Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Biology Recap

Borchert, Dr. Schapranow

Data Management for Digital Health
Winter 2023

Agenda
Pillars of the Lecture

Hasso Plattner Institut

Medical Use Cases

Biology Recap

Oncology

Nephrology

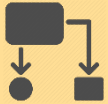
Infectious Diseases

Technology Foundation

Data Sources

Data Formats

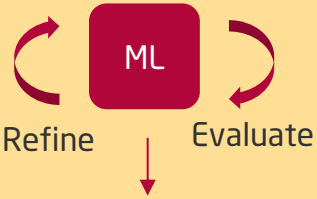
Processing and Analysis



Software Architectures

Machine Learning

Data

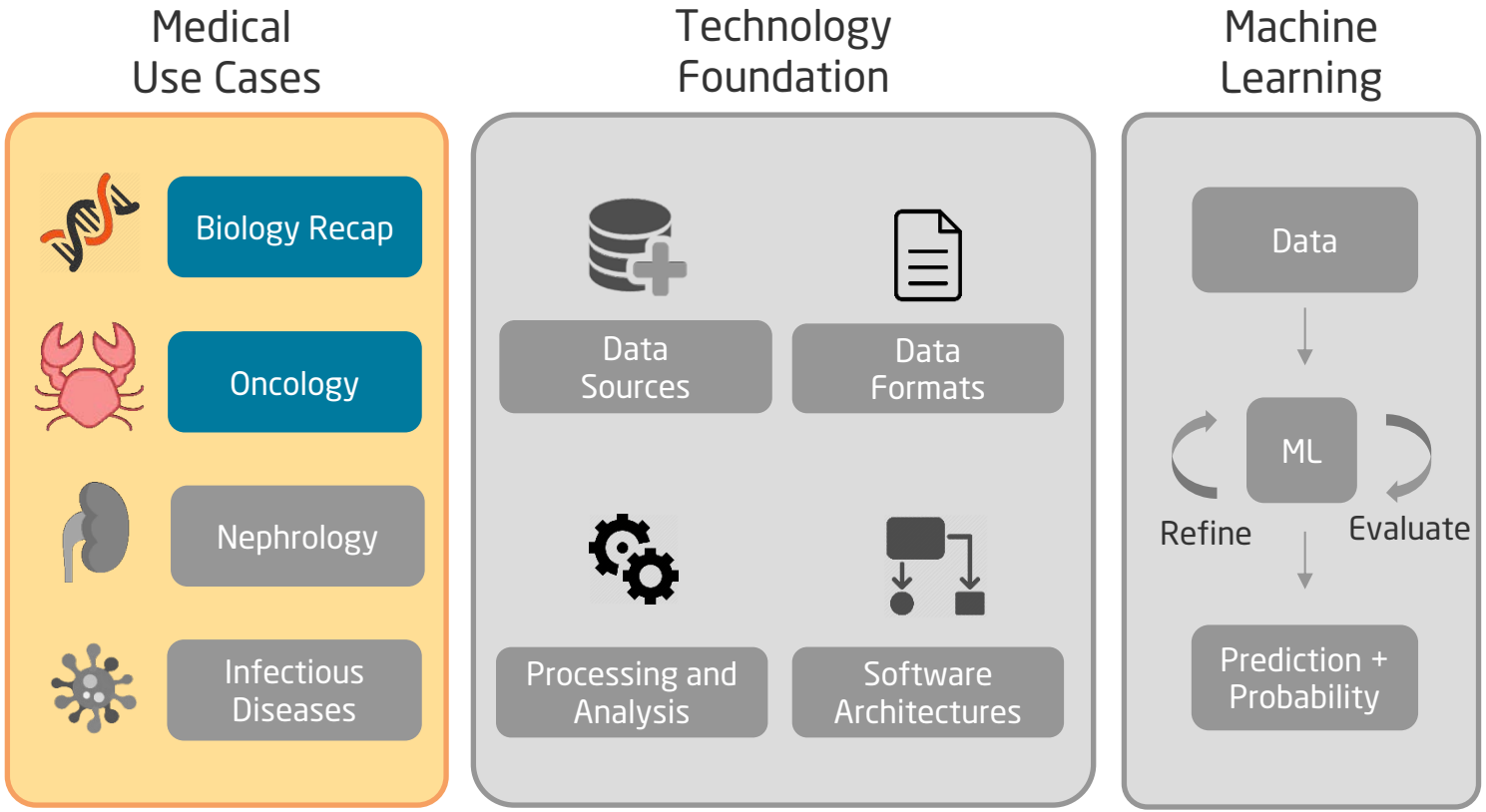


Prediction + Probability

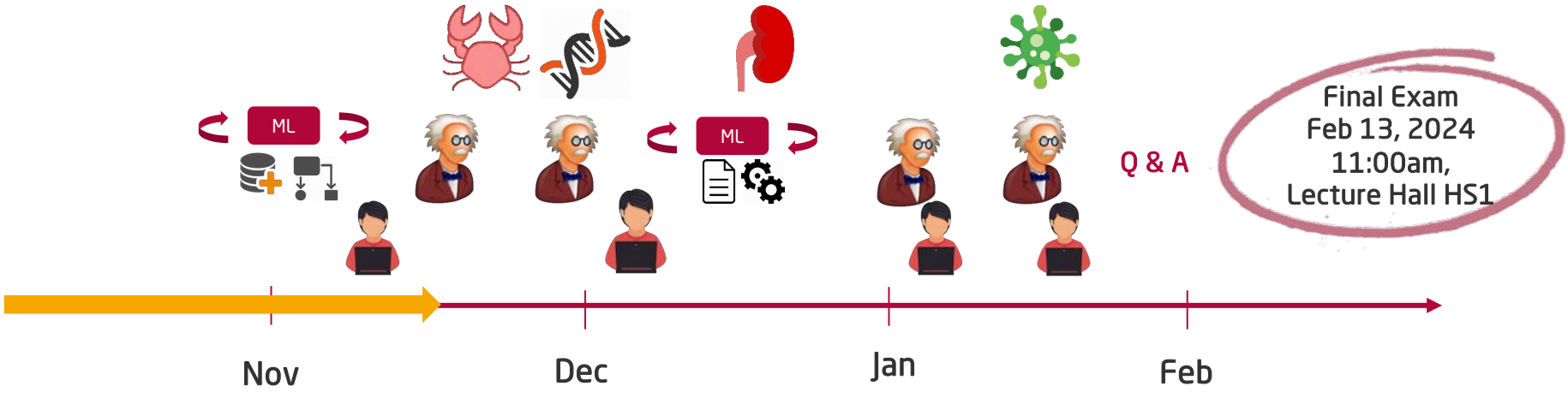
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Agenda
Pillars of the Lecture



Lecture Schedule



- Lecture Kickoff
- Actors in Healthcare
- Digital Health Data

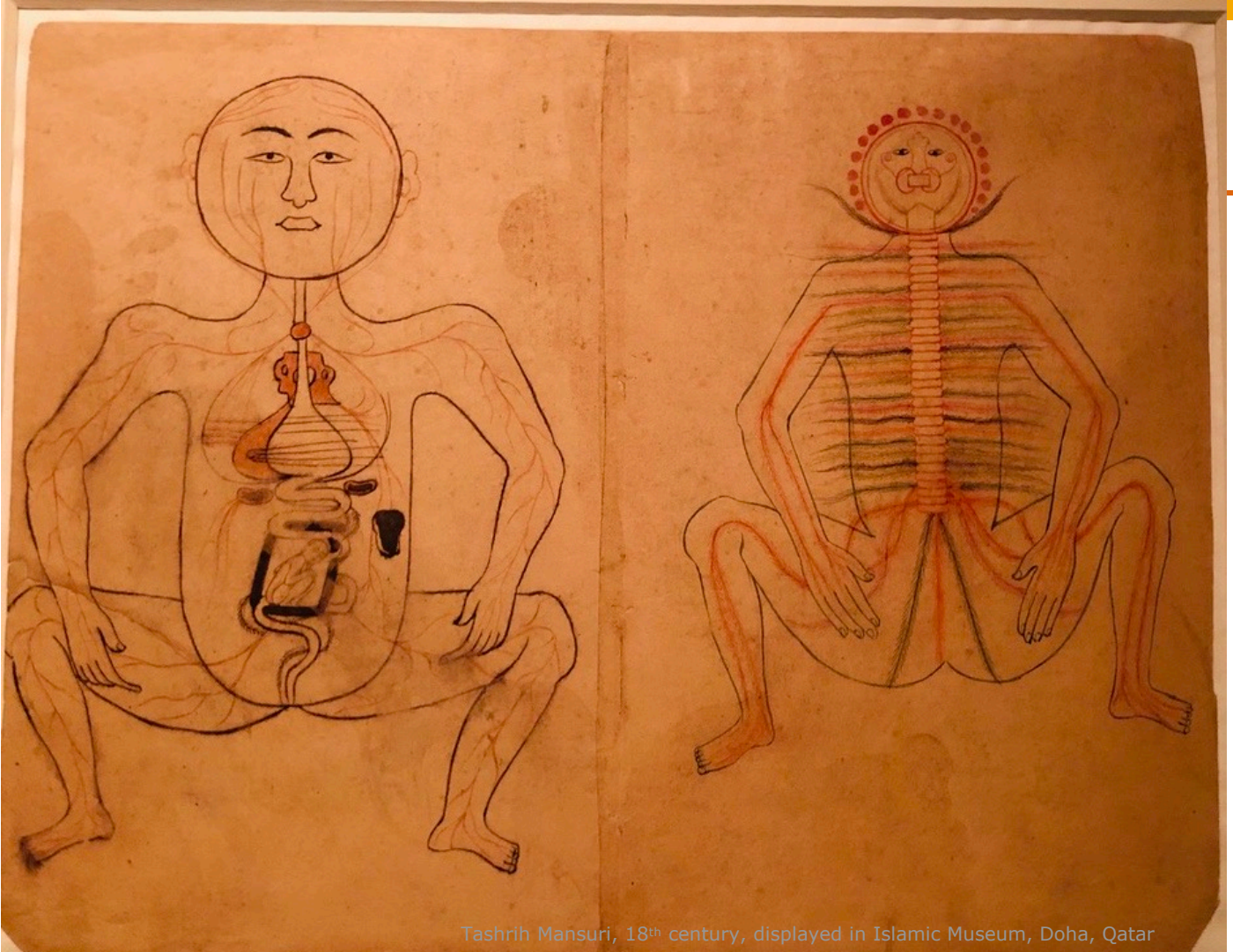
- Machine Learning (ML) Foundations
- Use Case Oncology
- Biology Recap

- Natural Language Processing
- Use Case Nephrology & Intensive Care
- Supervised ML & Deep Learning

- Use Case Infectious Diseases
- Unsupervised ML

Agenda

- Facts you should know
- Discovery of
  - Cells,
  - DNA/RNA structures,
  - The human genome
- Biology recap
  - Pro- vs. eukaryotes
  - Cell components
  - Genetic changes and defects
- Build your own DNA



Tashrih Mansuri, 18th century, displayed in Islamic Museum, Doha, Qatar

Hasso Plattner Institut

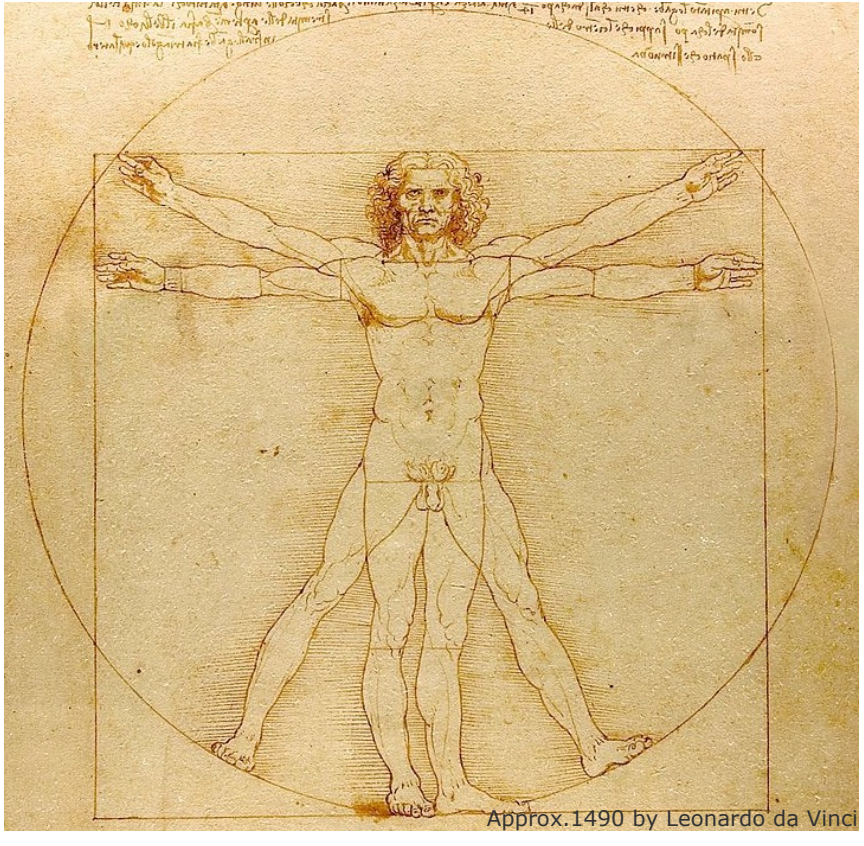
A global trend started centuries ago: Understanding internals of human-beings

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Humans in Numbers

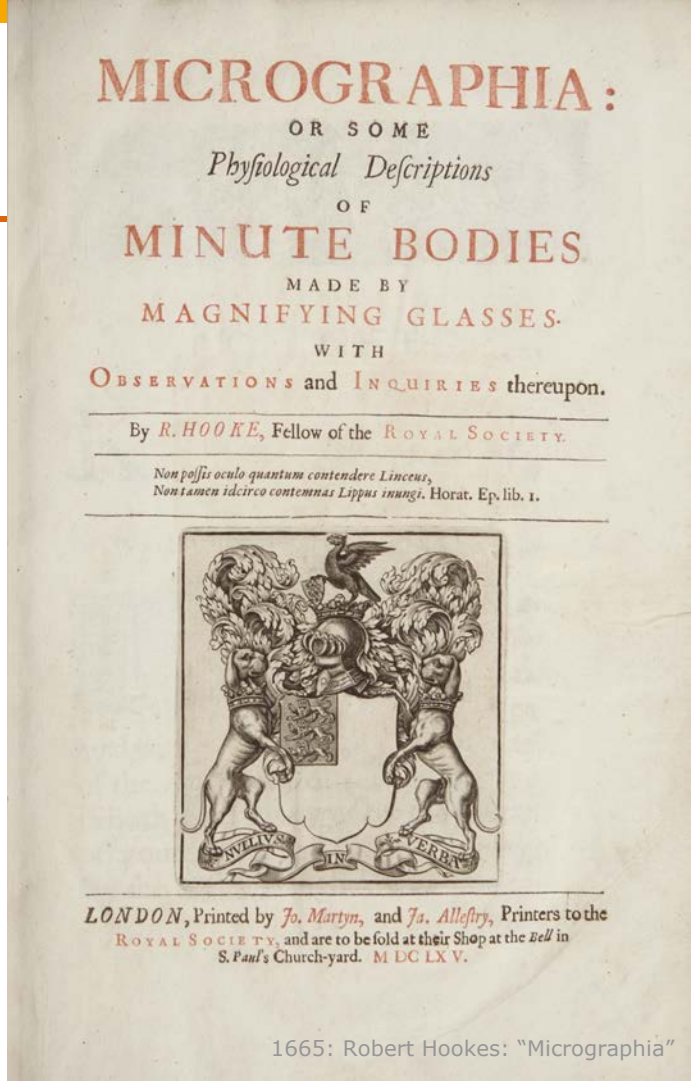
- Genome size: 3.2 Gbp
- Genes: approx. 20k-25k
- Chromosomes: 22 + XY
- Mean gene size: 27 kbp



Approx.1490 by Leonardo da Vinci

History of Cells

- 1665: Robert Hookes published his textbook “Micrographia”
- Shared the power of microscopic observations
- Defined term “cell” using the plant cork


MICROGRAPHIA:
OR SOME
Physiological Descriptions
OF
MINUTE BODIES
MADE BY
MAGNIFYING GLASSES.
WITH
OBSERVATIONS and INQUIRIES thereupon.
By *R. HOOKE*, Fellow of the ROYAL SOCIETY.
Non possis oculo quantum contendere Linceus,
Non tamen idcirco contemnas Lippus inungi. Horat. Ep. lib. 1.
NVLLIVS IN VERBA
LONDON, Printed by *Jo. Martyn*, and *Ja. Allestry*, Printers to the ROYAL SOCIETY, and are to be sold at their Shop at the *Bell* in S. *Paul's* Church-yard. M DC LX V.

1665: Robert Hookes: “Micrographia”

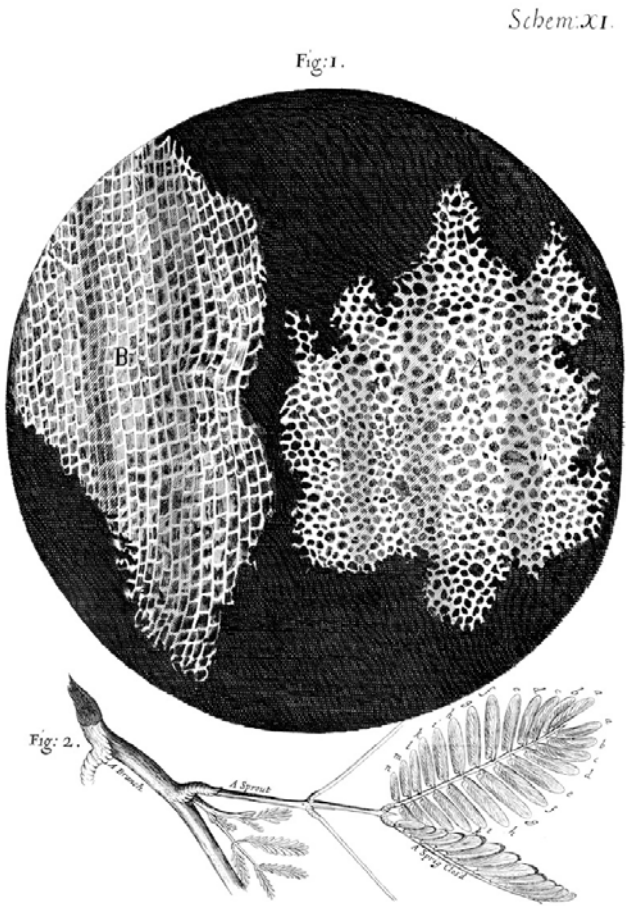
Hasso Plattner Institut

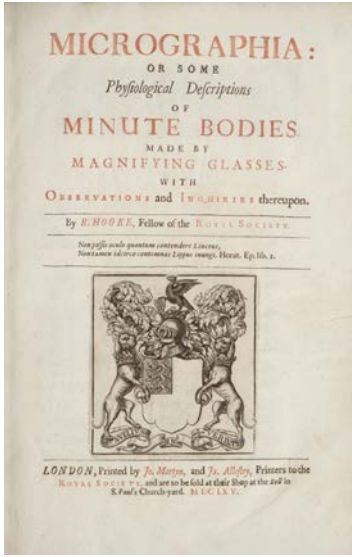
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

History of Cells

- "Observ. XVIII. Of the Schematisme or Texture of Cork, and of the Cells and Pores of some other such frothy Bodies."
- "... it had a very little solid substance..."
- "...for the *Interstitia*, or walls (as I may so call them) or partitions of those pores were neer as thin in proportion to their pores..."
- "Next, in that these pores, or cells, were not very deep, but consisted of a great many little Boxes, separated out of one continued long pore..."
- Continue reading at [1].





Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

[1] 1665: Robert Hookes: "Micrographia", online available at: http://www.gutenberg.org/files/15491/15491-h/15491-h.htm

Human Cells
<< QUIZ >>

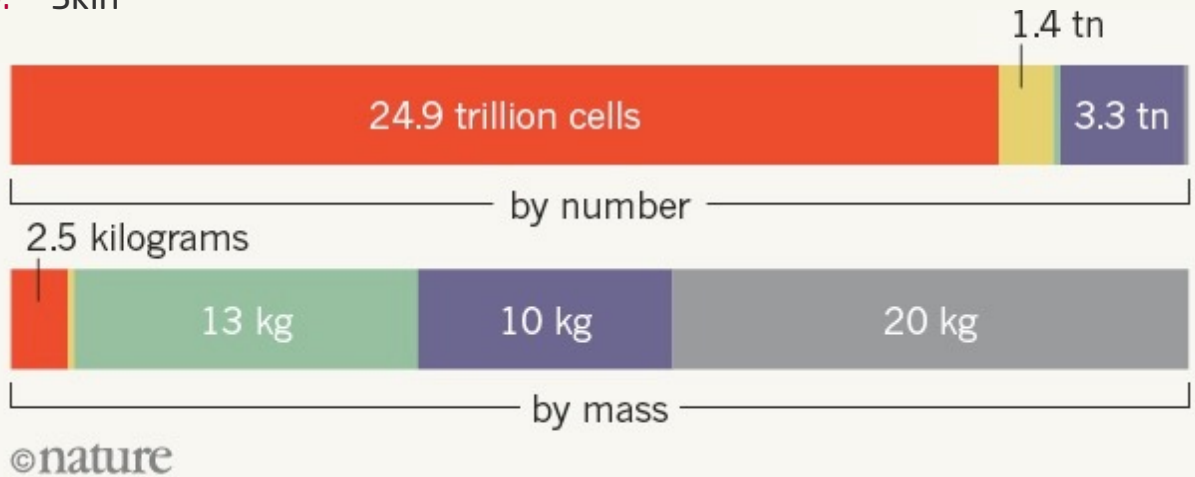
To what type do the most cells by number belong to?

