Master in Cellular and Molecular Biology

Medical and Cancer Genetics course

# MEDICAL GENETICS

Teacher: Claudia Giachino

**Lesson 1**

Genetics, genes and genomes

# GENETICS, GENES AND GENOMES

- The last 60 years of genetics
- Human genome and other genomes
- Genetic diseases
- Genetic polymorphisms

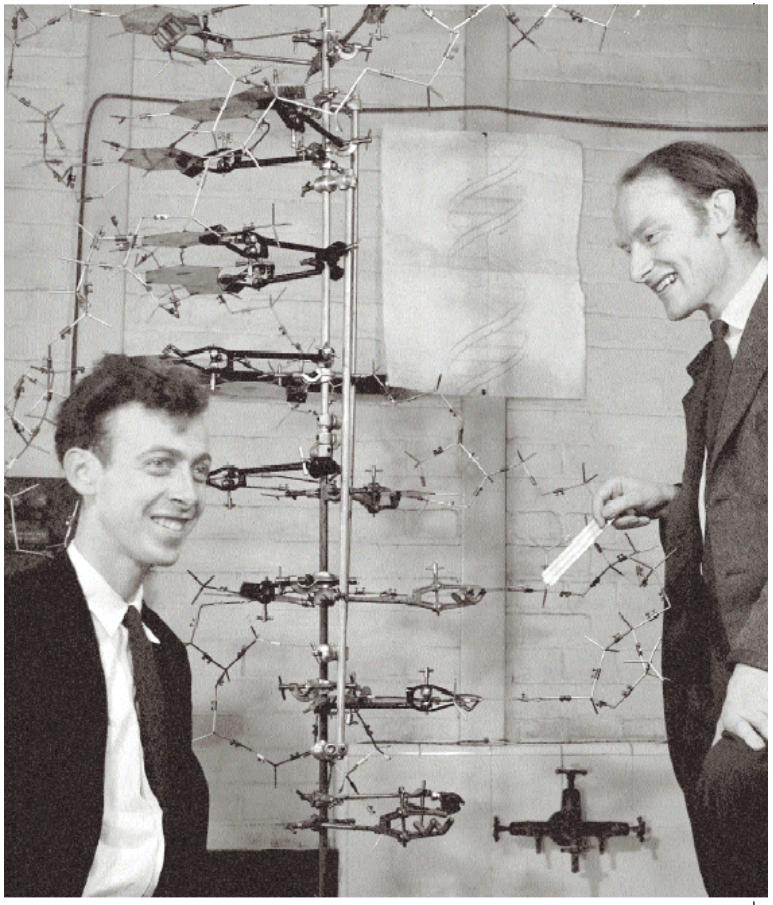


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

# GENETICS, GENES AND GENOMES

- **The last 60 years of genetics**
- **Human genome and other genomes**
- **Genetic diseases**
- **Genetic polymorphisms**

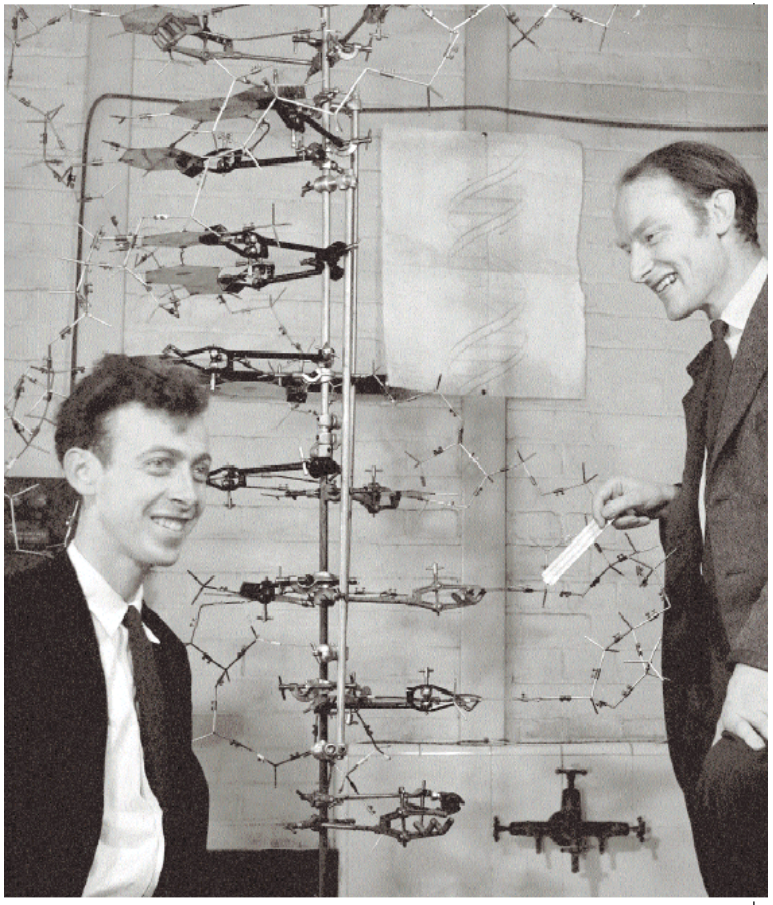


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

# Who is Watson and who is Crick?

Watson?

YES

NO

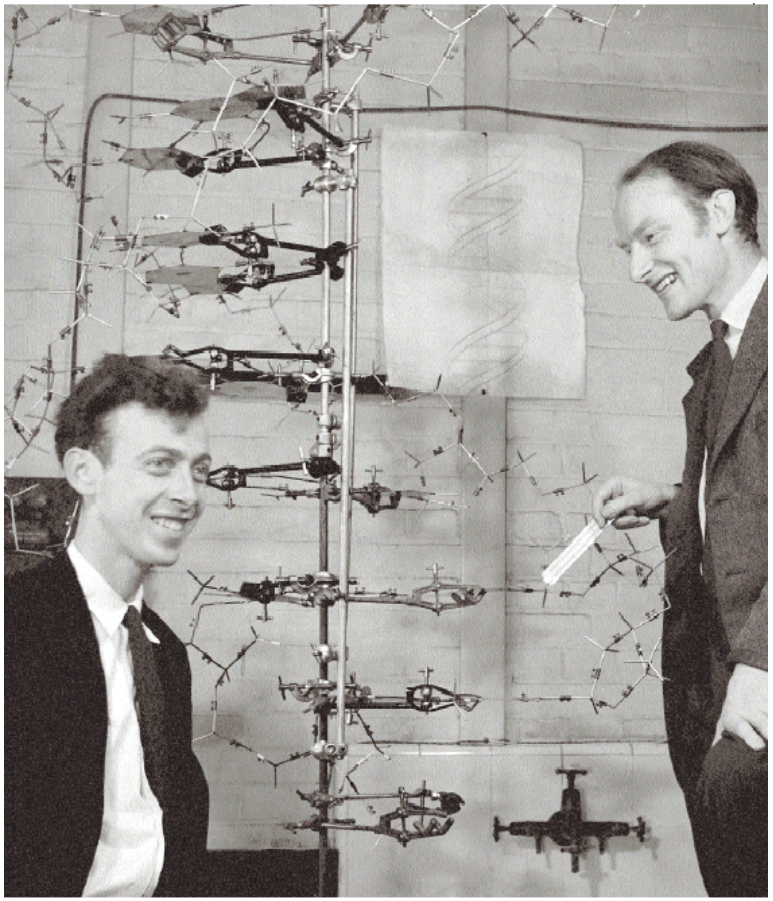


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

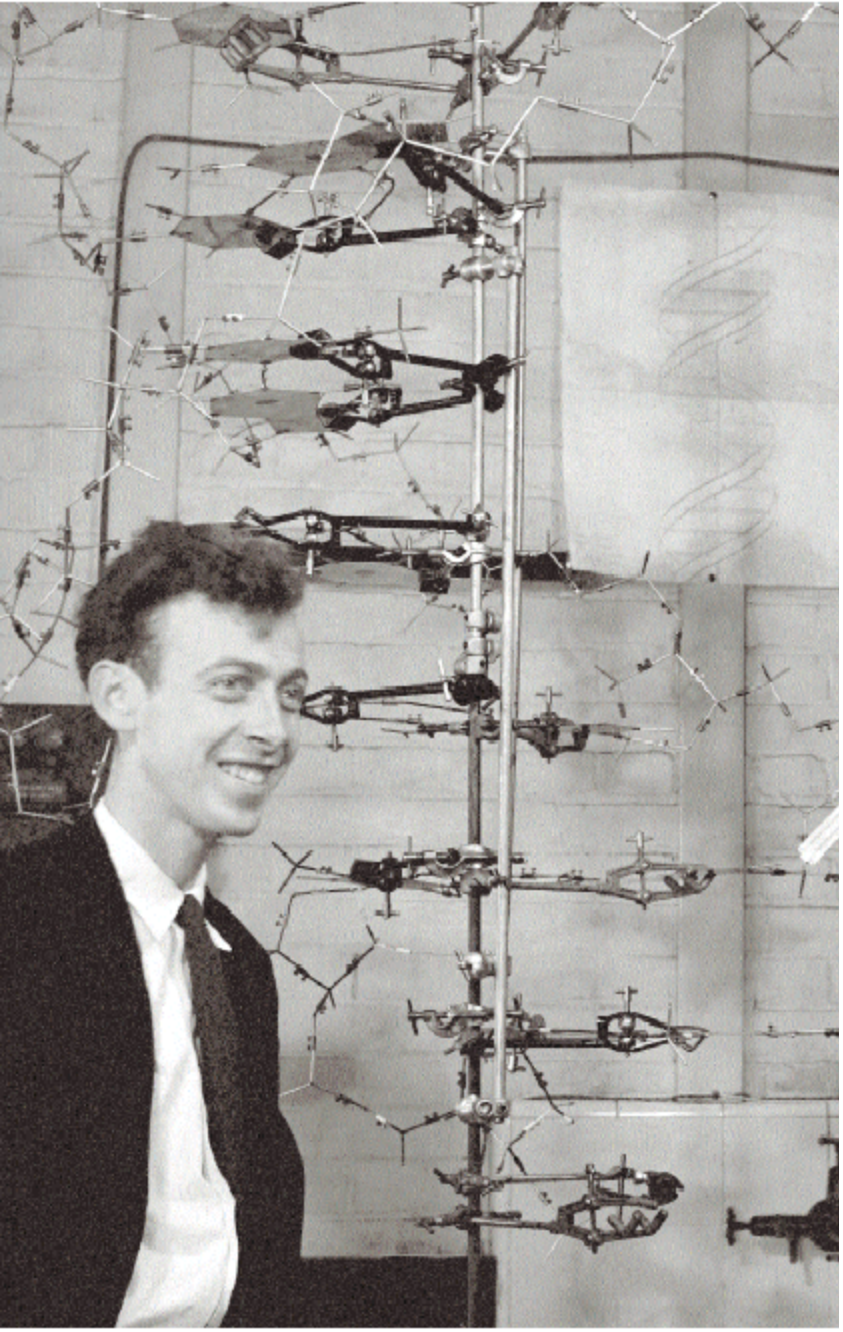
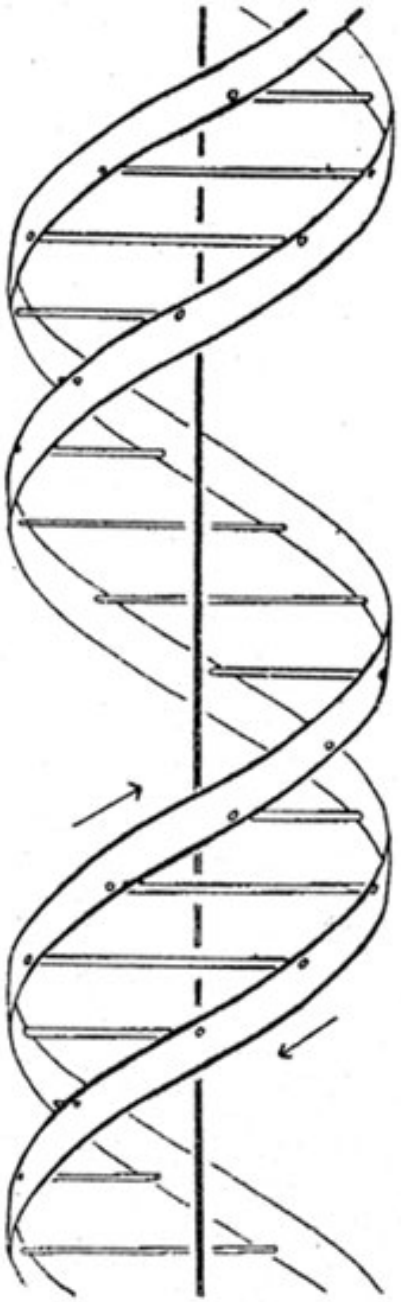


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DN
Laboratory in Cambridge, 21 May 1953.





is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

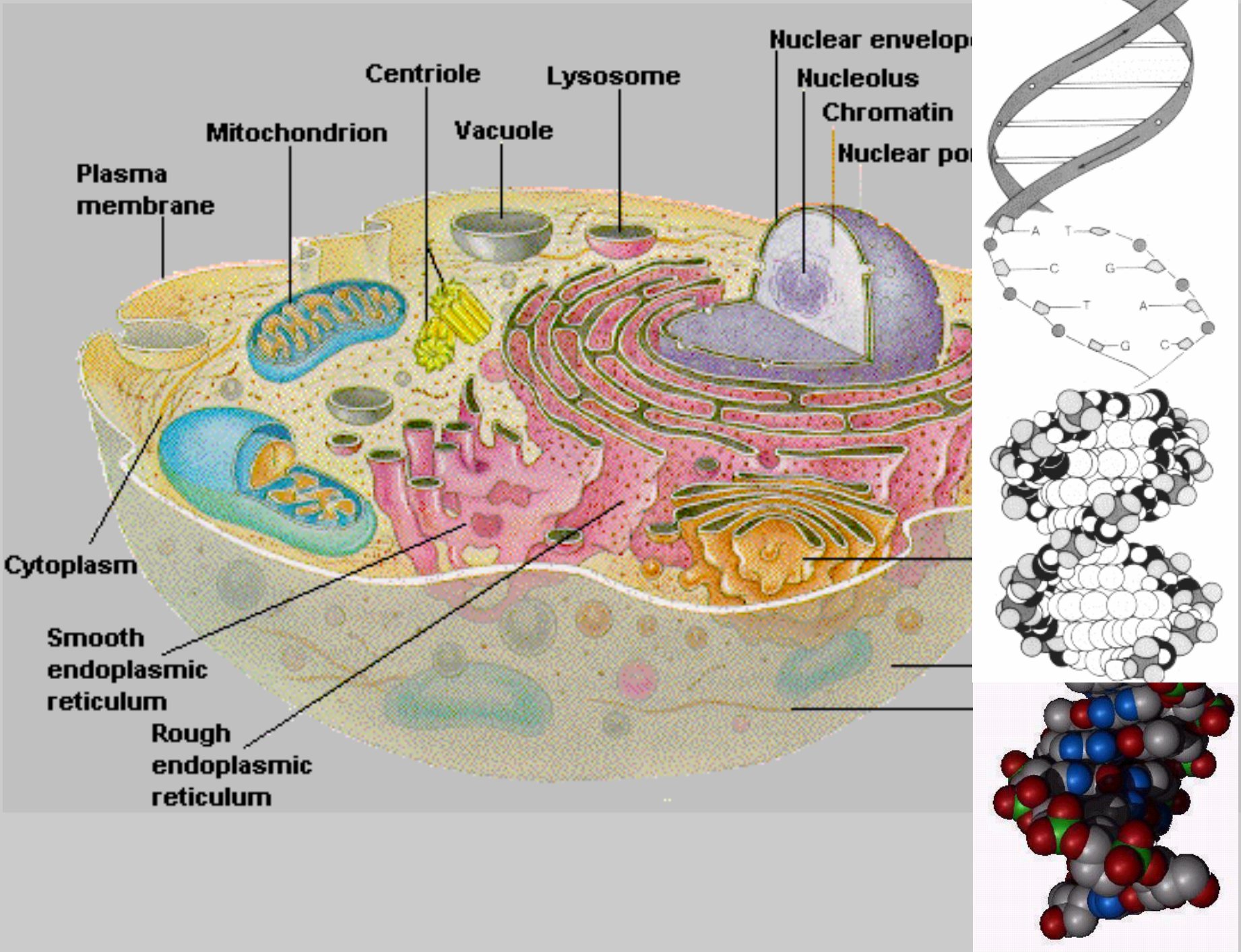
The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

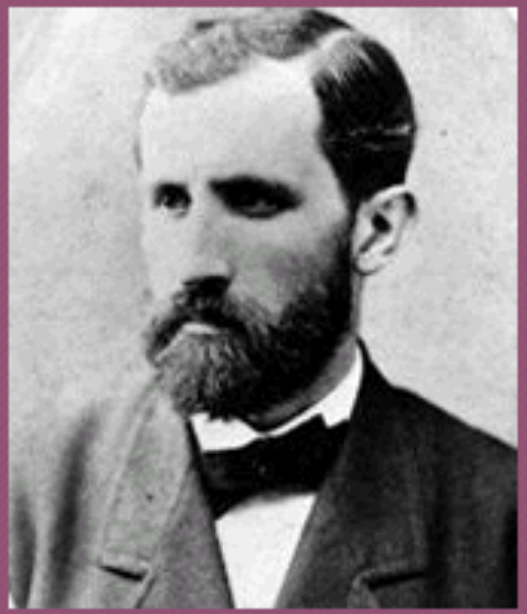
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at


1962, the Nobel Prize in Physiology or Medicine

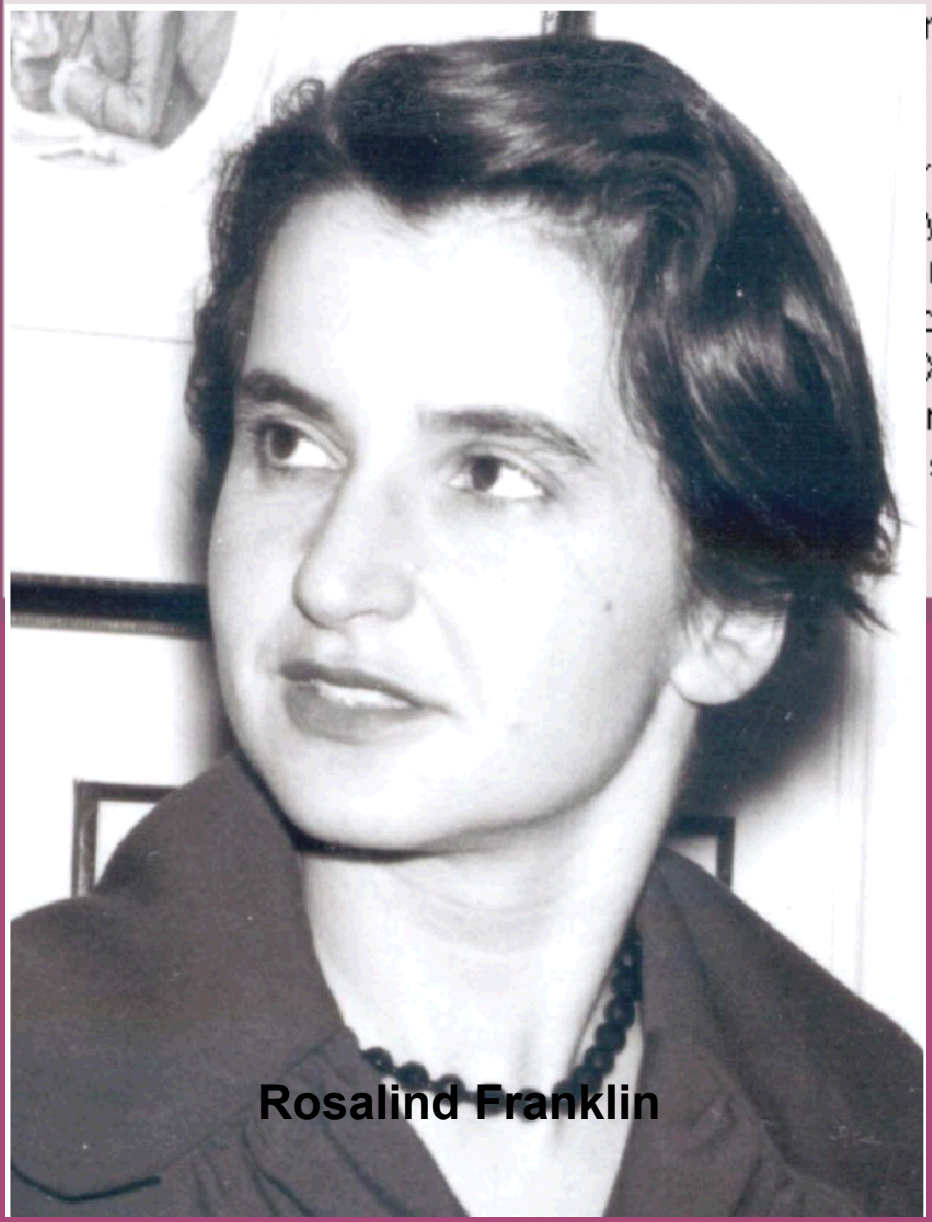

Nuclear envelop
Centriole
Lysosome
Nucleolus
Chromatin
Mitochondrion
Vacuole
Nuclear po
Plasma membrane
Cytoplasm
Smooth endoplasmic reticulum
Rough endoplasmic reticulum
A
T
C
G
T
A
G
C

# 1869 DNA first isolated



1879 picture of the laboratory where Miescher isolated nuclein. The lab, a part of the University of Tübingen in southern Germany, was run by Felix Hoppe-Seyler, and located in the vaults of an old castle

# 1943 X-ray diffraction of DNA

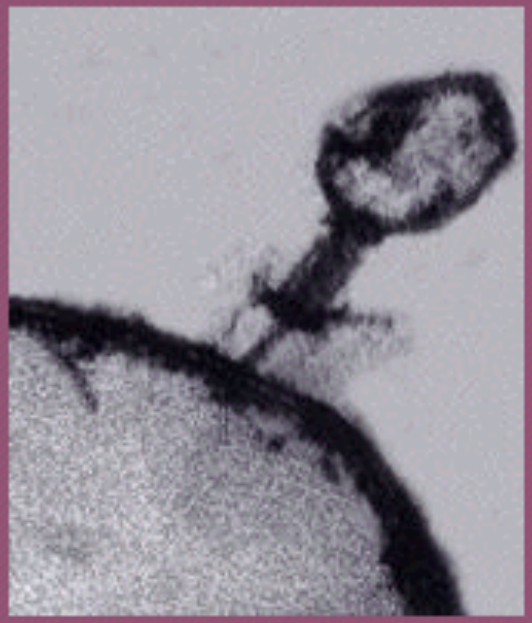


Rosalind Franklin

ray diffraction pattern

veal their
y diffraction patterns
n dipped a needle
d containing many
X-ray diffraction
regular, periodic
stacked on top of

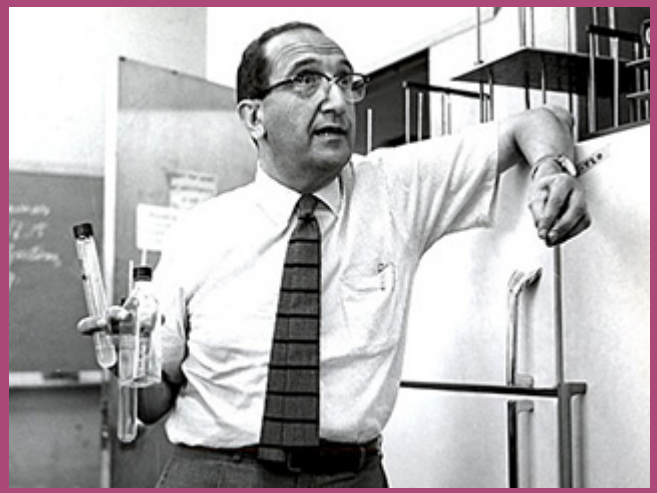
# 1952 Genes are made of DNA



Alfred Hershey and
needs to enter a ba

Their experiment p
of DNA. They firmly
tentatively propose

Electron microscop
bacteriophage T4 -
Chase figured that
to direct the produc



## The Nobel Prize in Physiology or Medicine 1969

"for their discoveries concerning the replication mechanism and the genetic structure of viruses"

**Max Delbrück**

1/3 of the prize

USA

California Institute of Technology
Pasadena, CA, USA

b. 1906
(in Berlin, Germany)
d. 1981

**Alfred D. Hershey**

1/3 of the prize

USA

Carnegie Institution of Washington
Long Island, New York, NY, USA

b. 1908
d. 1997

**Salvador E. Luria**

1/3 of the prize

USA

Massachusetts Institute of Technology (MIT)
Cambridge, MA, USA

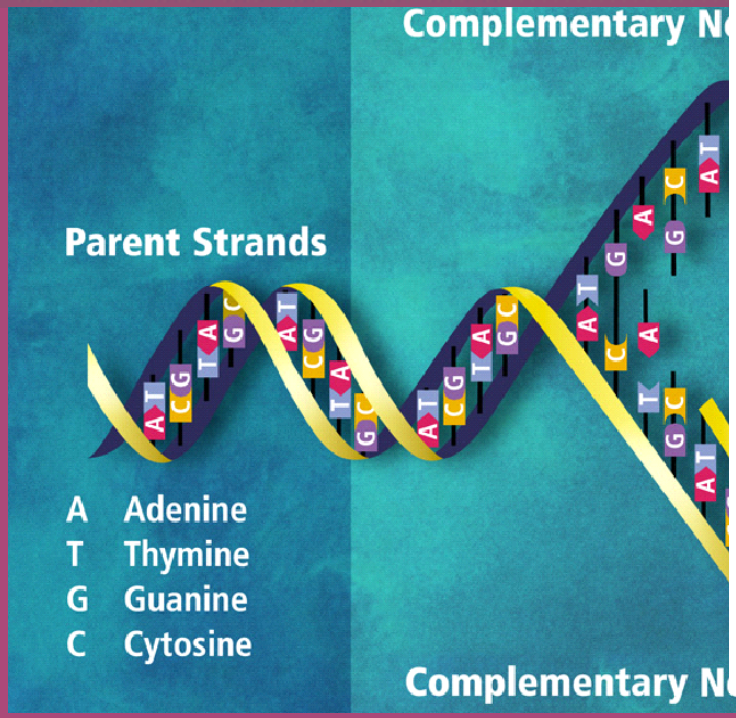
b. 1912
(in Torino, Italy)
d. 1991

# 1955 DNA copying enzyme


Arthur Kornbe
later used for

Kornberg's gr
added the pro
nucleotide bui

The enzyme th
polymerase in
repair DNA. Hi
job of replicati




## The Nobel Prize in Physiology or Medicine 1959

"for their discovery of the mechanisms in the biological synthesis of ribonucleic acid and deoxyribonucleic acid"

**Severo Ochoa**

1/2 of the prize

USA

New York University; College of Medicine
New York, NY, USA

b.1905
(in Luarca, Spain)
d.1993

**Arthur Kornberg**

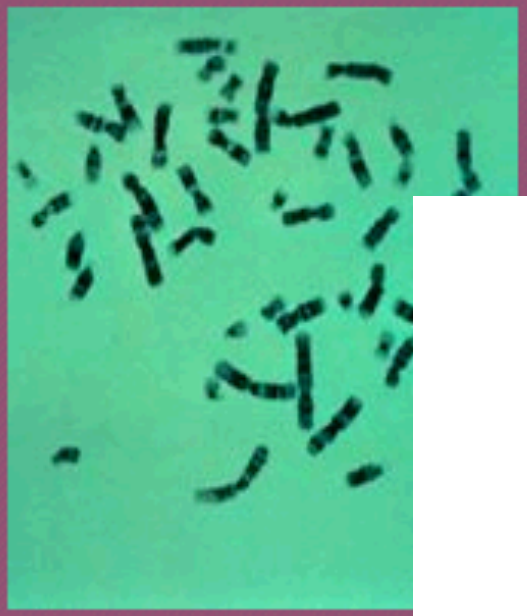
1/2 of the prize

USA

Stanford University
Stanford, CA, USA

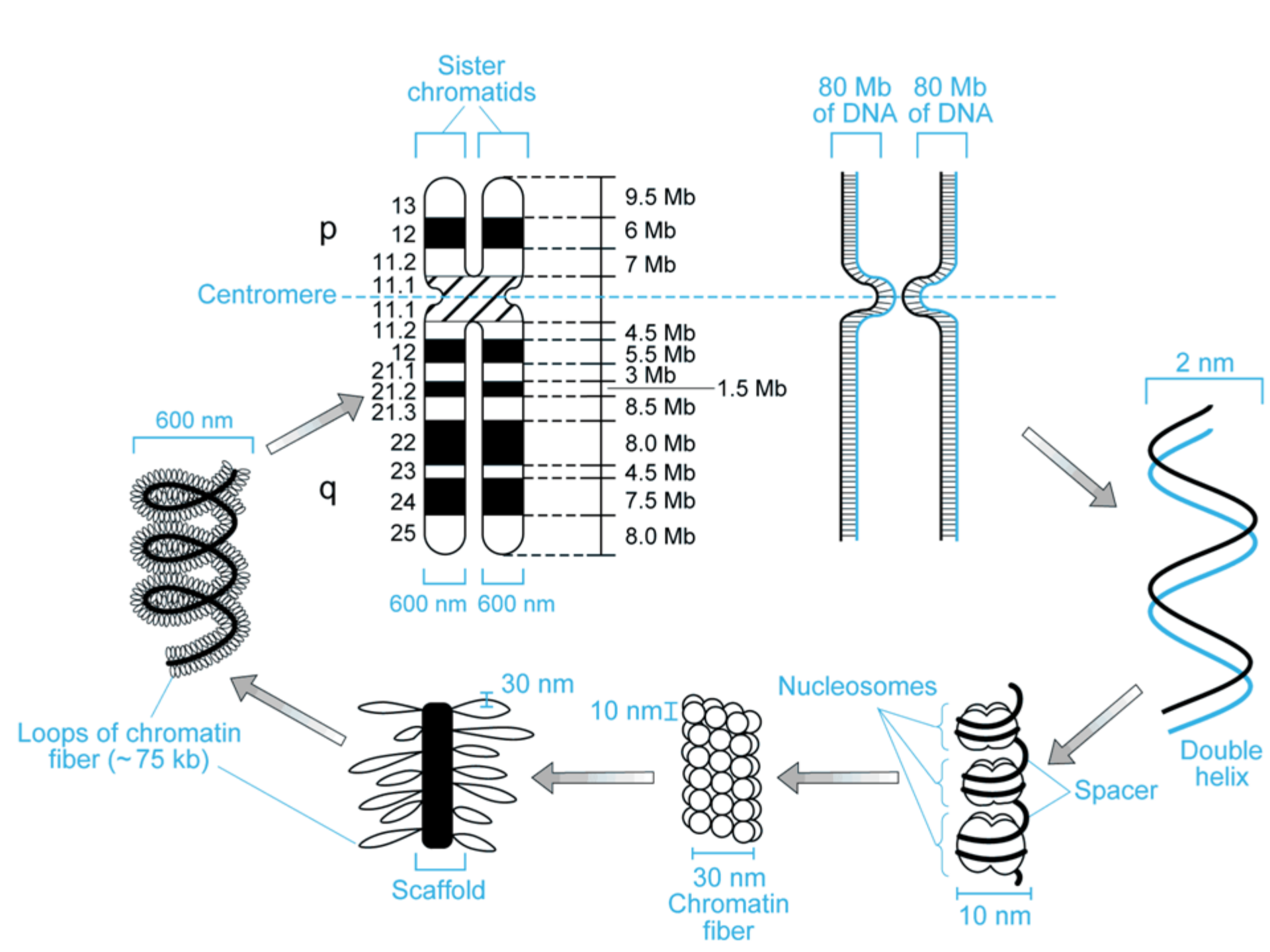
b.1918

# 1955 46 human chromosomes



Joe Hin Tjio defined 46 as the exact number of human chromosomes.

