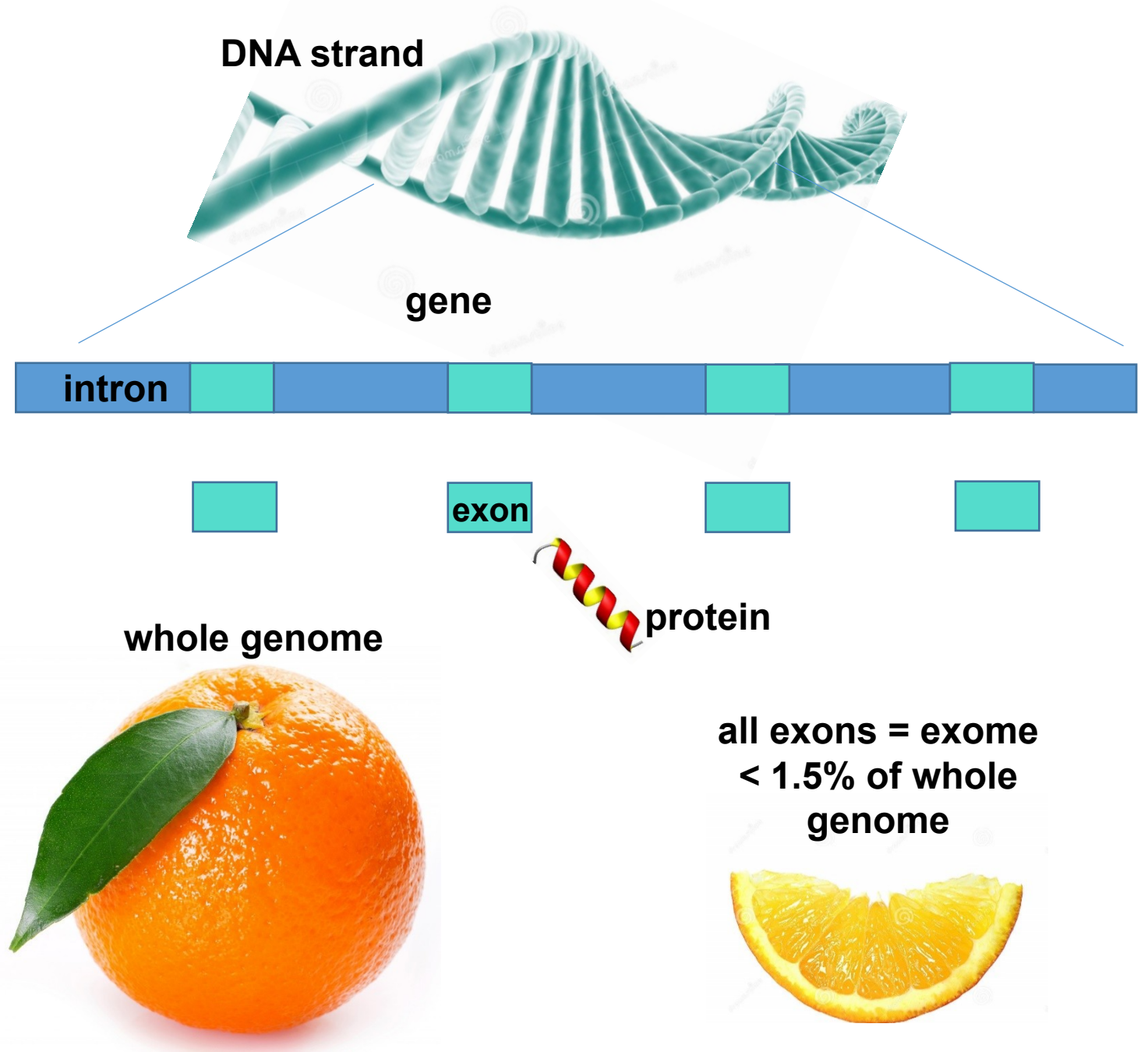
from genotype to phenotype

Modern techniques of genome analysis

Whole-Genome Sequencing vs. Whole-Exome Sequencing

- human genome = 3.2×10^9 bp
~ 20 000 genes

- **Exome** = < 1.5% of human genome
contains ~ 85% of known disease causing mutations



NGS – flexibility

whole genome



3 200 000 000 bp
30 x coverage

exome



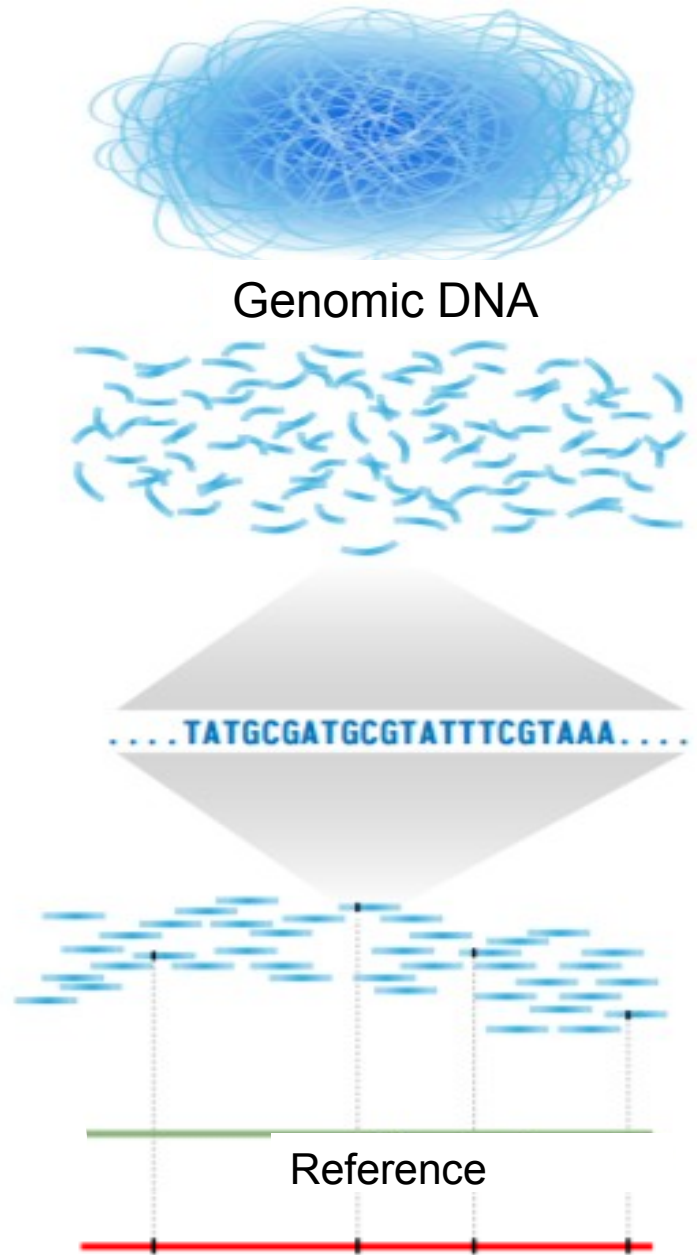
20 000 genes
100 x coverage

**targeted genes
or hotspots**



< 100 genes
≥ 1000 x coverage

Whole-Genome Sequencing



Generating a Person's Genome Sequence

Break genome into small pieces

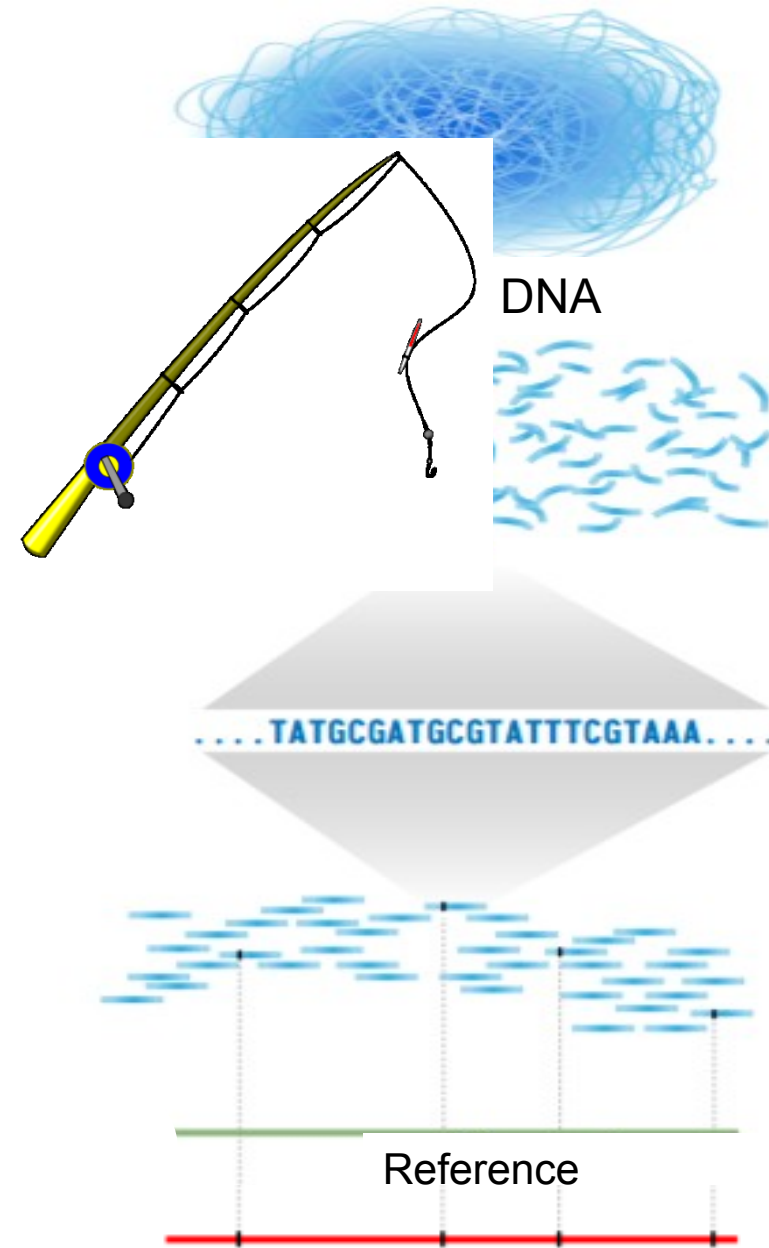
Capture library

Generate millions of sequence reads

Align sequence reads to establish reference sequence

Deduce starting sequence and identity differences from reference sequence

Whole-Exome Sequencing



Mutation vs. human genome variability

Mutation vs. human genome variability

- Every 1000th base could be mutated $\Rightarrow 3.2 \times 10^6$ variants
- One man has approx. 0.5×10^6 variants
- Exome analysis (1.5% of genome) \Rightarrow tens thousands of variants

Which of the found variants is the disease causing one?

Mutation vs. human genome variability

- Every 1000th base could be mutated $\Rightarrow 3.2 \times 10^6$ variants
- One man has approx. 0.5×10^6 variants
- Exome analysis (1.5% of genome) \Rightarrow tens of thousands of variants

**Which of the found variants is the
one?**

mutation frequency
 $1.1 - 1.3 \times 10^{-8}$

Mutation vs. human genome variability

- Every 1000th base could be mutated $\Rightarrow 3.2 \times 10^6$ variants
- One man has approx. 0.5×10^6 variants
- Exome analysis (1.5% of genome) \Rightarrow tens thousands of variants

mutation
x polymorphisms

pathogenic variants is the disease causing

Mutation vs. human genome variability

Mutations: **spontaneous vs. induced**

gene vs. chromosomal

Mutations: **missense**
nonsense (terminating triplet)
same sense
frameshift

Mutation vs. human genome variability

Single nucleotide polymorphisms (SNPs)

cgcgcggcctcctccttggtg^ccatcctggtcctcctaaaccacctggac

cgcgcggcctcctccttggtg^tcatcctggtcctcctaaaccacctggac

Insertions/deletions (indels)

cgcgcggcctcctccttggtggccatcctggtcctcctaaaccacctggac

cgcgcggcctcctccttggtg-----ctggtcctcctaaaccacctggac

Mutation vs. human genome variability

Microsatellites (STR)

cgcgcggcctcctccttggtgg**cacacacaca**catcctggtcctcctaaaccacctgga

cgcgcggcctcctccttggtgg**cacacacaca**catcctggtcctcctaaaccacctgga

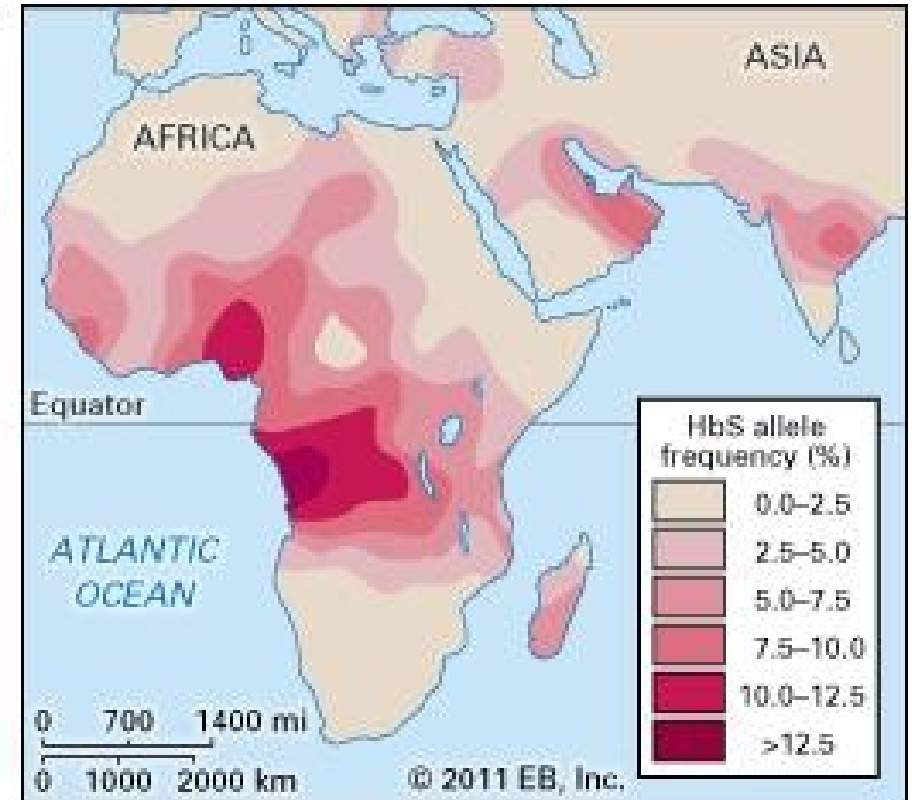
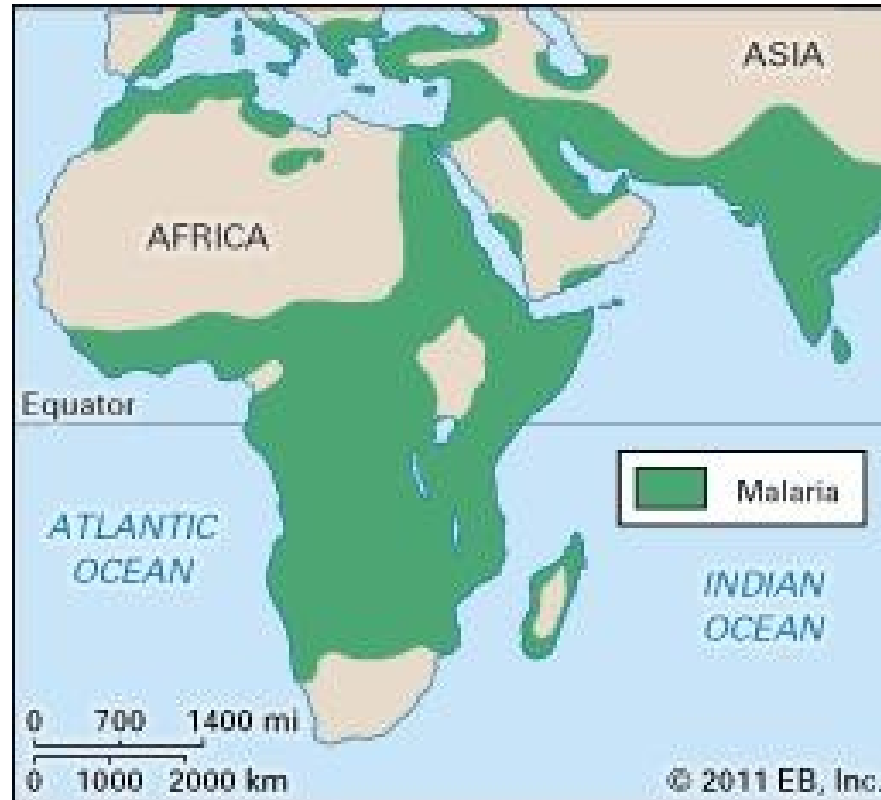
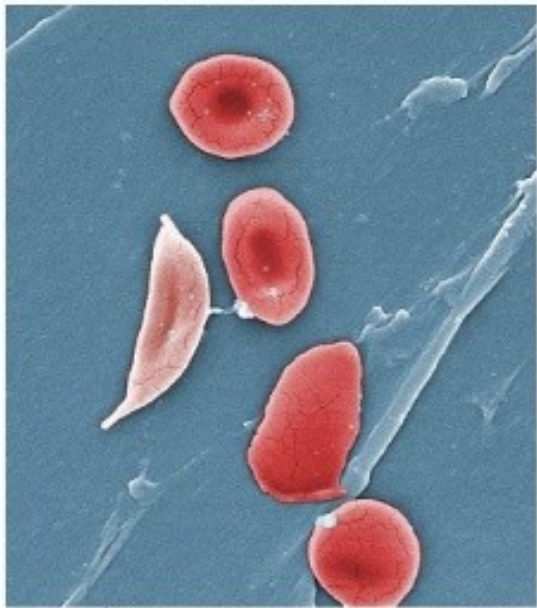
Copy number variants (CNV)