A. Brain

B. Hair

C. Blood

D. Skin

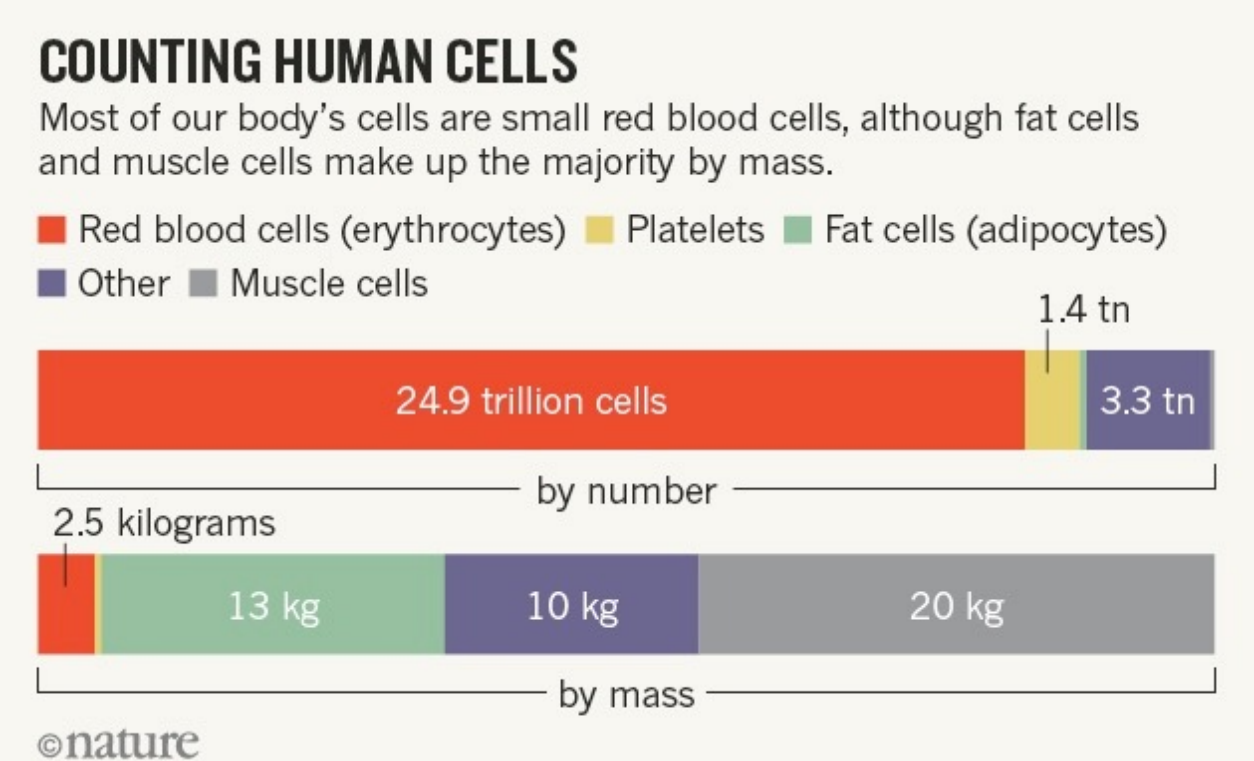


Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Ron Sender, Shai Fuchs, Ron Milo: "Revised estimates for the number of human and bacteria cells in the body", *PLOS Biology*, 2016

Human Cells
<< QUIZ >>

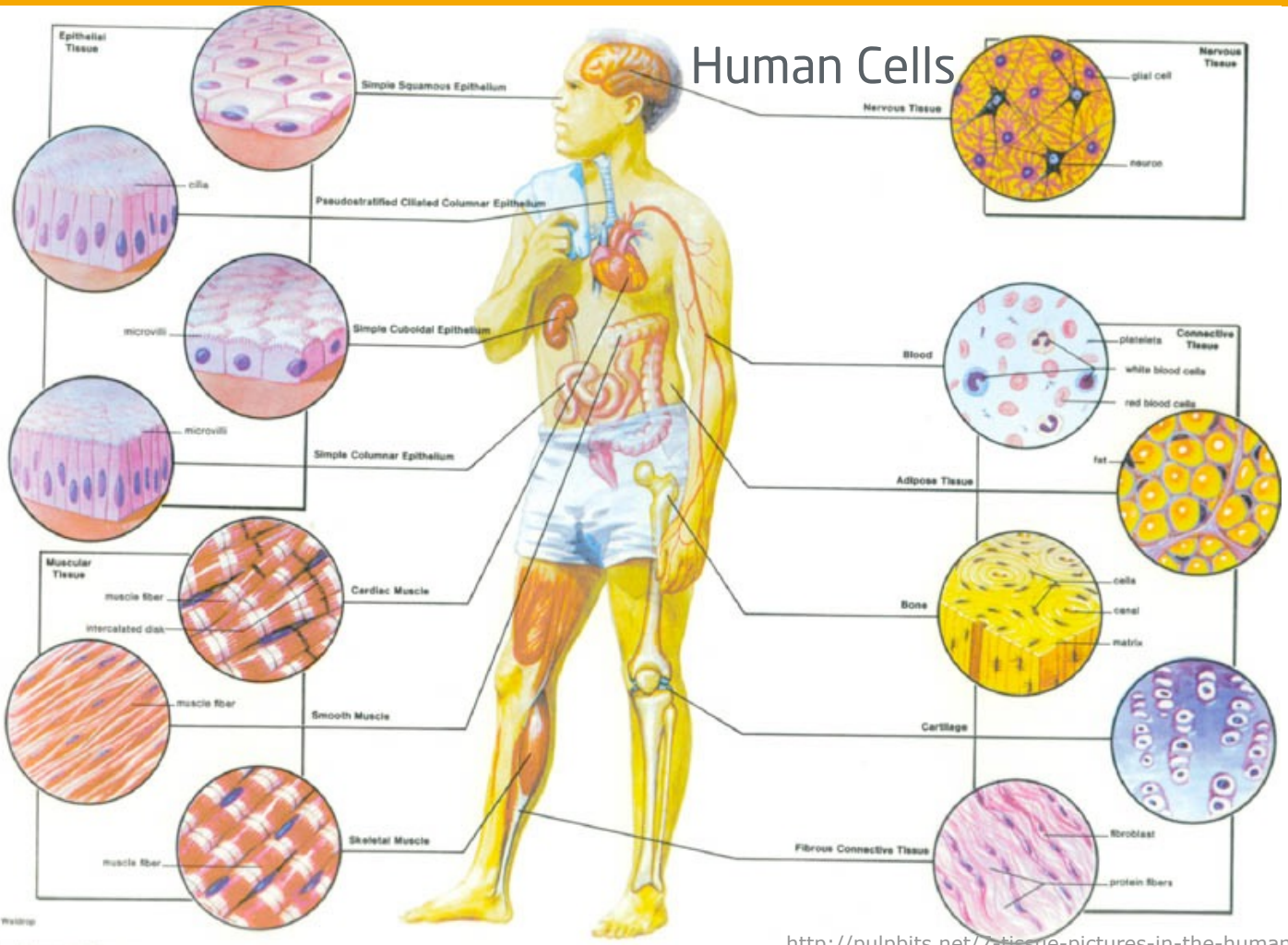


Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Ron Sender, Shai Fuchs, Ron Milo: "Revised estimates for the number of human and bacteria cells in the body", *PLOS Biology*, 2016

Human Cells



59 Human Tissues
Figure 9.1

Sylvia S. Mader, *Inquiry into Life*, 6th ed. Copyright © 1991 Wm. C. Brown Publishers, Dubuque, Iowa. All Rights Reserved.

http://pulpbits.net/7-tissue-pictures-in-the-human-body/

Hasso Plattner Institut

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Categorization of Cells

- Prokaryote := Absence of nucleus, single-cell organisms, e.g. bacteria

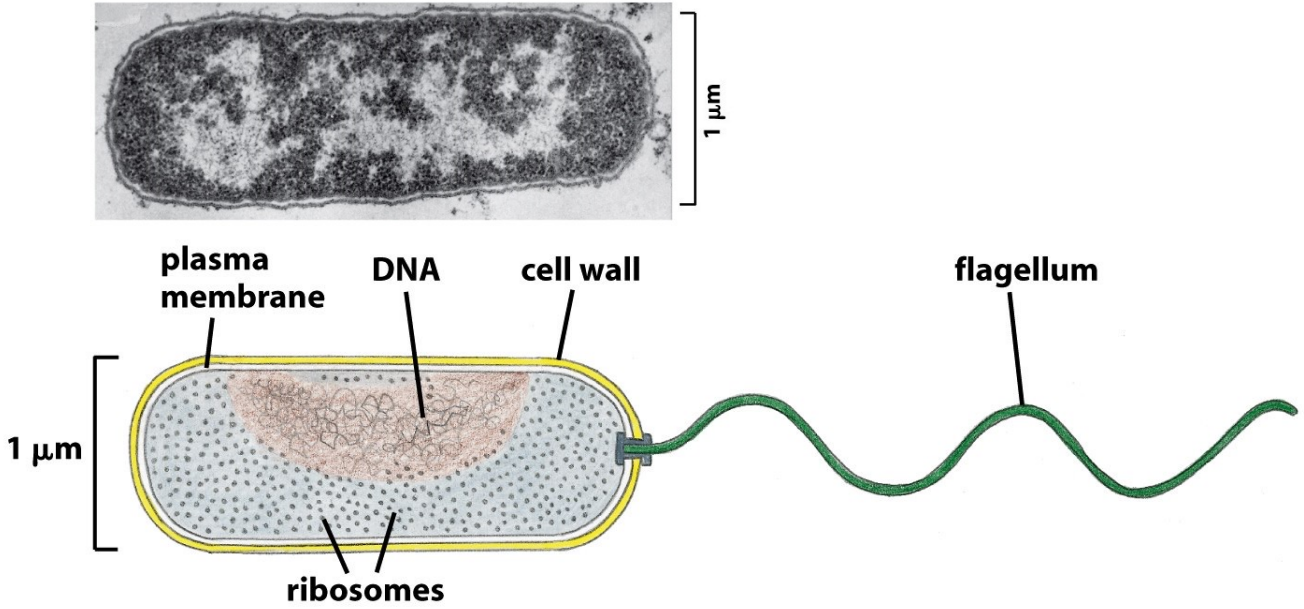


Figure 1-18 *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

Categorization of Cells

- Eukaryote := Cell nucleus + cell organelles within membranes, e.g. plants and animals

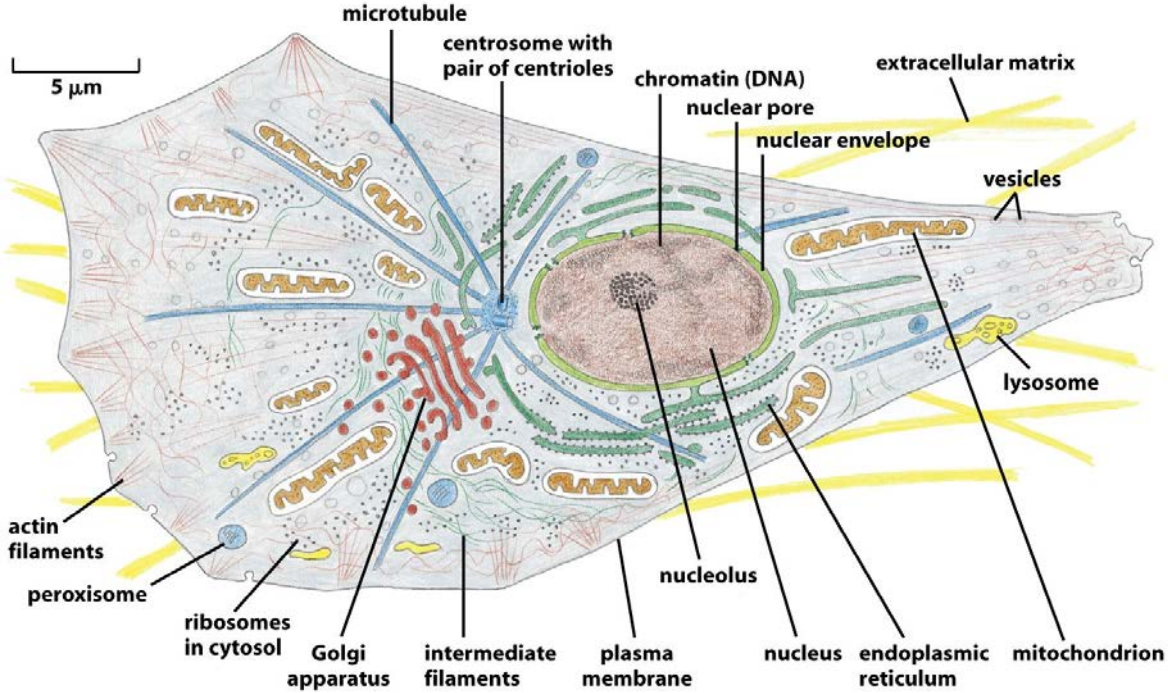


Figure 1-30 *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

Categorization of Cells

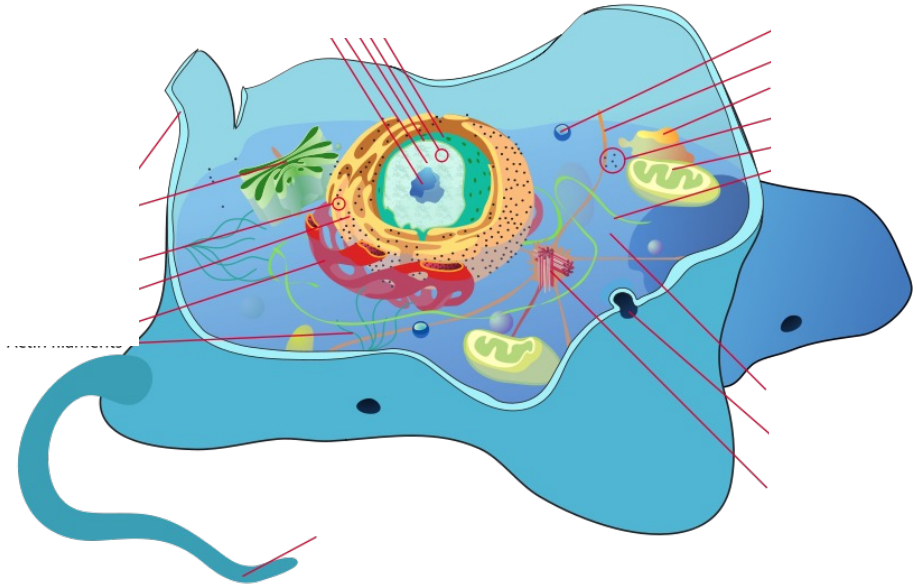
- Prokaryote := Absence of nucleus, single-cell organisms, e.g. bacteria
- Eukaryote := Cell nucleus + cell organelles within membranes, e.g. plants and animals

	Prokaryotes	Eukaryotes
Size	1–10 µm	10–100 µm
Organisms	bacteria, archaea	protists, fungi, plants, animals
DNA form	circular	linear

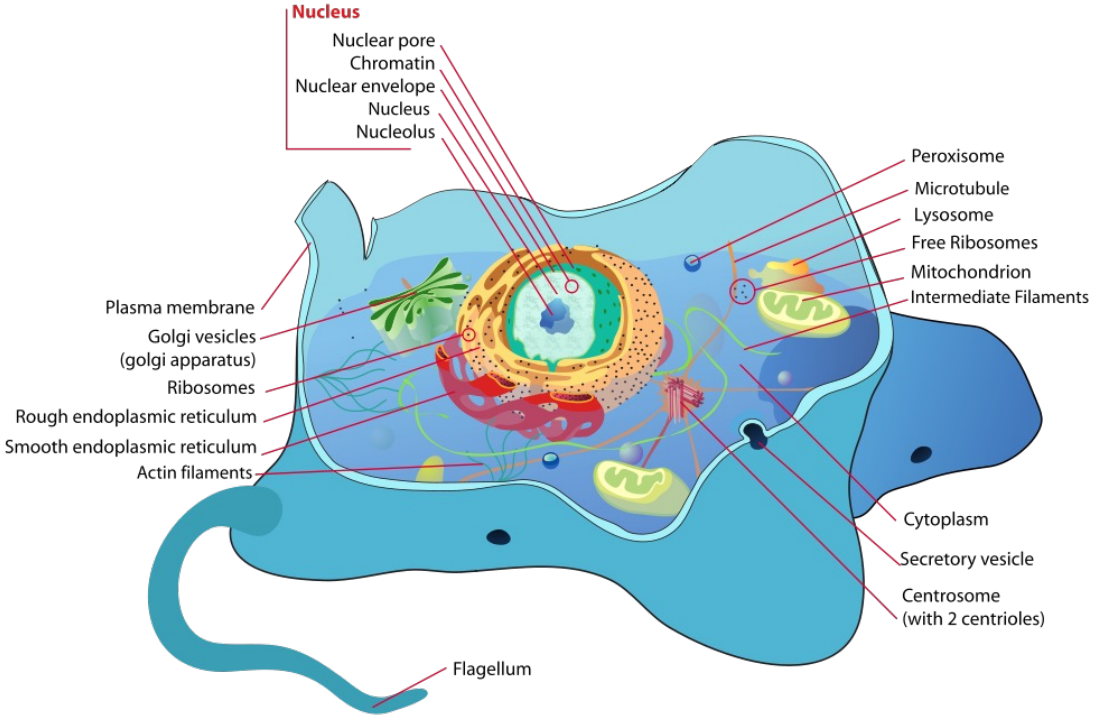
- Further classification:
  - Germ(-line) cells := Blueprint for differentiation of gametes
  - Gametes := Store genetic material for reproduction, e.g. ovozyte, sperm
  - Somatic cells := All remaining body cells, e.g. skin cell, tissue, organs, etc.

Components of Eukaryotic Cells (Organelles)
<< Brainstorming Sessions >>

- Task: What cell components (organelles) of eukaryotic cells do you remember?



Components of Eukaryotic Cells (Organelles)

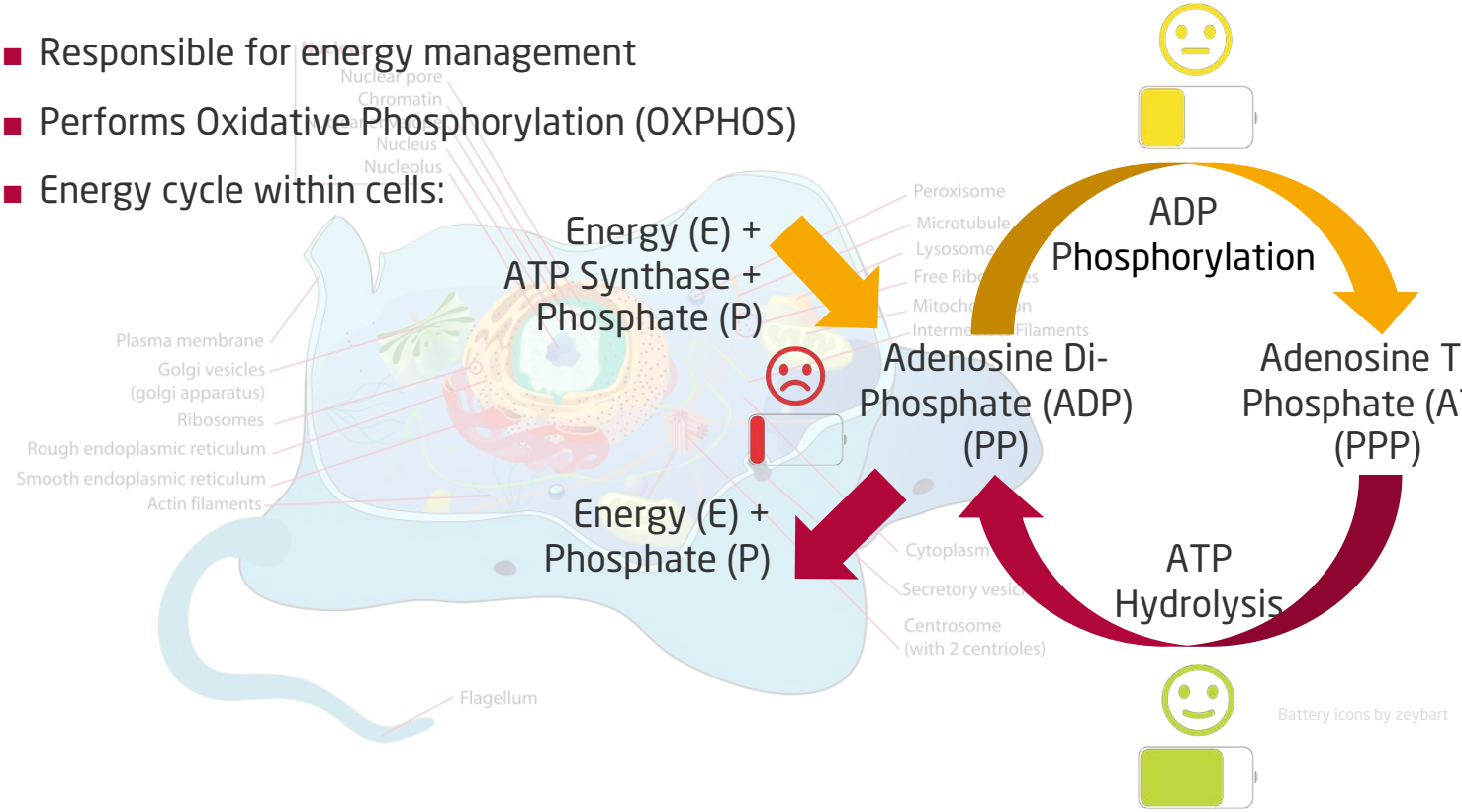


Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Components of Eukaryotic Cells (Organelles)
Mitochondria

- Responsible for energy management
- Performs Oxidative Phosphorylation (OXPHOS)
- Energy cycle within cells:

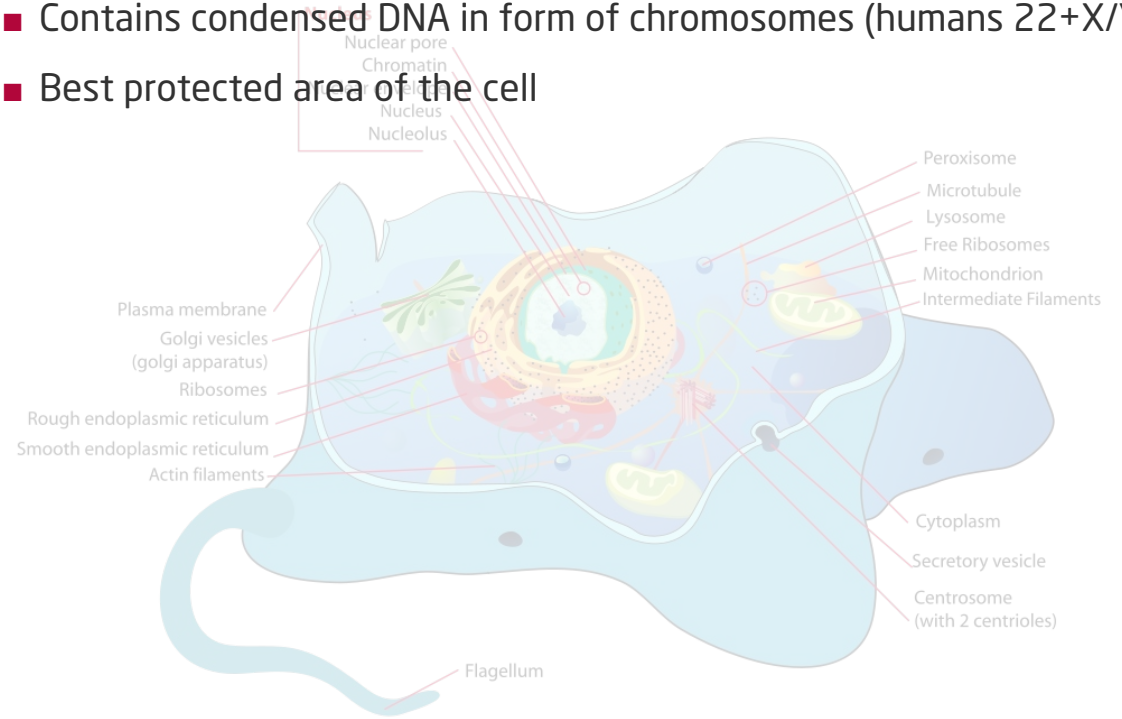


Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

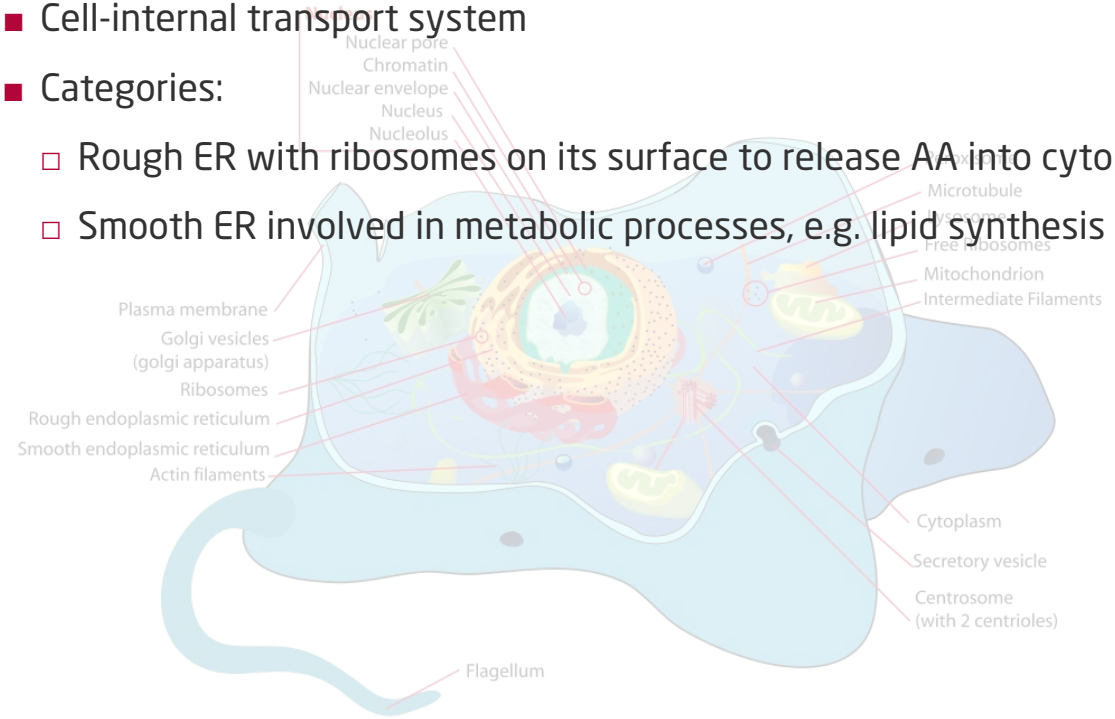
Components of Eukaryotic Cells (Organelles)
Cell Nucleus

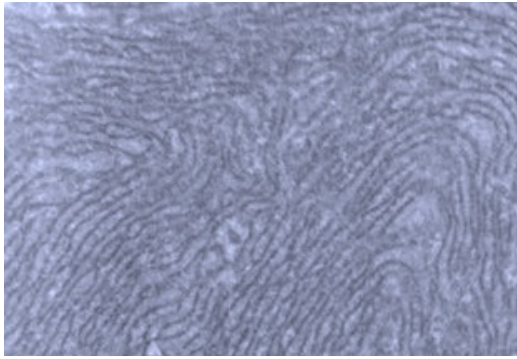
- Contains condensed DNA in form of chromosomes (humans 22+X/Y)
- Best protected area of the cell



