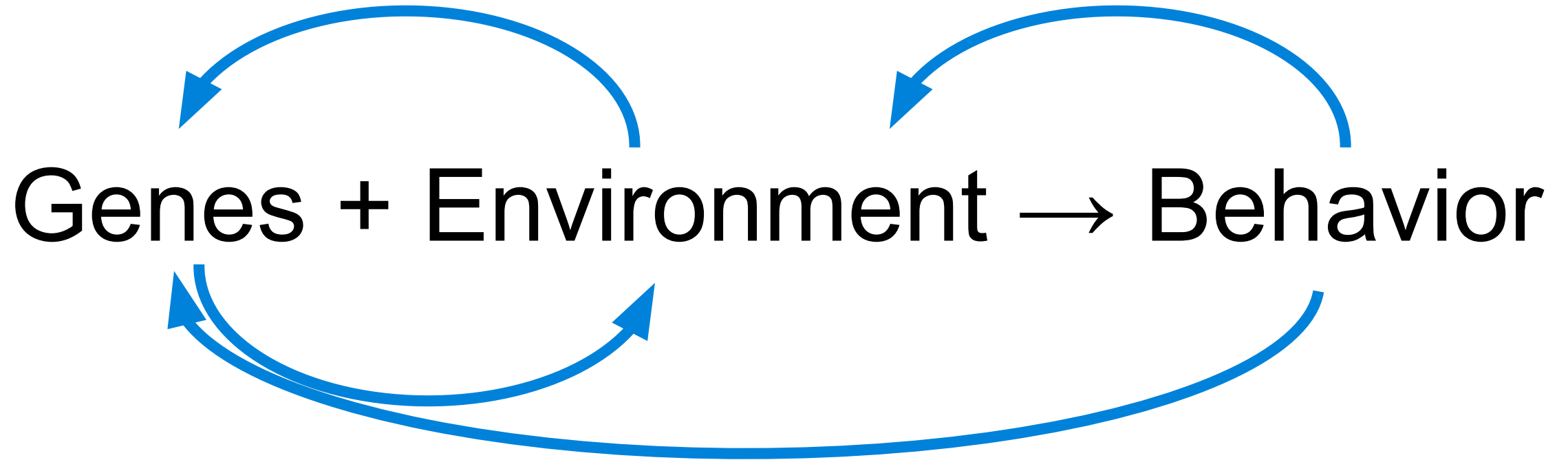


AMIDD 2023 Lecture 5: Biological Sequence Analysis



Dr. Jitao David Zhang, Computational Biologist

¹ Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

² Department of Mathematics and Informatics, University of Basel

Today's goals

- Drug discovery and development in social context
- Three views of drug discovery and development: the pipeline view, the multiscale modelling view, and the map/network view.
- Biological and mathematical foundations of biological sequence analysis
 - Biological functions and mathematical modelling of mutations
 - Dynamic programming
- Selected applications of biological sequence analysis for drug discovery

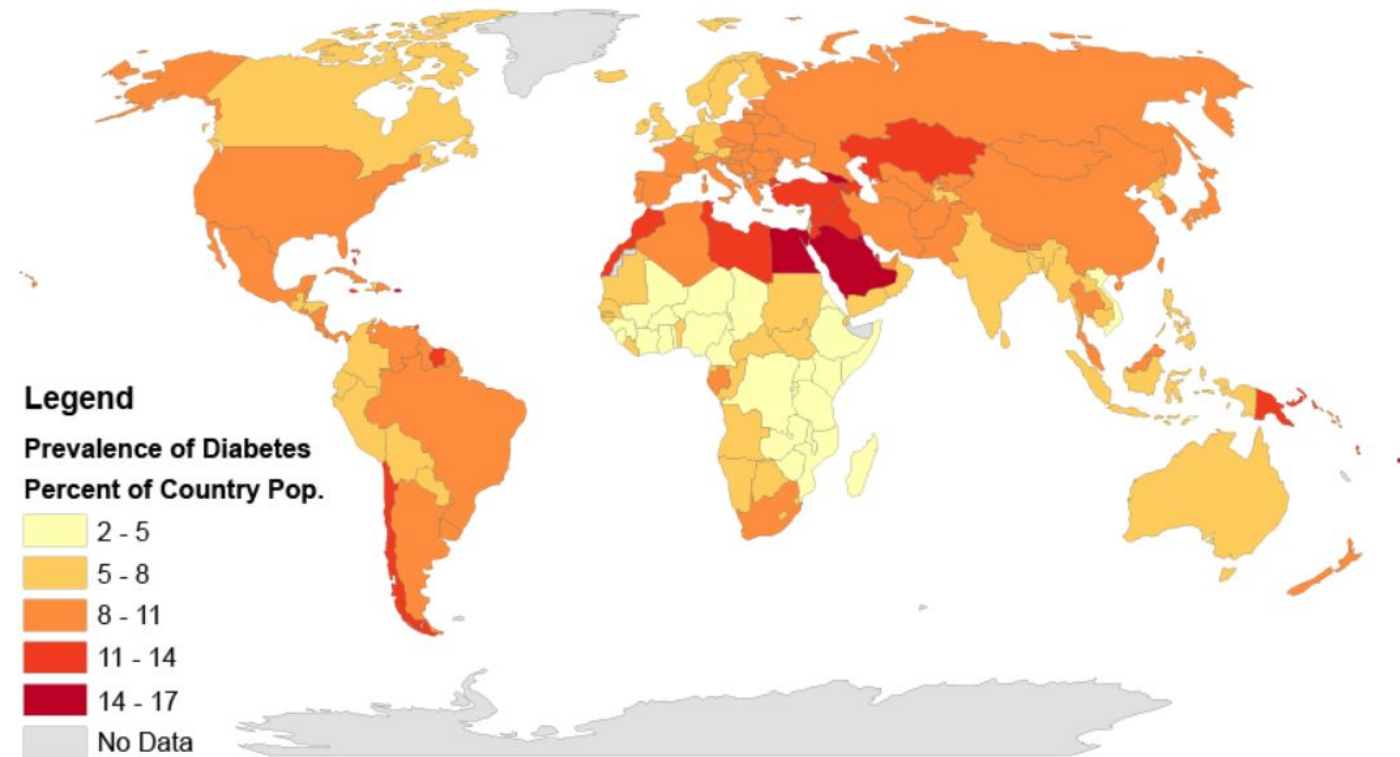
The social context of drug discovery: a role-playing game

We consider a case study of developing new drugs to treat complications caused by type II diabetes, which affects on average 9.2% of the world population.

We divide the classroom into five personas.

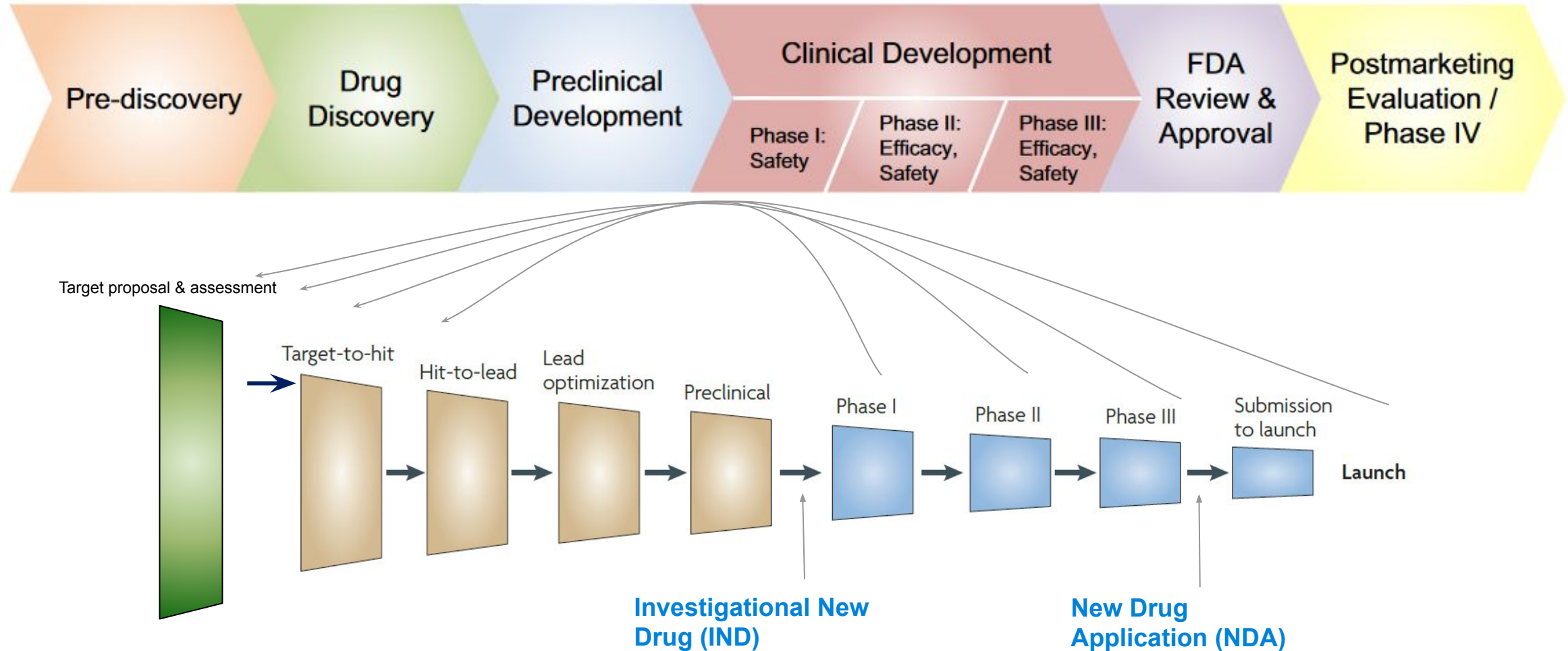
1. Patients
2. Medical doctors
3. Two drug discovery companies, *IntelliMed* and *BetterMed*.
4. An academic research institute
5. The regulatory agency

Questions for each group: (1) What are your main interests and concerns? (2) With which other group do you have to work together and why? (3) What are your priorities? (4) What are the ideal and worse scenarios for you?



[Global prevalence of diabetes from 2014](#), using data from 195 countries. Source: Wikimedia. Author: Walter Scott Wilkens. Reused CC-AS 4.0 license.