Tjio was interested in the chromosomes of cancer cells and how


Sister chromatids
80 Mb of DNA
80 Mb of DNA
p
q
Centromere
13
12
11.2
11.1
11.1
11.2
12
21.1
21.2
21.3
22
23
24
25
9.5 Mb
6 Mb
7 Mb
4.5 Mb
5.5 Mb
3 Mb
1.5 Mb
8.5 Mb
8.0 Mb
4.5 Mb
7.5 Mb
8.0 Mb
600 nm
600 nm
600 nm
2 nm
Double helix
Spacer
Nucleosomes
10 nm
10 nm
30 nm Chromatin fiber
30 nm
Scaffold
Loops of chromatin fiber (~75 kb)

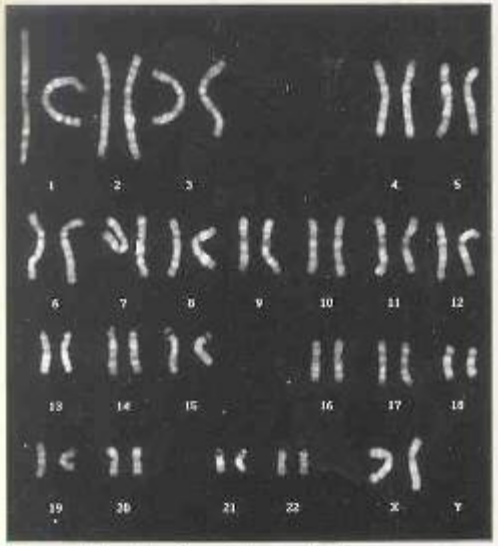


Figura 10a *Cariotipo femminile normale, 46xx*

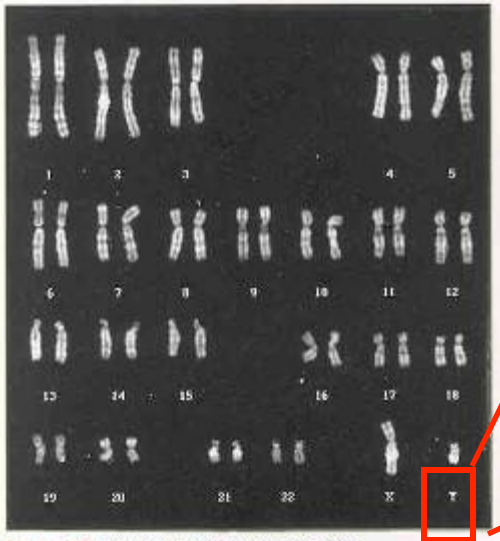
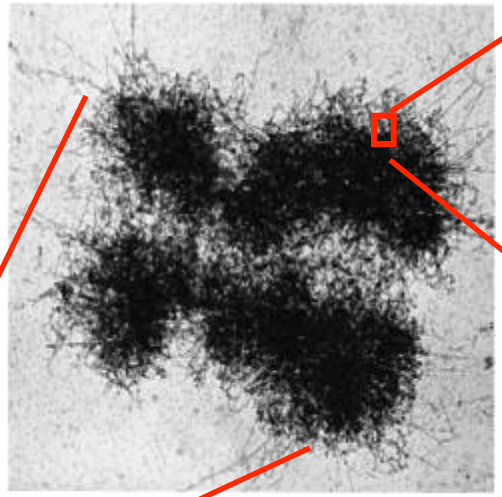
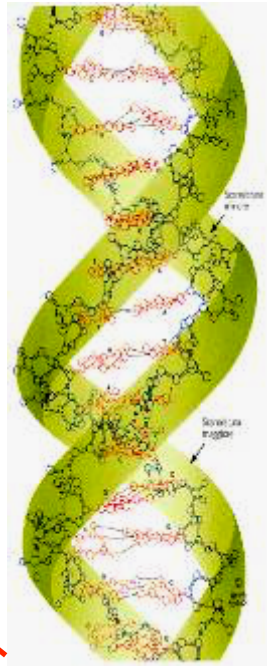


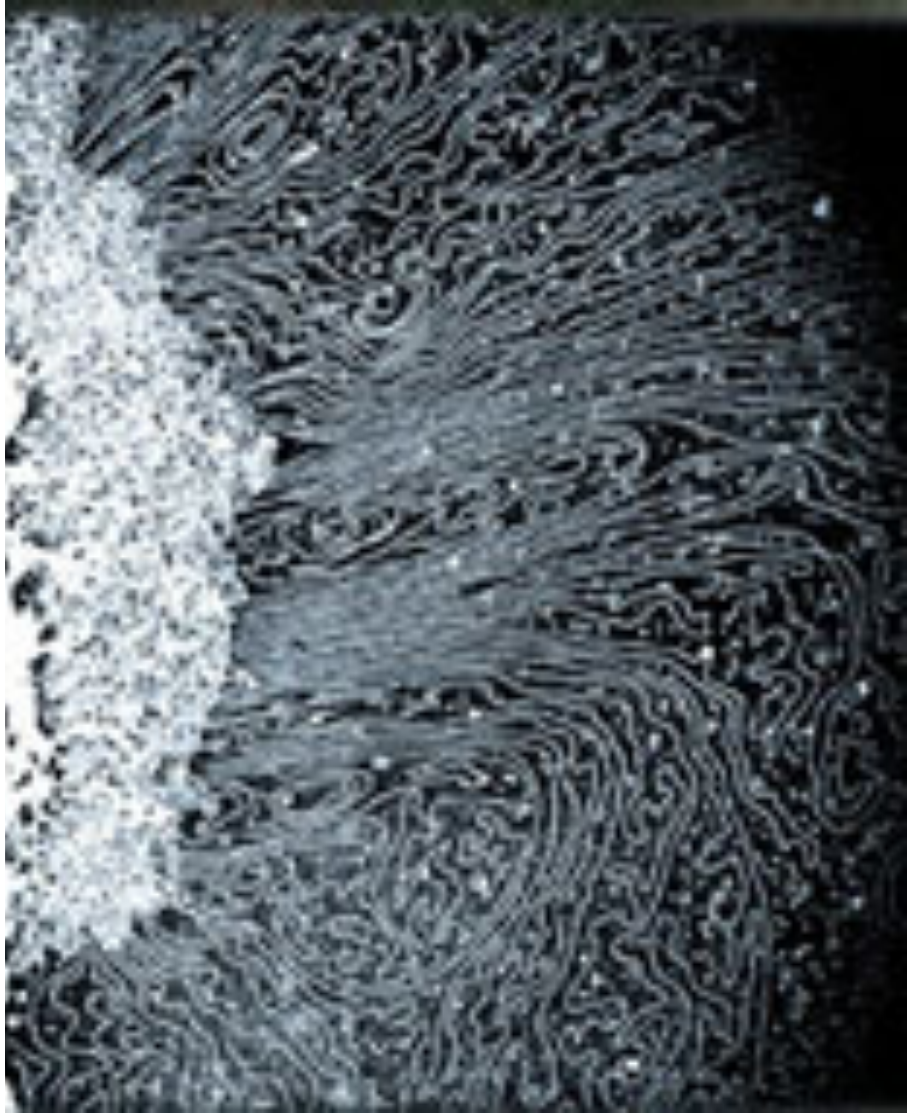
Figura 10b *Cariotipo maschile normale, 46xy*



A chromosome



A DNA tract (gene)

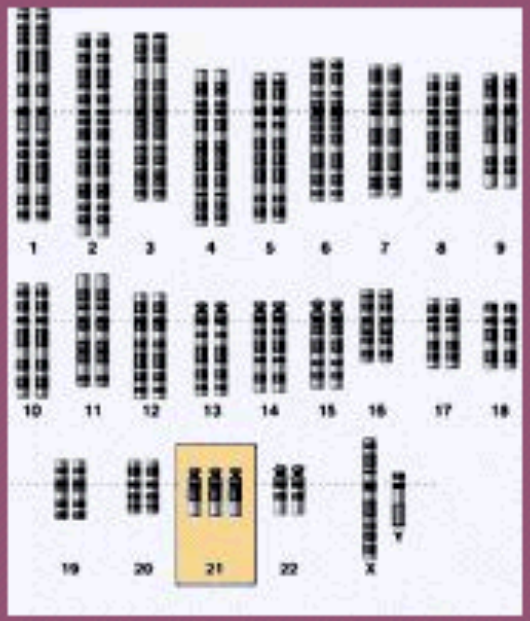


# 1956 Cause of disease traced to alteration

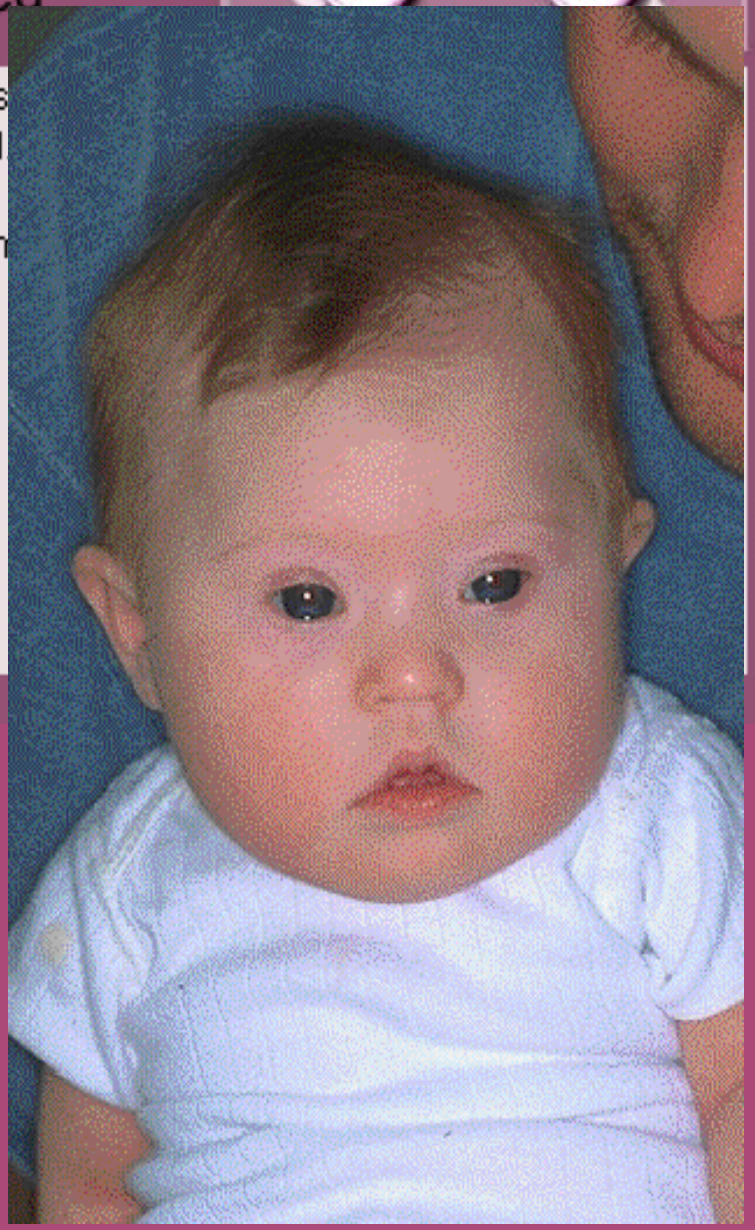
Sickle cell anemia was first described in 1910, after Ernest E. Irons, an intern at Chicago's Presbyterian Hospital, noticed elongated cells in a blood smear from Walter C. Noel, a dental student with severe anemia.

In the late 1940s, Linus Pauling and his colleagues realized that sickle cell anemia stems from a change in the structure of hemoglobin. In 1956, Vernon Ingram discovered that a specific chemical alteration in a hemoglobin protein — the substitution of valine for glutamic acid in the sixth amino acid in beta globin — is the root of the disease. Other mutations in hemoglobin can also cause sickle cell anemia, and were sorted out later.

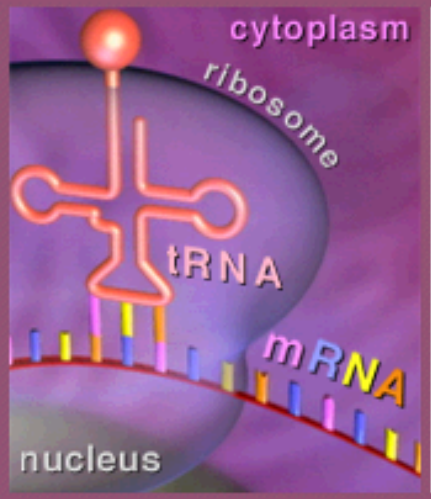
# 1959 Chromosome abnormalities identified




Professor Jerome Lejeune and his
syndrome, first classified by J. L. H
— that is, having three instead of 2
copies of the genes on chromoson
and body.



# 1961 mRNA ferries information

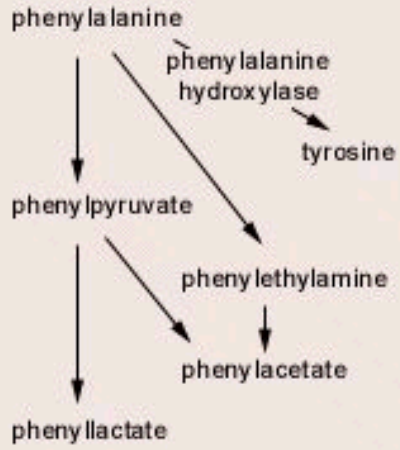



Sydney Brenner, Francois Jacc
mRNA is the molecule that tak
the protein-making machinery

When DNA was determined to
wondered how DNA, which is s
formation of proteins, when pr
is the intermediary molecule. S
ferries information from the nu

RNA is chemically similar to D

# 1961 1st screen for metabolic defec




In 1961, Robert Guthrie, a c
Buffalo Children's Hospital
babies have phenylketonur
acid phenylalanine. He test
the blood by seeing whethe
growth of a bacterial strain
grow.

Excess phenylalanine in th
mental retardation. But if ba
on a diet that avoids the amino acid phenylalanine and evade almost all
symptoms of the condition. Mass screening for PKU became routine

## The Nobel Prize in Physiology or Medicine 2002

"for their discoveries concerning 'genetic regulation of organ development and programmed cell death'"

**Sydney Brenner**
1/3 of the prize
United Kingdom
The Molecular Sciences Institute
Berkeley, CA, USA
b. 1927
(in Union of South Africa)

**H. Robert Horvitz**
1/3 of the prize
USA
Massachusetts Institute of Technology (MIT)
Cambridge, MA, USA
b. 1947

**John E. Sulston**
1/3 of the prize
United Kingdom
The Wellcome Trust Sanger Institute
Cambridge, United Kingdom
b. 1942

# 1966 Genetic code cracked

Nirenberg's notes, on elucidation of the first "word" of the genetic code. (Nirenberg still works at NIH.) →

## Genetic code table

| | | | |
|---|---|---|---|
| UUU Phenyl-alanine | UCU Serine | UAU Tyrosine | UGU Cysteine |
| UUC | UCC | UAC | UGC |
| UUA Leucine | UCA | UAA Terminator | UGA Termination |
| UUG | UCG | UAG | UGG Tryptophan |
| CUU Leucine | CCU Proline | CAU Histidine | CGU Arginine |
| CUC | CCC | CAC | CGC |
| CUA | CCA | CAA Glutamine | CGA |
| CUG | CCG | CAG | CGG |
| AUU Isoleucine | ACU Threonine | AAU Asparagine | AGU Serine |
| AUC | ACC | AAC | AGC |
| AUA | ACA | AAA Lysine | AGA Arginine |
| AUG Methionine (Initiator) | ACG | AAG | AGG |
| GUU Valine | GCU Alanine | GAU Aspartic acid | GGU Glycine |
| GUC | GCC | GAC | GGC |
| GUA | GCA | GAA Glutamic acid | GGA |
| GUG | GCG | GAG | GGG |

## The Nobel Prize in Physiology or Medicine 1968

"for their interpretation of the genetic code and its function in protein synthesis"

**Robert W. Holley**

1/3 of the prize

USA

Cornell University
Ithaca, NY, USA

b. 1922
d. 1993

**Har Gobind Khorana**

1/3 of the prize

USA

University of Wisconsin
Madison, WI, USA

b. 1922
(in Raipur, India)

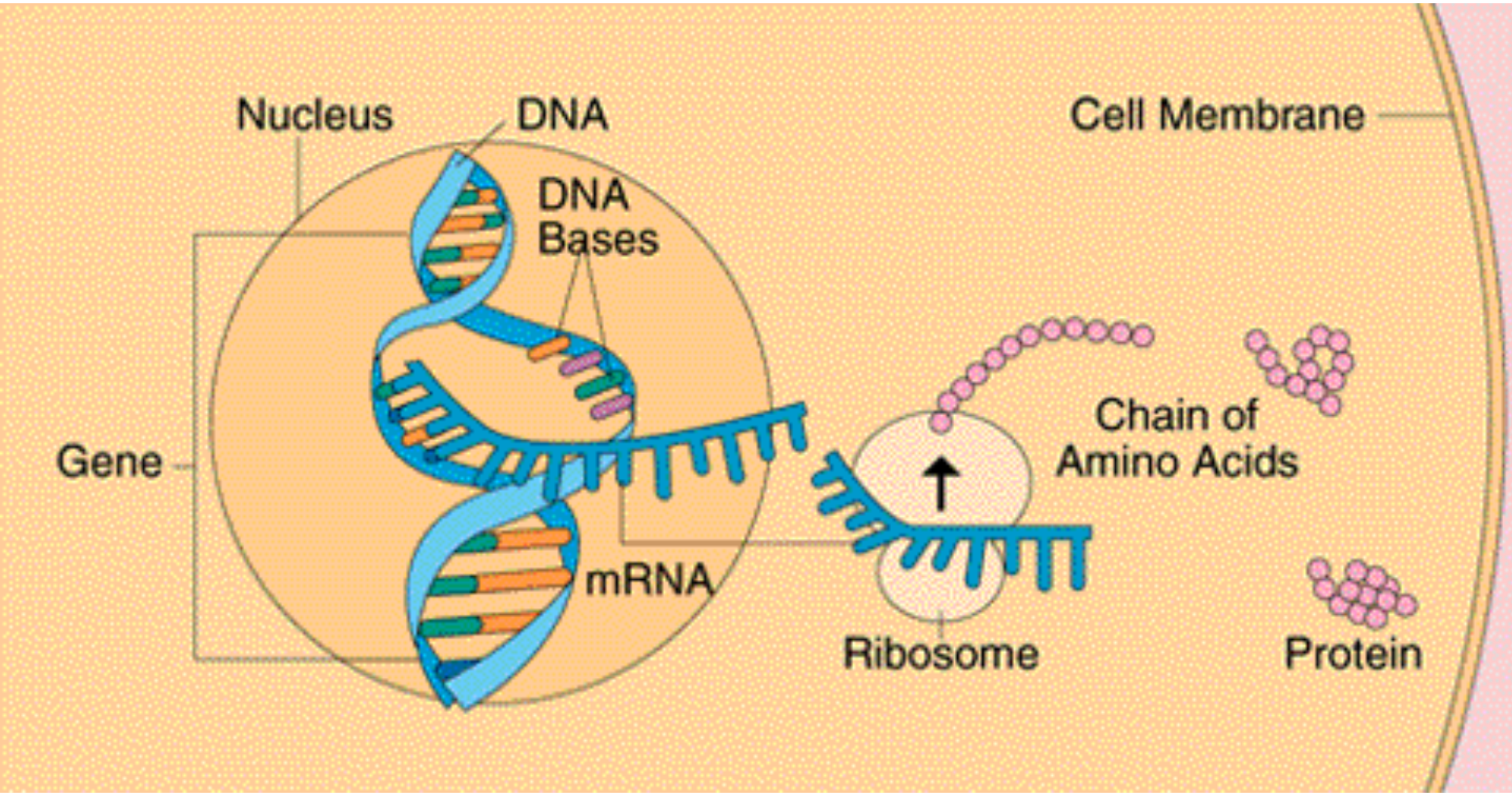
**Marshall W. Nirenberg**

1/3 of the prize

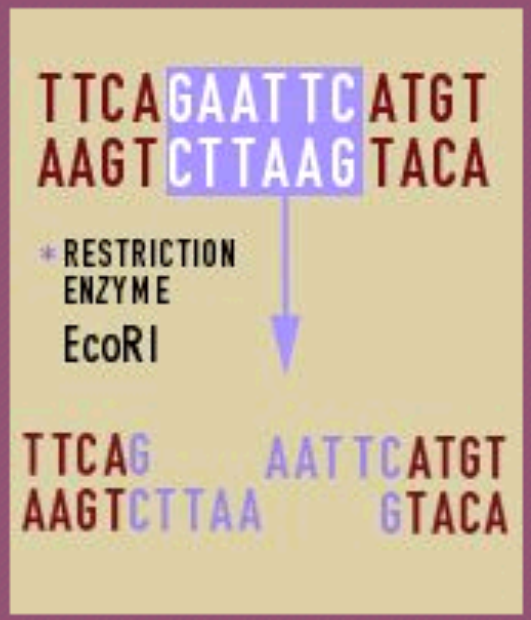
USA

National Institutes of Health
Bethesda, MD, USA

b. 1927


Nucleus
DNA
Cell Membrane
DNA
Bases
Gene
Chain of
Amino Acids
mRNA
Ribosome
Protein

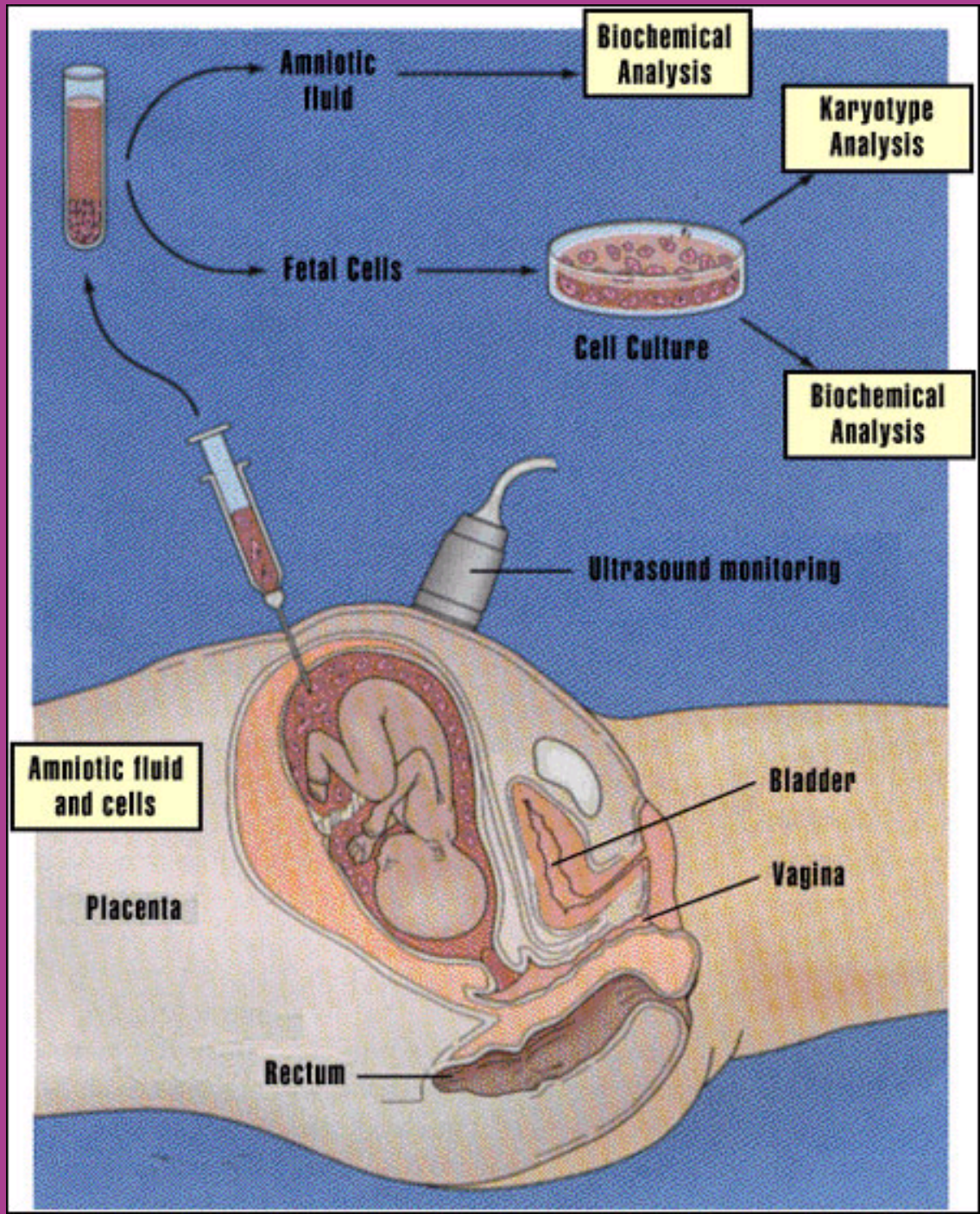
# 1968 First restriction enzymes described



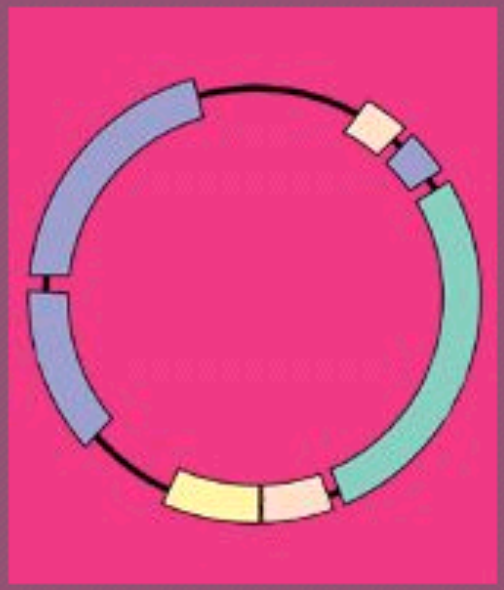

Several groups of researchers — including M. Meselson's group at Harvard and H. O. Smith, K. W. Wilcox, and T. J. Kelley at Johns Hopkins — studied and characterized the first restriction nucleases, enzymes that revolutionized molecular biologists' ability to manipulate DNA.

