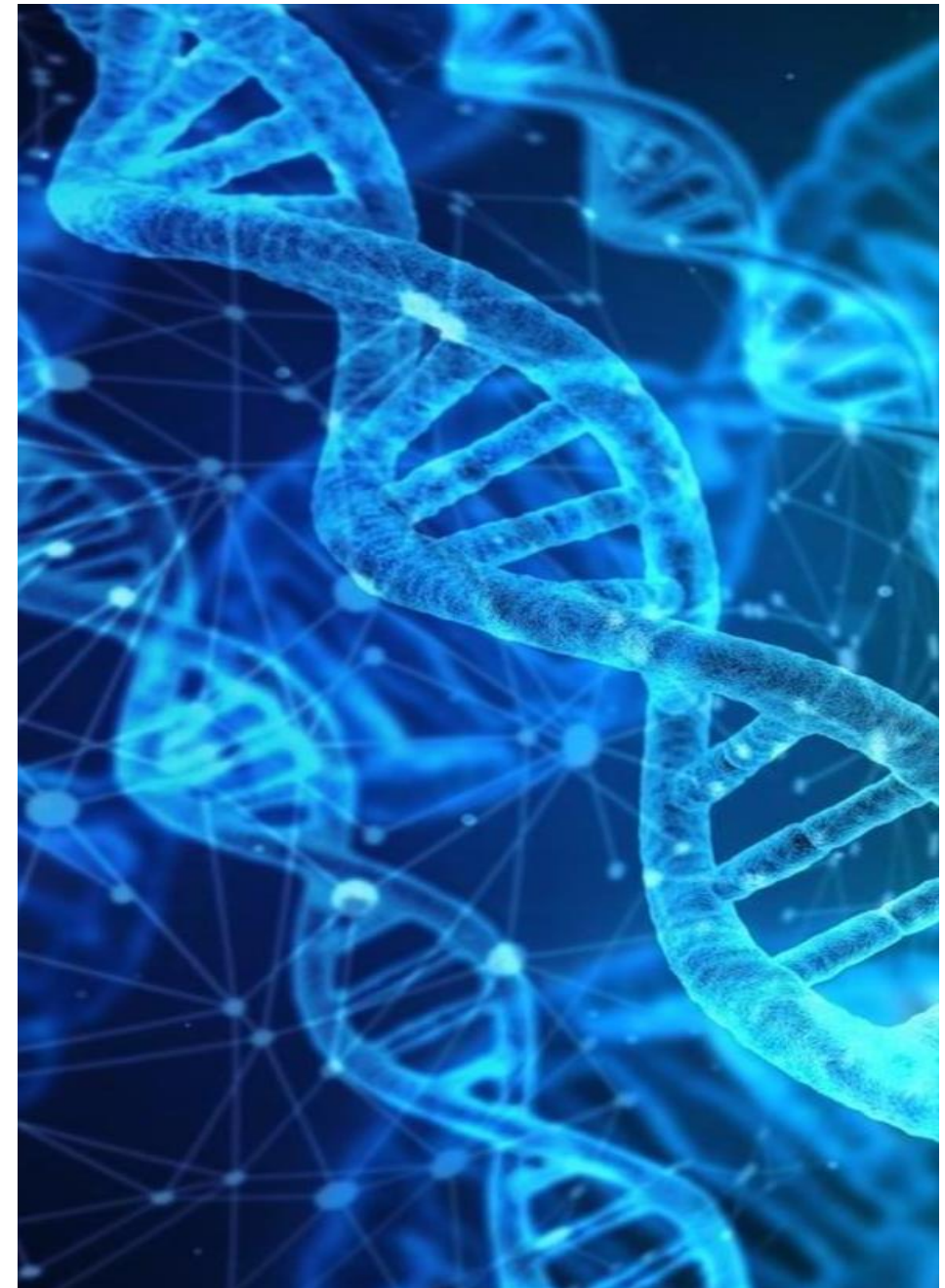


ARTIFICIAL INTELLIGENCE FOR GENOMIC MEDICINE AND HEALTH

Dr. Sara Botti , PhD

JIB Paris - 2 December 2022

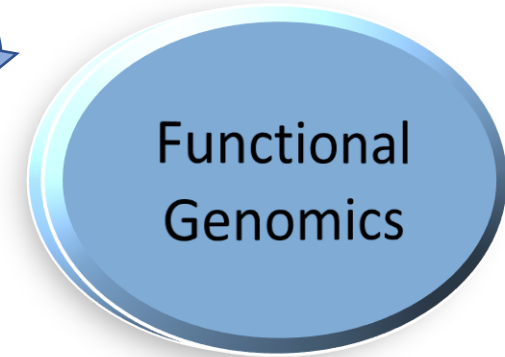
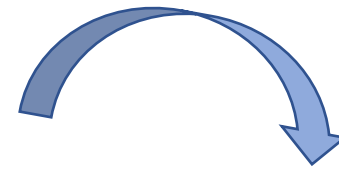
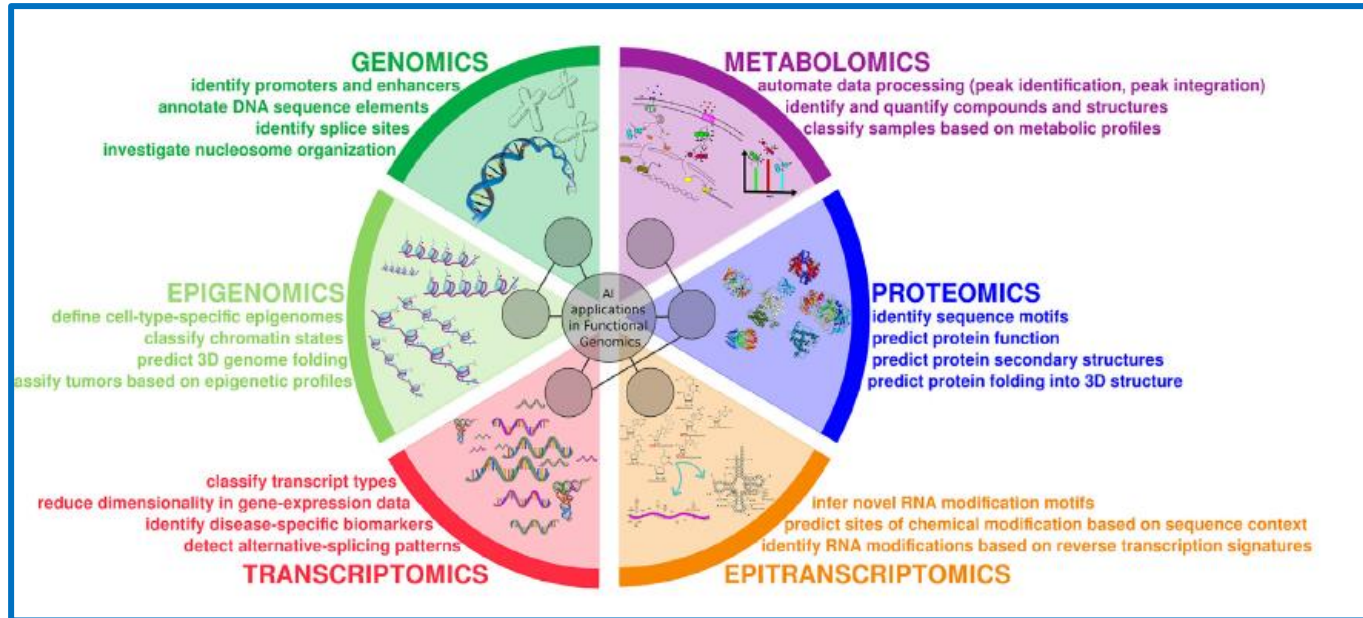


Content of the presentation

- ✓ Omics Sciences
- ✓ Artificial Intelligence (AI): Machine Learning e Deep Learning
- ✓ AlphaFold e RoseTTAFold: milestones for protein structure
- ✓ Application of AI:
 - Innovative vaccine based on AI
 - Identification of genetic diseases from facial images
 - Cancer diagnosis and prognosis
 - Personalized Medicine : Polygenic risk score