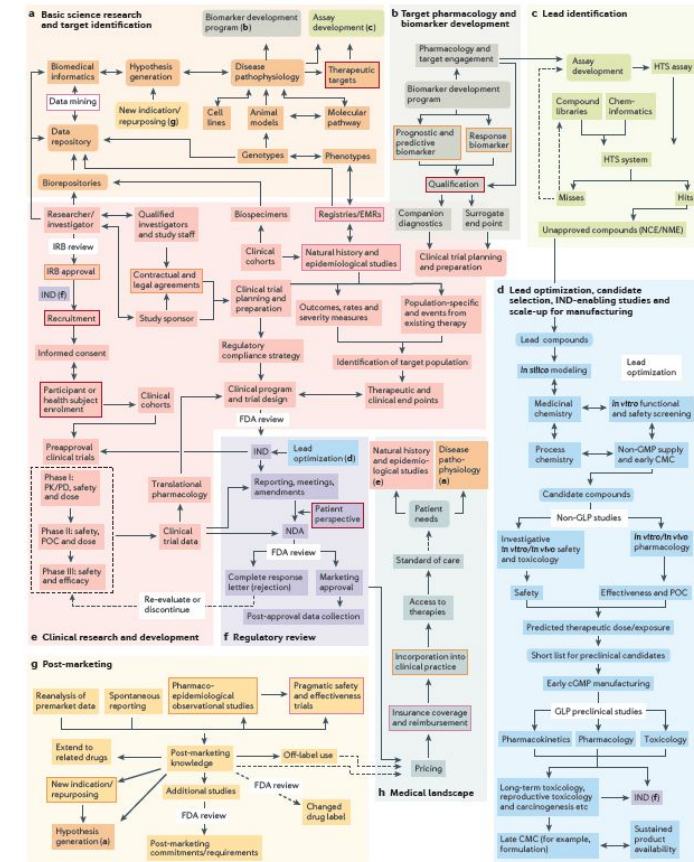
Chevron diagrams as a pipeline view of drug discovery and development



FDA: US Food and Drug Administration. Top: Wagner, J. A. et al. [Application of a Dynamic Map for Learning, Communicating, Navigating, and Improving Therapeutic Development](#). Clinical and Translational Science 11, 166–174 (2018). Bottom: Adapted from Paul et al. [How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge](#). Nature Reviews Drug Discovery, 2010.

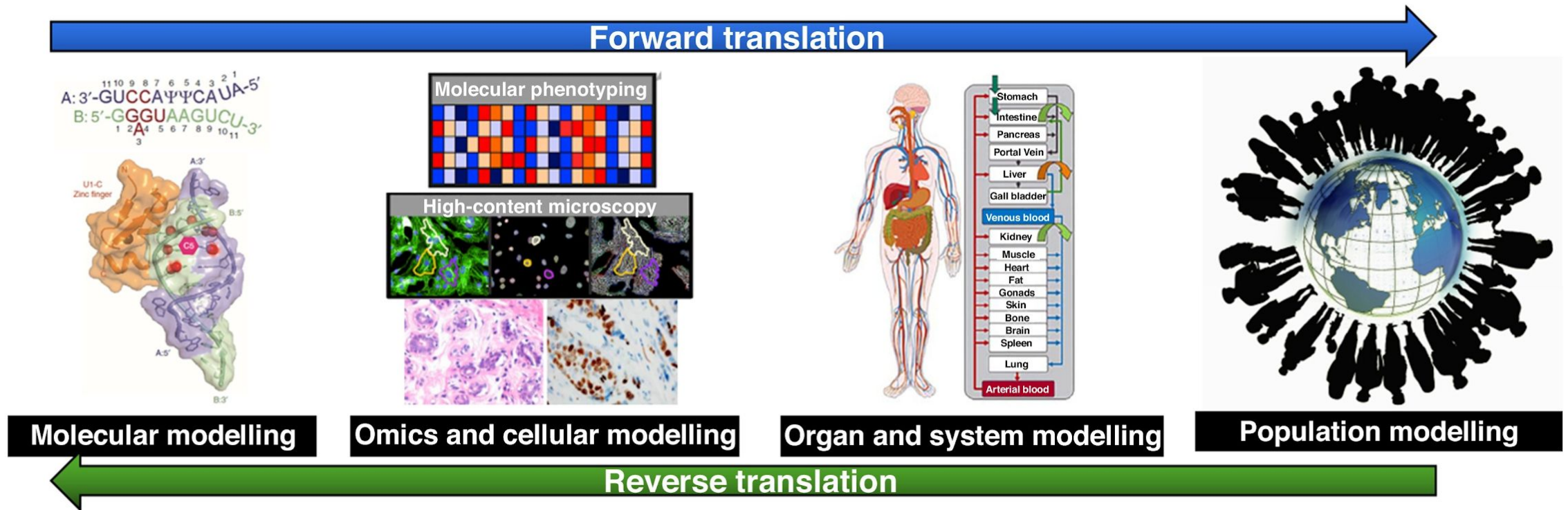
A dynamic map for drug discovery, development, and deployment

1. **Basic science research and target identification.** *What causes the disease? What do we want to achieve? Which protein can I target with which modality?*
2. **Target pharmacology and biomarker development.** *What is the effect of targeting the protein? What we can measure to confirm that the protein is properly targeted?*
3. **Lead identification.** *How can we find a starting point of a new drug?*
4. **Lead optimization and clinical candidate selection.** *What are criteria to define a good drug? How can I improve the starting material?*
5. **Clinical research and development.** *Does it work in human? How about efficacy and safety profiles?*
6. **Regulatory review.** *Should we approve the drug?*
7. **Post marketing.** *How does the drug work in real world?*



Wagner, J. et al. [A dynamic map for learning, communicating, navigating and improving therapeutic development.](#) Nat Rev Drug Discov 17, 150–150 (2018).

The multiscale modelling view of drug discovery



Drug Discovery Today

Zhang, Jitao David, Lisa Sach-Peltason, Christian Kramer, Ken Wang, and Martin Ebeling. 2020. "Multiscale Modelling of Drug Mechanism and Safety." *Drug Discovery Today* 25 (3): 519–34. <https://doi.org/10.1016/j.drudis.2019.12.009>.

Six key questions to ask before we start

1. For what indication do we want to discover and develop a new drug?
2. What problems met by patients do we want to solve?
3. What is the target of our drug?
4. Where should the drug go and how long do they stay in patients body?
5. What is the safety profile of the drug in light of its benefits?
6. Who are particularly responsive to the drug?

Right target

- Strong link between target and disease
- Differentiated efficacy
- Available and predictive biomarkers

Right tissue

- Adequate bioavailability and tissue exposure
- Definition of PD biomarkers
- Clear understanding of preclinical and clinical PK/PD
- Understanding of drug–drug interactions

Right safety

- Differentiated and clear safety margins
- Understanding of secondary pharmacology risk
- Understanding of reactive metabolites, genotoxicity and drug–drug interactions
- Understanding of target liability

Right patient

- Identification of the most responsive patient population
- Definition of risk–benefit for a given population

Right commercial potential

- Differentiated value proposition versus future standard of care
- Focus on market access, payer and provider
- Personalized health-care strategy, including diagnostics and biomarkers

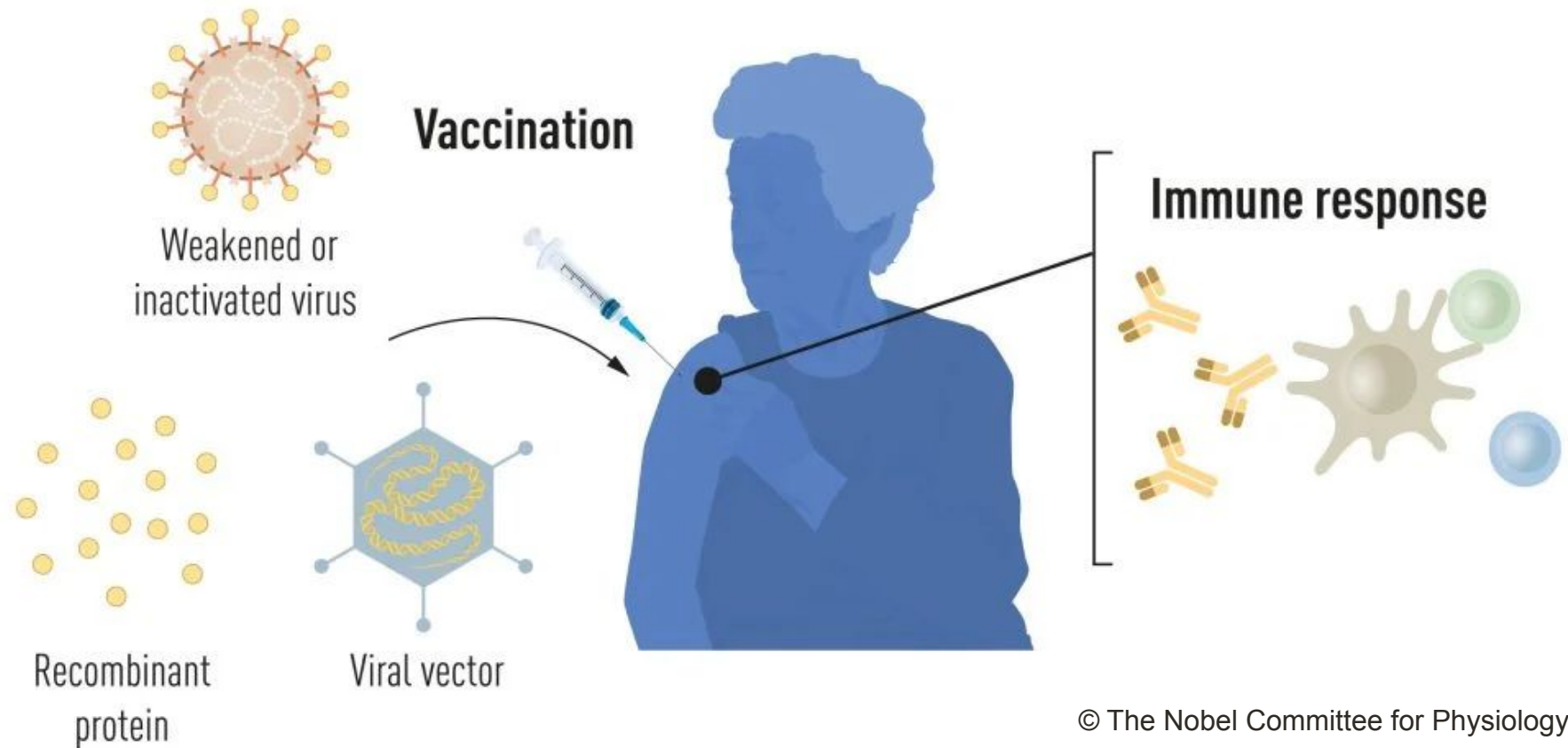
Morgan, P. et al. [Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nature Reviews Drug Discovery](#) 17, 167–181 (2018).

Nobel Prize in Physiology or Medicine 2023 was awarded to Katalin Karikó and Drew Weissman for “their discoveries concerning nucleoside base modifications that enabled the development of effective mRNA vaccines against COVID-19”



Methods for vaccine production before the COVID-19 pandemic.

