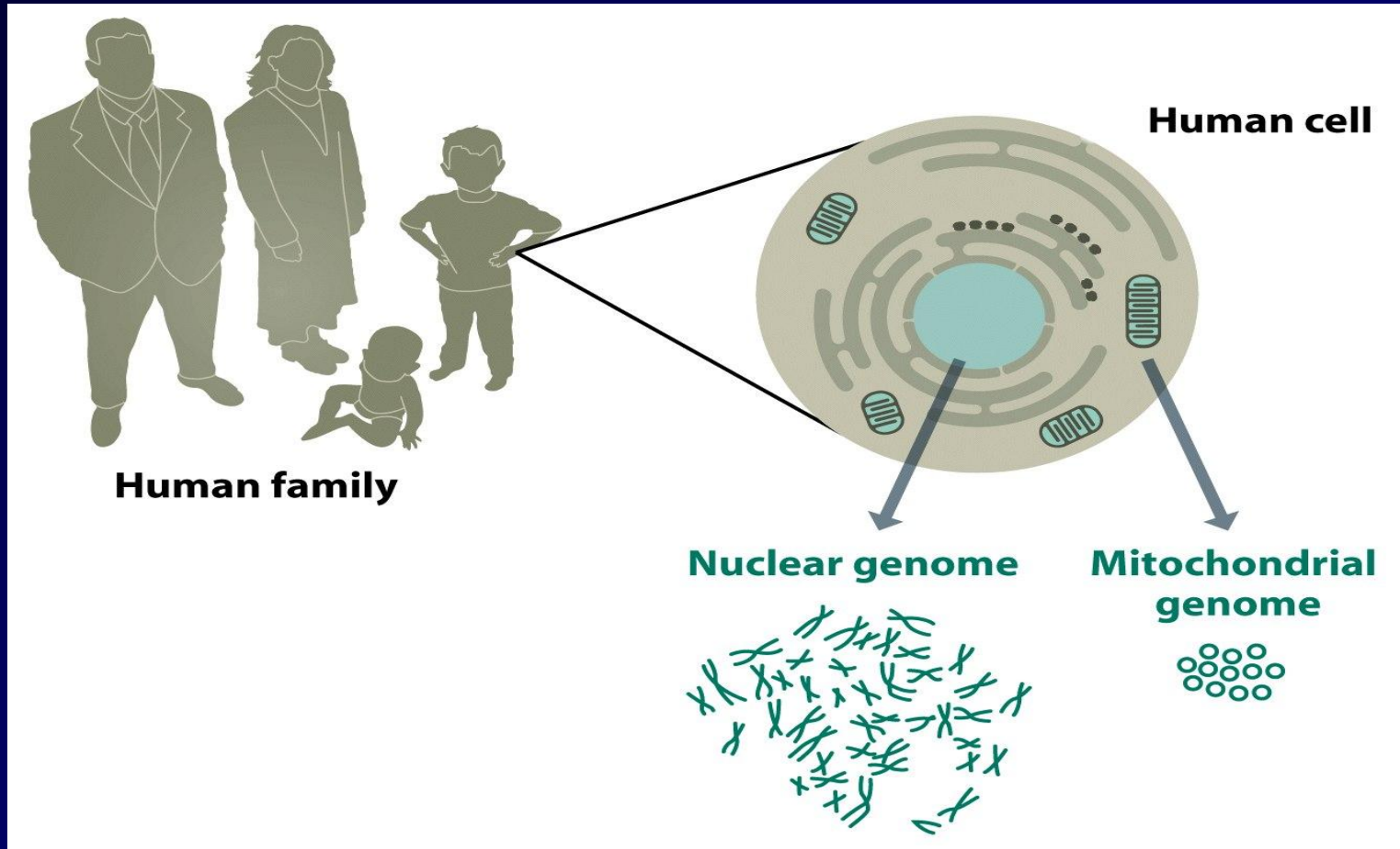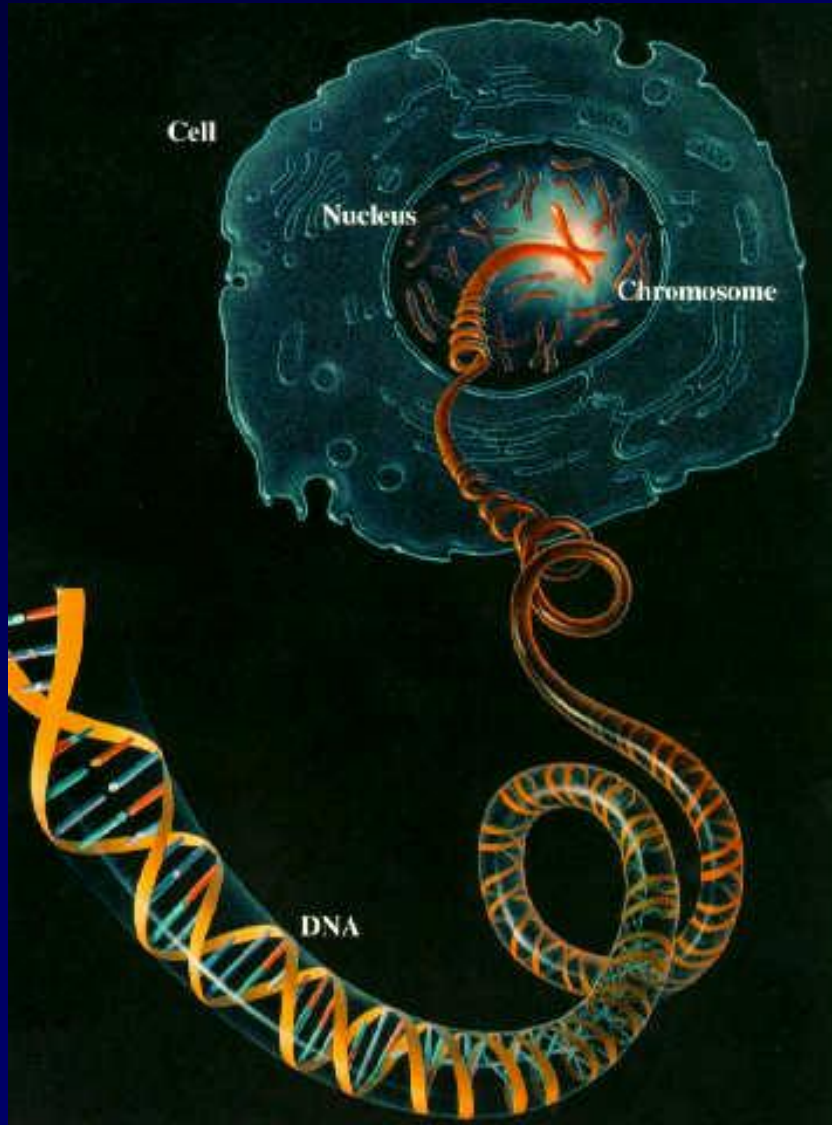# HUMAN GENOME