>1 kb – 1Mgb

Mutation vs. human genome variability



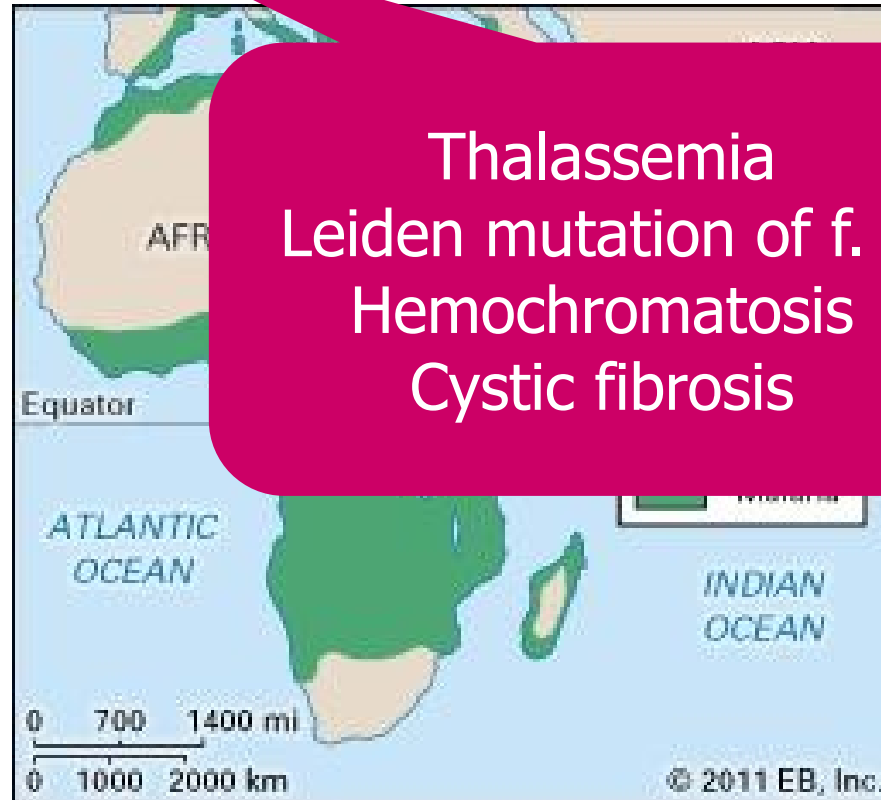
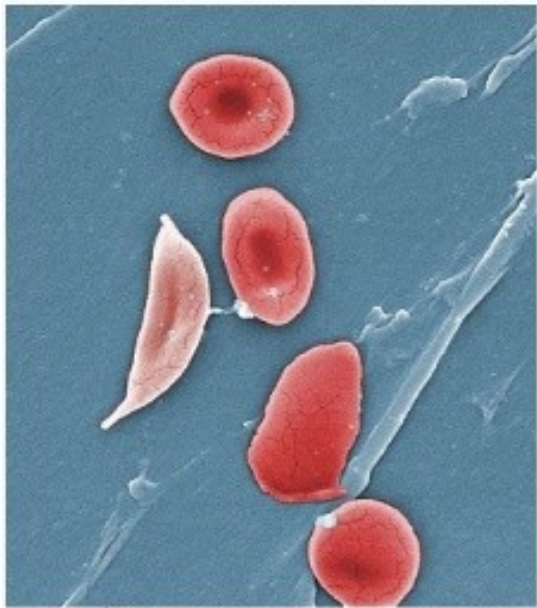
Sickle-cell anemia



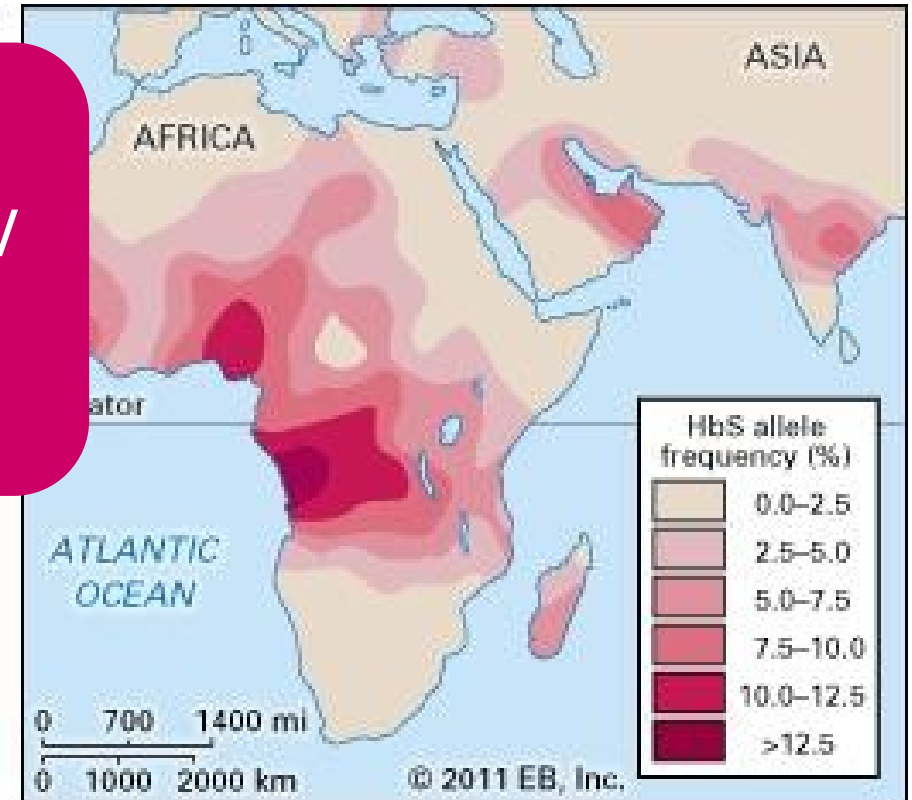
Mutation vs. human genome variability



Sickle-cell anemia



Thalassemia
Leiden mutation of f. V
Hemochromatosis
Cystic fibrosis



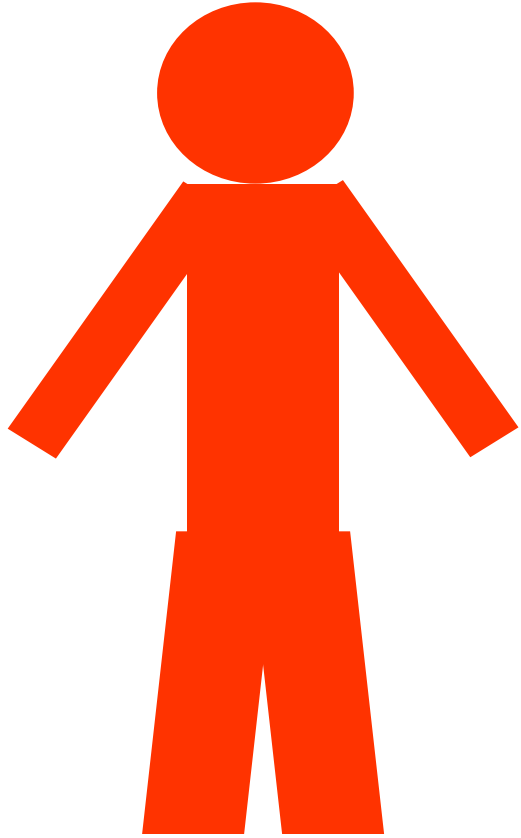
Positive mutations



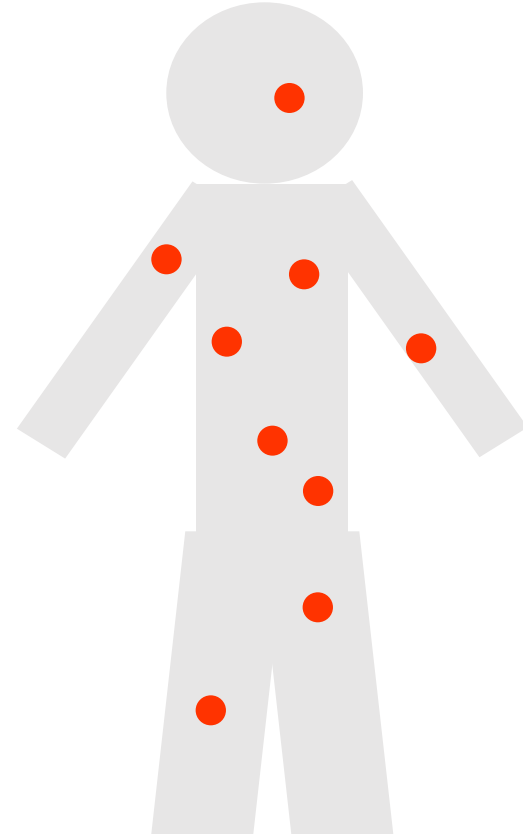
mutation

Germinal vs. somatic mutations

Germinal vs. somatic mutation



Germinal mutation



Somatic mutation

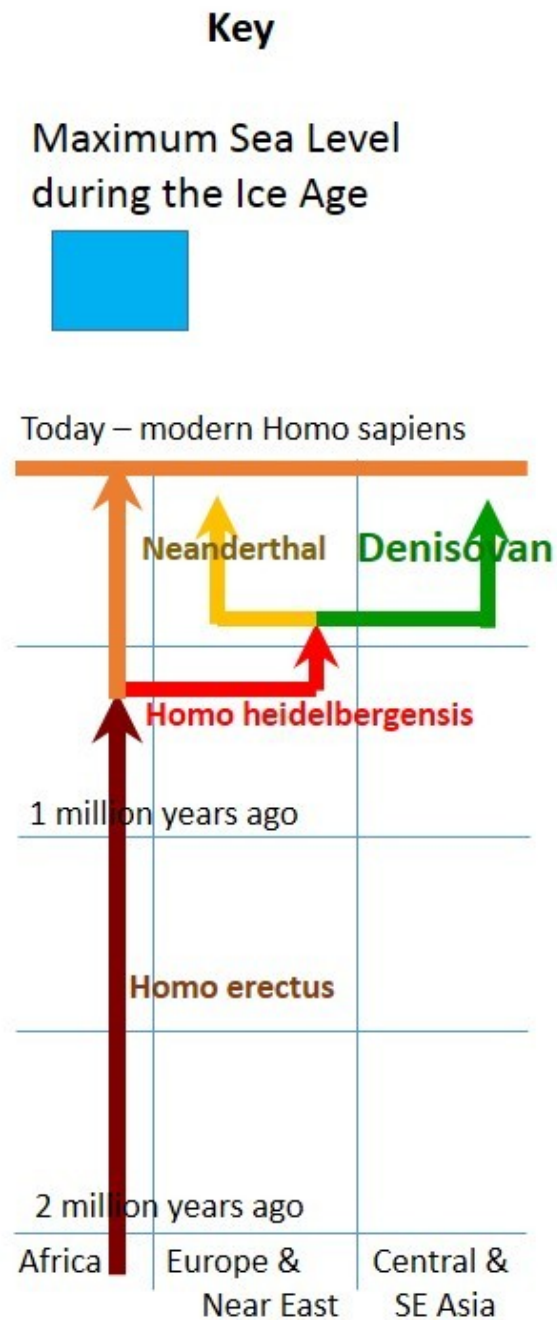
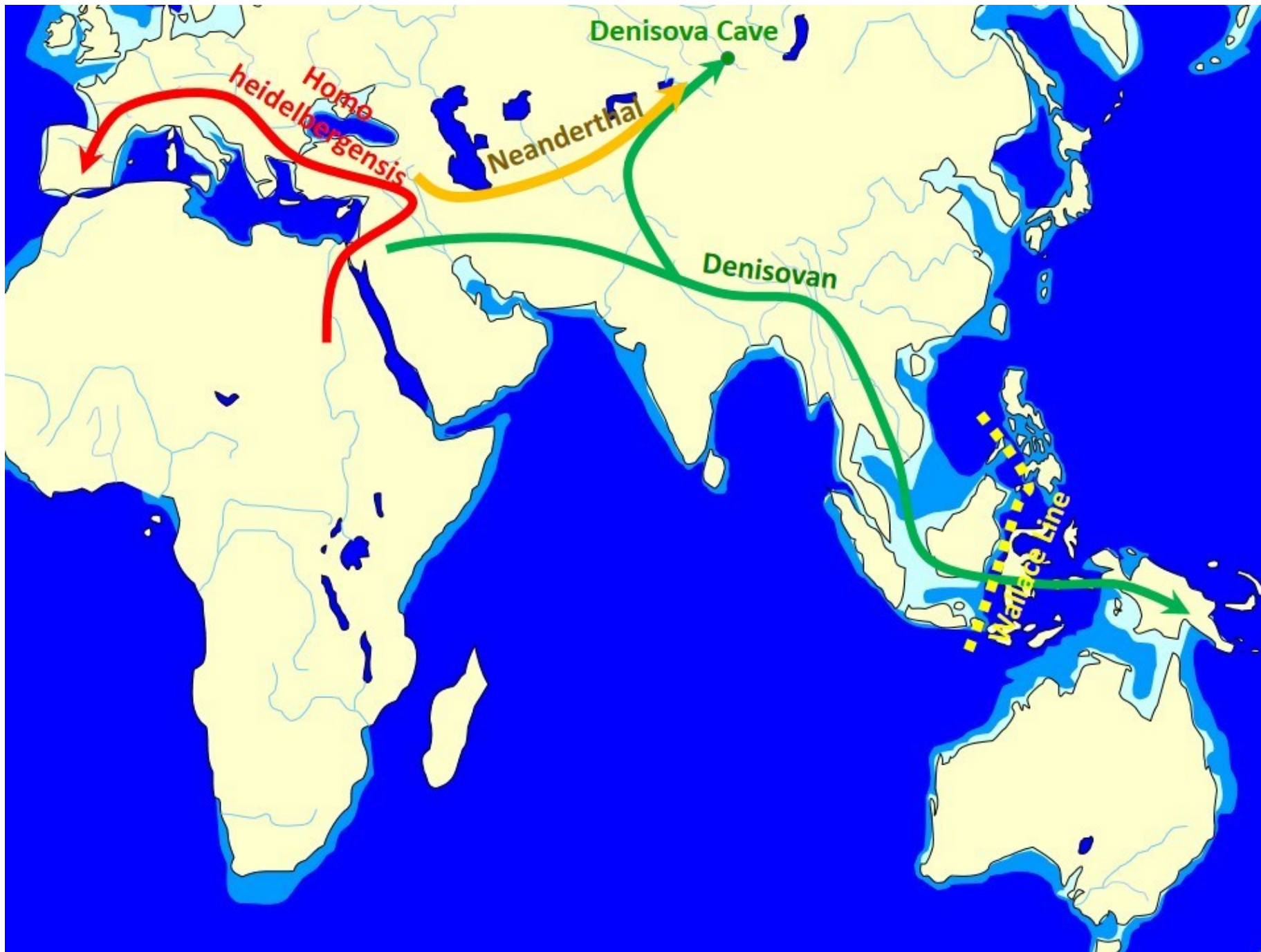
Are we Homo sapiens?

Are we Homo sapiens?

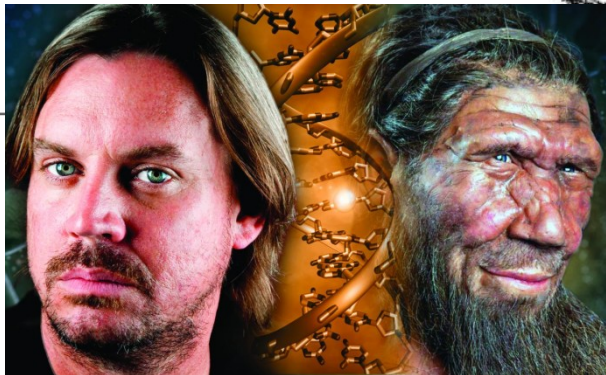


Denisova hominins, 41 000 years ago
mtDNA

*Reich a kol.
Nature 2010*



Are we Homo sapiens?