Issues: large-scale cell culture required, which limits the possibilities for rapid production in response to pandemics.



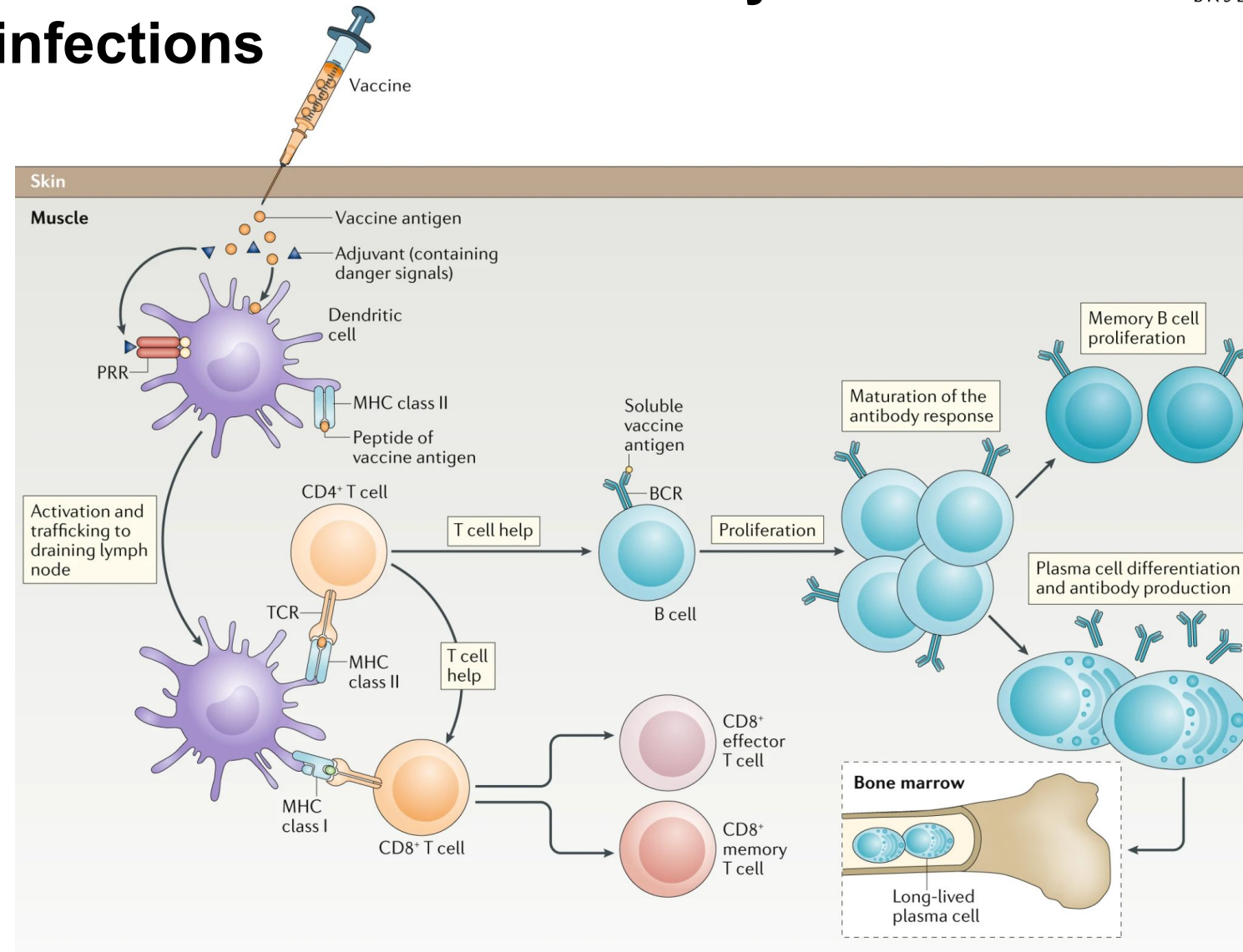
© The Nobel Committee for Physiology or Medicine. III. Mattias Karlén

Vaccine mimics viral infection to activate the immune system to protect body from future infections

Vaccine mimics a viral infection to activate innate and adaptive immune system, while minimizing the pathogenic effects.

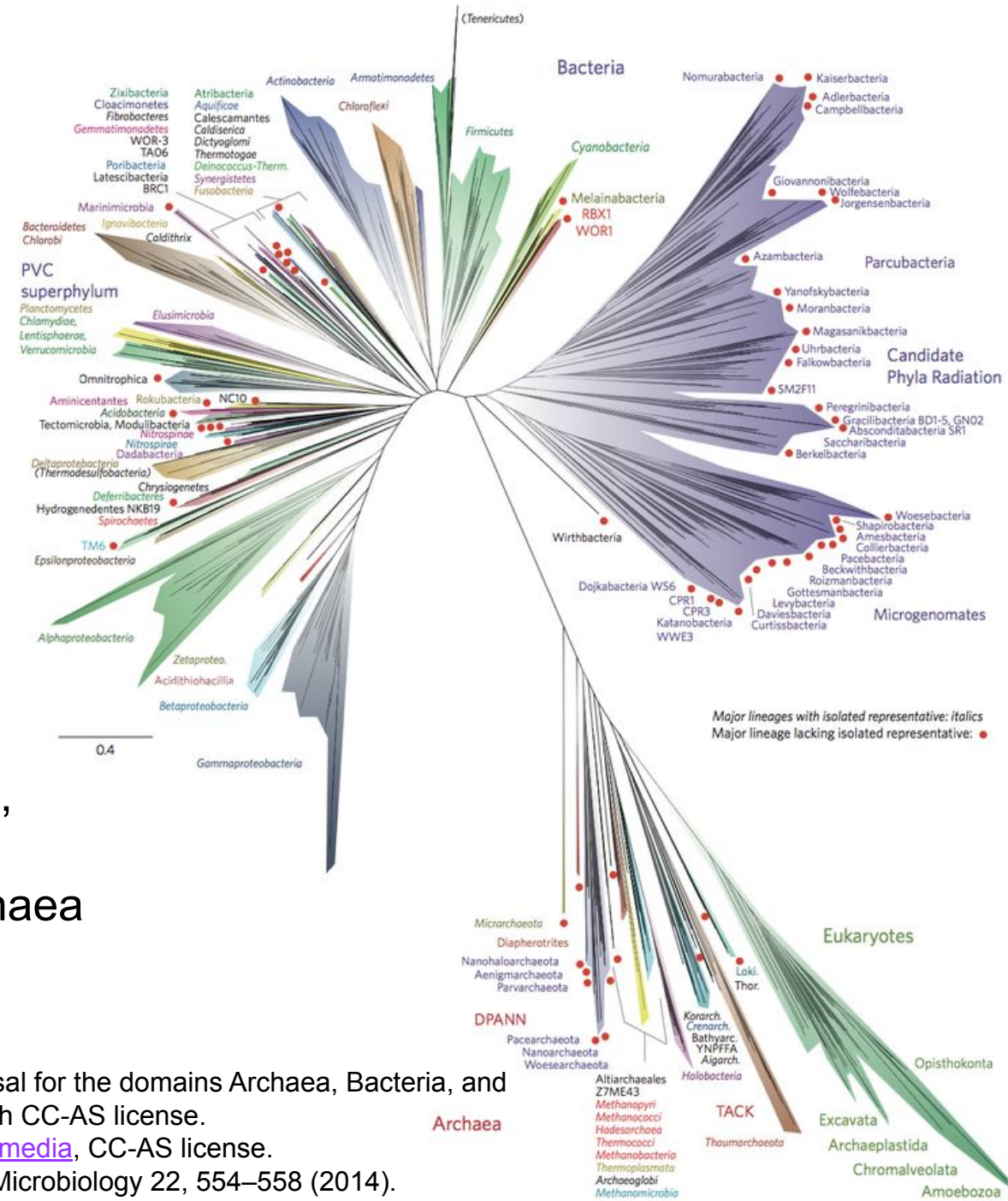
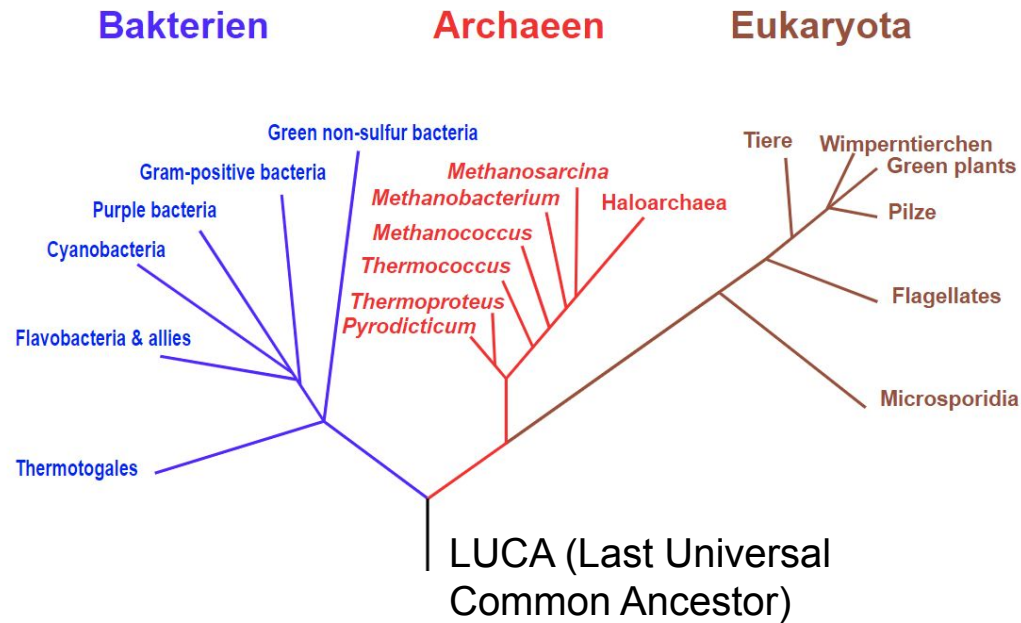
Key players in the game:

1. Viral proteins as *antigens*
2. Antigen-presenting cells (e.g. dendritic cells)
3. T cells (T comes from Thymus, because they mature there)
4. B cells (B comes from bone marrow).



Pollard, A. J. & Bijker, E. M. A guide to vaccinology: from basic principles to new developments. Nature Reviews Immunology 21, 83–100 (2021).

Virus is special



- The three-domain model of *cellular* life: (eu-)bacteria, archaeobacteria, and eukaryotes.
- The two-domain model: bacteria as one branch, archaea and eukaryotes as the other.
- Virus fits in no domain of neither models.