**Department of Medical Genetics
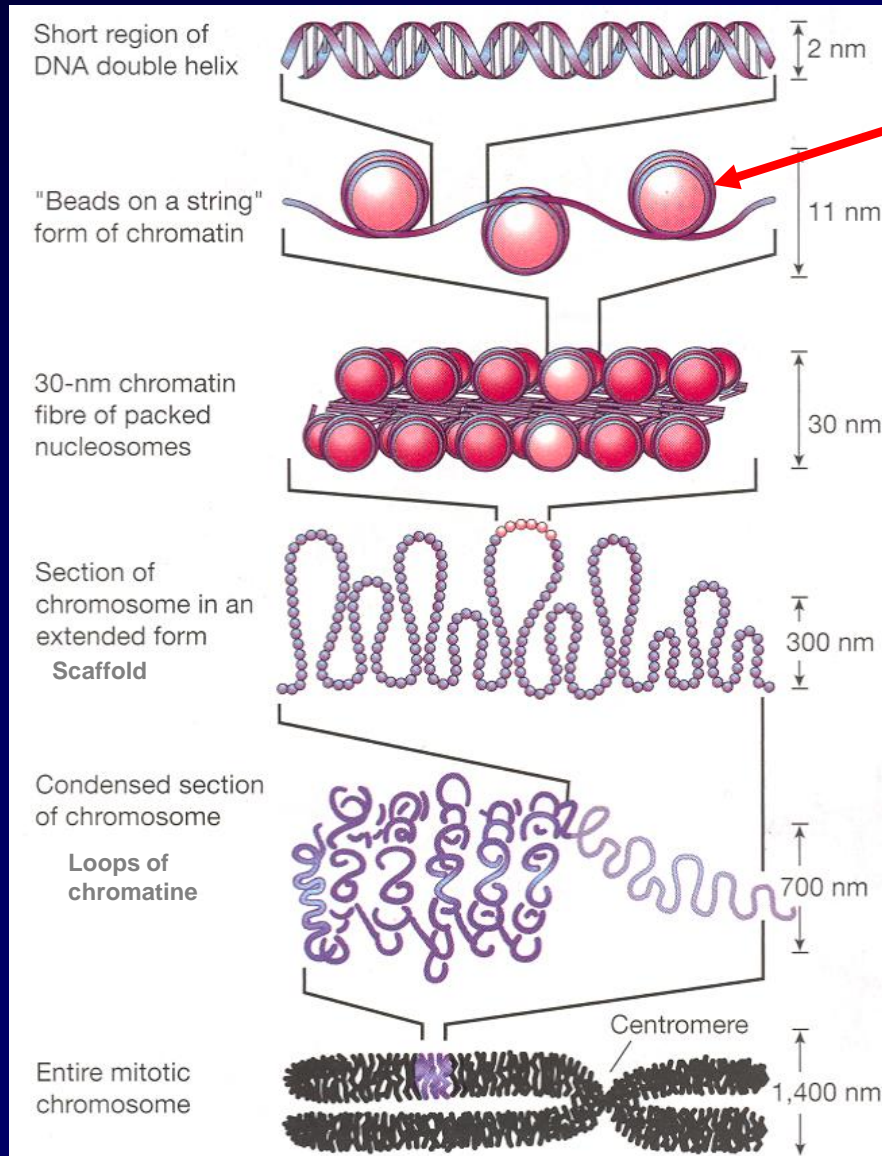Poznan University of Medical Sciences
Katarzyna Wicher M.Sc.**

The **human genome** is the complete set of genetic information for humans.
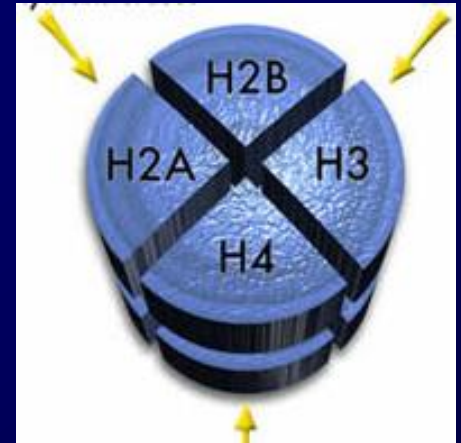
Nuclear genome

- Nuclear genome contains over **3 billion base pairs** organized into 23 paired chromosomes.

- It is **a double-stranded, linear** DNA molecule occuring within the cell nucleus.

- The haploid nuclear genome contains approximately **21,000** protein-coding genes.

Chromatine structure/ DNA condensation

Short region of DNA double helix — 2 nm

"Beads on a string" form of chromatin — 11 nm

nucleosome

30-nm chromatin fibre of packed nucleosomes — 30 nm

Section of chromosome in an extended form
Scaffold — 300 nm

Condensed section of chromosome
Loops of chromatine — 700 nm

Entire mitotic chromosome
Centromere — 1,400 nm

H2B
H2A
H3
H4

Gene structure

5' 3'