Restriction enzymes recognize and cut specific short sequences of DNA. They're found in bacteria, which use the enzymes to digest invading DNA. The bacteria add methyl groups to their own DNA to protect them from digestion. Molecular biologists began using these enzymes, along with DNA polymerase and DNA ligase (an enzyme that sticks fragments of DNA together), in the early 1970s to cut, manipulate, and analyze pieces of DNA in a predictable and reproducible way. The enzymes became an important,

From 1970 is introduced the option of prenatal diagnosis of Down syndrome on amniotic fluid through chromosome analysis



# 1972 First recombinant DNA



The first p
enzymes,

Recombi
species a
bacterium
enzymes
the cut st
Stanford
DNA tech
Genentec

## The Nobel Prize in Physiology or Medicine 1986

"for their discoveries of growth factors"

**Stanley Cohen**

1/2 of the prize

USA

Vanderbilt University School of Medicine
Nashville, TN, USA

b. 1922

**Rita Levi-Montalcini**

1/2 of the prize

Italy and USA

Institute of Cell Biology of the C.N.R
Rome, Italy

b. 1909
(in Turin, Italy)

# 1975-77 DNA seque

Sanger a
sequenci

Sanger a
sequenci
where a r
commonl
ends of D

## The Nobel Prize in Chemistry 1980

"for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA"

"for their contributions concerning the determination of base sequences in nucleic acids"

**Paul Berg**

1/2 of the prize

USA

Stanford University
Stanford, CA, USA

b. 1926

**Walter Gilbert**

1/4 of the prize

USA

Biological Laboratories
Cambridge, MA, USA

b. 1932

**Frederick Sanger**

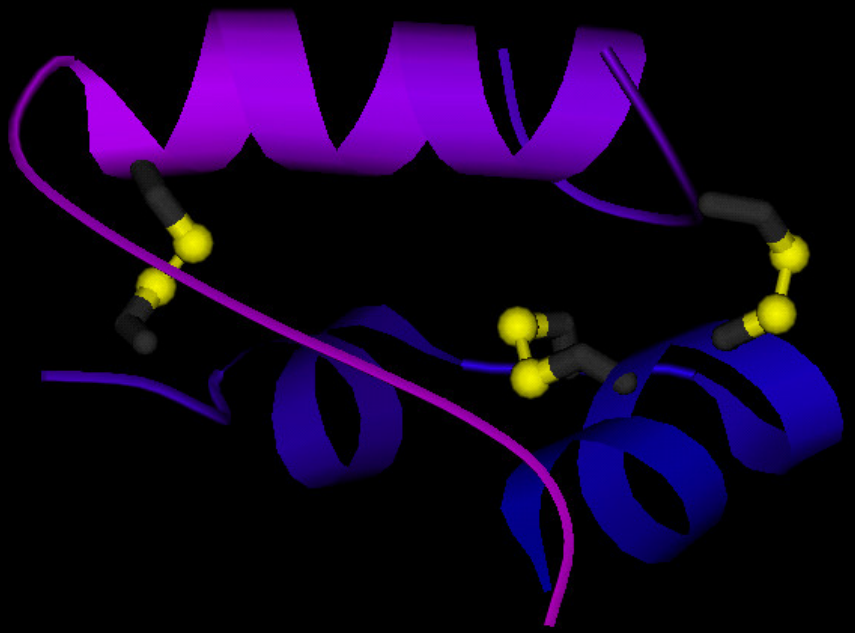
1/4 of the prize

United Kingdom

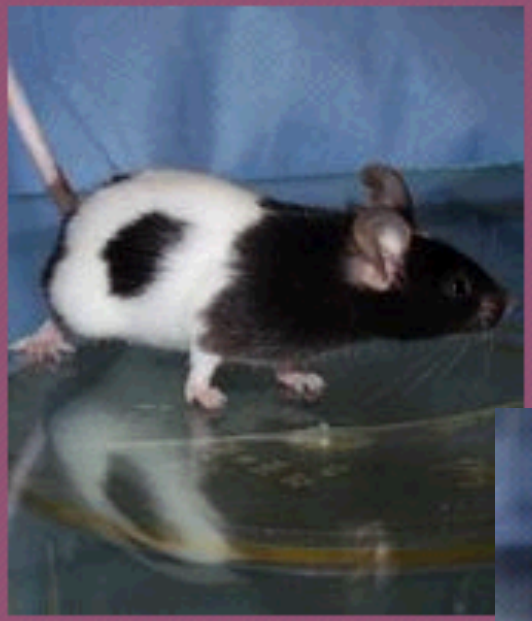
MRC Laboratory of Molecular Biology
Cambridge, United Kingdom

b. 1918

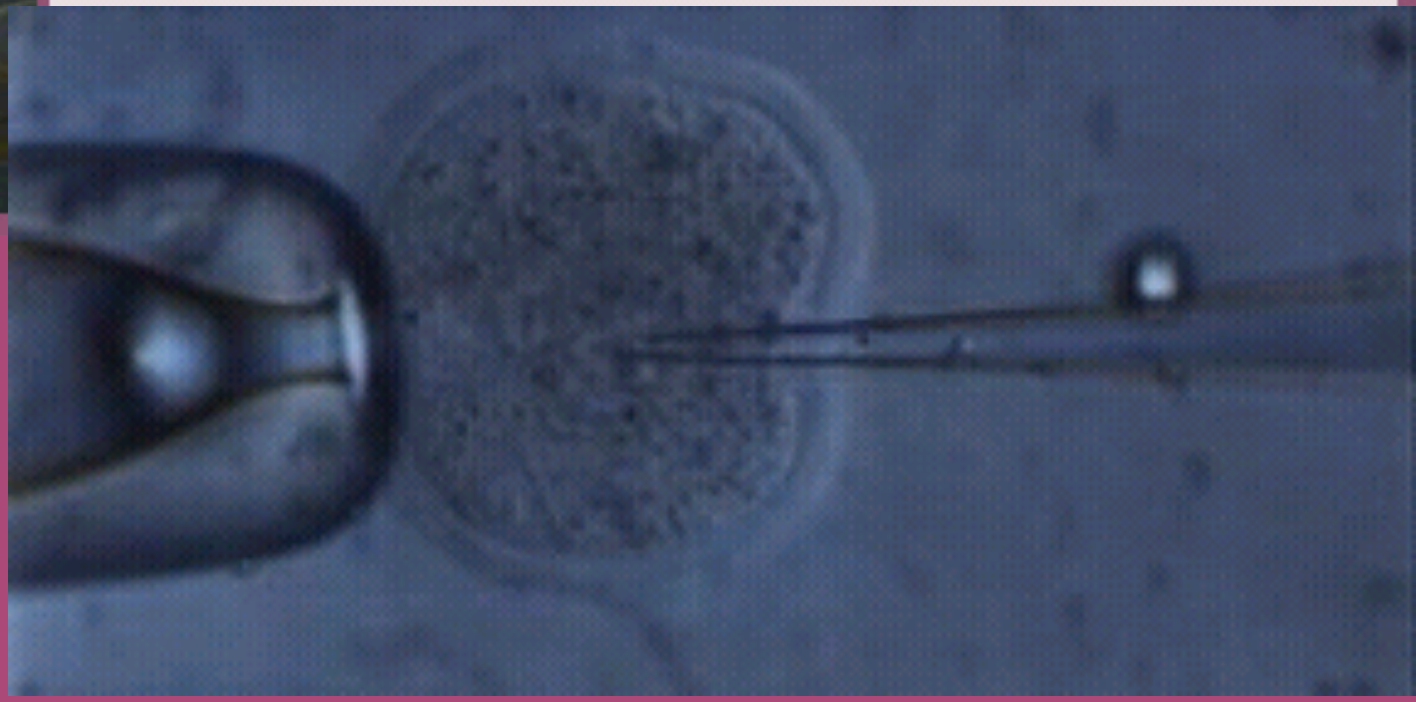
In 1980, the first human biosynthetic insulin has been injected (recombinant DNA produced by Eli Lilly: genetic engineering on the cell of the bacterium Escherichia coli) to Sandy Atherton, 37 years old, from Wichita, Kansas.



# 1981-82 First transgenic mice and fruit flies



Scientists had been able to add new genes to bacterial cells for several years. In the early 1980s, they figured out how to add stably-inherited new genes to animals. The first such "transgenic animals" were mice and fruit flies. By adding foreign genes or genes spelled slightly differently than normal, scientists had a new way to test the functions of genes.



# The Nobel Prize in Physiology or Medicine 2007

"for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells"

Photo: U. Montan

**Mario R. Capecchi**

1/3 of the prize

USA

University of Utah
Salt Lake City, UT, USA;
Howard Hughes Medical Institute

b. 1937
(in Italy)

Photo: U. Montan

**Sir Martin J. Evans**

1/3 of the prize

United Kingdom

Cardiff University
Cardiff, United Kingdom

b. 1941

Photo: U. Montan

**Oliver Smithies**

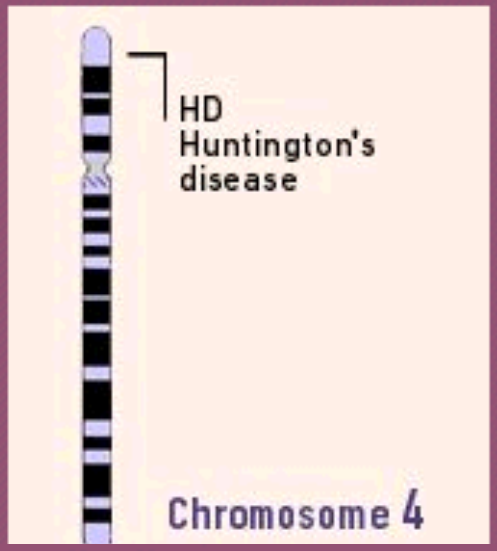
1/3 of the prize

USA

University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

b. 1925
(in United Kingdom)

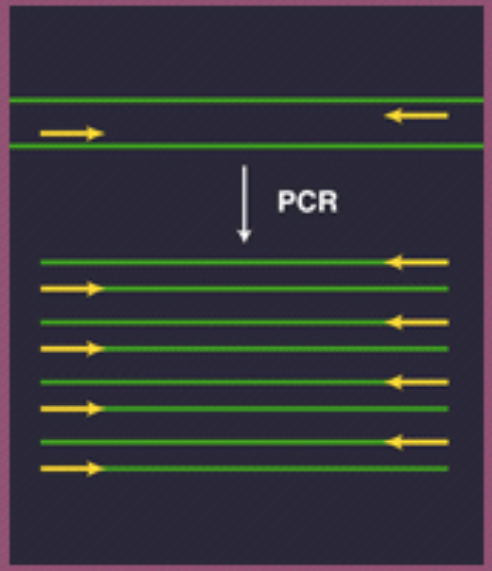
# 1983 First disease gene mapped




A genetic marker linked to Huntington's disease was found on chromosome 4 in 1983, making Huntington's disease, or HD, the first genetic disease mapped using DNA polymorphisms.

HD is inherited as an autosomal dominant disease. (In other words, only one of the two copies of the gene needs to be mutated to cause disease.) HD causes the death of specific neurons, leading — usually in midlife — to characteristic jerky movements, physical rigidity, and dementia, symptoms that worsen progressively. The gene was finally isolated in 1993. A collaborative group of 58 researchers in 6 research groups isolated a gene called *huntingtin* located on chromosome 4p16.3.

# 1983 PCR invented



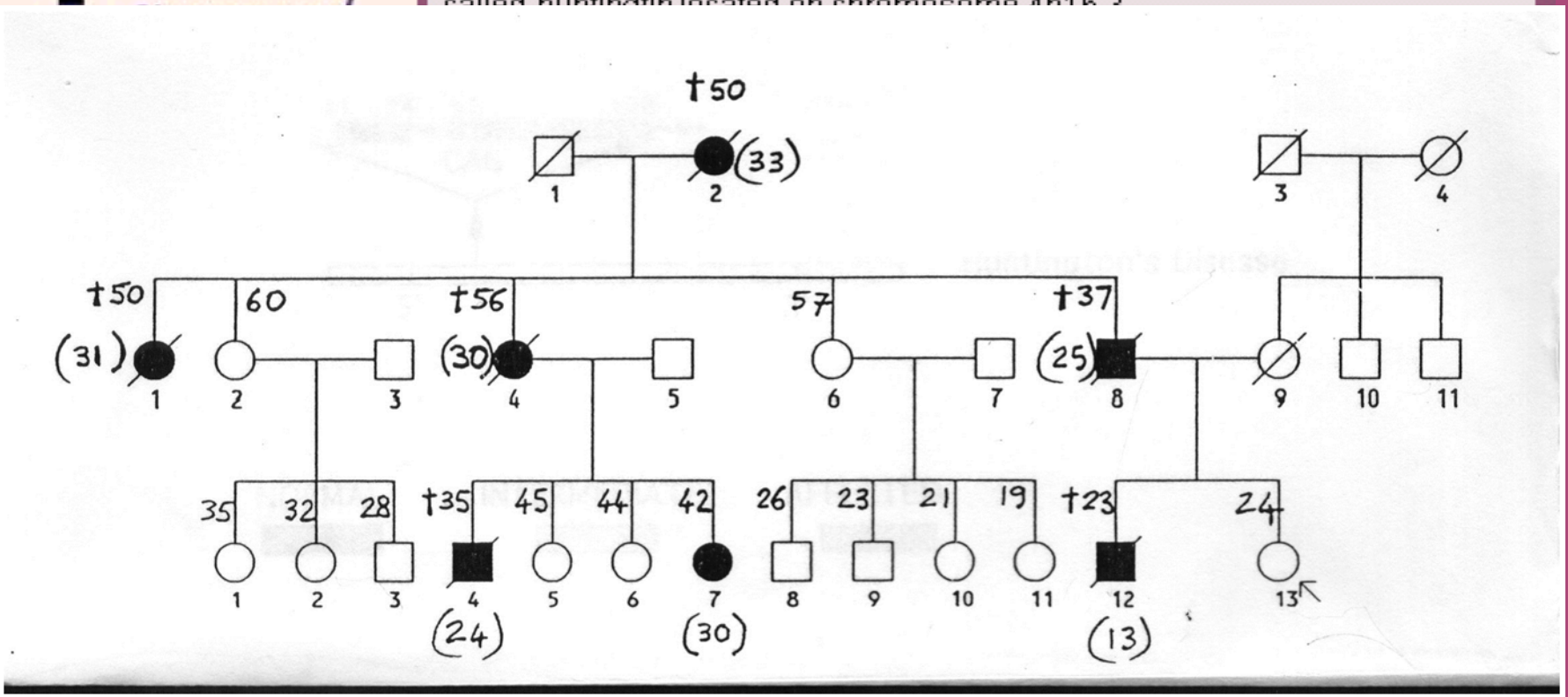

PCR — the polymerase chain reaction — is a technique for amplifying DNA that dramatically boosted the pace of genetic research. In a matter of a few hours, PCR can make billions of copies of a specific segment of DNA.

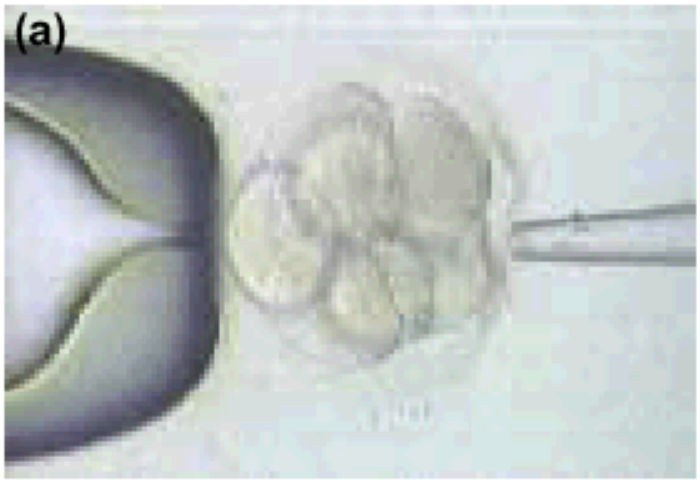
The 1993 Nobel Prize in Chemistry was given for the invention of PCR.

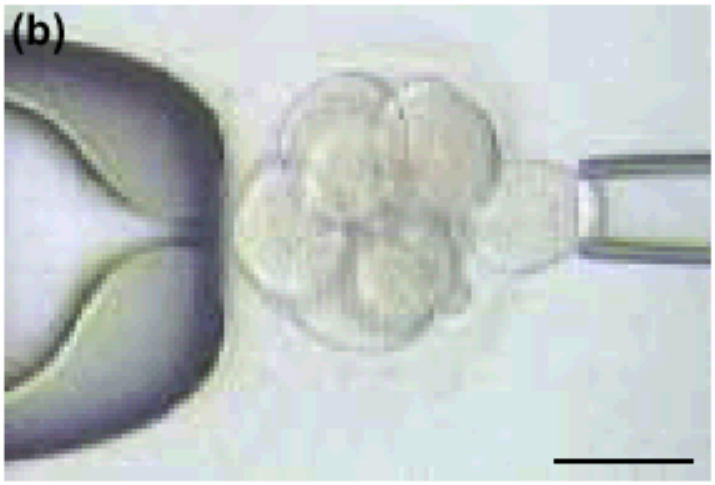
# 1983 First disease gene mapped


A genetic marker linked to Huntington's disease was found on chromosome 4 in 1983, making Huntington's disease, or HD, the first genetic disease mapped using DNA polymorphisms.

HD is inherited as an autosomal dominant disease. (In other words, only one of the two copies of the gene needs to be mutated to cause disease.) HD causes the death of specific neurons, leading — usually in midlife — to characteristic jerky movements, physical rigidity, and dementia, symptoms that worsen progressively. The gene was finally isolated in 1993. A collaborative group of 58 researchers in 6 research groups isolated a gene called huntingtin located on chromosome 4p16.3






ased

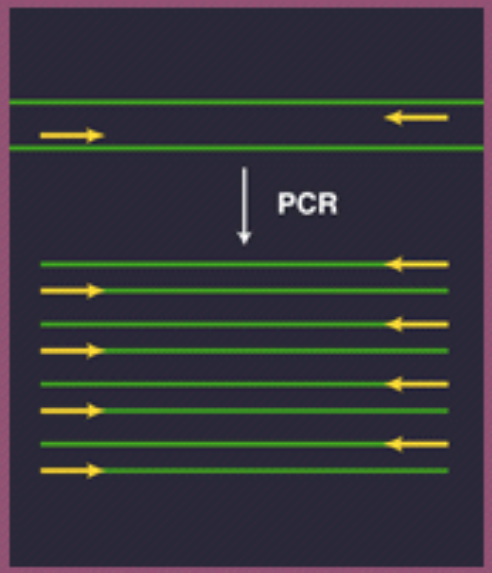
utions

directed
pment

Chromosome 4

called *hunting*

# 1983 PCR invented




PCR — the poly
that dramaticall
hours, PCR car

The 1993 Nobe

**Kary B. Mullis**

1/2 of the prize

USA

La Jolla, CA, USA

b. 1944

**Michael Smith**

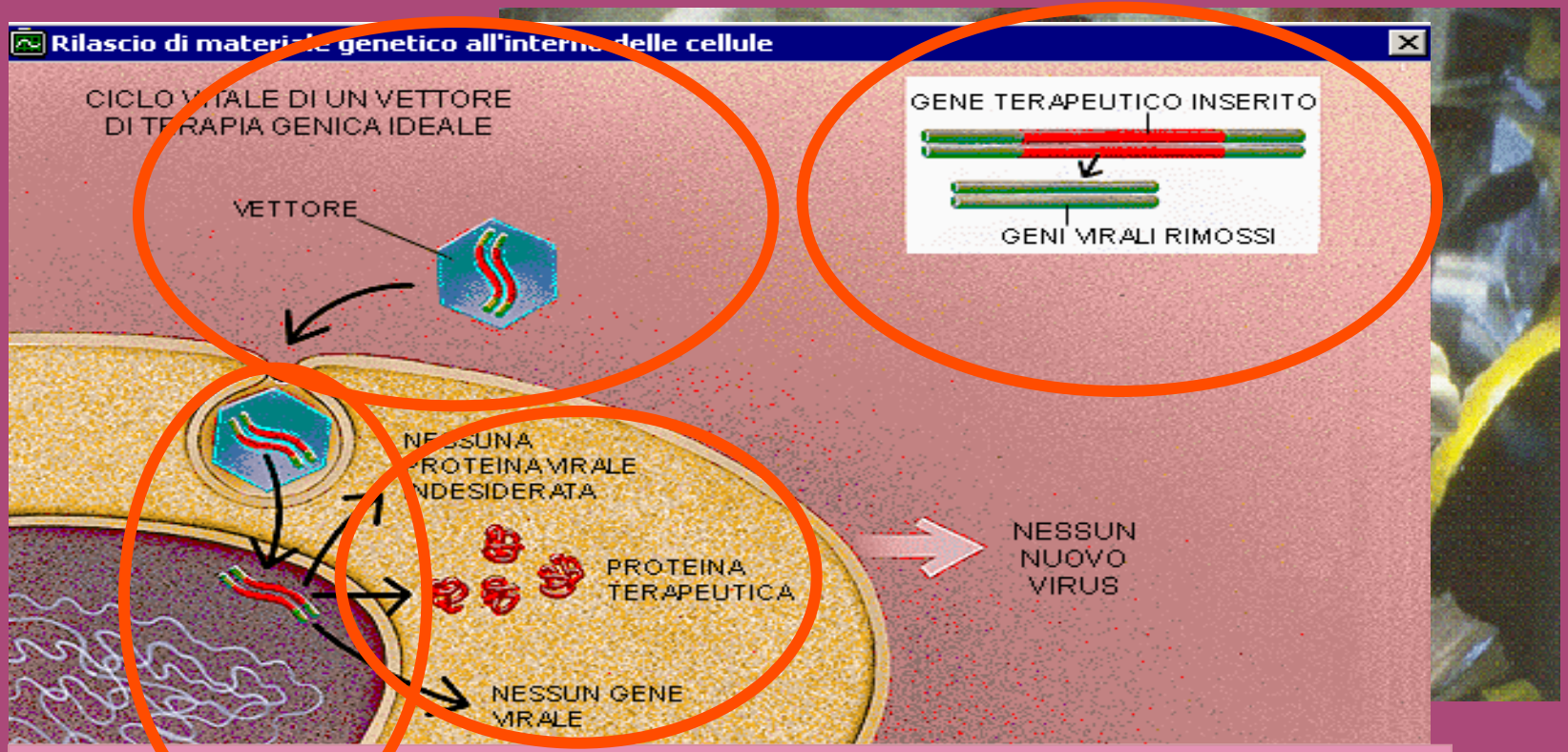
1/2 of the prize

Canada

University of British Columbia
Vancouver, Canada

b. 1932
(in Blackpool, United Kingdom)
d. 2000

# 1990: Michael Blaese in the United States applies the first gene therapy procedure on a child with a SCID, a hereditary severe combined immunodeficiency



# 1990 Launch of the Human Genome Project

The Human Genome Project officially began in 1990. Beginning in December 1984, the U.S. Department of Energy (DOE), National Institutes of Health (NIH), and international groups had sponsored meetings to consider the feasibility and usefulness of mapping and sequencing the human genome. The DOE had become interested in studying the human genome as a way of aiding the detection of mutations that nuclear radiation might cause. Groups like the NIH and the Wellcome Trust in Britain had longstanding interest in understanding biology for the sake of advancing medicine.

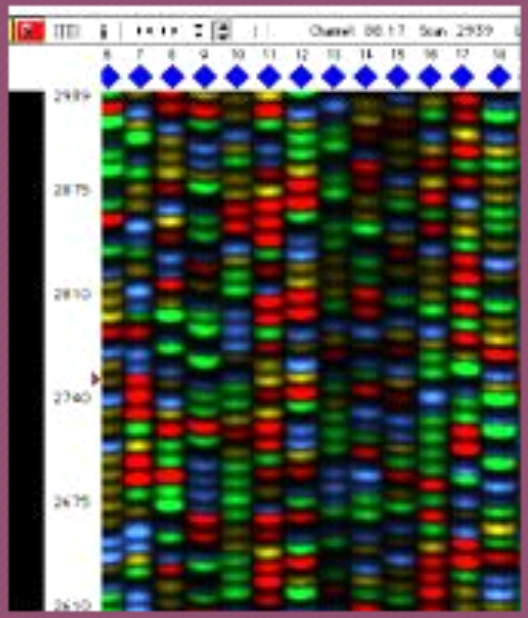
In 1987, DOE had proposed a Human Genome Initiative to Congress. Meanwhile, NIH had started funding occasional grants for genome

# 1994 FLAVR SAVR tomato

The FDA approved the sale of the first genetically modified food — the FLAVR SAVR tomato, deeming it as safe as conventionally-bred tomatoes.

The FDA's decision on the FLAVR SAVR tomato — marketed by Calgene, Inc. of Davis, California — marked the first time the agency evaluated a food that was genetically engineered. FLAVR SAVR tomatoes are modified to stay firm after harvest, so they can be left on the vine longer before shipping. The FDA decided the change in the tomatoes was not great enough to warrant mandated labeling describing the alteration.

# 1996 Human DNA sequencing begins



In 1996, the National Human Genome Research Institute funded pilot projects to find efficient strategies for completely sequencing the human genome. The pilot projects tested the feasibility of large-scale sequencing, and explored how accurate and how costly alternative approaches might be.

