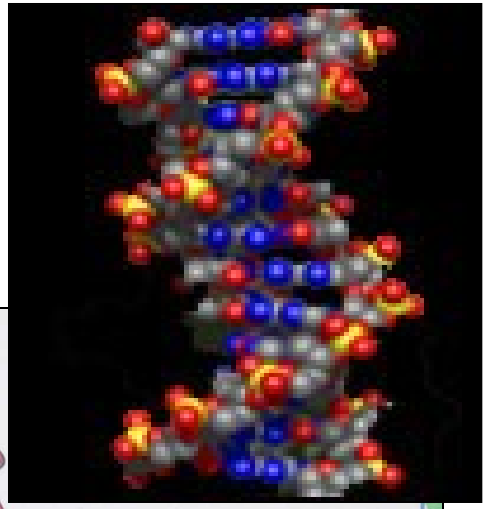
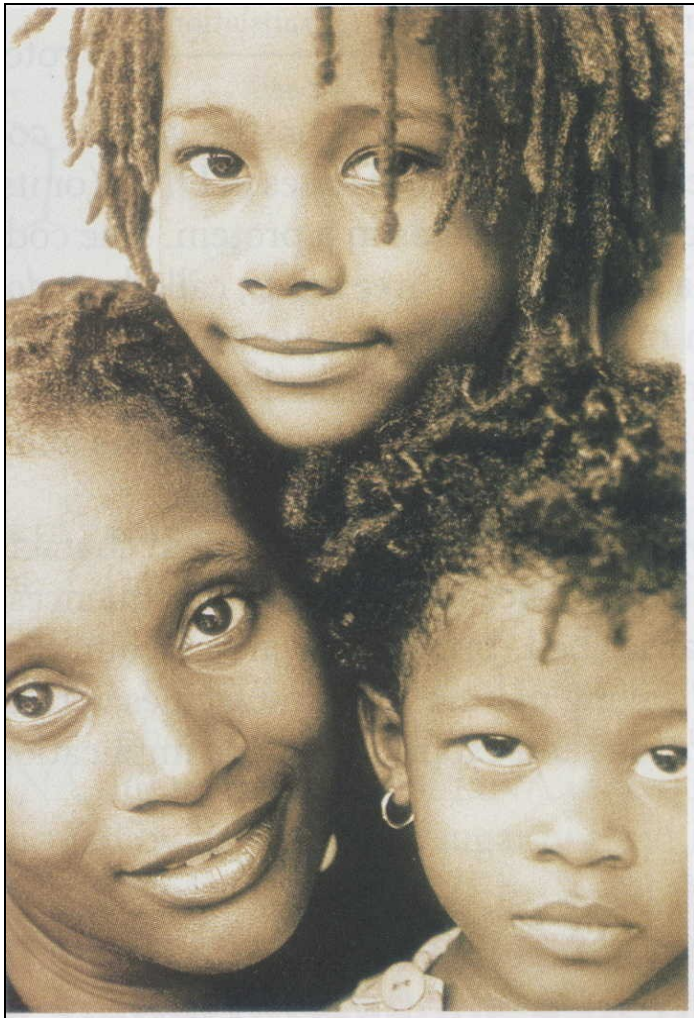
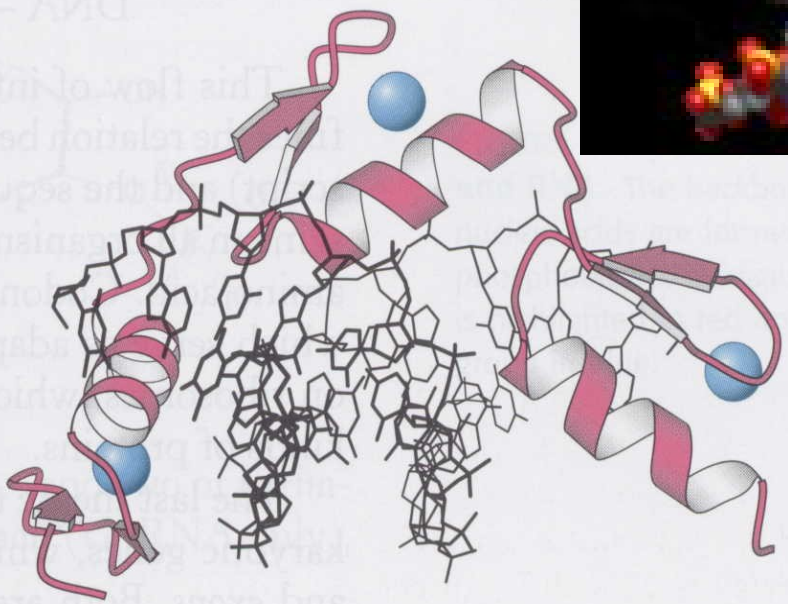
Nucleic acids

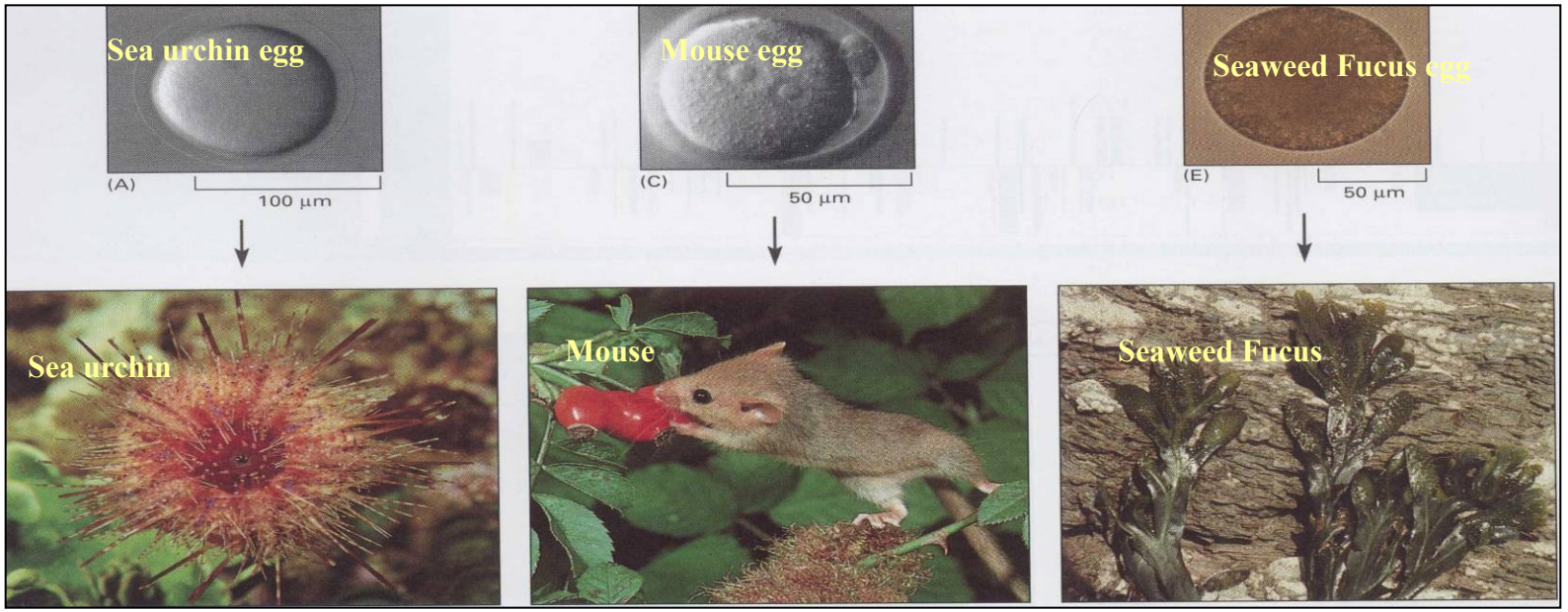






Having genes in common accounts for the resemblance of a mother and her daughters. Genes must be expressed to exert an effect, and proteins regulate such expression. One such regulatory protein, a zinc-finger protein (zinc ion is blue, protein is red), is shown bound to a control or promoter region of DNA (black).
[(Left) Barnaby Hall/Photonica.]

DNA determines the nature of the future organism



- **The hereditary information** in the egg cell determines the nature of the whole multicellular organism (mouse egg---mouse)
- Each species is different, and **each reproduces itself faithfully**, yielding progeny that belong to the same species; the parent organism hands down information specifying, in extraordinary detail, the characteristics that the offspring shall have
- **Heredity is a central part of definition of life**: it distinguishes life from other processes (the growth of crystals, the burning of a candle, the formation of waves on water), in which **orderly structures** are generated but without the **same link** between the peculiarities of **parents and peculiarities** of offspring

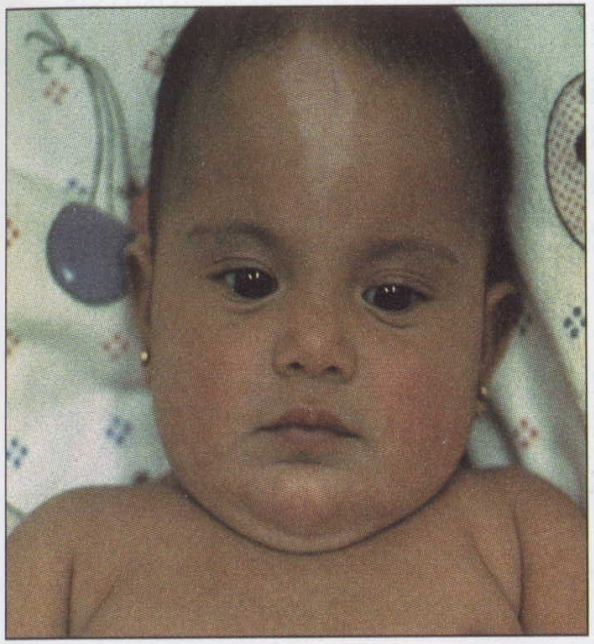


Figure 1–53 Human and mouse: similar genes and similar development. The human baby and the mouse shown here have similar white patches on their foreheads because both have mutations in the same gene (called *kit*), required for the development and maintenance of pigment cells. (From R.A. Fleischman, *Proc. Natl. Acad. Sci. USA* 88:10885–10889, 1991. © National Academy of Sciences.)

```
GTTCCGGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCC  human
GCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTTAAAGGAATTGGCGGGGGAGCACTACAACGGGTGGAGCCTGCGGTTTAATTGGATTCAACGCCGGGCATCTTACCA  Methanococcus
ACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGC.ACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCT  E. coli
GTTCCGGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCC  human
```

Figure 1–22 Genetic information conserved since the beginnings of life. A part of the gene for the smaller of the two main RNA components of the ribosome is shown. Corresponding segments of nucleotide sequence from an archaean *(Methanococcus jannaschii)*, a eubacterium *(Escherichia coli)* and a eucaryote *(Homo sapiens)* are aligned in parallel. Sites where the nucleotides are identical between species are indicated by a vertical line; the human sequence is repeated at the bottom of the alignment so that all three two-way comparisons can be seen. A dot halfway along the *E. coli* sequence denotes a site where a nucleotide has been either deleted from the eubacterial lineage in the course of evolution, or inserted in the other two lineages. Note that the sequences from these three organisms, representative of the three domains of the living world, all differ from one another to a roughly similar degree, while still retaining unmistakable similarities.

- All cells store their hereditary information in the same linear chemical code (DNA)
- All cells replicate their hereditary information by templated polymerization
- All cells trnscribe portions of their hereditary information into the same intermediary form
- All cells use proteins as catalysts
- All cells translate RNA into protein in the same way
- In all cells: one gene = one protein

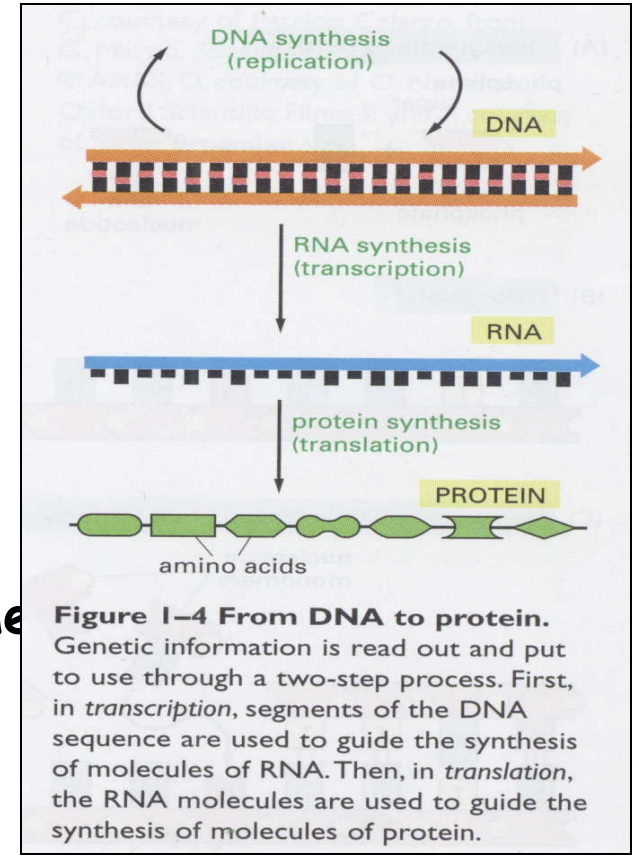
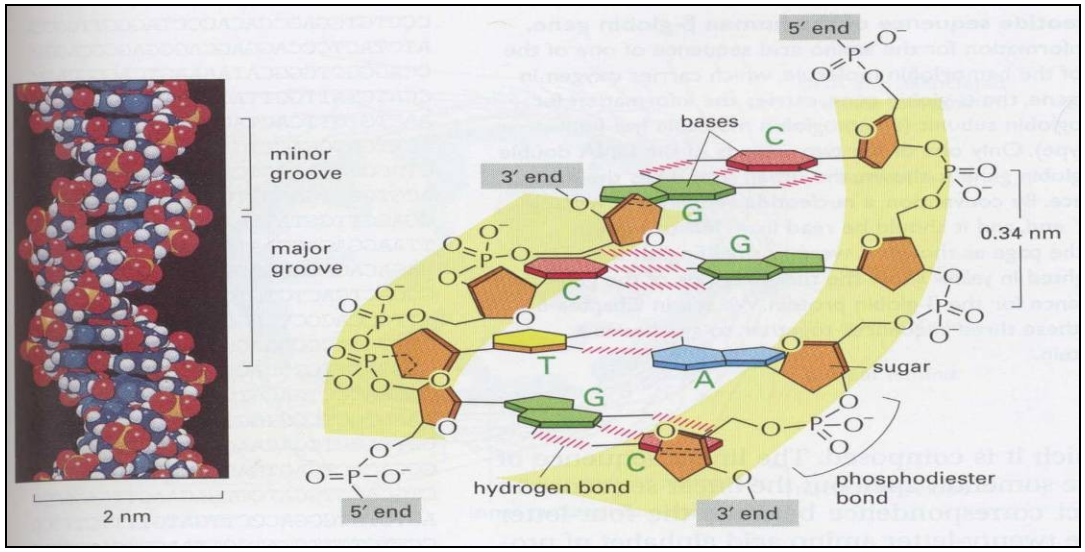
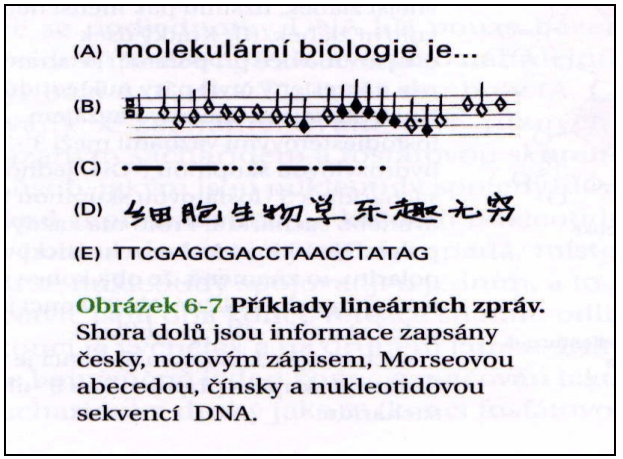


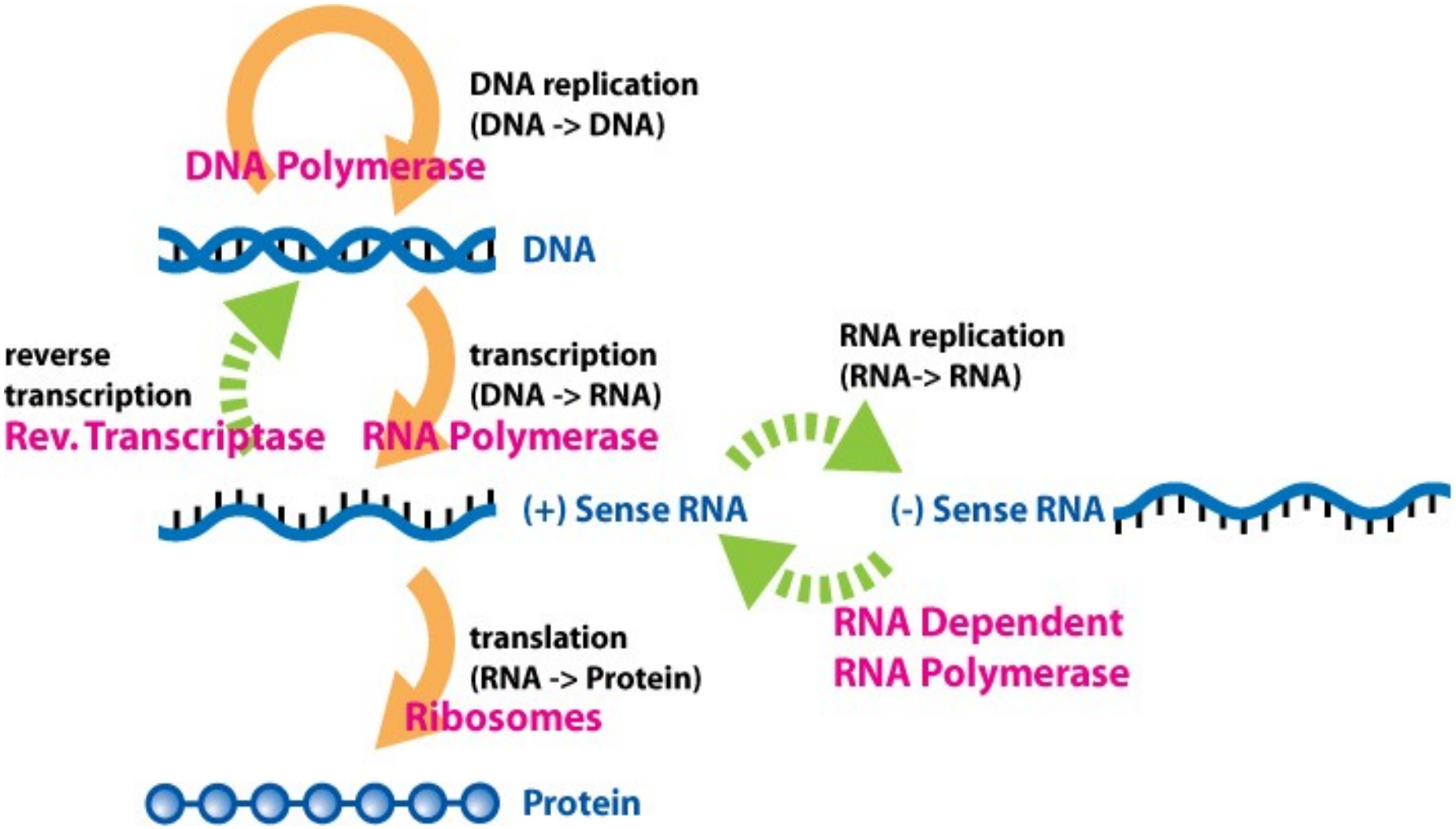
Figure 1–4 From DNA to protein. Genetic information is read out and put to use through a two-step process. First, in *transcription*, segments of the DNA sequence are used to guide the synthesis of molecules of RNA. Then, in *translation*, the RNA molecules are used to guide the synthesis of molecules of protein.





Obrázek 6-7 Příklady lineárních zpráv. Shora dolů jsou informace zapsány česky, notovým zápisem, Morseovou abecedou, čínsky a nukleotidovou sekvencí DNA.

Central dogma of molecular biology



https://es.m.wikipedia.org/wiki/Archivo:Extended_Central_Dogma_with_Enzymes.jpg

Basic principles of molecular biology

1. The information encoded within DNA, which directs the functioning of living cells and is transmitted to offspring, consists of a **specific sequence** of **nitrogenous bases.**

2. The physiological and genetic **function of DNA** requires the synthesis of relatively **error-free copies. DNA** synthesis involves the complementary pairing of nucleotide bases**.**

3. The **mechanism** by which genetic information is **utilized** to direct cellular processes involves the **synthesis** of another type of nucleic acid called **ribonucleic acid (RNA).**

Basic principles of molecular biology

4. RNA synthesis occurs through the **complementary pairing** of ribonucleotide bases with the bases in a DNA molecule.

5. Several types of RNA are responsible for the synthesis of the enzymes, structural proteins, and other polypeptides, that are required for organismal function.
6. **Central dogma of molecular biology“** describes the flow of genetic information from DNA through RNA and eventually to proteins.