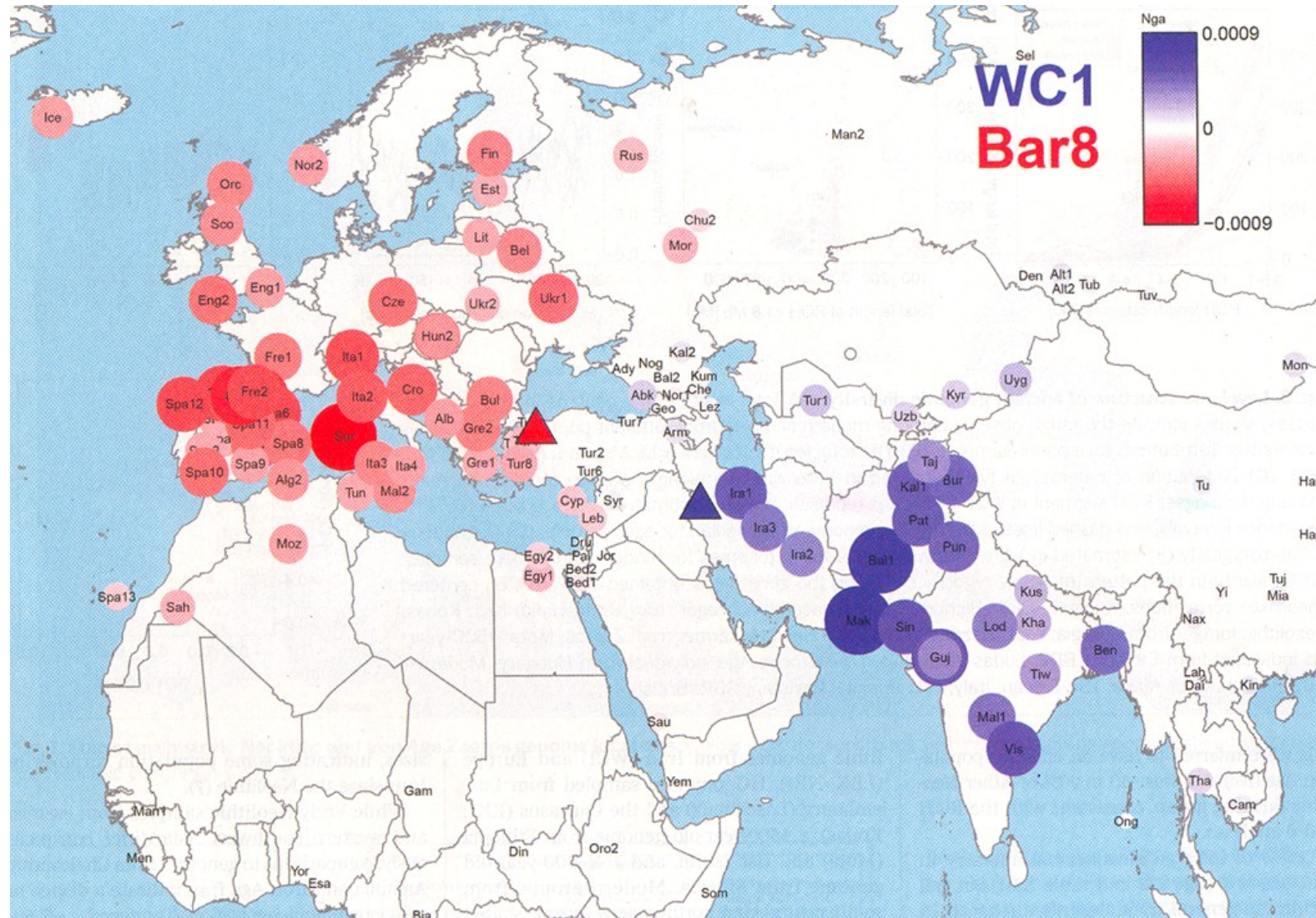
Europe 1 – 3% of the genome
Some diseases of Neanderthal origin – depression,
skin disorders?

Ancient genomes analysis

Genome from the Younger Stone Age and Iron Age - Zagros, Iran



Two ancient genomes in modern humans



Broushaki a kol., Science 2017

Famous ancient genomes

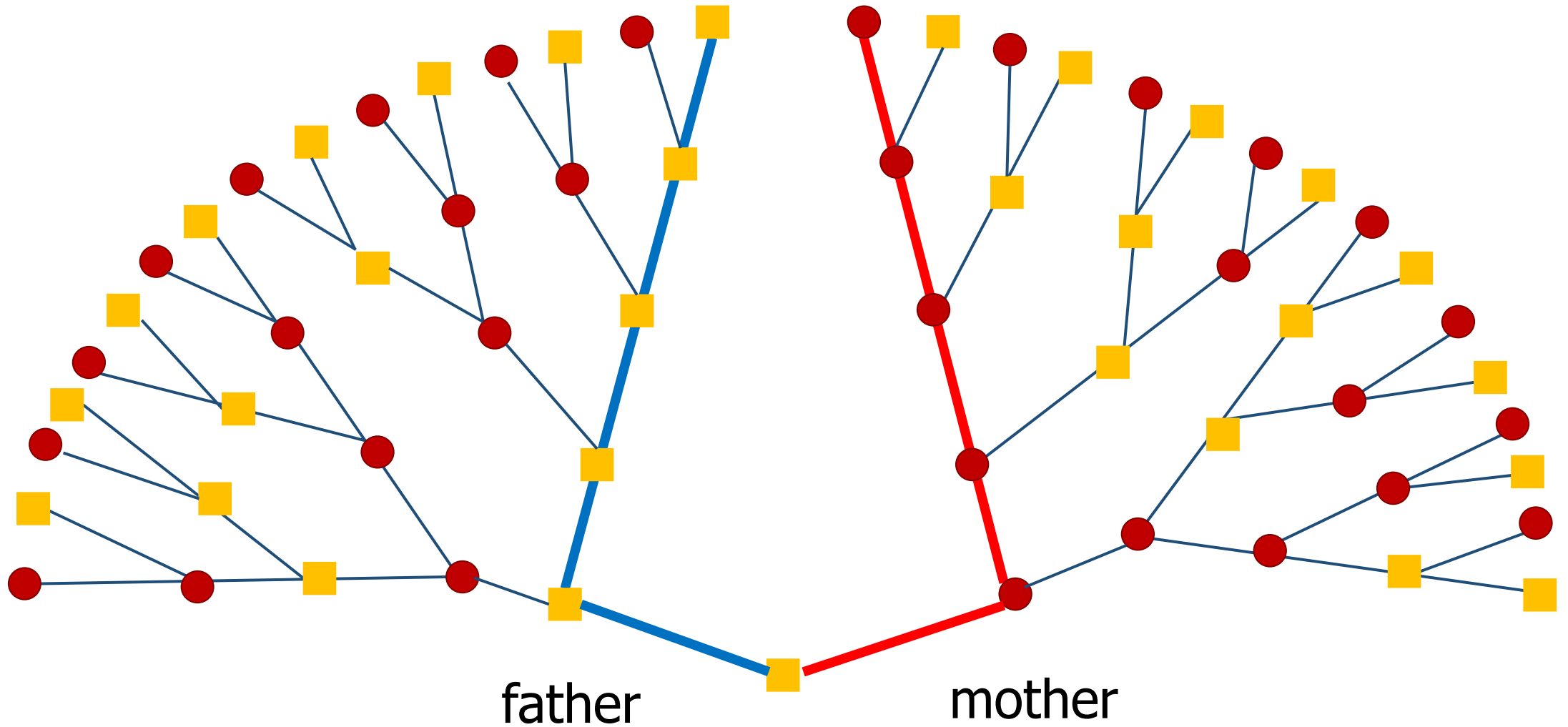
Ötzi

Cheddar man

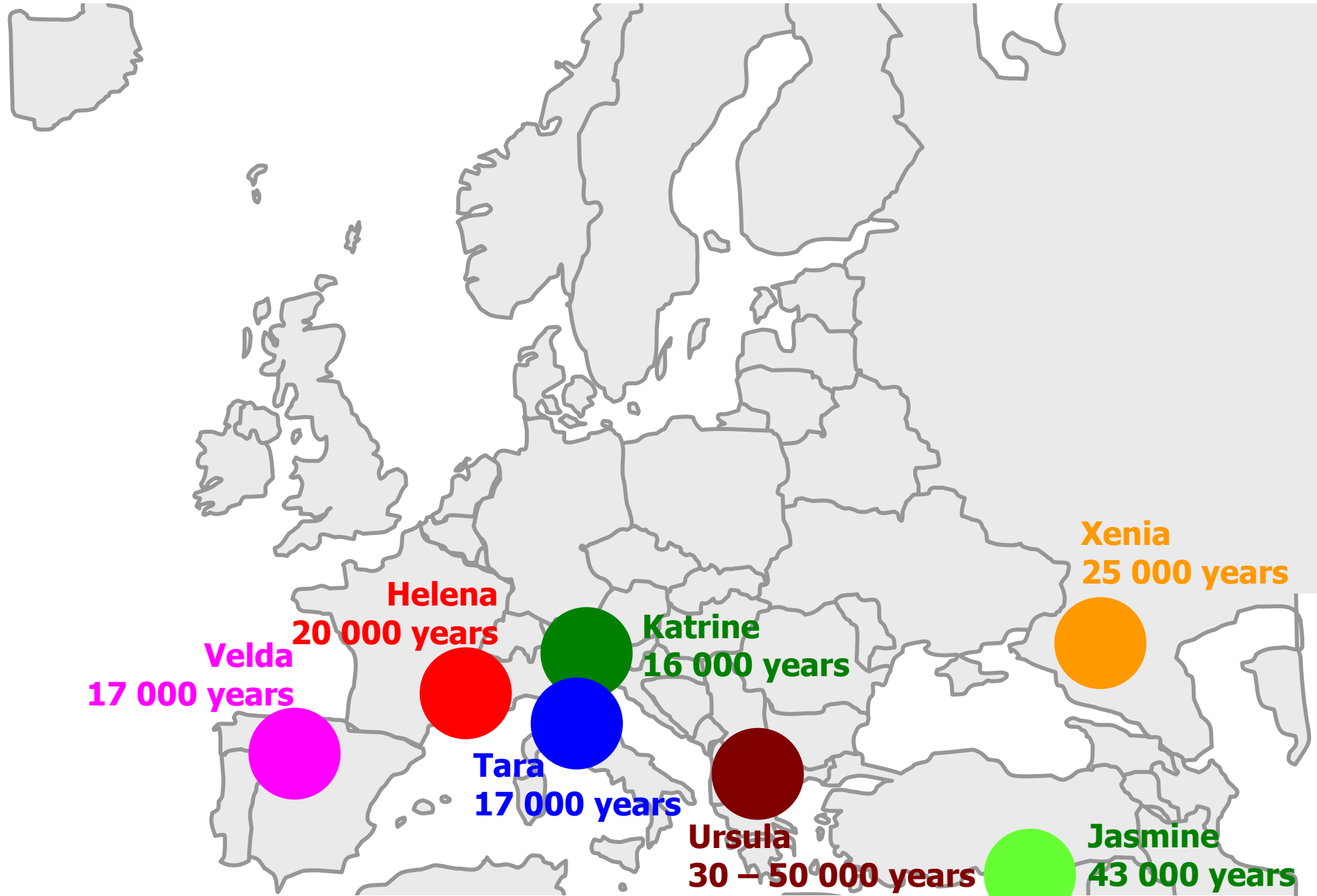


Origin based on mitochondrial DNA

Mitochondrial and Y-inheritance



The seven daughters of Eve



The seven daughters of Eve

- **mitochondrial Eve (140 000 years ago in Ethiopia)**
- **7 main mitochondrial haplotypes in Europe**
- **29 haplotypes worldwide**

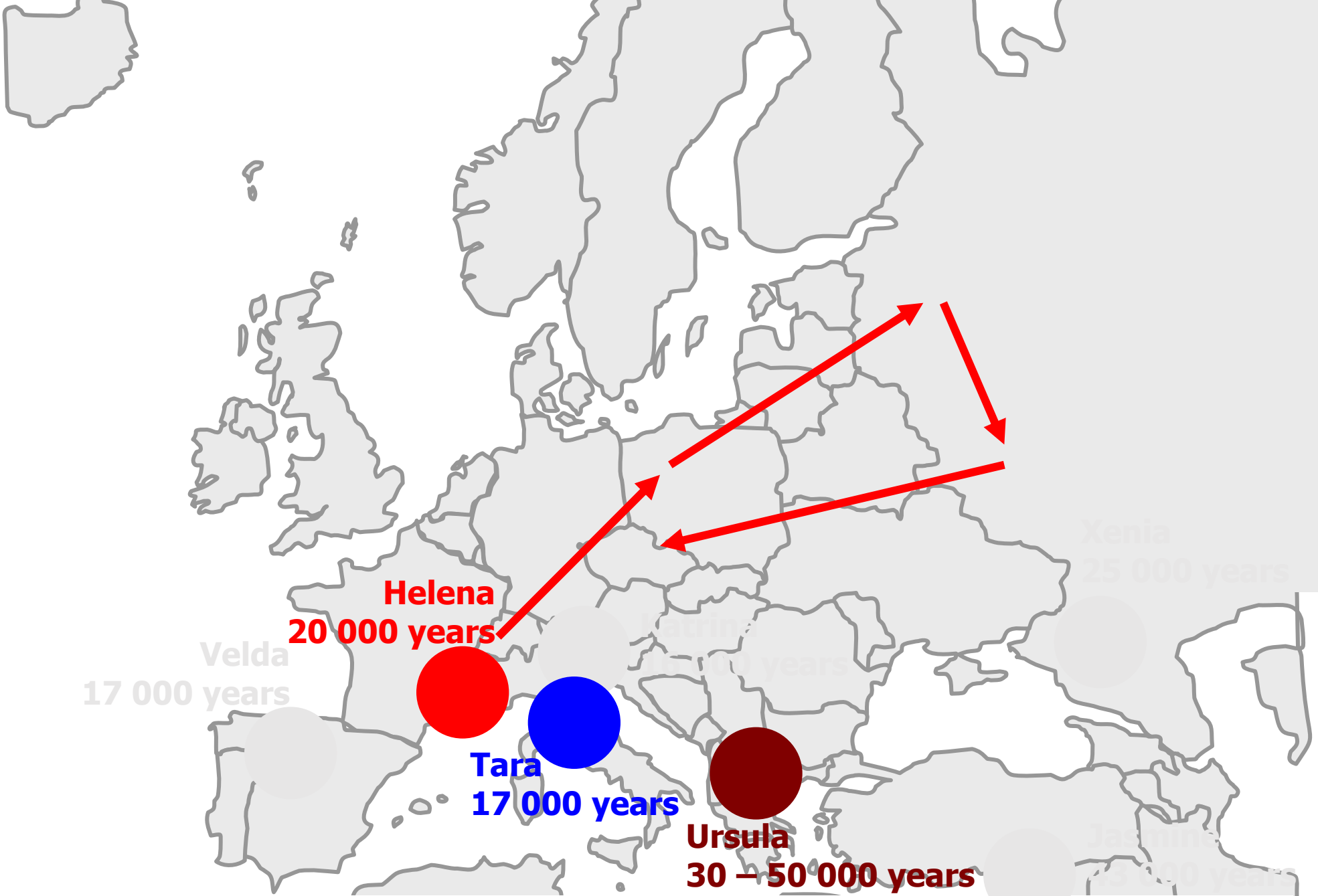
- **Results are not as accurate as from other methods – mtDNA is very similar**

offspring of the mitochondrial Eve may still live

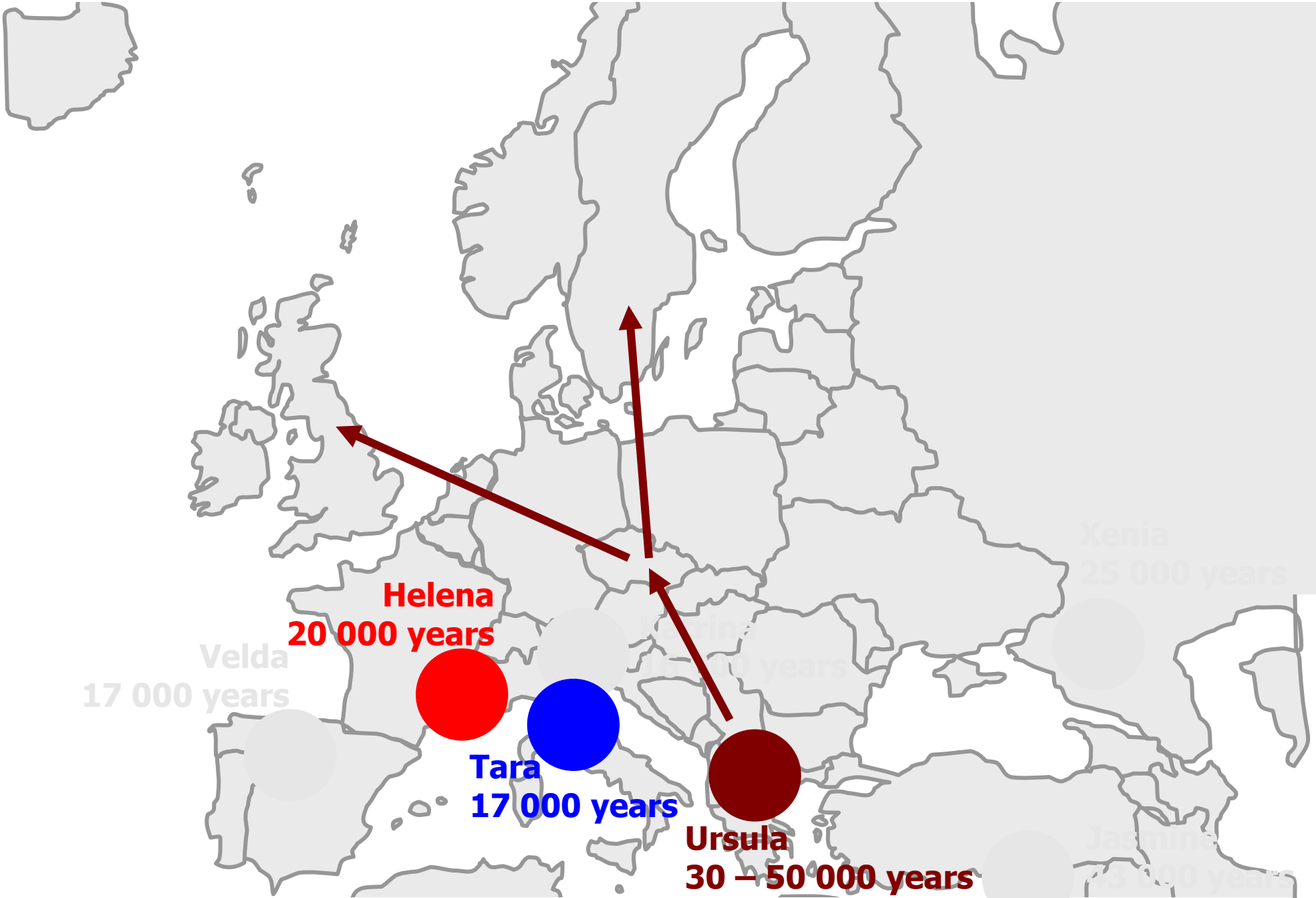
The seven daughters of Eve in the Czech population