# 1998 Company announces sequencing plan

In May 1998, the company Celera Genomics was formed to sequence much of the human genome in three years. Unlike the Human Genome Project, which built detailed maps before sequencing defined regions, Celera used a shotgun sequencing strategy, in which the entire genome is fragmented and random segments are sequenced and then put in order.

# 1999 Chromosome 22


In December 1999, the Human Genome Project completed the first finished, full-length sequence of a human chromosome — chromosome 22. This accomplishment demonstrated the ability of the Human Genome Project method of clone-by-clone sequencing to obtain large amounts of highly accurate sequence. In the clone-by-clone approach, clones of human DNA, such as bacterial artificial chromosomes (BACs), that have a precisely known location on a physical map are the starting points for DNA sequencing reactions. Researchers chose to finish chromosome 22 first because it is relatively small and because highly detailed maps of 22 had already been constructed.

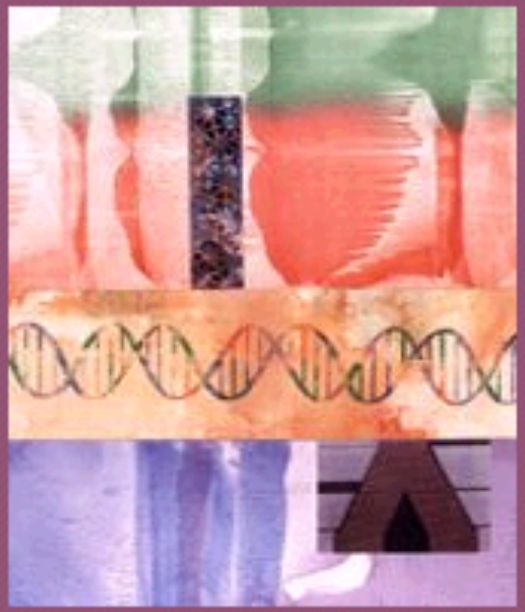
The sequence of chromosome 22 gave scientists their first ever view of the

# 2000 Chromosome 21


In May 2000, Human Genome Project scientists led by German and Japanese teams described the finished genome sequence of human chromosome 21, the second human chromosome to be fully sequenced. An extra copy of chromosome 21 causes Down syndrome, and genes on the chromosome have been linked to diseases like Alzheimer's and certain forms of cancer.

The published sequence contained approximately 33,500,000 base pairs. Chromosome 21 appears to have less than 300 genes.

# 2000 Free access to genomic information



In March 2000, U.S. President Clinton and U.K. Prime Minister Tony Blair stated that raw, fundamental data about human genome sequence and its variations should be freely available.

The President and Prime Minister Blair issued a Joint Statement in an effort to ensure that the public derive the maximum possible benefit from the sequence of the human genome. They agreed that the best course of action would be to urge the following: First, that raw fundamental sequence data — the letters that make up the book of human life — be distributed as widely as possible without barriers to its use. And, second, that private investment in gene-based technologies be encouraged, so this fundamental knowledge can be turned into useful medical products as

# GENETICS, GENES AND GENOMES

- The last 60 years of genetics
- Human genome and other genomes
- Genetic diseases
- Genetic polymorphisms

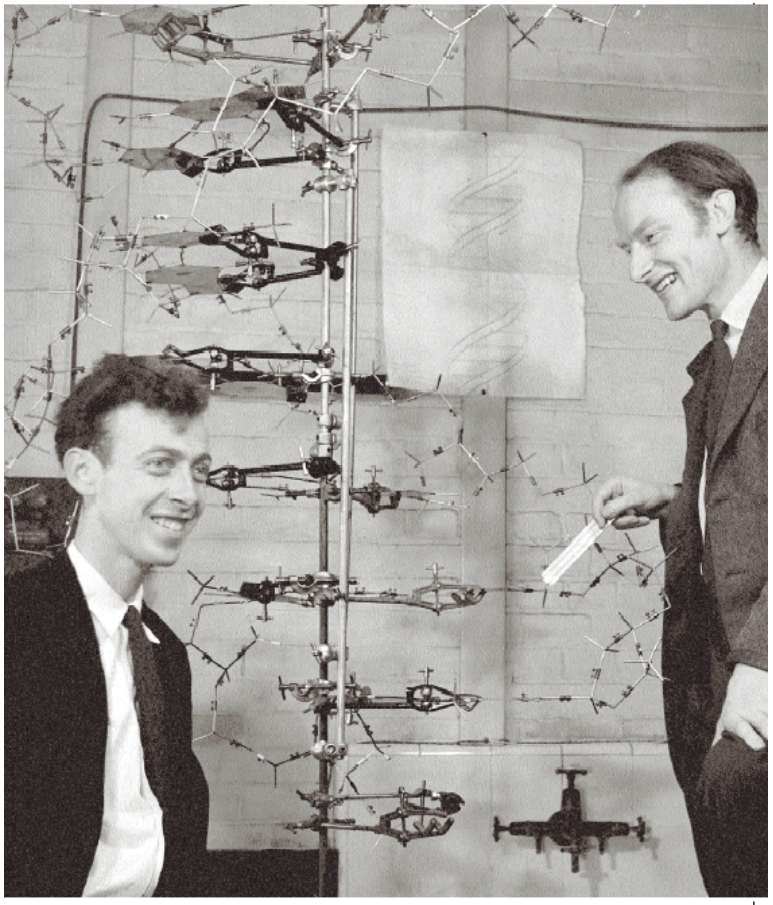


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.


15 February 2001
nature
www.nature.com
the human genome
Nuclear fission
Five-dimensional energy landscapes
Seafloor spreading
The view from under the Arctic ice
Career prospects
Sequence creates new opportunities
naturejobs
genomics special


Science
THE HUMAN GENOME


PRESIDENT OF THE UNITED

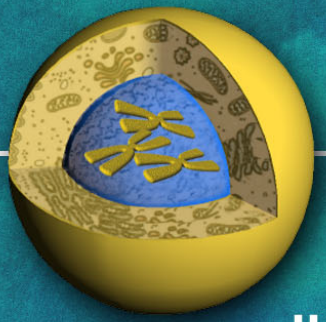
# DNA the molec

## Trillions of cells

Each cell:

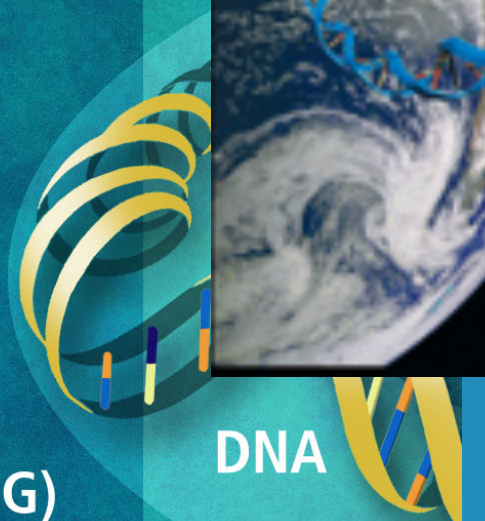
- 46 human chromosomes
- 2 meters of DNA
- 3 billion DNA subunits (the bases: A, T, C, G)
-

How many genes?

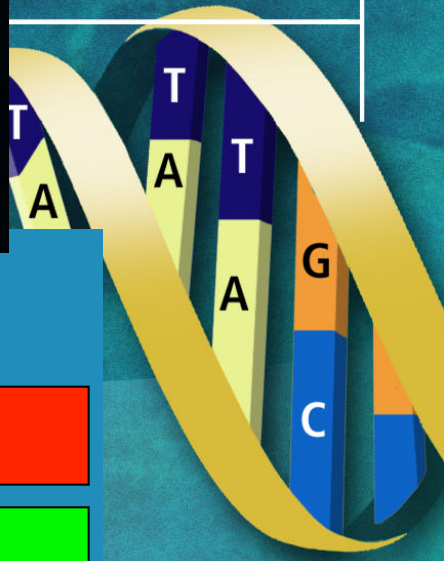


cell

ne



DNA




250.000

25.000

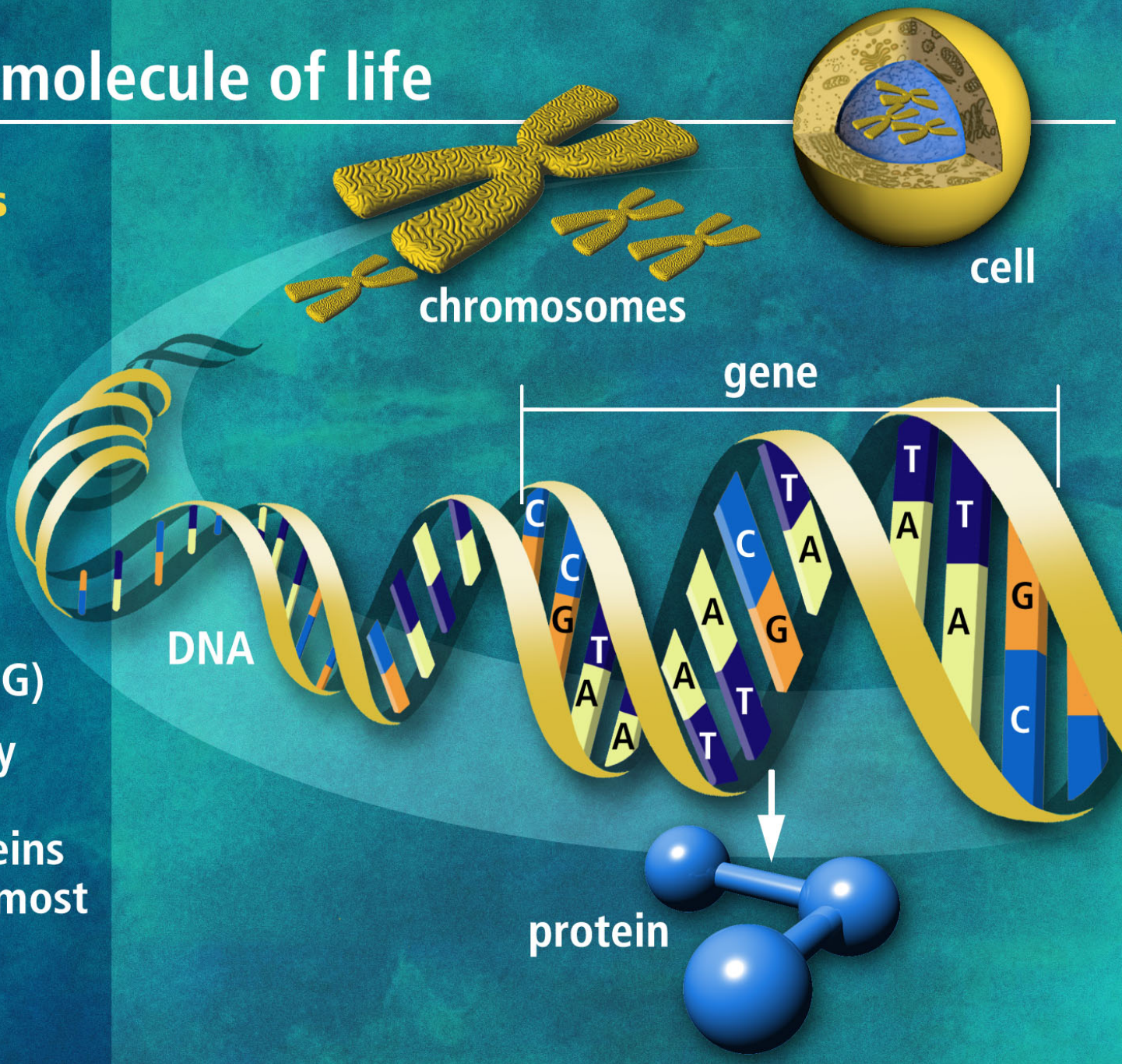


# DNA the molecule of life

## Trillions of cells

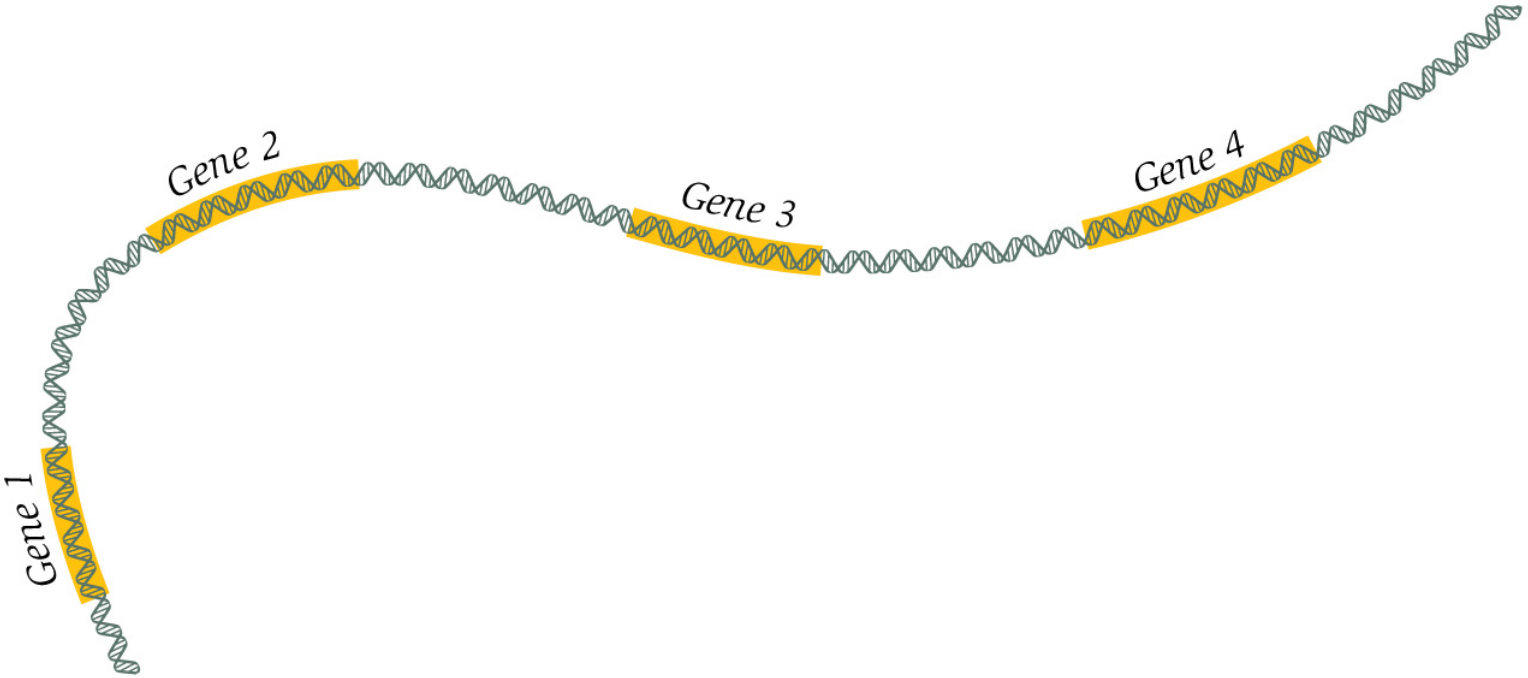
Each cell:

- 46 human chromosomes
- 2 meters of DNA
- 3 billion DNA subunits (the bases: A, T, C, G)
- Approximately 30,000 genes code for proteins that perform most life functions

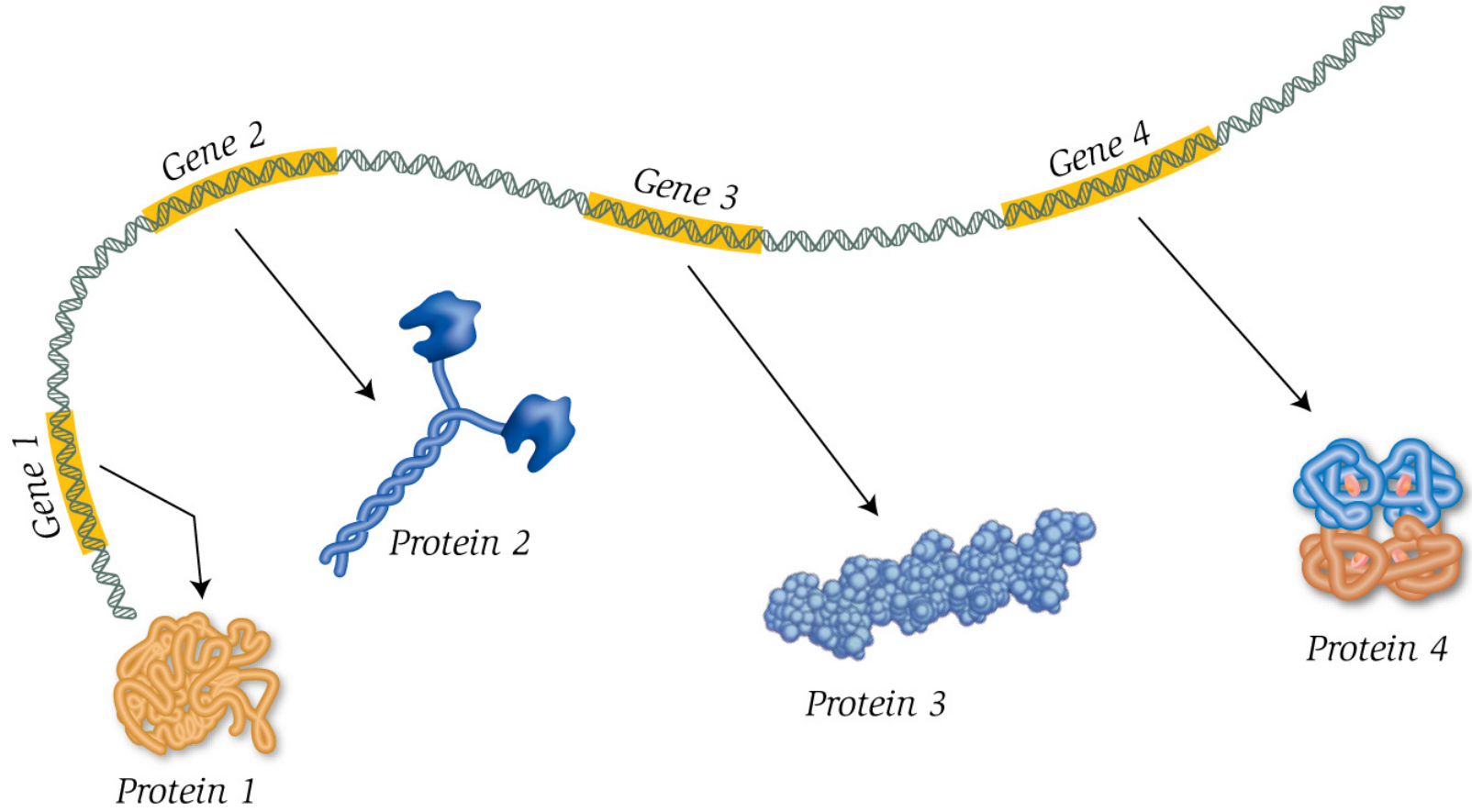



Y-GG 01-0085

# DNA is organized into genes



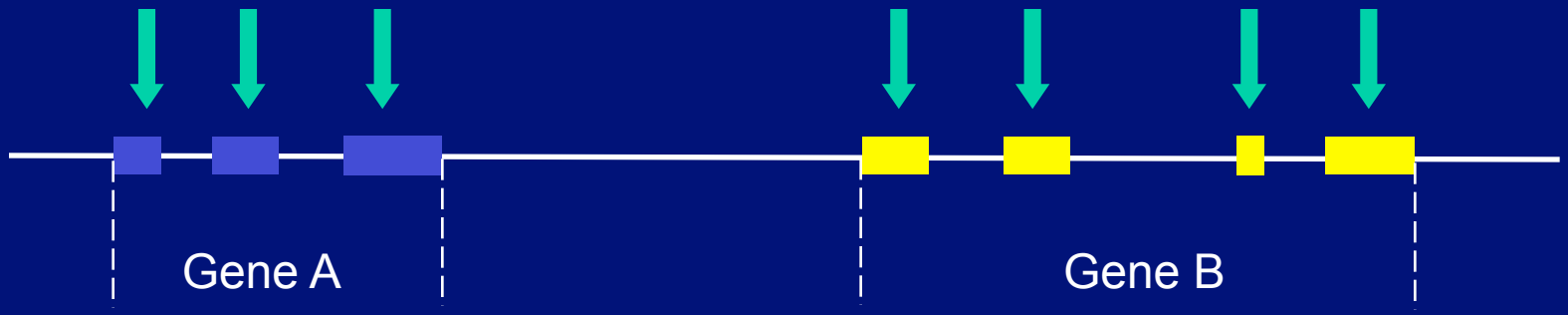
# Cells decode the information in genes to build proteins

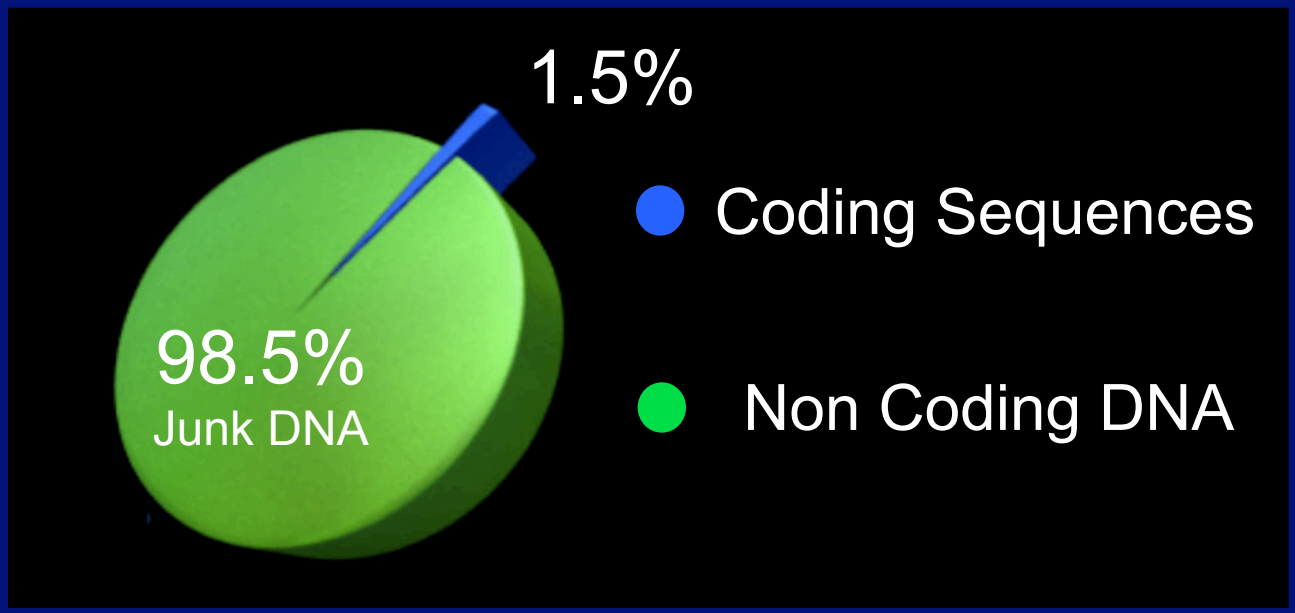


## Gene dimensions

- Gene size
- Internal exons have a mean size of 145 bp vs 218 of the worm
- Short exons require intronic splice enhancers
- The dystrophin  is the largest one (2.4Mb)
- The titin gene has the longest coding sequence (80,780bp), the largest number of exons (178) and the longest single exon (17,106)

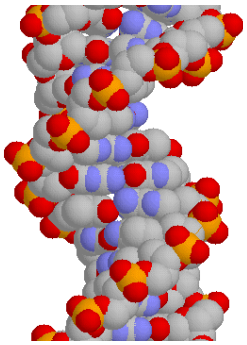
# The exons represent only about 1.5% of the entire genome



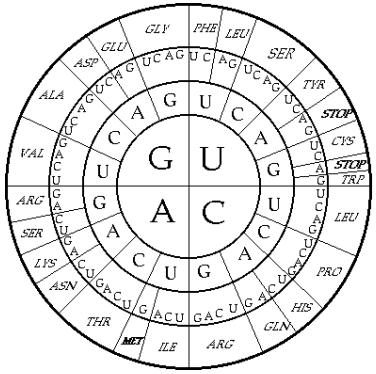
# There are between 25,000 genes in the human genome

DNA

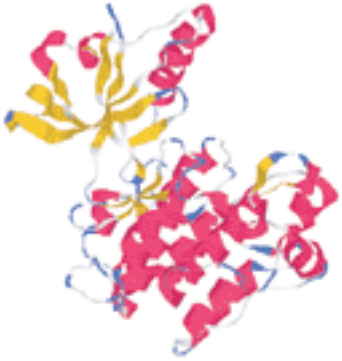


*

genetic code



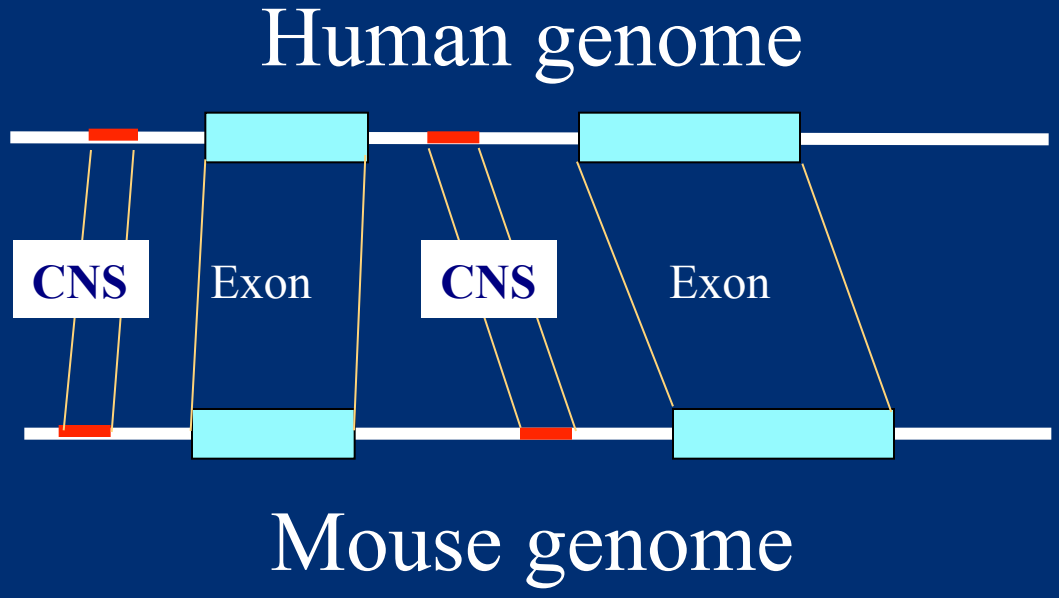

protein

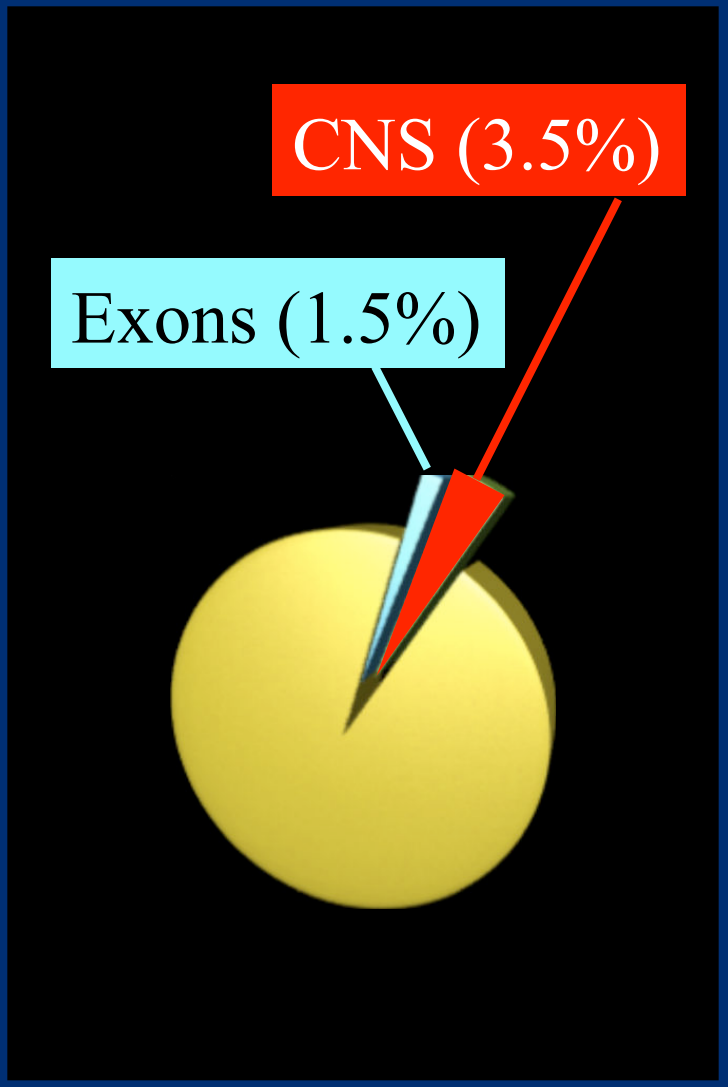


The human gene inventory corresponds to ~1.5% of the genome (coding regions)