DNA synthesis = replication; DNA-dependent RNA synthesis = transcription, protein synthesis = translation

Proof of DNA as a carrier of genetic information

Indirect evidence:

- **DNA** is localized on **chromosomes**, RNA and proteins are in cytoplasm

- **The amount of DNA** in somatic cells is in **correlation** with number of c**hromosomes**, sex cells carry half of the DNA amount

- DNA is **more stable** than RNA or proteins

Direct evidence:

- **Transformation** in *Streptococcus pneumoniae* – **change** of v virulence

- Analysis of Enterobacteria phage T2 – into the bacterial **cell comes only DNA, not proteins**

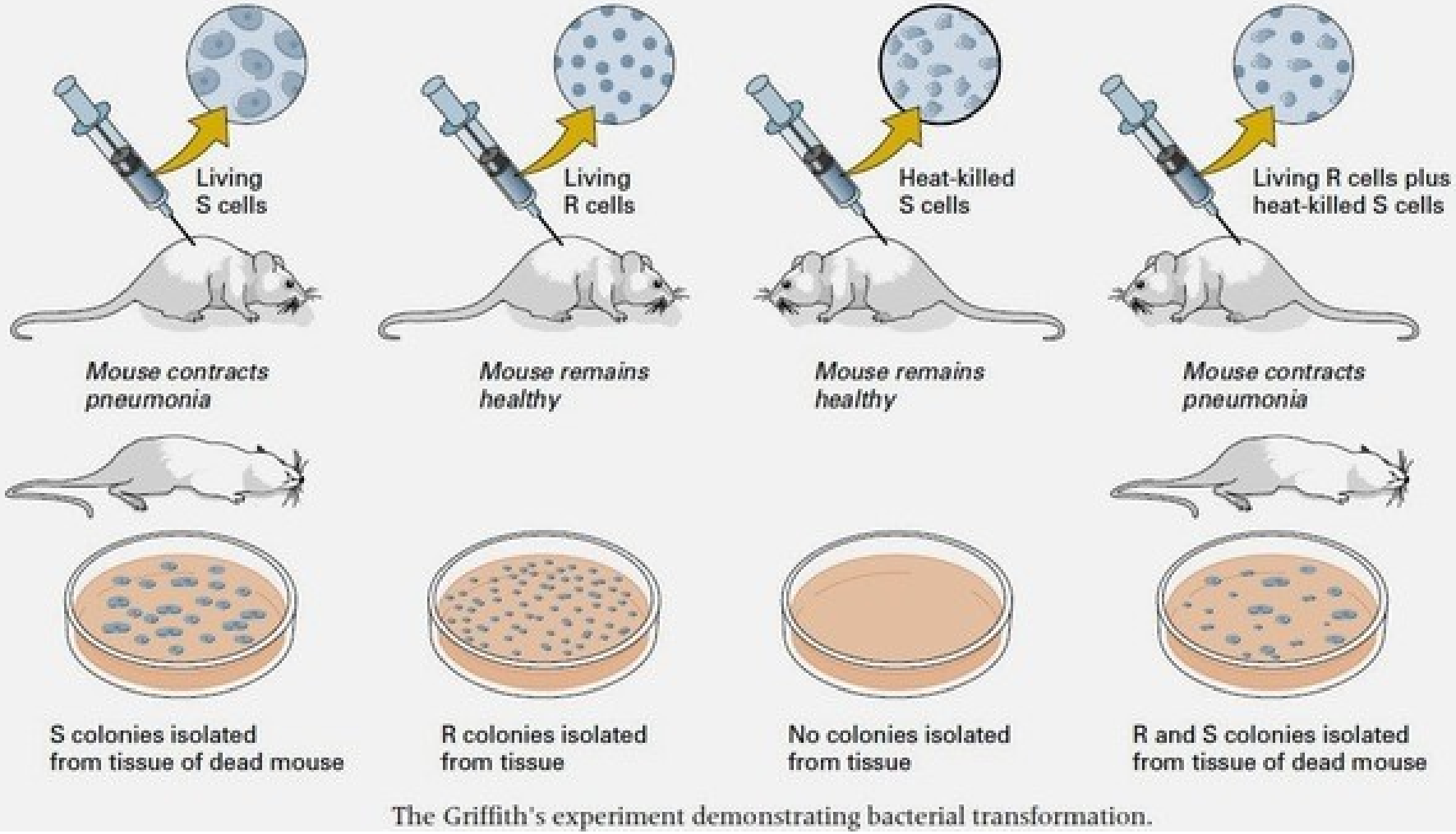
- RNA viruses: the carrier of genetic information is RNA (coronaviruses, HIV)

* 1928 British scientist – Frederick Griffith

  * Wanted to know how bacteria made people sick, especially pneumonia

* Griffith isolated 2 types of pneumonia bacteria:
  * S strain – harmful bacteria, smooth edges 
  * R strain – harmless bacteria, rough edges 

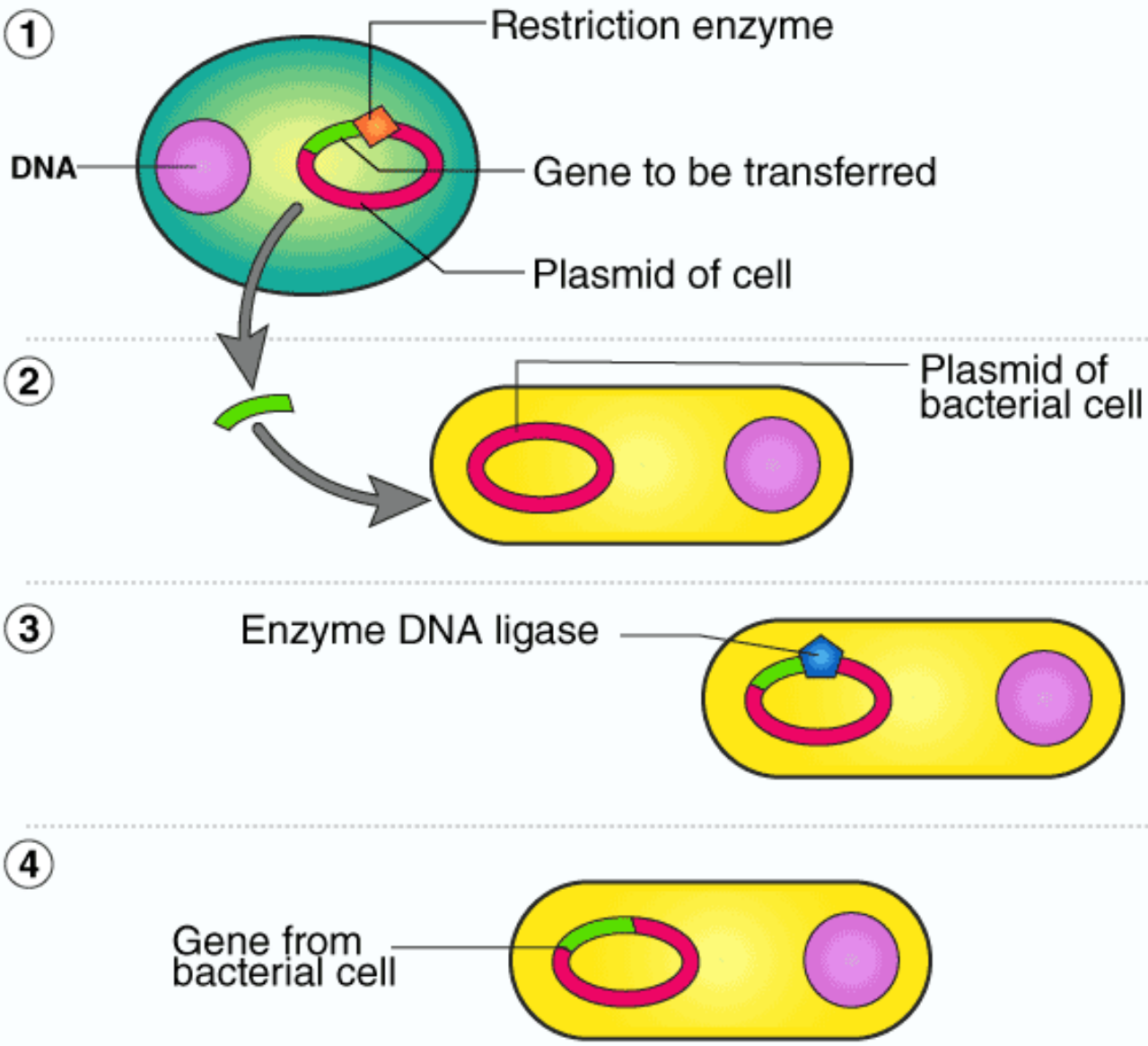
Transformation in *Streptococcus pneumoniae* *(Griffith 1928) – „transforming factor“*



The Griffith's experiment demonstrating bacterial transformation.

https://byjus.com/biology/griffith-experiment-genetic-material/

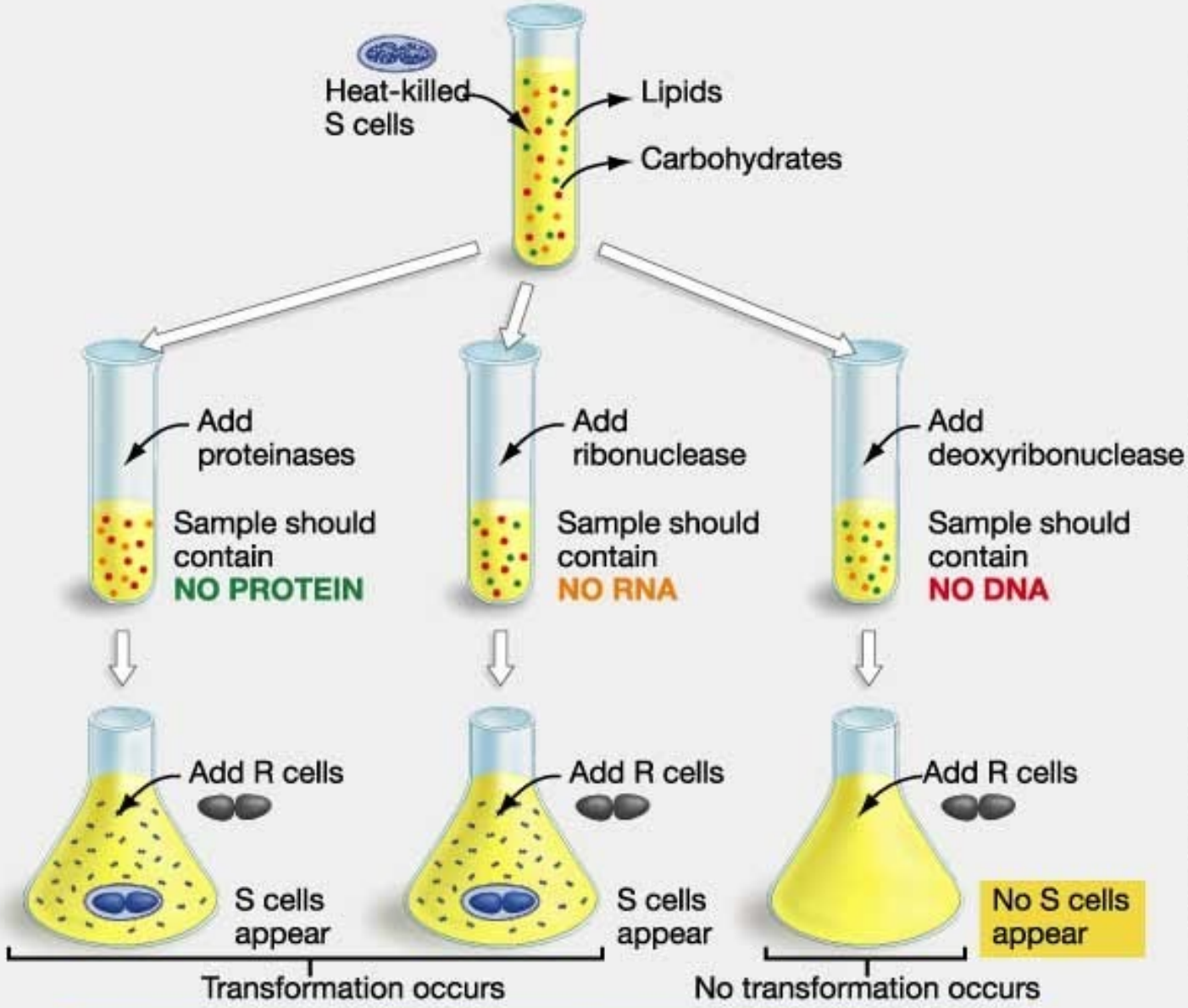
BACTERIAL TRANSFORMATION



What substance is the transforming factor?

(O. Avery, C. Mac Leod, M. McCarty 1944)

DETERMINING THAT DNA IS THE HEREDITARY MATERIAL



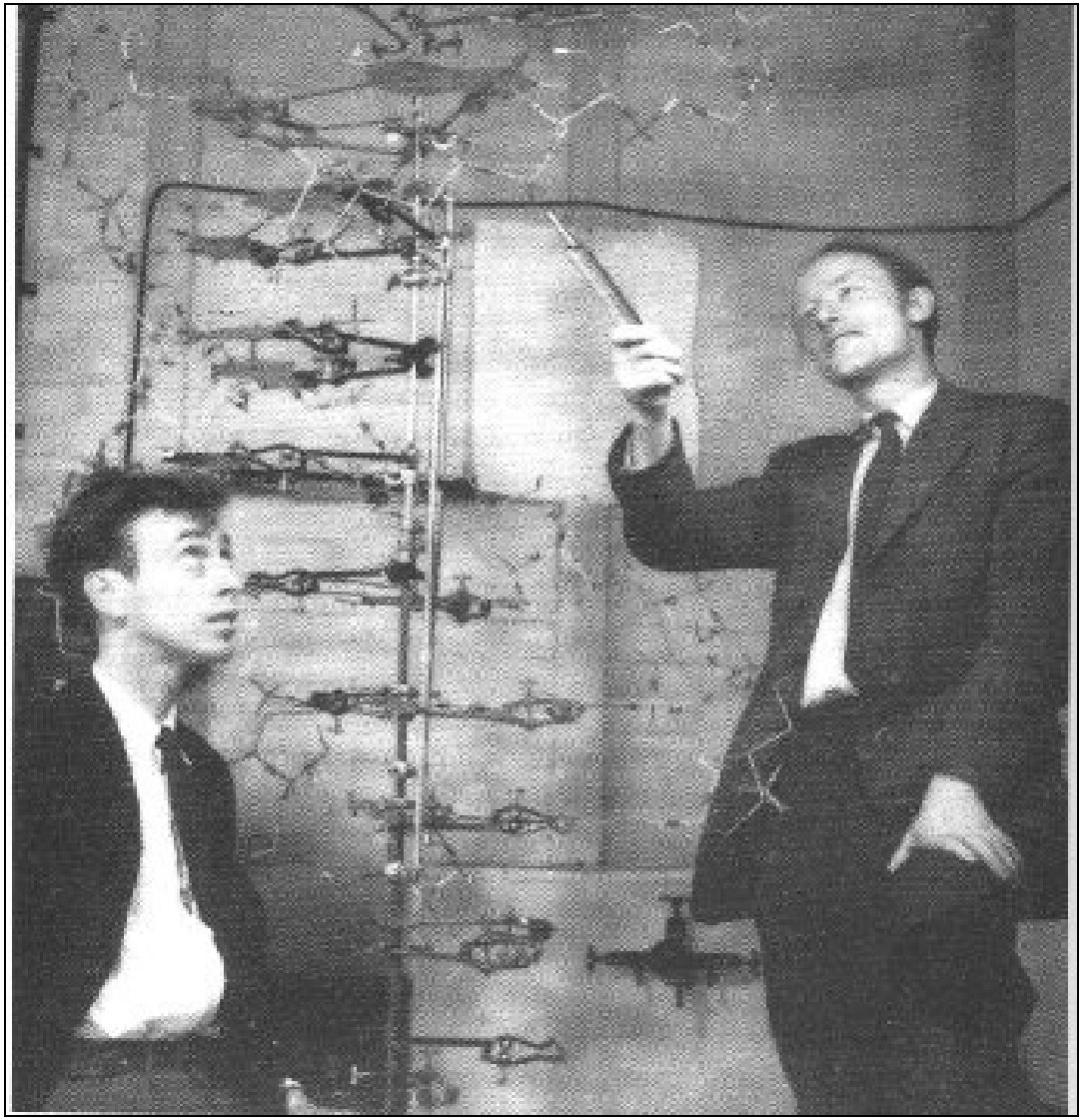
1. Remove the lipids and carbohydrates from a solution of heat-killed S cells. Proteins, RNA, and DNA remain.

2. Subject the solution to treatments of enzymes to destroy either the proteins, RNA, or DNA.

3. Add a small portion of each sample to a culture containing R cells. Observe whether transformation has occurred by testing for the presence of virulent S cells.

Conclusion: Transformation cannot occur unless DNA is present. Therefore, DNA must be the hereditary material.

1953 - J. Watson; F. Crick; M. Wilkins, R. Franklin (1962 – the Nobel Prize)



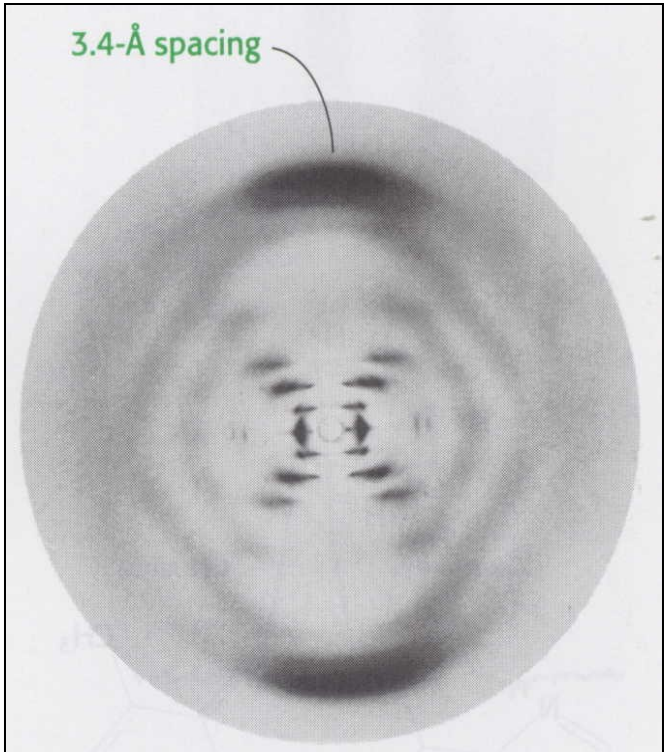


FIGURE 5.10 X-ray diffraction photograph of a hydrated DNA fiber. The central cross is diagnostic of a helical structure. The strong arcs on the meridian arise from the stack of nucleotide bases, which are 3.4 Å apart. [Courtesy of Dr. Maurice Wilkins.]

The information used to construct the Watson-Crick model included the following

- The chemical structures ans molecular dimensions of **deoxyribose, the nitrogenous bases and phosphate**
- The **1:1 ratios of adenine:thymine and guanine:cytosine** in the DNA isolated from a wide variety of species investigated by Erwin **Chargaff - Chargaff s rules**
- Superb X-ray diffraction studies performed by Rosalind Franklin indicating that **DNA is a symmetrical molecule and probably a helix**
- The diameter and pitch of the helix estimatd by Maurice Wilkins and his colleague Alex Stokes from other X-ray diffraction studies

Deoxyribonucleic acid - structure

- **The antiparallel orientation** of the two polynucleotide strands allows the formation of **hydrogen bonds** between the nitrogenous bases that are **oriented toward the helix interior**

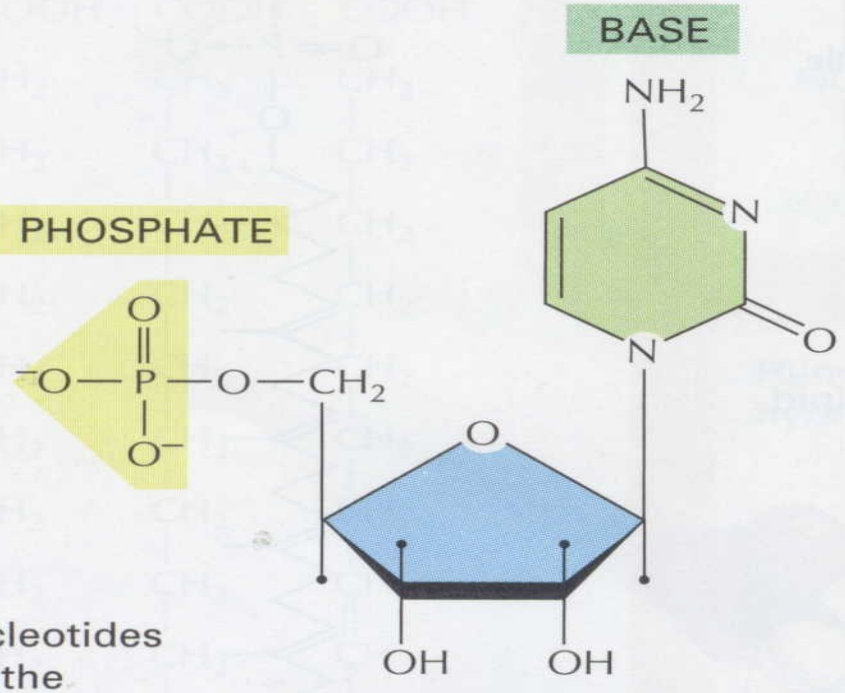
- There are two types of base pair (bp) in DNA:
  A-T (adenine-thymine), G-C (guanine-cytosine)

- Because **each base** pair is oriented **at a right angle** to the long axis of the helix, the **overall structure** of DNA resembles a **twisted staircase**

Nucleotides

NUCLEOTIDES

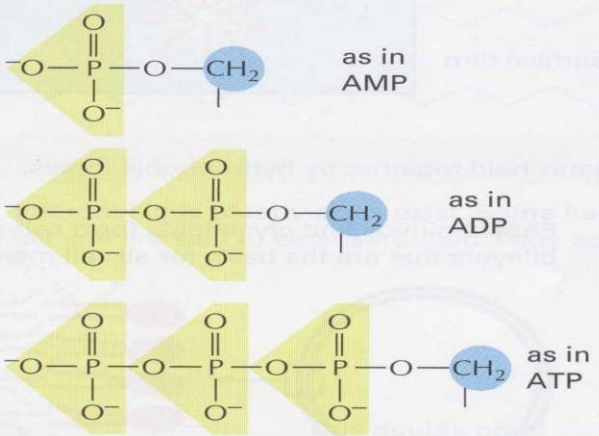
A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.



Nucleotides are the subunits of the nucleic acids.

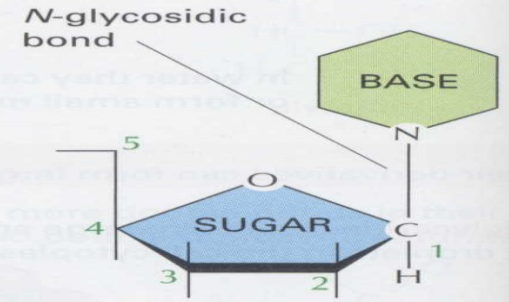
PHOSPHATES

The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar (designated 5'). Mono-, di-, and triphosphates are common.



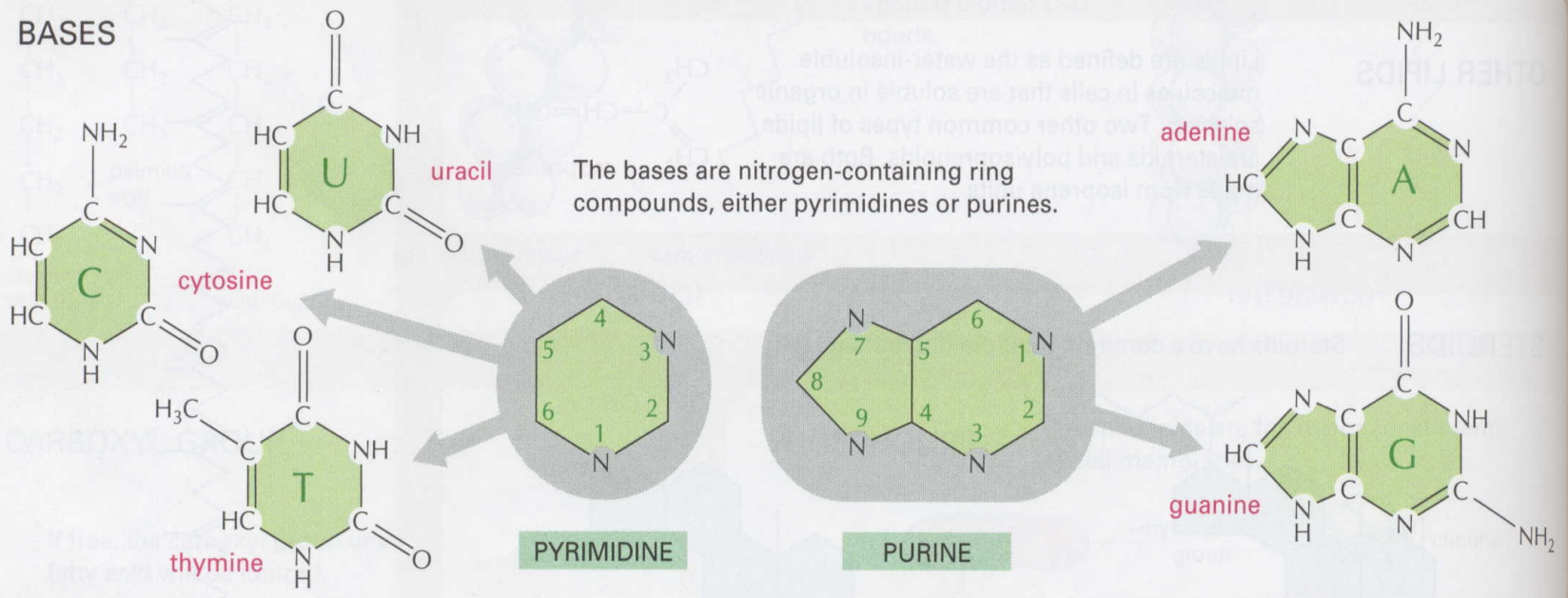
The phosphate makes a nucleotide negatively charged.