 - Microbiome studies
 - Automatic analysis of genetic data (early diagnosis of newborns)
 - Cancer diagnosis, treatment outcome, prognosis

Omic Sciences

The Omics Sciences are independent disciplines aimed at collecting and analyzing large sets of biological data. The data and knowledge produced by them converge in an ambitious goal of Functional Genomics.

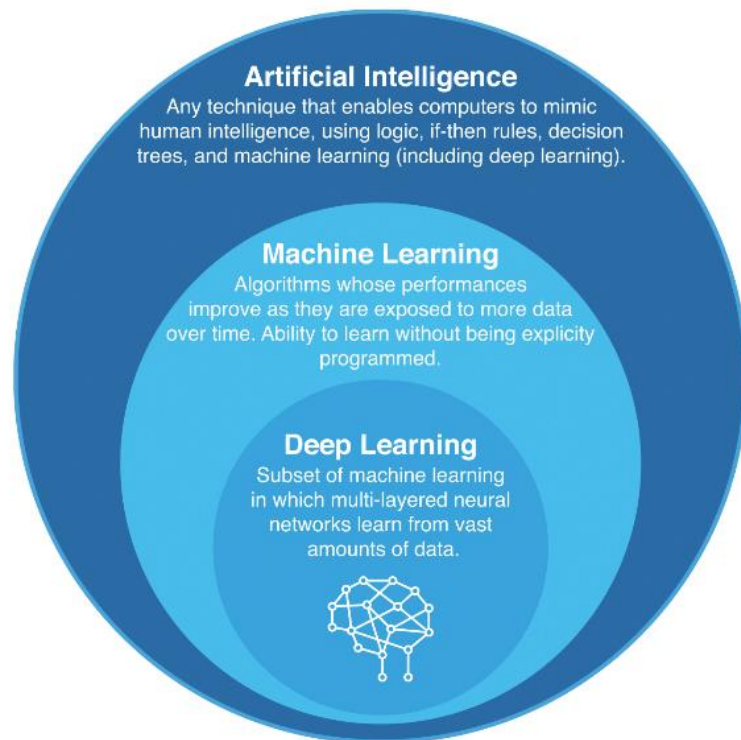


A field of research aimed at characterizing the action and interaction of the main actors (DNA, proteins, metabolites and their modifications) that link the phenotype to the genotype and environmental conditions.