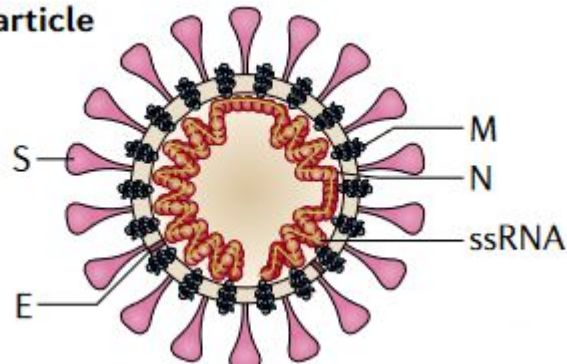
1. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87, 4576–4579 (1990). Figure from [Wikimedia](#), reused with CC-AS license.

1. Hug, L. A. et al. A new view of the tree of life. Nat Microbiol 1, 1–6 (2016). Figure from [Wikimedia](#), CC-AS license.

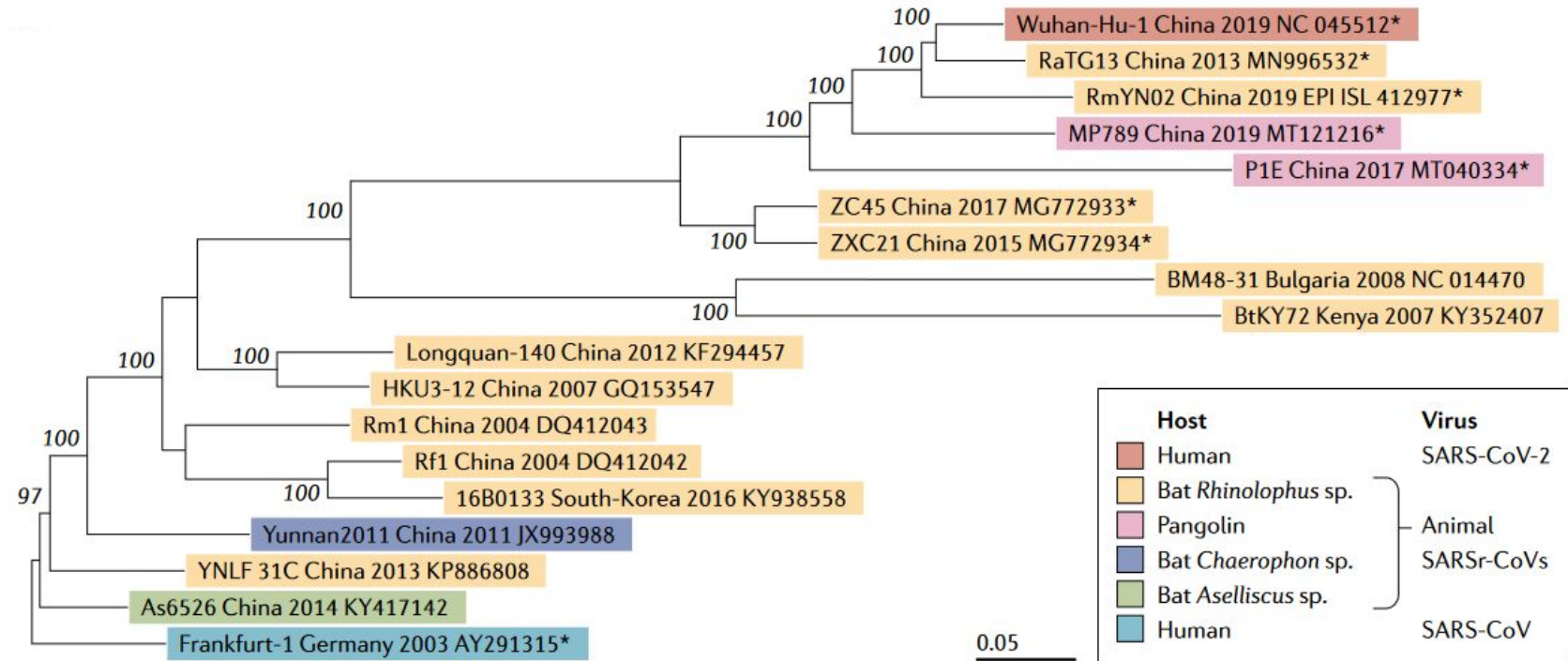
1. Forterre, P., Krupovic, M. & Prangishvili, D. Cellular domains and viral lineages. *Trends in Microbiology* 22, 554–558 (2014).

Coronavirus is a RNA virus infecting human and other species

Viral particle



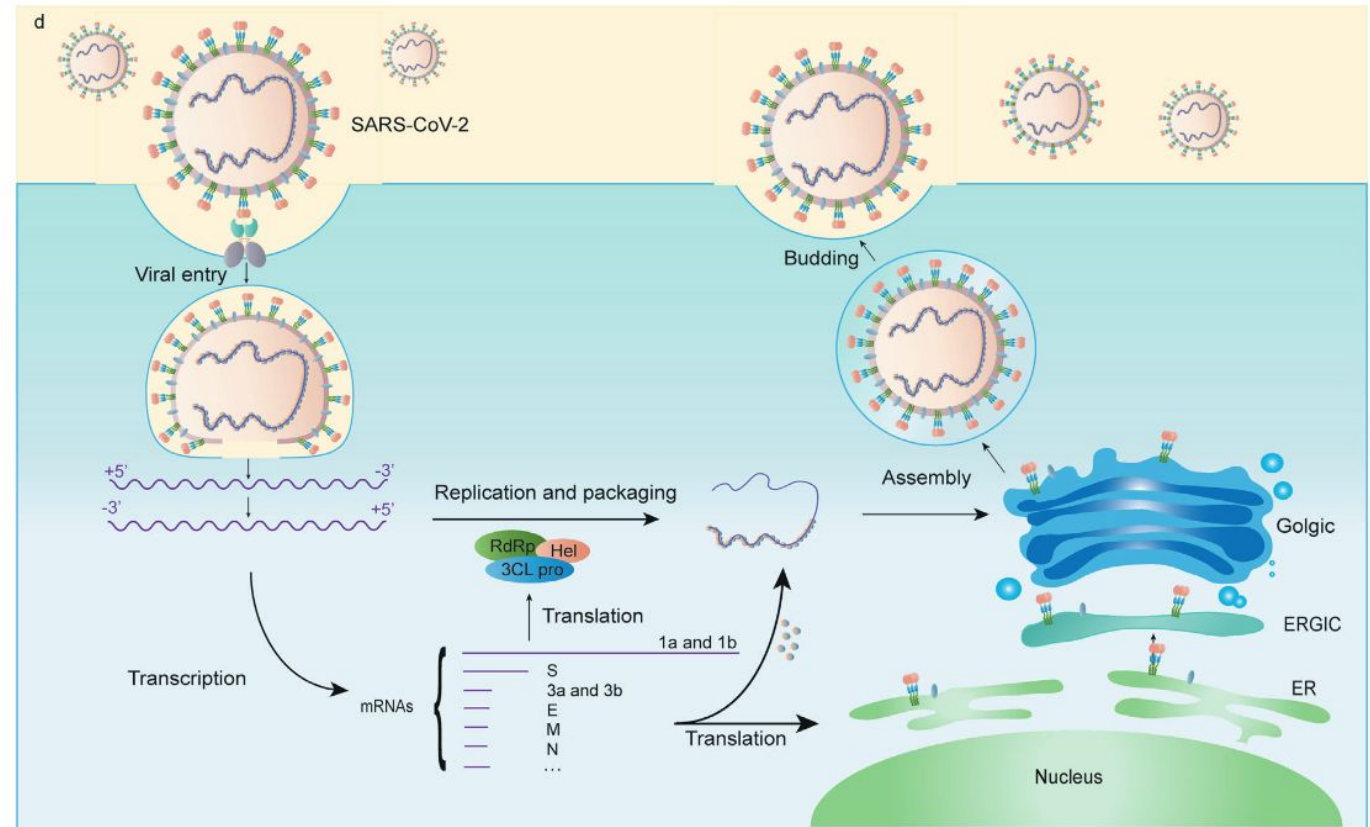
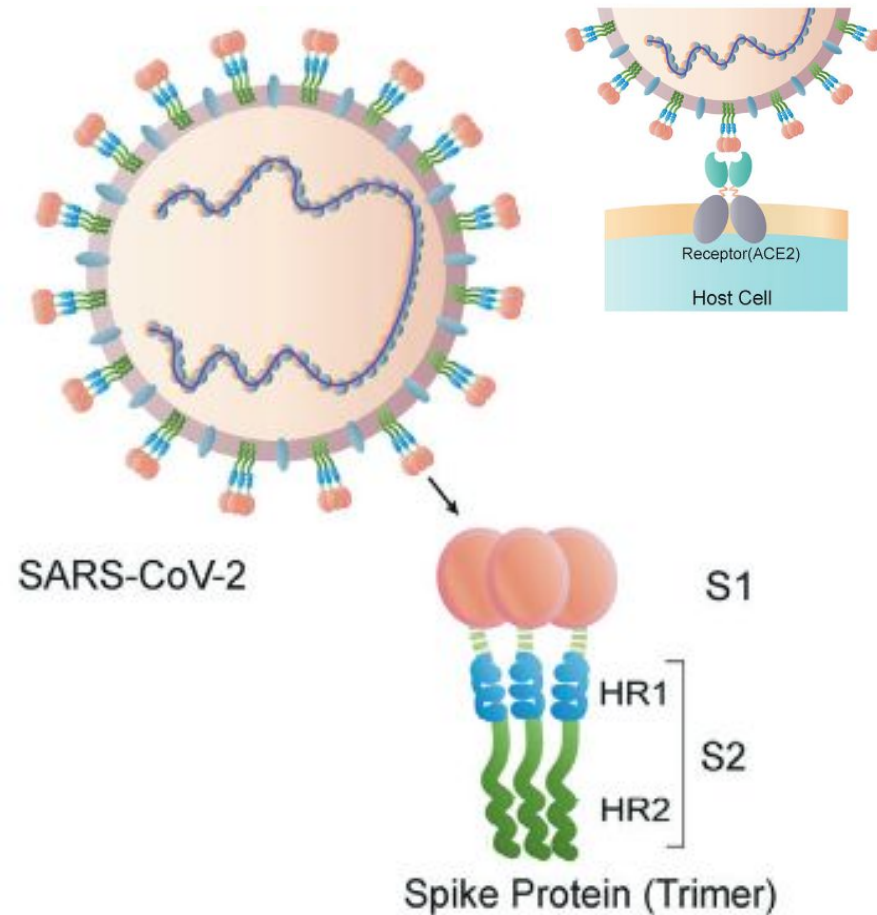
The coronavirus virion consists of structural proteins, namely spike (S), envelope (E), membrane (M), nucleocapsid (N)



Phylogenetic relationships of representative members of the species Severe acute respiratory syndrome-related coronavirus

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19, 155–170 (2021).