BASIC SUGAR LINKAGE



The base is linked to the same carbon (C1) used in sugar–sugar bonds.

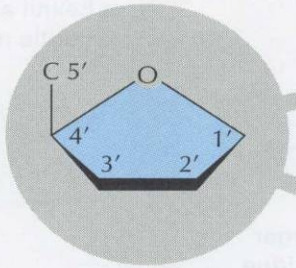
BASES



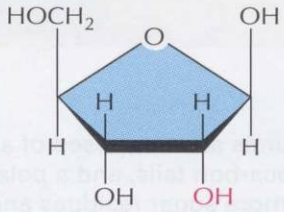
The bases are nitrogen-containing ring compounds, either pyrimidines or purines.

SUGARS

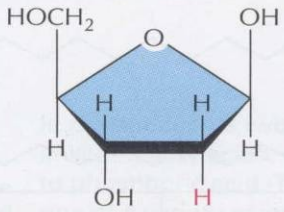
PENTOSE
a five-carbon sugar



two kinds are used



β-D-ribose
used in ribonucleic acid



β-D-2-deoxyribose
used in deoxyribonucleic acid

Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of the "5-prime carbon," etc.

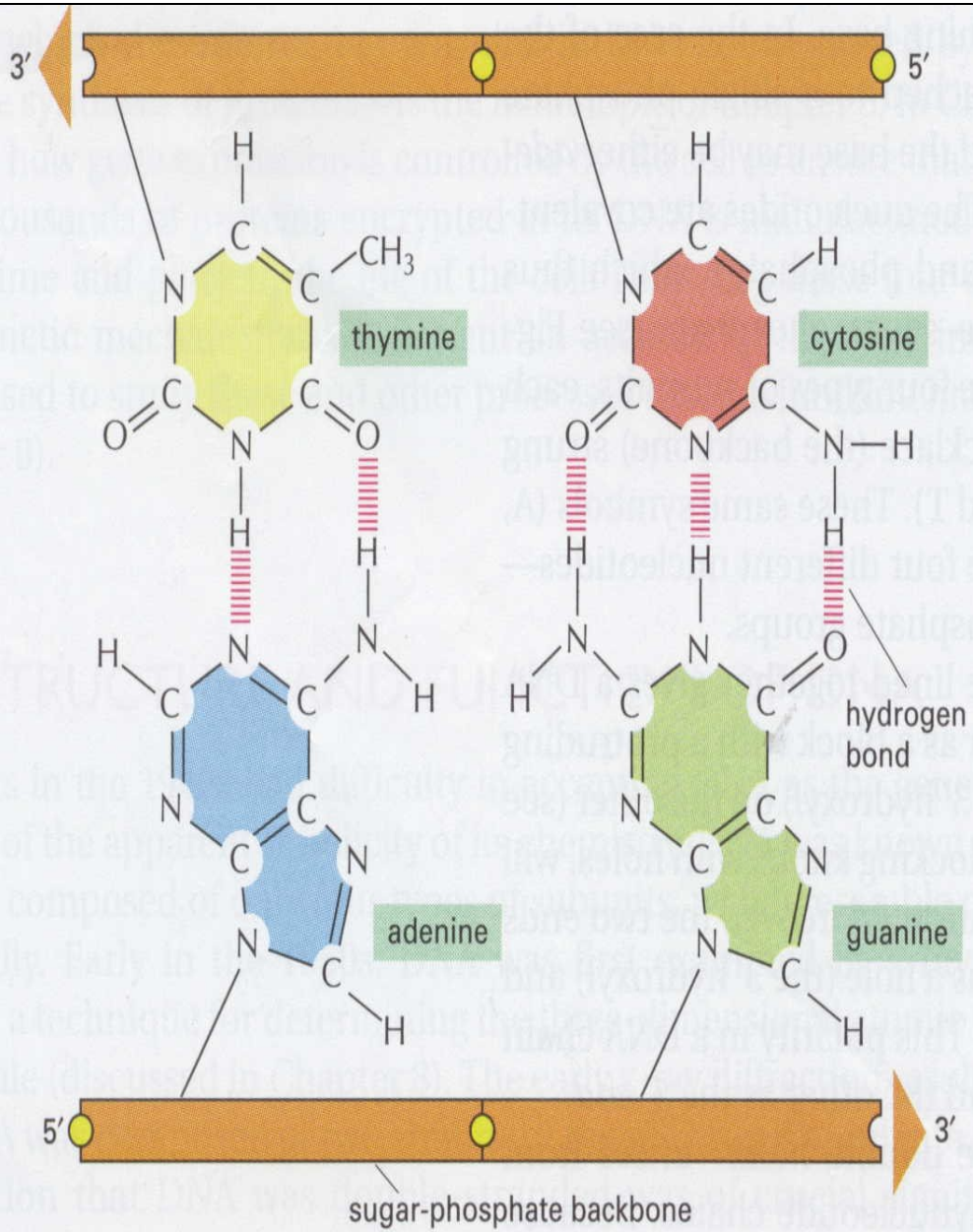


Figure 4–4 Complementary base pairs in the DNA double helix. The shapes and chemical structure of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C, where atoms that are able to form hydrogen bonds (see Panel 2–3, pp. 114–115) can be brought close together without distorting the double helix. As indicated, two hydrogen bonds form between A and T, while three form between G and C. The bases can pair in this way only if the two polynucleotide chains that contain them are antiparallel to each other.

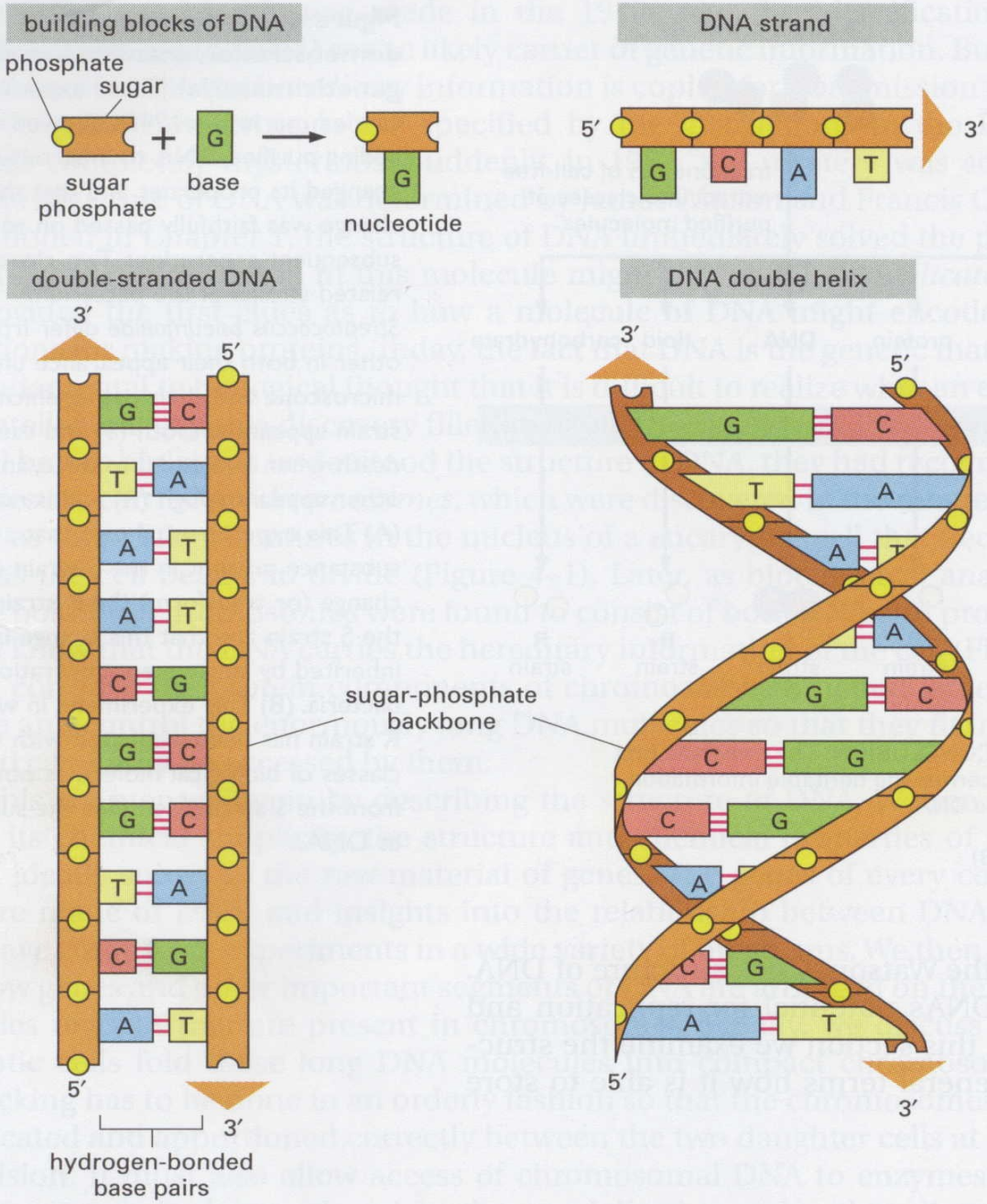


Figure 4–3 DNA and its building blocks. DNA is made of four types of nucleotides, which are linked covalently into a polynucleotide chain (a DNA strand) with a sugar-phosphate backbone from which the bases (A, C, G, and T) extend. A DNA molecule is composed of two DNA strands held together by hydrogen bonds between the paired bases. The *arrowheads* at the ends of the DNA strands indicate the polarities of the two strands, which run antiparallel to each other in the DNA molecule. In the diagram at the bottom left of the figure, the DNA molecule is shown straightened out; in reality, it is twisted into a double helix, as shown on the right. For details, see Figure 4–5.

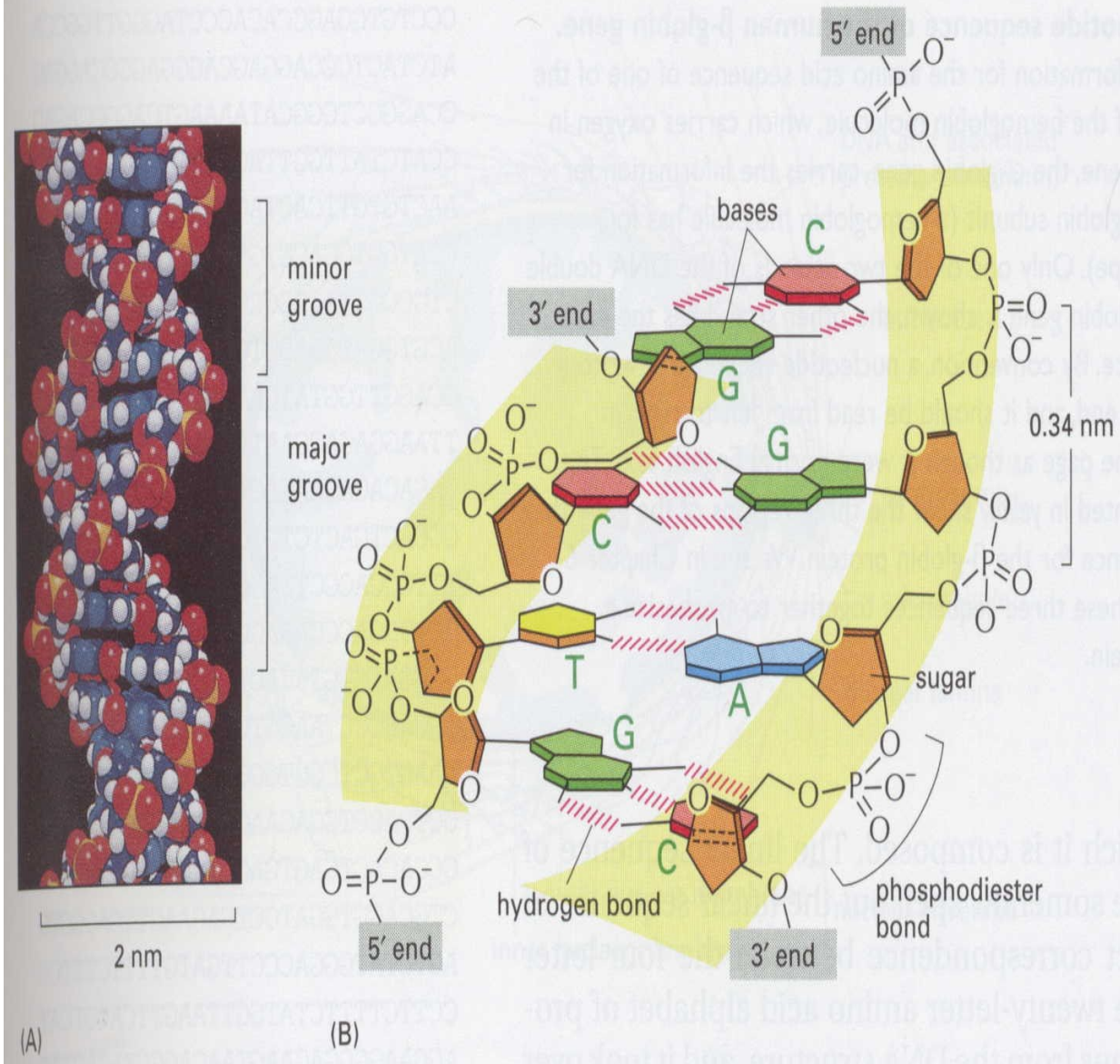


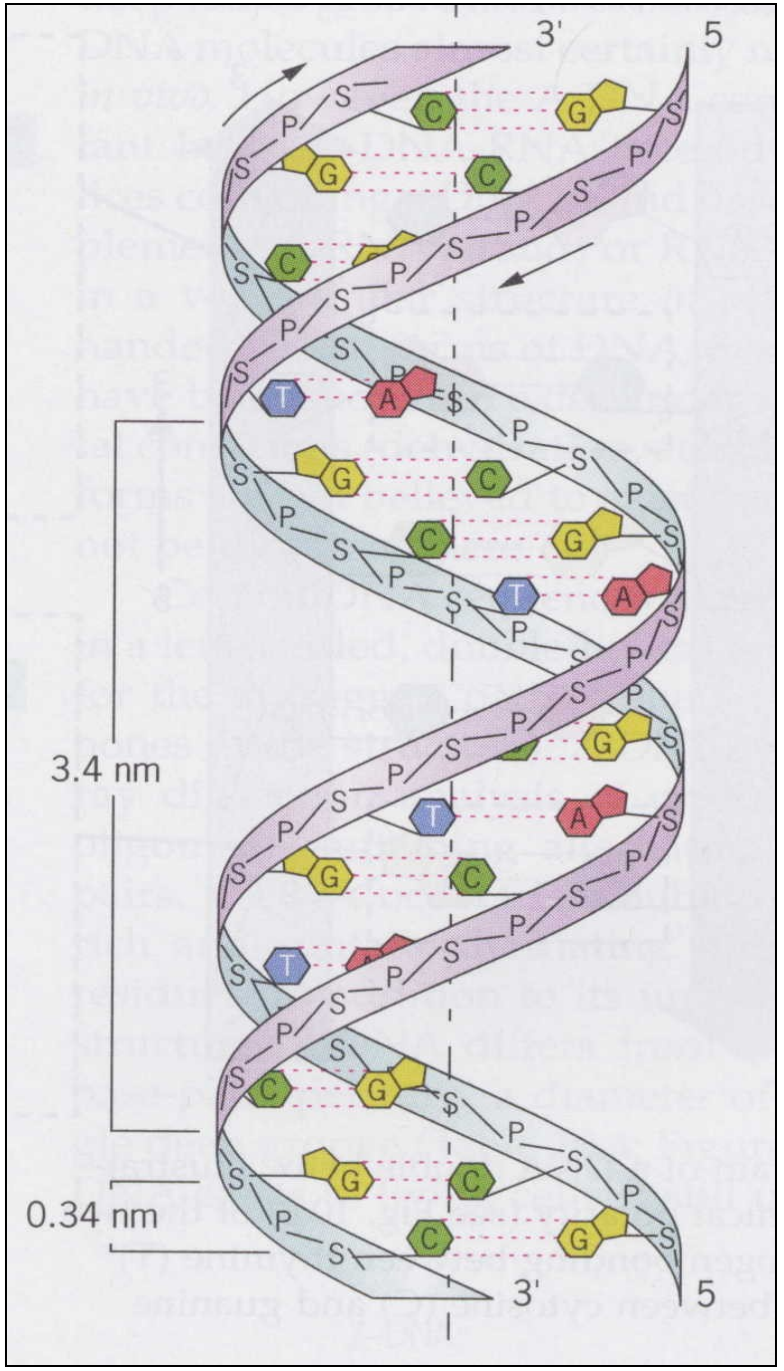
Figure 4–5 The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs and the center-to-center distance between adjacent nucleotide pairs is 3.4 nm. The coiling of the two strands around each other creates two grooves in the double helix. As indicated in the figure, the wider groove is called the major groove, and the smaller the minor groove. (B) A short section of the double helix viewed from its side, showing four base pairs. The nucleotides are linked together covalently by phosphodiester bonds through the 3′-hydroxyl (–OH) group of one sugar and the 5′-phosphate (P) of the next. Thus, each polynucleotide strand has a chemical polarity; that is, its two ends are chemically different. The 3′ end carries an unlinked –OH group attached to the 3′ position on the sugar ring; the 5′ end carries a free phosphate group attached to the 5′ position on the sugar ring.

The dimensions of crystalline DNA have been precisely measured

- **One turn of** the double helix spans **3.4 nm** and consists of approximately **10.4 bp**

- The **diameter** of the double helix is **2 nm**; there is sufficient space in the double helix interior only for base pairing between a purine and a pyrimidine

- The distance between adjacent base pairs is **0.34**



DNA stability

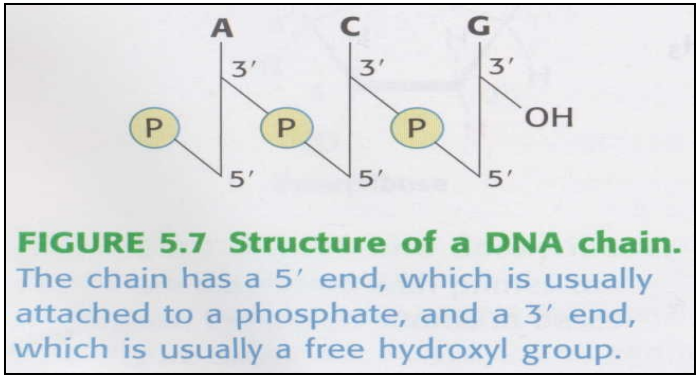
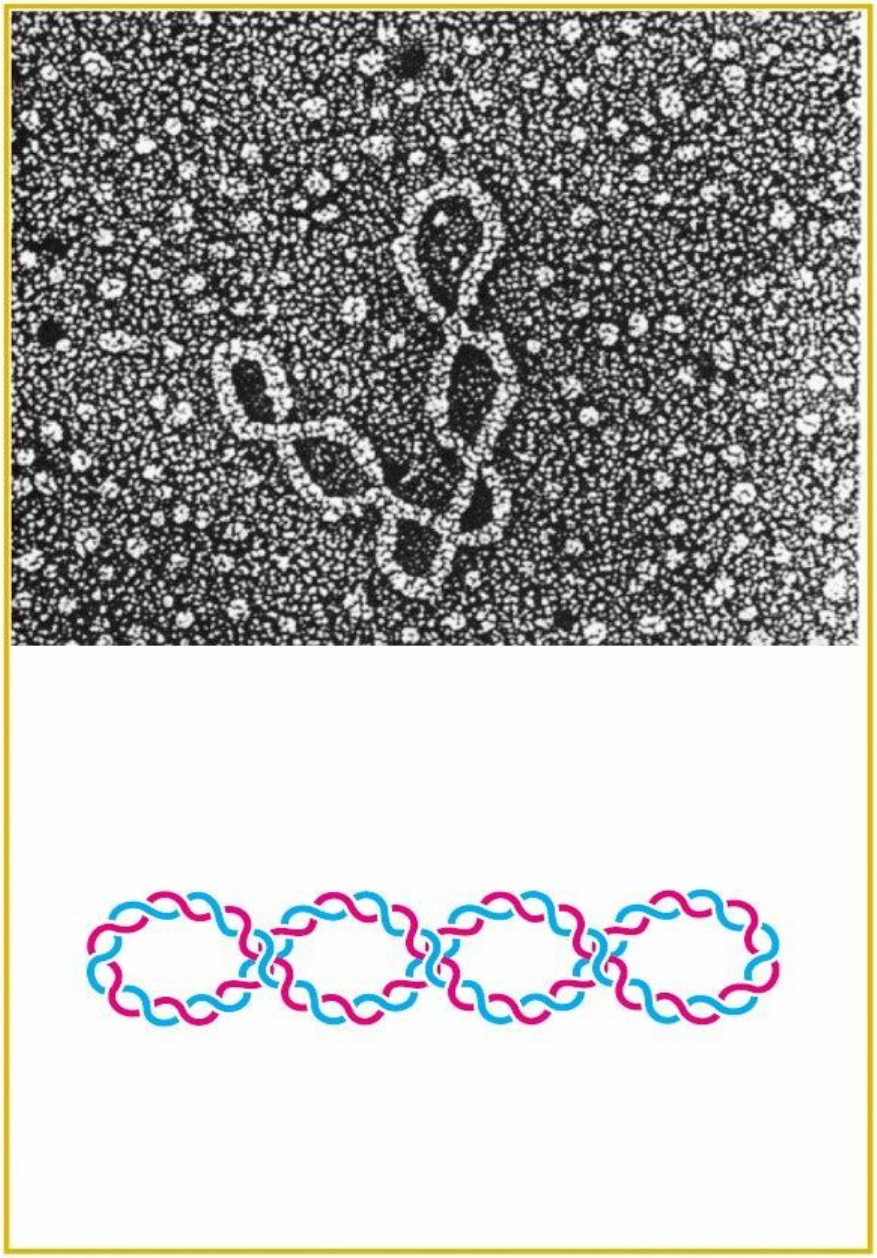


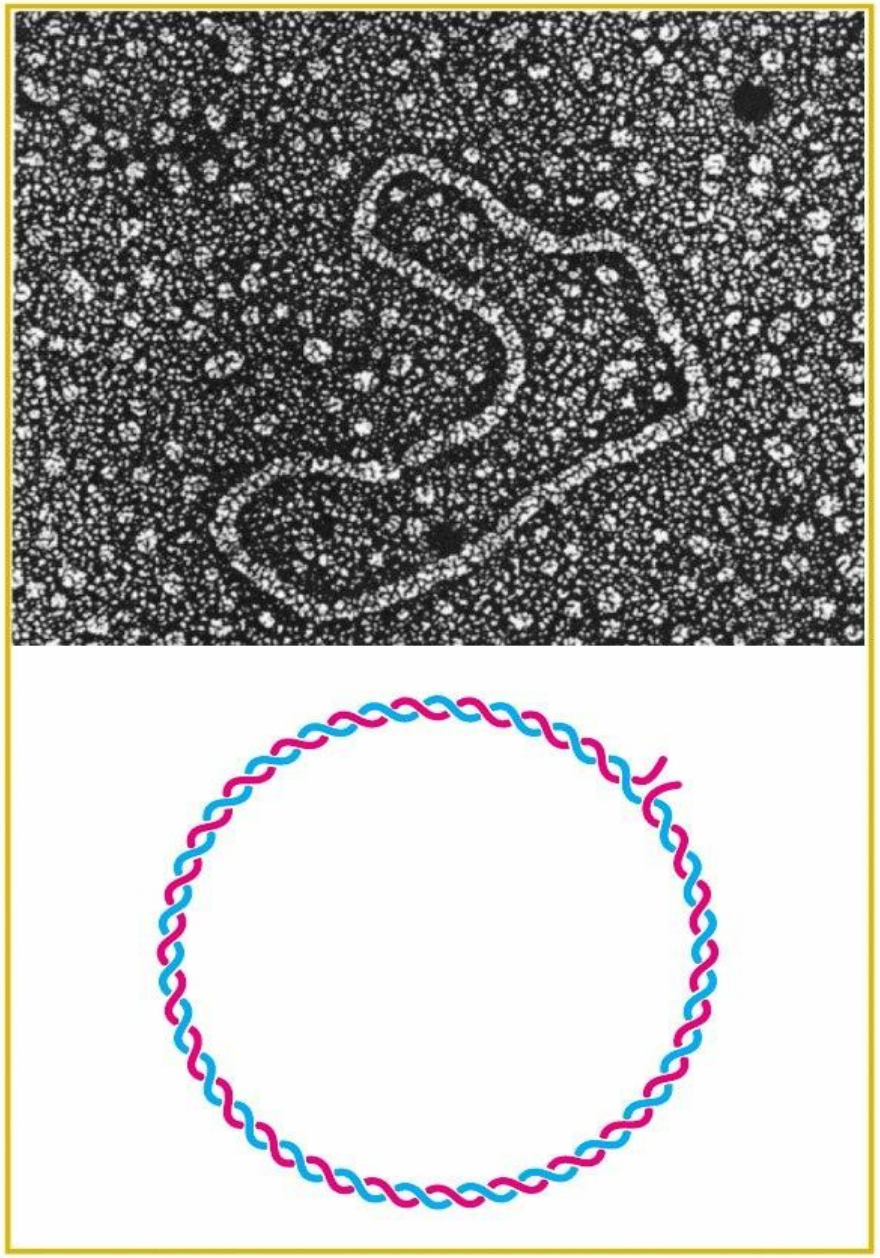
FIGURE 5.7 Structure of a DNA chain. The chain has a 5′ end, which is usually attached to a phosphate, and a 3′ end, which is usually a free hydroxyl group.