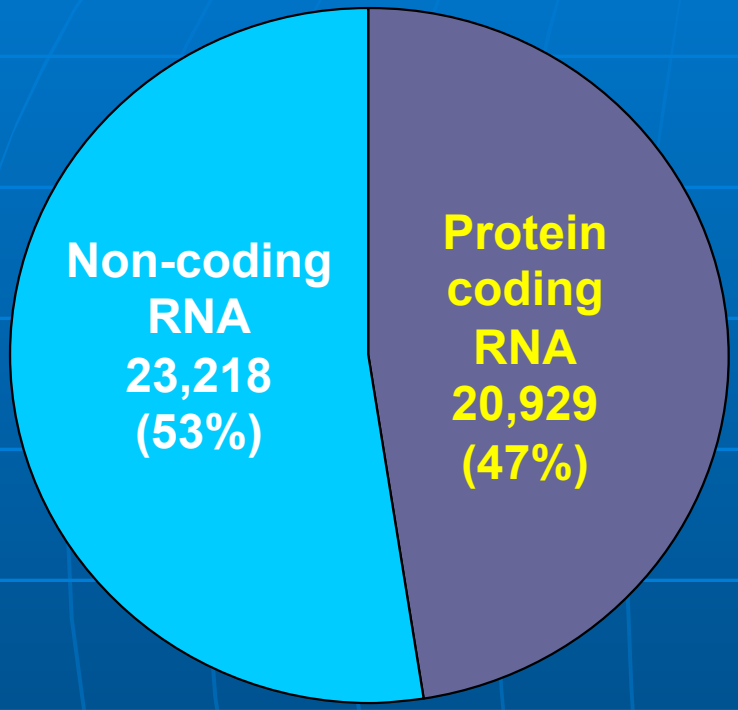
# Conserved Non-Coding Sequences

About 3.5% of small segments in the human genome are under evolutionary selection for biological functions common to human and mouse

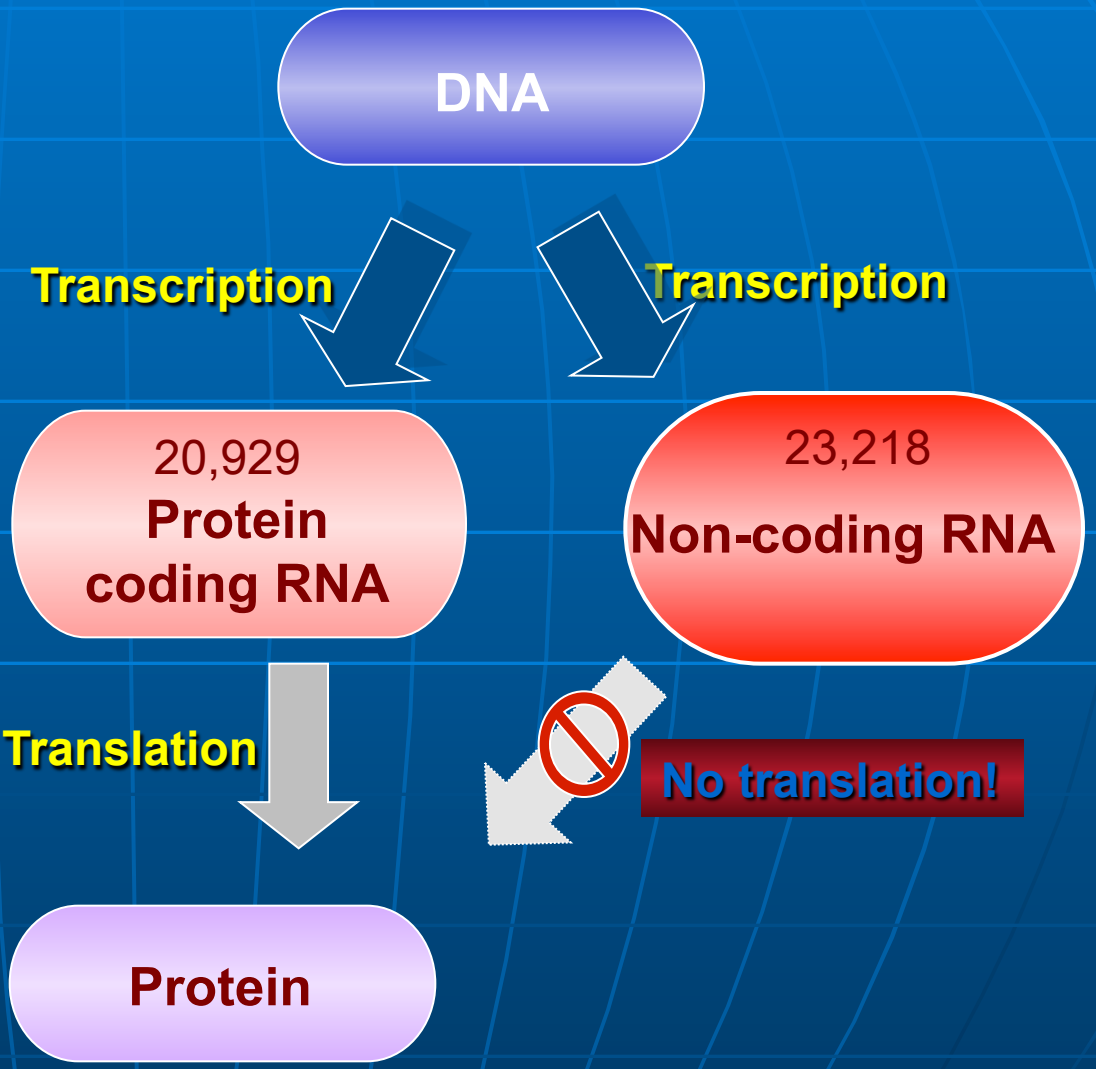


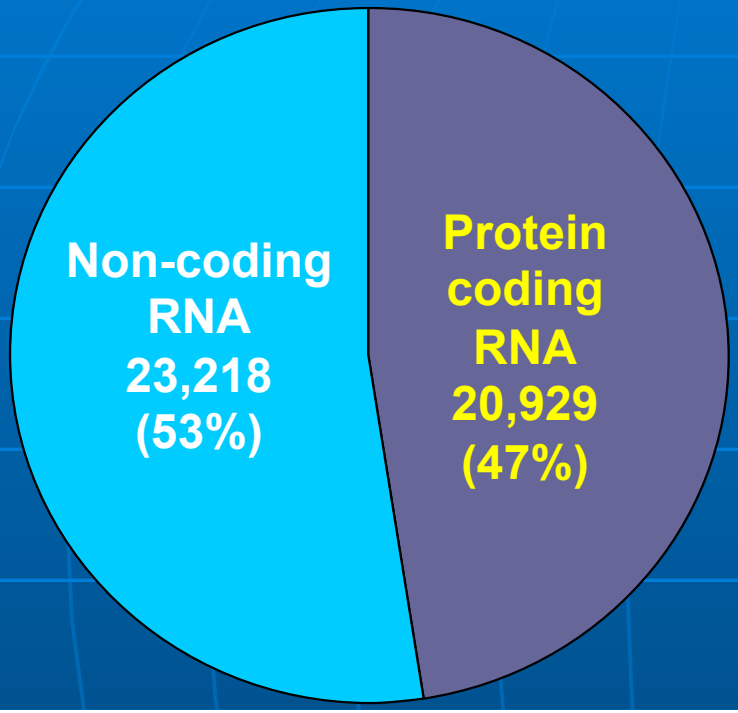
# Discovery of the “RNA continent”




Until FANTOM2 in 2002, only ~ 100 ncRNAs genes were reported except tRNAs and rRNAs.
More than half of the genes have been missing from the gene maps till that time.



# Discovery of th




Tra

articles

## Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

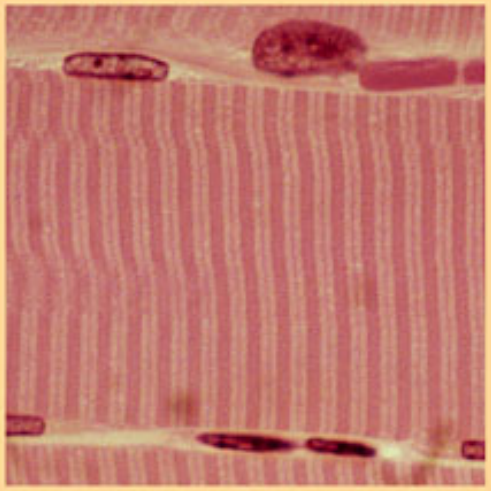
**The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team***

**A full list of authors appears at the end of this paper*

**Only a small proportion of the mouse genome is transcribed into mature messenger RNA transcripts. There is an international collaborative effort to identify all full-length mRNA transcripts from the mouse, and to ensure that each is represented in a physical collection of clones. Here we report the manual annotation of 60,770 full-length mouse complementary DNA sequences. These are clustered into 33,409 'transcriptional units', contributing 90.1% of a newly established mouse transcriptome database. Of these transcriptional units, 4,258 are new protein-coding and 11,665 are new non-coding messages, indicating that non-coding RNA is a major component of the transcriptome. 41% of all transcriptional units showed evidence of alternative splicing. In protein-coding transcripts, 79% of splice variations altered the protein product. Whole-transcriptome analyses resulted in the identification of 2,431 sense–antisense pairs. The present work, completely supported by physical clones, provides the most comprehensive survey of a mammalian transcriptome so far, and is a valuable resource for functional genomics.**


*Nature*, volume 420, (December 5, 2002)

## PATTERNS OF GENE EXPRESSION IN FIVE TYPES OF CELLS

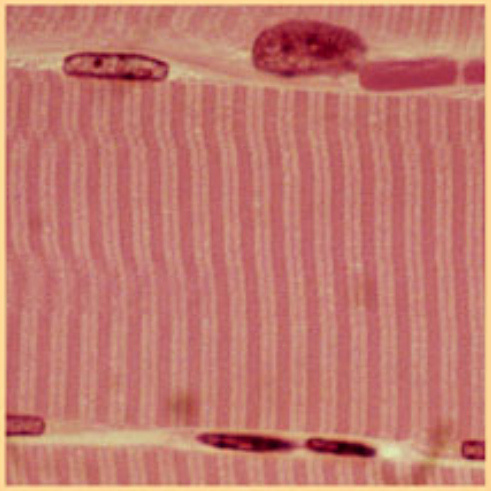


**Muscle Cell**

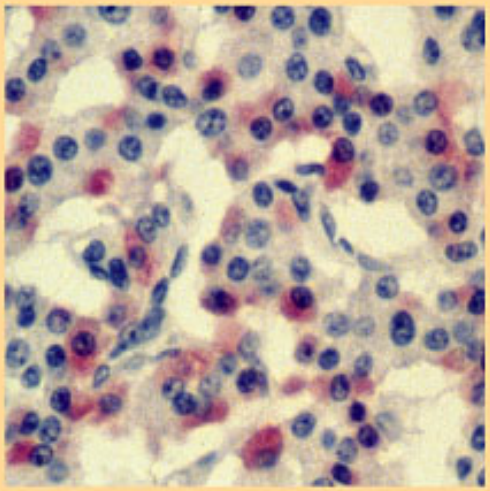
**Genes for...**

| | |
|---|---|
| Glycolysis enzymes | On |
| Muscle contraction proteins | On |
| Glucagon | Off |
| Insulin | Off |
| Hemoglobin | Off |

## PATTERNS OF GENE EXPRESSION IN FIVE TYPES OF CELLS



**Muscle Cell**



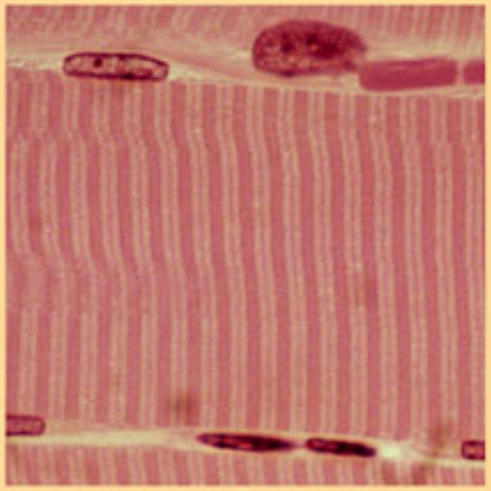
**Pancreas Cells**

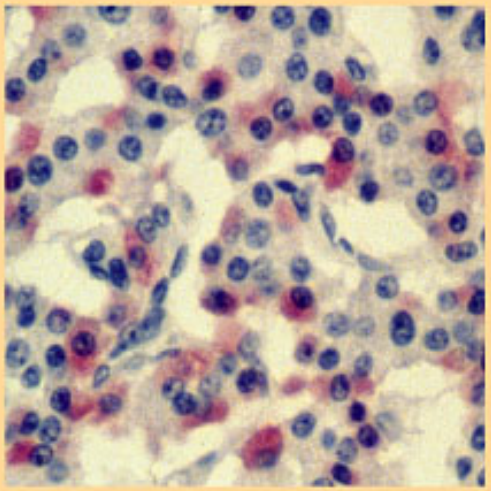
Alpha Cells | Beta Cells

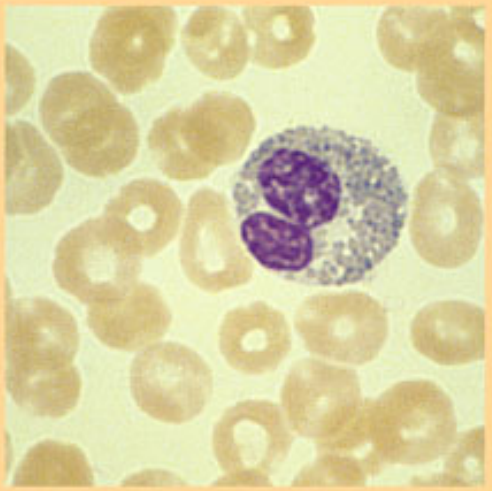
**Genes for...**

| | | | |
|---|---|---|---|
| Glycolysis enzymes | On | On | On |
| Muscle contraction proteins | On | Off | Off |
| Glucagon | Off | On | Off |
| Insulin | Off | Off | On |
| Hemoglobin | Off | Off | Off |

## PATTERNS OF GENE EXPRESSION IN FIVE TYPES OF CELLS



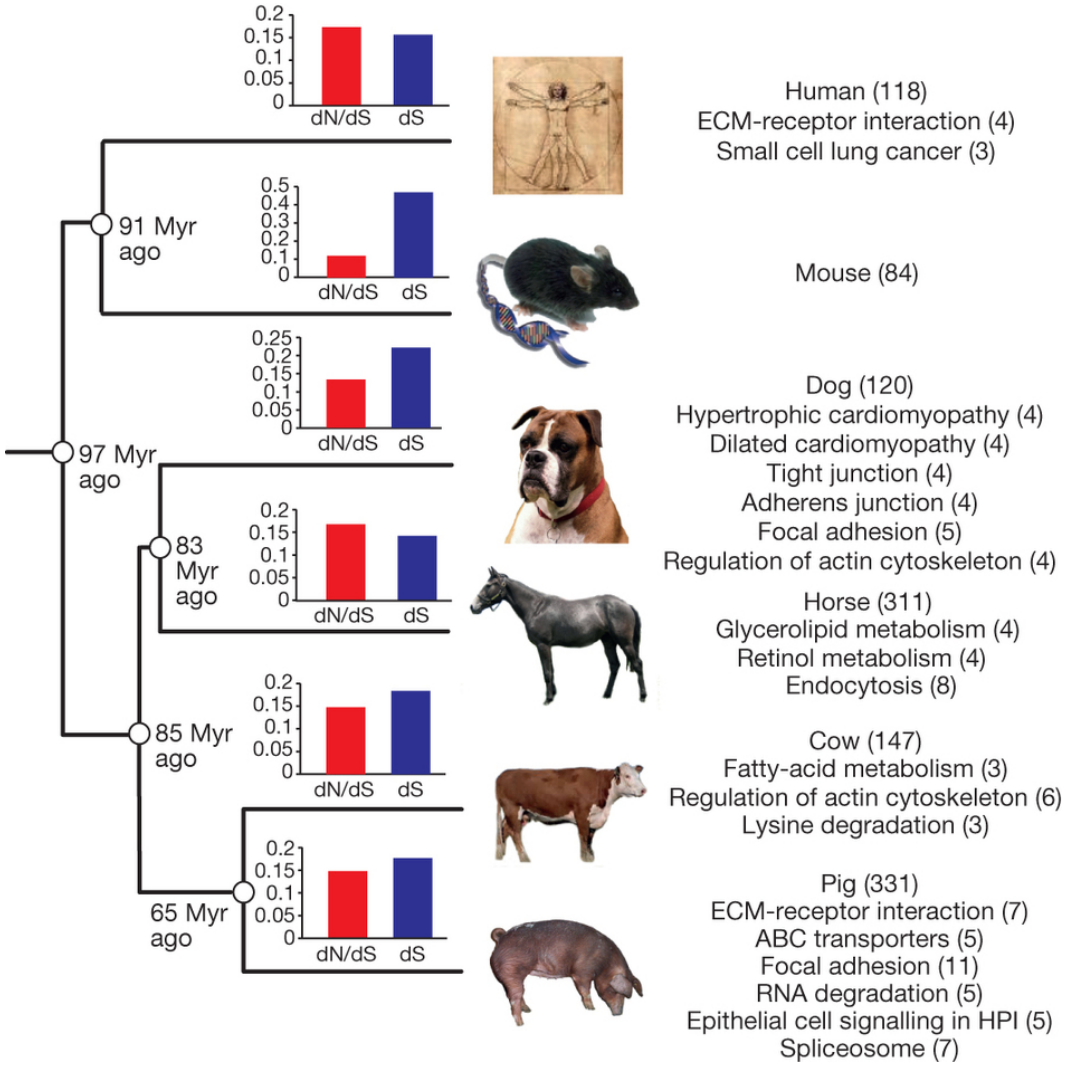




| Genes for... | Muscle Cell | Pancreas Cells | | Blood Cells | |
|---|---|---|---|---|---|
| | | Alpha Cells | Beta Cells | White Cells | Red Cells (Immature) |
| Glycolysis enzymes | On | On | On | On | On |
| Muscle contraction proteins | On | Off | Off | Off | Off |
| Glucagon | Off | On | Off | Off | Off |
| Insulin | Off | Off | On | Off | Off |
| Hemoglobin | Off | Off | Off | Off | On |

# Genome Projects on other organisms

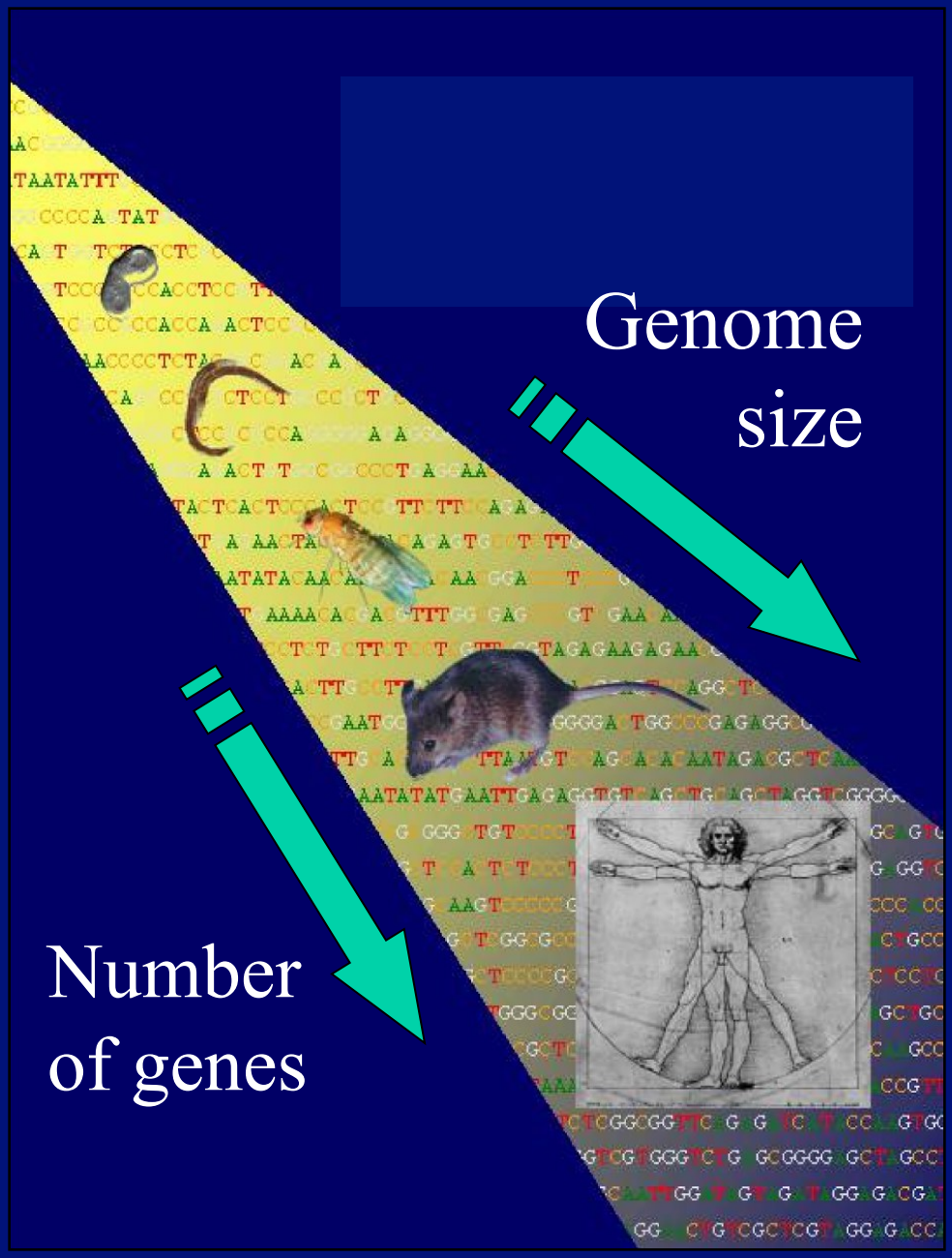

nature
BRINGING HOME THE BACON

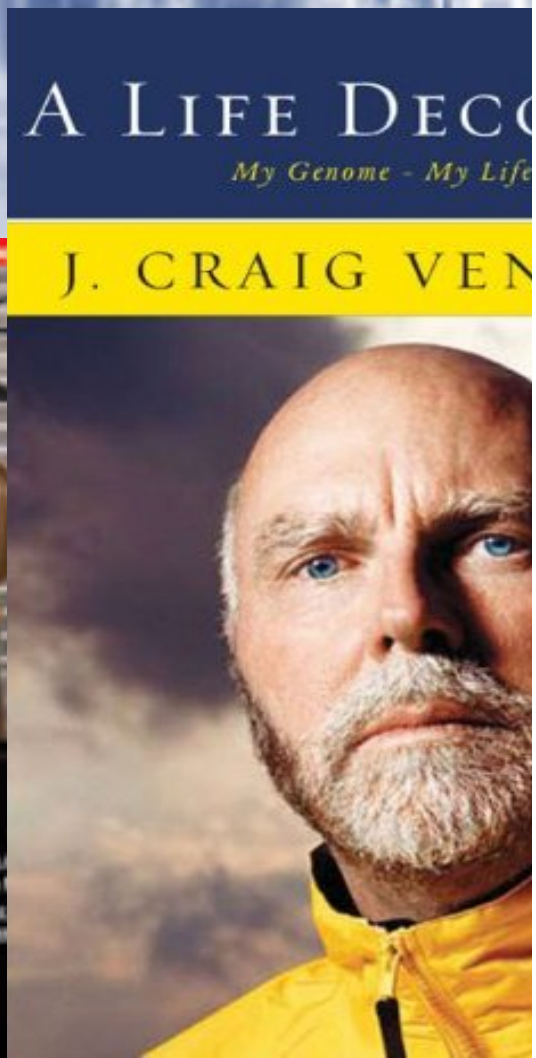

0.2
0.15
0.1
0.05
0
dN/dS
dS
91 Myr ago
0.5
0.4
0.3
0.2
0.1
0
97 Myr ago
0.25
0.2
0.15
0.1
0.05
83 Myr ago
85 Myr ago
65 Myr ago
Human (118)
ECM-receptor interaction (4)
Small cell lung cancer (3)
Mouse (84)
Dog (120)
Hypertrophic cardiomyopathy (4)
Dilated cardiomyopathy (4)
Tight junction (4)
Adherens junction (4)
Focal adhesion (5)
Regulation of actin cytoskeleton (4)
Horse (311)
Glycerolipid metabolism (4)
Retinol metabolism (4)
Endocytosis (8)
Cow (147)
Fatty-acid metabolism (3)
Regulation of actin cytoskeleton (6)
Lysine degradation (3)
Pig (331)
ECM-receptor interaction (7)
ABC transporters (5)
Focal adhesion (11)
RNA degradation (5)
Epithelial cell signalling in HPI (5)
Spliceosome (7)

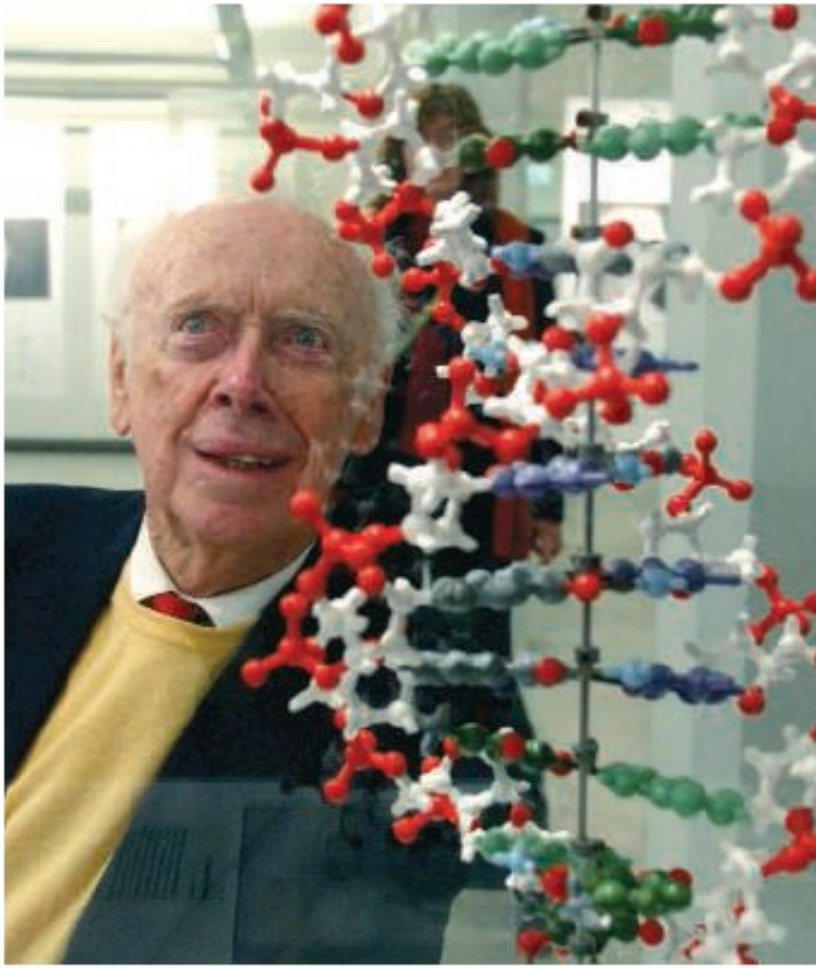
# Organism complexity and gene number

| | |
|---|---|
| *Escherichia coli* | *3,200* |
| *Saccharomyces cerevisiae* | *6,300* |
| *Caenorhabditis elegans* | *19,100* |
| *Drosophila melanogaster* | *13,600* |
| *Danio rerio* | *21,322* |
| *Mus musculus* | *22,000* |
| *Homo sapiens* | 22,000 |
| *Fugu rubripes* | 26,700 |
| *Arabidopsis thaliana* | *25,000* |
| *Tetrahymena thermophila* | 27,000 |

Gene number does not always correlate with evolutionary status



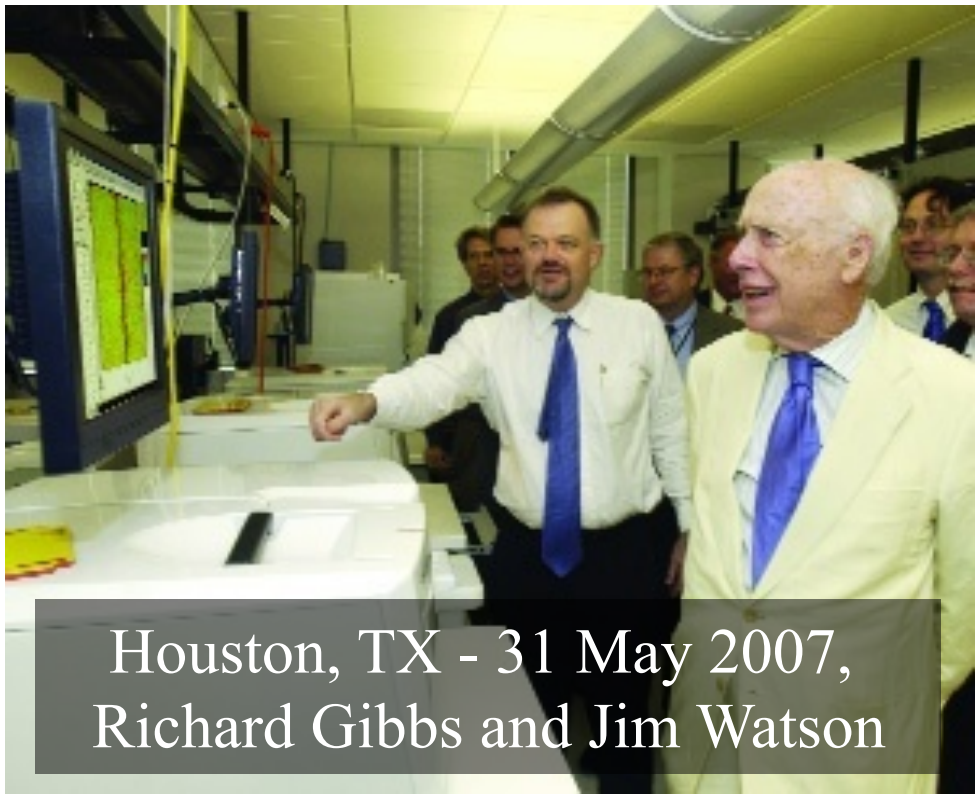



**Know thyself.** Nobelist James Watson is planning to receive—and possibly share—a complete copy of his own genome sequence this year.