Spike protein of coronavirus is conserved and responsible for viral entry into human cells



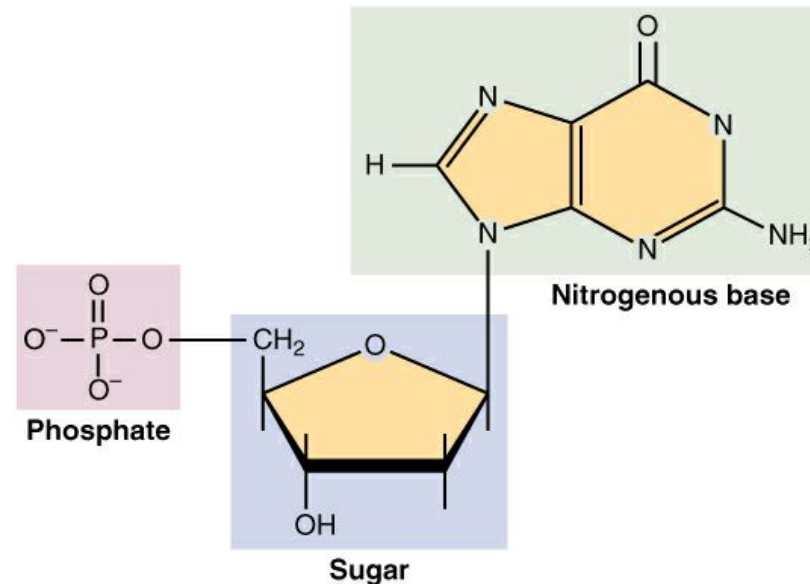
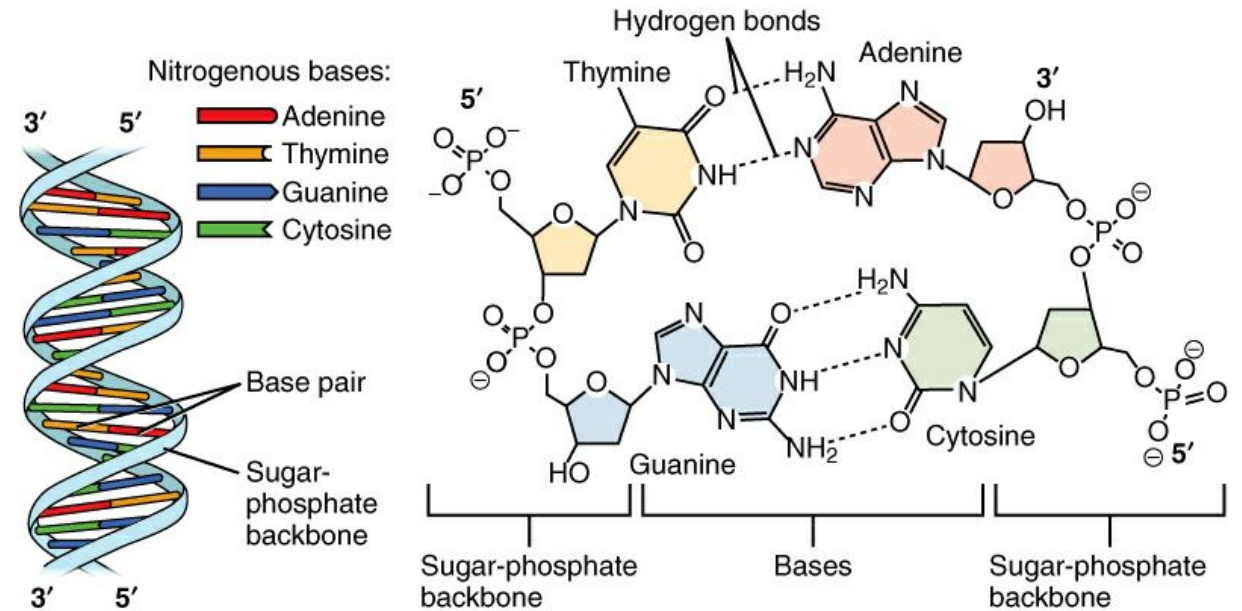
1. Huang, Y., Yang, C., Xu, X., Xu, W. & Liu, S. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* 41, 1141–1149 (2020).

DNA encode genetic information of all life forms that we know, including viruses



Photo 51, X-ray diffraction image of DNA

Franklin R, Gosling RG (1953)
"Molecular Configuration in Sodium Thymonucleate". *Nature* 171: 740–741.



From the textbook OpenStax Anatomy and Physiology, discovered through Wikimedia, reused under the CC license.