Promoter

START codon

intron

exon

STOP codon

AUG

5'UTR

UAA
UAG
UGA

3'UTR
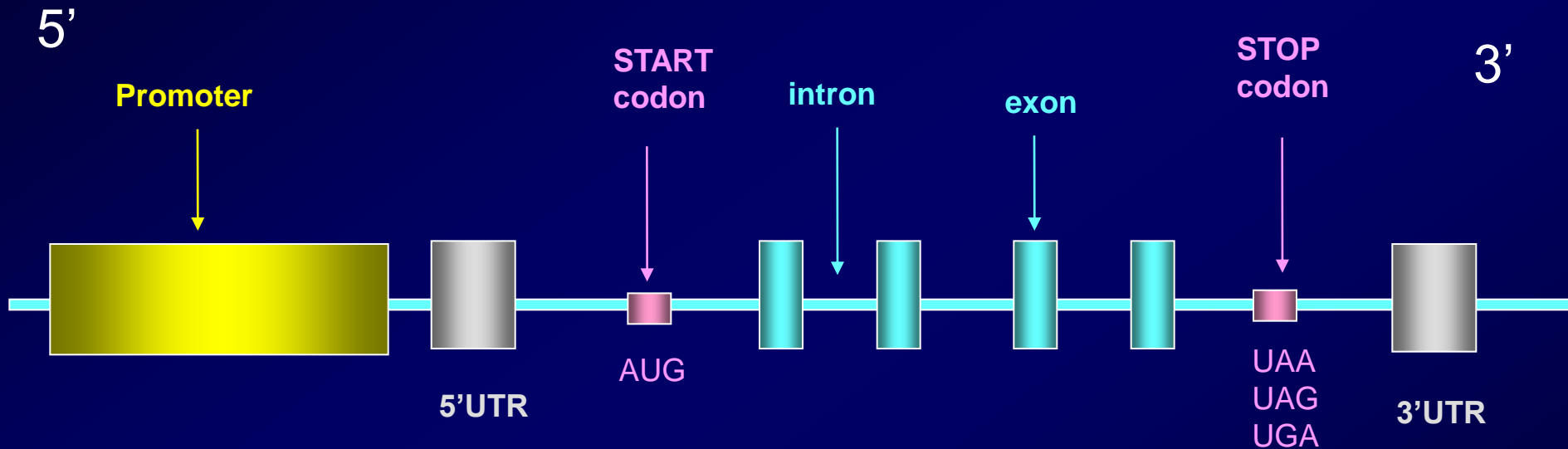
GENE is a fragment of DNA, localized in a specific place on the chromosome (locus) and coding a specific phenotypic trait.
It is a hereditary factor, which is transmited from parents to children.

Gene structure

An <u>allele</u> is one of two or more versions of a gene. An individual inherits two alleles for each gene, one from each parent. If the two alleles are the same, the individual is homozygous for that gene. If the alleles are different, the individual is heterozygous.
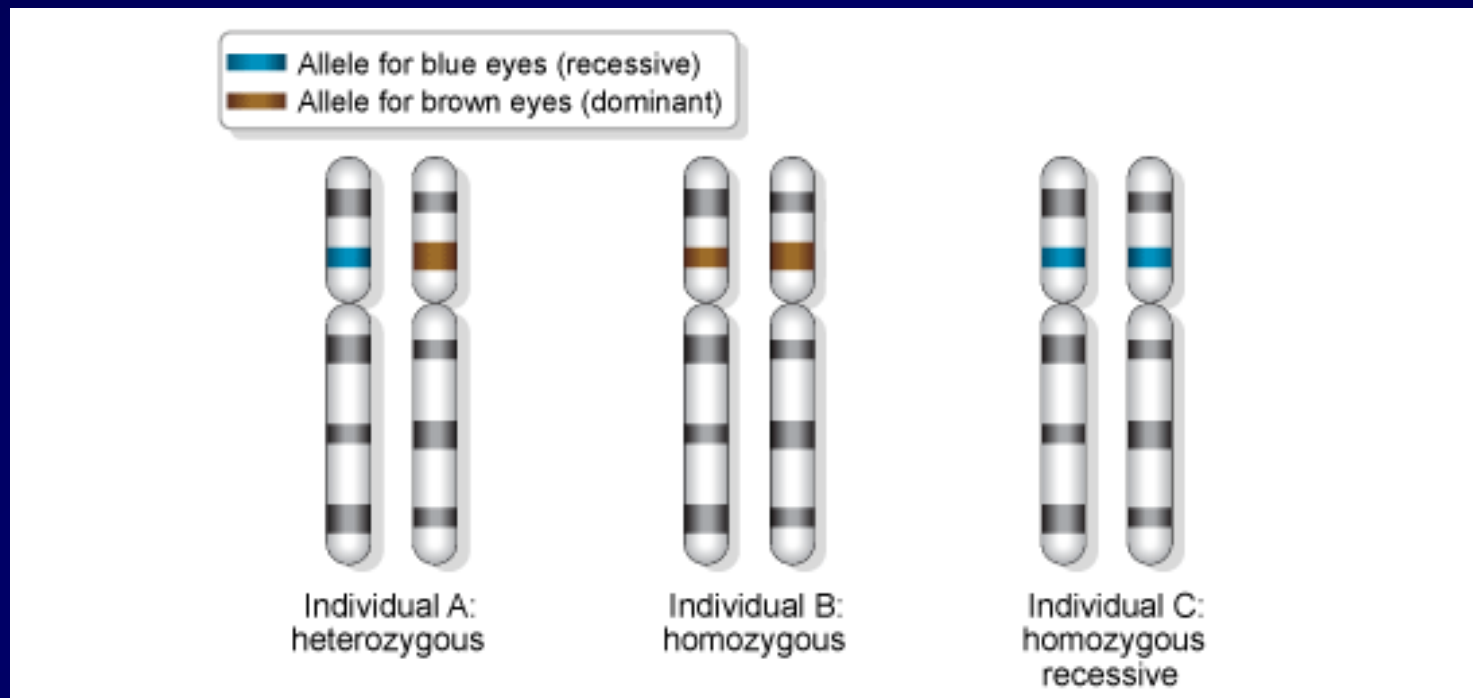
A gene may have more than two alleles

Consider Human blood groups:

1) The gene for the ABO blood type has three alleles: I^A; I^B; i

2) I^A specifies an enzyme that adds sugar A (dominant)
   I^B specifies an enzyme that adds sugar B (dominant)
   i does not produce a functional sugar-adding enzyme (recessive)

Genotypes ----------------------→ Phenotypes

$I^A I^A$; I^A i

$I^A I^B$???????????

$I^B I^B$; I^B i

i i

Mutations in the DNA are the source of new alleles

1) Mutation is the process whereby genes change from one allelic form to another. The creation of entirely new alleles can occur.

2) Genes mutate randomly, at any time and in any cell of an organism.

3) Mutations occur during normal replication; can also occur due to a mutagen, and due to erroneous repair following a exposure to a mutagen.

Mutation or polymorphism?