- **Helena 43.51%** (dominant lineage from Poland and european part of Russia)
- **Ursula 17.6%** (mainly UK and Scandinavia)
- **Tara 11.17%**
- **Jasmine 8.78%**
- **Katrine 5.89%** (Ashkenazi Jews)
- **Velda 4%**
- **Xenia 3%**

The seven daughters of Eve after the Ice Age



The seven daughters of Eve after the Ice Age



The role of genome in the disease onset

- **Mendelian hereditary diseases 8%**
- **Multifactorial 90%**
- **Others 2%**

The role of genome in the disease onset

➤ **Mendelian hereditary diseases 8%**

➤ **Multifactorial 90%**

➤ **Others 2%**

⇒ **genetic background plays almost always a role in the disease onset**

Inheritance types

Inheritance types

Mendelian

monogenic: one gene \Rightarrow one feature

X-linked and Y-linked (sex-linked disorders)

Polygenic

several genes \Rightarrow one feature

Mitochondrial

Environmental factors

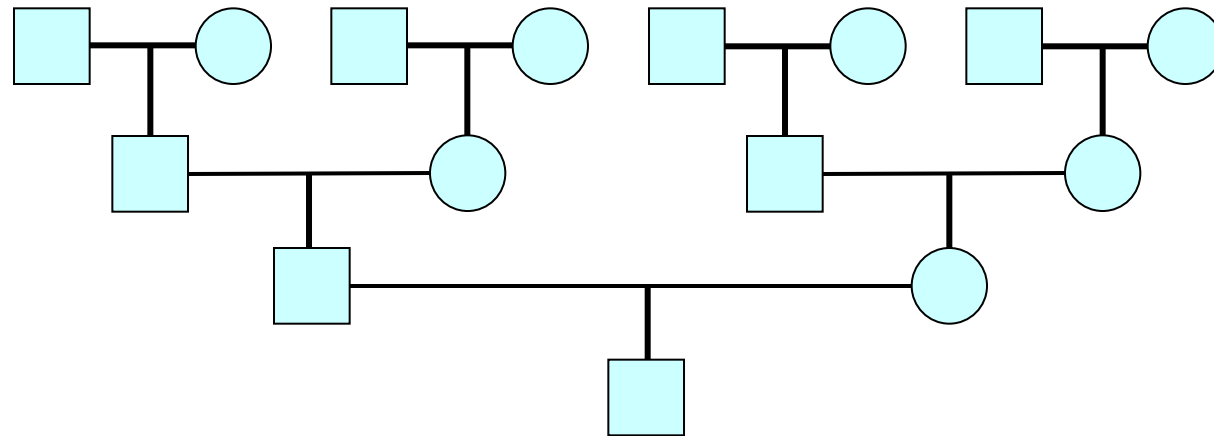
What is the procedure of hereditary diseases tracing?

- **family studies:**
 - pedigree
 - monozygous twins
 - odds ratio
 - relative risk
- **disease frequency in population**
- **molecular biology methods**
- **genetic linkage and functional tests**



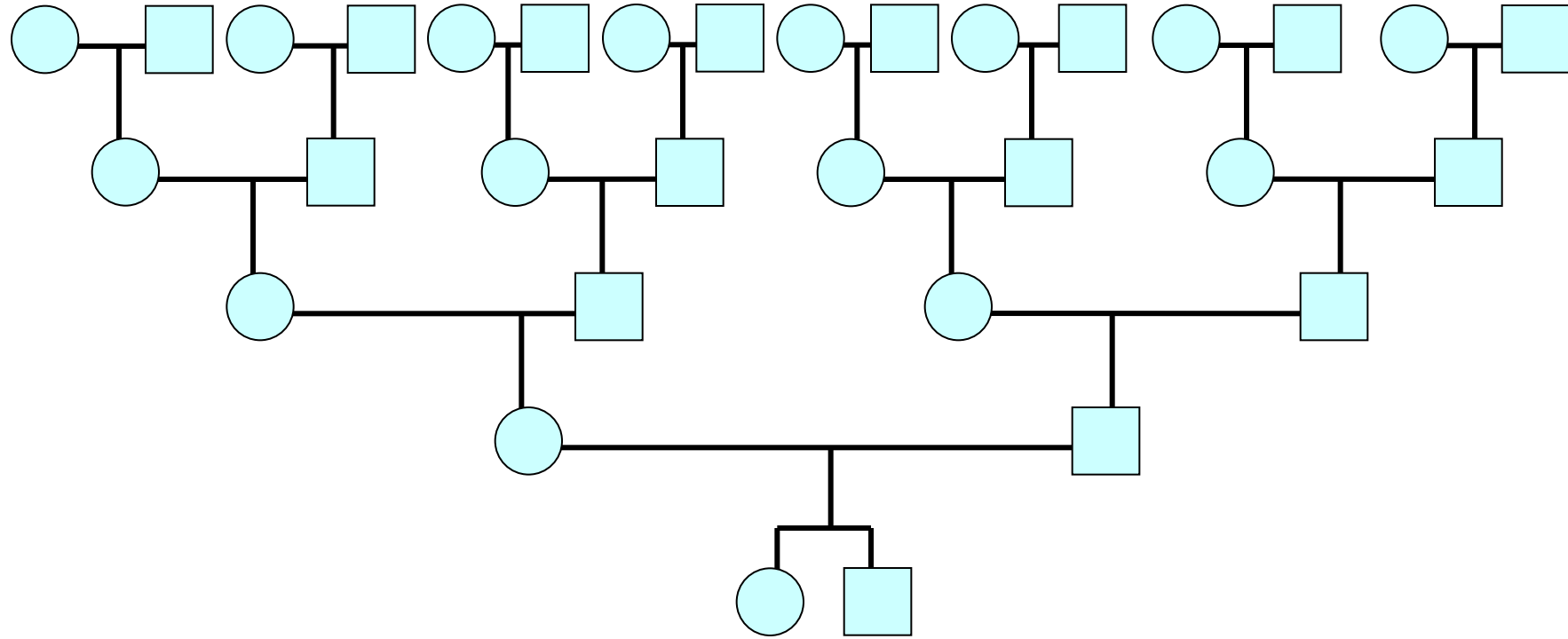
Common ancestor

- two common ancestors in previous generation: parents
- 4 grandparents, 8 great-grandparents
- the number of ancestors in generation n is 2^n

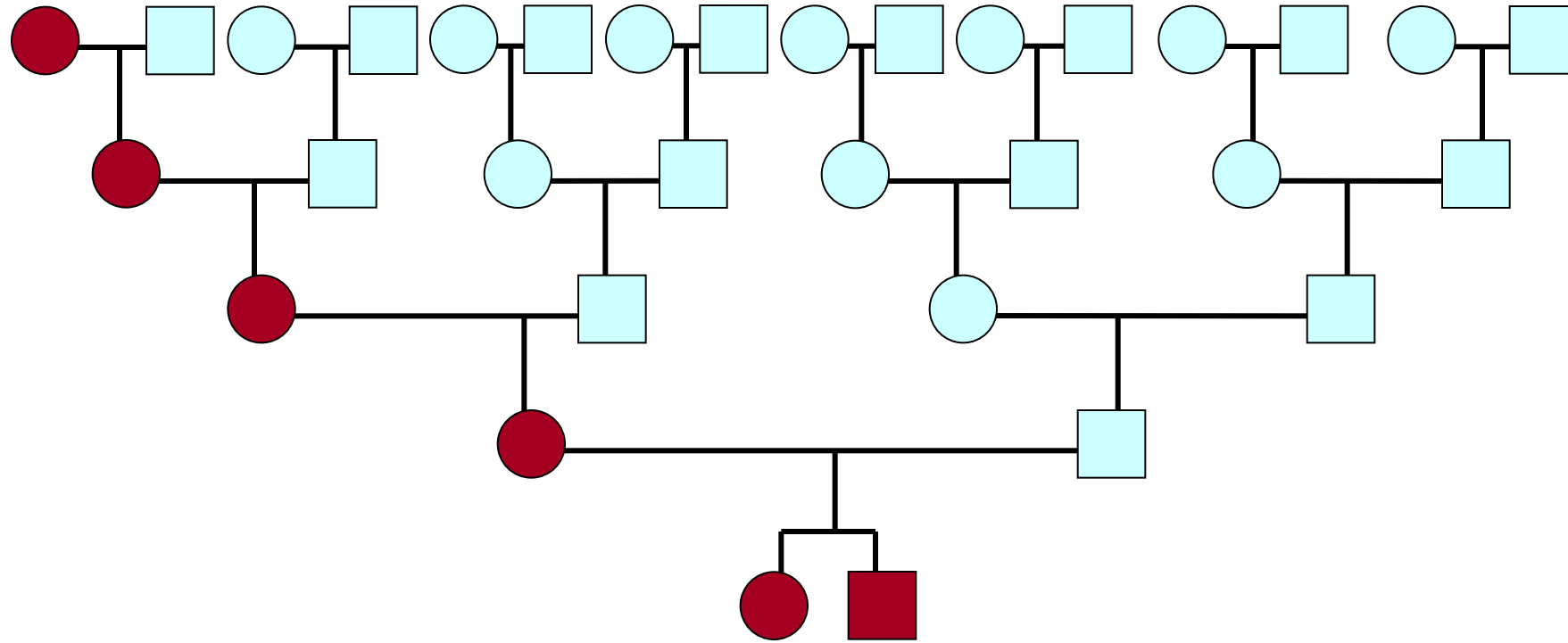


- 40th generation (1000 years back): $2^{40} = 1.09 \times 10^{12}$
- so many people didn't live on this planet (7.0×10^9)