Components of Eukaryotic Cells (Organelles)
Endoplasmic Reticulum (ER)

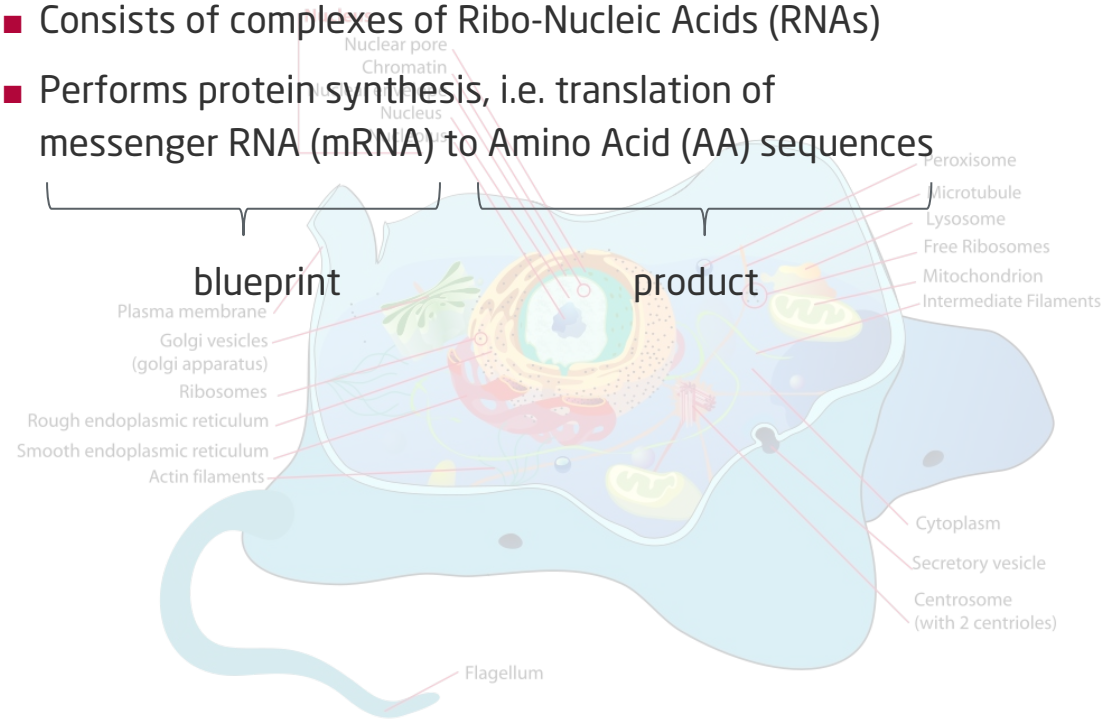
- Cell-internal transport system
- Categories:
  - Rough ER with ribosomes on its surface to release AA into cytoplasm
  - Smooth ER involved in metabolic processes, e.g. lipid synthesis





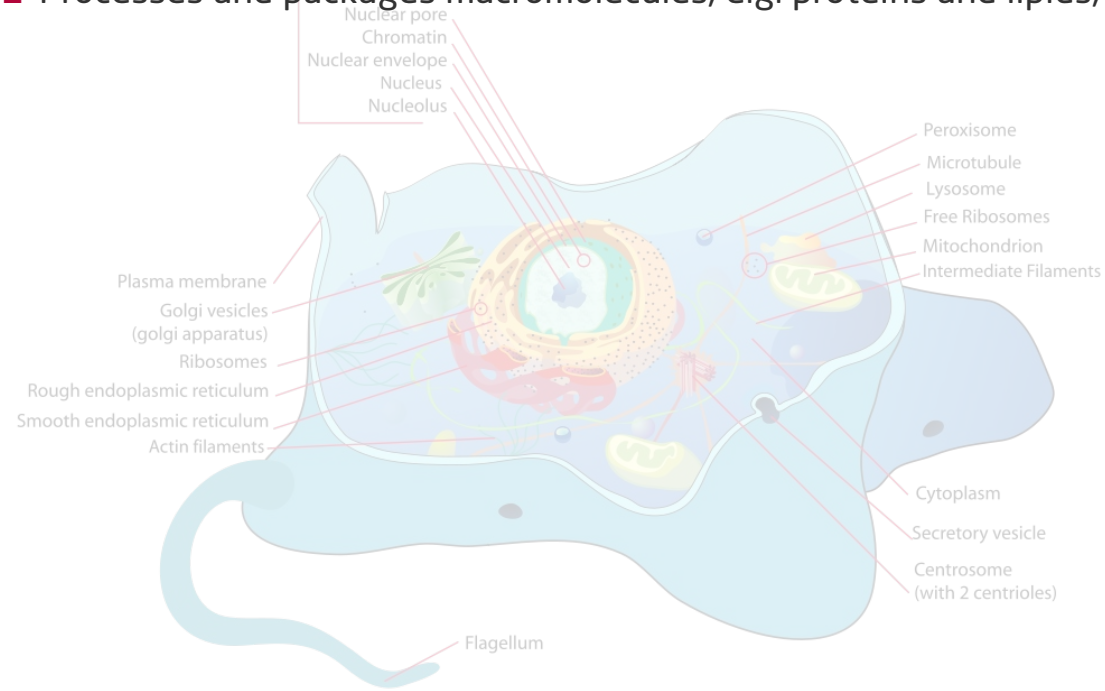
Components of Eukaryotic Cells (Organelles)
Ribosome

- Consists of complexes of Ribo-Nucleic Acids (RNAs)
- Performs protein synthesis, i.e. translation of messenger RNA (mRNA) to Amino Acid (AA) sequences



Components of Eukaryotic Cells (Organelles)
Golgi Apparatus

- Processes and packages macromolecules, e.g. proteins and lipids, prior to transport



What to take Home?

- Mitochondria: Power supply for the cell

- Cell core: Contains source code, i.e. DNA

- Endoplasmic reticulum: Provides transport network

- Ribosomes: Compiler, e.g. mRNA to AA

- Golgi apparatus: Packaging, e.g. proteins

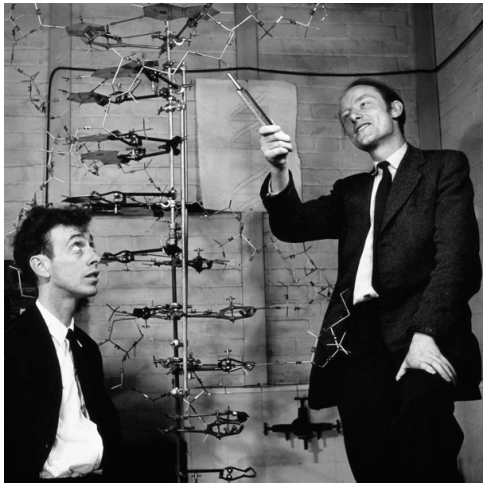
Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Discovery of the Human Genome

Discovery of the Human Genome
<< QUIZ >>

- When was the human DNA model discovered as we know it today?

A. 1890s
B. 1910s
C. 1930s
D. 1950s



Medical Use Case Oncology

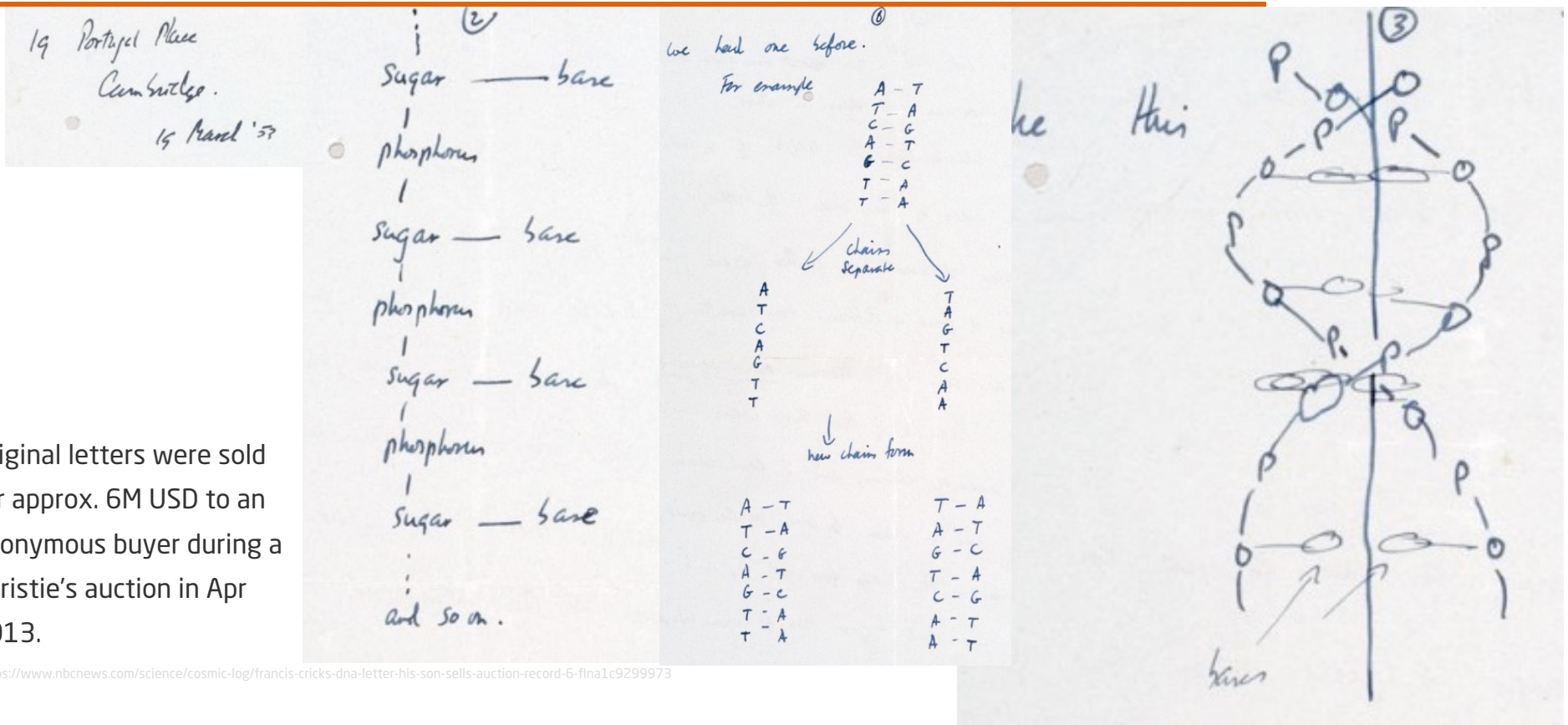
Data Management for Digital Health, Winter 2023

https://www.theguardian.com/science/2015/jun/23/sexism-in-science-did-watson-and-crick-really-steal-rosalind-franklins-data
https://www.chemheritage.org/historical-profile/james-watson-francis-crick-maurice-wilkins-and-rosalind-franklin

Discovery of the Human Genome

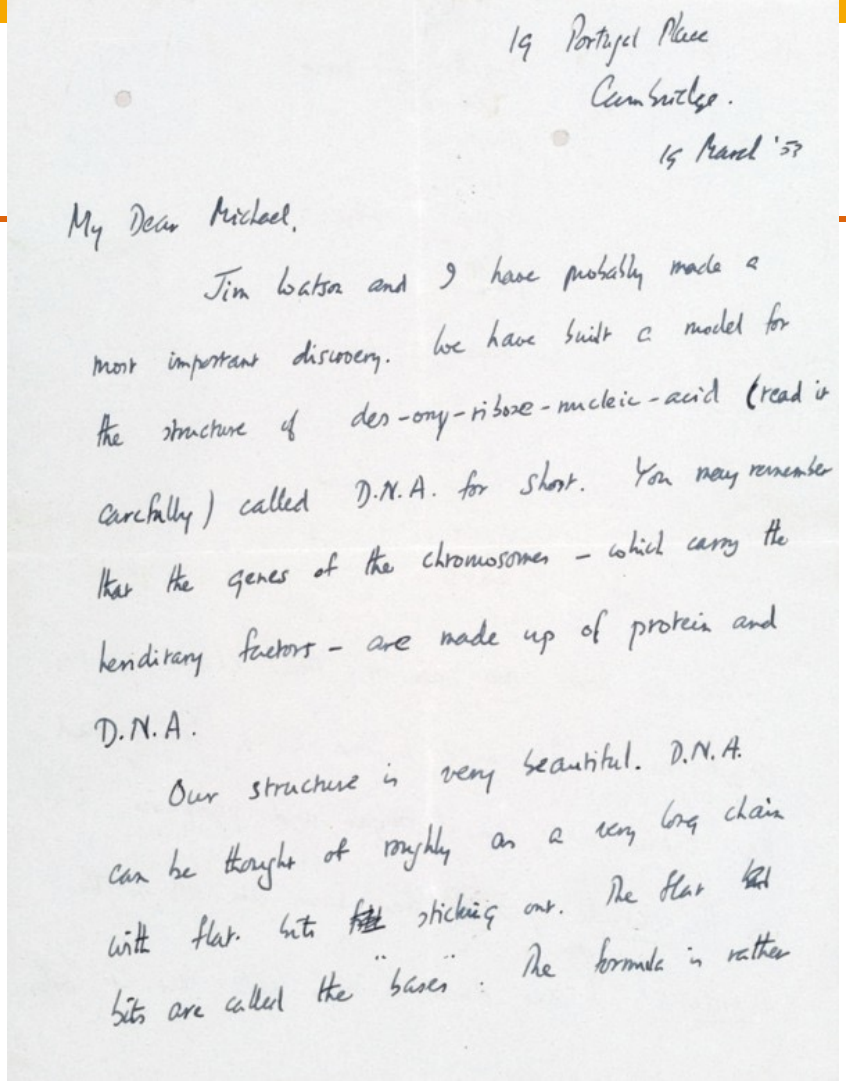
- 1869: Swiss physiological chemist *Friedrich Miescher* accidentally discovered nuclein whilst investigating proteins of leukocytes → nucleic acid
- 1919: Russian biochemist *Phoebus Levene* defined polynucleotide model consisting of four bases, sugar, and phosphate following same repetition
- 1944: *Oswald Avery* discovered that DNA composes hereditary units → genes
- 1950: *Erwin Chargaff* discovered that DNA varies across species whilst the amount of A,T and C,G keeps balanced
- 1953: Physicist *Francis Crick* & biologist *James Watson* define model for structure of Des-oxy-ribose-nucleic-acid (D.N.A.)

Discovery of the Human Genome
1953 Crick's Letter to His Son Michael



Original letters were sold for approx. 6M USD to an anonymous buyer during a Christie's auction in Apr 2013.

https://www.nbcnews.com/science/cosmic-log/francis-cricks-dna-letter-his-son-sells-auction-record-6-flna1c9299973



19 Portugal Place
Cambridge.
19 March '53

My Dear Michael,

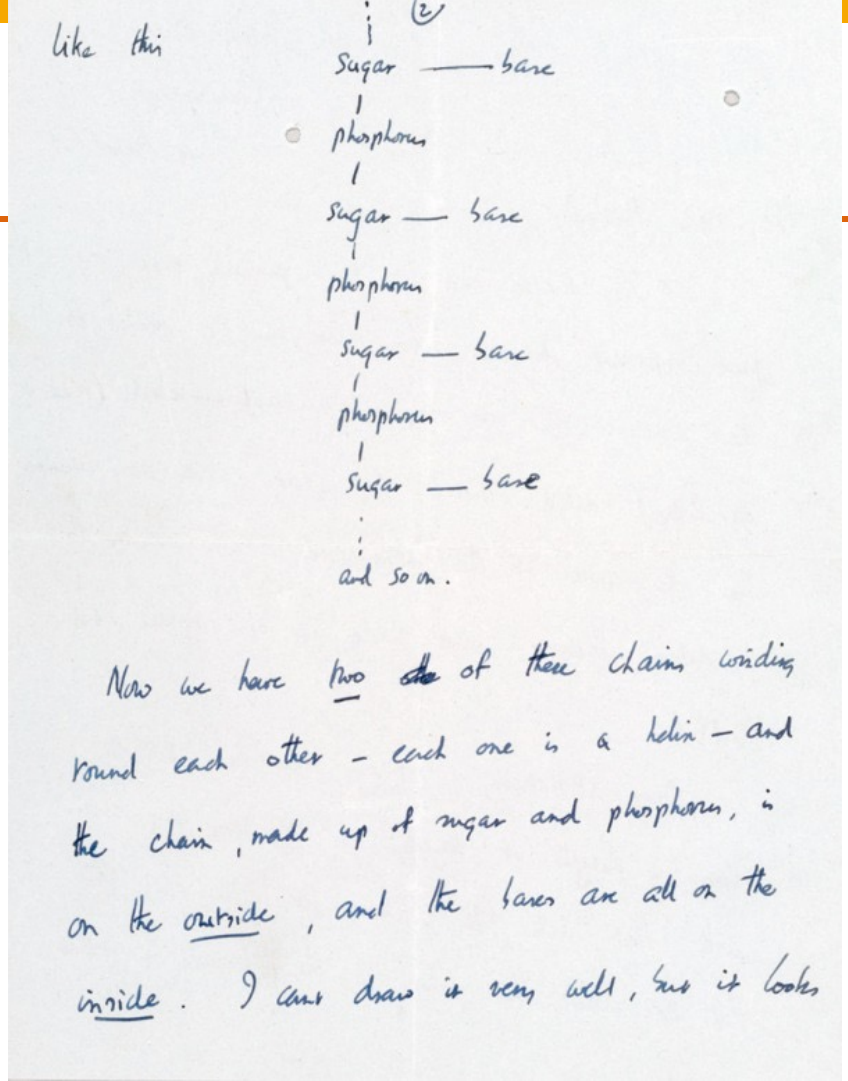
Jim Watson and I have probably made a most important discovery. We have built a model for the structure of des-oxy-ribose-nucleic-acid (read it carefully) called D.N.A. for short. You may remember that the genes of the chromosomes - which carry the hereditary factors - are made up of protein and D.N.A.

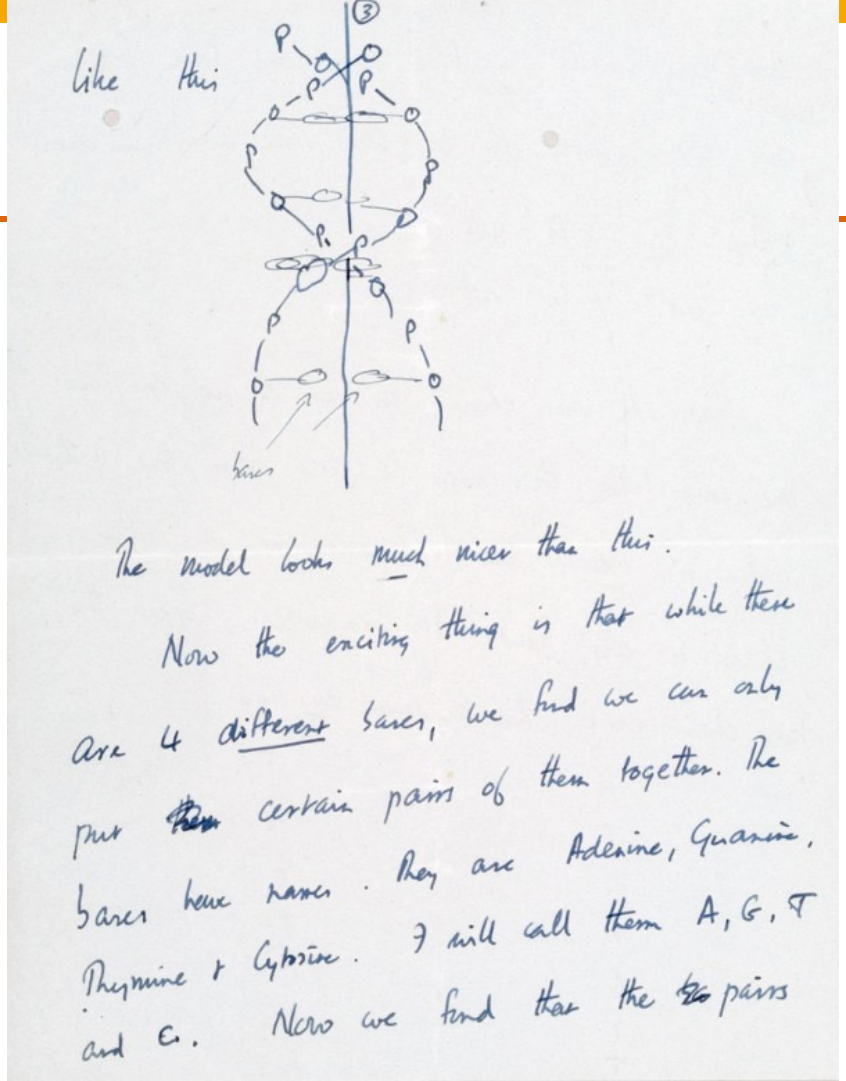
Our structure is very beautiful. D.N.A. can be thought of roughly as a very long chain with flat bits sticking out. The flat bits are called the bases. The formula is rather

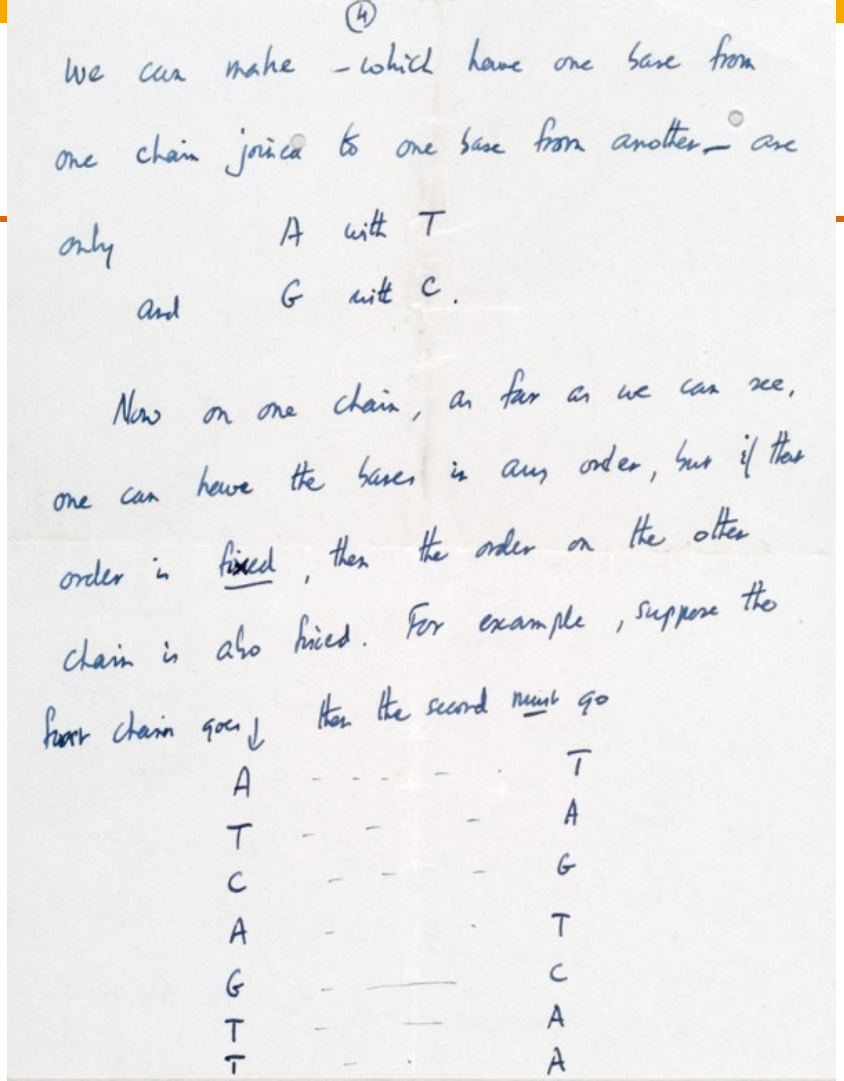
Hasso Plattner Institut

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023


like this
Sugar — base
phosphorus
Sugar — base
phosphorus
Sugar — base
phosphorus
Sugar — base
and so on.
Now we have two ~~the~~ of these chains winding round each other - each one is a helix - and the chain, made up of sugar and phosphorus, is on the outside, and the bases are all on the inside. I can't draw it very well, but it looks


Like this
bases
The model looks much nicer than this.
Now the exciting thing is that while there are 4 different bases, we find we can only put certain pairs of them together. The bases have names. They are Adenine, Guanine, Thymine & Cytosine. I will call them A, G, T and C. Now we find that the pairs

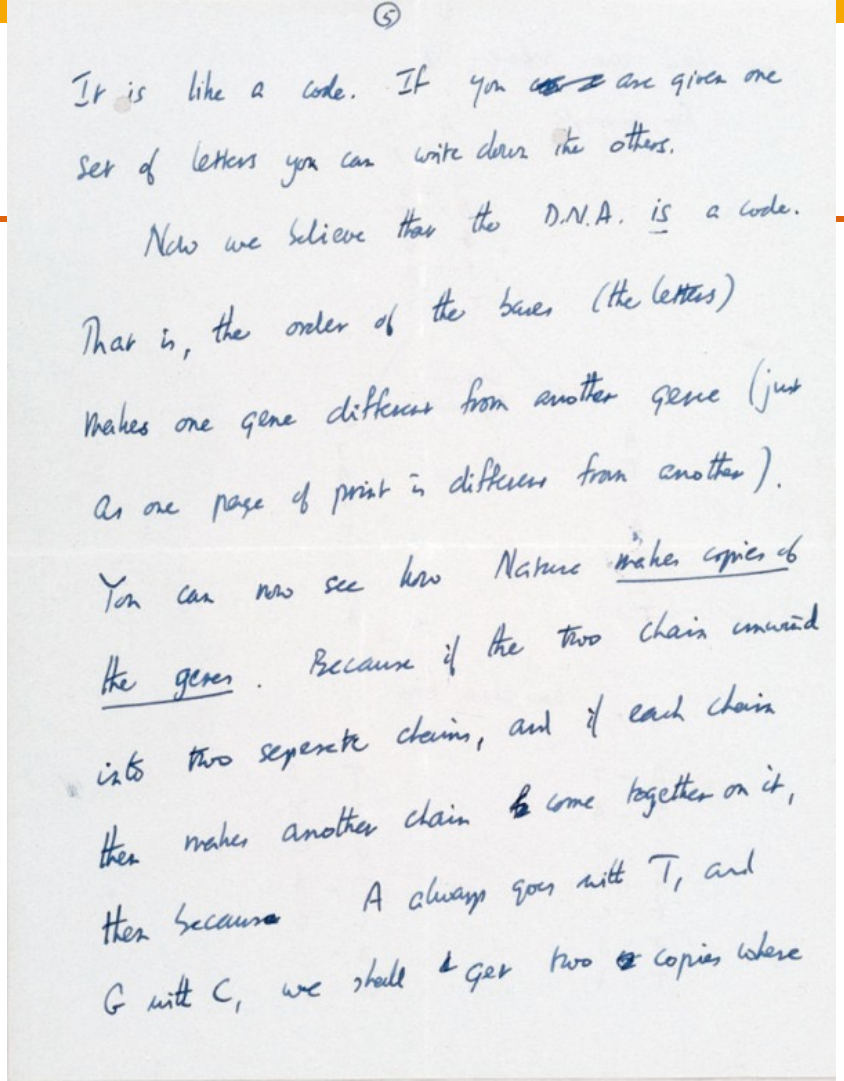


(4)

We can make - which have one base from one chain joined to one base from another - are only A with T and G with C.