- DNA is relatively stable molecule
- Several types of noncovalent bonding contribute to this stability:
  - **Hydrophobic interacions** beween he sacked base pairs in the double helix – an imporant (but poorly understood) role in stabilizing DNA: **p-p** interactions
  - **Sugar-phosphate backbone** is hydrophilic and therefore DNA s external surface is solvated with water
  - **Hydrogen bonding** between complementary bases promotes stability as well as providing a mechanism for accurate pairings between the bases

(a) Supercoiled



(b) Relaxed circle



„Quarternary structure“ of DNA – nucleosome, chromatin

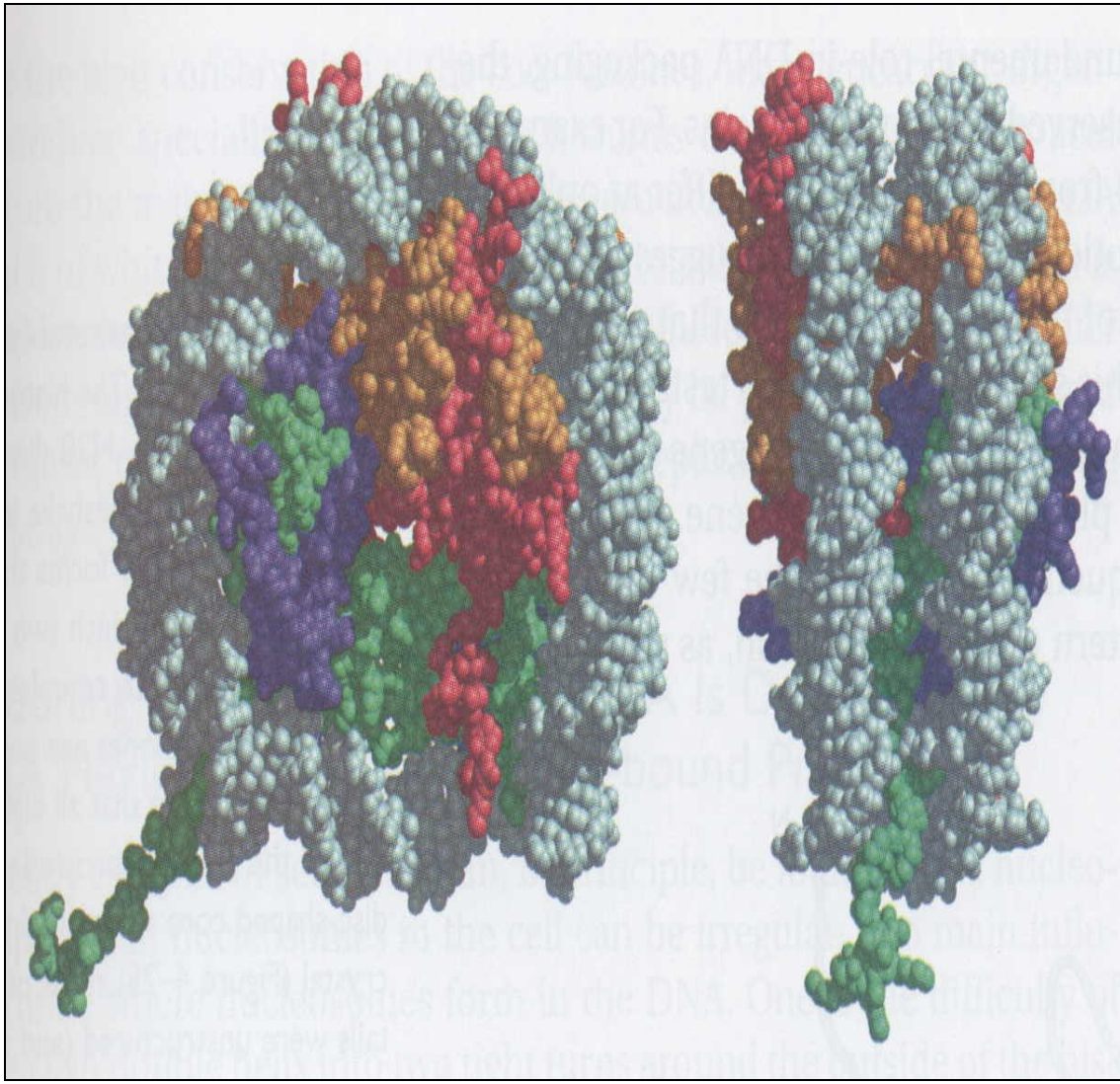
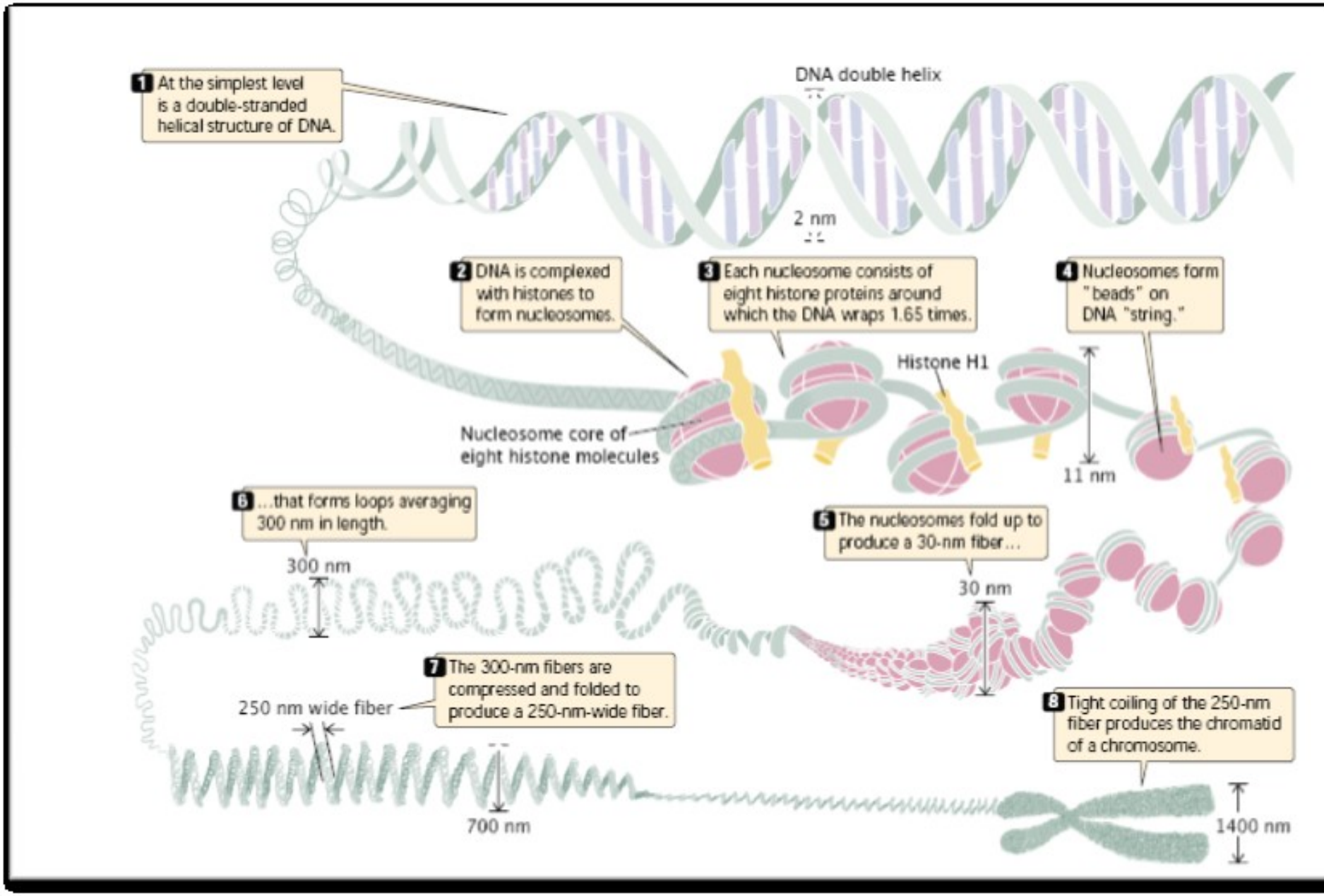
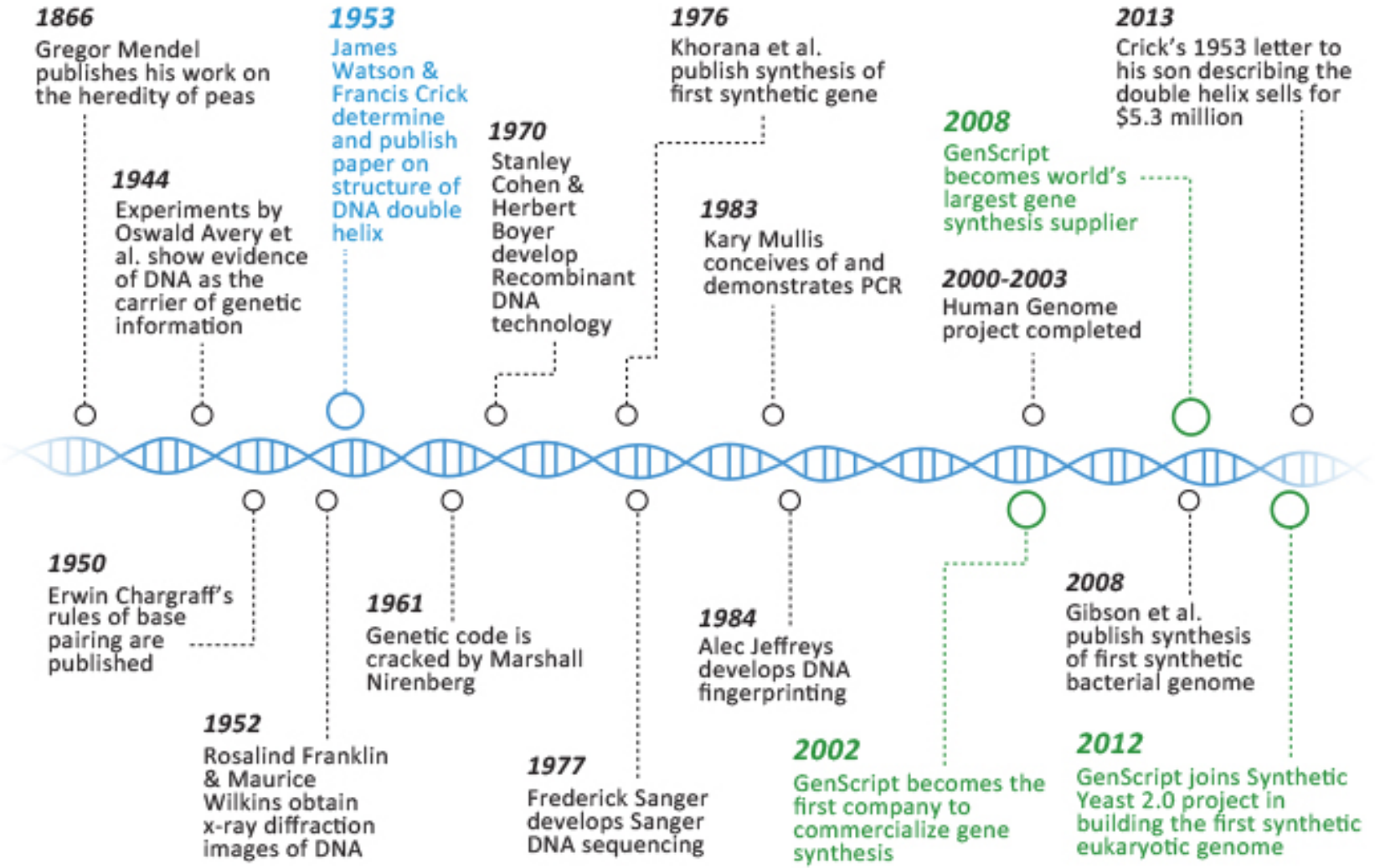


Figure 4–25 The structure of a nucleosome core particle, as determined by x-ray diffraction analyses of crystals. Each histone is colored according to the scheme of Figure 4–24, with the DNA double helix in *light gray*. (Reprinted by permission from K. Luger et al., *Nature* 389:251–260, 1997. © Macmillan Magazines Ltd.)

Packaiging of DNA



Analysis of DNA



In situ hybridization

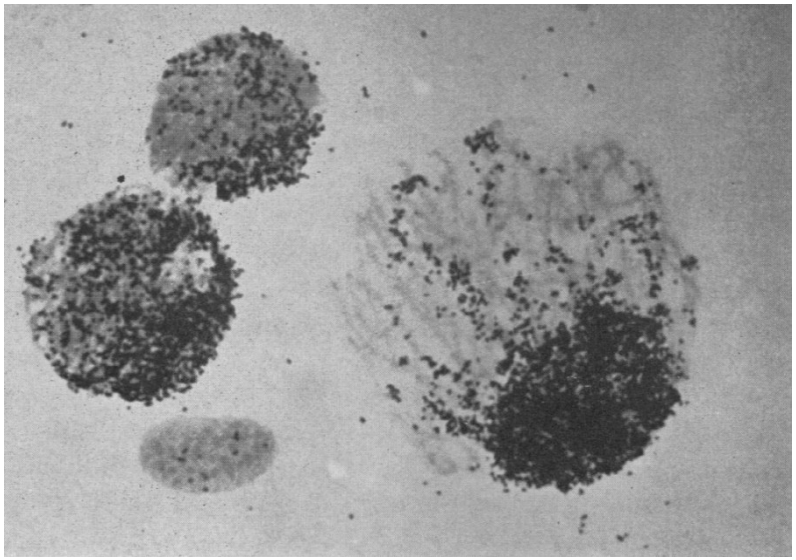
*FORMATION AND DETECTION OF RNA-DNA HYBRID MOLECULES IN CYTOLOGICAL PREPARATIONS**

By Joseph G. Gall and Mary Lou Pardue

Kline Biology Tower, Yale University

Communicated by Norman H. Giles, March 27, 1969

Abstract.—A technique is described for forming molecular hybrids between RNA in solution and the DNA of intact cytological preparations. Cells in a conventional tissue squash are immobilized under a thin layer of agar. Next they are treated with alkali to denature the DNA and then incubated with tritium-labeled RNA. The hybrids are detected by autoradiography. The technique is illustrated by the hybridization of ribosomal RNA to the amplified ribosomal genes in oocytes of the toad *Xenopus*. A low level of gene amplification was also detected in premeiotic nuclei (oogonia).



[8] Bergman, J. W., and Maan, S. S., *Fourth int. Wheat Genet. Symp.*, Missouri, 329–335 (1973).
[9] Davis, J., *Ann. N.Y. Acad. Sci.*, 121, 404–427 (1964).
[10] Davidson, E. H., and Britten, R. J., *Q. Rev. Biol.*, 48, 565–606 (1973).
[11] Schwartz, D., *Genetics*, 67, 411–425 (1970).

High resolution detection of DNA–RNA hybrids *in situ* by indirect immunofluorescence

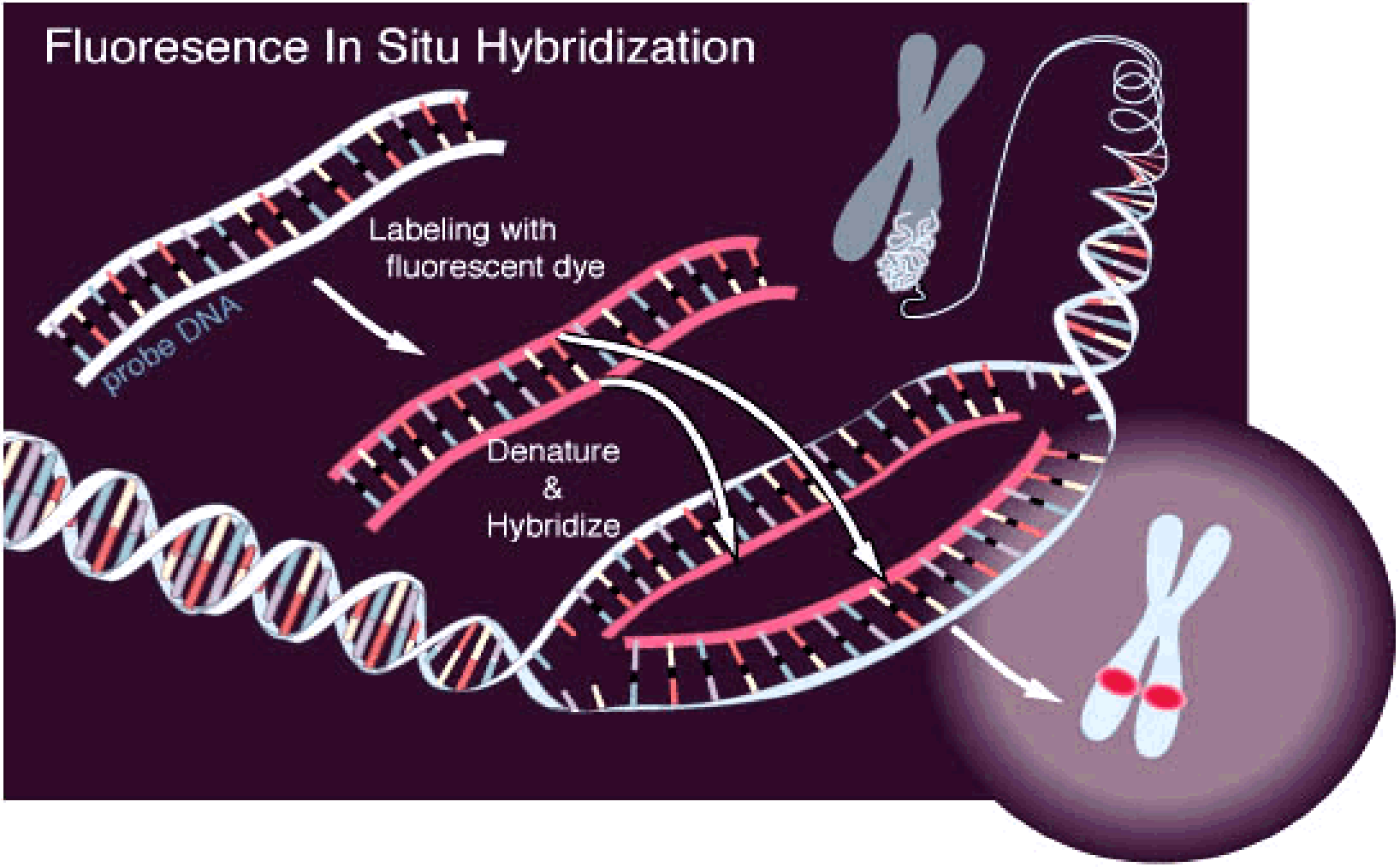
We describe here a new method for the detection of RNA–DNA hybrids in cytological preparations with which we have revealed the locations of hybrid molecules on polytene chromosomes. The critical reagent is an antiserum raised in rabbits against poly(rA)·poly(dT) complexed with methylated bovine serum albumin, originally described by Stollar[1]. The specificity and resolving power of the indirect immunofluorescence procedure are demonstrated using *in situ* hybridisation of 5S rRNA (ribosomal RNA) to polytene chromosomes of *Drosophila melanogaster* as a model system. The method has significant advantages over the autoradiographic procedures[2–5] used so far.

The procedure for visualising the *in situ* hybrids follows Alfageme *et al.*[6] It consists of exposing the cytological preparation to the rabbit anti-hybrid antiserum, then to anti-rabbit IgG prepared in goat and tagged with rhodamine, followed by examination in a fluorescence microscope (see legend to Fig. 1). Our test objects were polytene chromosomes of *Drosophila melanogaster* (giant phenotype) to which 5S rRNA had been serum, the hybrid was the only reactive polynucleotide class, even in assays with undiluted serum.