A mutation is a physical event in a single individual/cell.
A polymorphism is a population attribute..
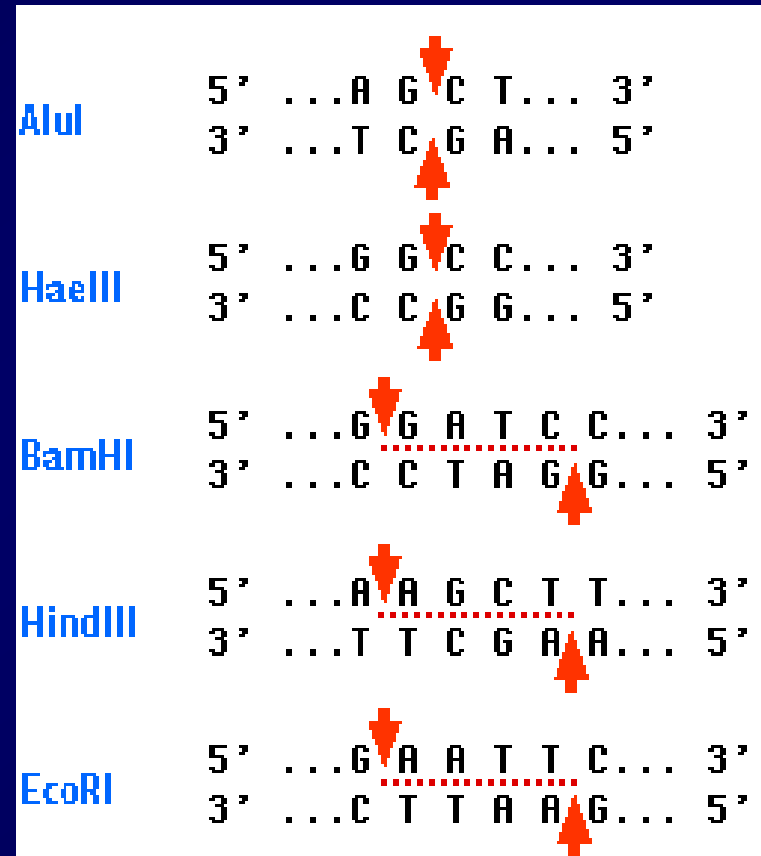
Definition of polymorphism

A polymorphic locus is one at which there are at least two alleles, each with a frequency greater than 1%. Alleles with frequencies less than 1% are referred to as mutants.

Types of DNA polymorphisms

1) **<u>Restriction fragment length polymorphism (RFLP)</u>** – alter the lenght of restriction fragments after cutting

2) Short tandem repeat polymorphism (STR):

   a) Minisatellite repeat polymorphism (or Variable Numbers of Tandem Repeats; VNTRs) - 7-300 bp motifs

   b) Microsatellite repeat polymorphism - 1-6 bp motifs

3) Single nucleotide polymorphism (SNP) – 1 bp

Restriction Enzymes

• Restriction enzymes are DNA-cutting enzymes found in bacteria. Because they cut within the molecule, they are often called restriction endonucleases.

• A restriction enzyme recognizes and cuts DNA only at a particular sequence of nucleotides.

• The recognition site is commonly 4 or 6 bp in lenght.

AluI
5' ...A G C T... 3'
3' ...T C G A... 5'

HaeIII
5' ...G G C C... 3'
3' ...C C G G... 5'

BamHI
5' ...G G A T C C... 3'
3' ...C C T A G G... 5'

HindIII
5' ...A A G C T T... 3'
3' ...T T C G A A... 5'

EcoRI
5' ...G A A T T C... 3'
3' ...C T T A A G... 5'

AluI and HaeIII produce blunt ends

BamHI HindIII and EcoRI produce "sticky" ends

Restriction Fragment Length Polymorphism – RFLP

This method allows detection of nucleotide changes that alter a restriction site in amplified sequence

PCR amplification of a portion of the ß-globin (HBB) gene and digestion with MstII in a sickle-cell disease
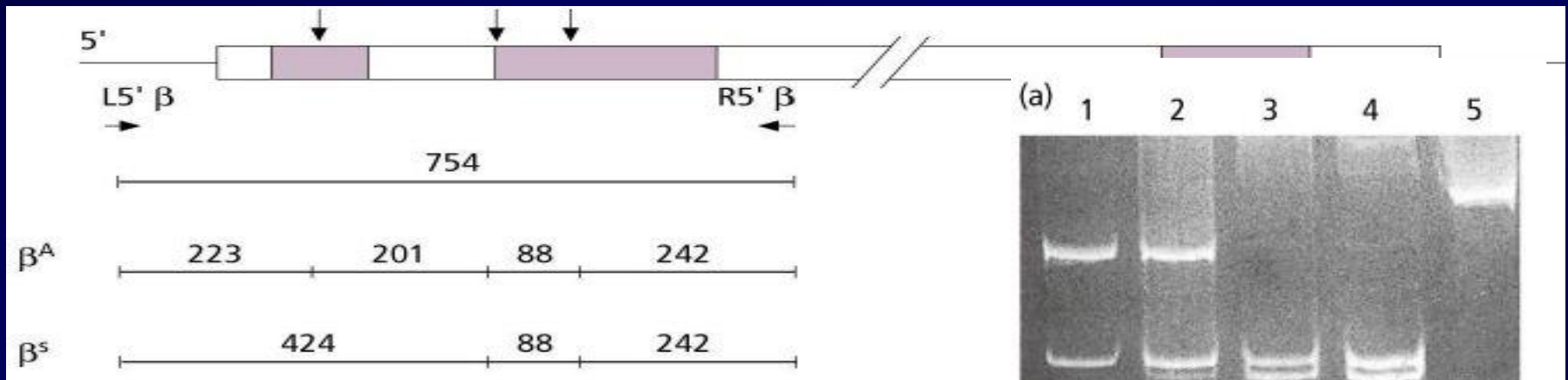
lane 1 – homozygote