Now on one chain, as far as we can see, one can have the bases in any order, but if that order is fixed, then the order on the other chain is also fixed. For example, suppose the first chain goes ↓ then the second must go

A - - - - T
T - - - A
C - - - G
A - - T
G - - C
T - - A
T - A



⑤

It is like a code. If you ~~are~~ are given one set of letters you can write down the others.

Now we believe that the D.N.A. is a code.

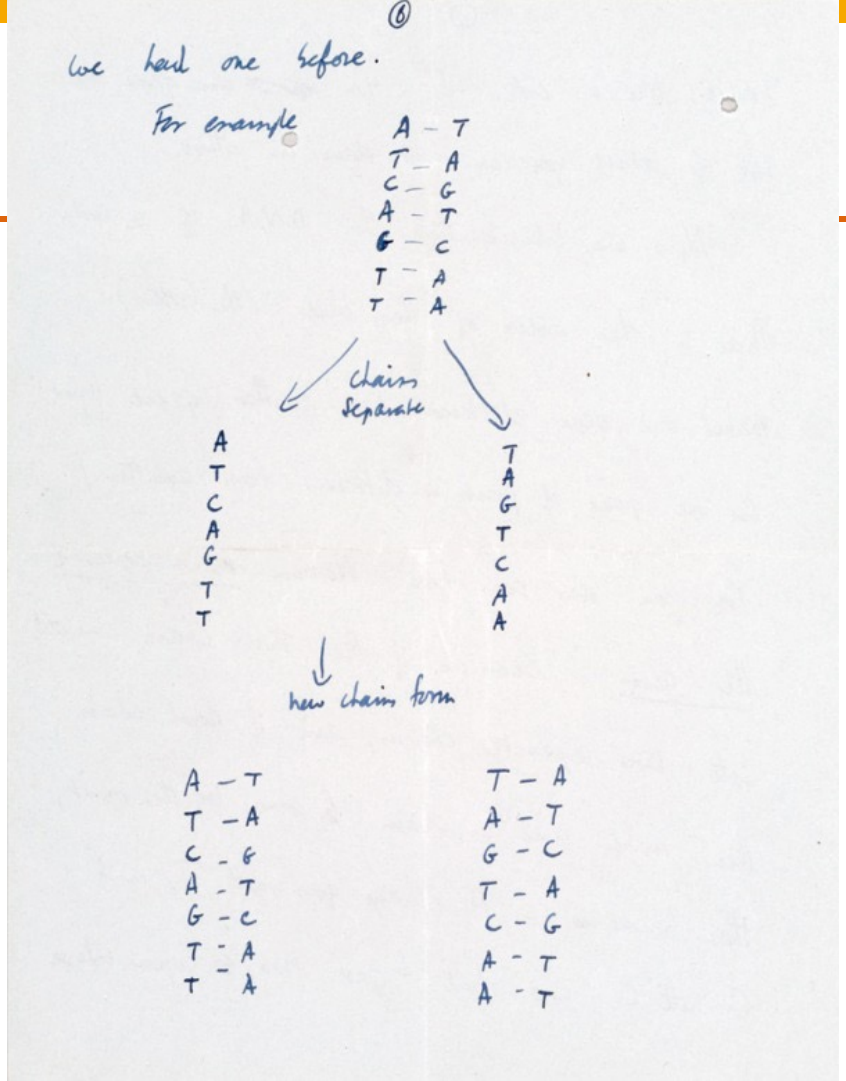
That is, the order of the bases (the letters) makes one gene different from another gene (just as one page of print is different from another).

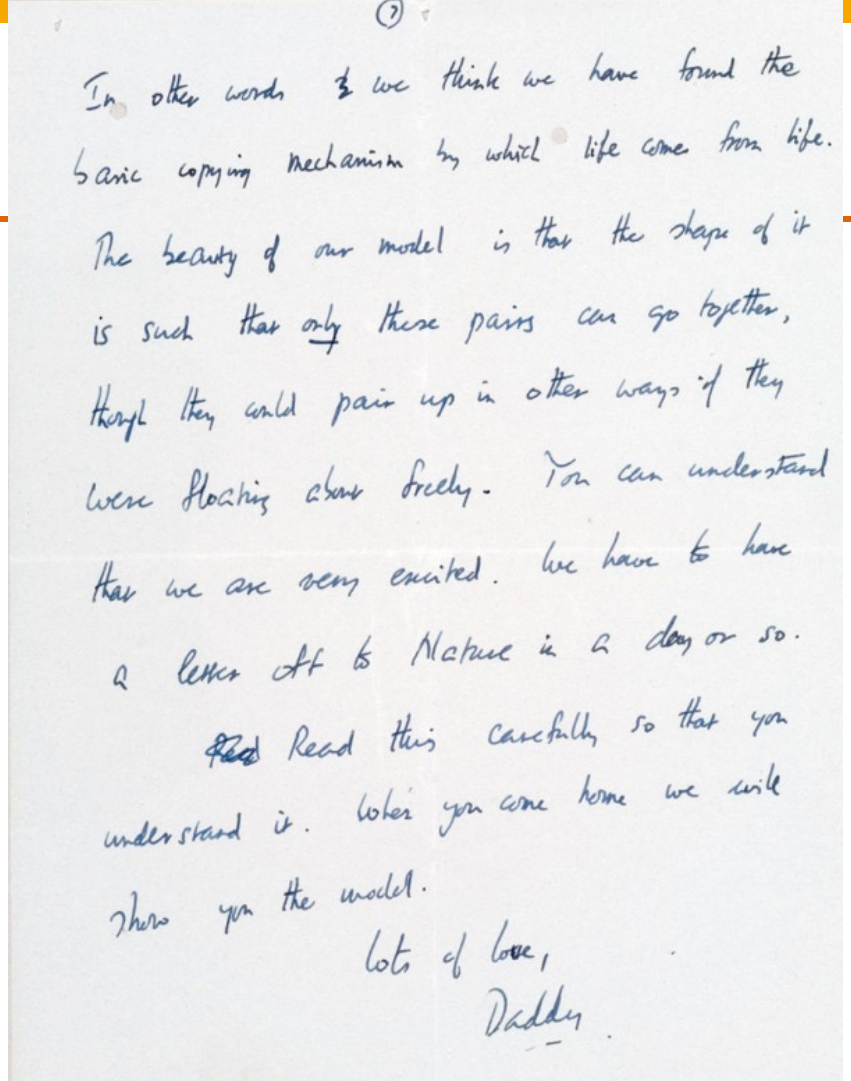
You can now see how Nature makes copies of the genes. Because if the two chain unwind into two seperate chains, and if each chain then makes another chain to come together on it, then because A always goes with T, and G with C, we shall get two copies where

Hasso Plattner Institut

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023


6
we had one before.
For example
A – T
T – A
C – G
A – T
G – C
T – A
T – A
chains separate
A
T
C
A
G
T
T
T
A
G
T
C
A
A
new chains form
A – T
T – A
C – G
A – T
G – C
T – A
T – A
T – A
A – T
G – C
T – A
C – G
A – T
A – T



(7)

In other words if we think we have found the basic copying mechanism by which life come from life.

The beauty of our model is that the shape of it is such that only these pairs can go together, though they could pair up in other ways if they were floating about freely. You can understand that we are very excited. We have to have a letter out to Nature in a day or so.

Read this carefully so that you understand it. When you come home we will show you the model.

Lots of love,

Daddy

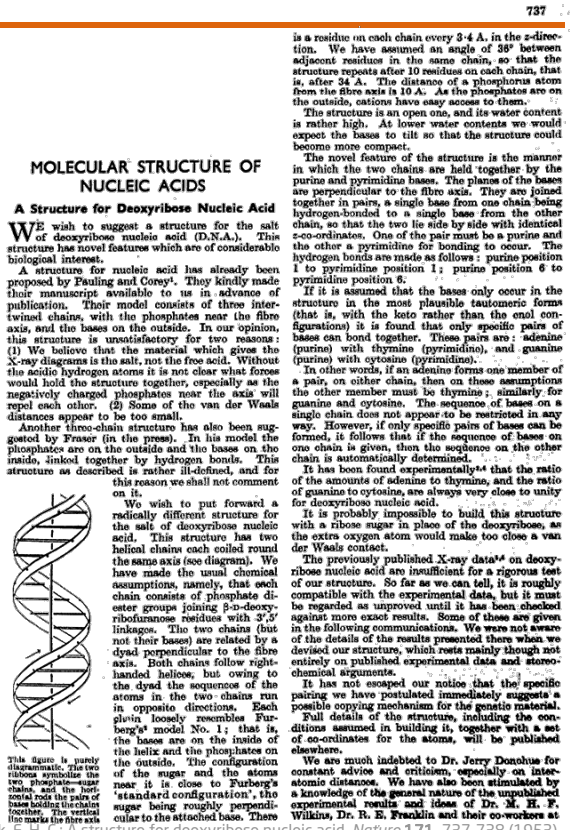
Hasso Plattner Institut

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Discovery of the Human Genome
1953 Scientific Discovery of DNA and its Structure

- *Francis Crick* (Physicist) and *James Watson* (Biologist) define model for structure of Des-oxy-ribose-nucleic-acid (DNA)
- Findings:
  - Three-dimensional, double-helix model of DNA
  - Specific base bindings, i.e. A-T and C-G (proofing Chargaff)
  - Anti-parallel structure, i.e. 5' end is bound to 3' end of complementary strand



737

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There 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

Watson, J. D., & Crick, F. H. C.: A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953)

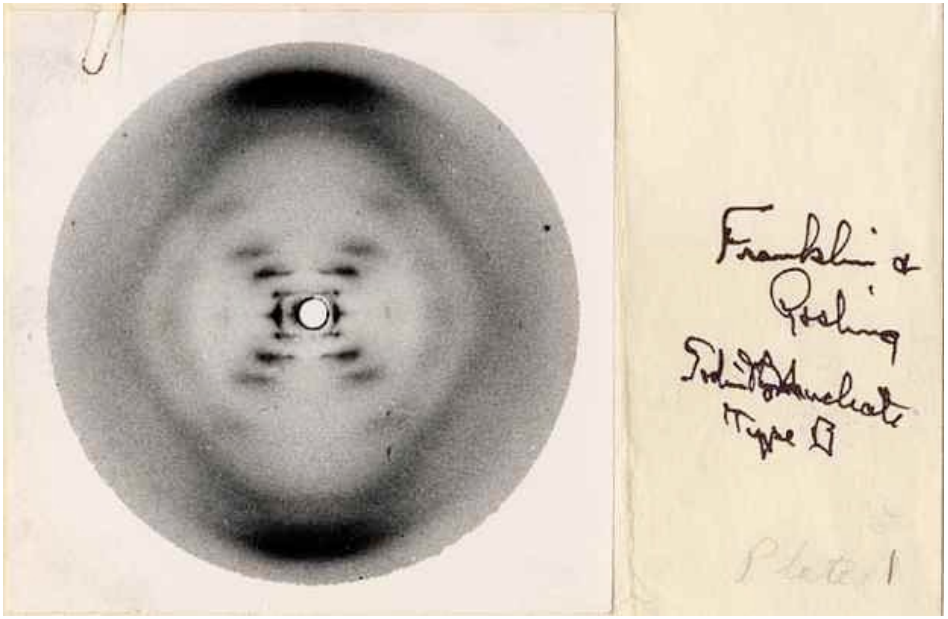
https://www.chemheritage.org/historical-profile/james-watson-francis-crick-maurice-wilkins-and-rosalind-franklin

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Discovery of the Human Genome
1953 Scientific Discovery of DNA and its Structure

- *Crick* and *Watson* incorporated X-ray crystallography work of *Rosalind Franklin* and *Maurice Wilkins* already done in 1952



Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

https://www.theguardian.com/science/2015/jun/23/sexism-in-science-did-watson-and-crick-really-steal-rosalind-franklins-data

Discovery of the Human Genome
Recent Decades

- 1977: DNA sequencing methods was designed by Francis Sanger
- 1984: Alta Summit: "DNA available on the Internet" → Idea of the global Human Genome Project (HGP)
- 1990: HGP initiated in the US, initial runtime 15 years (3 billion USD funding)
- 2000: Rough draft of the HG announced
- 2003: Human genome completely sequenced by HGP
- 2006: Sequence of the last and longest chr1 published
- 2015: U.S. Pres. Obama initiated Precision Medicine Initiative
- 2018: Many national genome projects, e.g. USA, UK, Estonia, Qatar

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Data Storage: Components of DNA and RNA

Data Storage: Components of DNA and RNA Overview

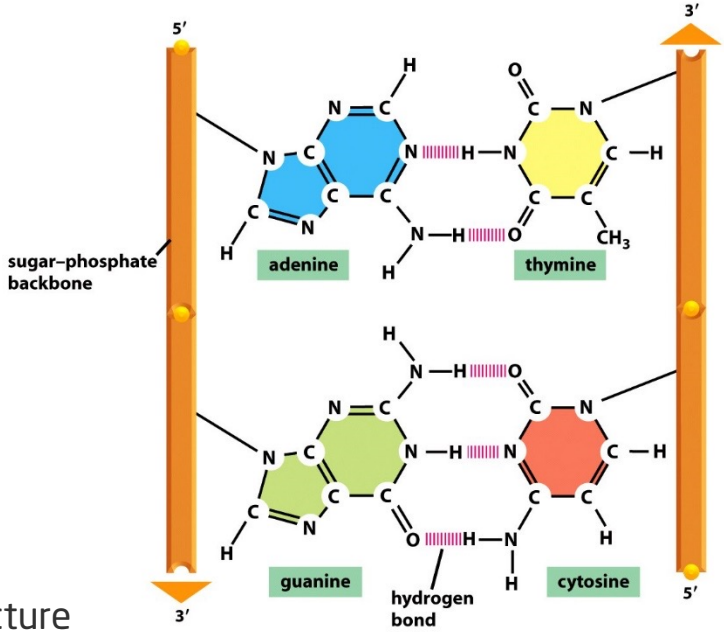
- Purpose: Understand components of DNA and its structure

- Deoxyribonucleic Acid (DNA) stores the blueprint of the cell
- Ribonucleic Acid (RNA) is used for transferring information from nucleus into the cell

Components of Deoxyribonucleic Acid (DNA)

- Each DNA strand consists of
  - Nucleobase,
    - Adenine (A),
    - Cytosine (C),
    - Guanine (G), or
    - Thymine (T)
  - Sugar: Deoxyribose, and
  - Phosphate group

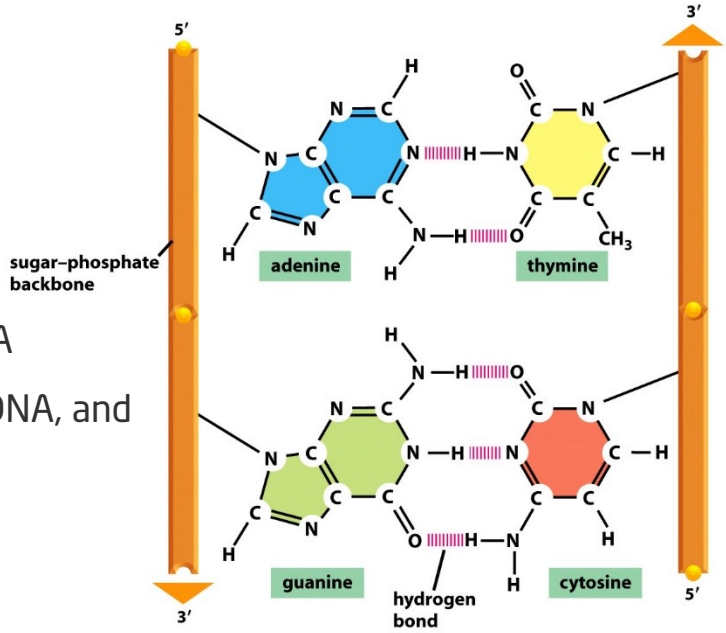
- Two strands of DNA form double-helix structure



Molecular Biology of the Cell (Garland Science 2008)

Components of Ribonucleic Acid (RNA)

- RNA consists of
  - Nucleobase,
    - Adenine (A),
    - Cytosine (C),
    - Guanine (G), or
    - Uracil (U) instead of Thymine (T) in DNA
  - Sugar: Ribose instead of deoxyribose in DNA, and
  - Phosphate group

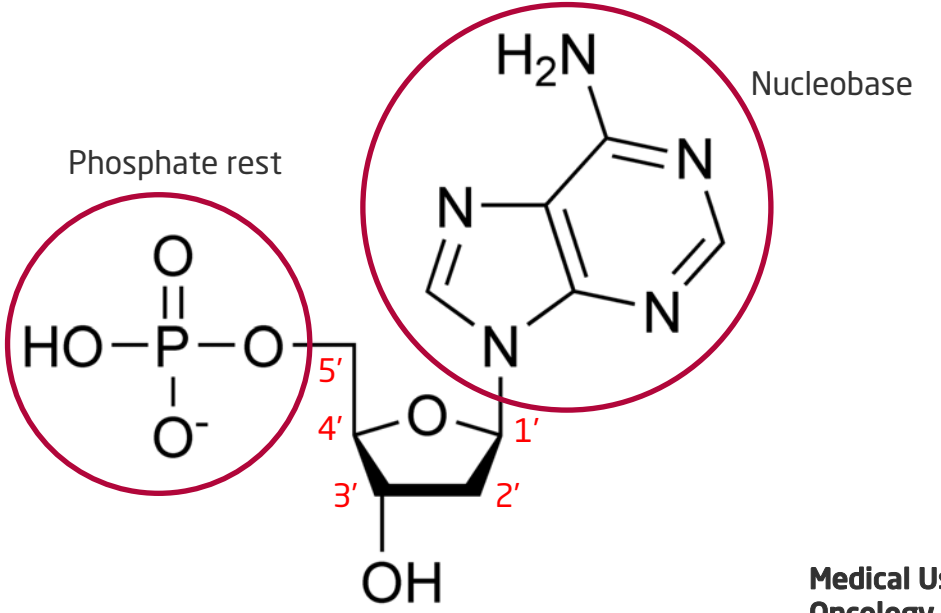


Molecular Biology of the Cell (Garland Science 2008)

Structure of Deoxyribose

- 1′ end binds nucleobase
- 3′ end contains OH hydroxyl group
- 5′ end replaced by phosphate rest

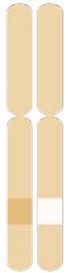
- Phosphate rest at 5′ binds to the 3′ carbon of the preceding deoxyribose



Genes

- Gene := Certain region on the DNA
  - Intron := Regions of a gene not used for RNA coding (a.k.a. non-coding)
  - Exon := Region of a gene responsible for RNA coding (a.k.a. coding)
- "Junk DNA" := Non-coding DNA regions; better: regions we do not know enough so far

- In humans:
  - Approx. 20k-25k genes
  - Length of genes range vary from few hundreds to millions of base pairs



Chromosomes
<< QUIZ >>

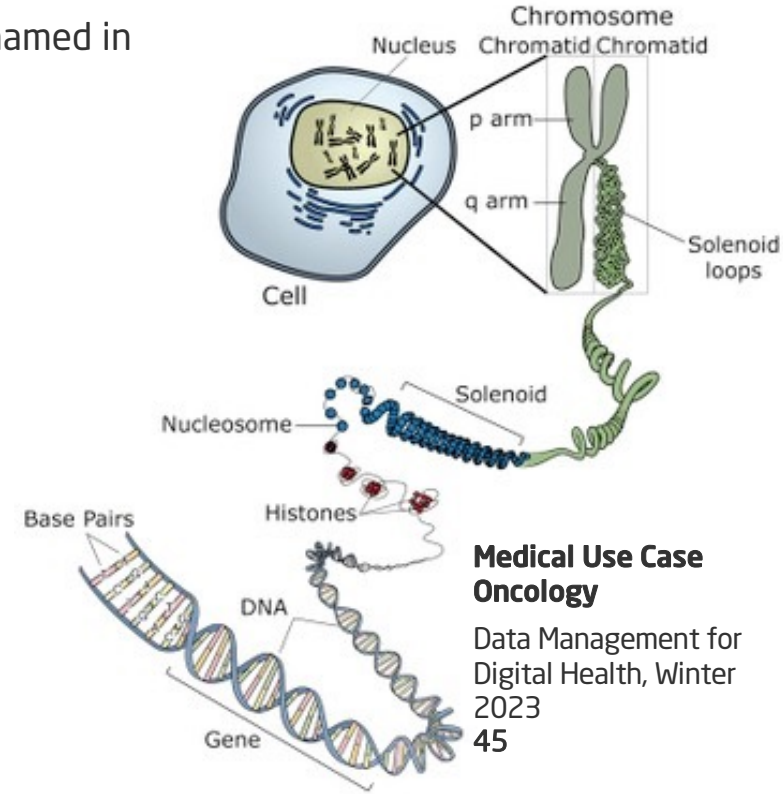
- Chromosome := DNA molecule storing parts of the genome; named in chronological order of length
- Please bring the chromosome sets of the following species in the correct numerical order (begin with the smallest).