RNA is transcribed from DNA and translated into protein

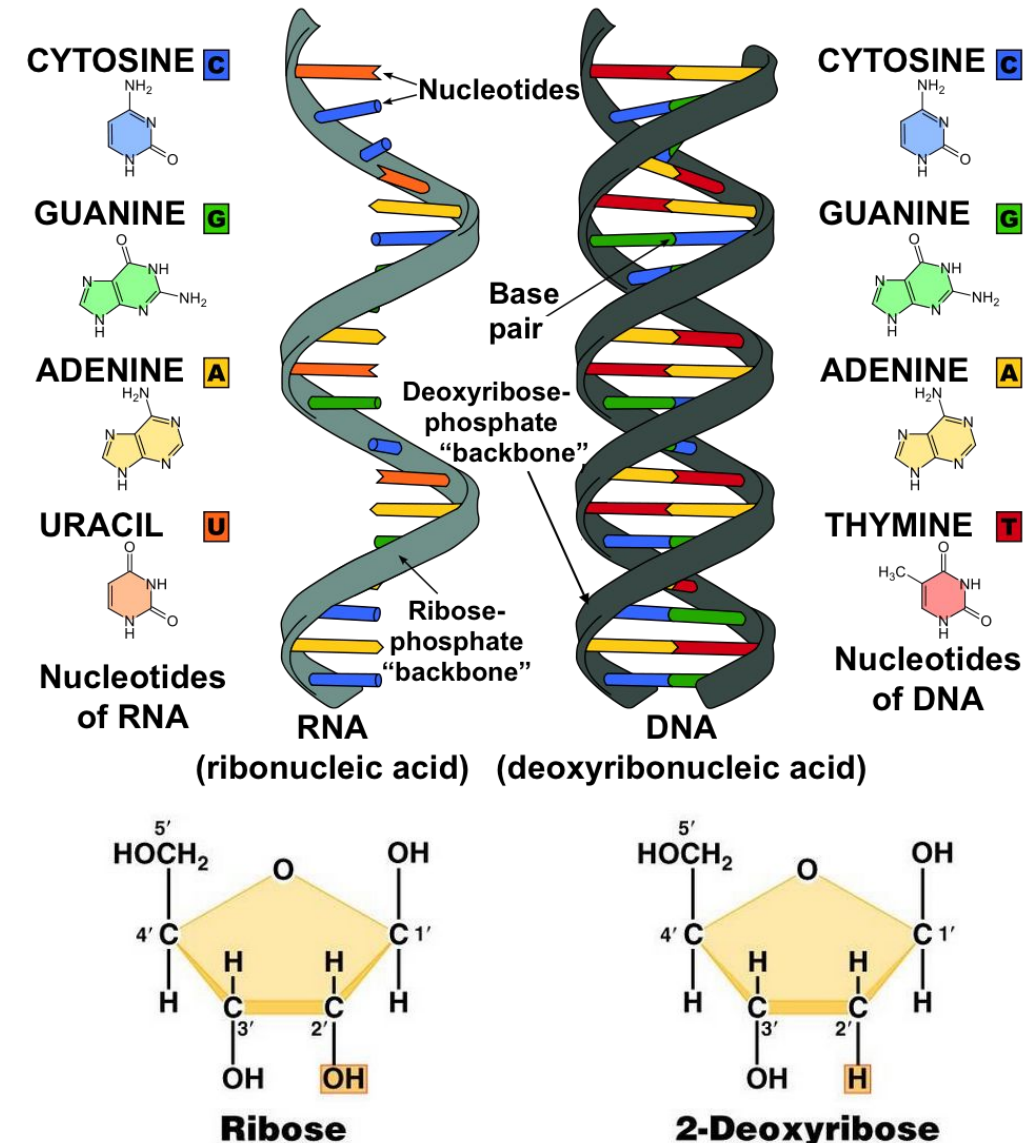
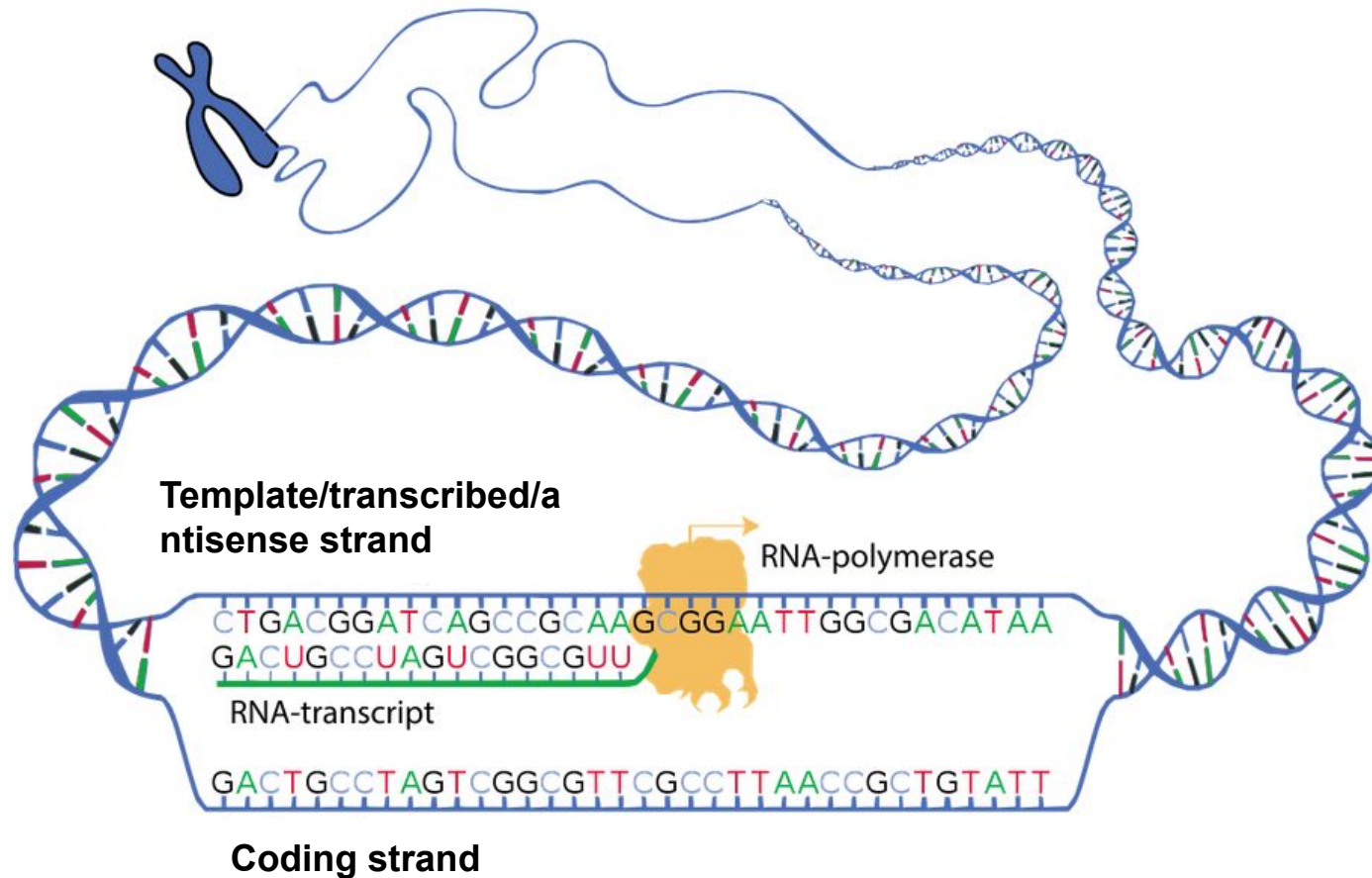
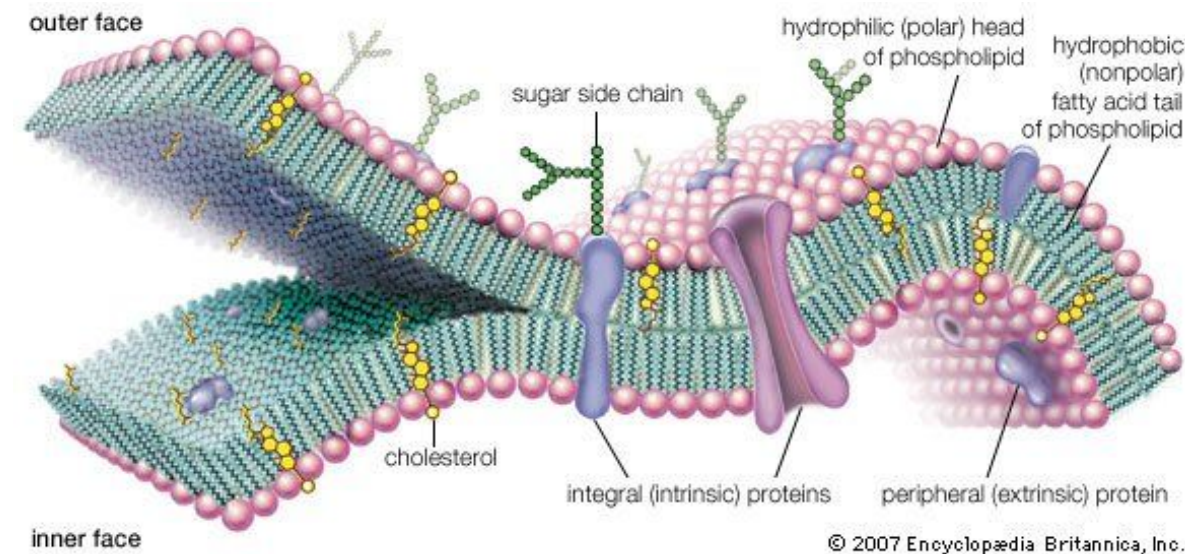
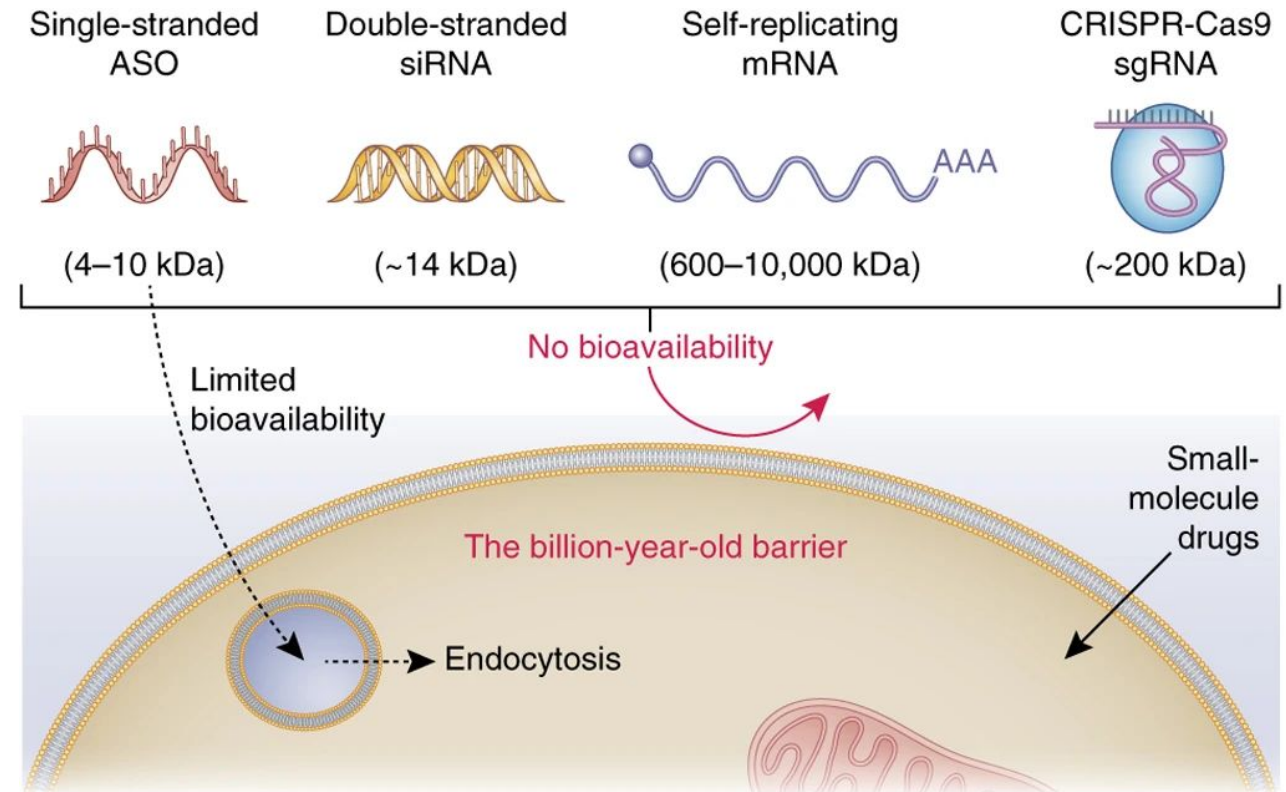


Figure: https://commons.wikimedia.org/wiki/File:DNA_transcriptie.svg and https://en.m.wikipedia.org/wiki/File%3AHAR1F_RF00635_rna_secondary_structure.jpg. Original work by wikipedia user: OrgreBot and user:Ppgardne. Used under CC-SA 3.0 license.

Delivery, stability, and immune responses are essential challenges for mRNA-based vaccines



- **mRNAs are too large and charged to pass lipid bilayers.**
- mRNAs are degradable by ribonucleases (RNases).
- mRNAs are rapidly cleared from liver and kidney
- *Exogenous* mRNAs induce *Immunogenicity*

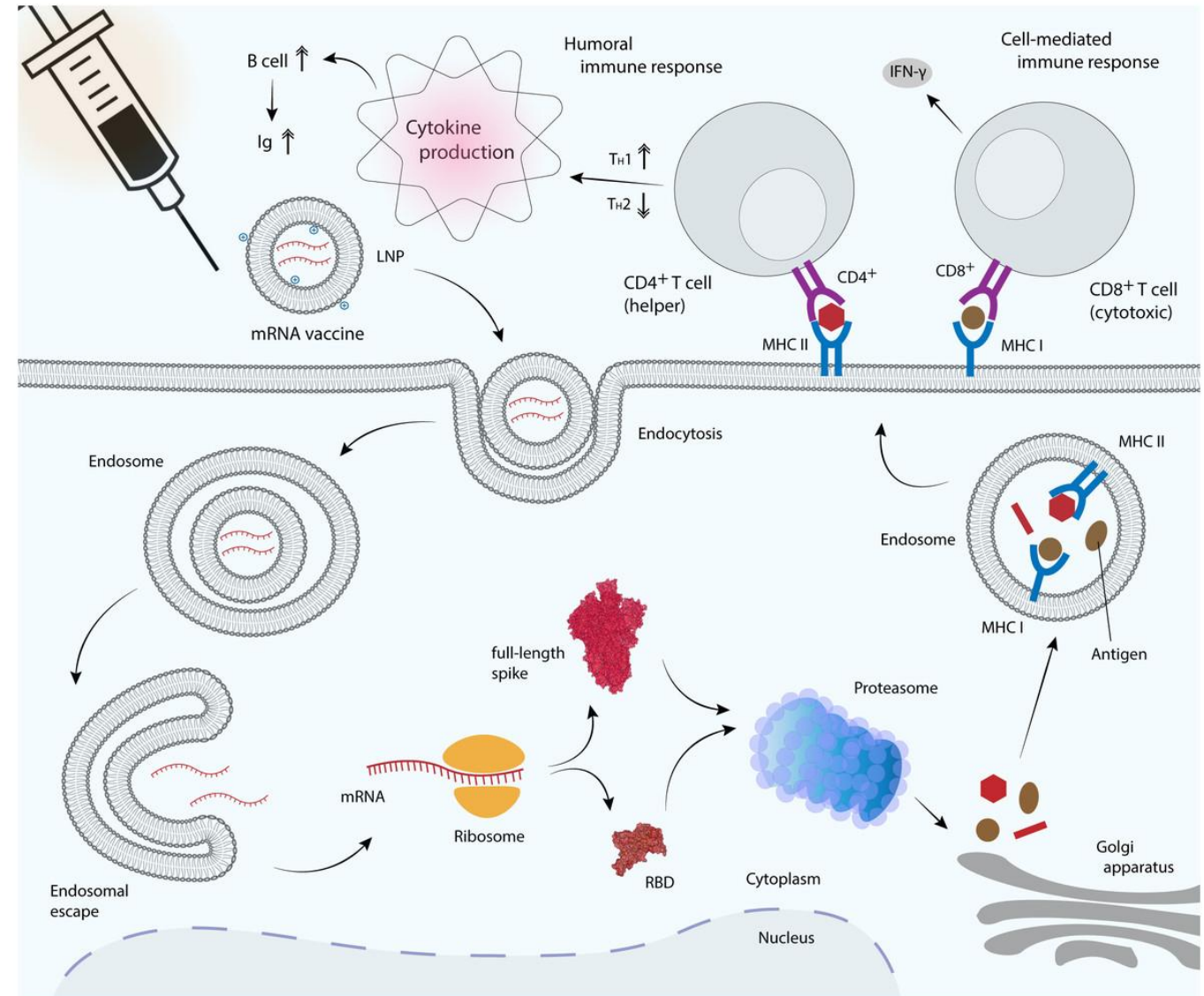


Left: Cell membrane, copyright of Encyclopædia Britannica, Inc.
Right: The four-billion-year-old barrier to RNA therapeutic