lane 2 – heterozygote

lane 3 – normal homozygote

lane 4 – normal homozygote

lane 5 – control containing amplified but undigested DNA
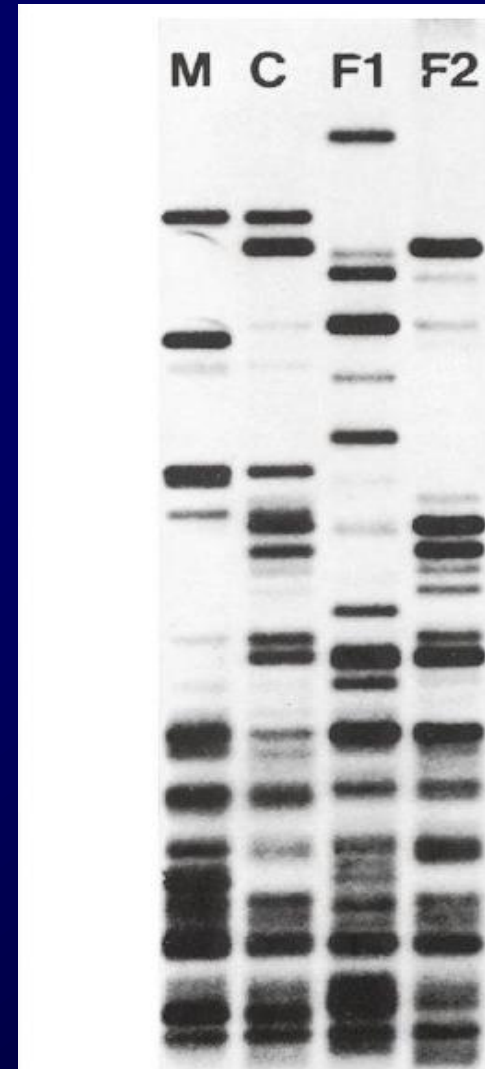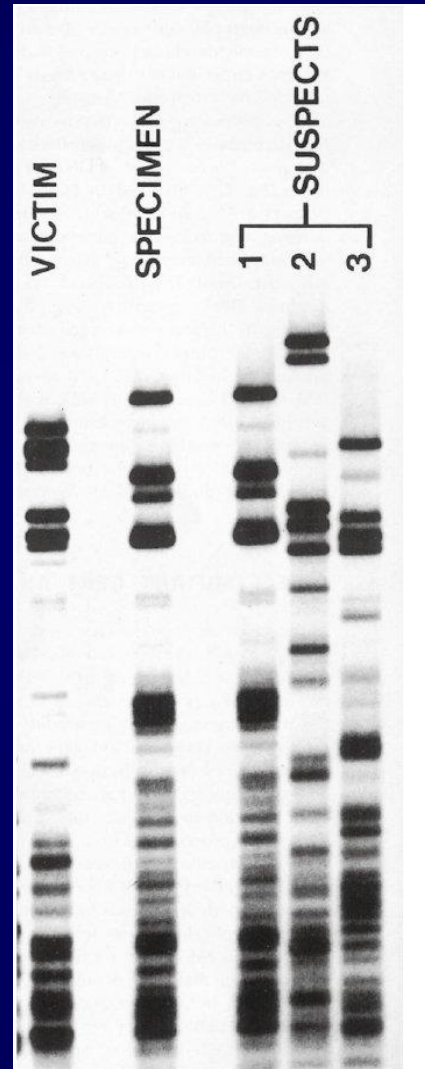
Types of DNA polymorphisms

1) Restriction fragment length polymorphism (RFLP)

2) **Short tandem repeat polymorphism (STR):**

 a) Minisatellite repeat polymorphism (or Variable Numbers of Tandem Repeats; VNTRs) - 7-300 bp motifs

 b) Microsatellite repeat polymorphism - 1-6 bp motifs

3) Single nucleotide polymorphism (SNP) – 1 bp – the most simple form, very common in the human genome.

DNA fingerprint

Useful for identification of individuals by their respective/unique DNA profiles

- criminalistics/forensic investigation

- paternity testing

VICTIM   SPECIMEN   SUSPECTS   1 2 3

M   C   F1   F2

Types of DNA polymorphisms

1) Restriction fragment length polymorphism (RFLP)

2) Short tandem repeat polymorphism (STR):

   a) Minisatellite repeat polymorphism (or Variable Numbers of Tandem Repeats; VNTRs) - 7-300 bp motifs

   b) Microsatellite repeat polymorphism - 1-6 bp motifs

3) **<u>Single nucleotide polymorphism (SNP)</u>** – 1 bp – the most simple form, very common in the human genome.

Genetic Variation Among People

Single nucleotide polymorphisms (SNPs)

GATTTAGATCGCGATAGAG
GATTTAGATCTCGATAGAG

Common in "normal" human genomes--major cause of phenotypic variation
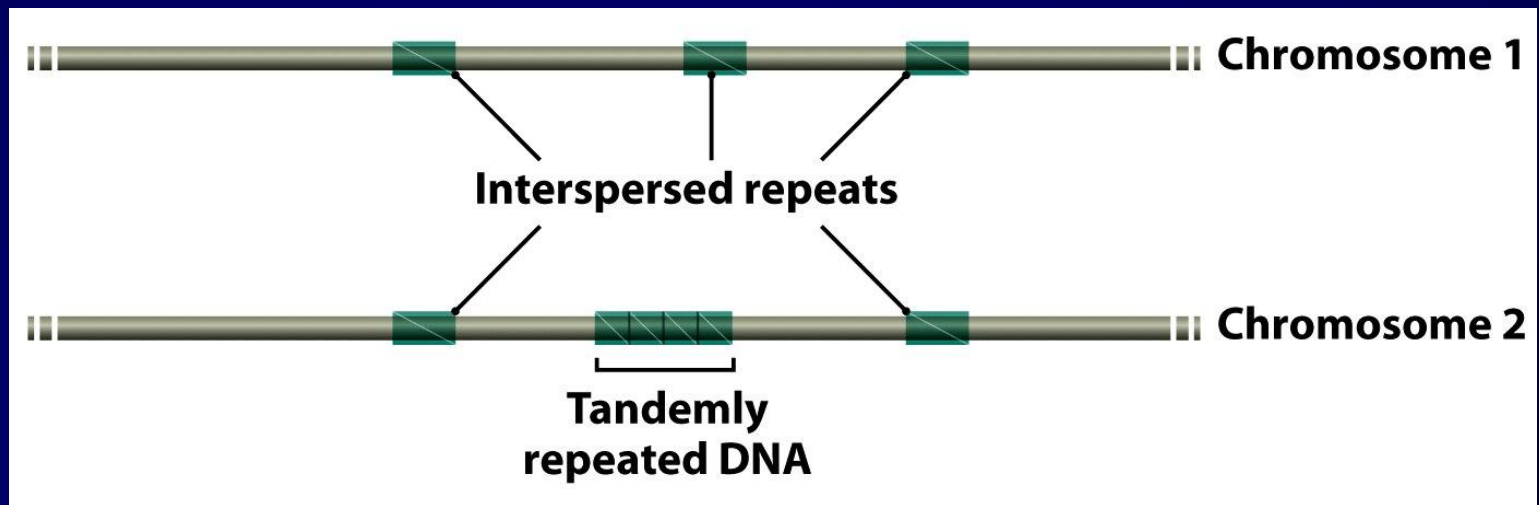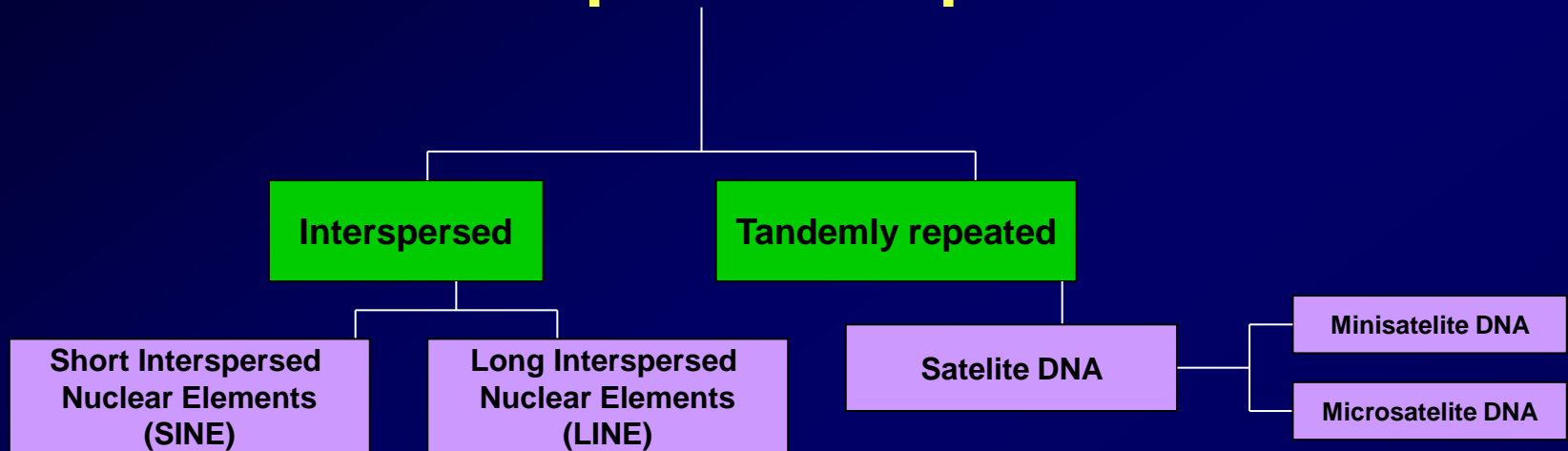
Common in certain diseases, particularly cancer