1. Cows
2. Hedgehogs
3. Drosophila
4. Humans

Chromosomes

- Chromosome := DNA molecule storing parts of the genome; named in chronological order of length

- Examples:
  3. Drosophila: 2n = 8 / 140 Mbp
  4. Humans: 2n = 46 / 3.2 Gbp
  1. Cows: 2n = 60 / 3.0 Gbp
  2. Hedgehogs: 2n = 90 / 2.3 Gbp



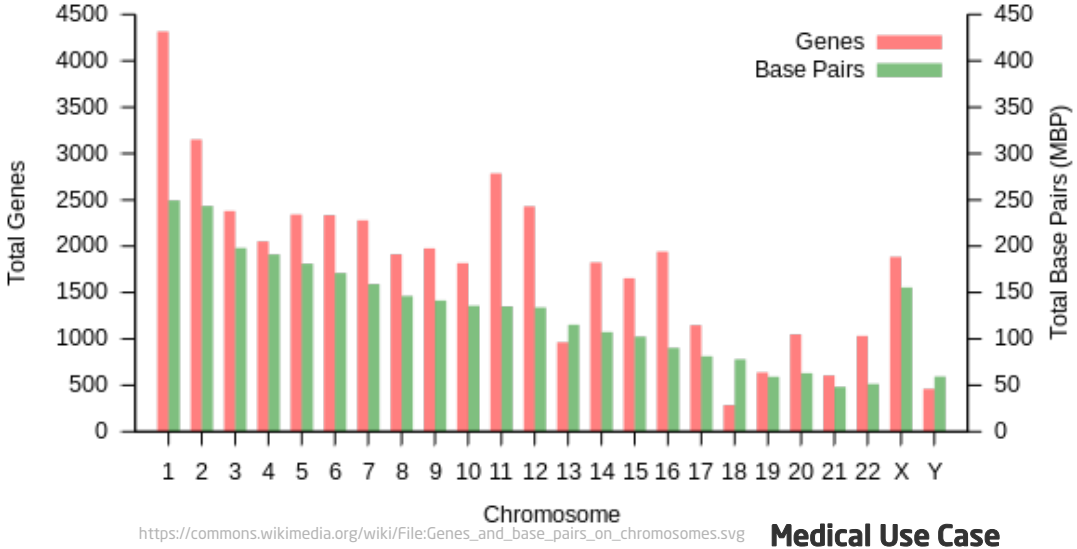
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

http://www2.le.ac.uk/projects/vgec/schoolscolleges/topics/dna-genes-chromosomes

What to take home?

- DNA consist of two strands
- 2nd DNA strand is reverse complementary to first
- Recap Chargaff: Bases bind in pairs, either A-T or C-G.

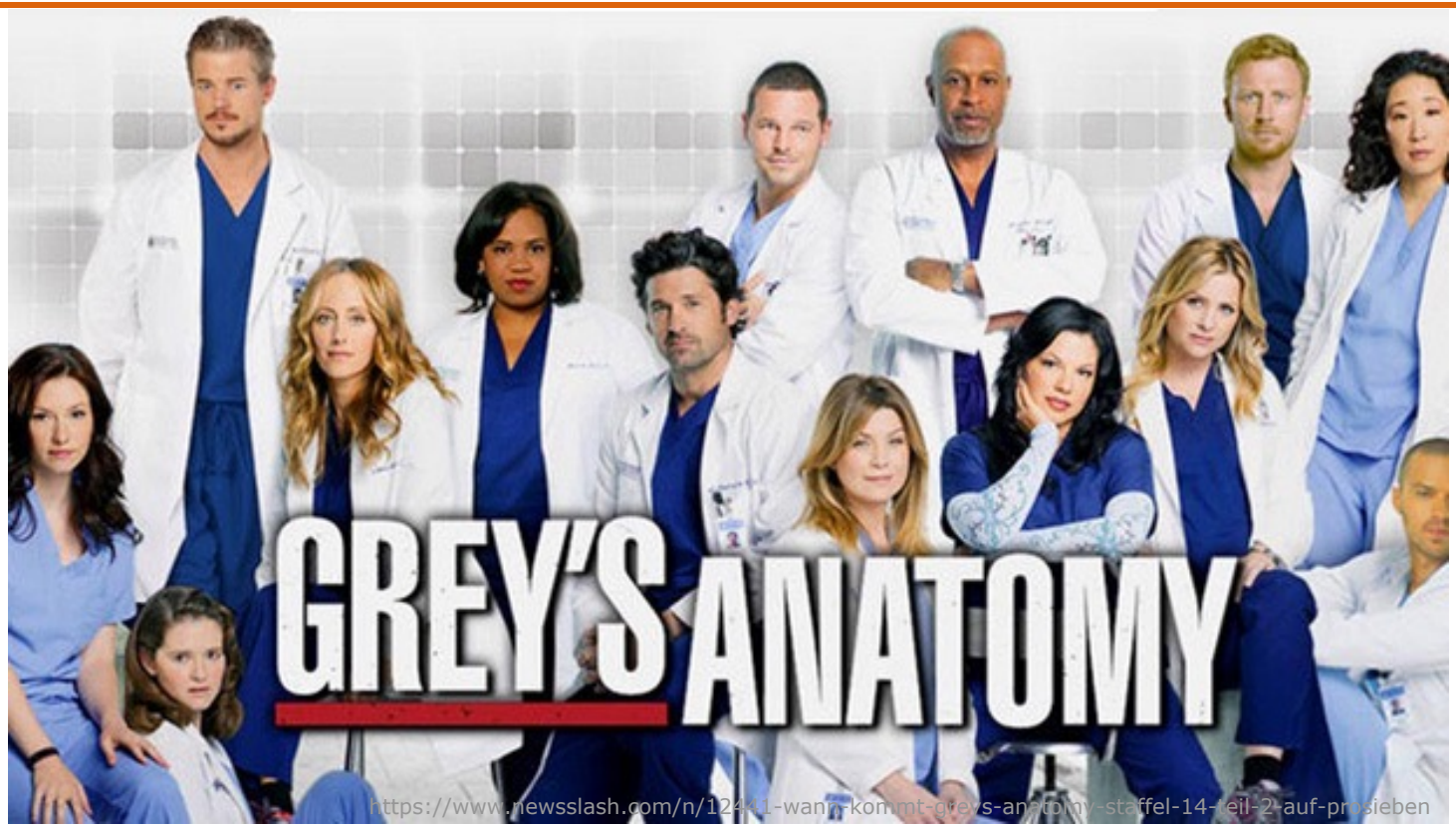


https://commons.wikimedia.org/wiki/File:Genes_and_base_pairs_on_chromosomes.svg

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

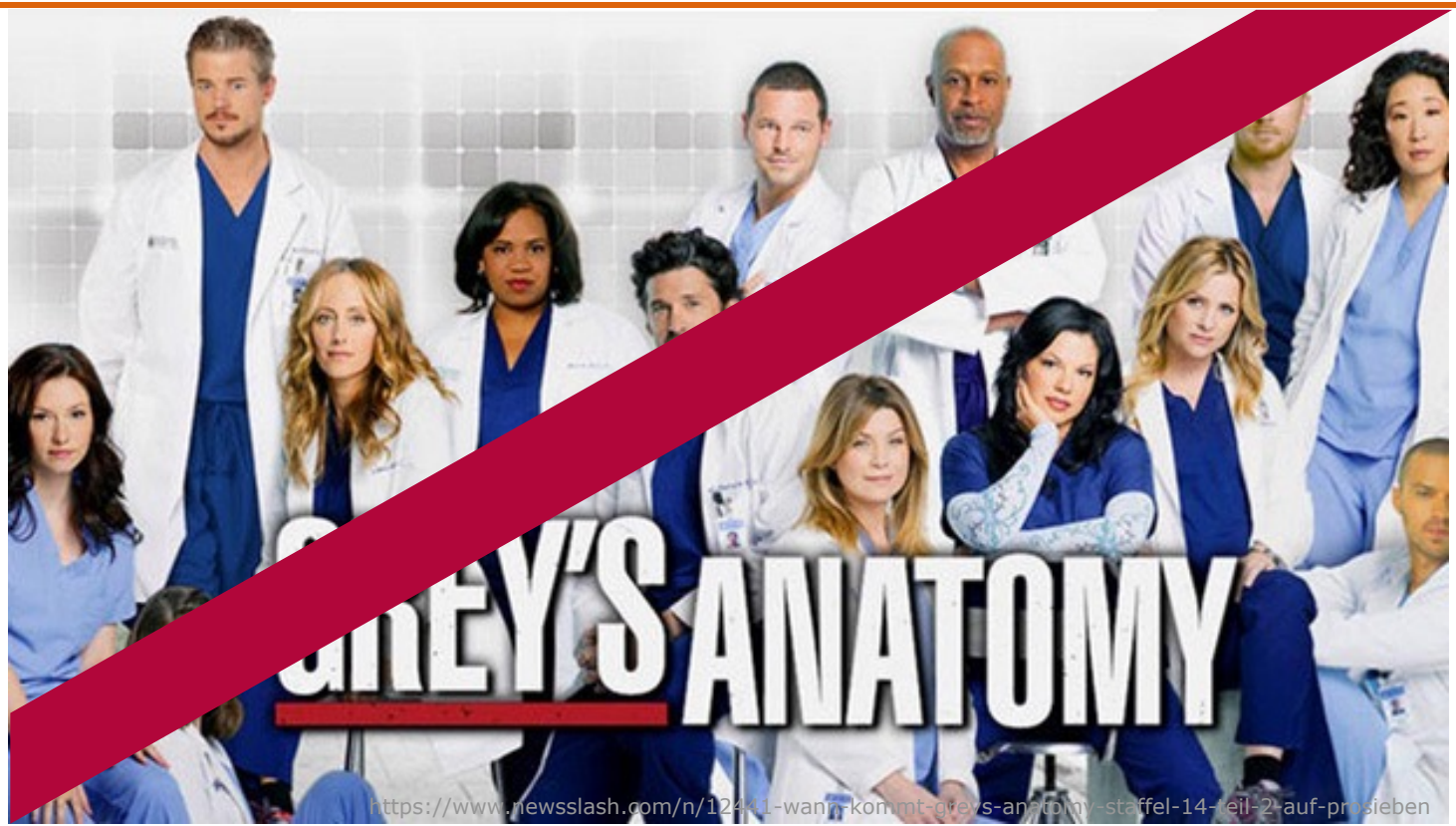
Lifecycle Management: Replication

Grey's Anatomy



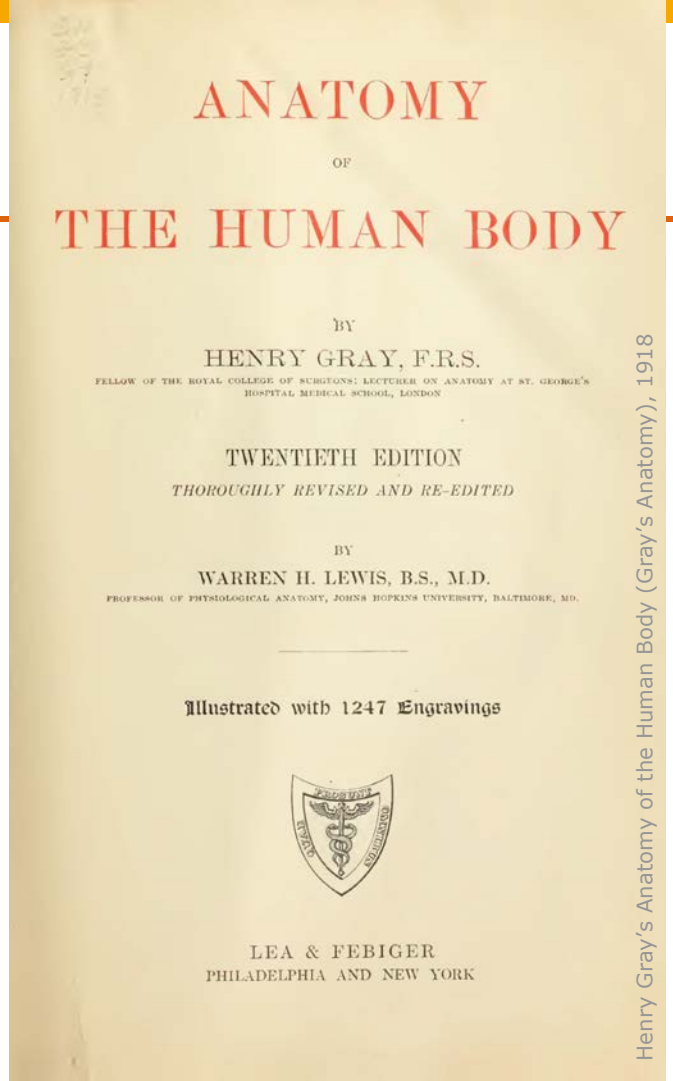
https://www.newsslash.com/n/12441-wann-kommt-greys-anatomy-staffel-14-teil-2-auf-prosieben

Grey's Anatomy



https://www.newsslash.com/n/12441-wann-kommt-greys-anatomy-staffel-14-teil-2-auf-prosieben

Gray's Anatomy 1918 ed.



ANATOMY

OF

THE HUMAN BODY

BY

HENRY GRAY, F.R.S.

FELLOW OF THE ROYAL COLLEGE OF SURGEONS; LECTURER ON ANATOMY AT ST. GEORGE'S HOSPITAL MEDICAL SCHOOL, LONDON

TWENTIETH EDITION

THOROUGHLY REVISED AND RE-EDITED

BY

WARREN H. LEWIS, B.S., M.D.

PROFESSOR OF PHYSIOLOGICAL ANATOMY, JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD.

Illustrated with 1247 Engravings

LEA & FEBIGER

PHILADELPHIA AND NEW YORK

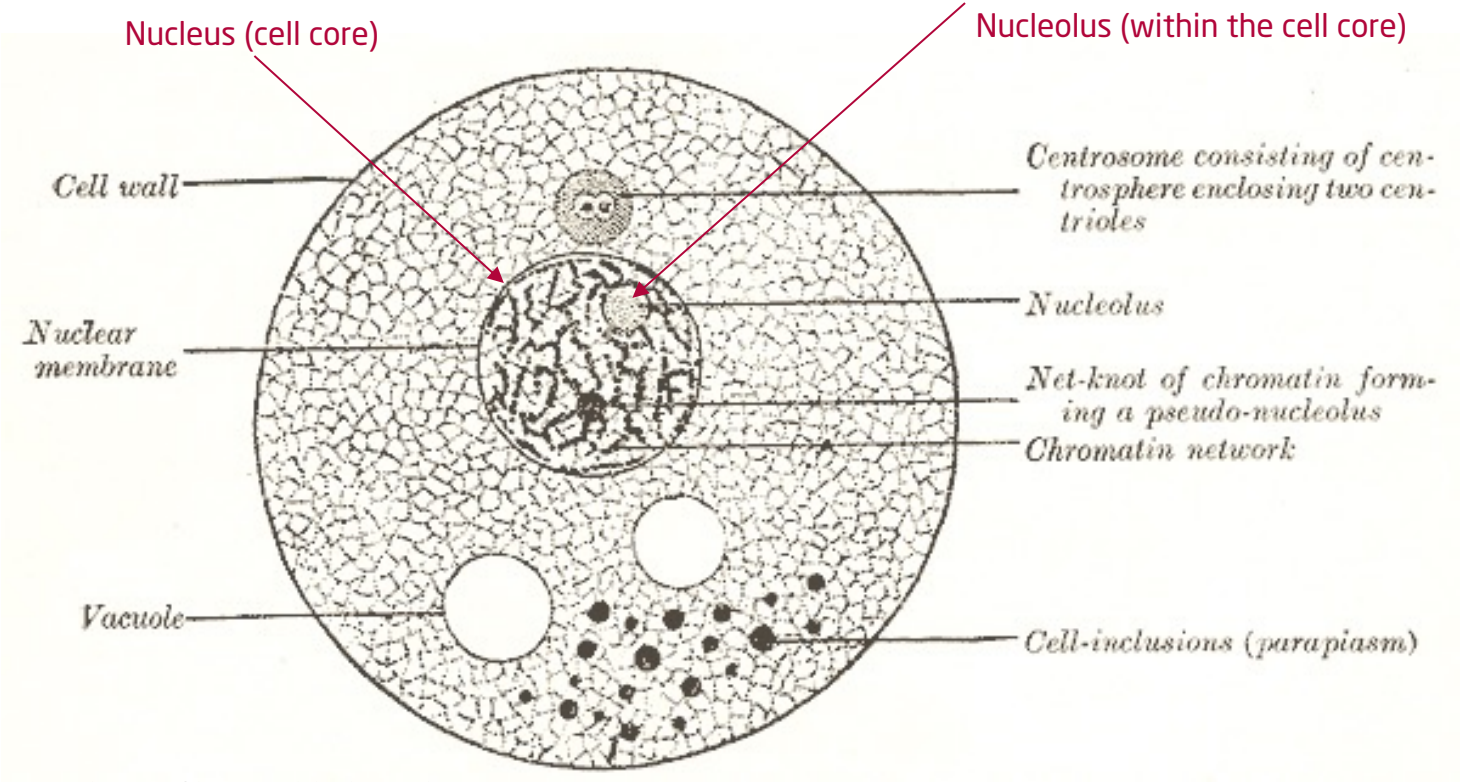
Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

Hasso Plattner Institut

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

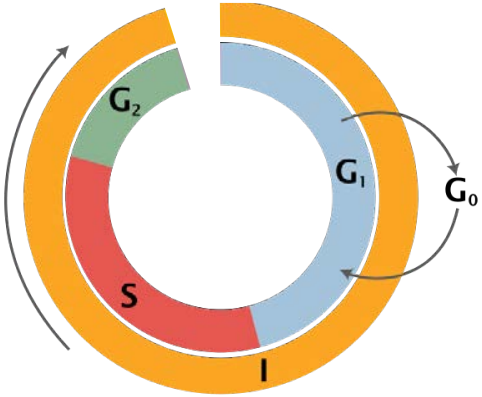
Gray's Anatomy
Recap: Cell Organelles



Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

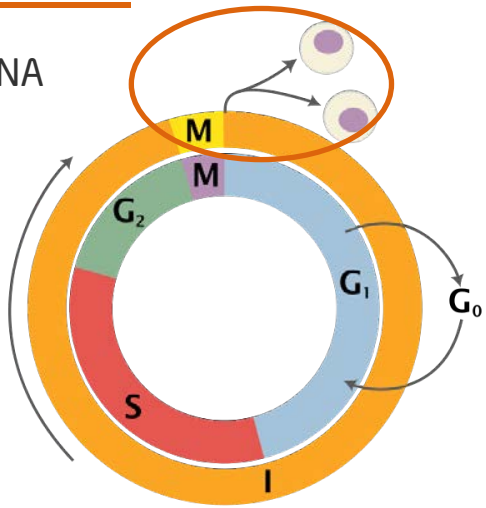
Cell Cycle
Interphase

- Interphase (I) consists of:
  - Gap 1 (G1), i.e. cell growth creating cell organelles
  - Gap 0 (G0) / Resting, i.e. cell stops division temporarily or forever
  - Synthesis (S) of DNA through replication of chromatids within the cell core
  - Gap 2 (G2), i.e. producing proteins for upcoming mitosis



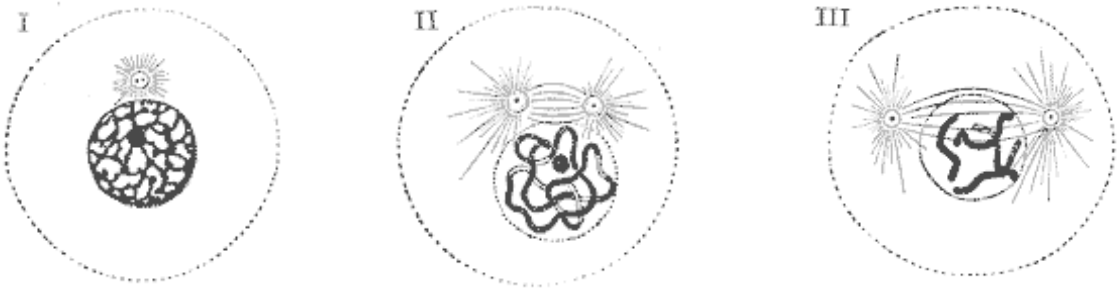
Cell Cycle
Mitosis

- Mitosis (M) := process of cell division into two daughter cells carrying identical DNA
  - Prophase
  - Metaphase
  - Anaphase
  - Telophase



Mitosis
Prophase

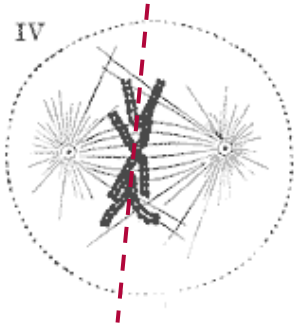
- Chromatin condenses into chromosomes
- Nucleolus, i.e. a part of the nucleus where ribosomes are made, disappears
- Spindle apparatus move to individual poles of the cell



Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

Mitosis
Metaphase

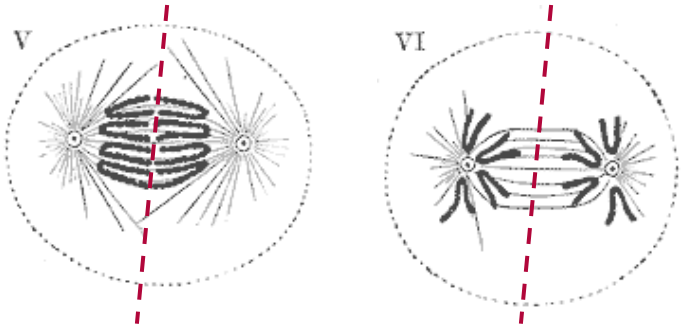
- Chromosomes line up along equatorial plane



Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

Mitosis
Anaphase

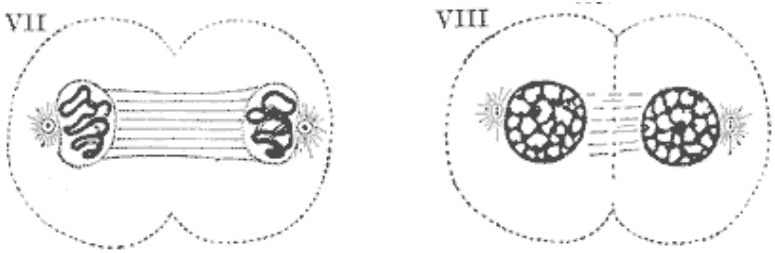
- Chromosomes break up at equatorial plane into individual chromatids
- Chromatids move to individual poles



Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

Mitosis
Telophase

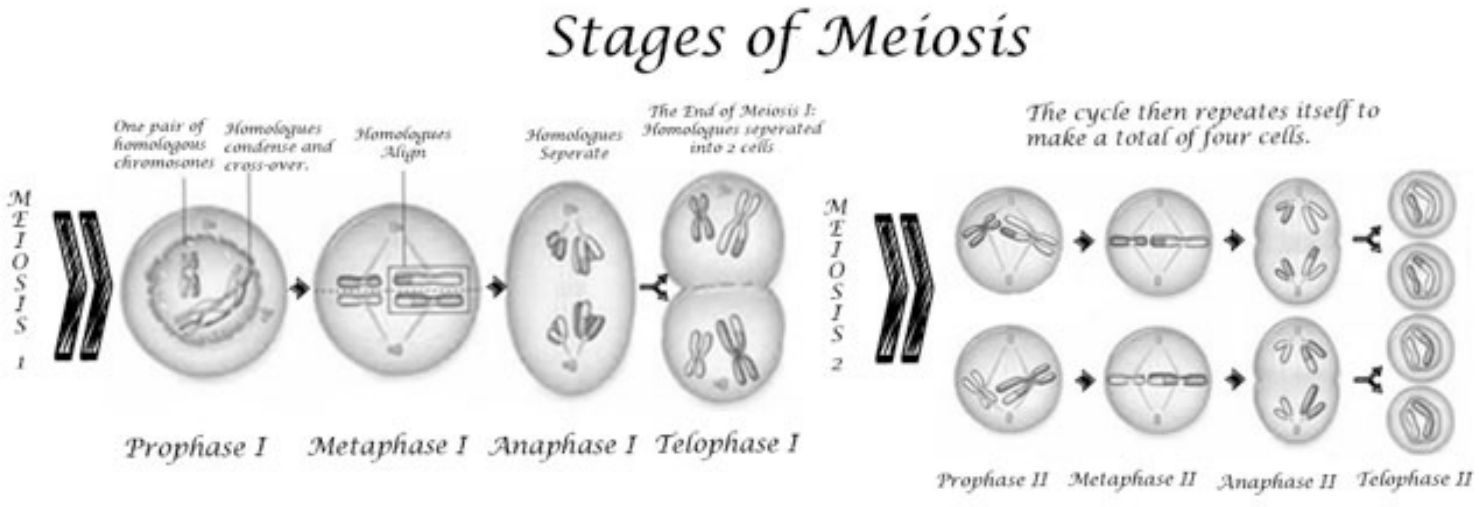
- Individual cell membranes form
- Nucleoli reappear
- Chromosomes unwind into more stable chromatin within nucleolus