Lipid Nanoparticles help delivering RNAs into cells

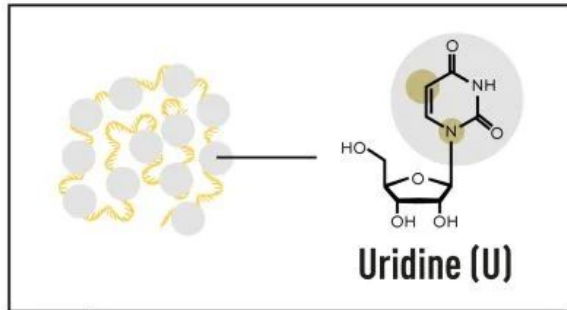
- Lipid nanoparticles can take mRNA vaccines as cargos, and deliver them into human cells.
- In the cell, mRNA encoding the part of the spike protein sequence is translated into proteins with the human protein translation mechanism.
- Synthesized proteins will be degraded and exposed on cell surface, which will be recognized by antigen presenting cells.

Salleh, Mohd Zulkifli *et al.* "[Immunogenicity Mechanism of mRNA Vaccines and Their Limitations in Promoting Adaptive Protection against SARS-CoV-2.](#)" PeerJ 10 (March 9, 2022)

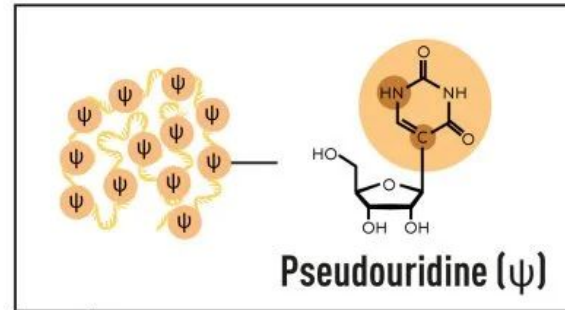


Delivery, stability, and immune responses are essential challenges for mRNA-based vaccines

Unmodified mRNA

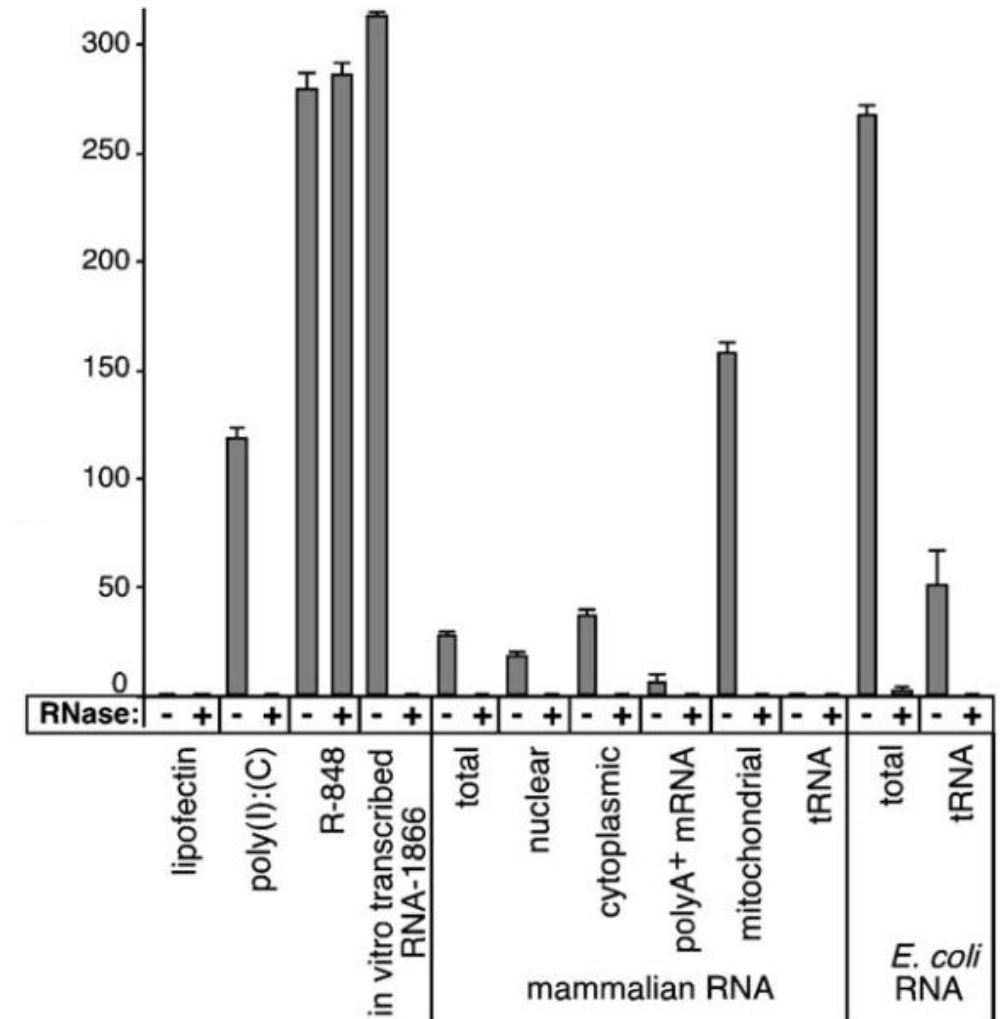


Base-modified mRNA

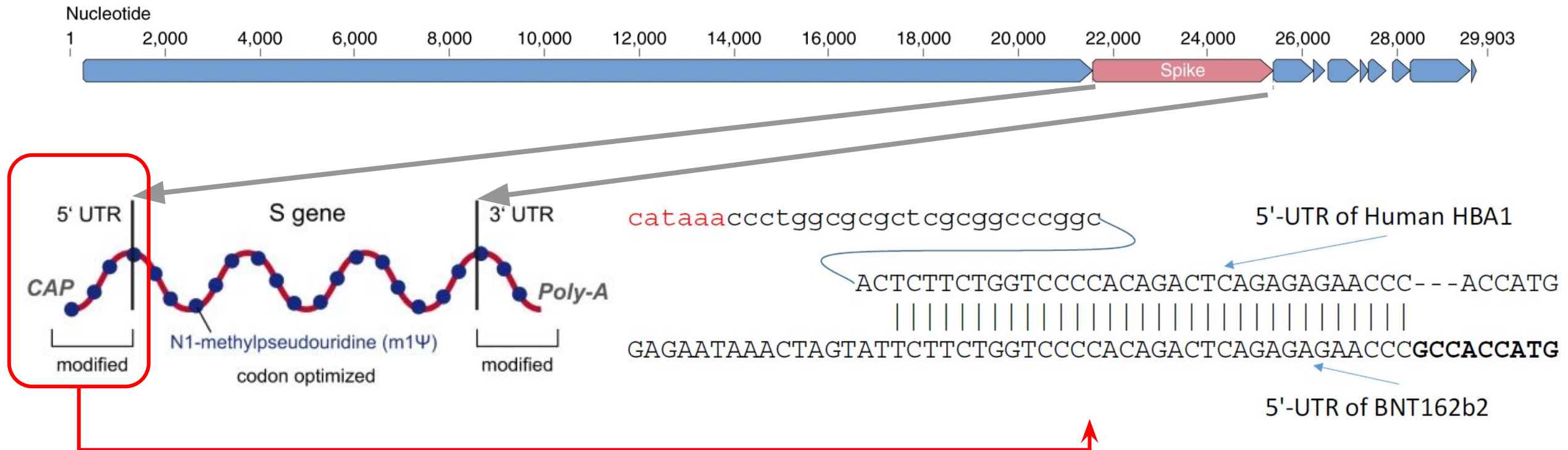


Exogeneous mRNAs induce Immunogenicity.

All RNAs is synthesize from four ribonucleotides: ATP, CTP, UTP, and GTP. Some nucleosides undergo modifications, for instance pseudouridine. Modified nucleosides suppress the potential of RNA to activate the immune cells in the immune system.



Modified RNAs are essential to make effective *and safe* vaccines against coronavirus