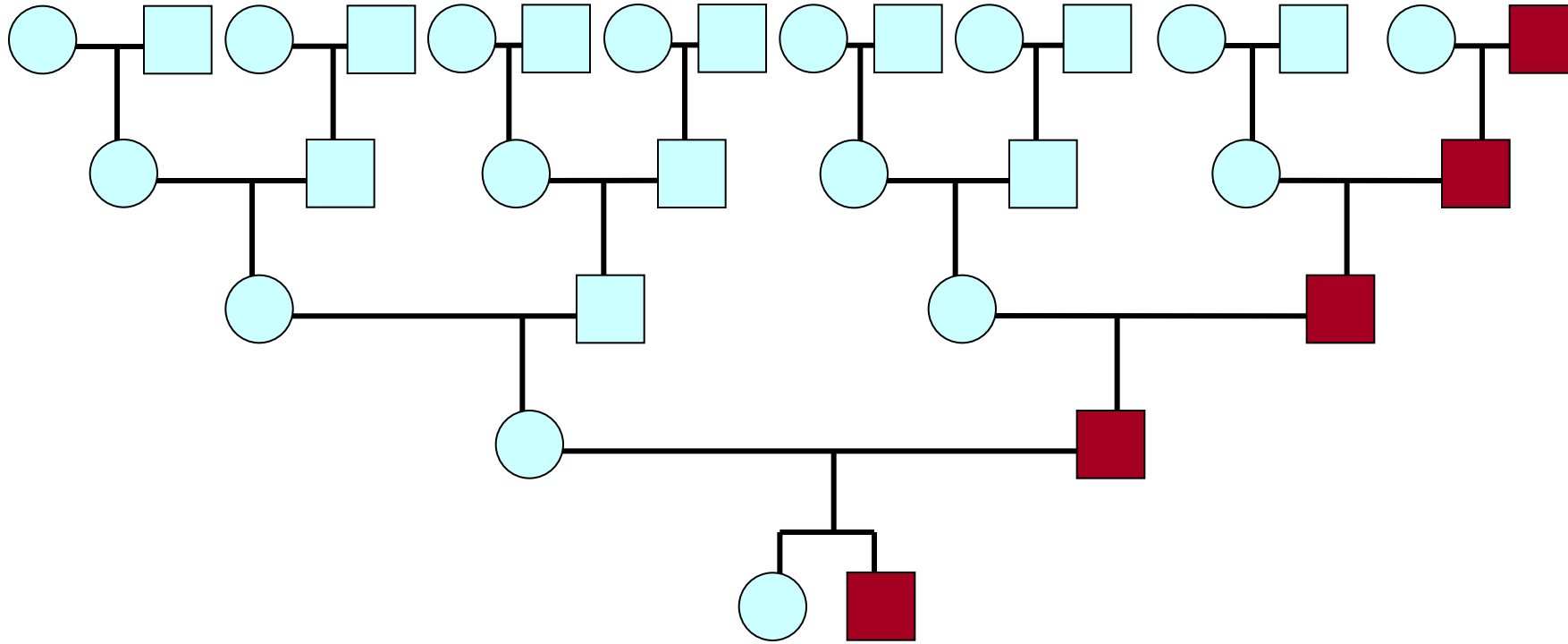
Pedigree



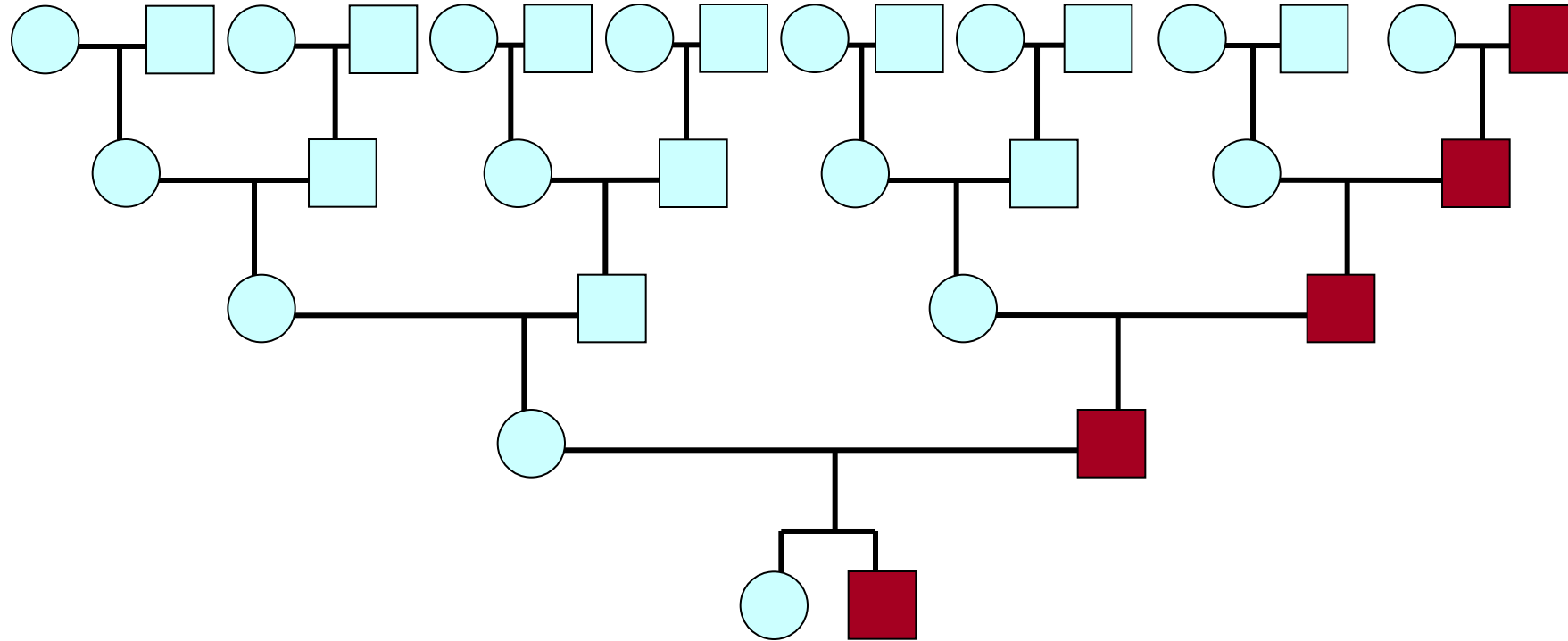
Mitochondrial inheritance



Y-chromosome inheritance

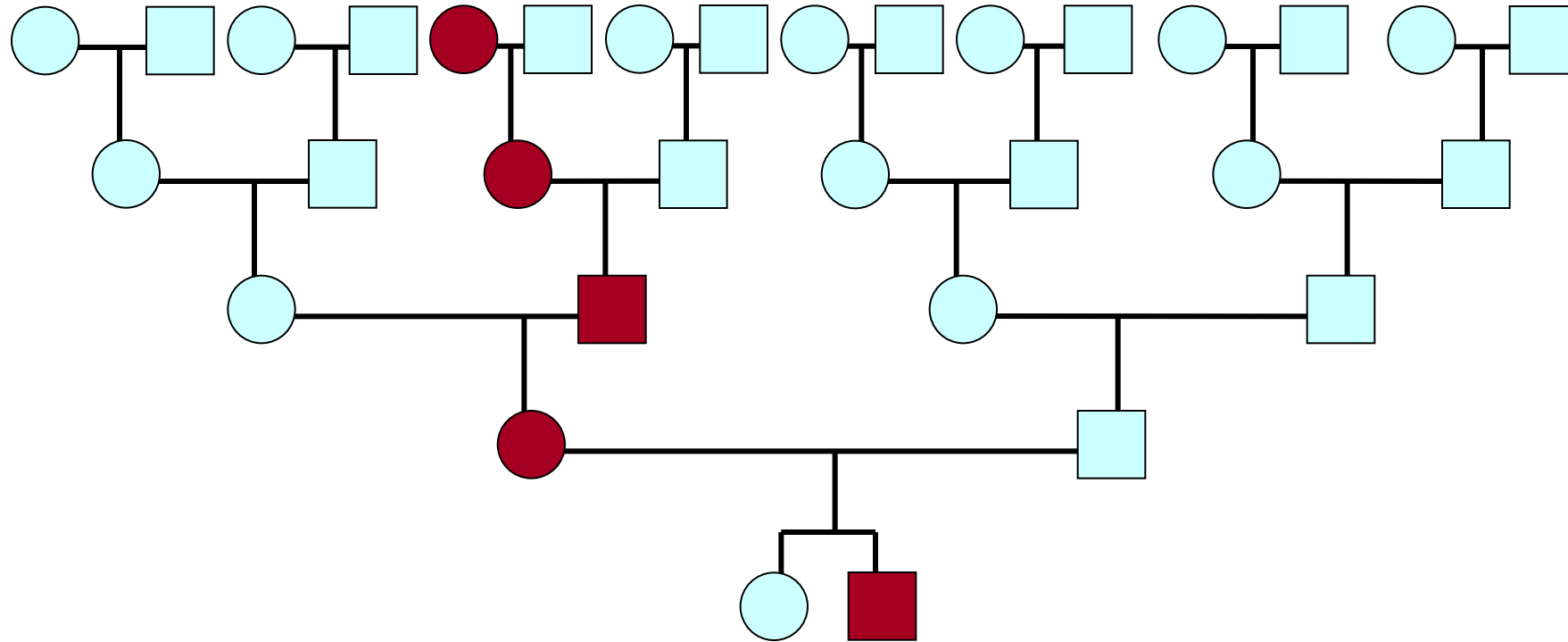


Y-chromosome inheritance

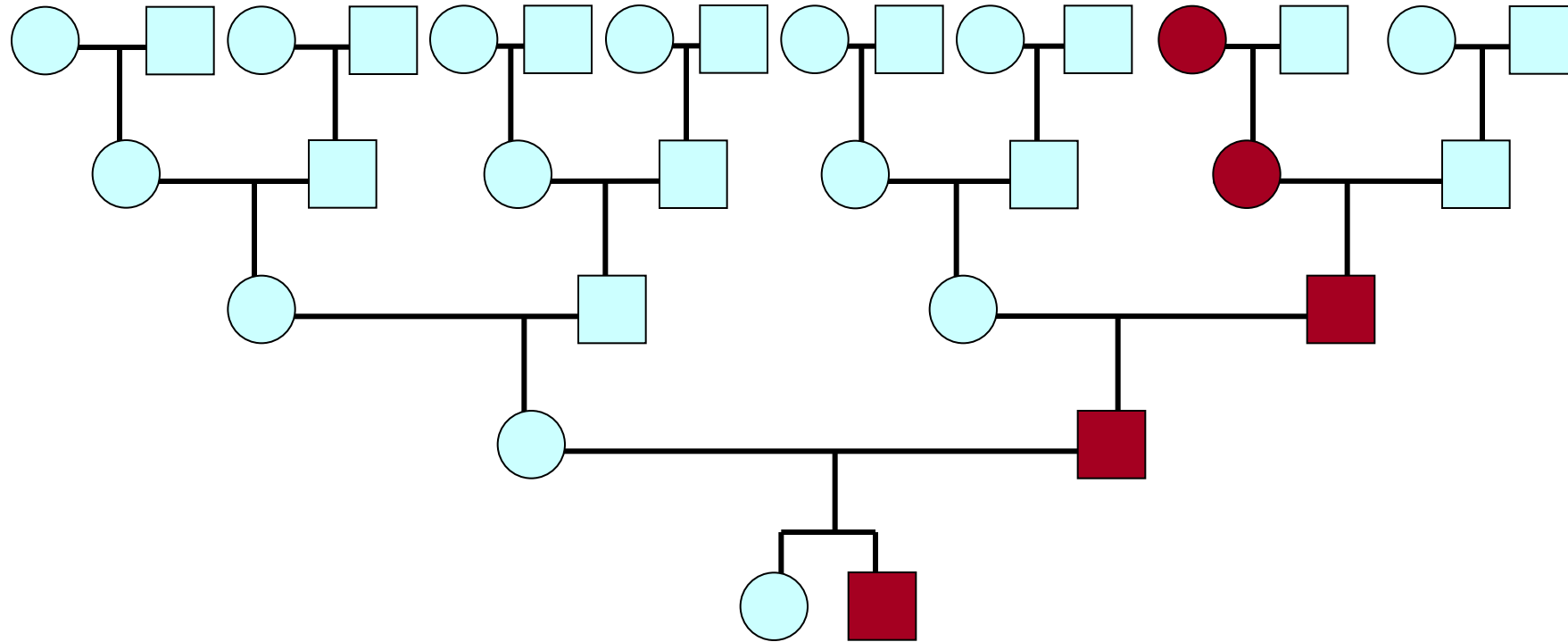


What Y-chromosome carries on?

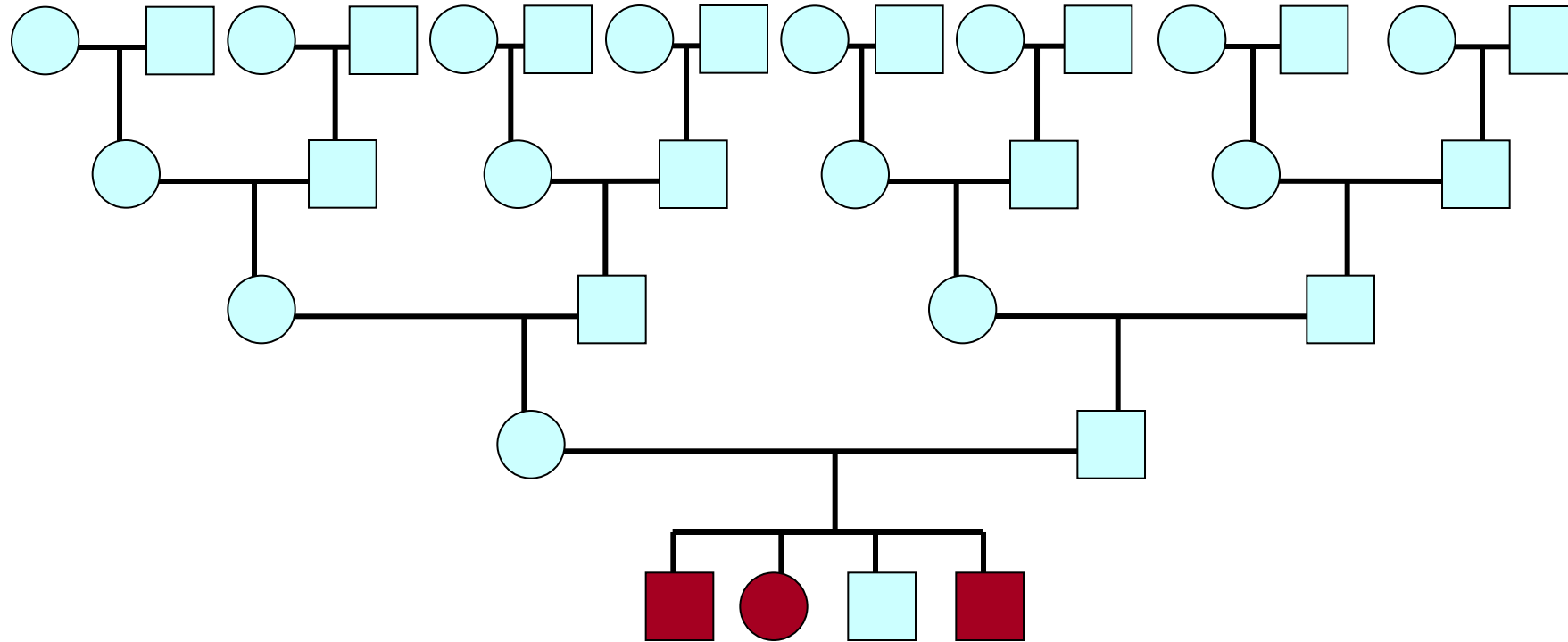
Autosomal dominant inheritance



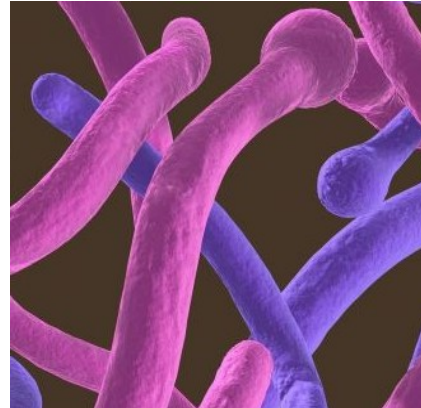
Autosomal dominant inheritance



Autosomal recessive inheritance



Environmental factors



Monogenic disorders

How many monogenic disorders exist?

≈ 1 000

≈ 10 000

≈ 100 000

How many monogenic disorders exist?

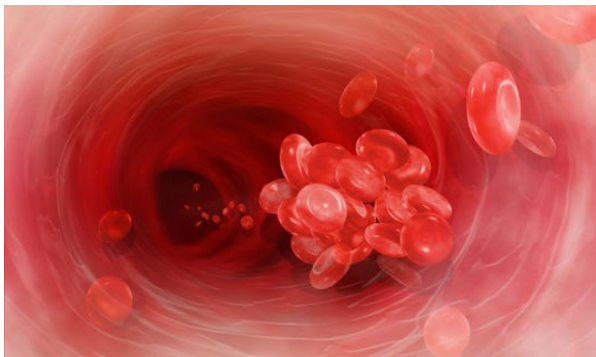
≈ 1 000

≈ 10 000

≈ 100 000

Recessive disorders

- **hemochromatosis (1:10)**
- **mutation of factor V Leiden (1:20)**
- **cystic fibrosis (1:25)**
- **spinal muscular atrophy (1:40)**



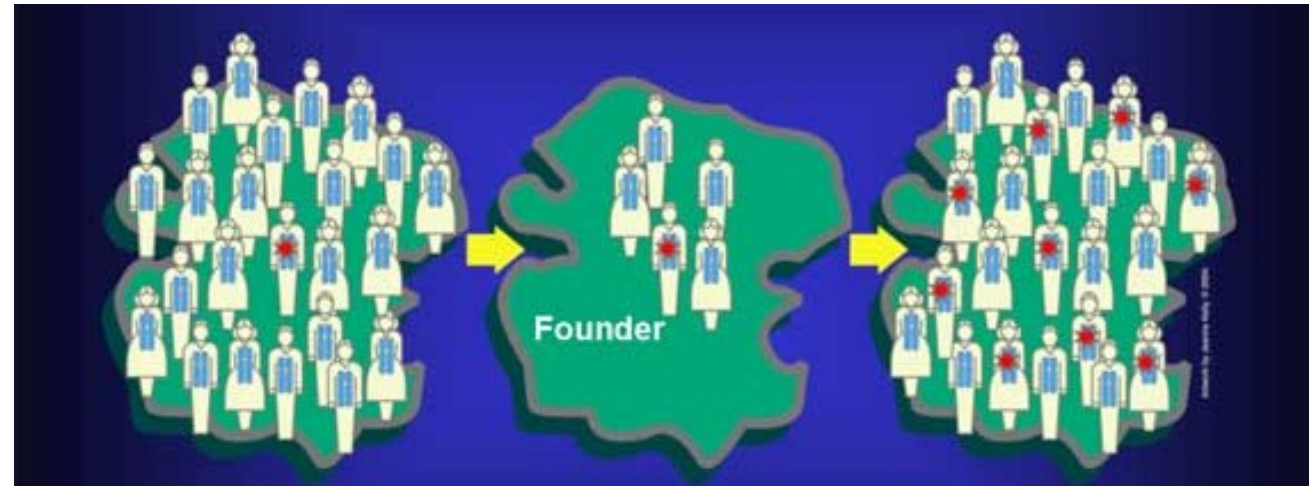
Dominant disorders

- **deafness**
- **polydactyly**
- **Huntington's Chorea**
- **Li-Fraumeni syndrome**
- **breast and ovarian cancer**

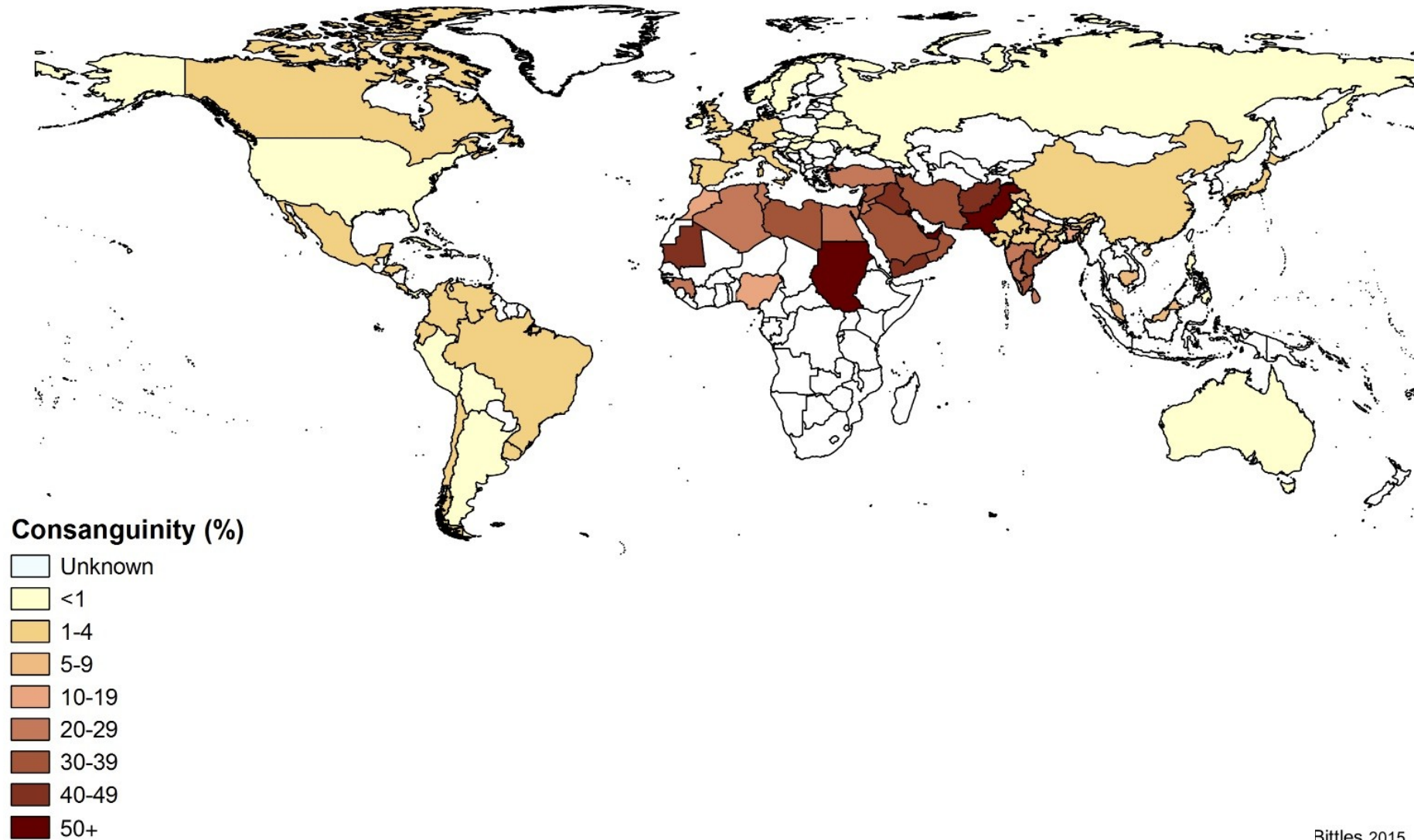


Origin by mutation type

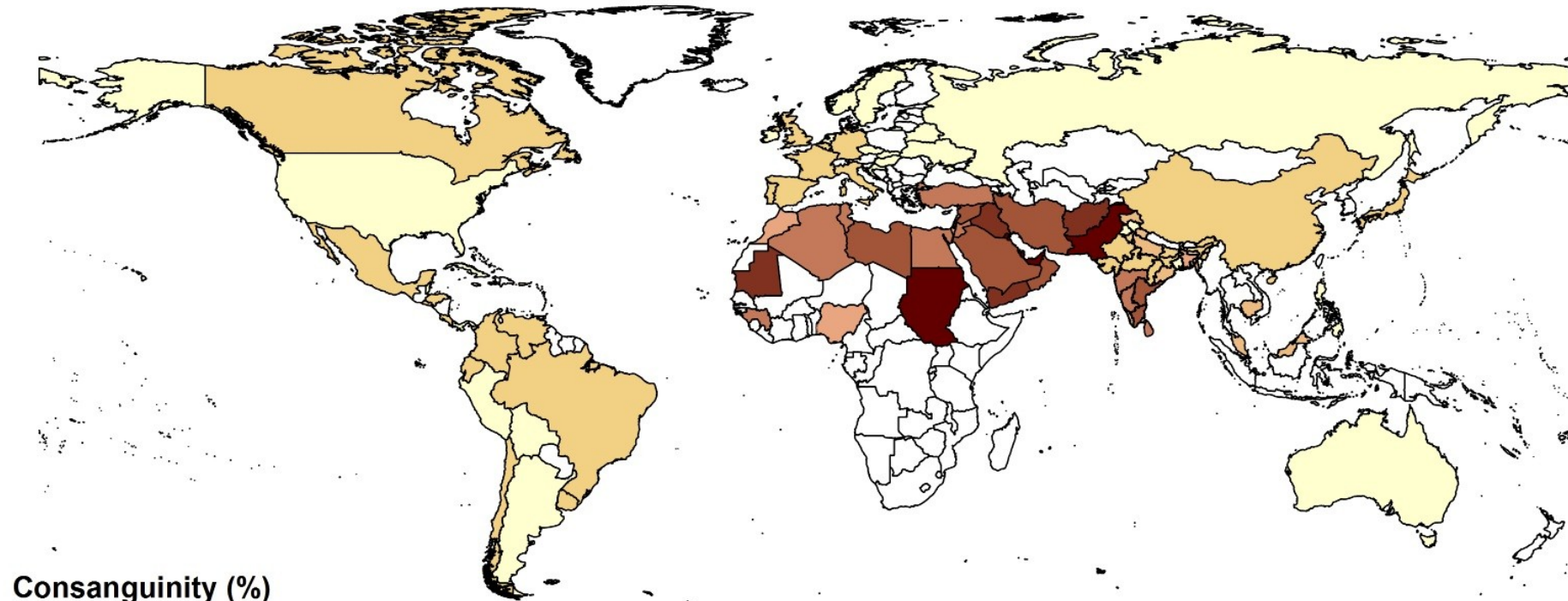
- **Founder effect**
- **Small closed populations:**
 - Ashkenazi Jews
 - franco-Canadiens
 - Iceland
 - surroundings of Maracaibo lake...
- **Marriages of relatives**