Henry Gray's Anatomy of the Human Body (Gray's Anatomy), 1918

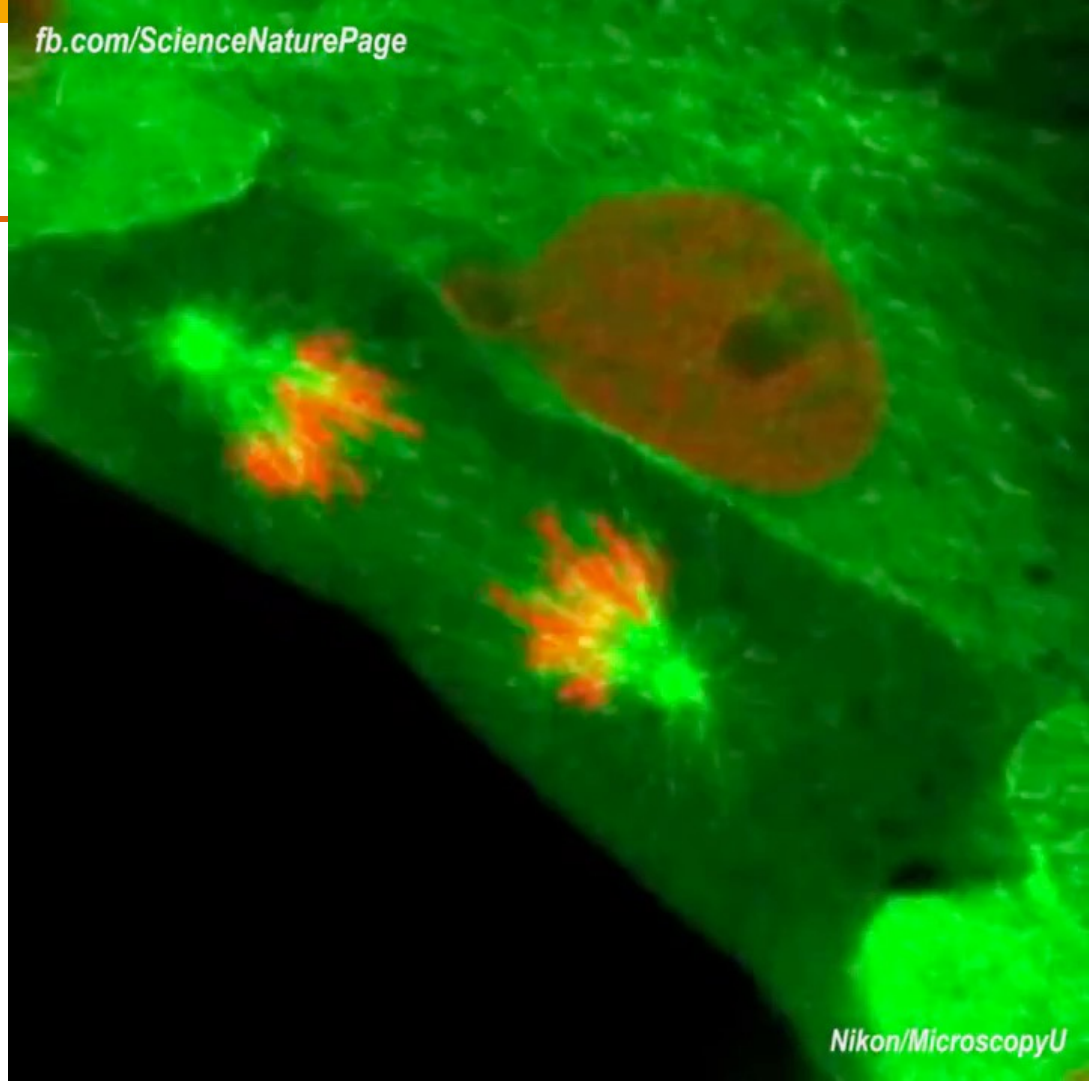
Meiosis

- Meiosis := two-level cell division (diploid) into four individual gametes carrying unique haploid DNA material
- Pro-, Meta-, Ana-, and Telophase are performed twice

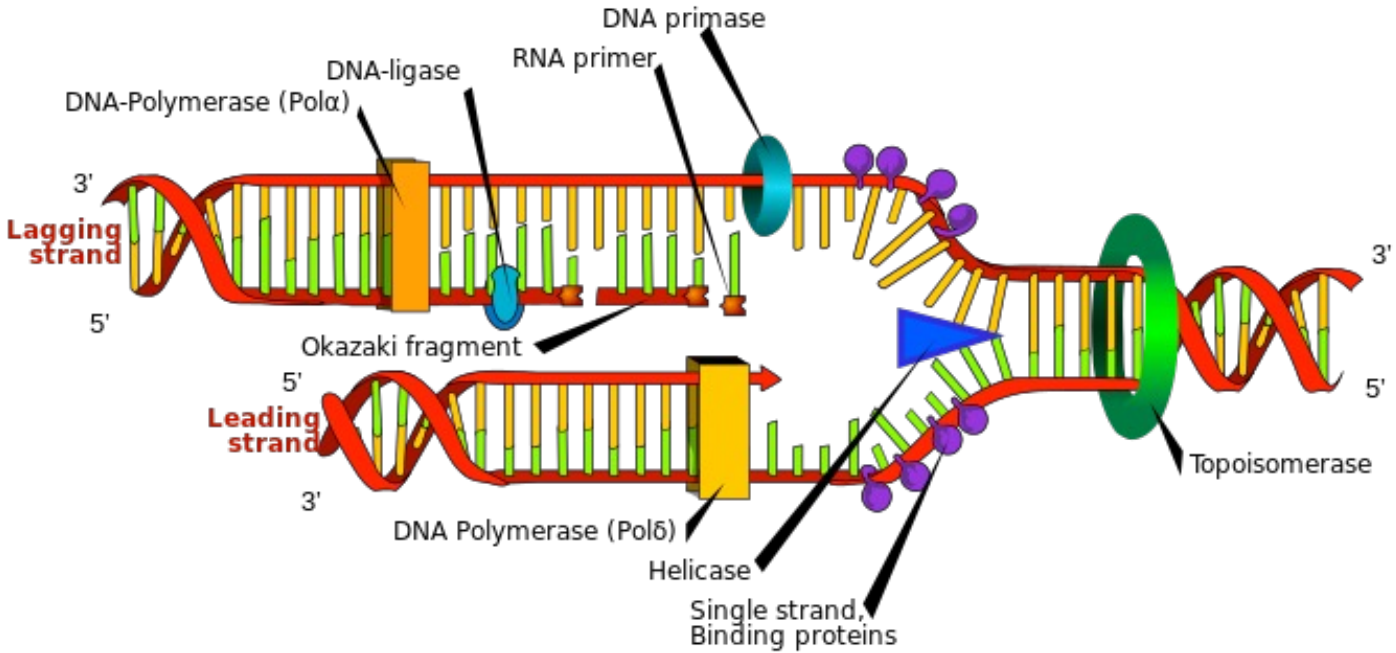


http://www.dbriers.com/tutorials/2012/11/comparing-mitosis-vs-meiosis/

Mitosis live (time lapse)

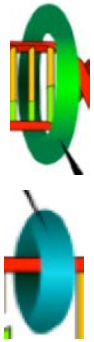


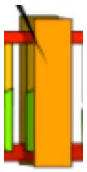
DNA Replication during Synthesis Phase

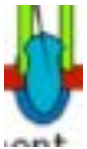


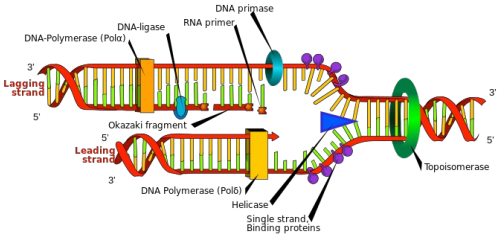
DNA Replication during Synthesis Phase

1. Initiation
   - Topoisomerase helps to prepare unwinding of DNA
   - Helicase unzips DNA at specific origins
   - Primase adds primer for binding of polymerase
2. Elongation
   - DNA polymerase
     - Extends DNA only in 5' → 3' direction using a template strand
     - Performs proofreading of replicated strand
   - DNA ligase seals strand breaks
3. Termination: Replication comes to an end

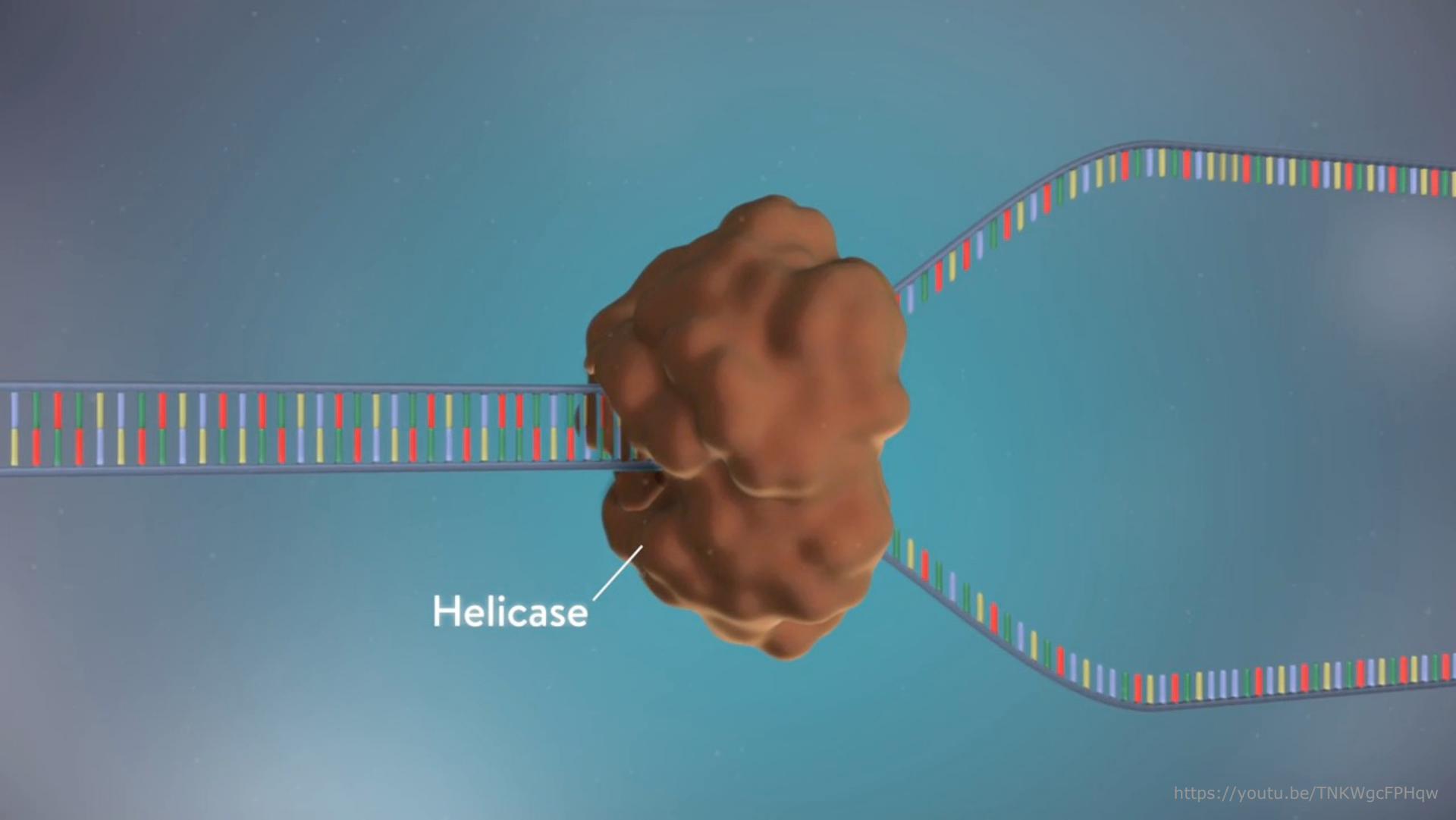








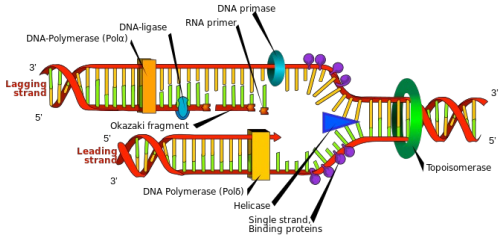
Medical Use Case Oncology



https://youtu.be/TNKWgcFPHqw

What to Take Home?

- Mitosis results in two daughter cells carrying identical DNA
- Meiosis is a two-level cell division of one diploid cell into 4 unique haploid gametes
- DNA polymerase performs proofreading of replicated strand
- Throughput of DNA polymerase:
  - Eukaryotes: Approx. 50-100 nucleotides / second
  - Prokaryotes: Approx. 1,000 nucleotides / second
- DNA replication is performed in parallel at different locations
  → DNA unwound only for a very short time



Recap the Recap
<< Open Mic Session >>

- Please name five cell organelles and their functions (speak up!).

Recap the Recap
<< Open Mic Session >>

- Please name five cell organelles and their functions (speak up!).

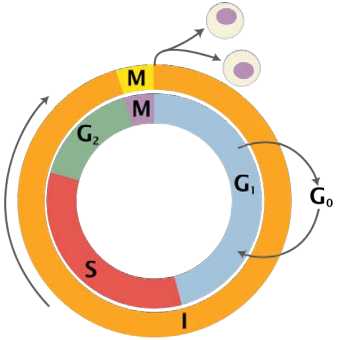
- Please name three properties of the DNA.

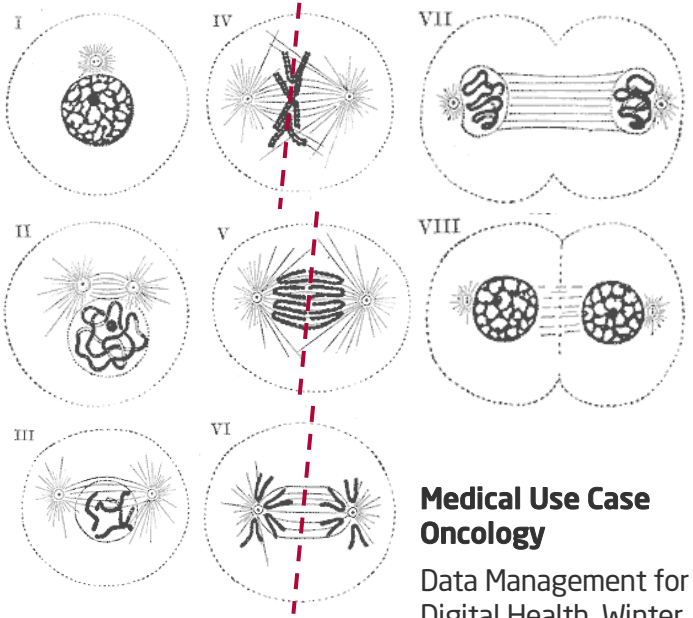
Recap the Recap
<< Open Mic Session >>

- Please name five cell organelles and their functions (speak up!).

- Please name three properties of the DNA.

- What is cell division and how does it work?
  - G1: 3-12h
  - S: 8-12h
  - G2: 1.5-3h
  - M: 0.5-1h





Medical Use Case Oncology

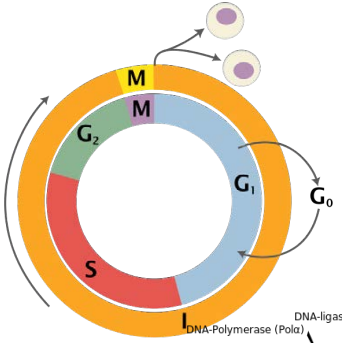
Data Management for Digital Health, Winter 2023

Recap the Recap
<< Open Mic Session >>

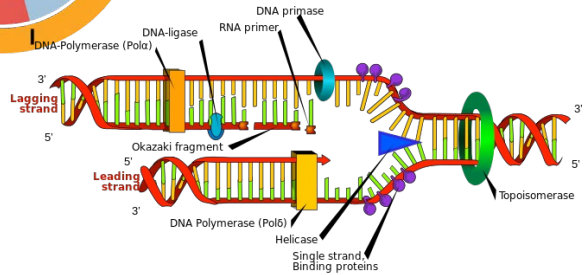
- Please name five cell organelles and their functions (speak up!).

- Please name three properties of the DNA.

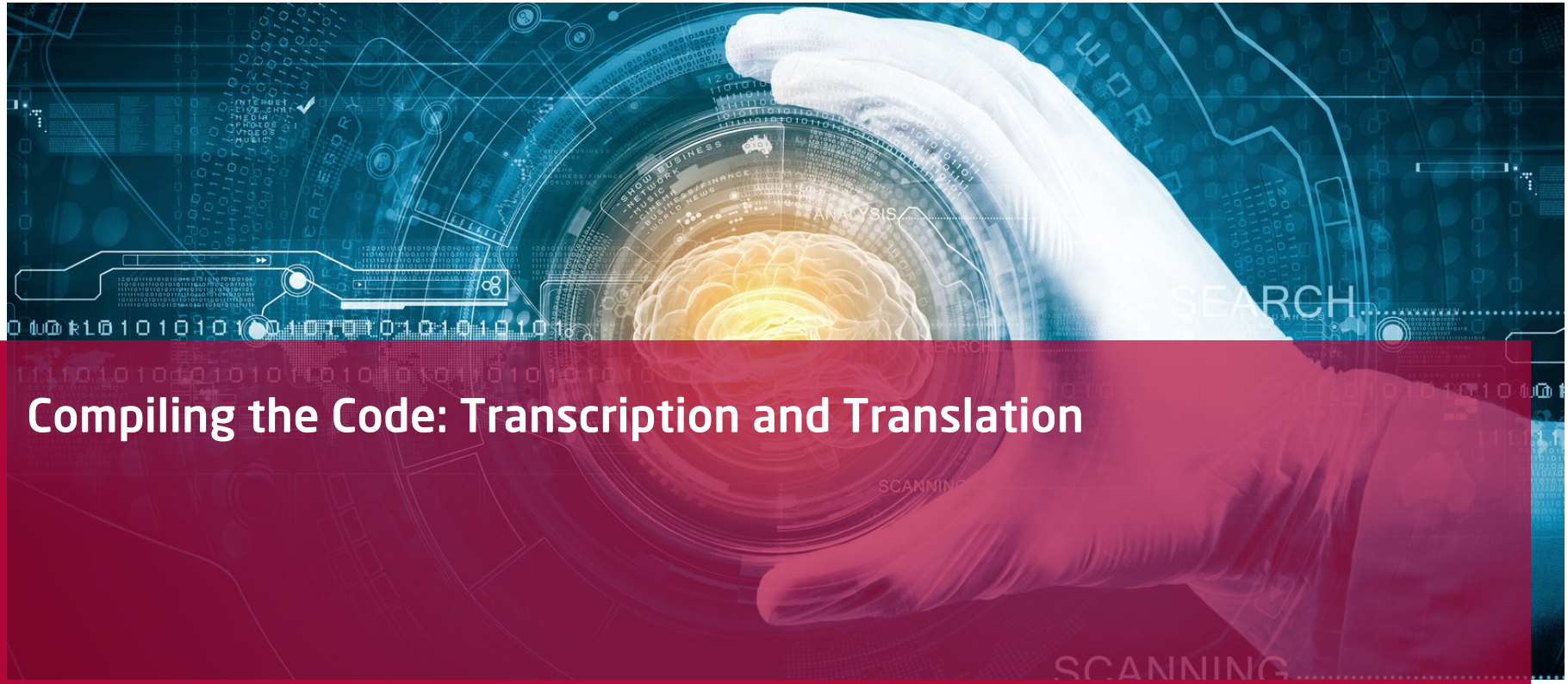
- What is cell division and how does it work?



- How does DNA synthesis work?



Hasso Plattner Institut
Digital Engineering · Universität Potsdam



Compiling the Code: Transcription and Translation

Transcription and Translation
<< QUIZ >>

- What are products produced by the ribosomes?

A. Water

B. Proteins, i.e. sequences of amino acids

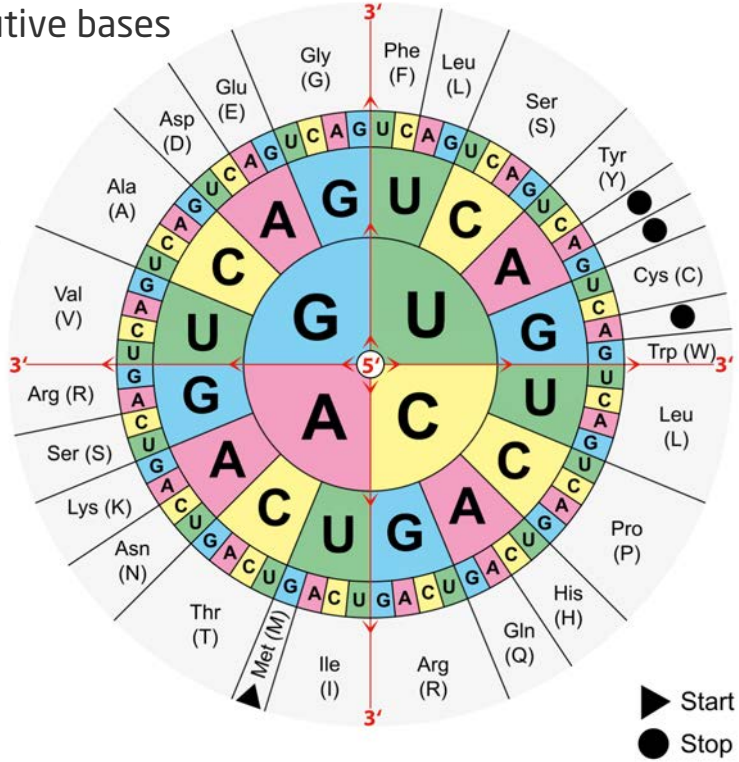
C. Organic yeast

D. Energy

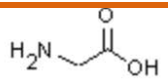
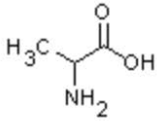
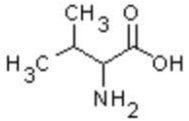
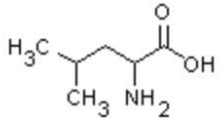
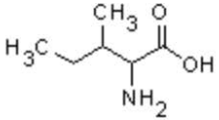
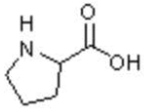
Amino Acid Coding Sun

- Codon := Triplet of bases, i.e. three consecutive bases
- Amino Acid (AA) := encoded by codons
- 20 canonical AAs in humans
- → Redundancy within the genetic code, i.e. multiple codons form the same AA
- Nine of the AAs cannot be synthesized → essential to be consumed in the diet, i.e.

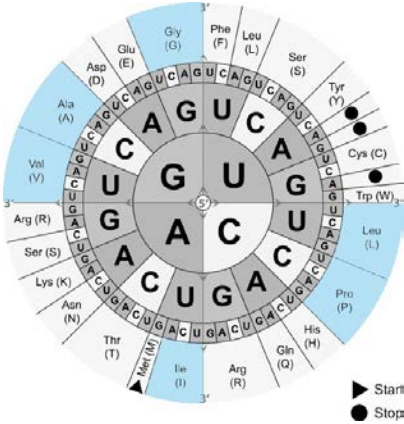
  Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Theronine, Tryptophan, and Valine.