Now showing up in rare disease; autism, schizophrenia

DNA repeated sequences

```
                          Interspersed                    Tandemly repeated

      Short Interspersed          Long Interspersed              Satelite DNA              Minisatelite DNA
      Nuclear Elements            Nuclear Elements
         (SINE)                      (LINE)                                                Microsatelite DNA
```


Chromosome 1

Interspersed repeats

Chromosome 2

Tandemly
repeated DNA

Interspersed repetitive noncoding DNA

Transposon-derived repeats make up >40% of the human genome and mostly arose through RNA intermediates
- known as transposable elements – transposons

They are organized into two grouped acording to the method of transposition:

1. Retrotransposons (retroposons). Via RNA transcripts and cellular reverse transcriptase (replicative transposition). Include three types: long interspersed nuclear elements (LINES); short interspersed nuclear elements (SINES); and retrovirus-like elements containing long terminal repeats.

2. DNA transposons. Migrate by conservative transposition. Sequence is excised and re-inserted elsewhere in the genome.

- LINES (LINE1 family) and SINES (Alu repeats) predominate.

Genetic code

The instruction in a gene that tells the cell how to make a specific protein. A, T (U), G, and C are the "letters" of the DNA code

The standard genetic code. Termination condons are indicated as 'Stop'.

The code defines how sequences of these nucleotide triplets, called **codons (64)**, specify which amino acid will be added next during protein synthesis.

Characteristics of genetic code

- composed of 4 nucleotides on mRNA (A,U,G,C)

- read in triplets (3 nucleotides / codon)

- non-overlapping: each nucleotide part of only 1 codon

- degenerated: each amino acid can be specified by ~ three different codons

- unambiguous: one codon can code just for ONE aminoacid

- commaless: there is no punctuation beetwen

- not quite universal: evolutionary divergence of organelle genetic codes (mitochondria)

History of medical genetics

- 1866: Gregor Mendel publishes Experiments in Plant Hybridization, which lays out the basic theory of genetics. It is widely ignored until 1900.
- 1900: rediscovery of Mendel's work by Robert Correns, Hugo de Vries, and Erich von Tschermak .
- 1902: Archibald Garrod discovers that alkaptonuria, a human disease, has a genetic basis. „Inborn errors in metabolism".
- 1910: Thomas Hunt Morgan proves that genes are located on the chromosomes (using Drosophila).
- 1953: Francis Crick and James Watson determine the structure of the DNA molecule, which leads directly to knowledge of how it replicates
- 1956: Joe Hin Tjio and Albert Levan establishe the correct chromosome number in humans to be 46.
- 2003: Sequence of the entire human genome is announced (U.S Department of Energy and the National Institute of Health)

Sequencing of Human Genome

- **2003** – **HUMAN GENOME PROJECT** - complete DNA sequence of Human Genome (13 years, 3 bilion dollars)

- **2007** –Craig Venter's DNA sequence was published

- **2007** - James Watson – first personal genome (1 milion dollars, 2 months, sequence was published: http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/, with the exception of **ApoE** gene)

- **2008** – first female genom - dr Marjolein Kriek from Netherlands (40 .000 euro, 6 months)
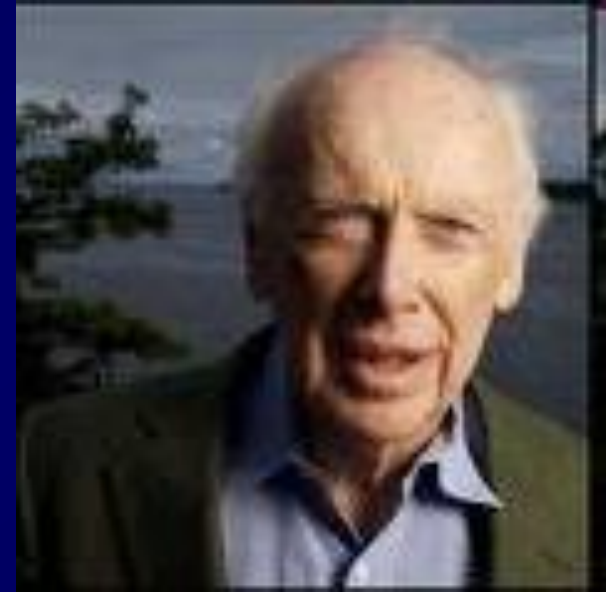
Personal Genome Project

- „More than 1,000 individuals have enrolled in the PGP and volunteered to share their DNA sequences, medical information, and other personal information with the research community and the general public.