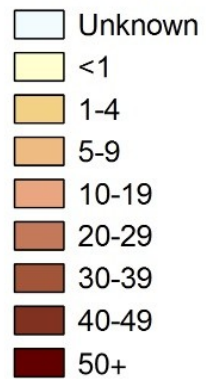
Consanguinity map



Consanguinity map

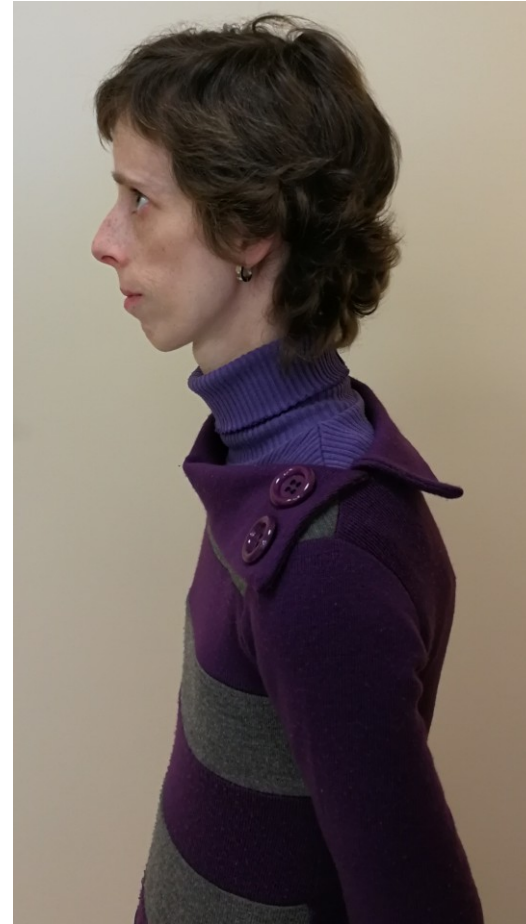


Consanguinity (%)



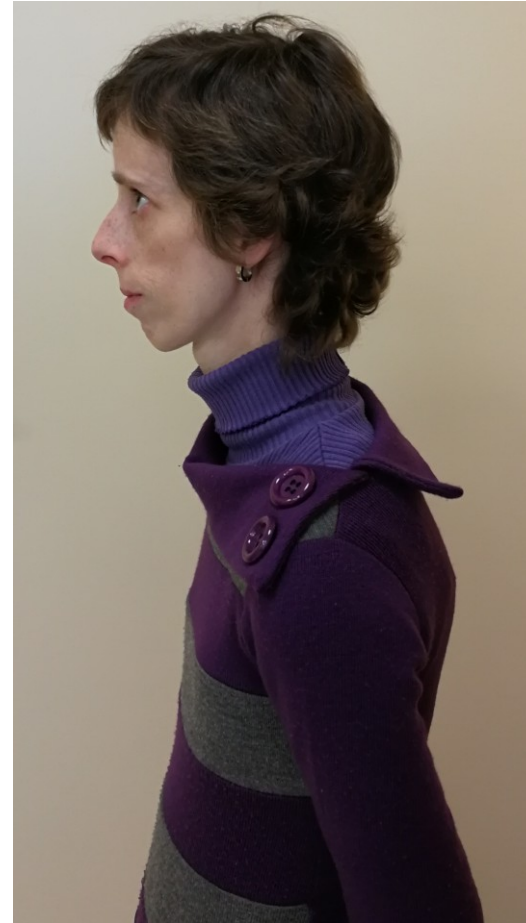
Mandatory genetic testing of partners:
Bahrain
Saudi Arabia
(Iceland)

Consanguinity example



Consanguinity example

Homozygous mutation
BLM gene
c.1642C>T, p.(Gln548*)



Genetic diseases in Czech population

Nijmegen breakage syndrome = Seeman syndrome (NBS)

NBN gene for nibrin in 8q21
Heterozygotes 1:130-150
Common ancestor

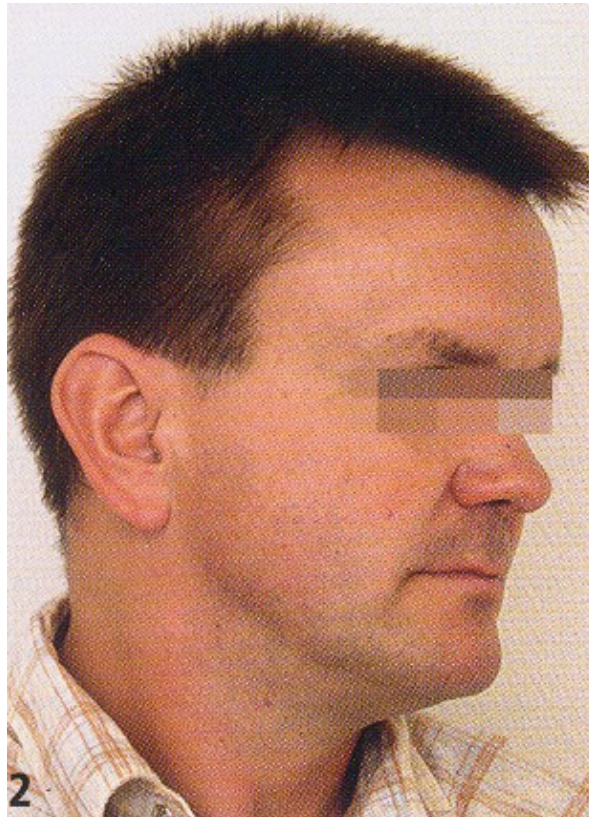


*Seemanová,
1985*

Czech dysplasia

COL2A1 gene

absence of ocular and orofacial anomalies
shortening of third and/or fourth toes

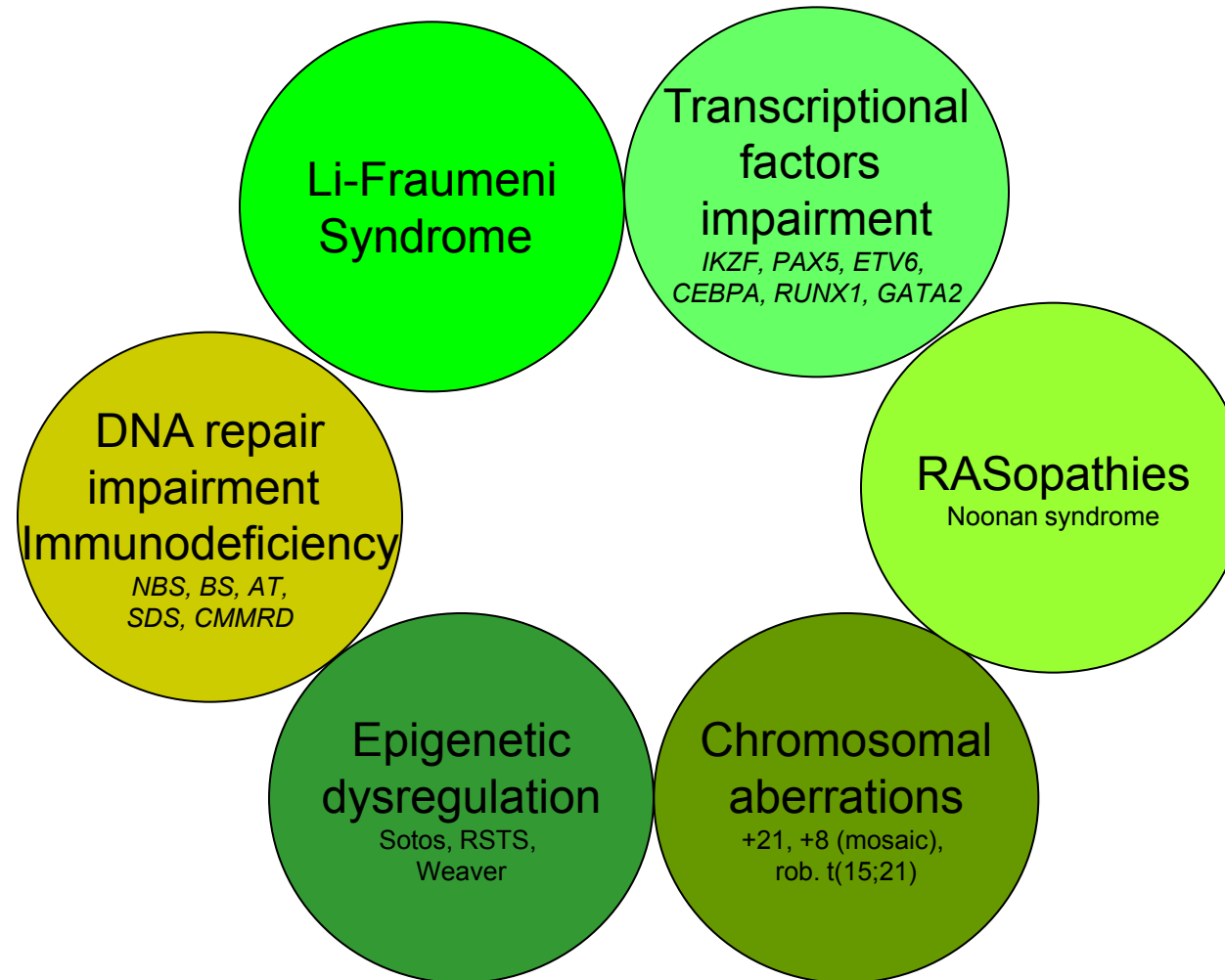


How do we search for new genetic disorders?

Molecular biology methods

- **Analysis of known disease-associated genes**
- **Comparing genetic information of healthy and affected family members**
- **Looking for new variants and “new” genes**
- **Verification by functional tests**

DNA variants and disorders



➤ **genes together with non-coding regions**



Twins



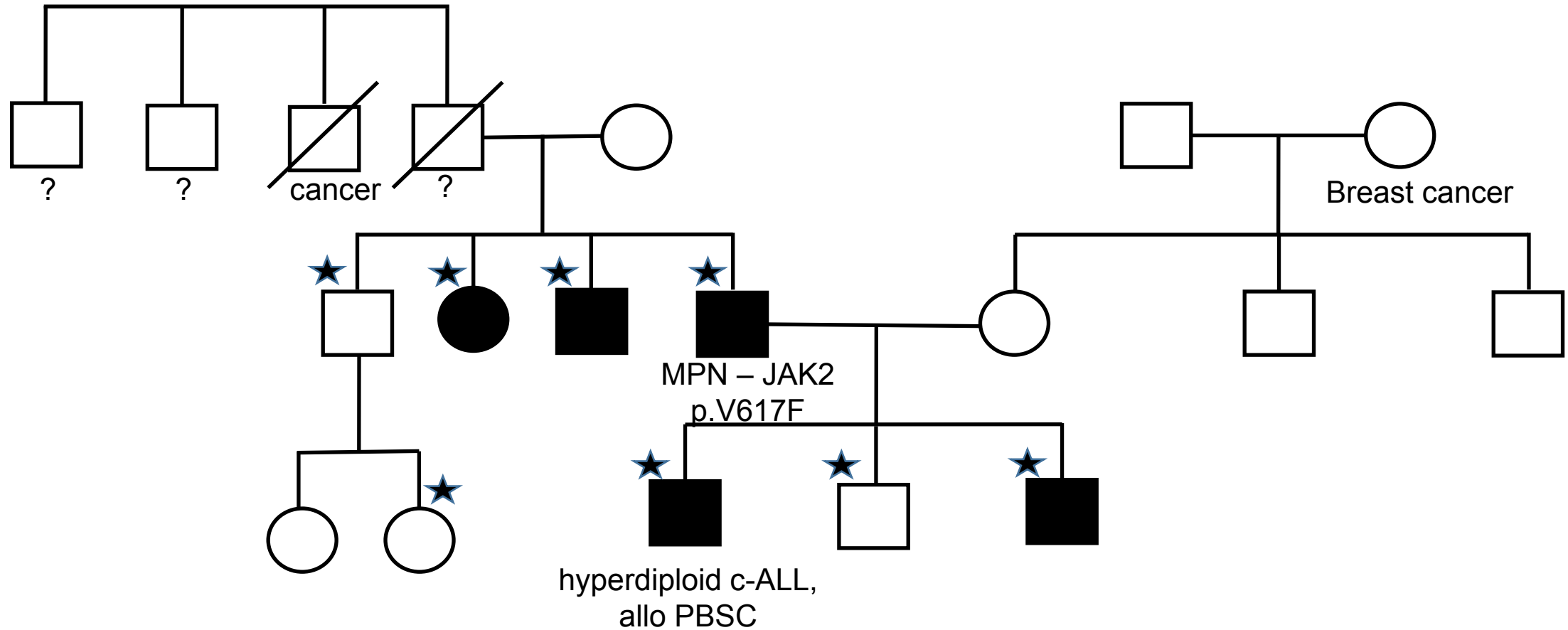
- Genetic vs. nongenetic influences
 - **monozygotic**: 100% of identical alleles
 - **dizygotic**: twins/siblings 50% of identical alleles

- Genetic influence: (concordance in MZ and DZ twins):
 - Diabetes mellitus
 - Schizophrenia
 - Lupus
 - Cleft
 - Sclerosis multiplex

Are monozygotic twins genetically identical?

Clinical cases

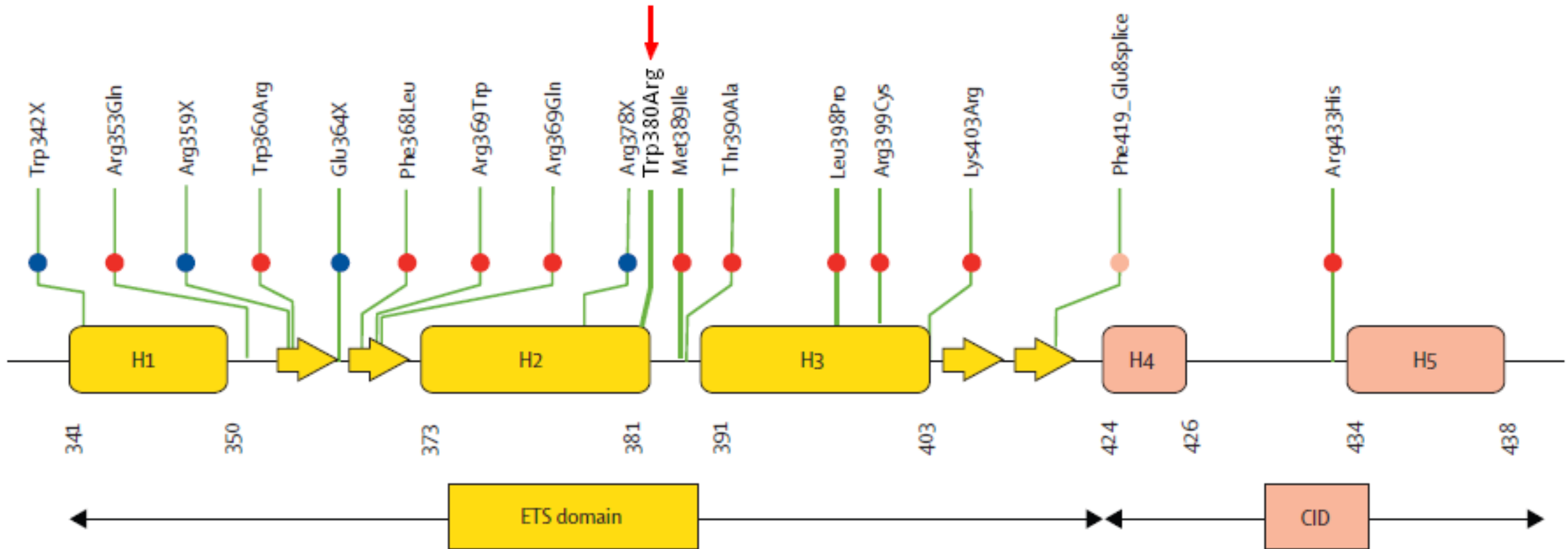
Clinical case: thrombocytopenia



★ analyzed samples

ETV6

- Exome sequencing – exome comparison of healthy and affected family members
- variant in gene *ETV6*: p.W380R



Bioinformatics and biostatistics

- Mapping on a reference sequence (BWA-mem)
- build-up of two variant files (Samtools mpileup):

1. file – affected family members
2. file – healthy family members

Exclusion of population and familial variants (VarScan):

Selection of variants present only in affected family members

Identification of potentially causal variants:

variant annotation (Annovar)