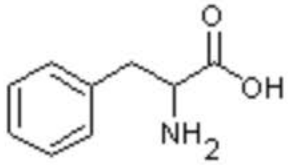
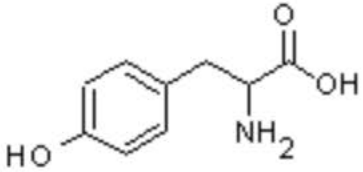
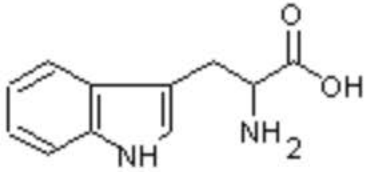
Amino Acids
Non-polar, aliphatic residues

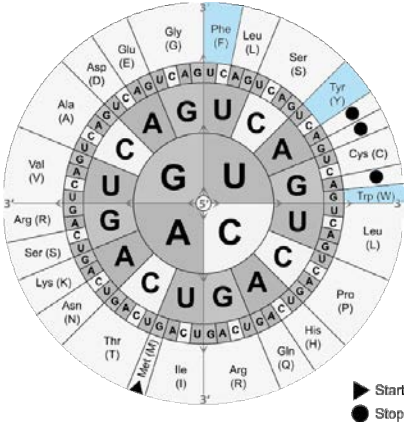
Glycine	Gly	G		GGU GGC GGA GGG
Alanine	Ala	A		GCU GCC GCA GCG
Valine	Val	V		GUU GUC GUA GUG
Leucine	Leu	L		UUA UUG CUU CUC CUA CUG
Isoleucine	Ile	I		AUU AUC AUA
Proline	Pro	P		CCU CCC CCA CCG

http://www.fr33.net/aminoacids.php



Amino Acids
Aromatic residues

Phenylalanine	Phe	F		UUU UUC
Tyrosine	Tyr	Y		UAU UAC
Tryptophan	Trp	W		UGG

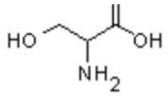
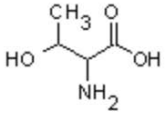
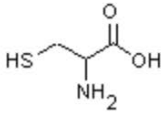
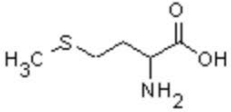
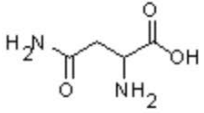
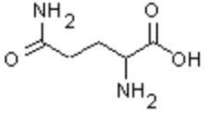


Medical Use Case Oncology

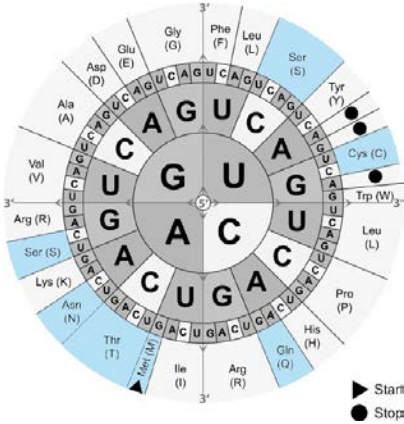
Data Management for Digital Health, Winter 2023

http://www.fr33.net/aminoacids.php

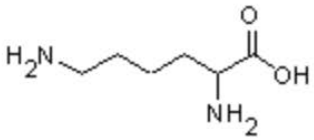
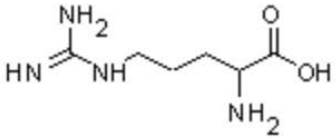
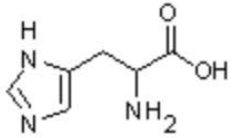
Amino Acids
Polar, non-charged residues

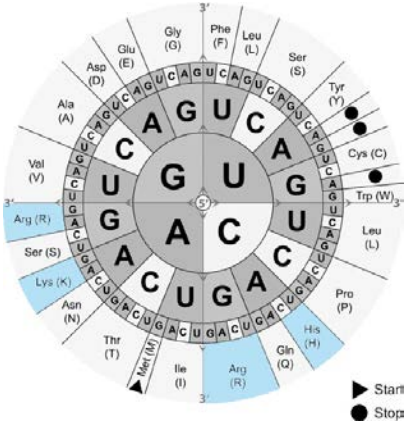
Serine	Ser	S		UCU UCC UCA UCG AGU AGC
Threonine	Thr	T		ACU ACC ACA ACG
Cysteine	Cys	C		UGU UGC
Methionine	Met	M		AUG
Asparagine	Asn	N		AAU AAC
Glutamine	Gln	Q		CAA CAG

http://www.fr33.net/aminoacids.php



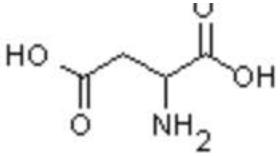
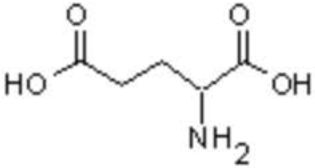
Amino Acids
Positively charged residues

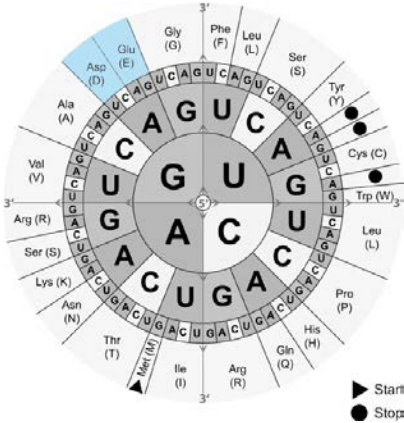
Lysine	Lys	K		AAA AAG
Arginine	Arg	R		CGU CGC CGA CGG AGA AGG
Histidine	His	H		CAU CAC



http://www.fr33.net/aminoacids.php

Amino Acids
Negatively charged residues

Aspartate	Asp	D		GAU GAC
Glutamate	Glu	E		GAA GAG



Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

http://www.fr33.net/aminoacids.php

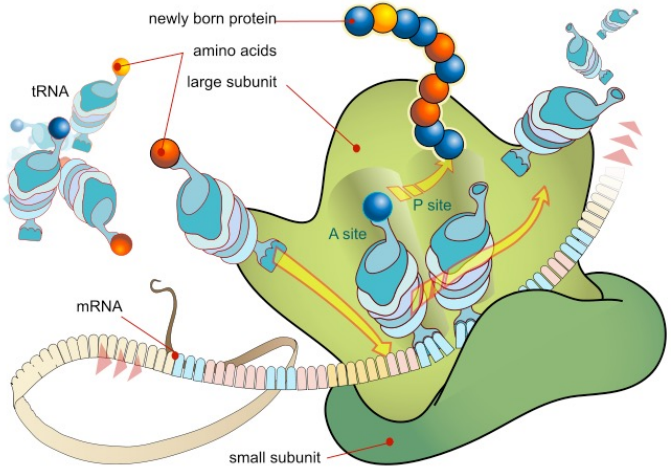
Transcription

- Transcription := Process of copying a segment of DNA into RNA to transport it from the nucleus into the cytoplasm, i.e. DNA (A,T,C,G) → RNA (A,U,C,G)
- Types of RNA:
  - Messenger RNA (mRNA): Exports a segment of a gene (code) from the cell core for processing (compiling) by ribosomes
  - Ribosomal RNA (rRNA): Source code for building ribosomes (compiler)
  - Transfer RNA (tRNA): Binds a specific AA from the cytoplasma for ribosomes

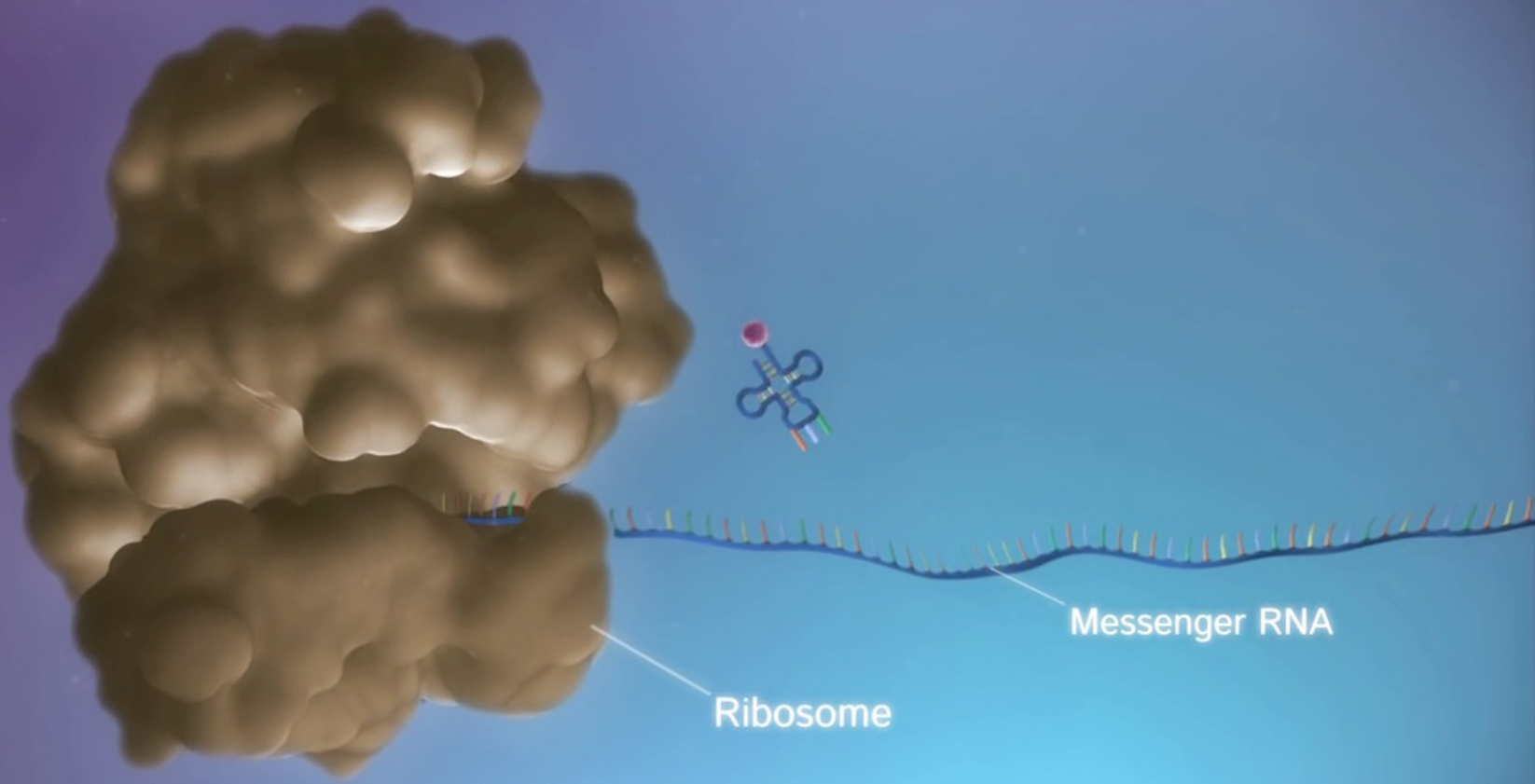
Translation

- Translation := Process of protein synthesis performed by ribosomes following a given template, i.e. RNA → AA sequence

1. Initiation: Detect start codon
2. Elongation: Bind amino acids defined by next codon, ribosome moves on
3. Termination: When a stop codon is detected, the ribosomes finishes its work

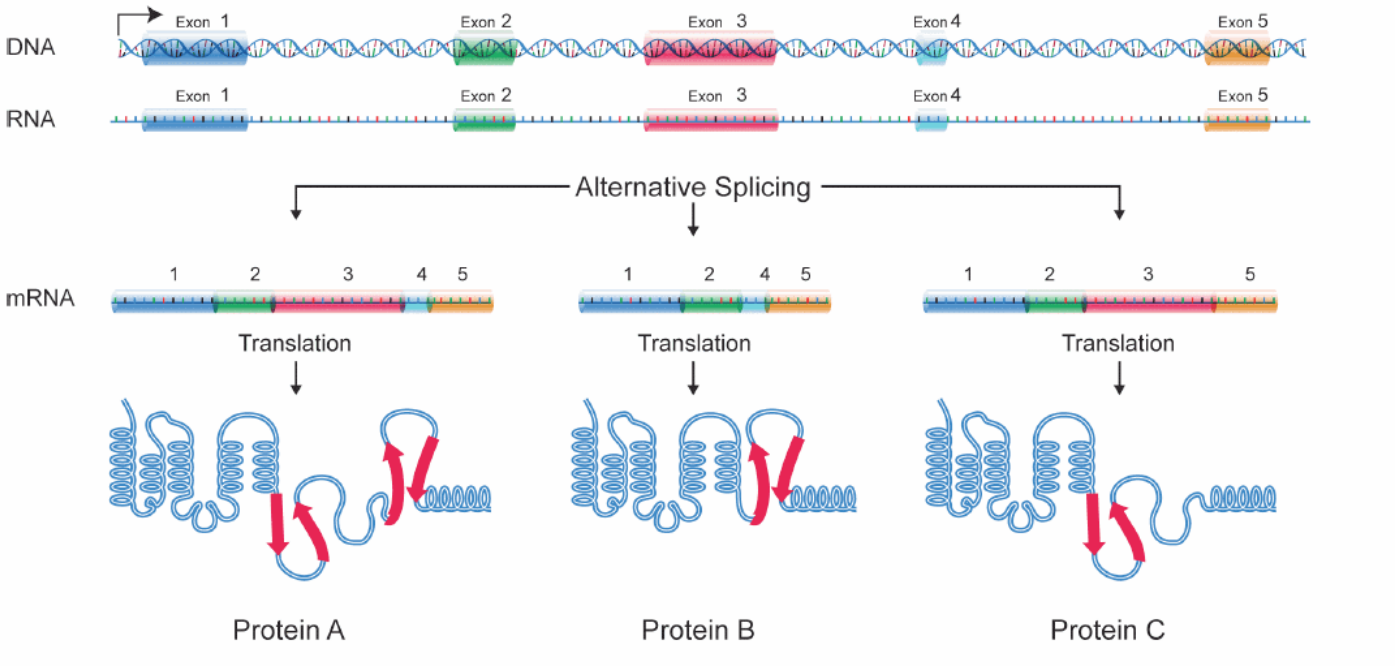


Mariana Ruiz Villarreal



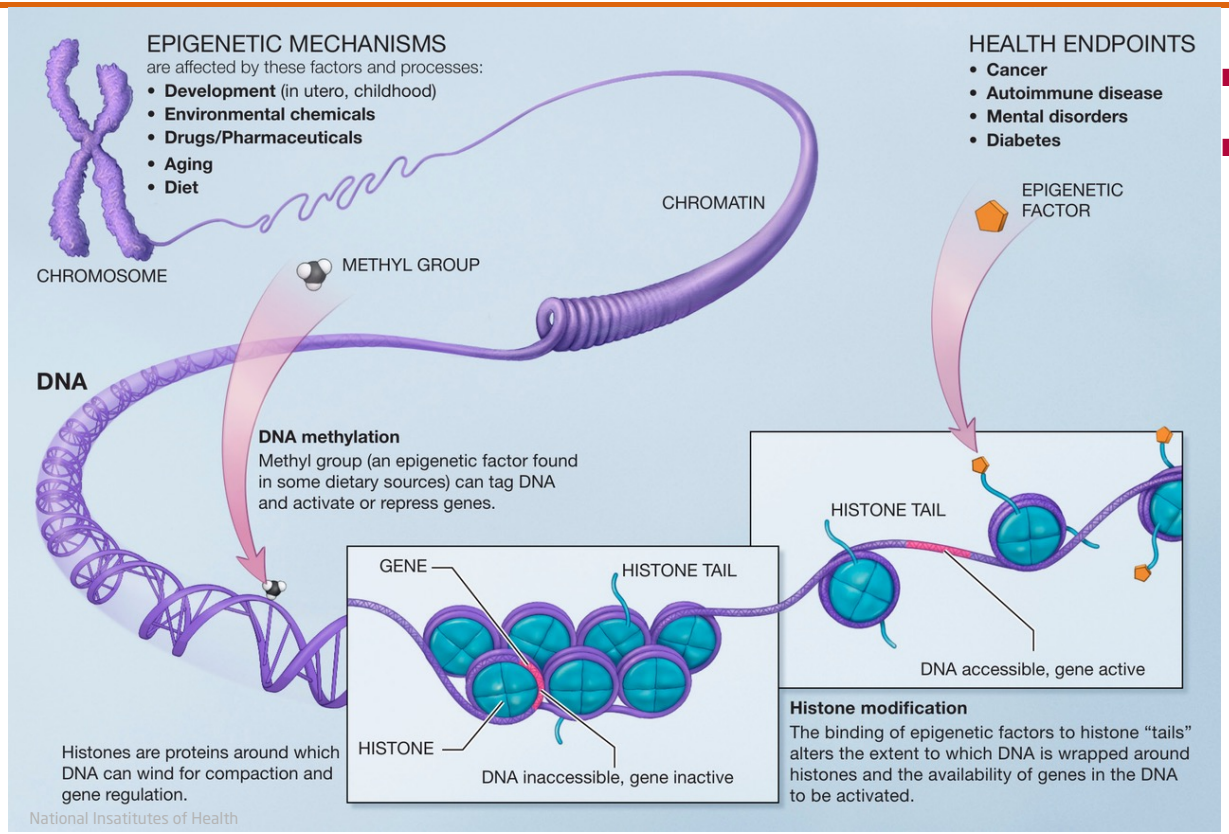
Increasing Variability through Alternative Splicing

- How to build different products out of the same DNA?



National Human Genome Research Institute

Epigentics: Additional Factors Impacting Gene Regulation



- DNA methylation
- Histone acetylation

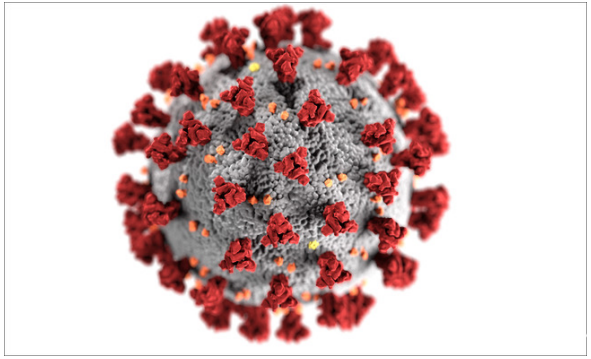
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Excursus: Replication of Viruses

- DNA virus need to be transcribed into viral mRNA before replication
- RNA virus are directly replicated by ribosomes of the host, e.g.
  - Severe Acute Respiratory Syndrome (SARS) viruses also SARS-CoV-2,
  - Influenza virus, and
  - Hep. C virus.

- Coronavirus:
  - RNA-based viruses surrounded by a hull
  - 30kbps single-stranded RNA genome with positive polarity, i.e. longest genome of all known RNA viruses



Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Bugs: Genetic Changes

Genetic Variants and Mutations
<< QUIZ >>

- What are properties of mutations?

A. Spontaneously

B. Undirected

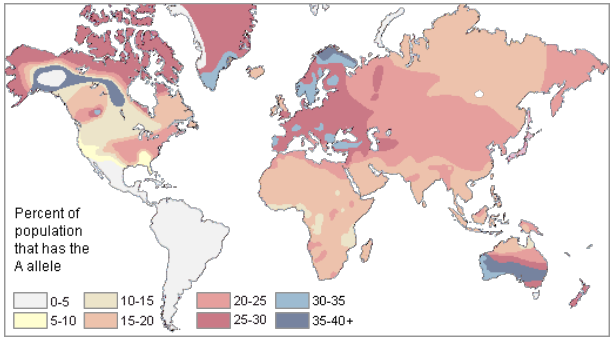
C. Triggered by mutagens, e.g. radiation or chemicals

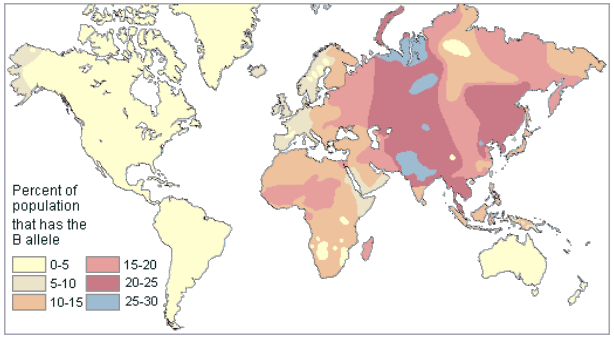
D. All of the above

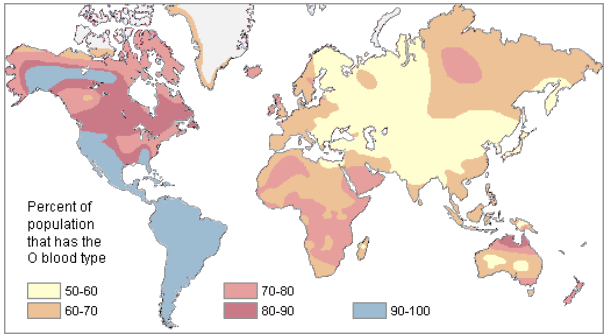
Genetic Variants and Mutations
Example: Individual blood groups

- Single Nucleotide Polymorphism (SNP) on the DNA strand
- Distribution of blood group by

http://anthro.palomar.edu/vary/vary_3.htm







A or AB

B or AB

0

Genetic Variants and Mutations

- Genetic variant := Polymorphism within the genetic code
- Mutation := Variants with measurable impact occurring spontaneously and undirected
- Mutagen := Component that may trigger changes in the genetic code, e.g. radiation, chemicals, temperature, pressure, etc.

Mutations

Affected Cell Types	Inheritable	Affects
Gametic	Yes	Offspring
Somatic	No	Current individual only

- Where mutation can occur, e.g.:
  - Gene, i.e. within a specific range on a chromosome
  - Chromosome, i.e. the structure of the chromosome is affected
  - Genome, i.e. the complete genome is affected, e.g. number of chromosomes

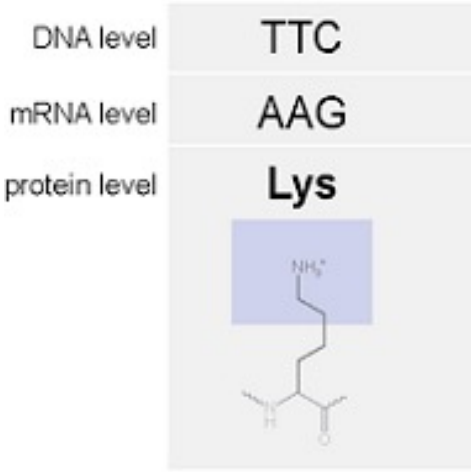
Point Mutations / Gene Mutations

- Single Nucleotide Polymorphism (SNP) := Affects a single locus on a gene, e.g. substitution of a single base
- In/Del := Insertion/Deletion of an arbitrary number of bases resulting in a frame shift
- Non-functional := No impact on products created from the affected genetic code, e.g. compensated through amino acids redundancy
- Functional := Impact on products built from affected genetic code, e.g.:

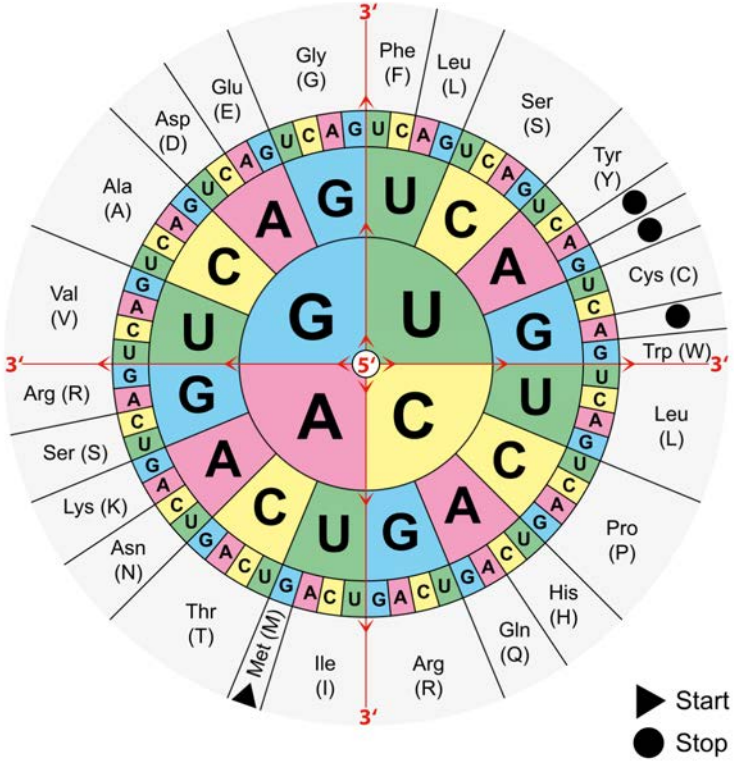
Size (before vs. after)	Type	Impact
=	Missense	Changes triplet, i.e. another amino acid chain is synthesized
>	Nonsense	Converts existing triplet to stop codon (+STOP)
<	Nonstop	Converts existing stop codon to another triplet (-STOP)

Point Mutations / Gene Mutations

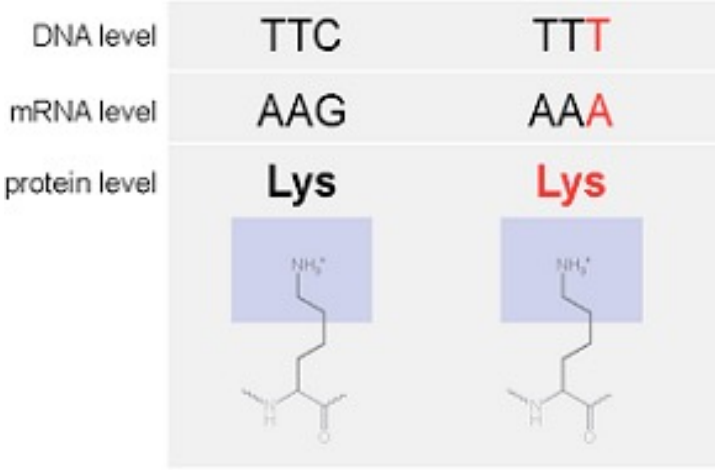
- Lysine, a.k.a. Lys (K)