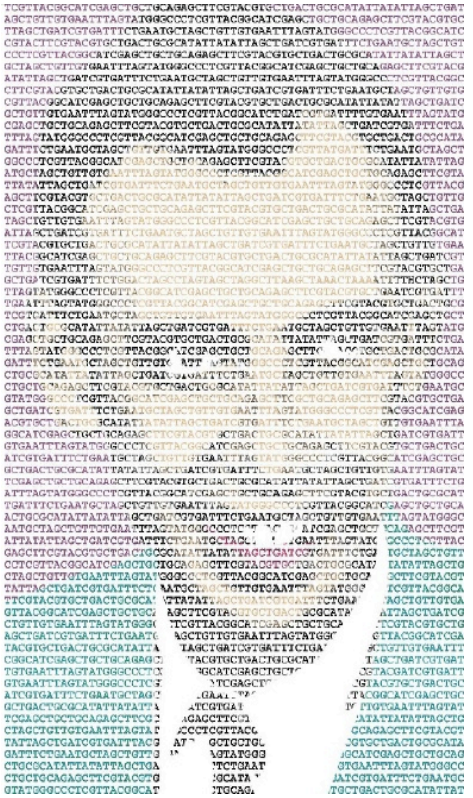
Vol 452|17 April 2008|doi:10.1038/nature06884

# The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yufeng Shen[1]*, Lei Chen[1], Amy McGuire[3], Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier Gomes[2], Karrie Tartaro[2]†, Faheem Niazi[2], Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chinault[4], Xing-zhi Song[1], Yue Liu[1], Ye Yuan[1], Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2], George M. Weinstock[1,4], Richard A. Gibbs[1,4] & Jonathan M. Rothberg[2]†



Houston, TX - 31 May 2007,
Richard Gibbs and Jim Watson

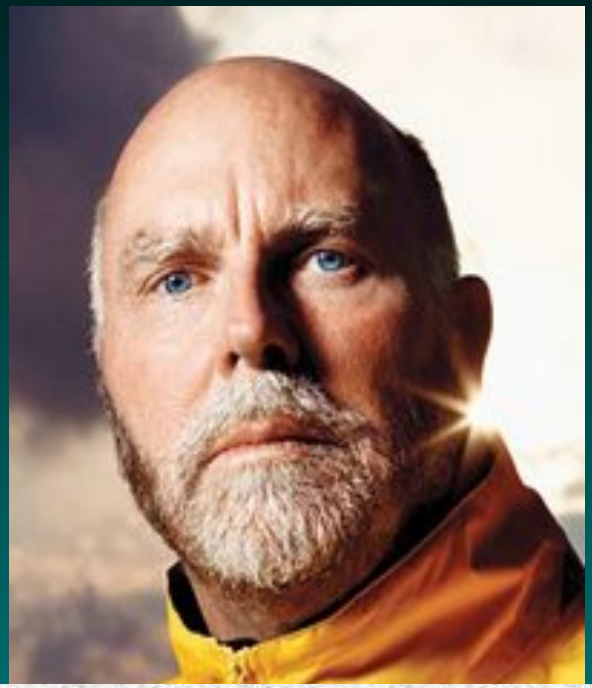


# Quicker, Smaller, Cheaper

| Genome sequenced (publication year) | HGP (2003) | Venter (2007) | Watson (2008) |
|---|---|---|---|
| Time taken (start to finish) | 13 years | 4 years | 4.5 months |
| Number of scientists listed as authors | > 2,800 | 31 | 27 |
| Cost of sequencing (start to finish) | $2.7 billion | $100 million | < $1.5 million |
| Coverage | 8–10 × | 7.5 × | 7.4 × |
| Number of institutes involved | 16 | 5 | 2 |
| Number of countries involved | 6 | 3 | 1 |

NATURE, Vol 452, 17 April 2008

# Who's next?



1

2


3

# The Yanhuang Project at BGI

- The Yanhuang Project, carried out by the Beijing Genomics Institute (BGI), will sequence the genomes of 100 Chinese volunteers for scientific purposes

- The sequencing of the first genome of a Han Chinese (a researcher) was finished in October 2007

- BGI has sequenced in Jan 2008 the genome of the first Chinese volunteer (who donated about 1.3 million U.S. dollars) as part of a project to create a database of Asian genomes

# March 2008: SOLiD Human HapMap Sample NA18507 Whole Genome Sequence

Scientists from Applied Biosystems have resequenced a human DNA sample that was included in the International HapMap Project (an anonymous African male of the Yoruba people of Ibadan, Nigeria) for less than $60,000

The team used the company's SOLiD System to generate 36 gigabases of sequence data in seven runs of the system

The sequence is already available in GenBank under the accession number SRA000272

SOLiD System

~ 50 bases/read

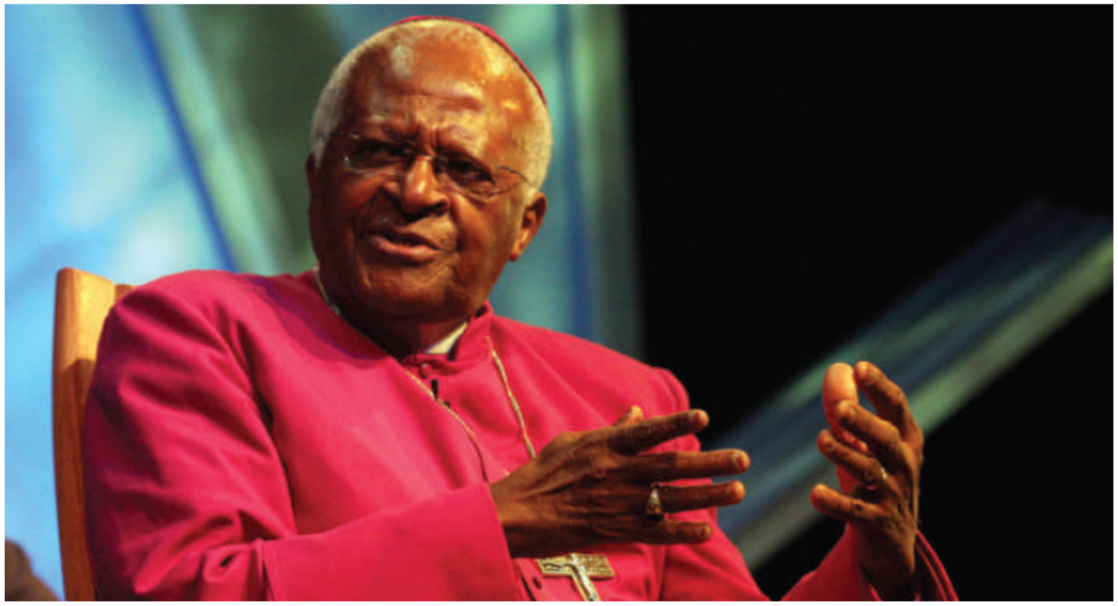
**SILICON WHISKERS CATCH SUN'S RAYS**
Tiny rods could make solar cells cheaper.
go.nature.com/1DRrnJ

J. SOHM/VISIONSOFAMERICA

# Africa yields two full human genomes

To the growing list of people with fully sequenced genomes, two memorable names have now been added: Archbishop Desmond Tutu, the South African civil-rights activist, and !Gubi, a Namibian hunter-gatherer.

!Gubi hails from the Khoisan community, one of the most ancient and diverse human populations. His is the first genome from a minority population in Africa to be sequenced. Comparing his sequence with the partial genome sequences of three other Khoisan shows that they are as different from one another as a European would be from an Asian, says team leader Stephan Schuster, a genome researcher at Pennsylvania State University in University Park. "This is despite the fact that they sometimes live within walking distance of one another," he adds.



REX FEATURES

Archbishop Desmond Tutu's genome was chosen to represent the Bantu peoples of southern Africa.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

# Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy

10.1056/NEJMoa0908094 NEJM.ORG

**Table 3. Individual Human Genomes Sequenced to Date.***

| Genome† | Technology Used |
|---|---|
| Venter | Sanger method |
| Watson | 454 Sequencing System (Roche) |
| Chinese (YH) | Genome Analyzer (Illumina) |
| African (NA18507) | Genome Analyzer (Illumina) |
| African (NA18507) | SOLiD system (Applied Biosystems) |
| Korean (SJK) | Genome Analyzer (Illumina) |
| Korean (AK1) | Genome Analyzer (Illumina) |
| Proband in this study | SOLiD system (Applied Biosystems) |

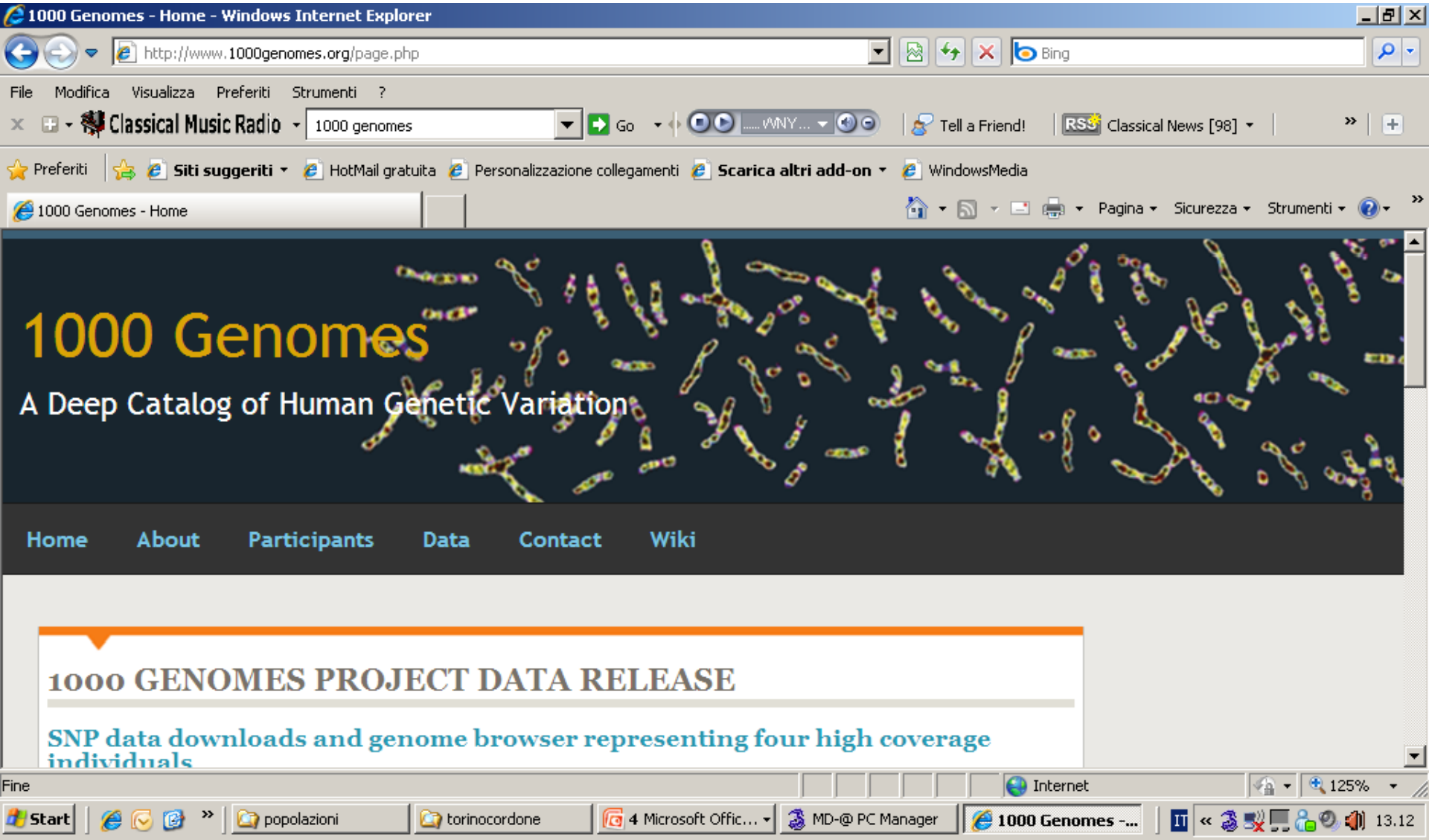
# Personal Genomics

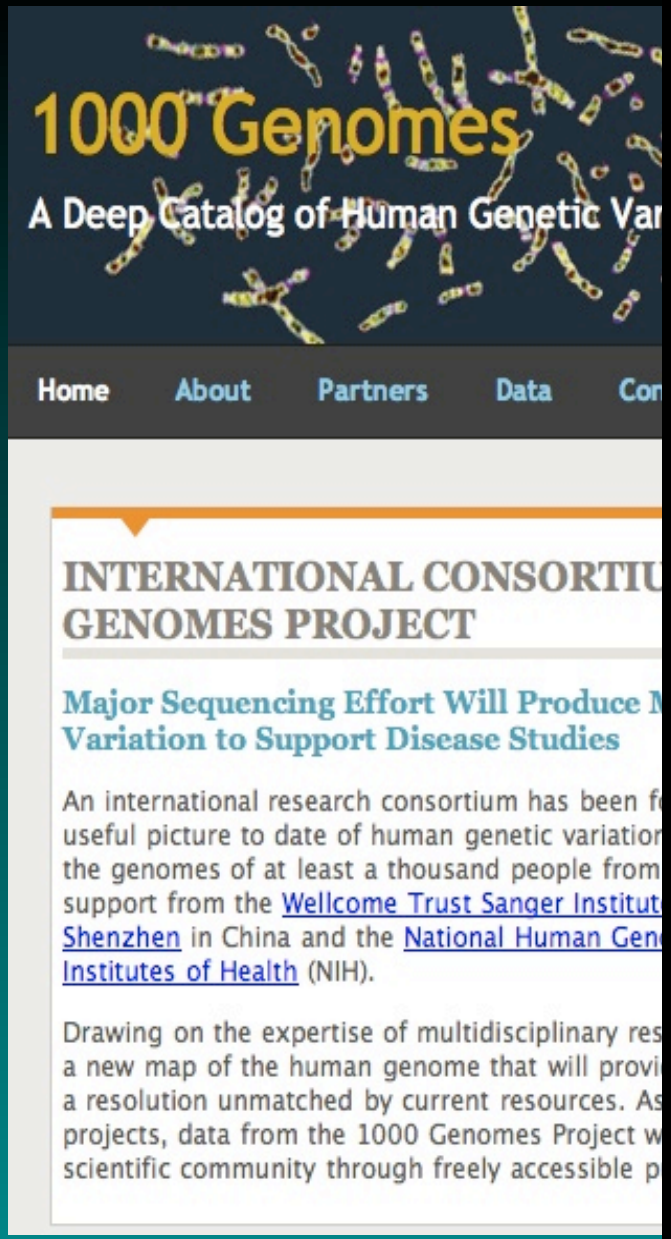
- Many scientists are competing in a challenge issued by the NIH to produce a sequencing method that costs:

  $100,000 less than per genome by 2009

  $1,000 or less by 2014


- X-Prize Foundation is offering $10 million - the largest medical prize in history - for the first private team that can decode 100 human genomes in 10 days


1000 Genomes - Home - Windows Internet Explorer
http://www.1000genomes.org/page.php
Bing
File Modifica Visualizza Preferiti Strumenti ?
Classical Music Radio
1000 genomes
Go
Tell a Friend!
Classical News [98]
Preferiti
Siti suggeriti
HotMail gratuita
Personalizzazione collegamenti
Scarica altri add-on
WindowsMedia
1000 Genomes - Home
Pagina
Sicurezza
Strumenti
1000 Genomes
A Deep Catalog of Human Genetic Variation
Home
About
Participants
Data
Contact
Wiki
1000 GENOMES PROJECT DATA RELEASE
SNP data downloads and genome browser representing four high coverage individuals
Fine
Internet
125%
Start
popolazioni
torinocordone
4 Microsoft Offic...
MD-@ PC Manager
1000 Genomes -...
13.12




http://www.1

# Steering Committee

**Richard Durbin (co-chair)** Sanger Institute

**David Altshuler (co-chair)** Broad / MGH / Harvard

**Goncalo Abecasis** University of Michigan

**Aravinda Chakravarti** Johns Hopkins

**Andrew Clark** Cornell University

**Francis Collins** National Human Genome Research Institute

**Peter Donnelly** Oxford University

**Paul Flicek** European Bioinformatics Institute

**Stacey Gabriel** Broad Institute

**Richard Gibbs** Baylor College of Medicine

**Bartha Knoppers** University of Montreal

**Eric Lander** Broad Institute

**Elaine Mardis** Washington University in St. Louis

**Gil McVean** Oxford University

**Debbie Nickerson** University of Washington

**Leena Peltonen** Sanger Institute

**Stephen Sherry** National Center for Biotechnology Information

**Rick Wilson** Washington University in St. Louis

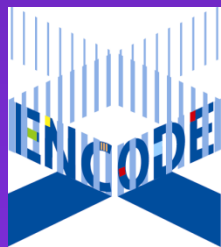
**Huanming (Henry) Yang** Beijing Genomics Institute

# Challenge

Compile a *comprehensive encyclopedia* of all of the sequence features in the human genome.

Approach:

- Apply lessons learned from the success of the Human Genome Project
- Start with well-defined pilot project
- Develop and test high-throughput technologies



ARTICLE

doi:10.1038/nature11247

# An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization
expansive resource of functional annotations for biomedical rese

The human genome sequence provides the underlying code for human biology. Despite intensive study, especially in identifying protein-coding genes, our understanding of the genome is far from complete, particularly with regard to non-coding RNAs, alternatively spliced transcripts and regulatory sequences. Systematic analyses of transcripts and regulatory information are essential for the identification of genes and regulatory regions, and are an important resource for the study of human biology and disease. Such analyses can also provide comprehensive views of the organization and variability of genes and regulatory information across cellular contexts, species and individuals.

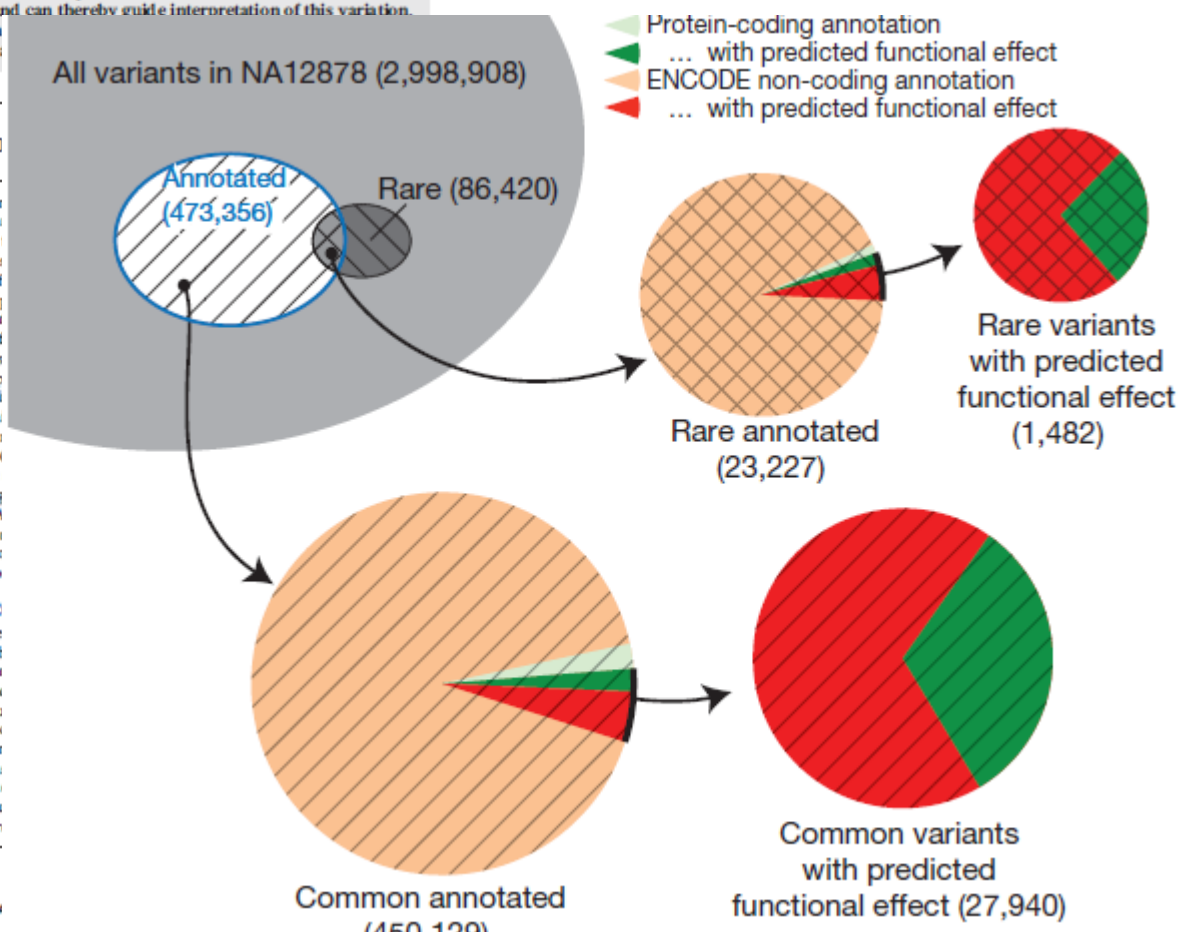
The Encyclopedia of DNA Elements (ENCODE) project aims to delineate all functional elements encoded in the human genome[1–5]. Operationally, we define a functional element as a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure). Comparative genomic studies suggest that 3–8% of bases are under purifying (negative) selection[6–8] and therefore may be functional, although other analyses have suggested much higher estimates[9–11]. In a pilot phase covering 1% of the genome, the ENCODE project annotated 60% of mammalian evolutionarily constrained bases, but also identified many additional putative functional elements without evidence of constraint[2]. The advent of more powerful DNA sequencing technologies now enables whole-genome and more precise analyses with a broad repertoire of functional assays.