  PG-10" – first ten volunteers whose data are publicly available:

  (James Sherley, Misha Angrist, John Halamka, Keith Batchelder, Rosalynn Gill, Esther Dyson, George Church, Kirk Maxey,Stan Lapidus, Steven Pinker)

- http://www.personalgenomes.org/mission.html

Genome Sequencing

- The Personal Genome Sequencing Service

- June 10, 2009-Illumina, Inc. today unveiled a service program to provide high-quality personal genome sequencing for consumers. This is the first service to offer complete coverage of the human genome sequence for under $50,000. The offering includes sequencing of an individual's DNA to 30 times depth, providing information on SNP variation and other structural characteristics of the genome such as insertions, deletions and rearrangements.

- Dr Jay Flatley (President of Illumina) -predicts that in 2019 the sequencing of the genome will be a routine technique performed shortly after birth, and the price of genome sequencing will be under 1000$
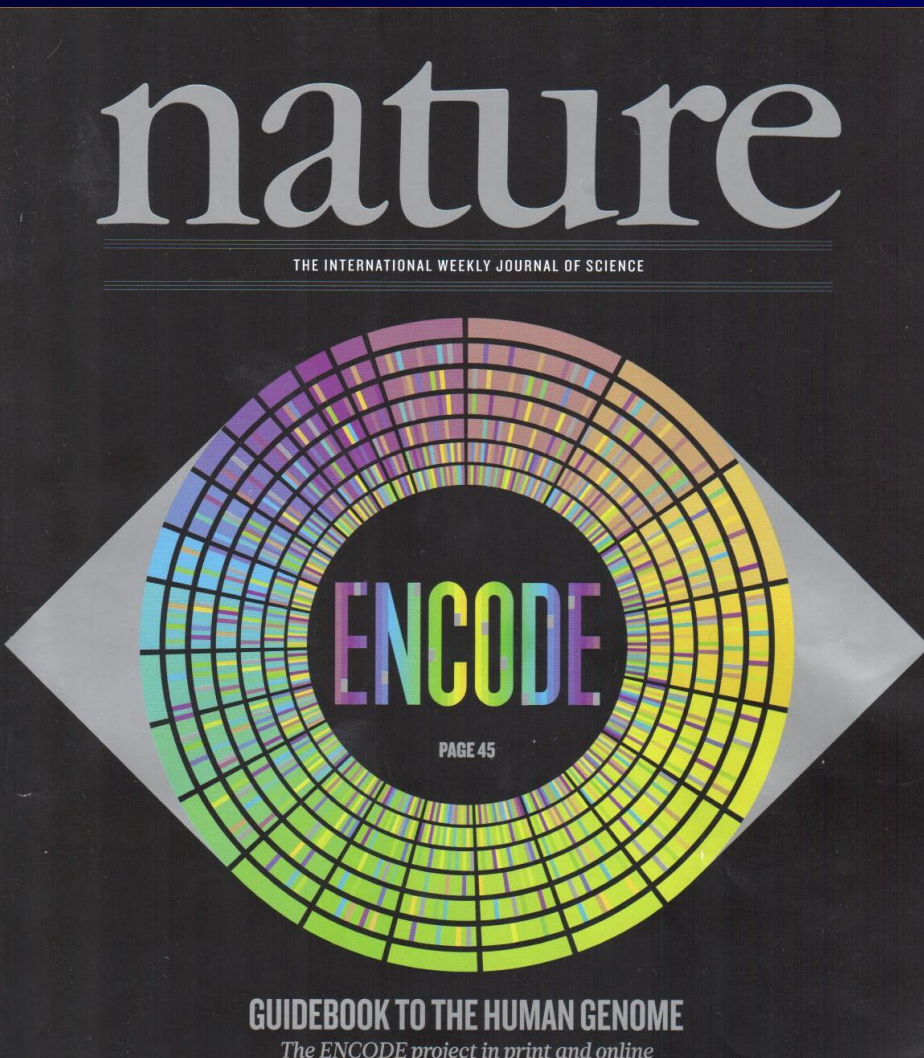
Sequencing Cost & Number of Sequenced genomes
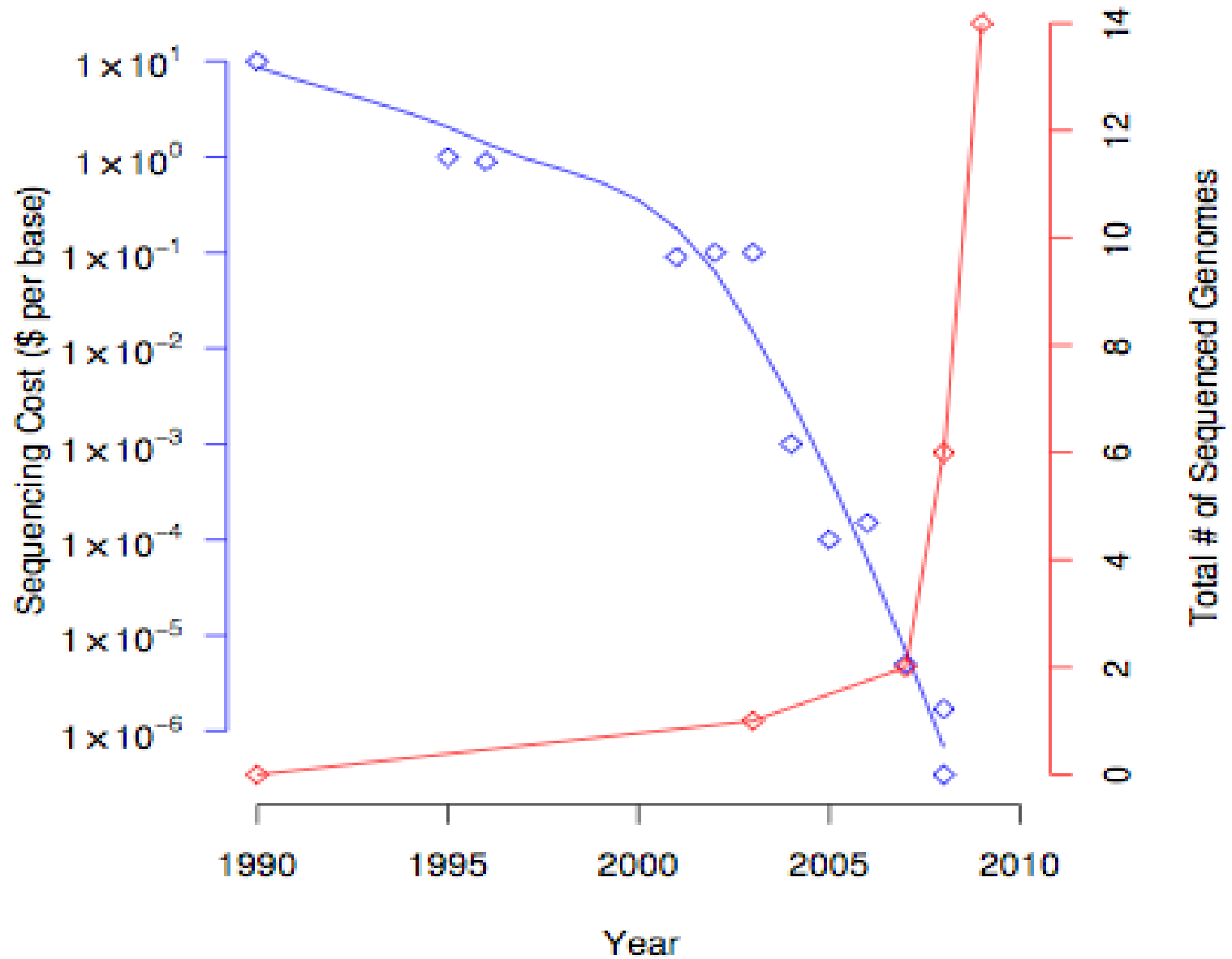
ENCODE project – no junk DNA?