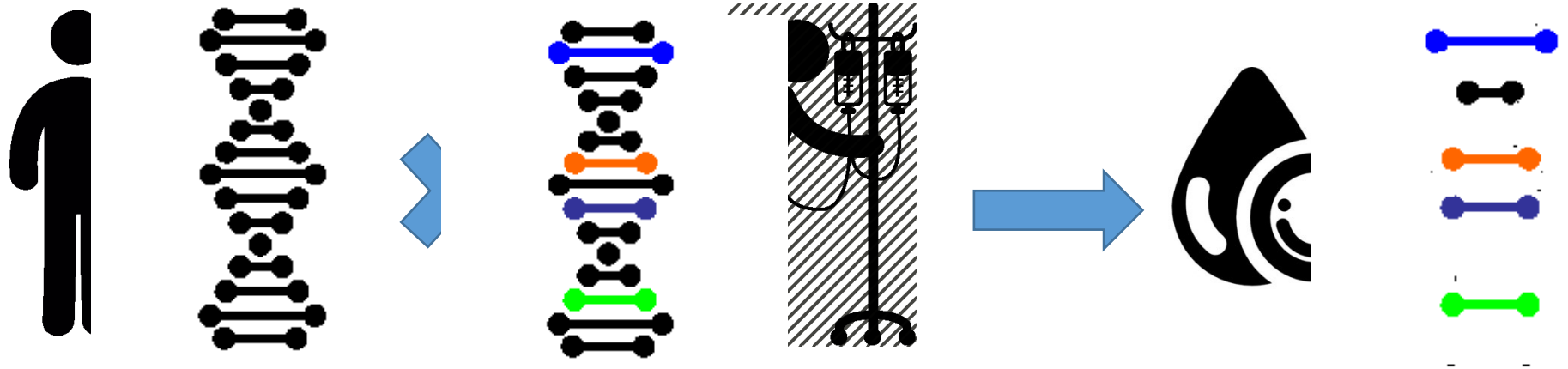
filters: coverage > 20

left only exonic, ncRNA exonic, downstream and upstream variants

left out synonymous variants

excluded variants annotated in dbSNP dtb. with rsXXXXX ID

Bioinformatics and biostatistics



Annotated mutations



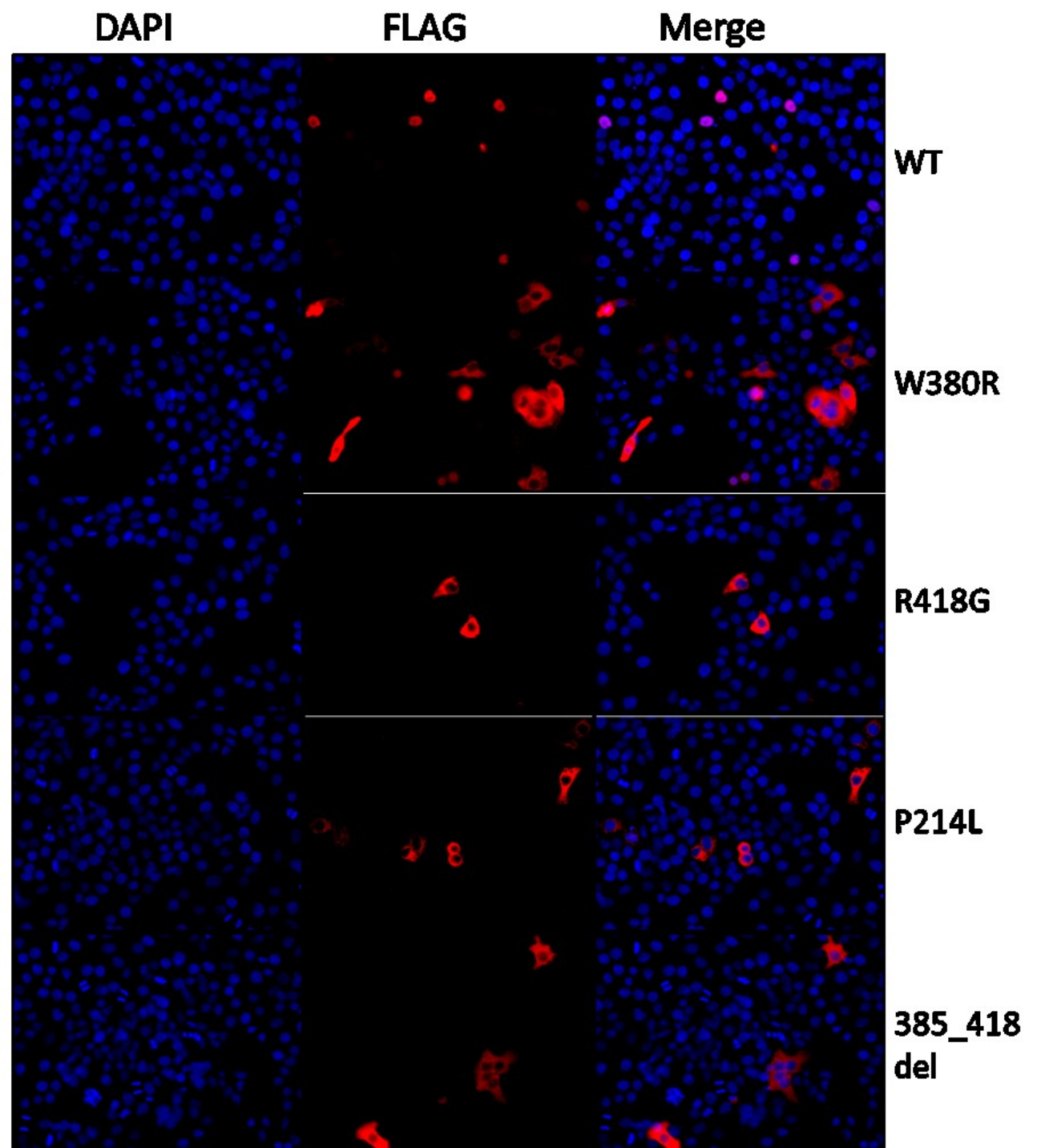
Variant frequency in population



Variant effect on the protein structure



Functional analysis of *ETV6* :
fluorescence microscopy

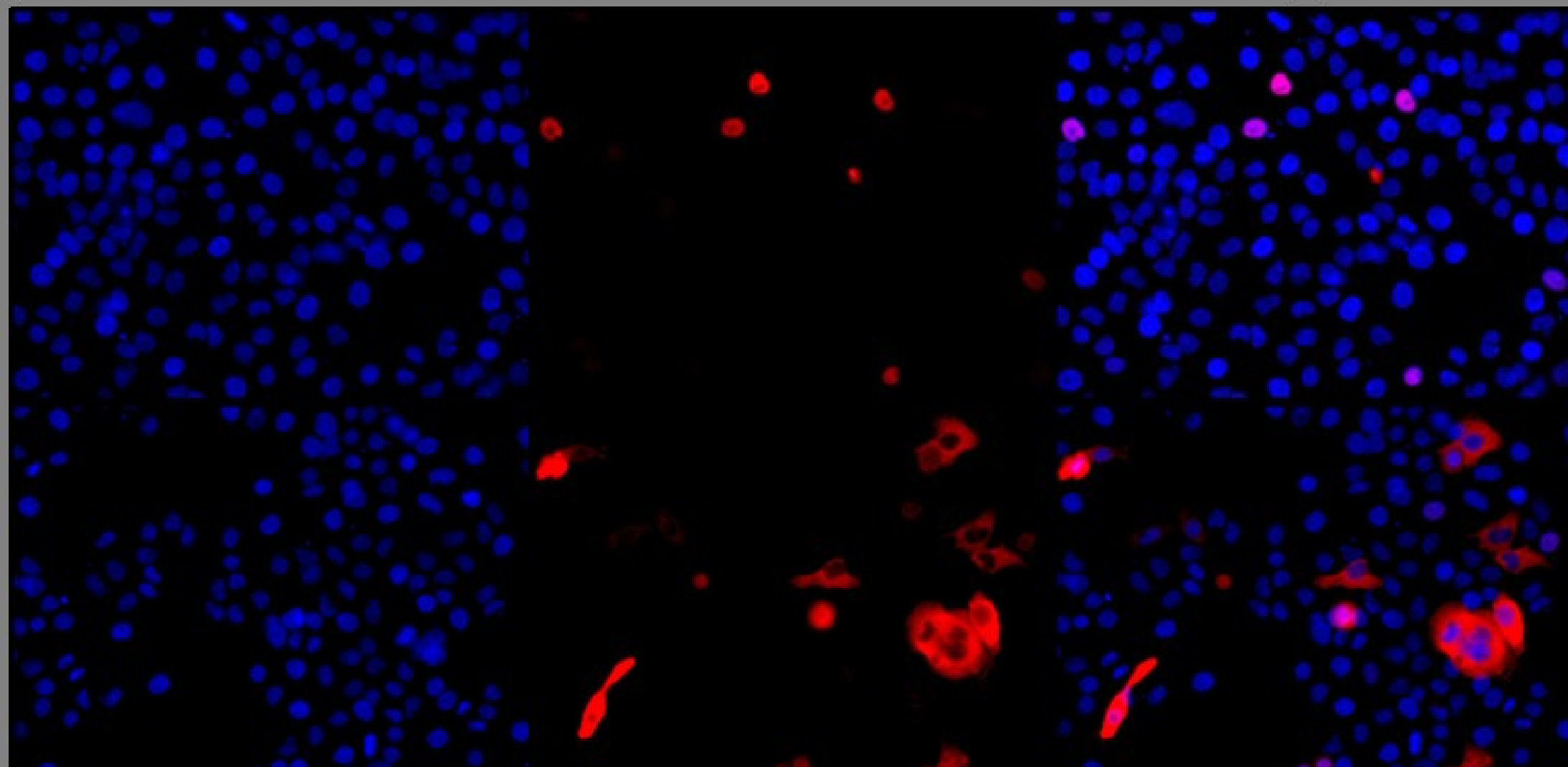


.2

DAPI

FLAG

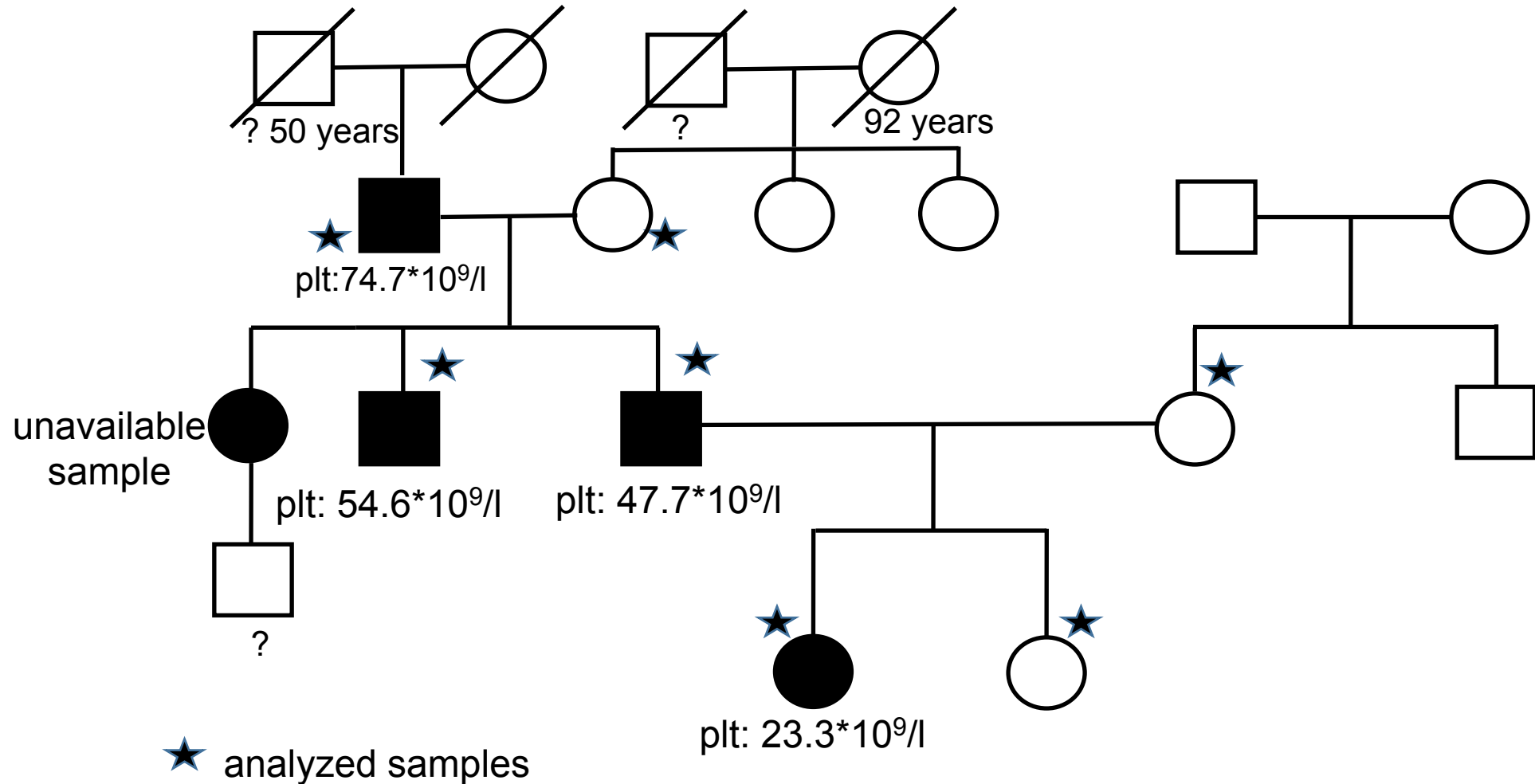
Merge

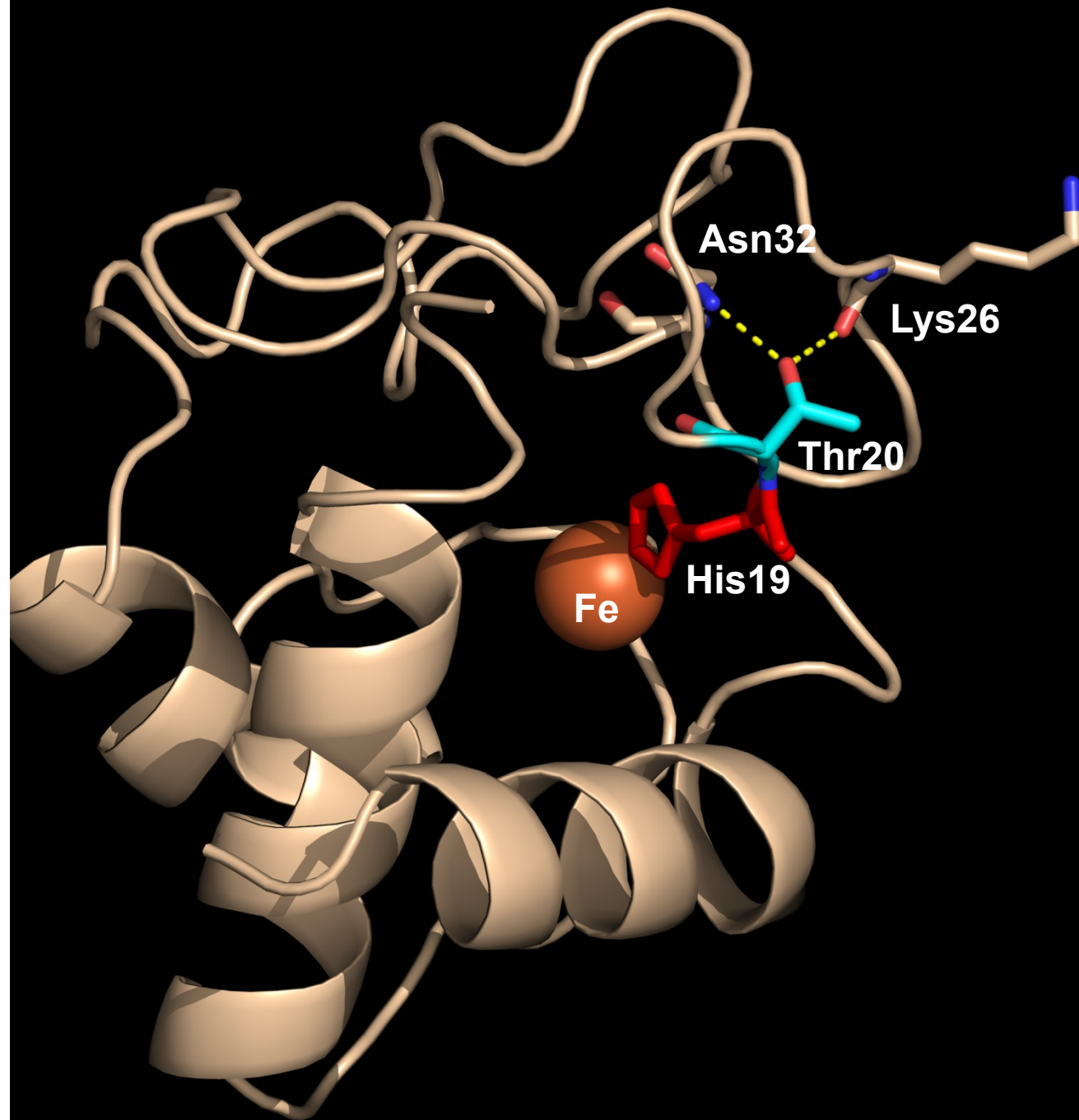


WT

W380R

CYCS: exon 2, p.T20I





CYCS: p.T20I



What are the skills of clinical geneticist?