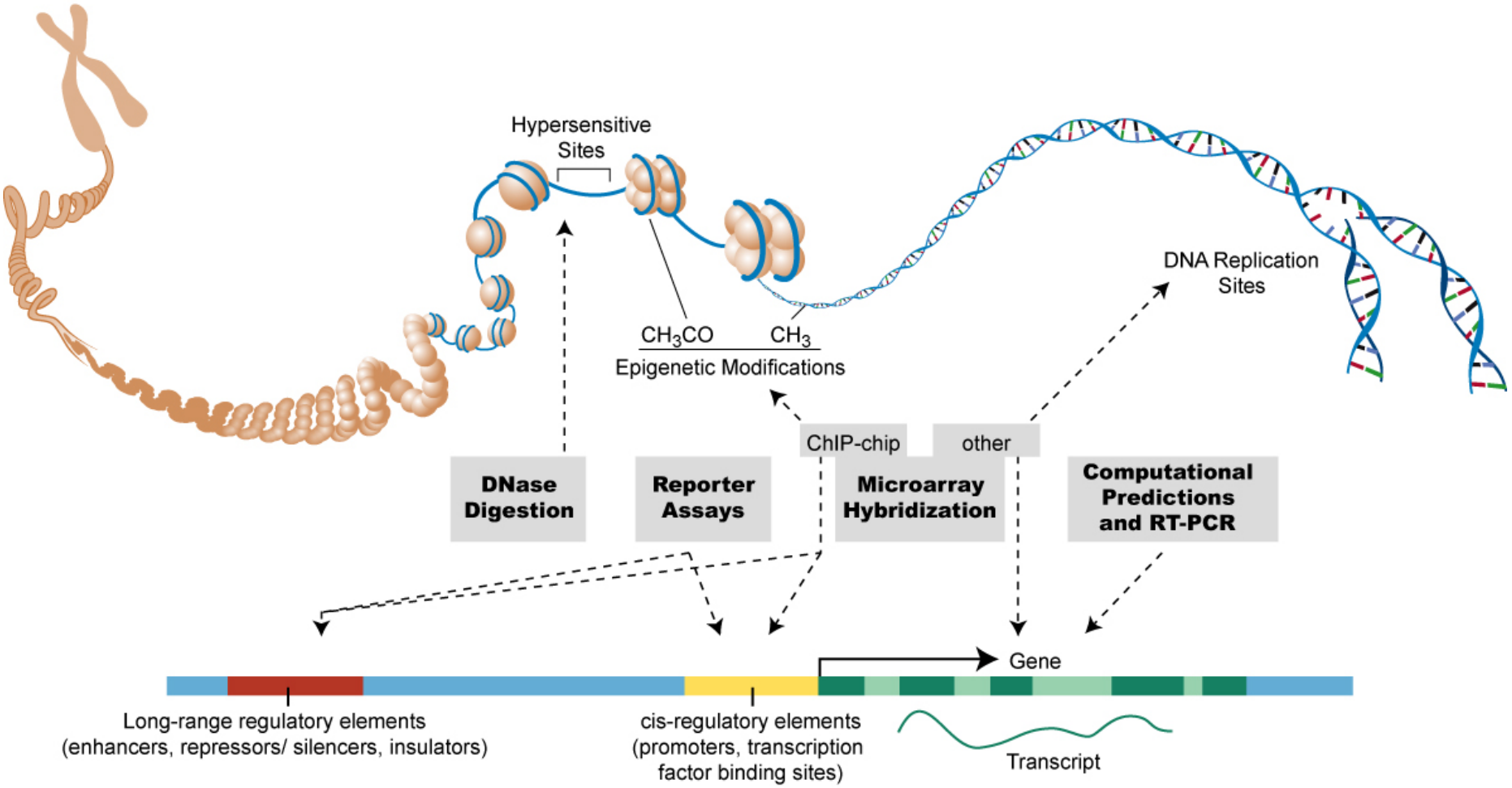
Here we describe the production and initial analysis of 1,640 data sets designed to annotate functional elements in the entire human genome. We integrate results from diverse experiments within cell types, related experiments involving 147 different cell types, and all ENCODE data with other resources, such as candidate regions from genome-wide association studies (GWAS) and evolutionarily constrained regions. Together, these efforts reveal important features about the organization and function of the human genome, summarized below.

• The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type. Much of the genome lies close to a regulatory event:

*Lists of participants and their affiliations appear at the end of the paper.








The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

# GENETICS, GENES AND GENOMES

- The last 60 years of genetics
- Human genome and other genomes
- Genetic diseases
- Genetic polymorphisms

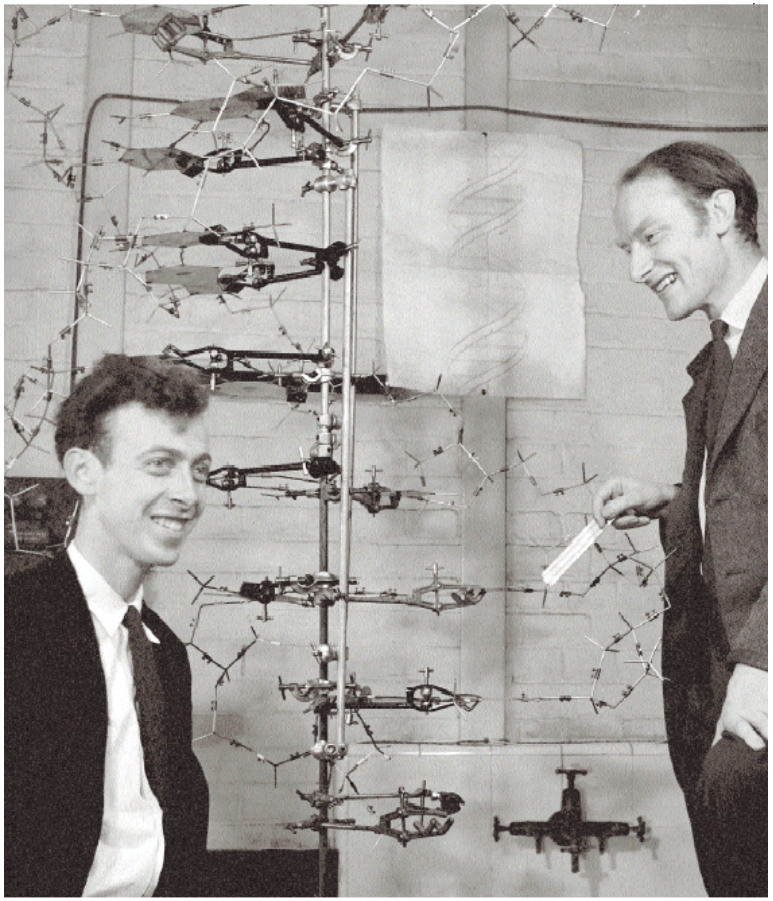
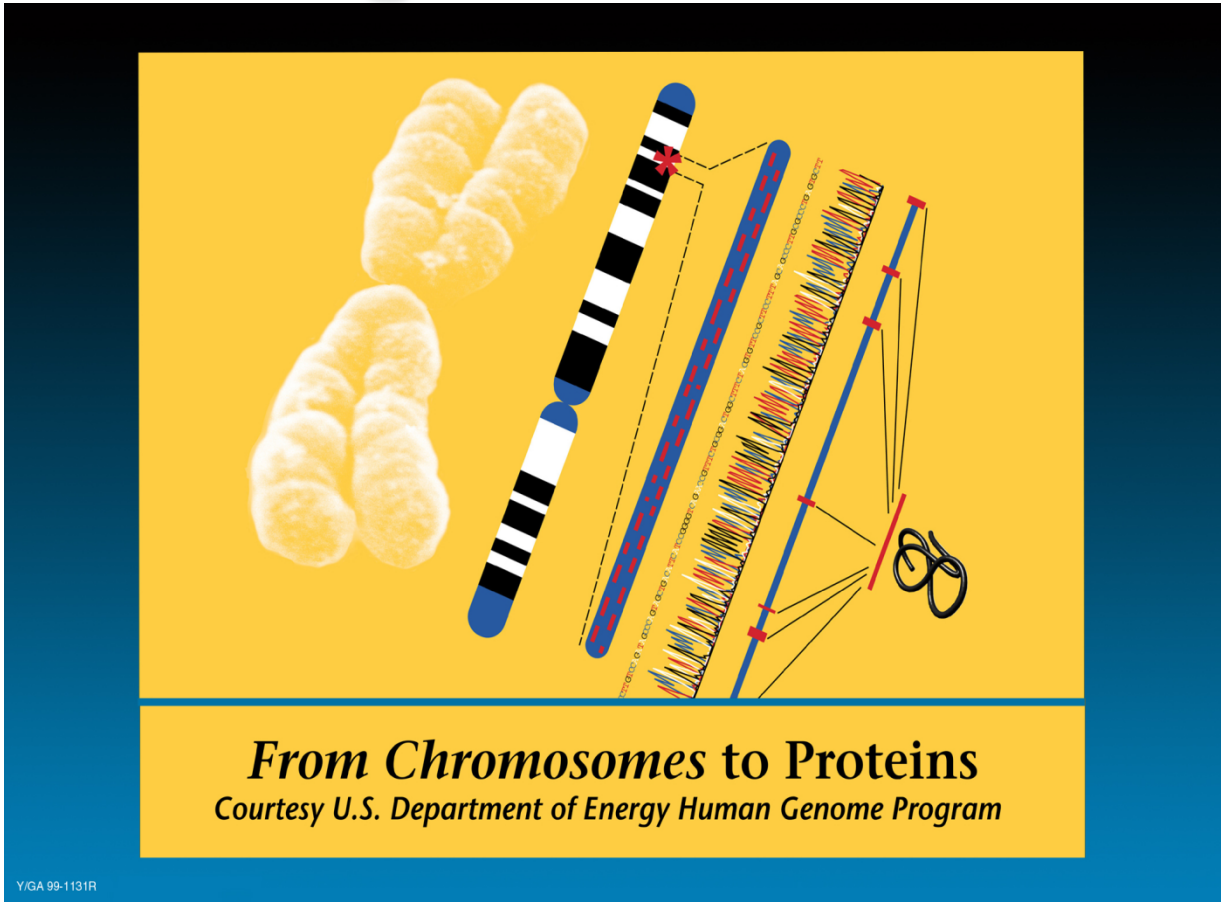


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

# And the implications for human health?

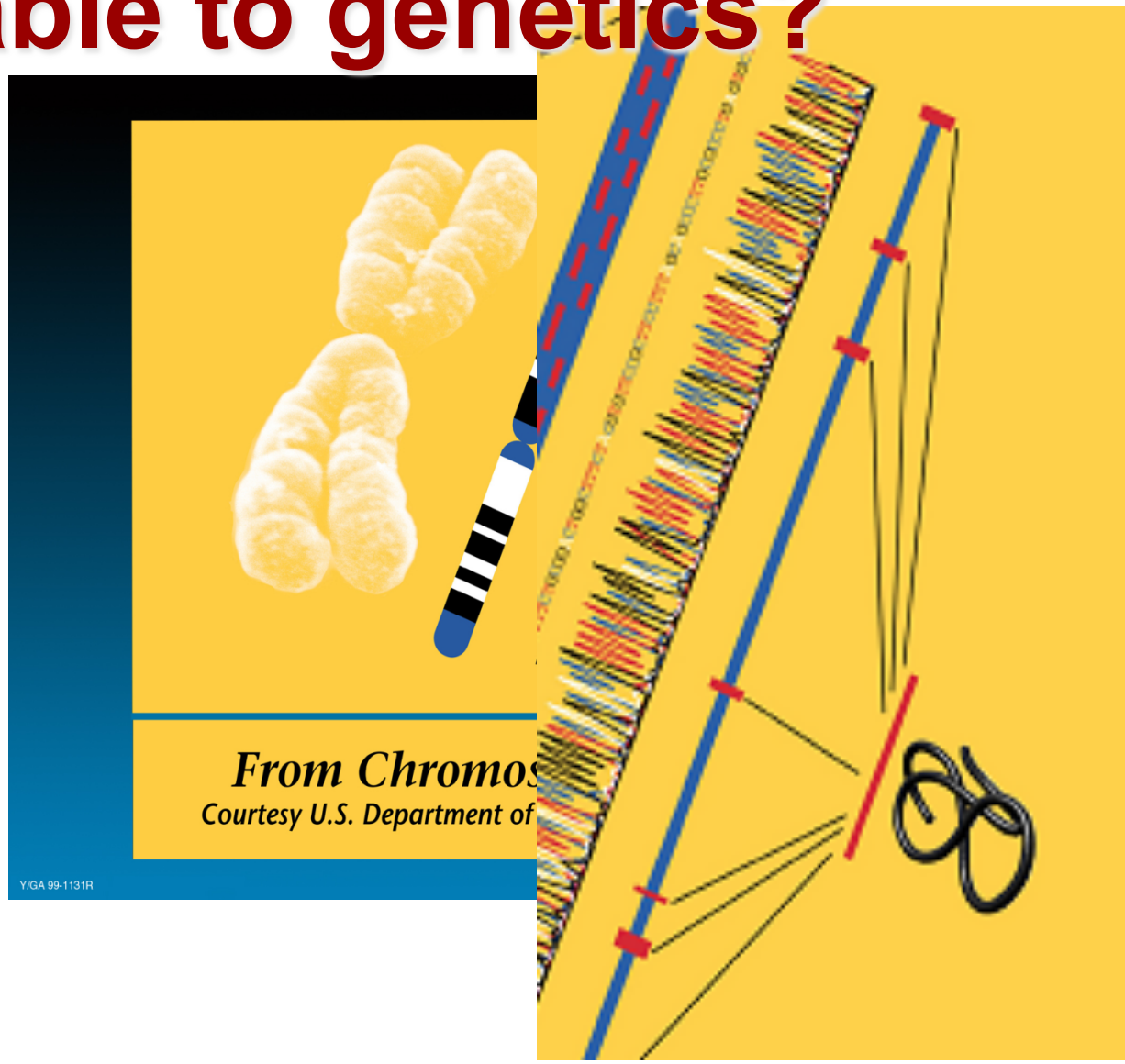
# What diseases are attributable to genetics?

- Monogenic or hereditary diseases
- Chromosomal diseases
- Multifactorial or complex diseases



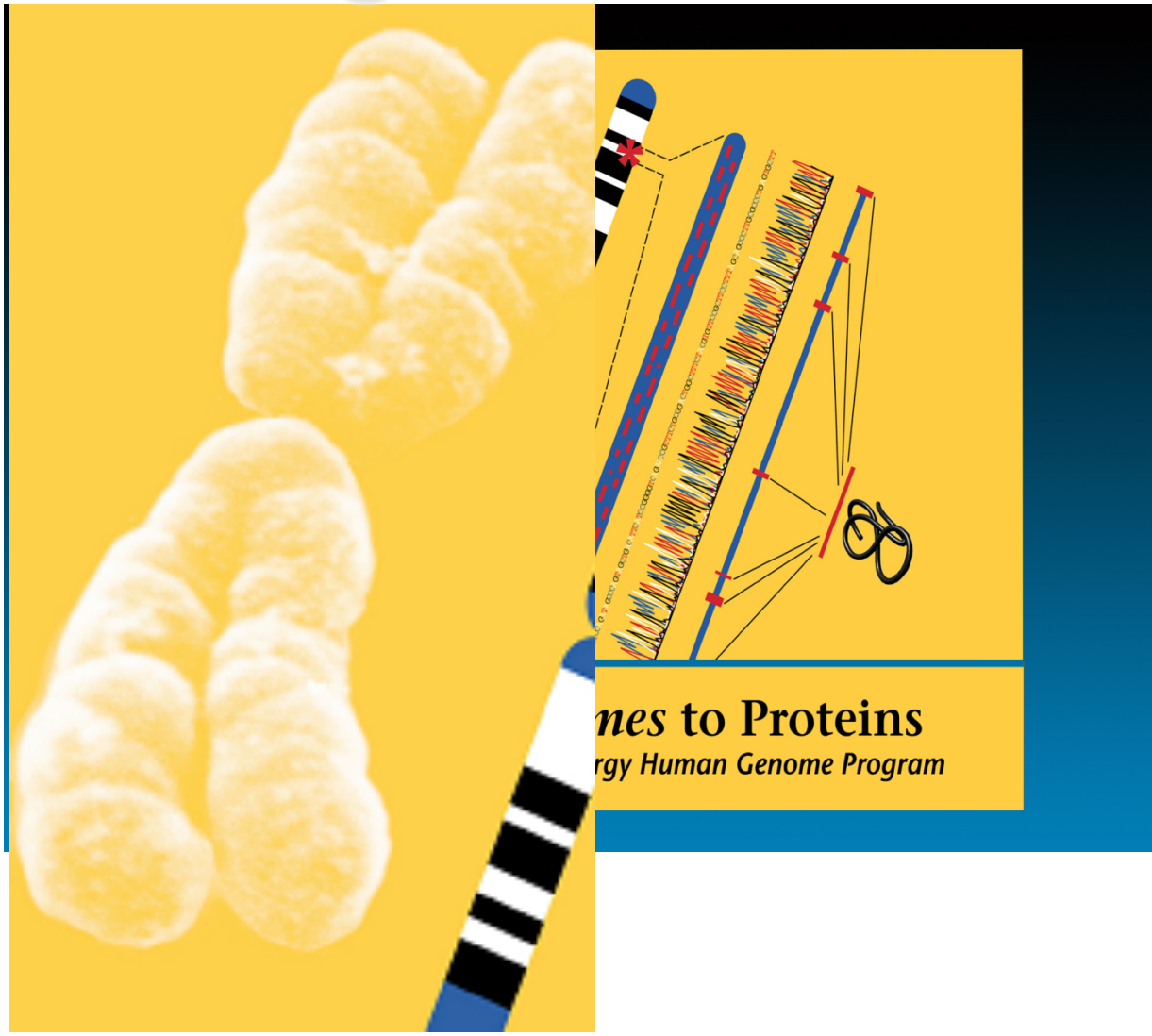
# What diseases are attributable to genetics?

- **Monogenic or hereditary diseases**
- **Chromosomal diseases**
- **Multifactorial or complex diseases**



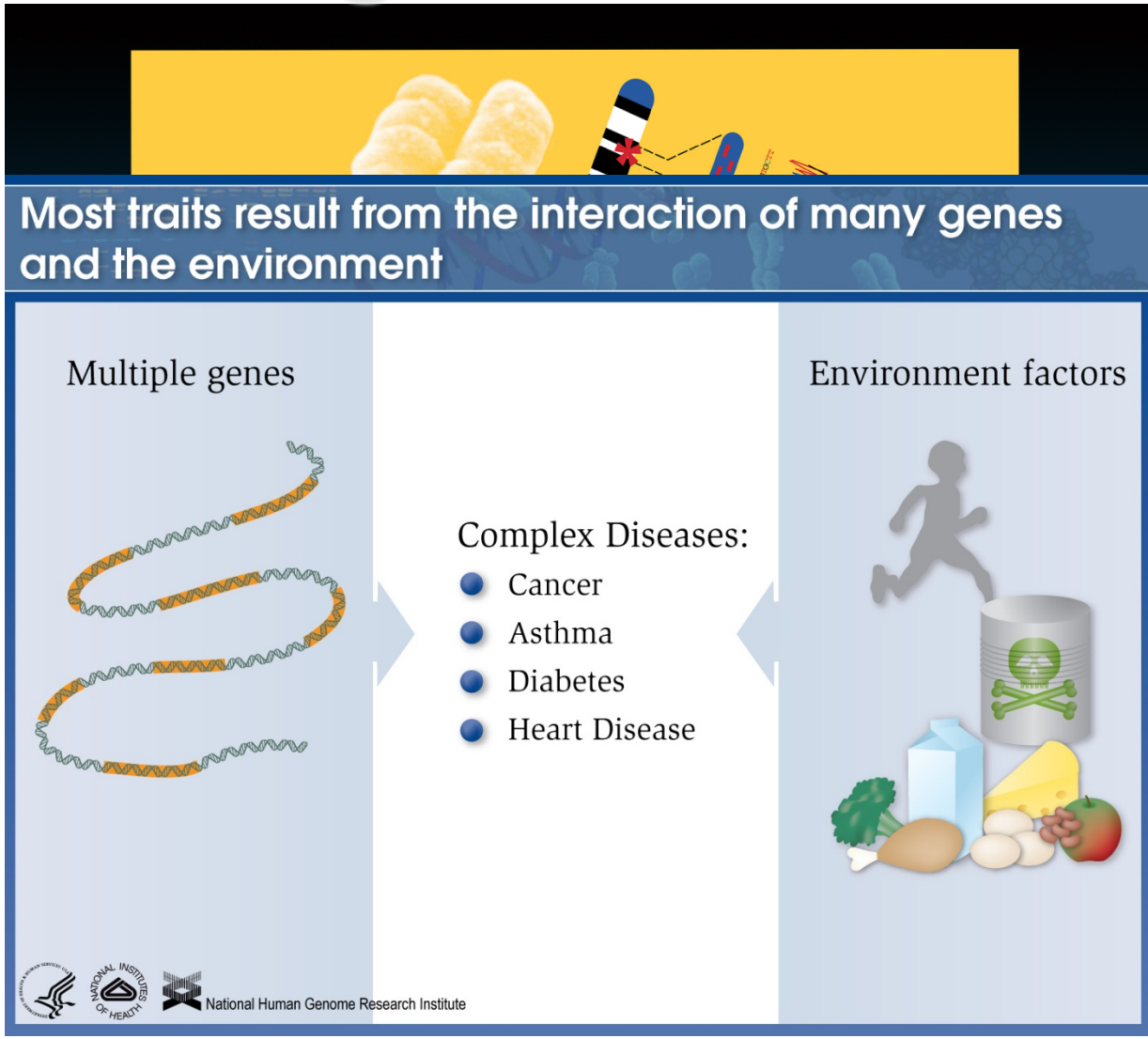
# What diseases are attributable to genetics?

- Monogenic or hereditary diseases
- Chromosomal diseases
- Multifactorial or complex diseases



# What diseases are attributable to genetics?

- Monogenic or hereditary diseases
- Chromosomal diseases
- Multifactorial or complex diseases



*How many genetic diseases?*

*What is their frequency?*

# The impact of genetic diseases

- The 2% of newborns has chromosomal abnormalities or defects in a single gene
- 50% of childhood blindness, deafness, and mental retardation is due to genetic defects
- 30% of hospital admissions in children and 50% of pediatric deaths are due to genetic diseases or birth defects
- 10% of the most common cancers has a genetic component
- 5% of adults experiences a disease in which genetic factors are important

# Age of onset of some genetic disorders

| Age | Disease |
|---|---|
| Lethal during prenatal life | Chromosomal aberrations, some serious malformations |
| Before birth | Congenital malformations, chromosomal aberrations |
| Soon after birth | Fenilchetonuria, cystic fibrosis |
| During first year | Duchenne muscular distrophy |
| At puberty | Other muscular distrophies |
| After puberty | Acute intermittent Porphyria, hereditary glaucoma |
| Variable age | Diabetes mellitus (0-80 years), Huntington disease (15-65 years) |

- **Genetic disease**
- **Congenital disease**

# Frequency of Different Types of Genetic Diseases

| *Type* | *Incidence at Birth (per 1,000)* | *Prevalence at Age 25 Years (per 1,000)* | *Population Prevalence (per 1,000)* |
|---|---|---|---|
| **Diseases due to genome/ chromosome mutations** | 6 | 1.8 | 3.8 |
| **Disease due to single gene mutations** | 10 | 3.6 | 20 |
| **Disease with multifactorial inheritance** | ~50 | ~50 | ~600 |

# GENETICS, GENES AND GENOMES

- The last 60 years of genetics
- Human genomes and other genomes
- Genetic diseases
- Genetic polymorphisms

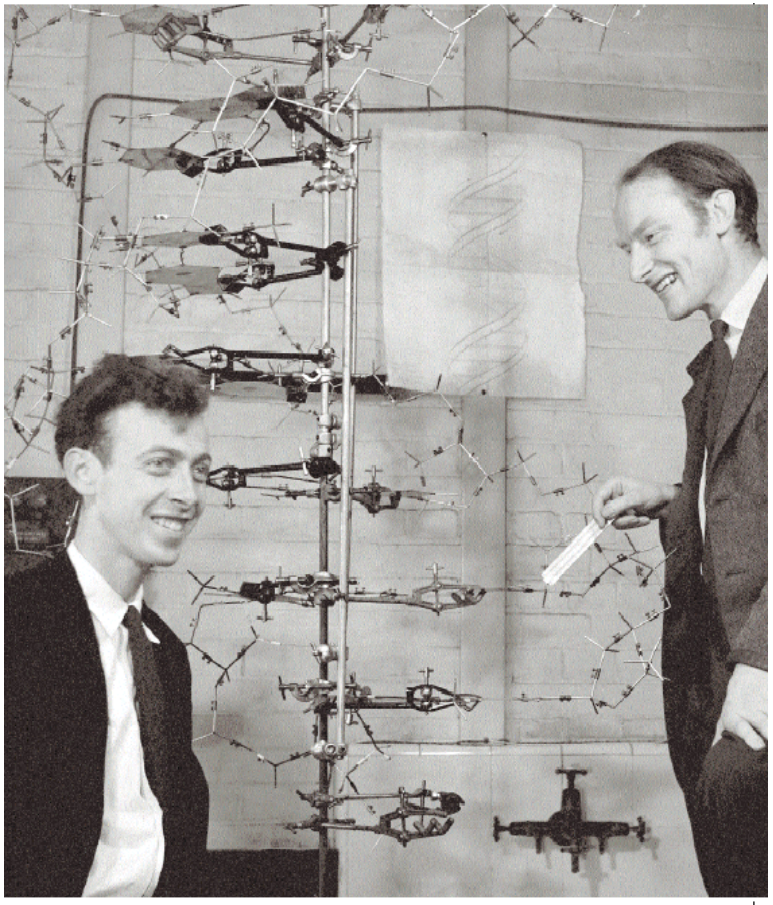
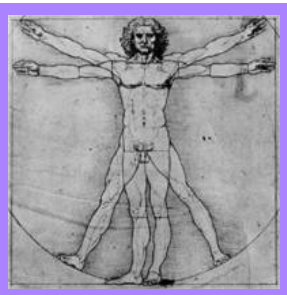
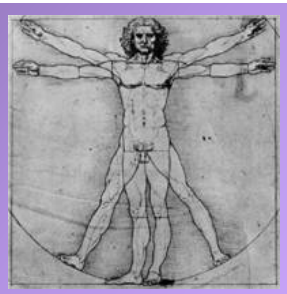