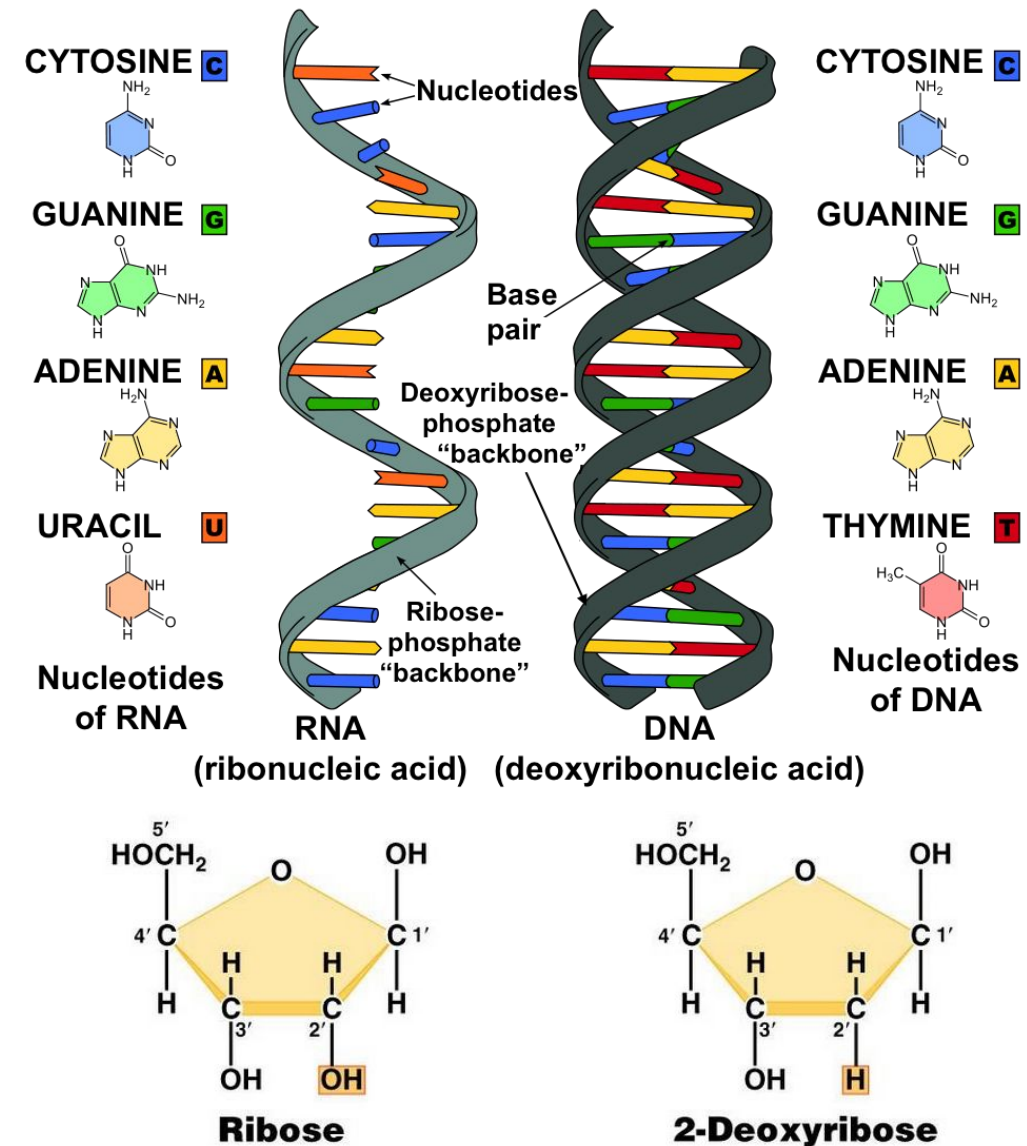
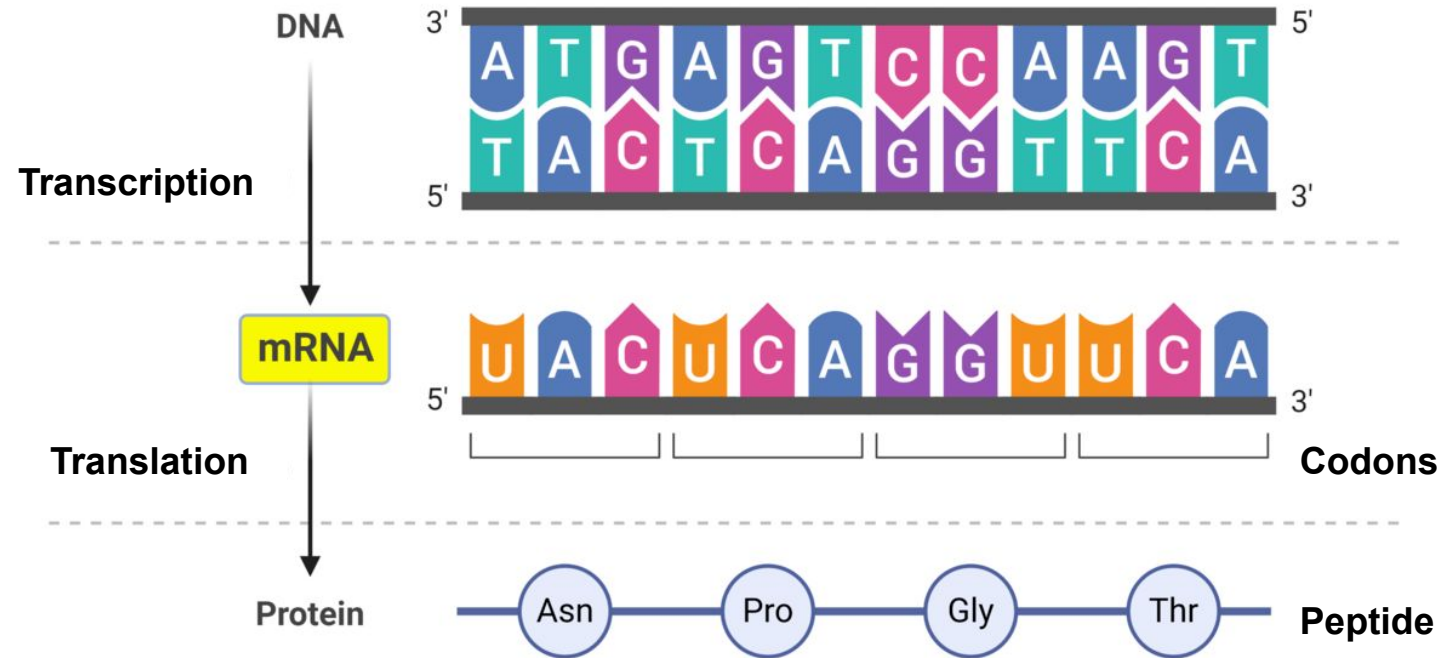
References: Heinz, Franz X., and Karin Stiasny. "Distinguishing Features of Current COVID-19 Vaccines: Knowns and Unknowns of Antigen Presentation and Modes of Action." *Npj Vaccines* 6, no. 1 (August 16, 2021): 1–13. <https://doi.org/10.1038/s41541-021-00369-6>; [Assemblies of putative SARS-CoV2-spike-encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273](#) (github.com/NAalytics); Xia, Xuhua. "Detailed Dissection and Critical Evaluation of the Pfizer/BioNTech and Moderna mRNA Vaccines." *Vaccines* 9, no. 7 (July 3, 2021): 734. <https://doi.org/10.3390/vaccines9070734>.

Offline activities

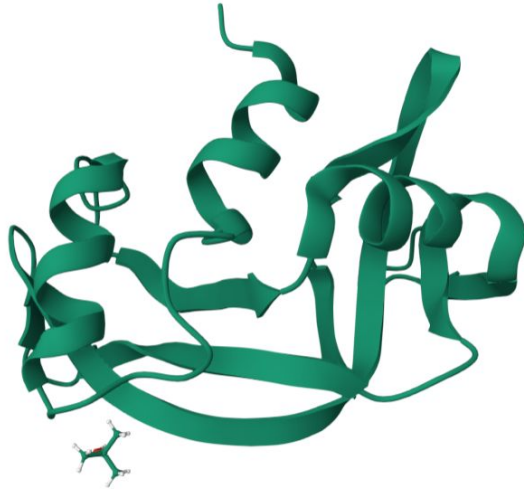
1. Read Wagner, J. et al. A dynamic map for learning, communicating, navigating and improving therapeutic development. Nat Rev Drug Discov 17, 150–150 (2018).
2. Read Bollag et al., 2010, preparing for the lecture next Friday. **This is required by next Friday.**

Backup slides

Proteins are translated from mRNAs, which are transcribed from DNAs



Delivery, stability, and immune responses are essential challenges for mRNA-based vaccines



- mRNAs are too large and charged to pass lipid bilayers.
- **mRNAs are degradable by ribonucleases (RNases).**
- mRNAs are rapidly cleared from liver and kidney.
- *Exogenous* mRNAs induce *immunogenicity*.

		<u>1</u>	<u>10</u>	<u>20</u>	<u>40</u>	
RNase 1	1	---	KESRAKKEFQROHMDSDSSPSSSTYCNQMRNRNMTQGRCKPV			
RNase 2	2	KPPQFTWAQWFETQHINMTSQ-----		QCTNAMQVINNYQRRCKNQ		
RNase 3	3	RPPQFTWAQWFAIQHISLNPP-----		RCTIAMRAINNYRWRCCKNQ		
RNase 4	4	--<QDGM YQRFRLQHVHPEET-GGSDRYCNLMQRRKMTLYHCKRF				
RNase 5	5	--<QDNSRYTHFLTOHYDAKPQ-GRDDRYCESIMRRRGLTS-PCKDI				
RNase 6	6	WPKRLTKAHWFETQHIQPSPL-----		QCNRAMSGINNYTQHKHQ		
RNase 7	7	KPKGMTSSQWFKIQHMQPSPO-----		ACNSAMKNINKHTKRCKDL		
RNase 8	8	KPKDMTSSQWFKTOHVQPSPO-----		ACNSAMSIINKYTERCKDL		
			↓		↓	
		<u>50</u>	<u>60</u>	<u>70</u>	<u>80</u>	<u>90</u>
RNase 1	1	NTTFVHEPLVDVQNVCFQEKVTCKNGQ--		GNCYKSNSSMHITDCRLTNGS		
RNase 2	2	NTFLLTTTFANVVNVCGNPNMTCPSNKTRKNCHHSGSQVPLIHCHLTPPS				
RNase 3	3	NTFLRTTFANVVNVCGNQSIKCPHNRTLNNCHRSRFRVPLHCHDLINPG				
RNase 4	4	NTTFIHEDIWNIRSIKSTTNIQCKNGK--		MNCHEGV--VKVTDCTDRTGSS		
RNase 5	5	NTTFIHGNKRSIKAICENKNG--		NPHRE-NLRISKSSFQVTTCKLHGGG		
RNase 6	6	NTFLHDSFQNVAAVCDLLSIVCKNRR--		HNCHQSSKPVNMTDCRLTSG-		
RNase 7	7	NTFLHEFPSSVAATCQTPKIACKNGD--		KNCHQSHGPPVSLTMCKLTSG-		
RNase 8	8	NTFLHEFPSSVAITCQTPNIACKNSC--		KNCHQSHGPPMSLTMGELTSG-		
				▲		▲
		<u>100</u>	<u>110</u>	<u>120</u>		
RNase 1	1	--RYPNCAVRTSPKERHIIIVACE-----		GSPYVPVHFDASVEDST		
RNase 2	2	PQNISNCRYAQTPANMFYIVACDNRDQRRDPPQYPVVPVHLDRII----				
RNase 3	3	AQNISNCTYADRPGRFRFYVACDNRDPR-DSPRYPVVPVHLDTTI----				
RNase 4	4	--RAPNCRYRAIASTRRVVIACE-----		GNPQVPVHFDG-----		
RNase 5	5	--PWPPCQYRATAGFRNVVACE-----		NG---LPVHLDQSIFRRP		
RNase 6	6	--KYPQCRYSAQAQYKFFIVACD-PPQKSDPP-YKLVPVHLDLSIL----				
RNase 7	7	--KYPNCRYKEKRQNKSYVACK-PPQKDSQQFHLVPVHLDRLV----				
RNase 8	8	--KYPNCRYKEKHLNTPYIVACD-PPQQDGP-YPLVPVHLDKVV----				
				↓		

Left: Structure of PDB [7RSA](#). Right: alignment of protein sequences of 8 canonical human RNases (ribonuclease A family). [Sorrentino FEBS Letters, 2010](#).