- **complex examination**
- **gene/s analysis indication**
 - exome sequencing
 - genome sequencing
 - functional tests
- **results interpretation**
(from practitioners to clinical geneticists)
- **therapeutic and preventive intervention proposal**
 - respecting wishes of affected individuals together with ethical aspects



What are the skills of clinical geneticist?

➤ **complex examination**

➤ **gene/s analysis indication**

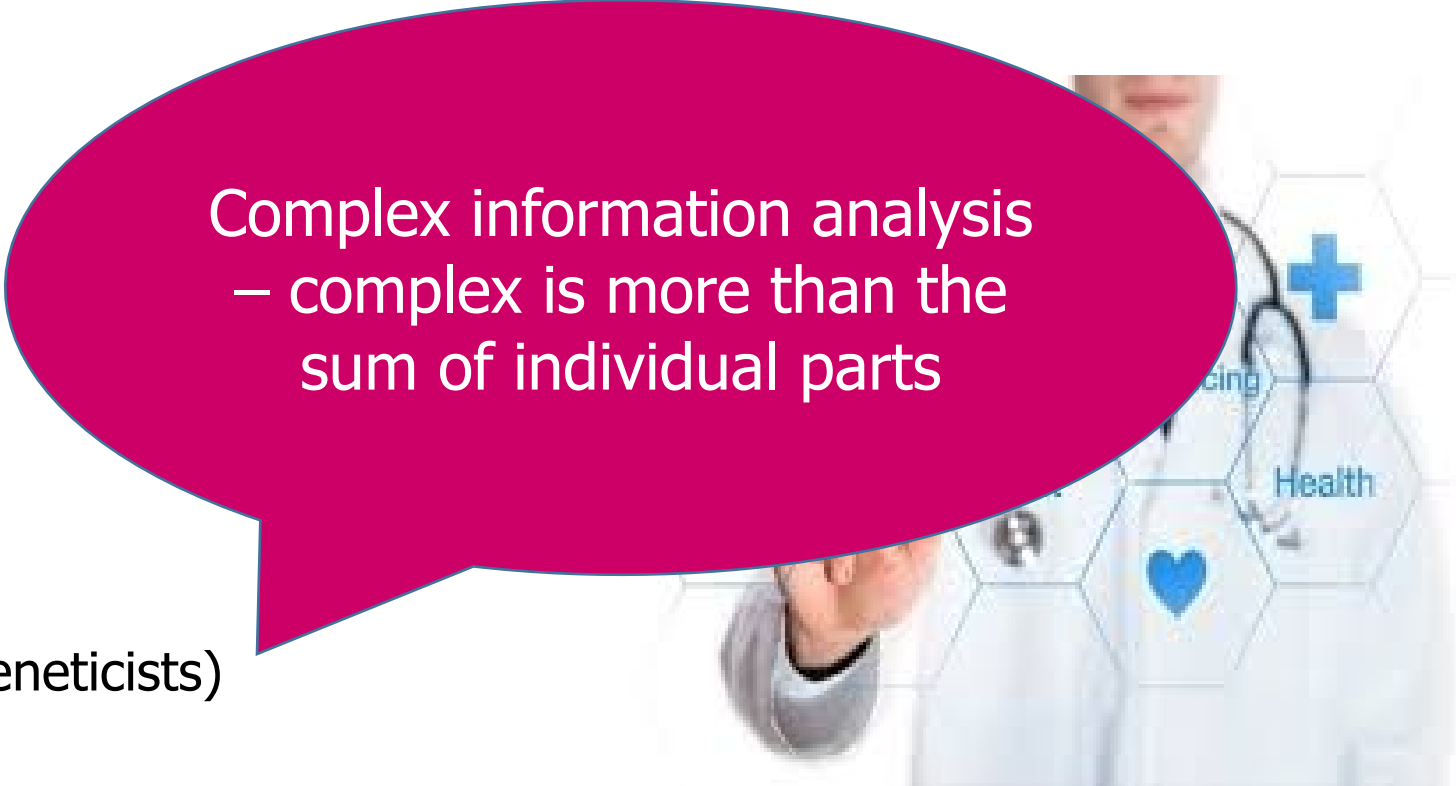
- exome sequencing
- genome sequencing
- functional tests

➤ **results interpretation**

(from practitioners to clinical geneticists)

➤ **therapeutic and preventive intervention proposal**

- respecting the desire of affected individuals together with ethical aspects



Complex information analysis
– complex is more than the
sum of individual parts

Why genetics?

- **Disease diagnostics:**
 - prenatal
 - preimplantational
 - genetic counselling
- **Therapy:**
 - pharmacogenetics
 - pharmacogenomics
 - immunogenetics
- **Prevention**
- **Gene therapy, genome editing**



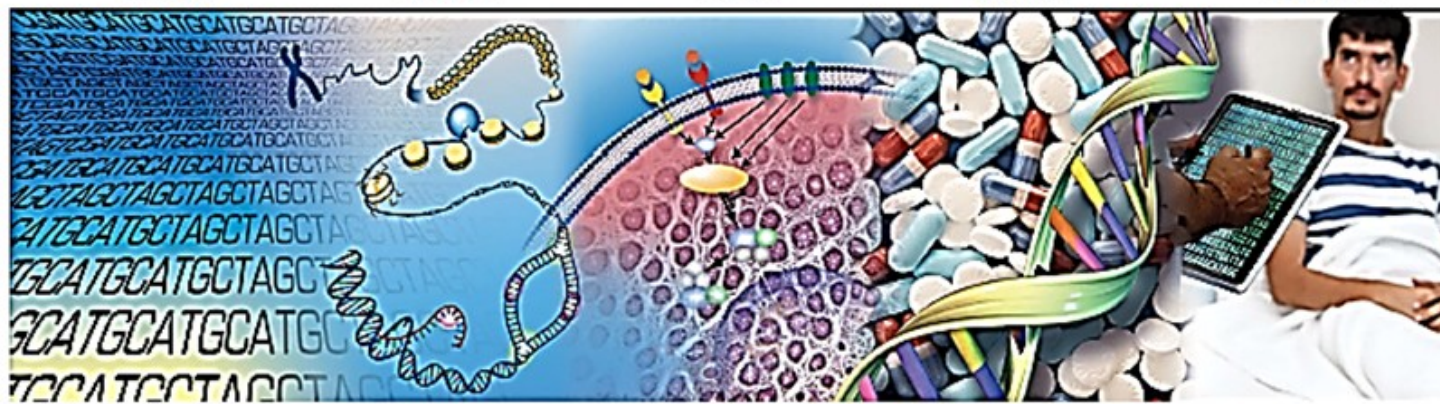
Understanding
the structure of
genomes

Understanding
the biology of
genomes

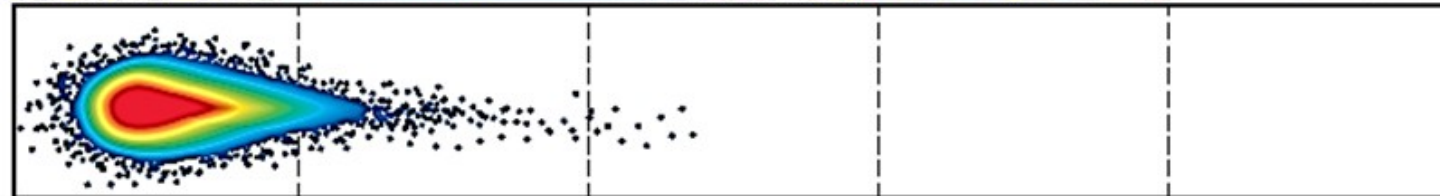
Understanding
the biology of
disease

Advancing
the science of
medicine

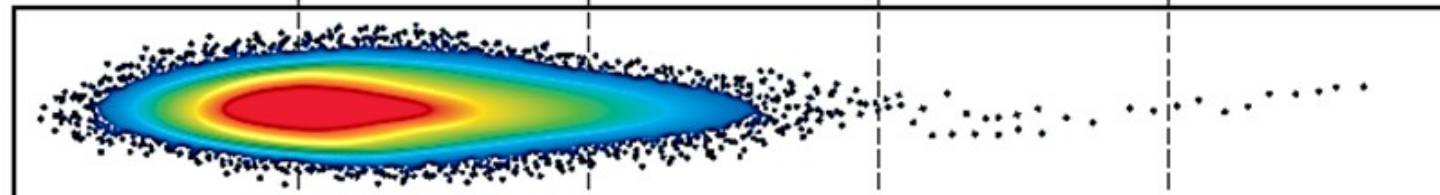
Improving the
effectiveness of
healthcare



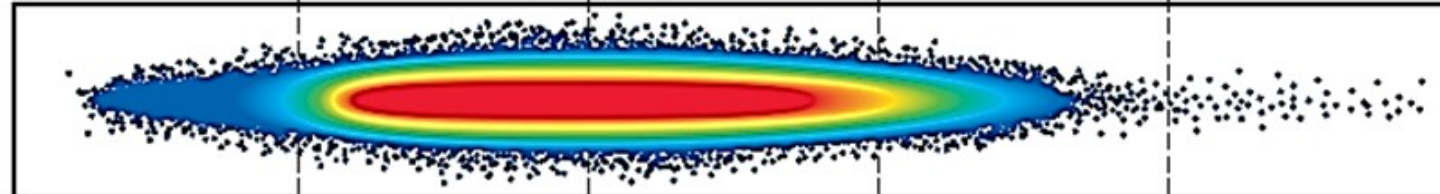
1990–2003
Human Genome Project



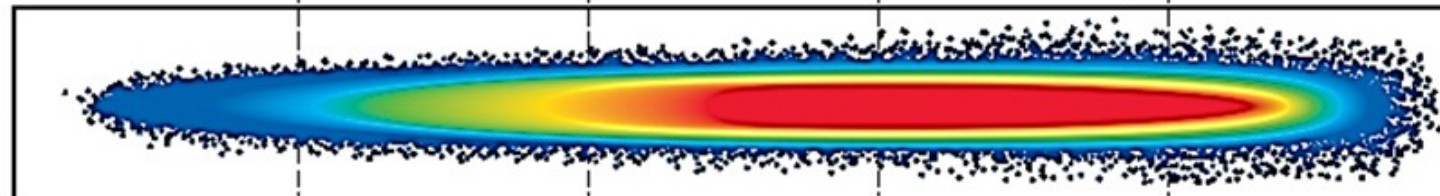
2004–2010



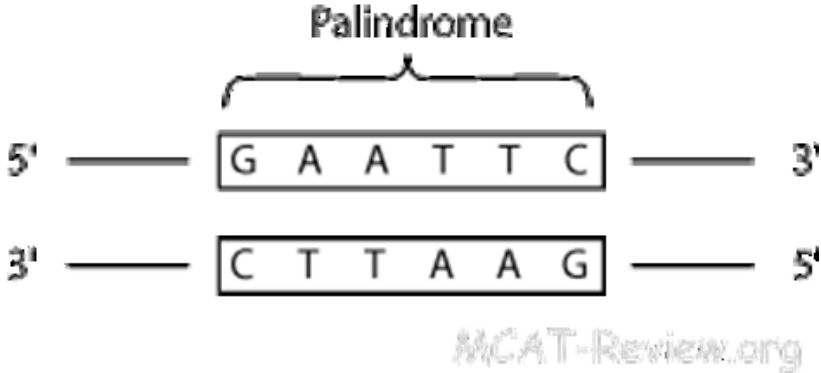
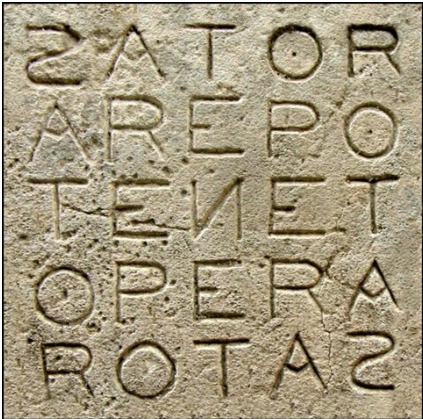
2011–2020



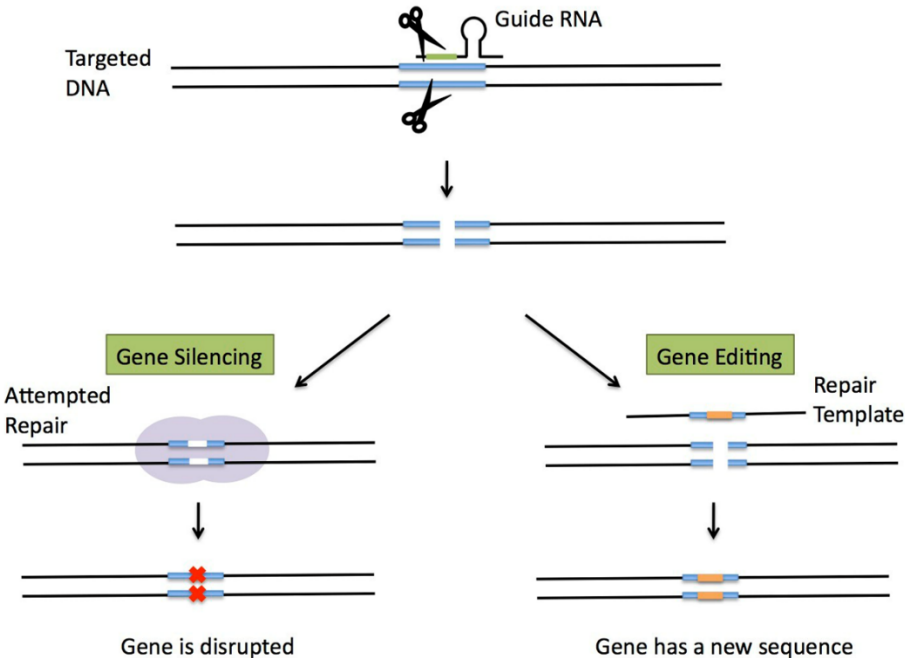
Beyond 2020



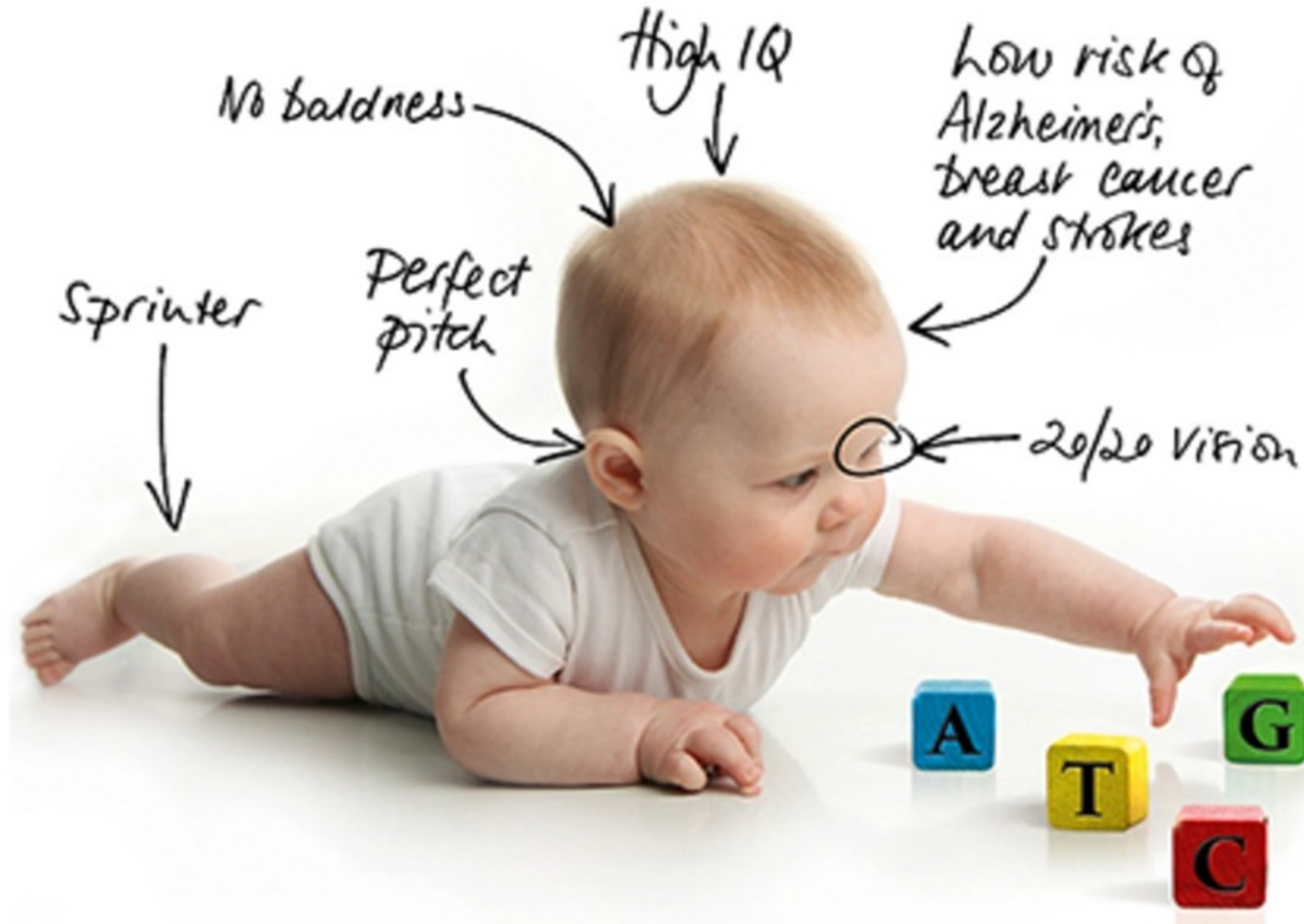
CRISPR-Cas9



Pompeie, 79AD



Made-to-order children?





**"Your weight problem is partly genetic
and partly Boston Cream pie."**