**Figure 3** Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

# How to build a *Homo sapiens*

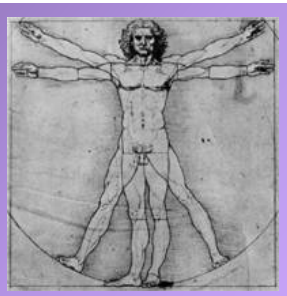


How to build a *Homo sapiens*



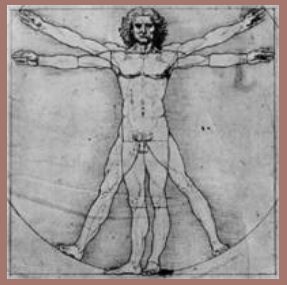
# Identical twins

How to build a *Homo sapiens*



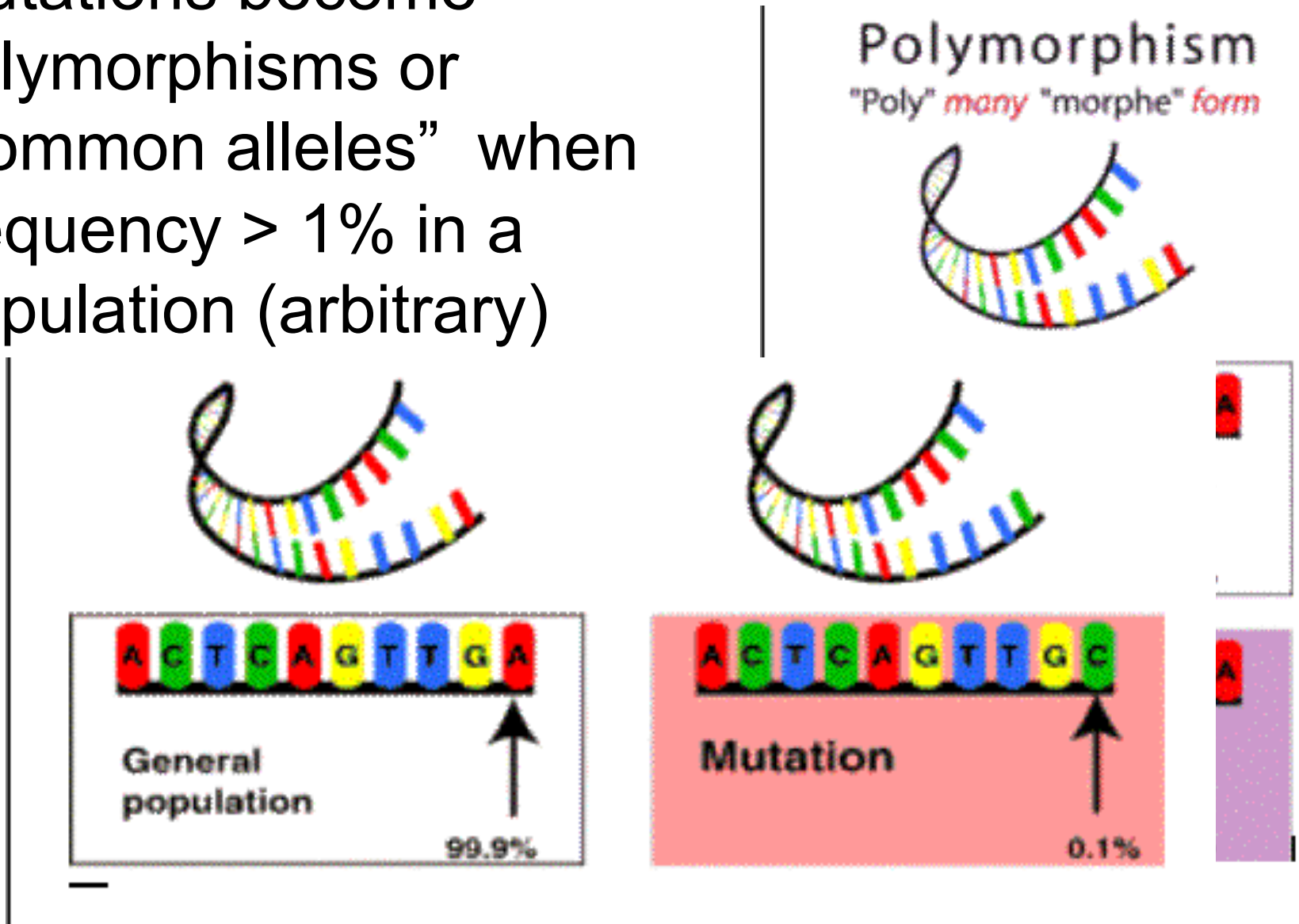
**How to build a *Homo sapiens***

**How to build a *Homo sapiens***

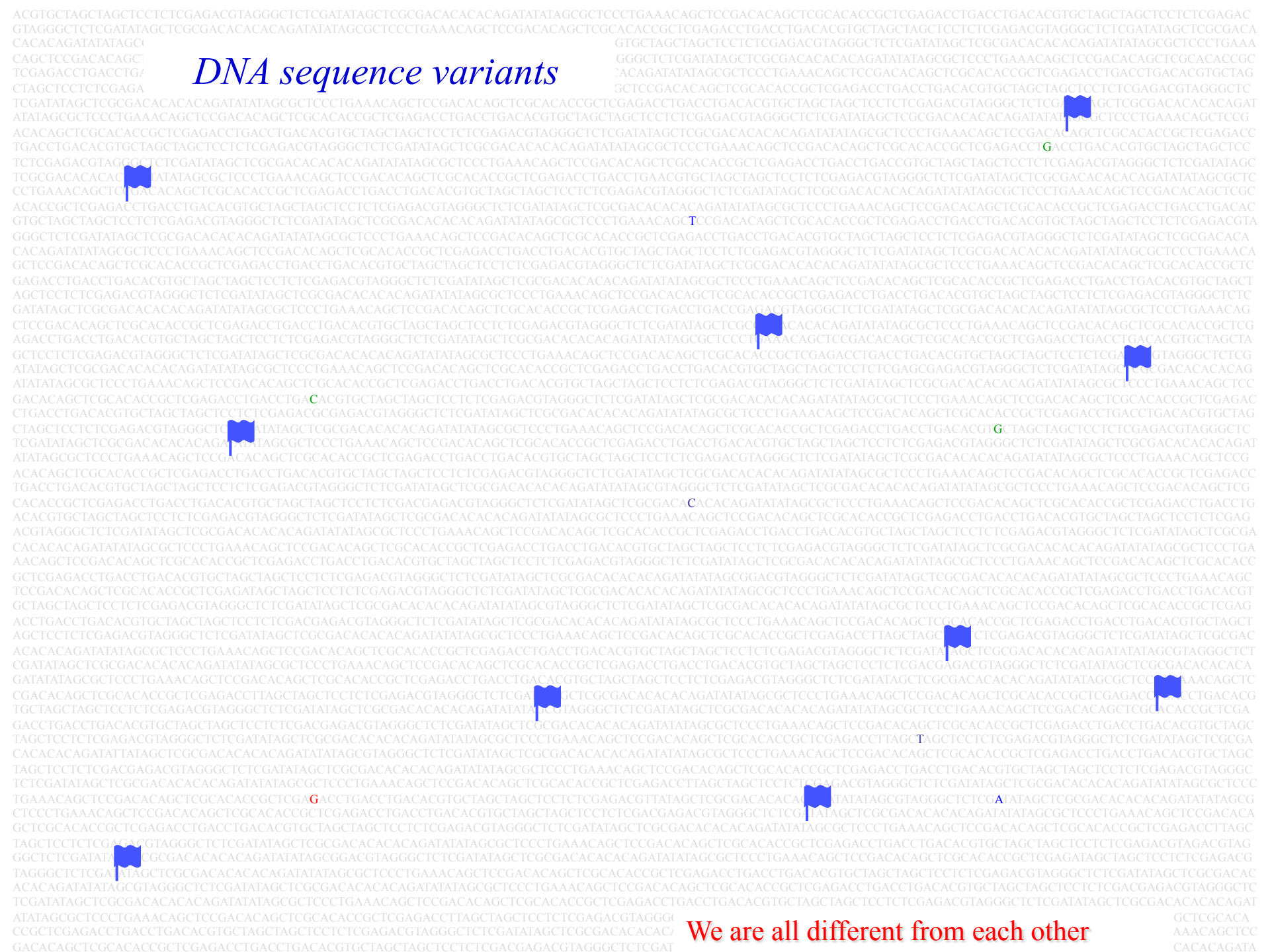


# Mutations & Polymorphisms

Mutations become polymorphisms or “common alleles” when frequency > 1% in a population (arbitrary)



# *DNA sequence variants*



We are all different from each other

# What is a SNP?

Example order of bases in a section of DNA on a chromosome:

...C C **A** T T G A C...
...G G **T** A A C T G...

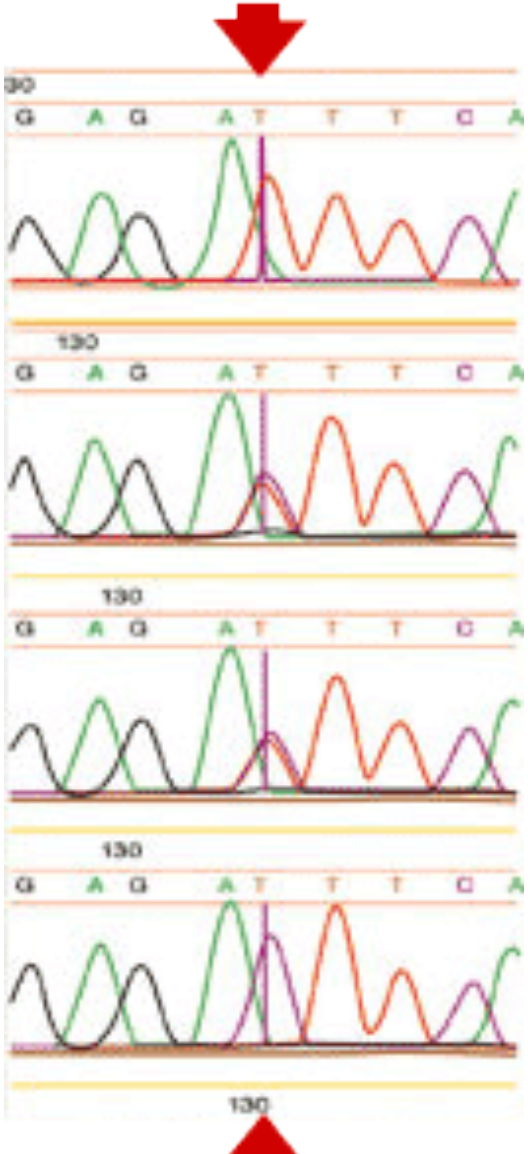
...C C **G** T T G A C...
...G G **C** A A C T G...

Some people have a different base at a given location

# SNPs

## single nucleotide polymorphisms

- Natural sequence variation occurs between any two copies of the human genome.
- Most variations are SNPs involving single base substitutions — the rest are insertions or deletions
- A SNP is detected by sequencing a particular region from different individuals, who may have identical (homozygous; T/T or C/C) or different (heterozygous; T/C) bases at the polymorphic site



# Each individual has a unique DNA sequence

## Our uniqueness lies in just 0.1% of our DNA sequence

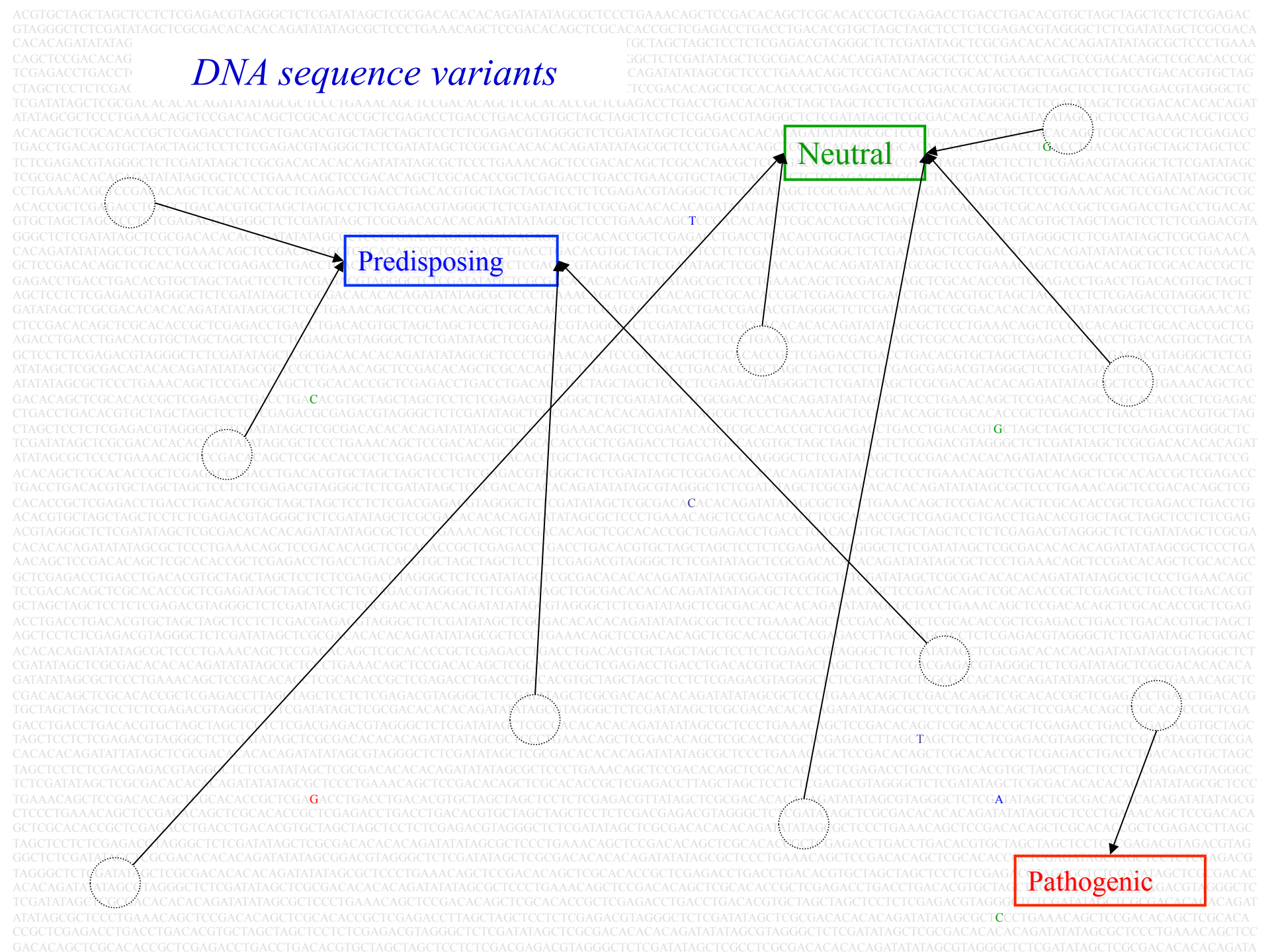
- The DNA sequences of any two people are 99.9% identical
- 1 difference in every 1,000 nucleotides
- 3 million total nucleotide differences

**DNA sequence variant 1:**

…AGGTTCAGGCATCAGATTCGCAATCGCTTG
AGCAATCGCTTGCAGATACGAAAGCTTATACC
TATGTCCTAGGTCAGTGTTTCAAAAAGTTTGT
TCCATAAAAAGTAACATTGTGCTGCAGGATTT
CTCAGACGGACCAGTTTGCTAAAGTACTCCGG
GTGTCTCCACAAAGCTTACATAGAATGTGAAG
CTTACAAAACTCATCAGACAAGAGATCATCTC
CTGGACTGAGTTTAAAACACAATTTGGAAA…

**DNA sequence variant 2:**

…AGGTTCAAGCATCAGATTCGCAATCGCTTG
AGCAATCGCTTGCAGATACGAAAGCTTATACC
TATGTCCTAGGTCAGTGTTTCAAAAAGTTTGT
TCCATAAAAAGTAACATTGTGCTGCAGGATTT
CTCAGACGGACCAGTTTGCTAAAGTACTCCGG
GTGTCTCCACAAAGCTTACATAGAATGTGAAG
CTTACAAAACTCATCAGACAAGAGAACATCTC
CTGGACTGAGTTTAAAACACAATTTGGAAA…


DNA sequence variants
Neutral
Predisposing
Pathogenic

# 1000 genomes project

- On average each individual has:
- 10.500-12.000 synonimous SNPs
- 10.000-10.800 SNPs non synonimous SNPs out of which:
  - Stop codon loss 4-14
  - New stop codons 67-100
  - Splice-sites loss 28-45
  - Frame-shift 192-280
  - Large deletions 33-49
- **Total deleterious variations 240-345**

# How do we differ? – Let me count the ways

- **Single nucleotide polymorphisms**
  - 1 every few hundred bp, mutation rate* $\approx 10^{-9}$

TGCATTGCGTAGGC
TGCATTCCGTAGGC

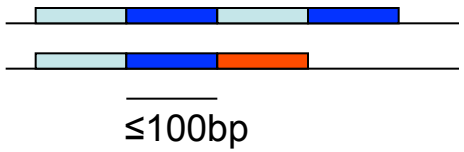
- **Short indels (=insertion/deletion)**
  - 1 every few kb, mutation rate v. variable

TGCATT---TAGGC
TGCATTCCGTAGGC

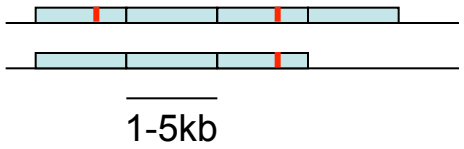
- **Microsatellite (STR) repeat number**
  - 1 every few kb, mutation rate $\leq 10^{-3}$

TGCTCATCATCATCAGC
TGCTCATCA------GC

- **Minisatellites**
  - 1 every few kb, mutation rate $\leq 10^{-1}$




- **Repeated genes**
  - rRNA, histones




- **Large inversions, deletions**
  - Rare, e.g. Y chromosome

*per generation

Science

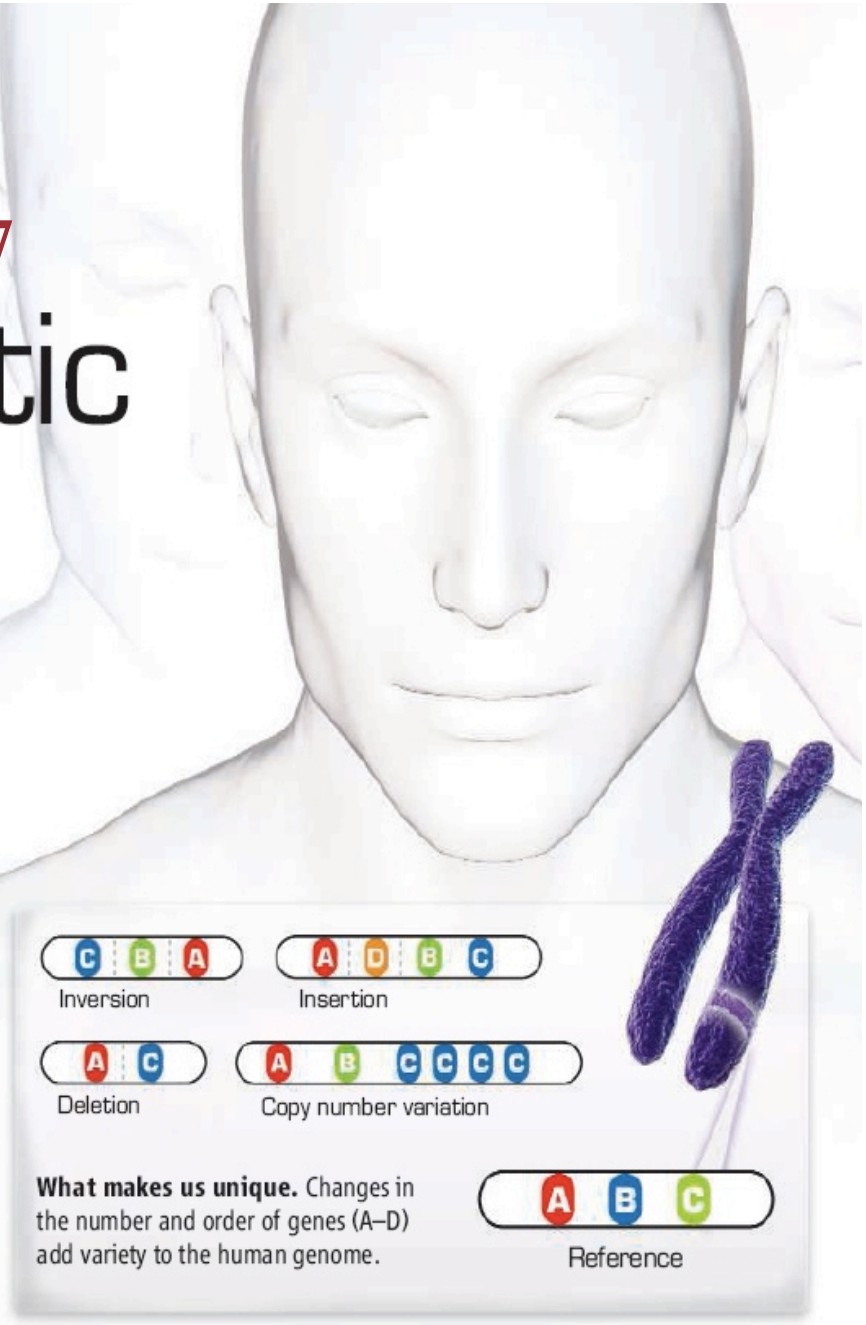
## BREAKTHROUGH OF THE YEAR 2007

# Human Genetic Variation

**Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another**

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.




**What makes us unique.** Changes in the number and order of genes (A–D) add variety to the human genome.

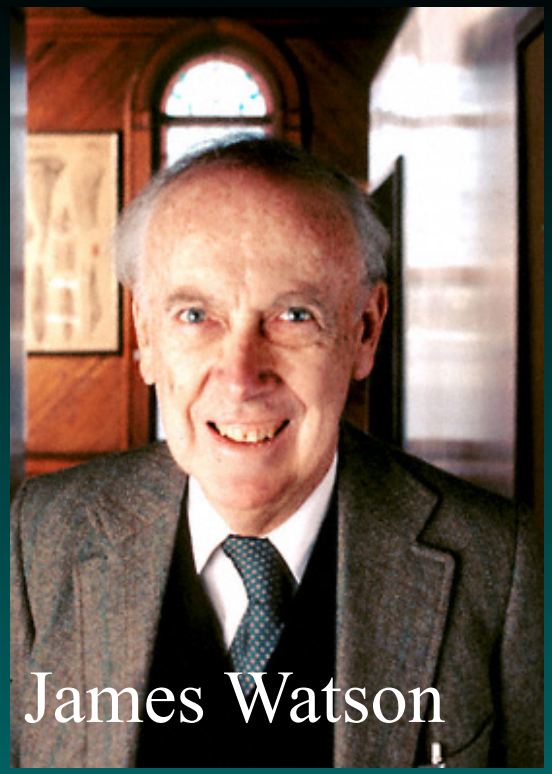
# If we consider variations in individual genes, the frequencies of different variations – or alleles – can be great among populations from different geographic origin

Rs 4821480

**Europeans**

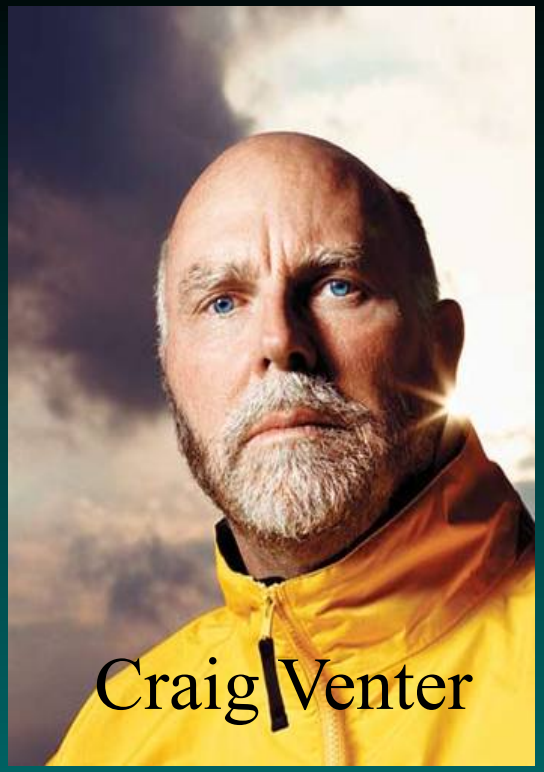

**Africans**


www.harunyahya.org



James Watson

7,648 protein coding changes



Craig Venter

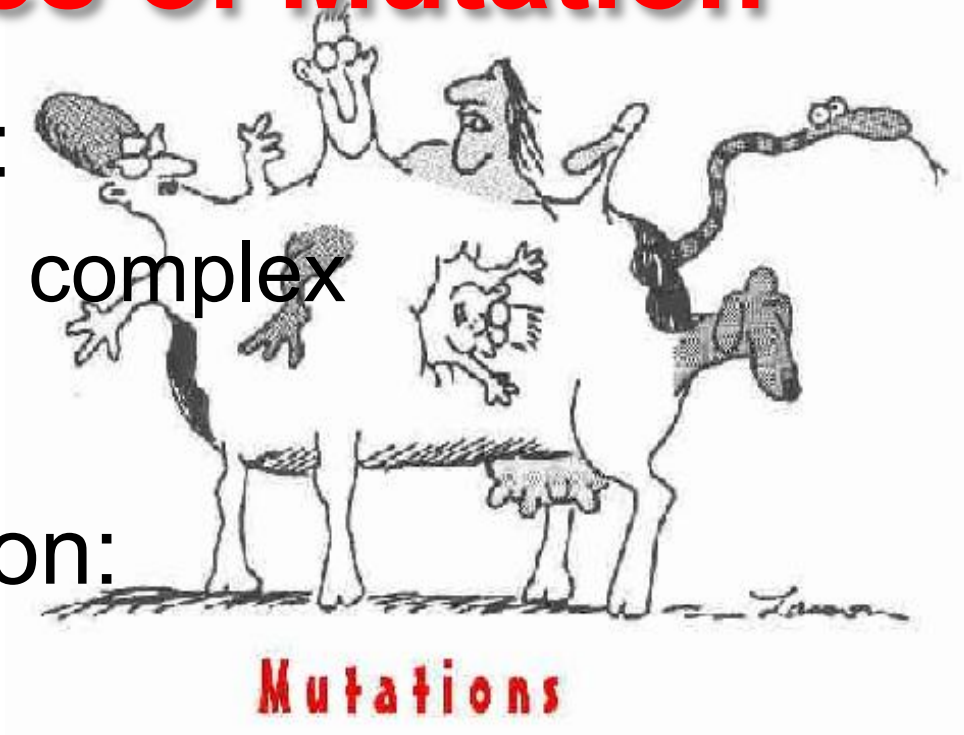
3,766 protein coding changes

3,882 protein coding changes

# Consequences of Mutation

- Genetic Variability:

- adaptation, evolution, complex traits




- Deleterious mutation:

- Germ cells

  hereditary disease

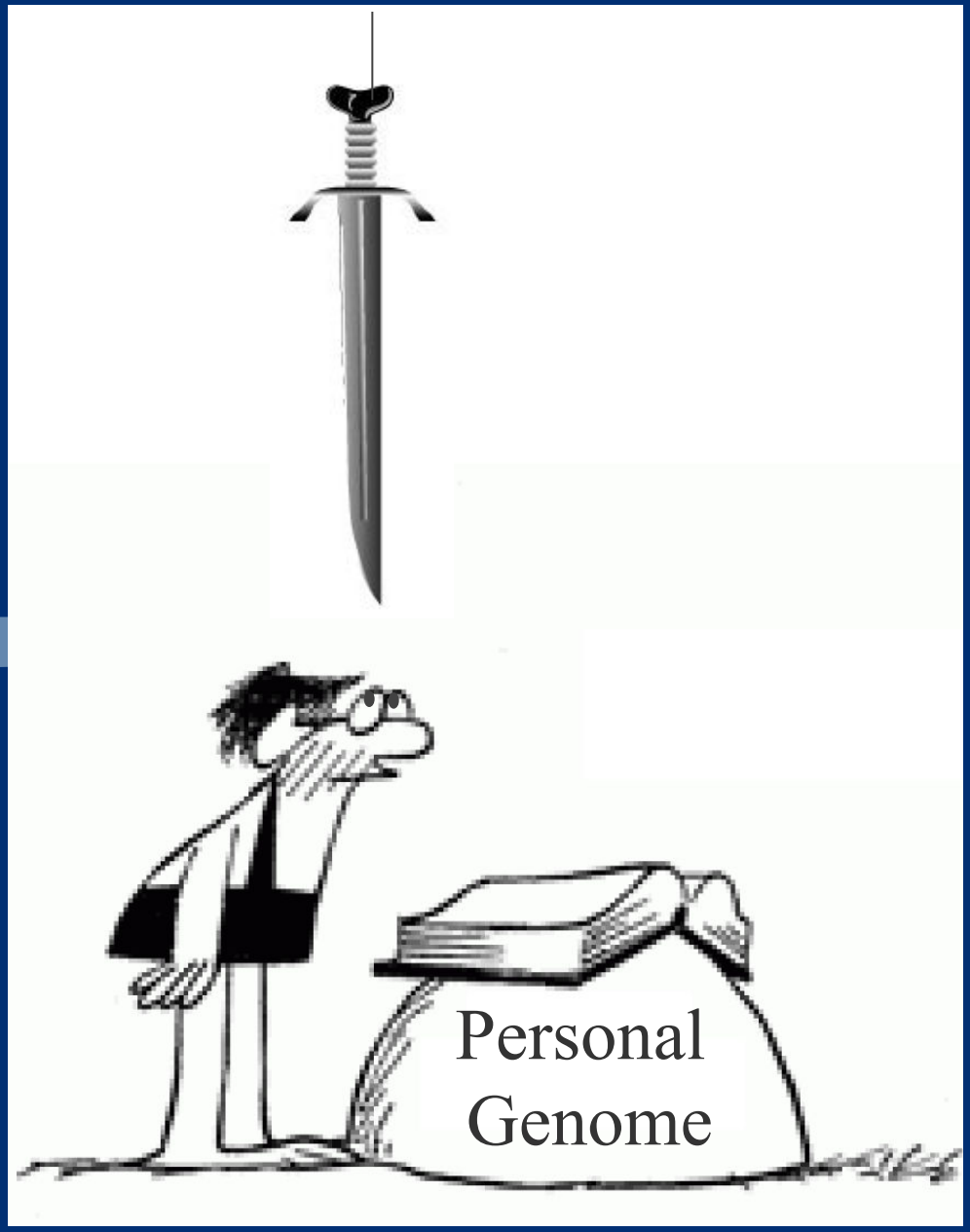
- Somatic cells

  malignancies, atherosclerosis, etc.

http://www.ebi.ac.uk/mutations/


IT'S ALL VERY SCIENTIFIC THESE DAYS

**Do we really want to know the sequence of our genome…..?**




claudia.giachino@unito.it