There is a variable amount of fluorescence in cytoplasmic components, the origin of which is not yet known. Experiments are projected to attempt to block it while leaving the activity against hybrid nucleic acids intact. Occasional pale fluorescence observed in nucleoli is attributed to contamination of the 5S rRNA probe with fragments of 18S and 28S nucleolar rRNA,

- Rudkin GT, Stollar BD. *High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence*. Nature. 1977 Feb 3;265(5593):472-3.

70´s – fluorescenční *in situ* hybridizace

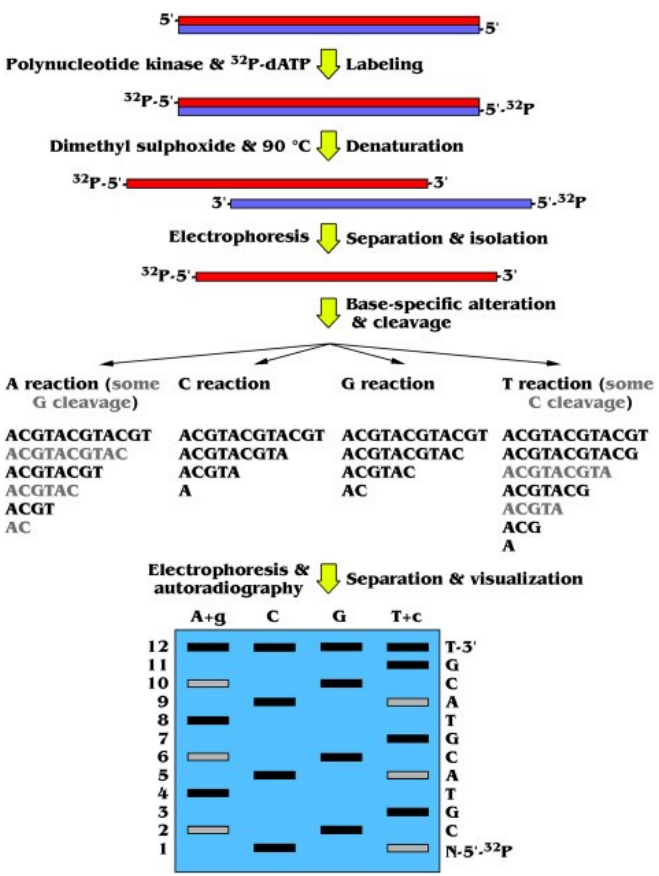


DNA sequenicng

- proces of determination of the order of nucleotides in DNA
- Idea – to vizualize the bases of DNA in ammner that they can be sorted and identified
- 70´s - 2 ways - „chemical“ and radioactive

Chemical sequencing (1976)

A. Maxam, P. Gilbert

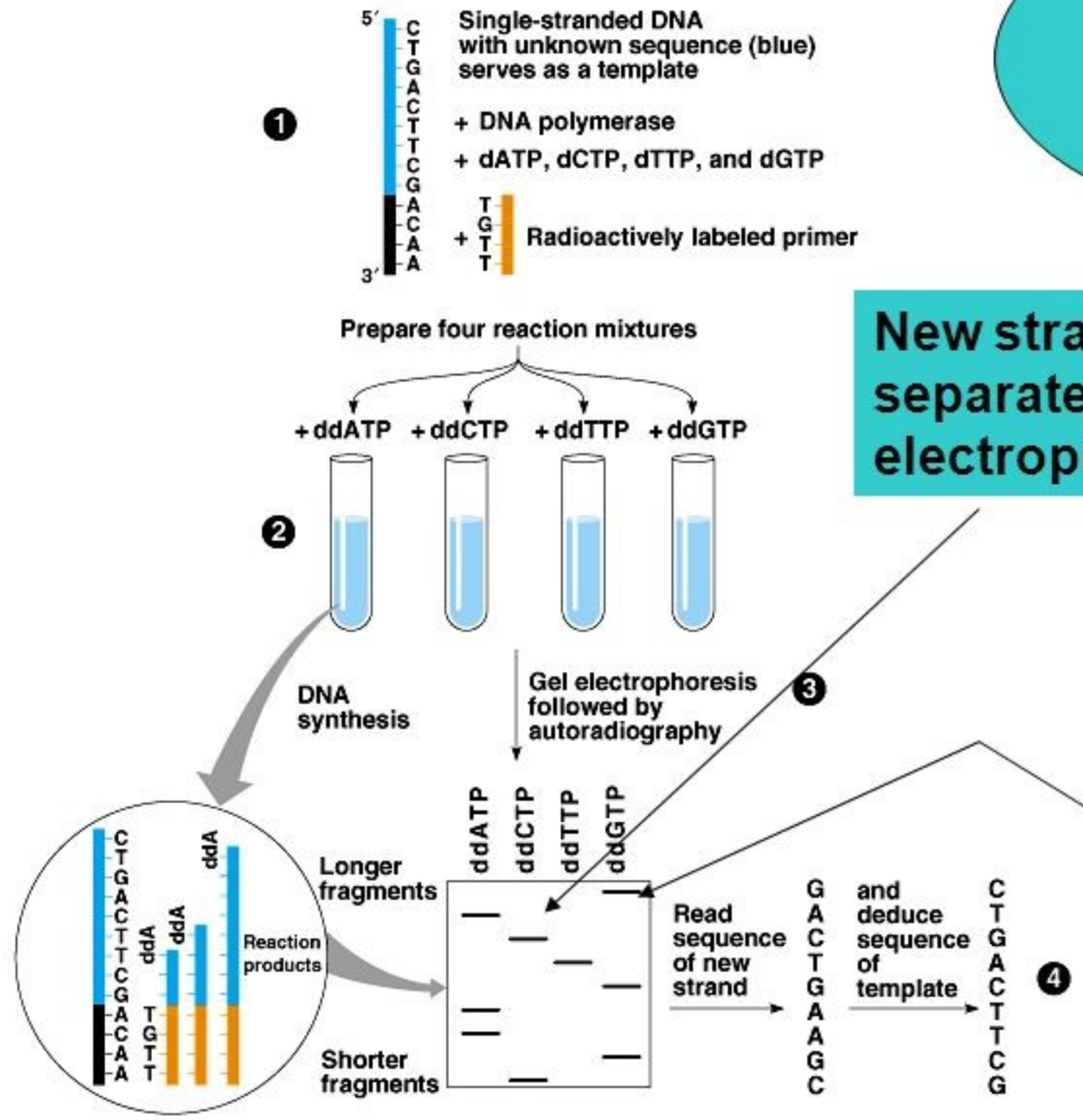


- based on chemical modification of DNA and subsequent cleavage at specific bases
- the method requires radioactive labelling at one end
- cchemical treatment generates breaks at a small proportions of one or two of the four nucleotide based in each of four reactions (G,A+G, C, C+T)
- series of labelled fragments is generated, from the radiolabeled end to the first ‘cut’ site in a molecule
- the fragments in the four reactions are arranged side by side in gel electrophoresis for size separation
- the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabeled DNA fragment, from which the sequence may be inferred.

Sanger (dideoxy termination) sequencing

- 1977 – F. Sanger and colleagues, Cambridge
- based on the random incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication
- Enzymatic (DNA polymerase), cost effective, less handwork, AUTOMATIZATION!

SANGER METHOD



New strands separated by electrophoresis

Sequence can be read from bands on autoradiograph and original template sequence deduced. Longest fragment ends with a ddG, so G must be the last base in the sequence

https://www.onlinebiologynotes.com/sangers-method-gene-sequencing/

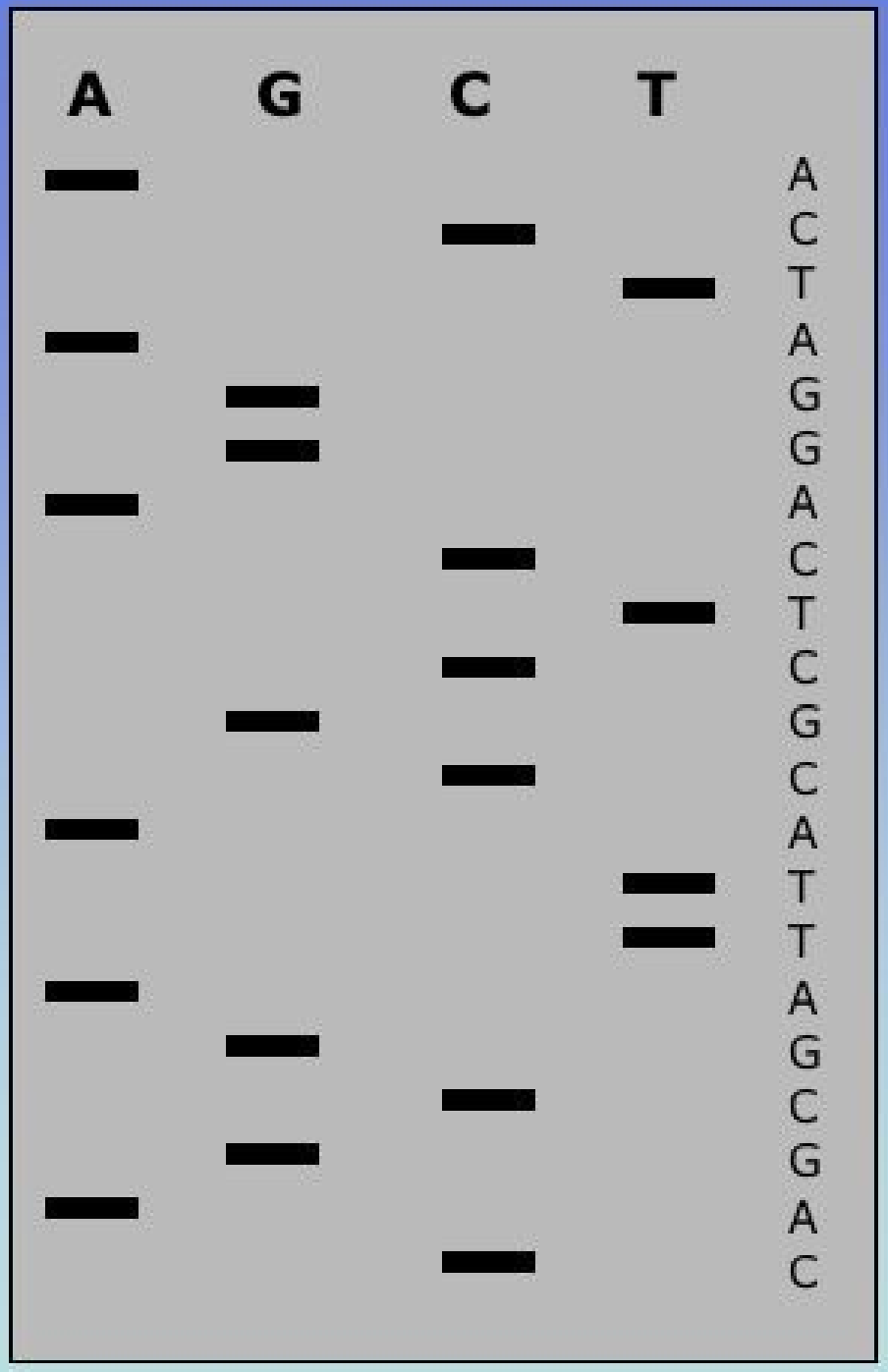
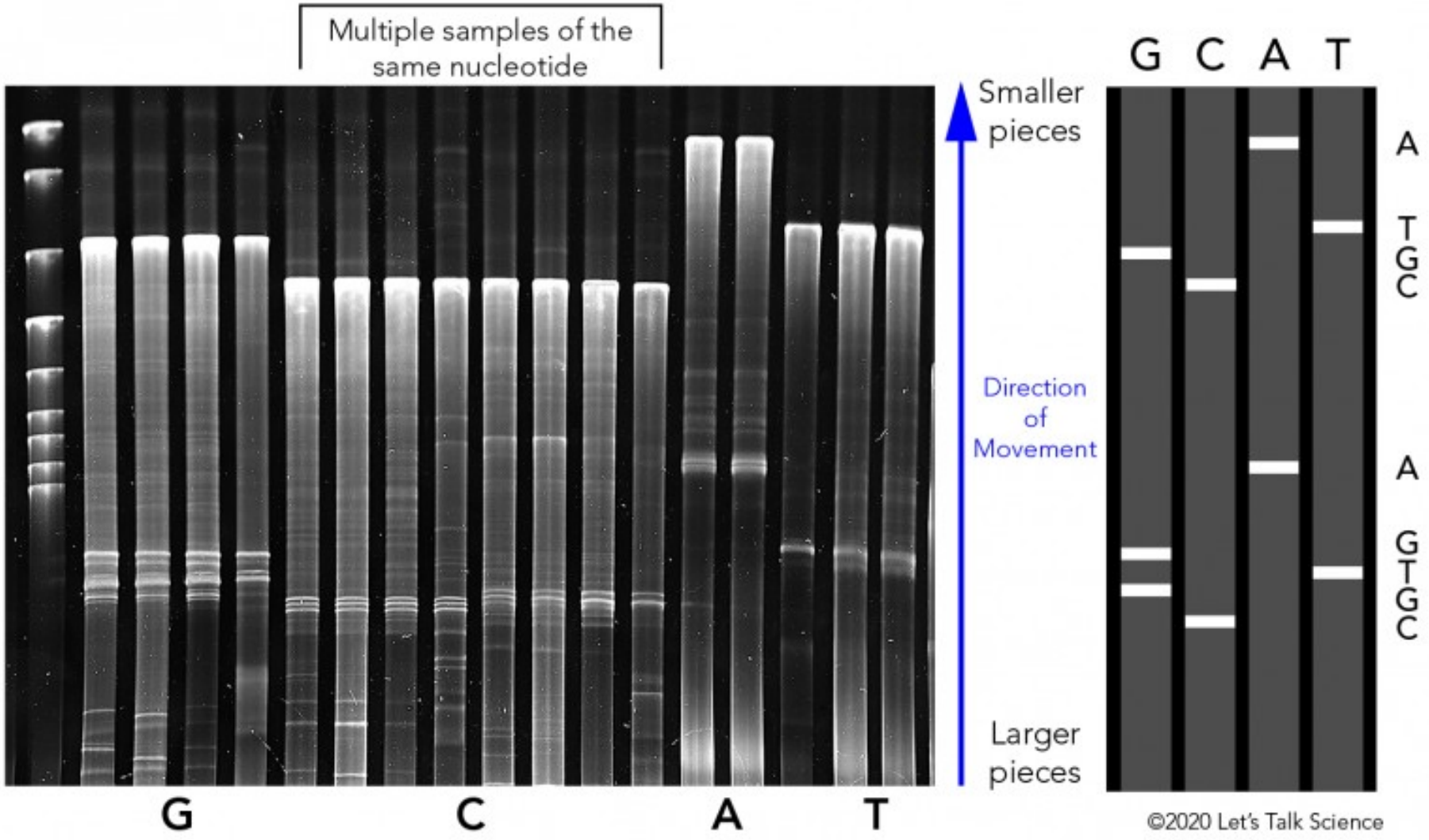


Figure 1: Manual Sequencing Using Radiolabeled ddNTPs

Manual sequencing using the Sanger method originally used radioactively labeled ddNTPs. Because there was no method to detect the difference between the A,G,C, and T bases, the reaction for each bases was done separately and loaded into separate lanes on a polyacrylamide gel as shown.

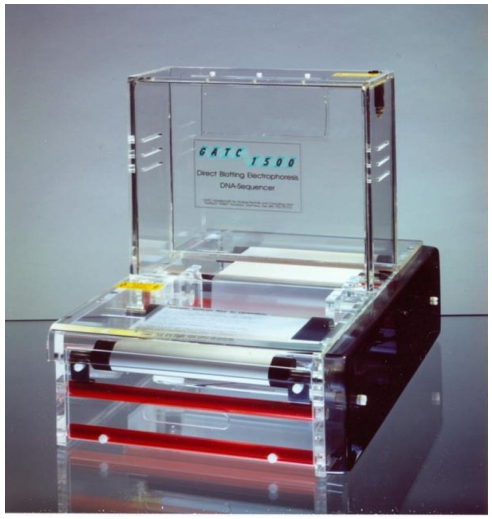

Multiple samples of the same nucleotide
Smaller pieces
Direction of Movement
Larger pieces
G C A T
A TGC A GTGC
G
C
A
T
©2020 Let's Talk Science

MAXAM GILBERT SEQUENCING

VERSUS

SANGER SEQUENCING

MAXAM GILBERT SEQUENCING	SANGER SEQUENCING
The method of DNA sequencing based on nucleotide specific partial chemical modification and subsequent DNA cleavage	The process of selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication
Developed by Allan Maxam and Walter Gilbert in 1977–1980	Developed by Frederick Sanger and colleagues in 1977
The chemical method of sequencing	The chain-termination method
Uses a large amount of hazardous chemical including radioactive material and hydrazine	Uses less hazardous chemicals
Highly sensitive and highly specific	Less sensitive and less specific

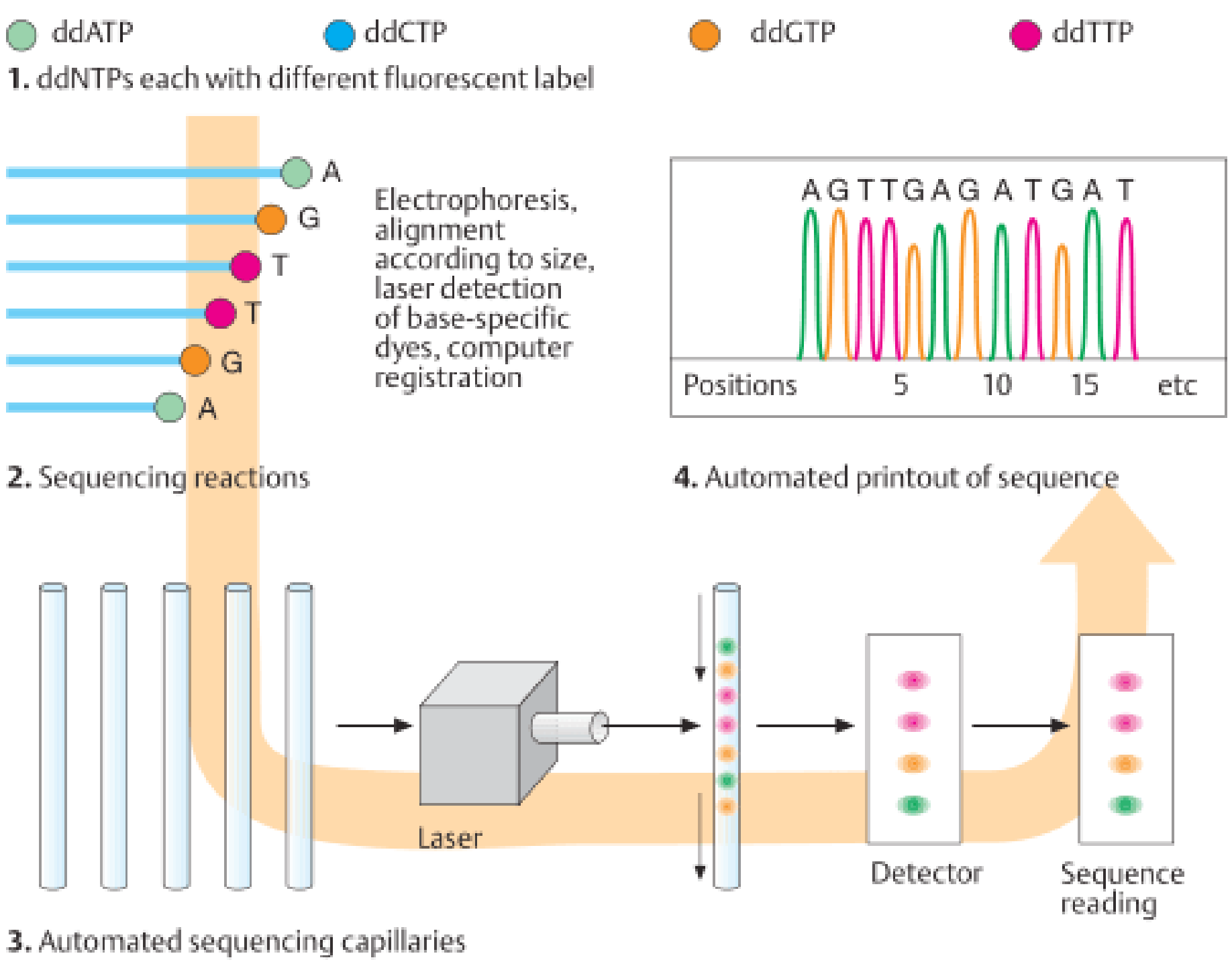


The direct blotting electrophoresis system GATC1500 (1984)

Applied Biosystems 370A Prototype Automated DNA Gene Sequencer (Hood, Hunkapiller 1987)

ABI 3730*xl* DNA Analyzer (384 well plate, 2010)

Applied Biosystems SeqStudio Genetic Analyzer


ddATP
ddCTP
ddGTP
ddTTP
1. ddNTPs each with different fluorescent label
A
G
T
T
G
A
Electrophoresis, alignment according to size, laser detection of base-specific dyes, computer registration
A G T T G A G A T G A T
Positions
5
10
15
etc
2. Sequencing reactions
4. Automated printout of sequence
Laser
Detector
Sequence reading
3. Automated sequencing capillaries

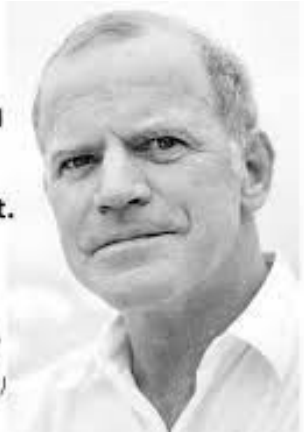
PCR – polymerase chain reaction

- Method widely used to rapidly **make millions to billions** of copies (complete or partial) of a specific DNA sample
- Discovery of **DNA polymerase** (1957 – Kornberg, mechanism of DNA replication
- Development of **synthetic DNA** oligonucleotides (early 60 s Khorana studying a genetic code)
- **Thermostable DNA** polymerase from *Thermus aquaticus* was isolated (1969 –Brock)
- PCR involves **using** short synthetic DNA fragments called **primers** to select a **segment of the genome** to be amplified, and then **multiple rounds** of DNA synthesis to **amplify** that segment
- Developed in CETUS Corparation by Kary B. Mullis in 1983 (Nobel prize – 1993)

"What if I had not taken LSD ever; would I have still invented PCR?

I don't know. I doubt it. I seriously doubt it."