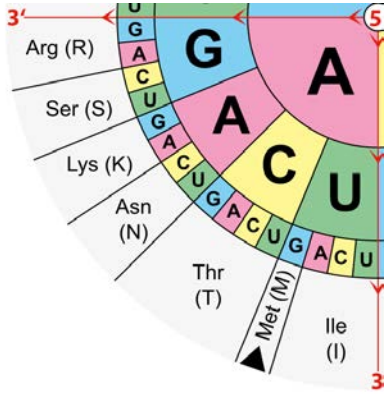
http://study.com/cimages/multimages/16/point_mutations.PNG



Point Mutations / Gene Mutations
Silent Mutation / Non-functional



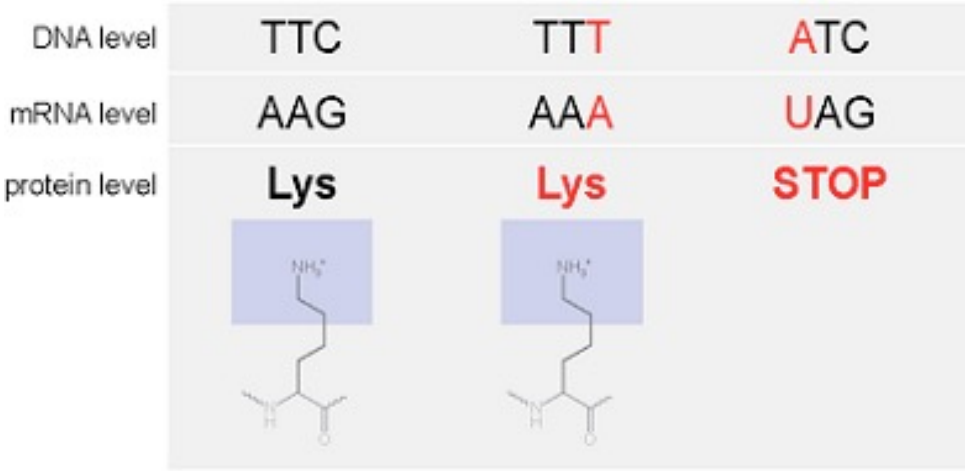
http://study.com/cimages/multimages/16/point_mutations.PNG



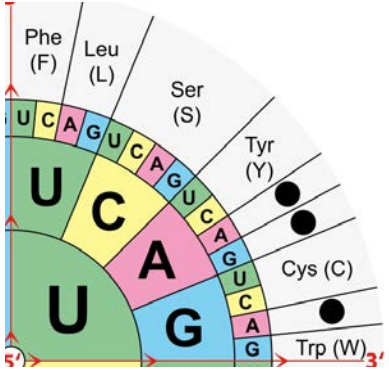
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Point Mutations / Gene Mutations
Nonsense Mutation / Functional



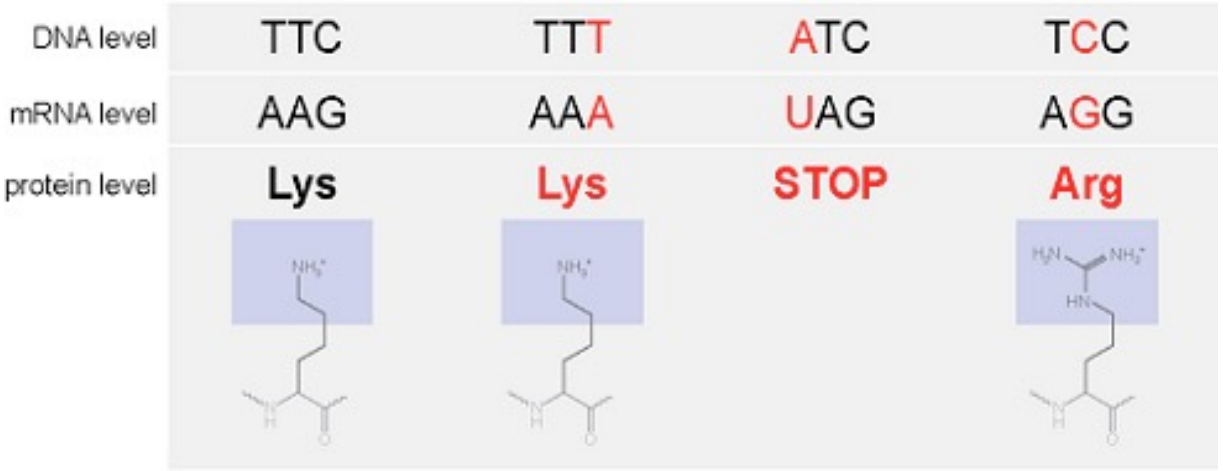
http://study.com/cimages/multimages/16/point_mutations.PNG



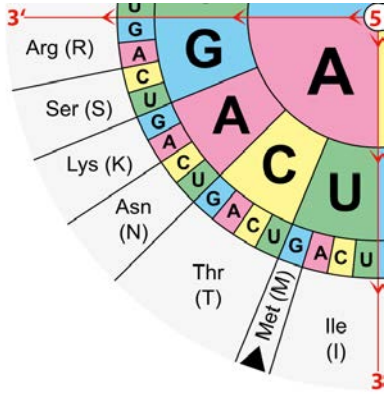
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Point Mutations / Gene Mutations
Conservative Missense Mutation / Functional



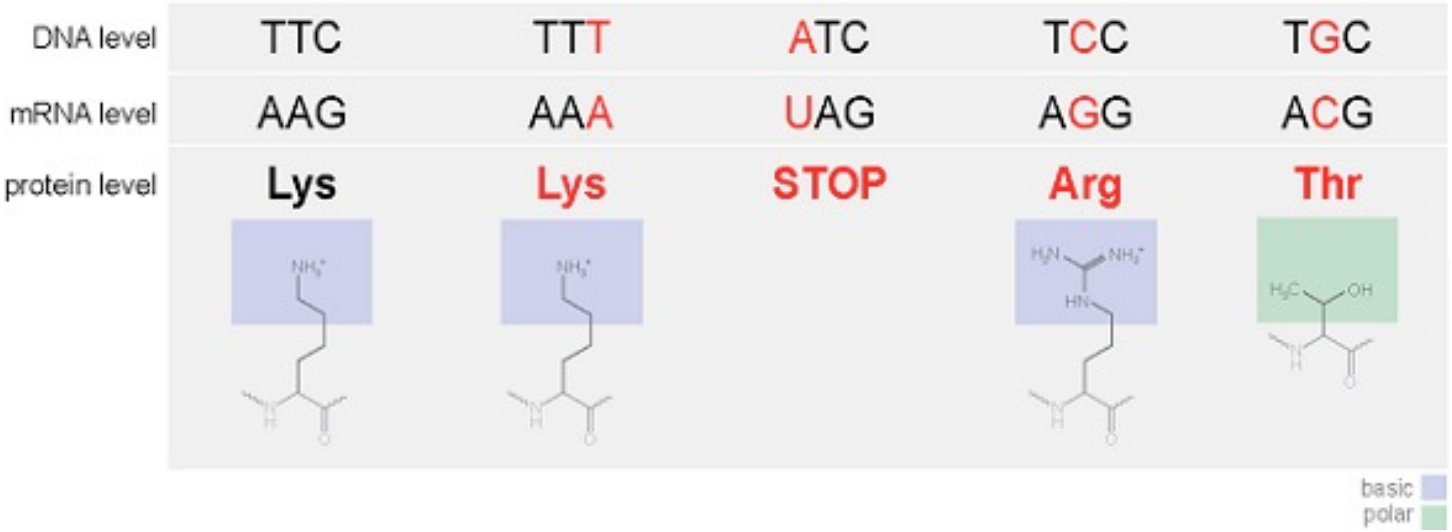
http://study.com/cimages/multimages/16/point_mutations.PNG



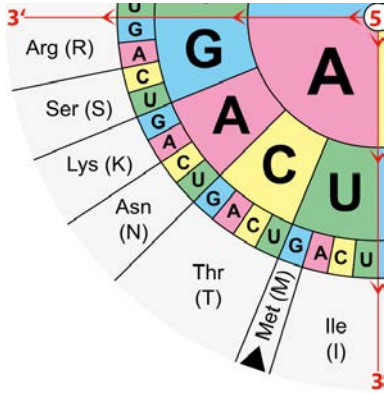
Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Point Mutations / Gene Mutations
Non-conservative Missense Mutation / Functional



http://study.com/cimages/multimages/16/point_mutations.PNG



Medical Use Case Oncology

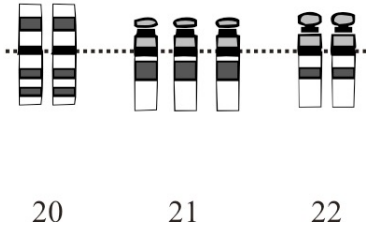
Data Management for Digital Health, Winter 2023

Chromosome Mutations

- In-/Del := Chromosome segment is inserted into/deleted from another chromosome
- Duplication := Chromosome segment is multiplied, e.g. amplying effect
- Inversion := Chromosome segment gets inverted on the chromosome
- Translocation := Segments of two chromosomes are exchanged
- Transposition := Copied segment is added to another chromosome

Genome Mutations

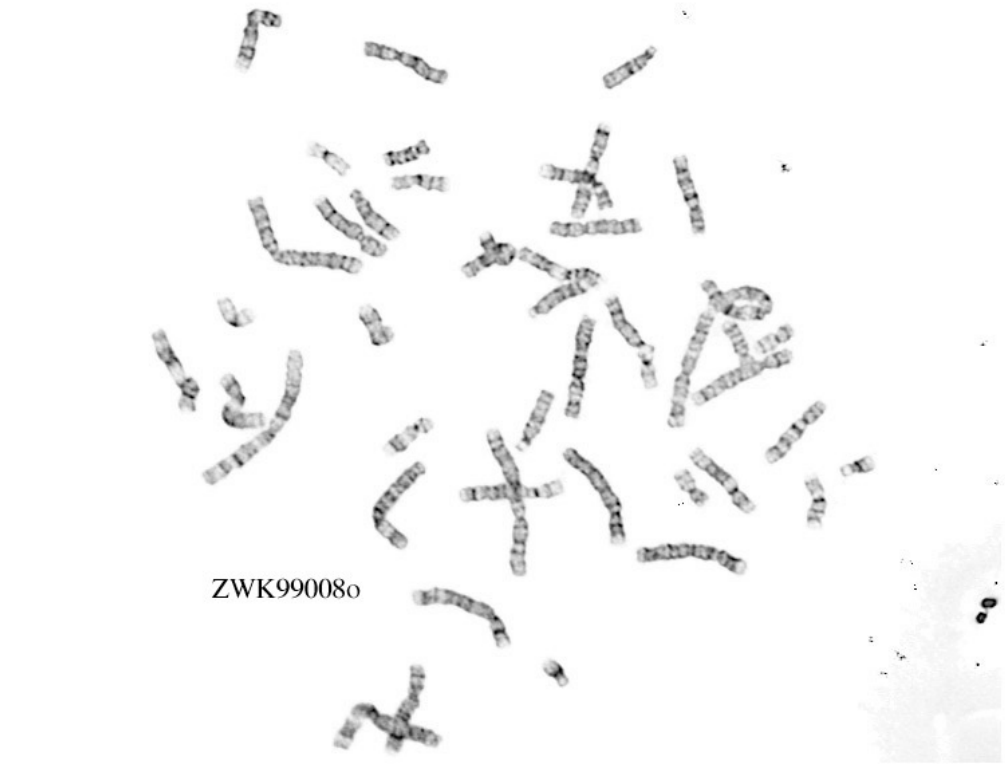
- Affects number of chromosomes



- Polyploidy := All chromosomes are multiplied
  - Diploid := All chromosomes exist twice, e.g. x=23, 2n=46 chromosomes
  - Triploid := All chromosomes exist three times, e.g. x=7, 3n=21 chromosomes
  - Tetraploid, etc.

- Aneuploidy := Number of selected chromosomes changed, e.g. trisomy 21, 2n+1

Genome Mutations
Karyograms

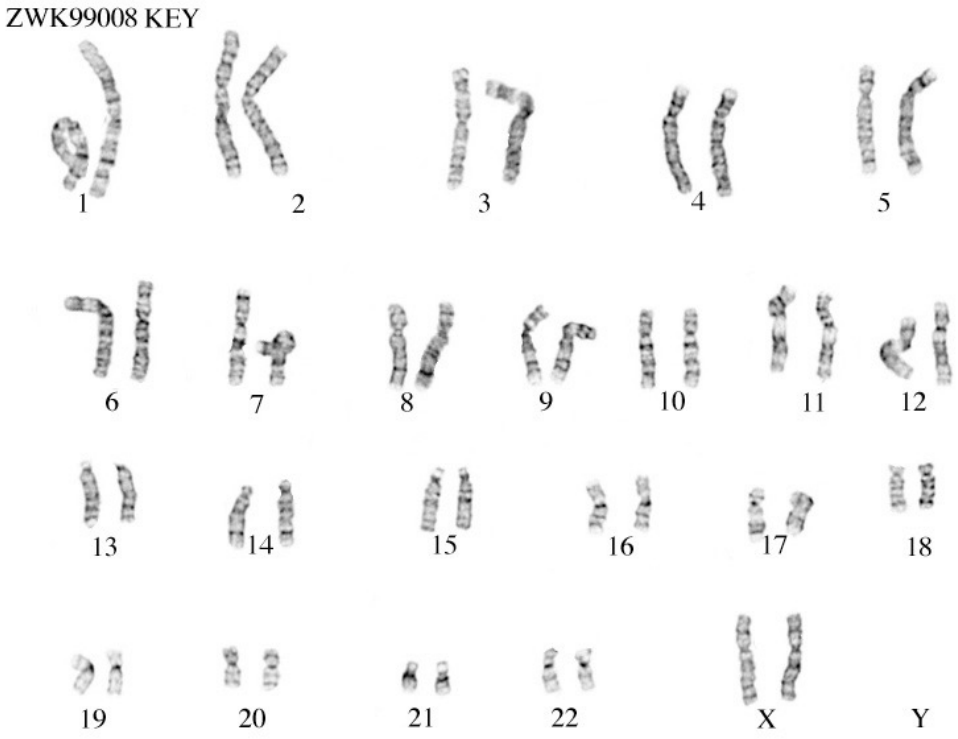


Wisconsin State Laboratory of Hygiene

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

Genome Mutations
Karyograms

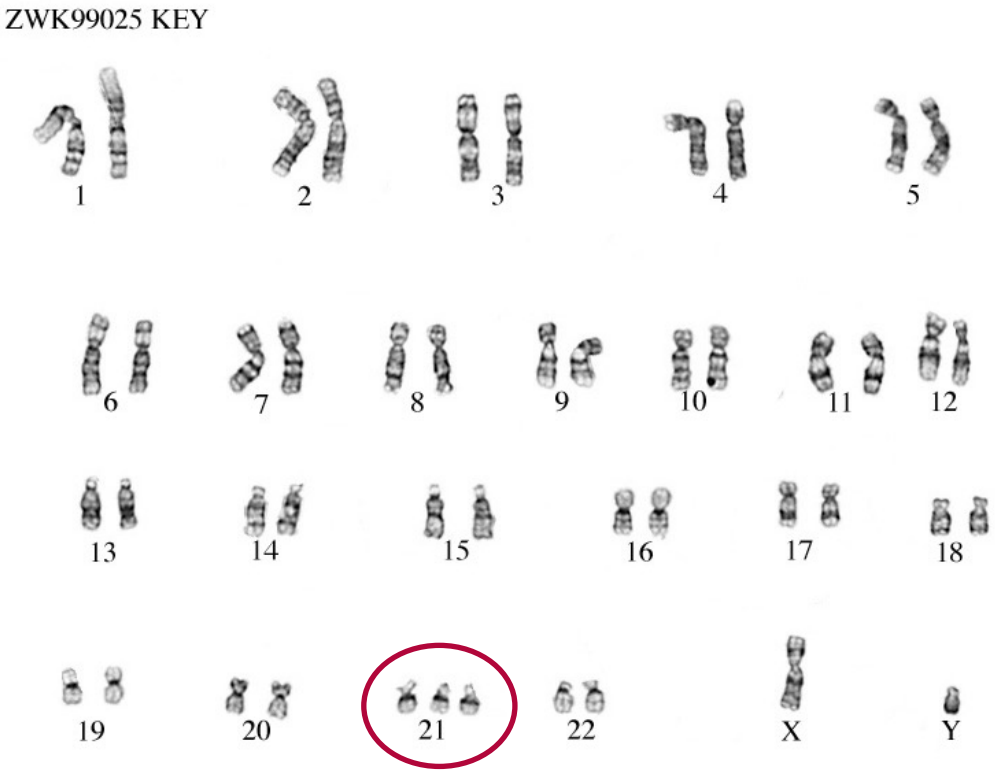


Wisconsin State Laboratory of Hygiene

Medical Use Case Oncology

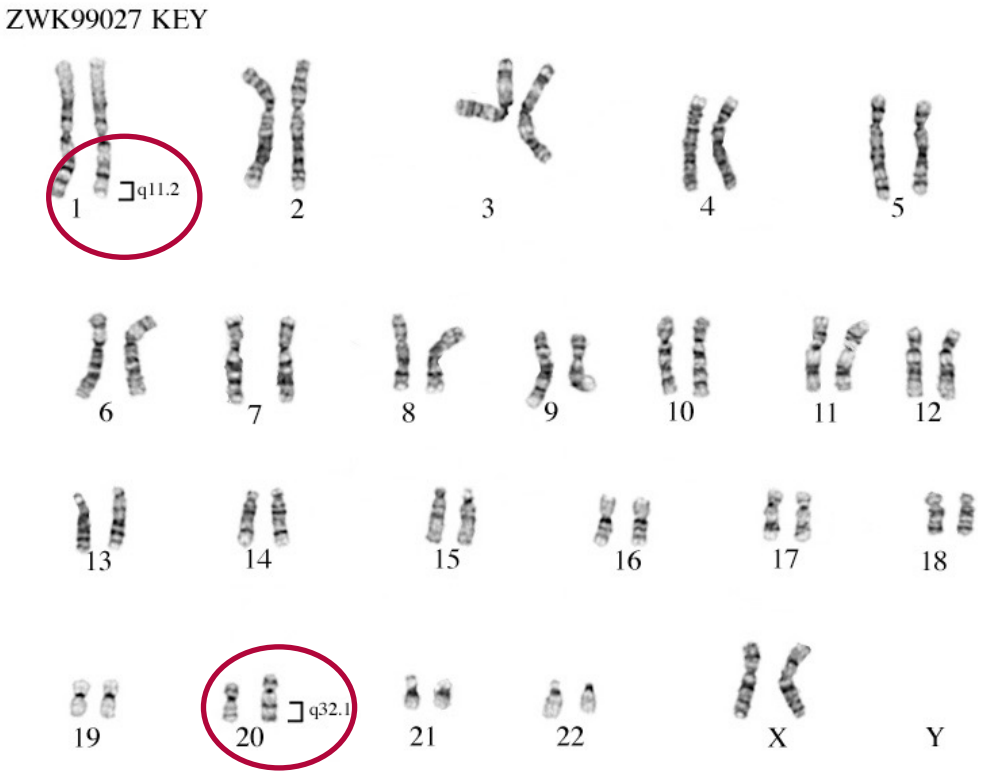
Data Management for Digital Health, Winter 2023

Genome Mutations
Karyograms



Wisconsin State Laboratory of Hygiene

Genome Mutations
Karyograms



Wisconsin State Laboratory of Hygiene

Medical Use Case Oncology

Data Management for Digital Health, Winter 2023

What to Take Home?

- Mutations
  - Occur spontaneously and undirected
  - Increase variability; are considered as foundation for evolution
- Changes in the genetic code can occur on various levels
- Many variants can be compensated, i.e. have no functional impact

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Hands-on: Understanding DNA

Build your own DNA
Objectives

- Understand
  - Phosphate, Deoxyribose, Nucleobase → Nucleotides
  - Complementary base pairing
  - Semiconservative replication
  - 3D-structure of DNA double strand
  - Transcription and translation
  - Different kinds of mutations and their effects

- **Have fun!**

Build your own DNA
Getting creative with PlayMais

A 3D-DNA Molecule Made of PlayMais

Massimo Caine, Ninon Horié, Sandrine Zuchuat, Aurélia Weber, Verena Ducret, Patrick Linder & Karl Perron

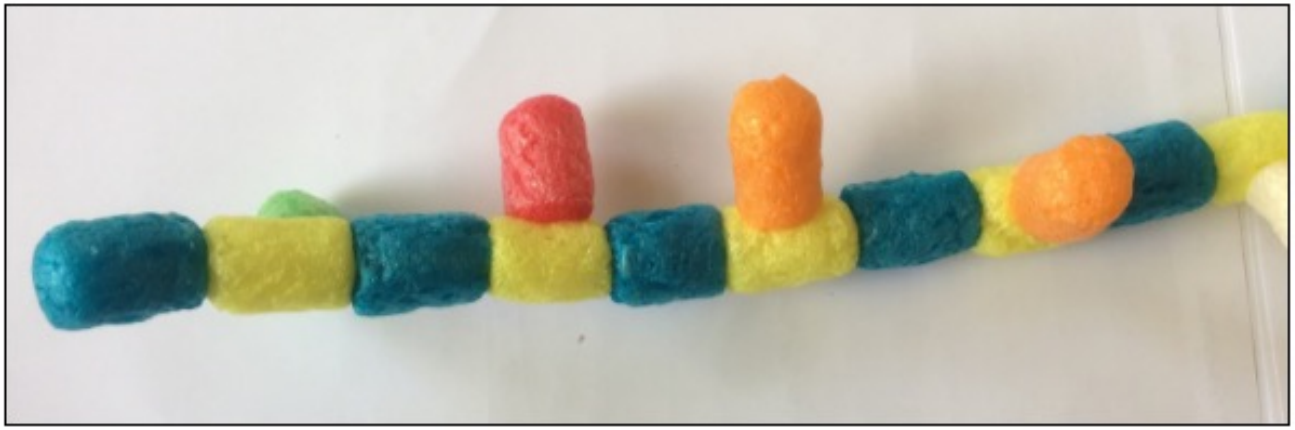
To cite this article: Massimo Caine, Ninon Horié, Sandrine Zuchuat, Aurélia Weber, Verena Ducret, Patrick Linder & Karl Perron (2015) A 3D-DNA Molecule Made of PlayMais, Science Activities: Classroom Projects and Curriculum Ideas, 52:2, 31-44, DOI: 10.1080/00368121.2015.1029427

To link to this article: http://dx.doi.org/10.1080/00368121.2015.1029427

Medical Use Case
Oncology

Data Management for Digital Health, Winter 2023

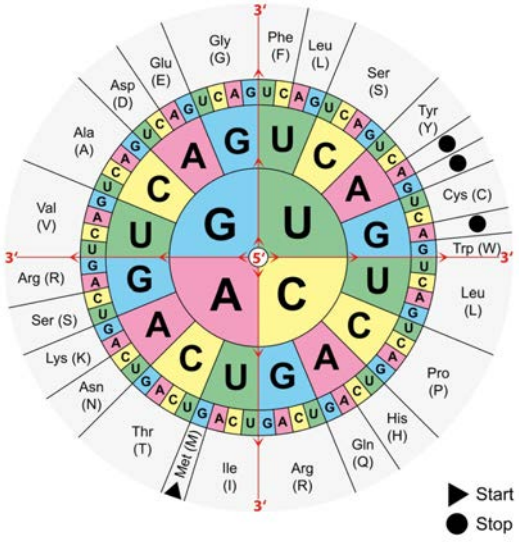
Part 1+2:



Part 3

Part 4 + 5

- Merge all sequences
- Transcribe and translate the complete sequence



- Introduce mutations and DNA modifications, e.g.
  - Point mutation
  - Indels
  - Premature STOP codons

Build your own DNA
Overall set-up

- Split into groups
- Receive your group target sequence
- Build the given sequence according to the manual
- Merge all sequences
- Translate the sequence
- Introduce Mutations
- Discuss the results

Homework

- Read the manual for the upcoming DIY workshop
- Have another look at the lecture slides and understand the biological concepts
- Have types of mutations at hand in our DIY workshop

Upcoming Week: DIY DNA

- Bear in mind:
  - Functions, e.g. data encoding
  - Stability, e.g. mechanical and functional
  - Structure, e.g. double helix