6 September 2012

the **Enc**yclopedia **O**f **D**NA **E**lements

- to identify all regions of transcription, transcription factor association, chromatin structure and histone modification in the human genome sequence.

- Thanks to the identification of these functional elements, **80%** of the components of the human genome now have at least one biochemical function associated with them.
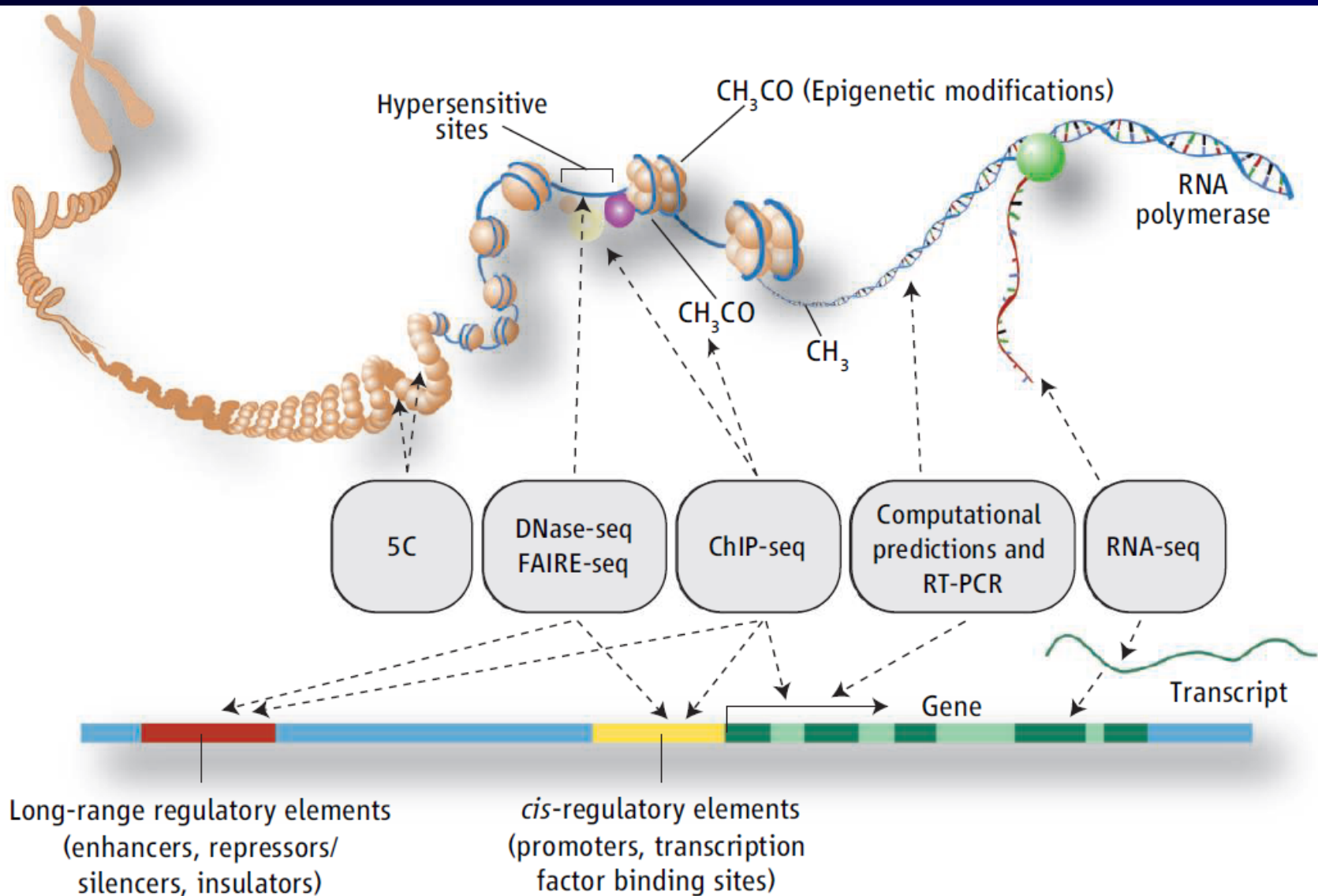
Zooming in. A diagram of DNA in ever-greater detail shows how ENCODE's various tests (gray boxes) translate DNA's features into functional elements along a chromosome.

Biologically active DNA

- About **21000** protein-coding genes – 3 % of the human genome

- **Noncoding genes**:

    1) **9600** long noncoding RNA molecules

    2) **8800** small RNA molecules

        (rRNA, tRNA, snRNA, snoRNA, miRNA, siRNA, ribosymes….)

    3) **11224** "pseudogenes"- disfunctional relatives of genes

- **Regulatory DNA elements:**

    4) promoters

    5) enhancers, silencers

    6) transcription regulatory elements on chromatin level (DNA methylation, histones modifications-reduce gene exp)

Thank you for your attention !!!