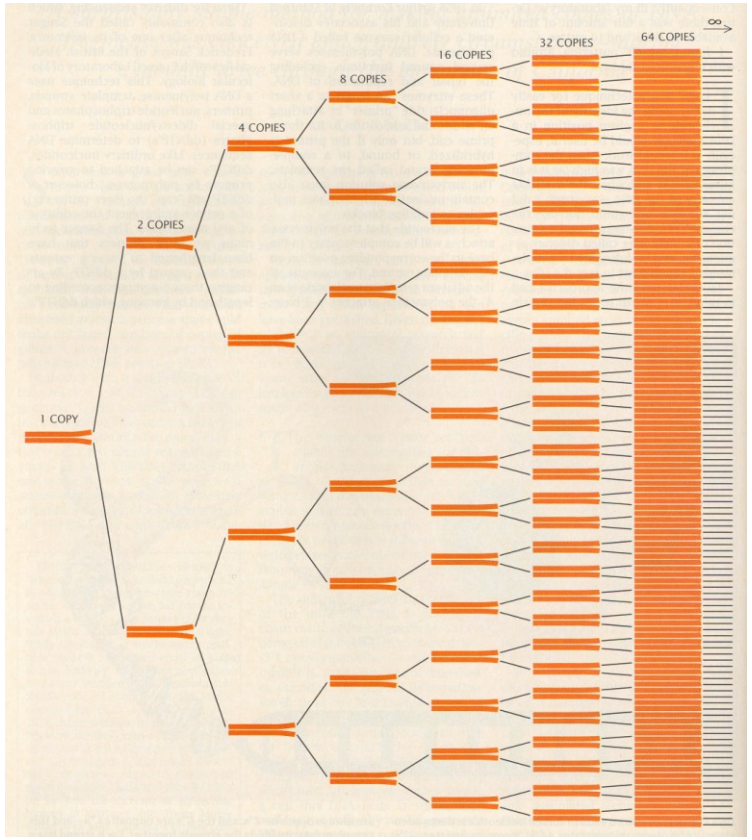
- *Kary Mullis*
(Winner of the 1993 Nobel Prize in Chemistry for his discovery of the polymerase chain reaction, or PCR)

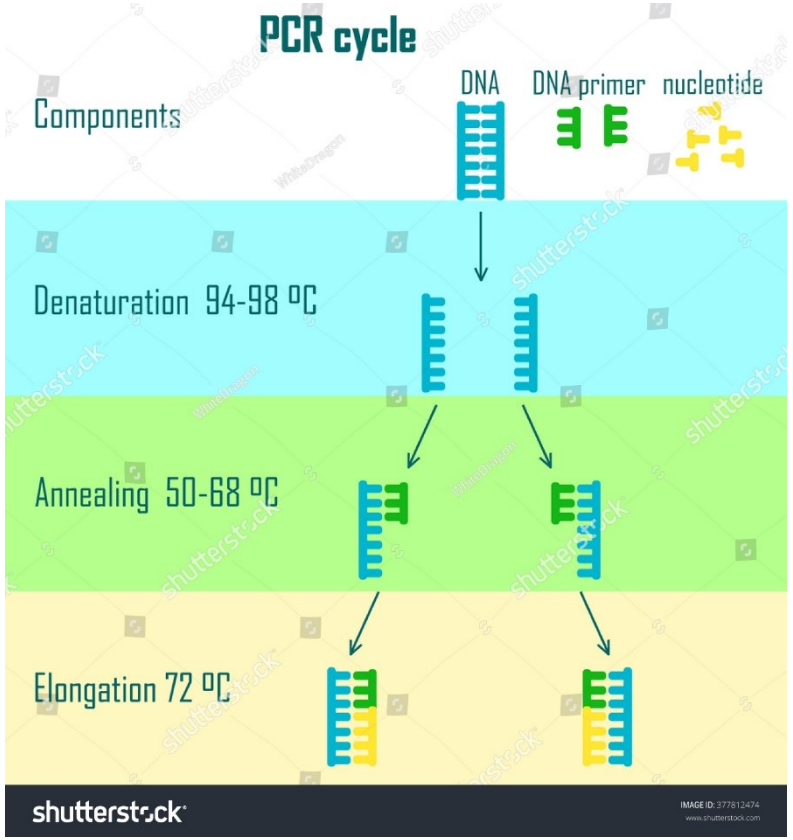


The Unusual Origin of the Polymerase Chain Reaction

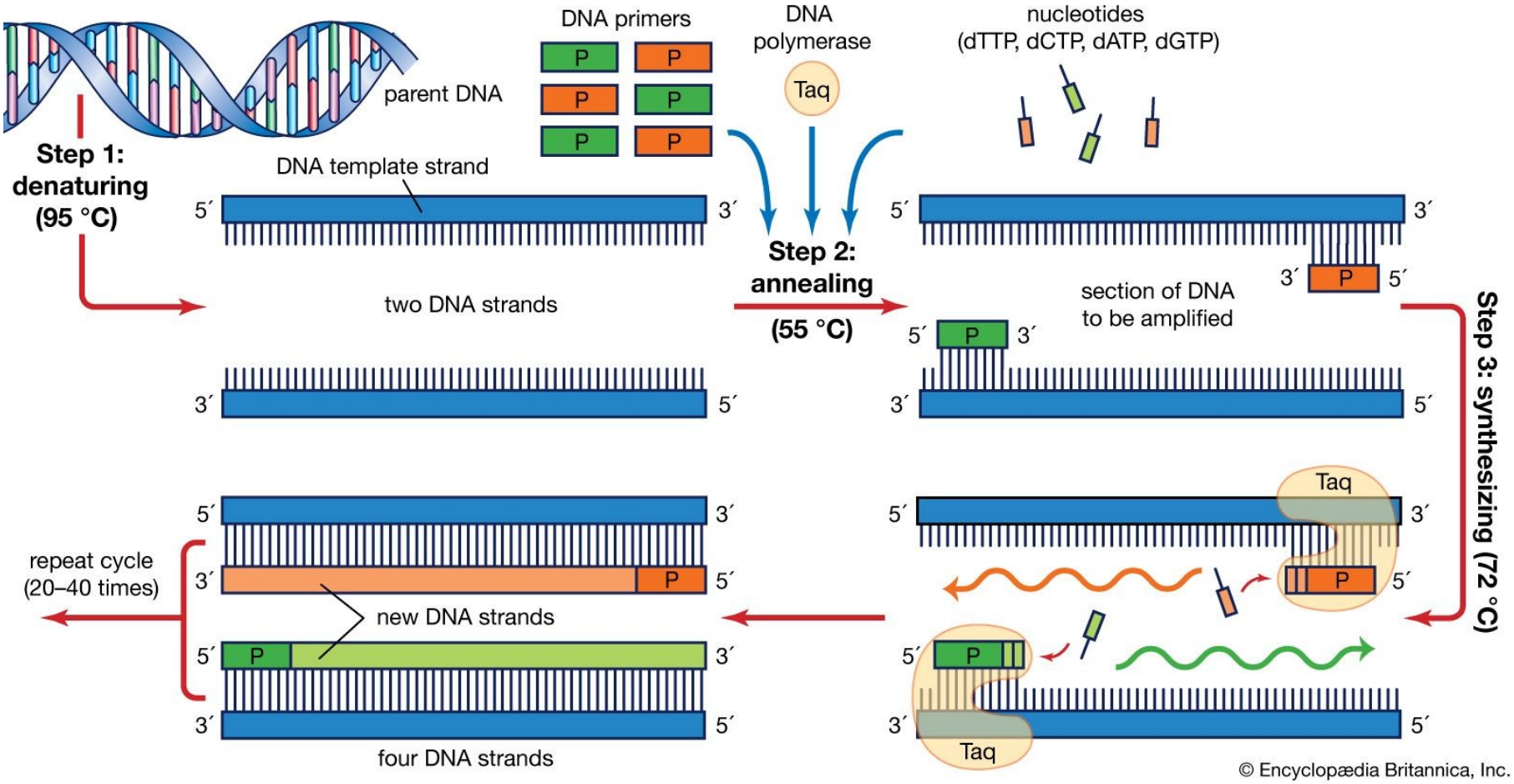
A surprisingly simple method for making unlimited copies of DNA fragments was conceived under unlikely circumstances—during a moonlit drive through the mountains of California

by Kary B. Mullis





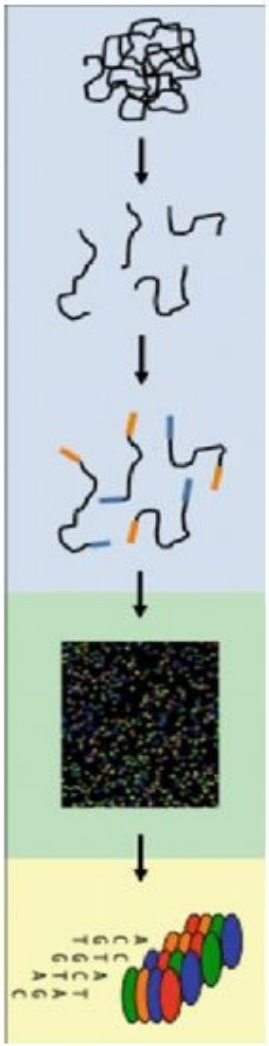
PCR – polymerase chain reaction

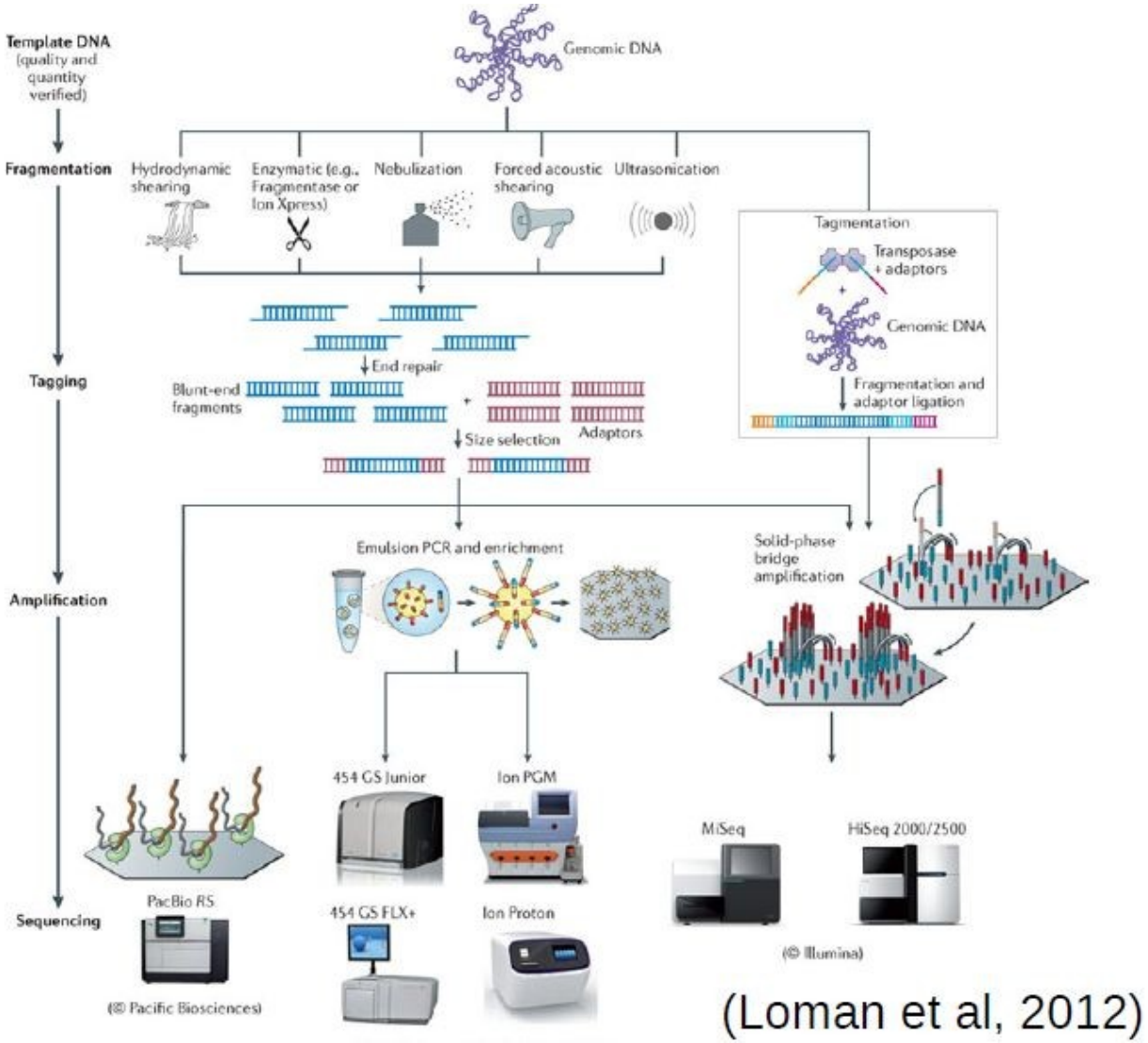


Next generation sequencing (NGS)

- Massive parallel (shotgun) sequencing
- Used for – analysis of large number of genes in one experiment
- Effective cost per base ratio
- Utilization of computer proccessing of data and bioinformatics

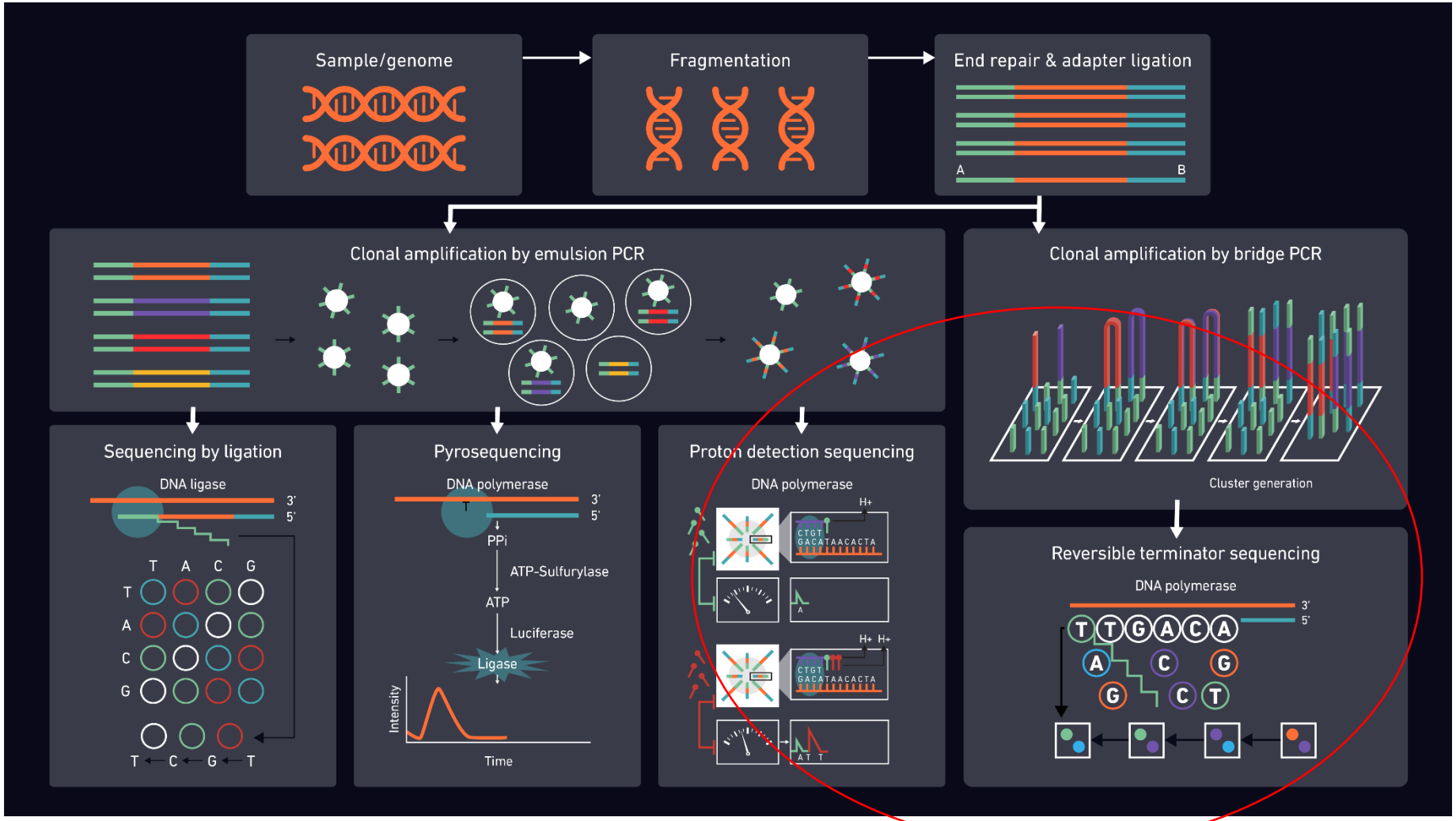
NGS workflow





(Loman et al, 2012)

NGS technologies



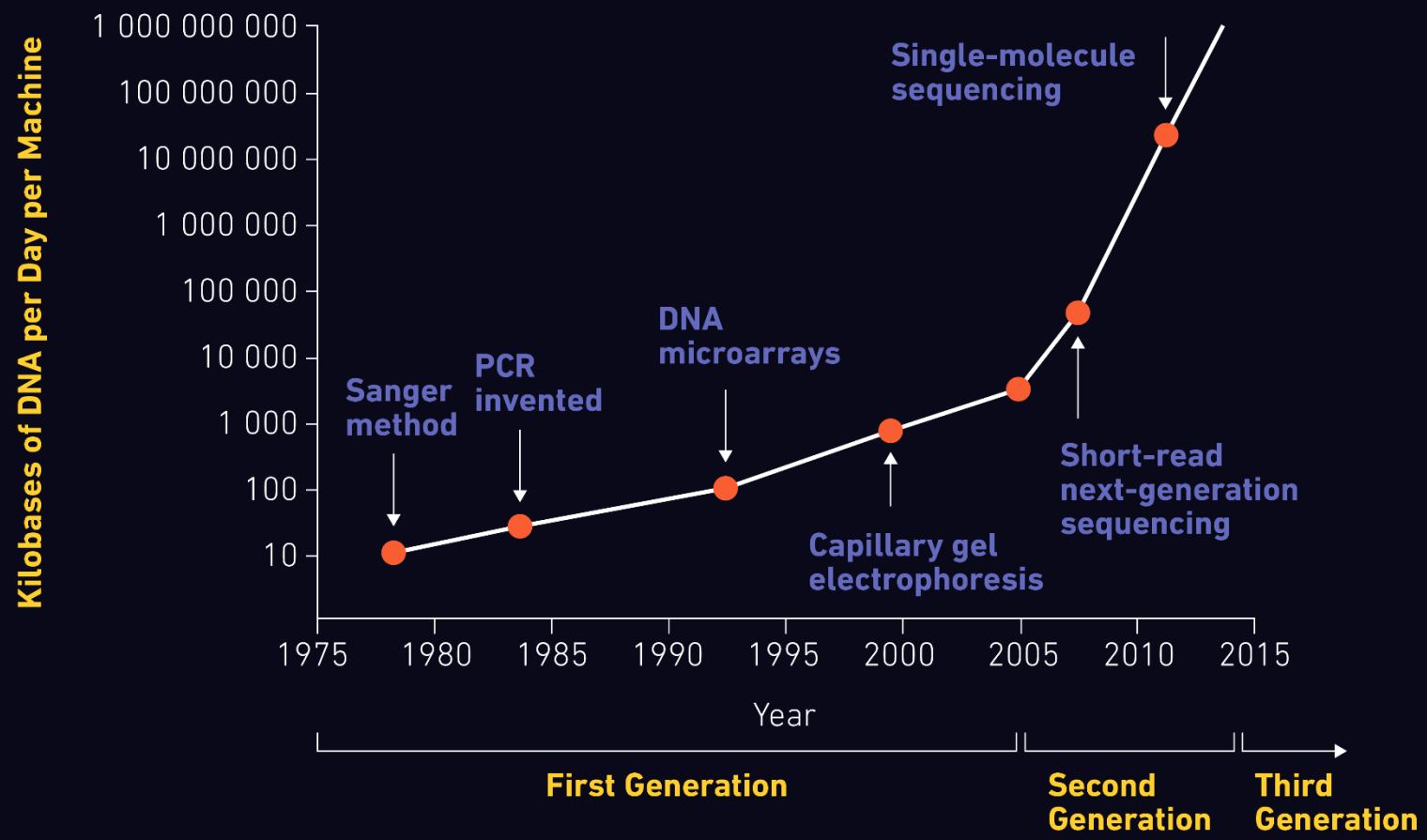
ABI SOLiD (Life)

454 Life Science (Roche Inc)

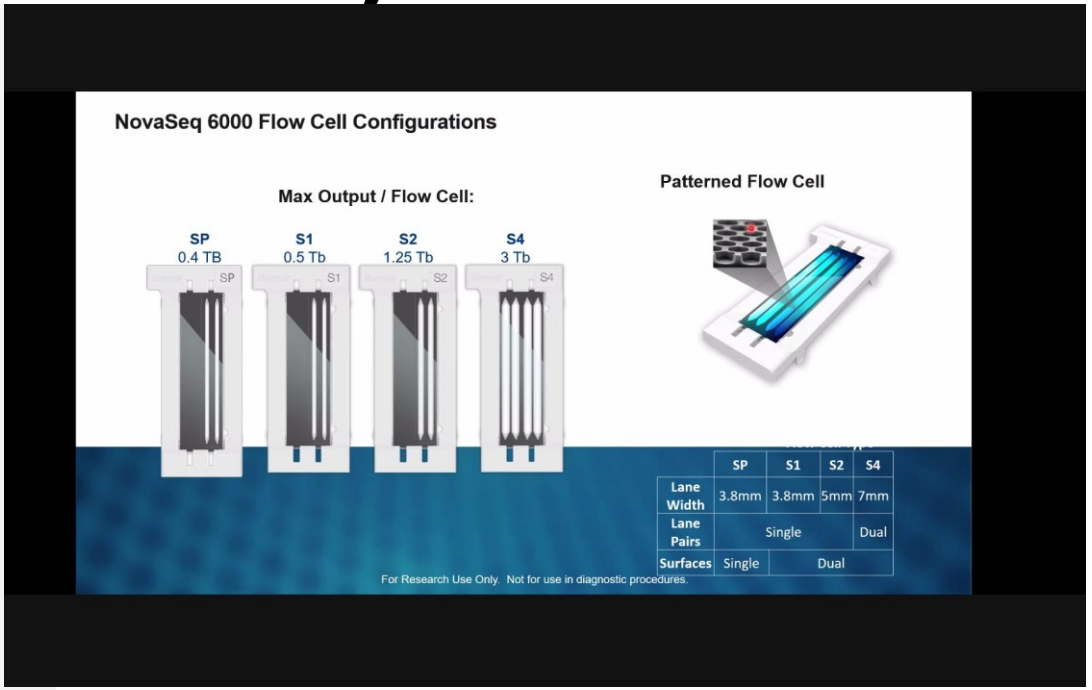
Semiconductor sequencing (Life)

Sequencing by synthesis (SBS) Illumina Inc.

NGS technologies – sequencing capaicity



Nowadays…**NovaSeq 6000 (Illumina)**



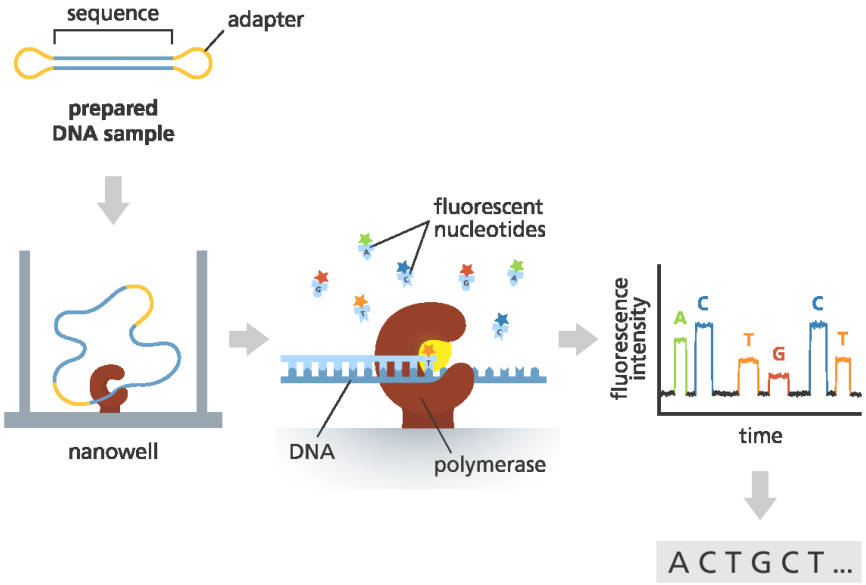
the most powerful high-throughput Illumina sequencing system to date

- **48 human genomes in 2 days**
- **Very variable and robust (2 flowcell port = cost effective)**
- **WGS, WES, panels, RNA**

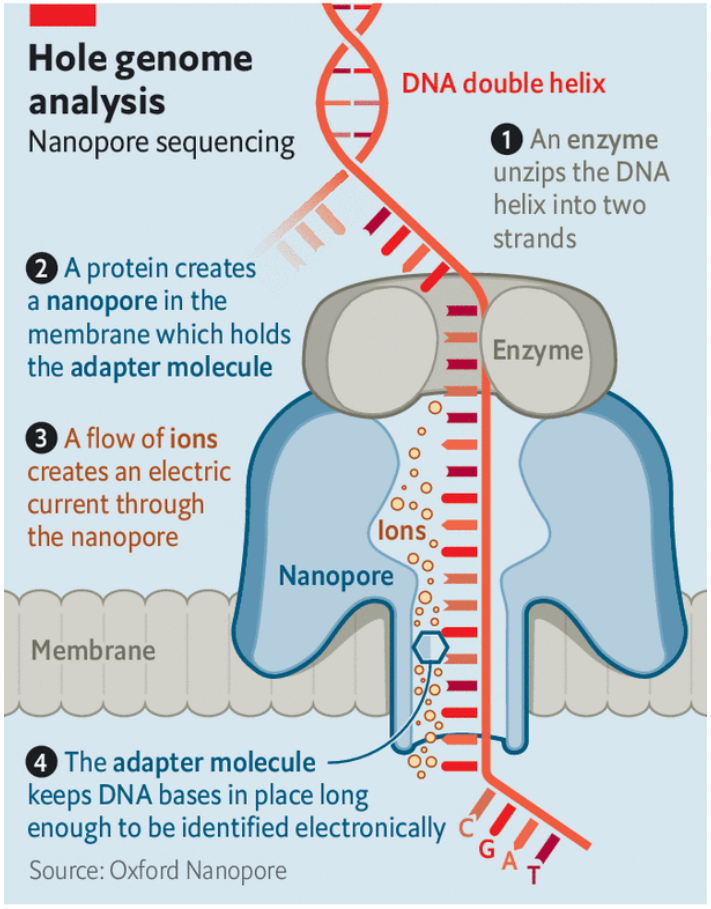
3rd generation of NGS

"single molecule NGS"

SMRT (Single Molecule, Real-Time Sequencing)

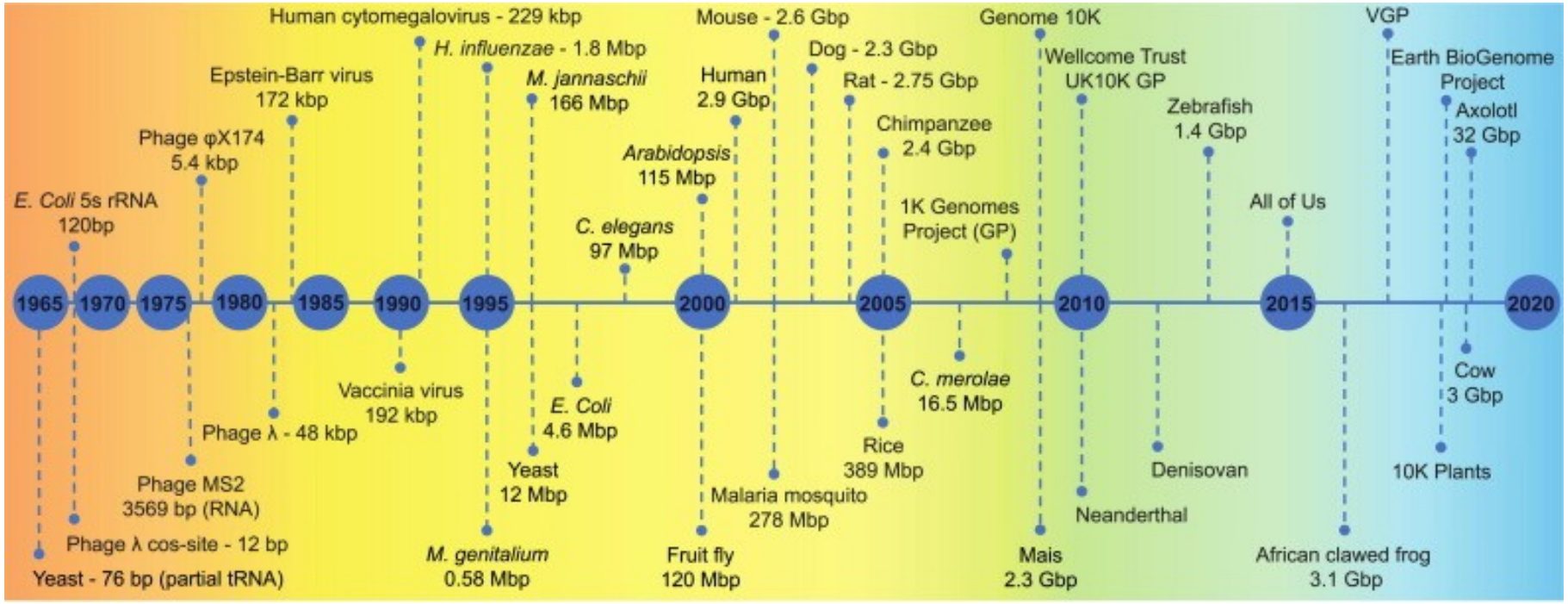


Oxford Nanopore



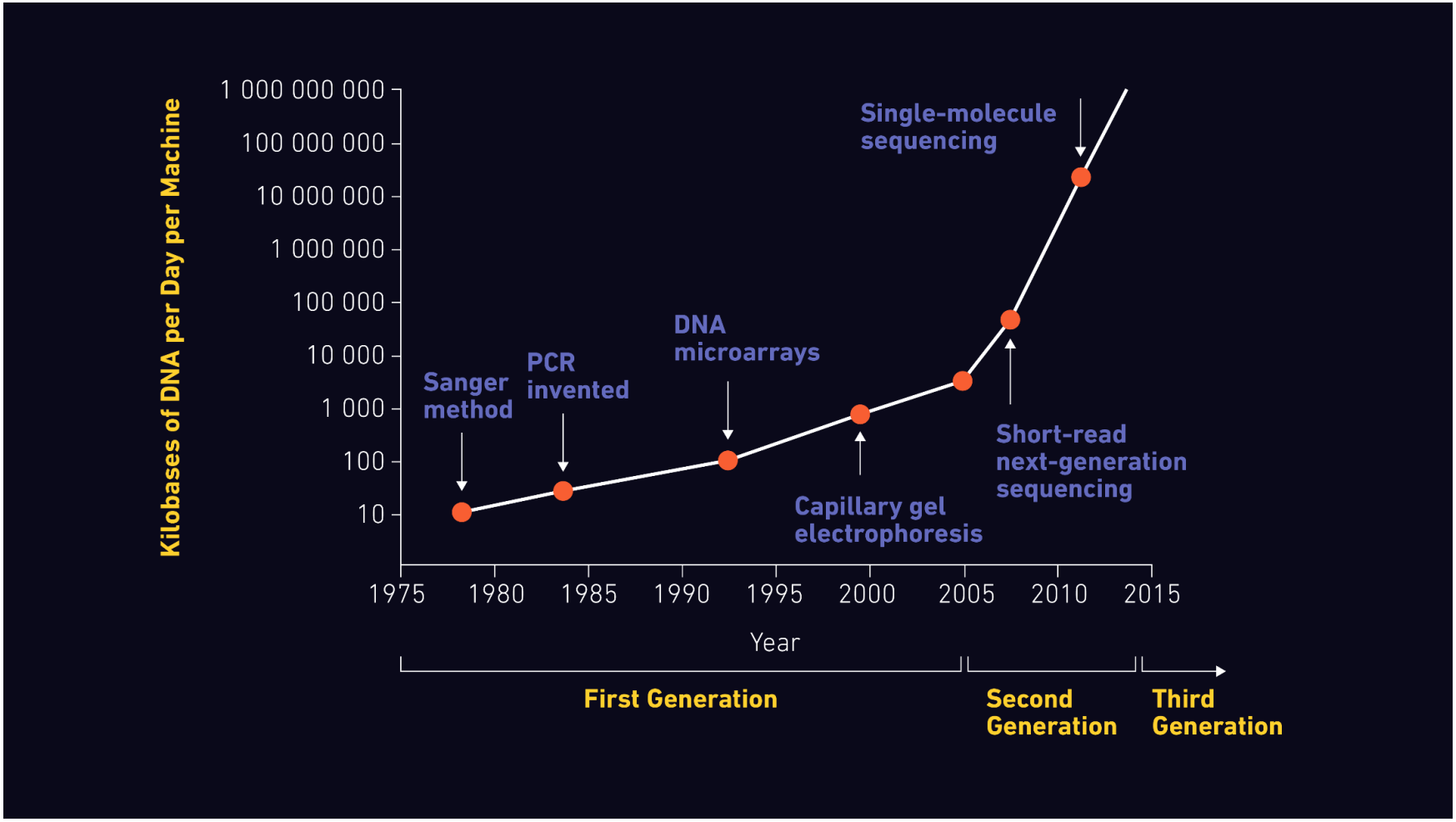
The Economist

Historiy of whole genome sequencing

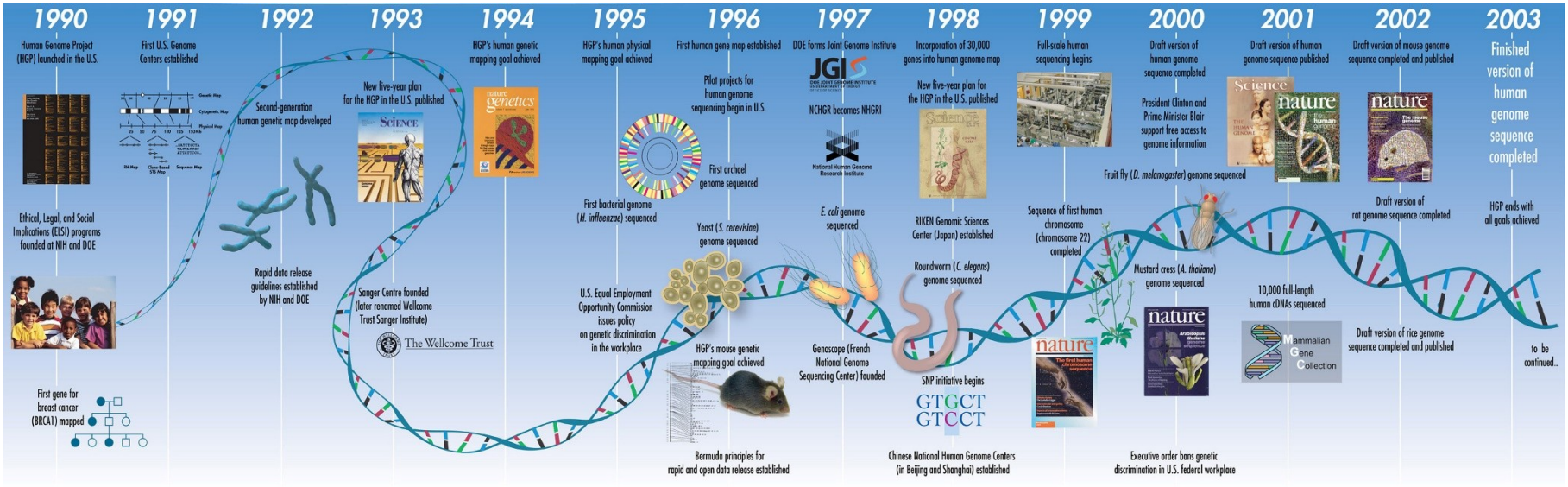


Gianni *et al*., Computational and Structural Biotechnology Journal, 2019

Historiy of whole genome sequencing



Human genome project (HGP)
1990-2003



Aim and Objective

Human Genome Project

To Sequence the whole genome with 3billion bps

Creating a physical Map of the Human genome

To identify all the Disease causing genes

To understand the Function of Genes

To make the informations Available for all Researchers

To develop tools For processing and Analysing data

Human genome project - origin

E. Southern

G. Church

Seeds of the Human Genome Project

Can we determine the germline mutation rate in humans?

Participants at the Alta Meeting, December 1984

David Botstein	Mortimer Mendelsohn
Elbert Branscomb	John Mulvihill
Charles R. Cantor	Richard Myers
C. Thomas Caskey	James V. Neel
George Church	Maynard Olson
John D. Delahanty	David A. Smith
Charles Edington	Edwin Southern
Raymond Gesteland	Sherman Weissman
Michael Gough	Raymond L. White
Leonard Lerman	

The Alta Summit, December 1984

The following article by Robert Cook-Deegan appeared in *Genomics*, Volume 5, pp 661-663 (October, 1989) by Academic Press, Inc.(1) . It describes the genesis of the U.S. Human Genome Project.

The Alta Summit, December 1984

Alta is a ski area nestled among the Wasatch Mountains (note: original text said Saguache Mountains) in Utah, a winding 40-minute drive southeast from Salt Lake City. From December 9 to 13, 1984, visitors were isolated by repeated blizzards. The slopes were covered most mornings with Utah's renowned fine light powder, which beckoned skiers to cut its virgin surface.

For those 5 days, Alta was also a capital of human genetics. Many historical threads in the fabric that later became the

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY

Richard M. Myers

HGP - backgronud

1984 – 1986 The U.S. Department of Energy (**DOE)** and the International Commission for Protection against Environmental Mutagens and Carcinogens **(ICPEMC**) initiatee the early meetings assess the feasibility of a Human Genome Project

1988

- The National Institutes of Health **(NIH)** assembles scientists, administrators and science policy experts to plan for a possible Human Genome Project
- Two published reports recommend creating an effort to sequence the human genome (National Research Council; he U.S. Congress Office of Technology Assessment)

1989

The National Center for Human Genome Research **(NCHGR**) is established to carry out the United States Human Genome Project. The center's first director is James D. Watson

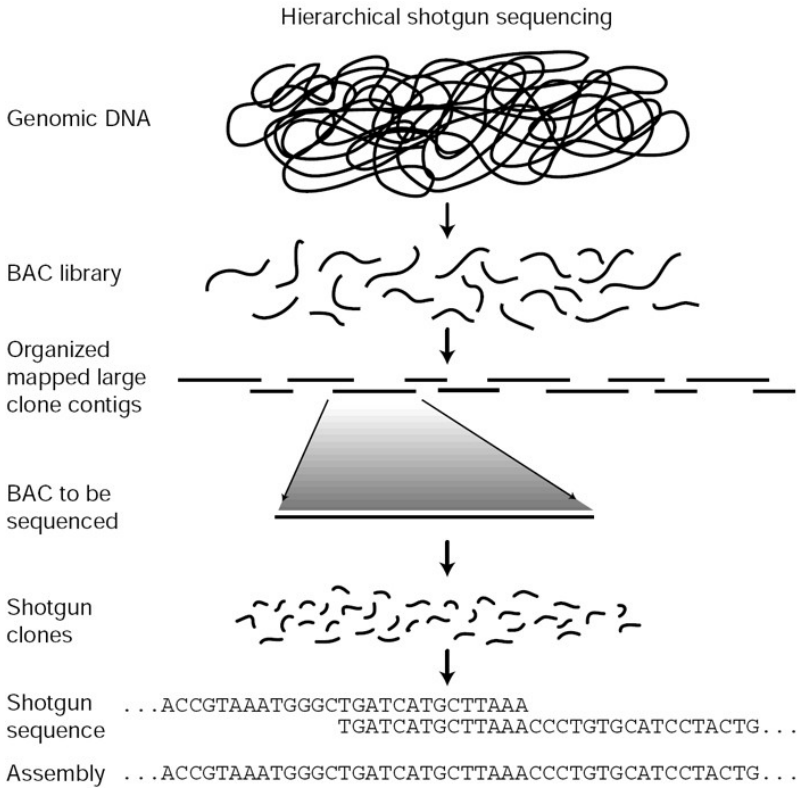
1990 The Human Genome Project begins with an initial five-year plan

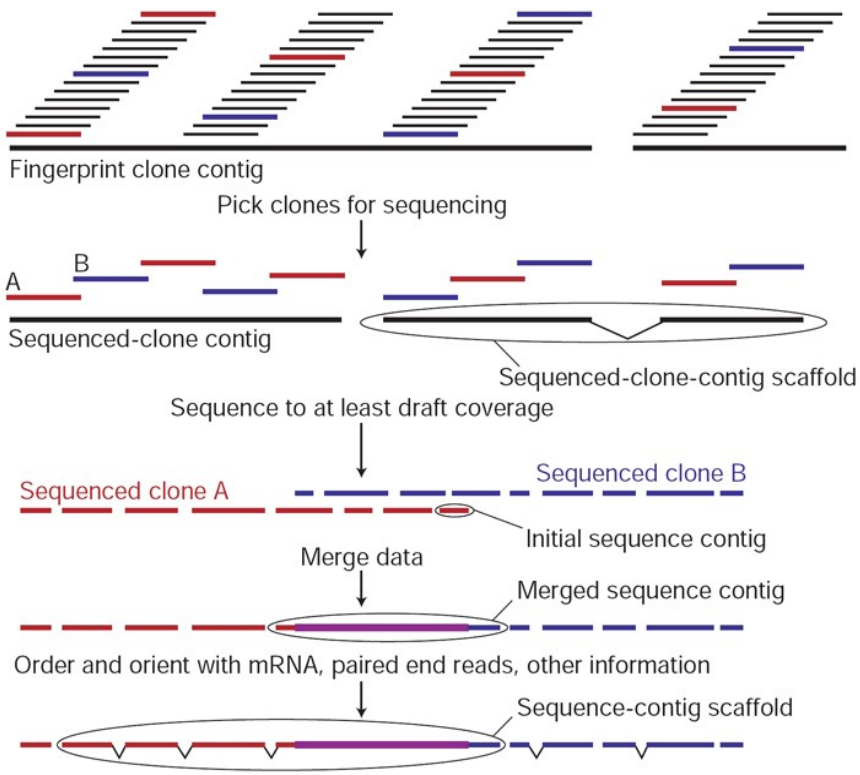
- NIH allocates the first funds to research grants aimed at developing the scientific approaches, technologies, and resources needed to map and sequence the human genome

The National Center for Human Genome Research (NCHGR)

1. The Whitehead Institute/MIT Center for Genome Research, Cambridge, Mass., U.S.
2. **The Wellcome Trust Sanger Institut**e, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, U. K.
3. Washington University School of Medicine Genome Sequencing Center, St. Louis, Mo., U.S.
4. United States DOE Joint Genome Institute, Walnut Creek, Calif., U.S.
5. Baylor College of Medicine Human Genome Sequencing Center, Department of Molecular and Human Genetics, Houston, Tex., U.S.
6. **RIKEN Genomic Sciences Center, Yokohama, Japan**
7. **Genoscope and CNRS UMR-8030, Evry, France**
8. GTC Sequencing Center, Genome Therapeutics Corporation, Waltham, Mass., USA
9. **Department of Genome Analysis, Institute of Molecular Biotechnology, Jena, Germany**
10. **Beijing Genomics Institute**/Human Genome Center, Institute of Genetics, Chinese Academy of Sciences, Beijing, China
11. Multimegabase Sequencing Center, The Institute for Systems Biology, Seattle, Wash.
12. **Stanford Genome** Technology Center, Stanford, Calif., U.S.
13. Stanford Human Genome Center and Department of Genetics, Stanford University School of Medicine, Stanford, Calif., U.S.
14. University of Washington Genome Center, Seattle, Wash., U.S.
15. **Department of Molecular Biology, Keio University School of Medicine, Tokyo, Japan**
16. University of Texas Southwestern Medical Center at Dallas, Dallas, Tex., U.S.
17. University of Oklahoma's Advanced Center for Genome Technology, Dept. of Chemistry and Biochemistry, University of Oklahoma, Norman, Okla., U.S.
18. **Max Planck Institute for Molecular Genetics, Berlin, Germany**
19. Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center, Cold Spring Harbor, N.Y., U.S.
20. GBF - German Research Centre for Biotechnology, Braunschweig, Germany

NCHGR - hierarchical shotgun method



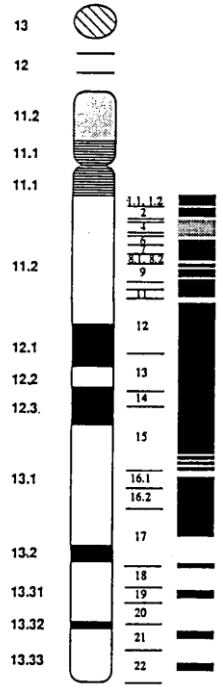


HGP – major achievements

1995 - complete physical map of human genome (=physical locations of identifiable landmarks on chromosomes backbone for genome assembly)

Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers

Callum J.Bell*, Marcia L.Budarf, Bart W.Nieuwenhuijsen[1], Barry L.Barnoski, Kenneth H.Buetow[2], Keely Campbell, Angela M.E.Colbert[3], Joelle Collins, Mark Daly[3] Philippe R.Desjardins[1], Todd DeZwaan[1], Barbara Eckman[1], Simon Foote[3,+], Kyle Hart[1], Kevin Hiester[1], Marius J.Van Het Hoog[1], Elizabeth Hopper, Alan Kaufman[3], Heather E.McDermid[4], G.Christian Overton[1], Mary Pat Reeve[3], David B.Searls[1], Lincoln Stein[3], Vinay H.Valmiki[1], Edward Watson, Sloan Williams, Rachel Winston[1], Robert L.Nussbaum[1,§], Eric S.Lander[3], Kenneth H.Fischbeck[1], Beverly S.Emanuel and Thomas J.Hudson[3]



HGP – major achievements

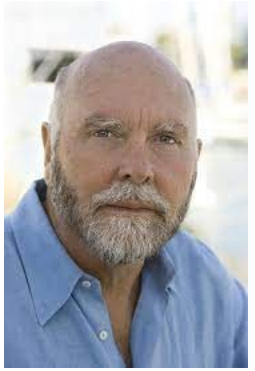
1996 - The Bermuda principles

„*all human genomic sequence information should be made freely available and placed in the public domain within 24 hours of being generated by federally funded large-scale human sequencing centers*“

1. Automatic release of sequence assemblies larger than 1 kb (preferably within 24 hours).
2. Immediate publication of finished annotated sequences.
3. Aim to make the entire sequence freely available in the public domain for both research and development in order to maximise benefits to society

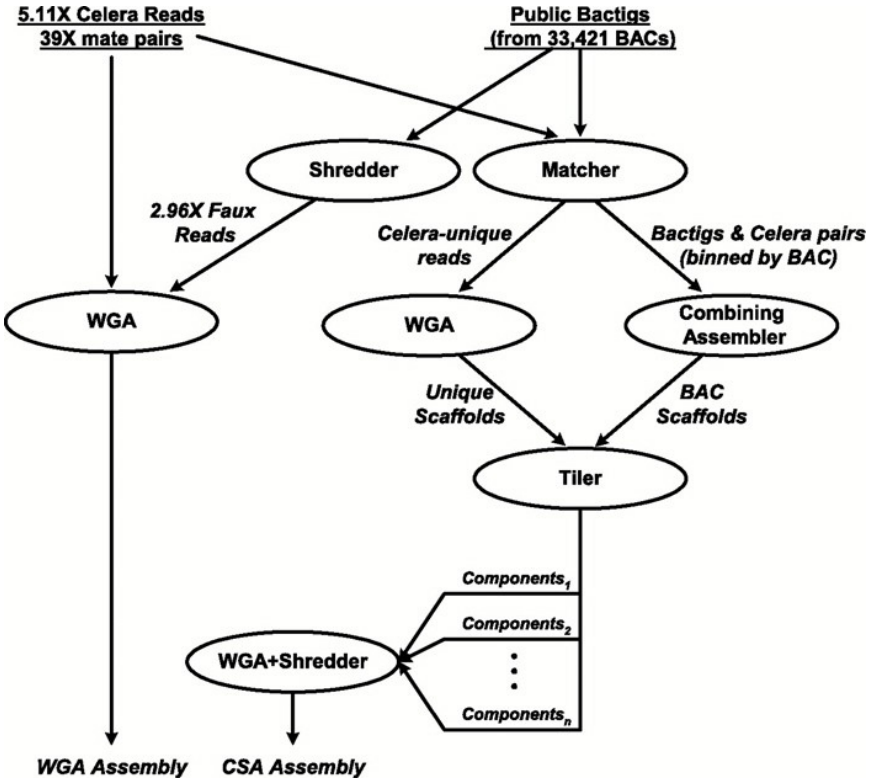
CELERA
A PE Corporation Business

HGP-The Celera story



- established in May 1998 by Dr. **J. Craig Venter** (former HGP) from The Institute for Genomic Research (TIGR)
- TIGR - „**shotgun**“ sequencing genome of *H. influenzae,*
- HGP **- $300** million of private funding for **3 years** (whole genome in 2001)
- intended to **s**ell subscriptions to its database, release data quarterly, and **obtain patents on genes** and related technologies.
- Political pressure - joint effort, draft of human genome published in 2001 in same time

HGP-The Celera story



- **Whole genome shotgun sequnecing**
- **two** independent **data sets** together with two distinct computational approaches
- Celera 27.27 million DNA sequence reads, each with an average length of 543 base pairs, derived from five different individuals
- Bactigs - DNA from HGP (GeneBank) 16.05 million sequence reads
- Whole genome assembly from 43.32 million sequence reads

HGP – major achievements

1999 - 1st human whole chromosome sequence obtained – **chromosome 22**

articles

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chissoe *et al.*†

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

Knowledge of the complete genomic DNA sequence of an organism allows a systematic approach to defining its genetic components. The genomic sequence provides access to the complete structures of all genes, including those without known function, their control elements, and, by inference, the proteins they encode, as well as all other biologically important sequences. Furthermore, the sequence is a rich and permanent source of information for the design of further biological studies of the organism and for the study of evolution through cross-species sequence comparison. The power of this approach has been amply demonstrated by the determination of the sequences of a number of microbial and model organisms. The next step is to obtain the complete sequence of the entire human genome. Here we report the sequence of the euchromatic part of human chromosome 22. The sequence obtained consists of 12 contiguous segments spanning 33.4 megabases, contains at least 545 genes and 134 pseudogenes, and provides the first view of the complex chromosomal landscapes that will be found in the rest of the genome.

HGP – major achievements

On Feb. 12, 2001 HGP and Celera announced a working draft of the sequence of the human genome — the genetic blueprint for a human being announced (HGP – Nature; Celera – Science)

President Bill Clinton holds a ceremony at the White House to announce this achievement.

15 February 2001

16 February 2001

HGP – major achievements

- **On April 14, 2003**, the IHGSC announced the successful completion of the Human Genome Project
- Project finished more than two years ahead of schedule with $2.7 billion total budget

THE WHITE HOUSE
WASHINGTON

April 11, 2003

I send greetings to the International Human Genome Sequencing Consortium as you mark the successful completion of the Human Genome Project. This historic milestone is also the 50th anniversary of the description of the double-helix structure of DNA.

The American system of medicine is a model of skill and innovation and is adding good years to countless lives. The completion of the sequencing of the human genome, begins another era of medical progress. New gene-based screening tools are alerting patients when they have an elevated risk of diseases so they can take an active role in preventing them. Scientists believe many new therapies will be tailor-made to an individual's genetic makeup, resulting in fewer adverse effects. The scientific accomplishments and international collaboration of the Consortium bring hope and promise to countless individuals who suffer from disease and others at risk.

My Administration has demonstrated our strong commitment to medical research by completing a five-year doubling of the National Institutes of Health (NIH) budget to more than $27 billion. As a result of this increase, the NIH now trains 1,500 more scientists per year and issues 10,000 more research grants than it did in 1998. This investment will help turn today's research opportunities into more of tomorrow's medical success stories.

I commend the scientists, researchers, and all others involved in the project for your tireless work to attain new scientific breakthroughs that enhance lives. Your efforts contribute to an improved system of medicine and will benefit the health and well-being of all mankind.

Laura joins me in sending our best wishes.

HGP - results

- The human **genome** contains roughly **3.2 billion** base pairs,
- The human genome contains **97%** repetitive junk DNA content, o**nly 2 to 3%** portion of the genome **encodes proteins**
- There are round **25,000 to 30,000 genes** protein coding genes,
- The average human gene consists of **3,000 nucleotide bases**, but sizes vary greatly (largest gene is the **dystrophin having 2.4Mb** in size)
- **Gene-rich** areas of the genome are predominantly made up of **G and C** bases, whereas gene-poor regions are mainly composed of A and T bases
- Chromosome **1** has the **most genes** (2968), whereas the Y chromosome has the least (231)
- The order of **99.9%** of nucleotide bases is exactly **the same in all people**
- The genome of us has **1.4** million known **SNPs.**

HGP- outcomes

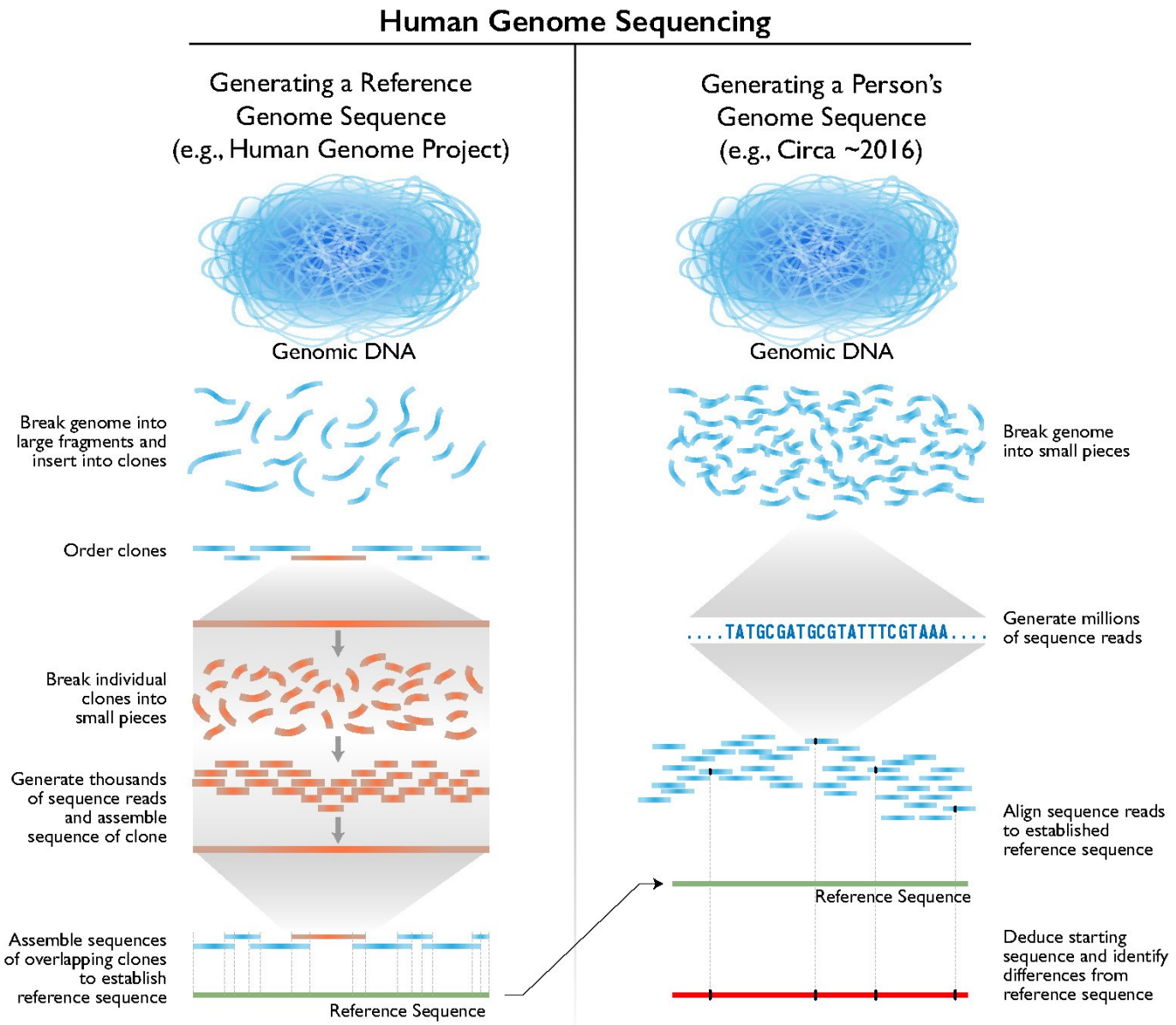
Area	Goal	Achieved	Date
Genetic Map	2- to 5-cMresolution map (600 - 1,500 markers)	1-cM resolution map(3,000 markers)	September 1994
Physical Map	30,000 STSs	52,000 STSs	October 1998
DNA Sequence	95% of gene-containing part of human sequence finished to 99.99% accuracy	99% of gene-containing part of human sequence finished to 99.99% accuracy	April 2003
Capacity and Cost of Finished Sequence	Sequence 500 Mb/year at < $0.25 per finished base	Sequence >1,400Mb/year at <$0.09 per finished base	November 2002
Human Sequence Variation	100,000 mapped human SNPs	3.7 million mapped human SNPs	February 2003
Gene Identification	Full-length human cDNAs	15,000 full-lengthhuman cDNAs	March 2003
Model Organisms	Complete genome sequences of *E. coli, S .cerevisiae, C. elegans, D. melanogaster*	Finished genome sequences of *E. coli, S. cerevisiae, C. elegans, D. melanogaster,* plus whole-genome drafts of several others, including *C. briggsae, D. pseudoobscura,* mouse and rat	April 2003
Functional Analysis	Develop genomic-scale technologies	High-throughput oligonucleotide synthesis DNA microarrays Eukaryotic, whole-genome knockouts (yeast) Scale-up of two-hybrid system for protein-protein interaction	1994 1996 1999 200

HGP - outcomes

- **Ethical, legal and social implications (ELSI)**
  - **5%** of the **annual budget** of the NHGRI was dedicated to develop ELSI
  - Genes **can't be patented**
  - **Informed consent** that should be guaranteed to those who have a **genetic test**
  - Issues related to **privacy and confidentiality** of genetic information of a person **must b**e taken care of
  - The ELSI program at NHGRI now serves as a model for large, publicly funded science efforts

"Every dollar we invested to map the human genome returned $140 to our economy — every dollar," (*B. Obama, 2013*)

https://link.springer.com/book/10.1007/978-90-481-3261-4



https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

Following projects – HapMap (2003)

The International HapMap Project

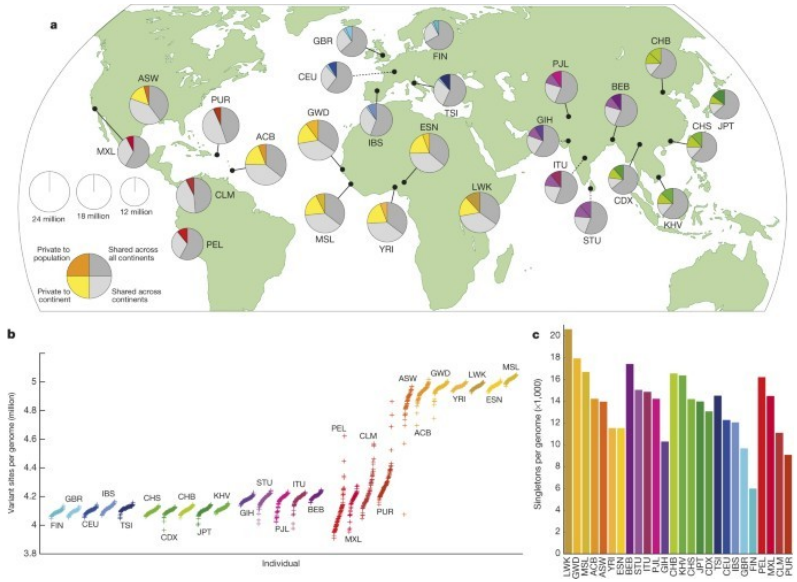
"…Determine the common patterns of DNA sequence variation in the human genome, by characterizing sequence variants, their frequencies, and correlations between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe." *Nature (2003)*

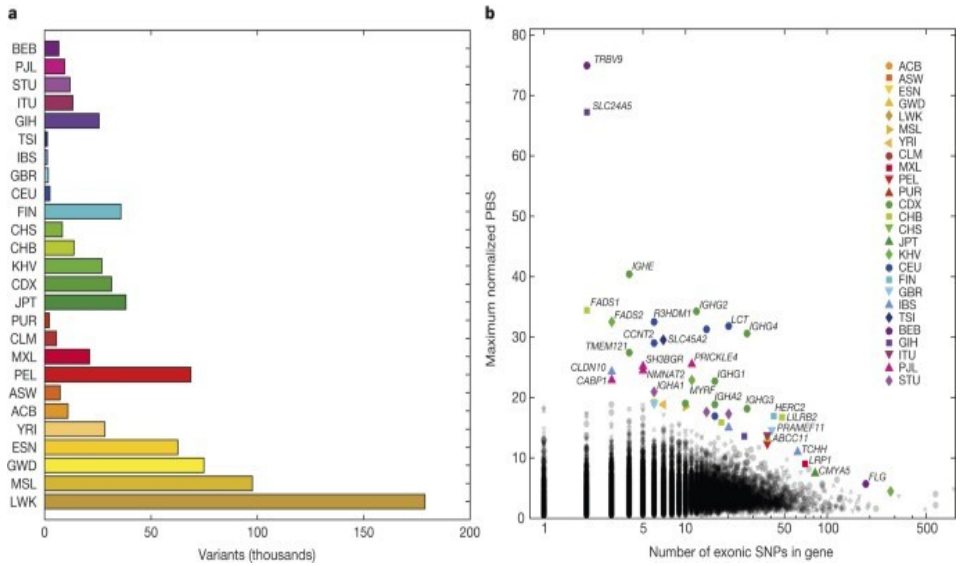
- Population-specific sequence variation
- Allele frequencies
- Linkage disequilibrium patterns
- Haplotype information
- Tag SNPs
- Structural genome variation
- Better understanding of human population dynamics and of the history of human populations
- Cell lines available from Coriell Inst. for Medical Research
- A rich resource for biomedical genetic analysis

https://slideplayer.com/slide/3053410/

Following projects – 1000Genomes

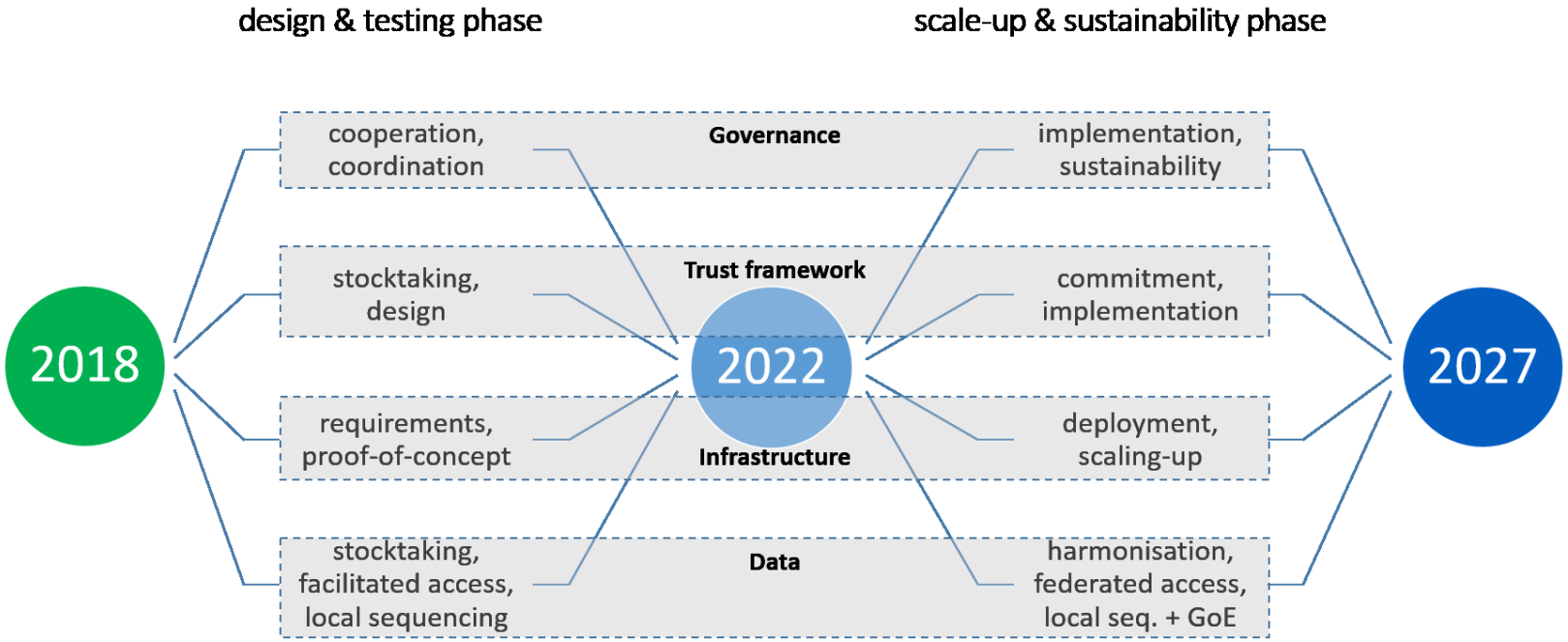
(2008-2015)





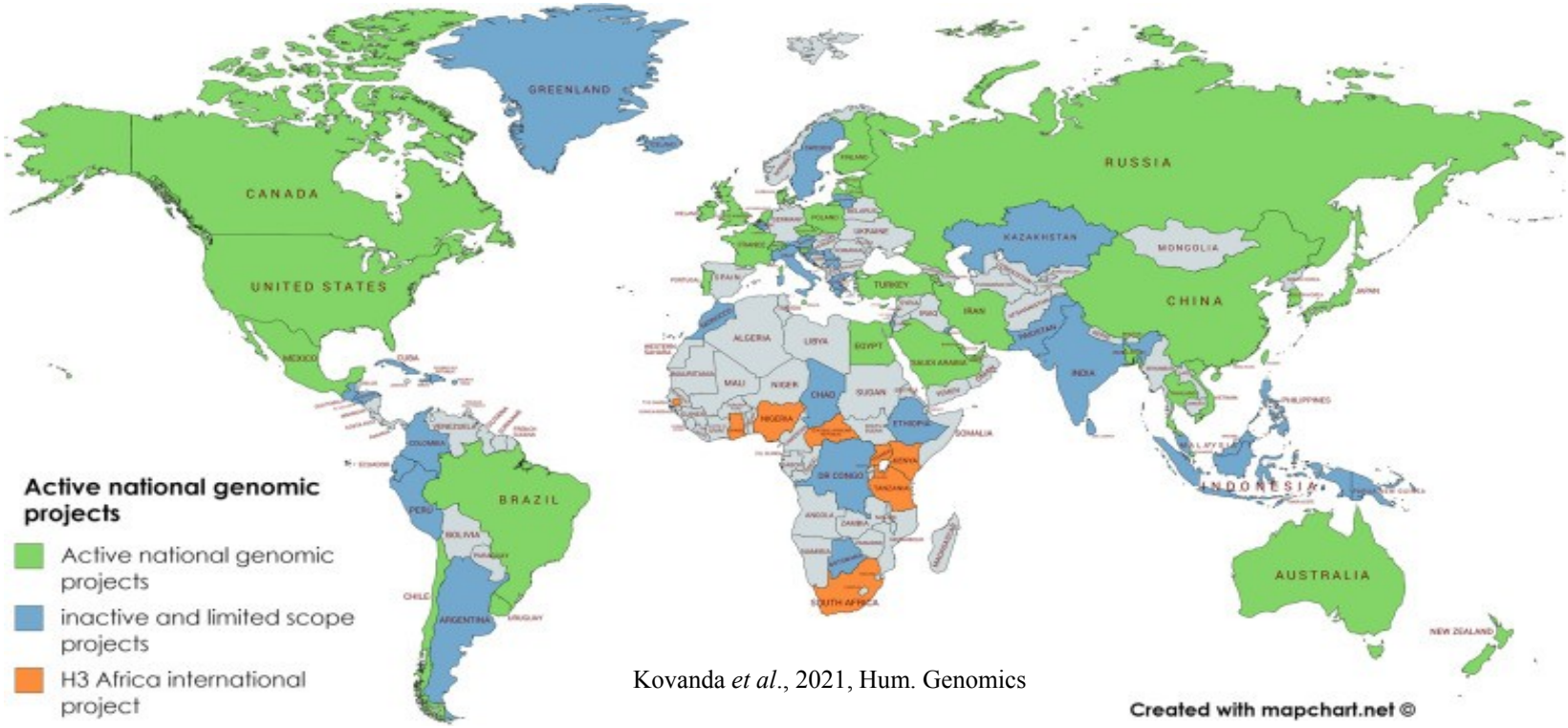
- The goal of the 1000 Genomes Project was to find common genetic variants with frequencies of at least 1% in the populations studied
- WGS of **2,504** individuals from 26 populations
- **Creation of global human genome reference**

+1 Million Genomes



22 EU countries, the UK and Norway signed Member States' **declaration** on stepping up efforts towards creating a European **data infrastructure** for **genomic data** and implementing common national **rules** enabling federated **data access**

Local population genome projects



Kovanda *et al*., 2021, Hum. Genomics

Worldwide, there are 86 or more project focused on improve genetic diagnostics and to pave the way for the integration of precision medicine into health systems