Artificial Intelligence

The study of complex systems using data from Omics Sciences requires advanced tools for analysis, interpretation of results and data integration (complex algorithms, in-depth knowledge in mathematics, statistics, bioinformatics, powerful calculation tools).

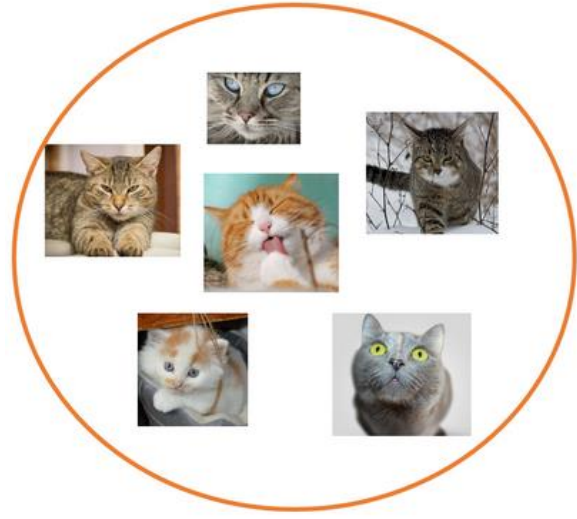
The use of Artificial Intelligence has been encouraged



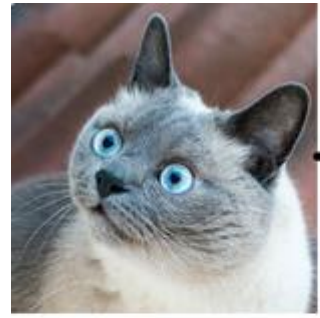
In the field of clinical diagnostics, Artificial Intelligence is any computerized system that can correctly interpret health data especially in the original form as observed by man (i.e. images).

Artificial Intelligence

MACHINE LEARNING: creates algorithm that can learn from data and make decisions based on the observed pattern (i.e. image recognition).



Cat



MACHINE



“cat”



Dog

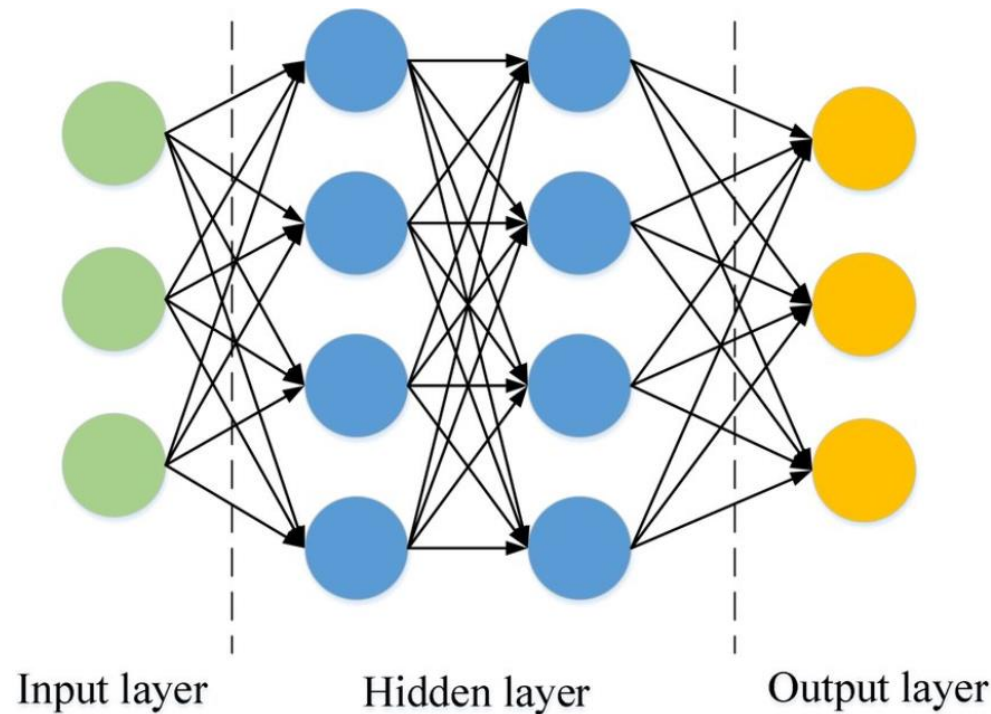


LEARNING

“dog”

Artificial Intelligence

DEEP LEARNING: a machine learning approach that simulates networks of neurons in the human cerebral cortex to reach accurate conclusions without human intervention.
 (e.g. automated guidance, text translation, speech recognition, etc.)

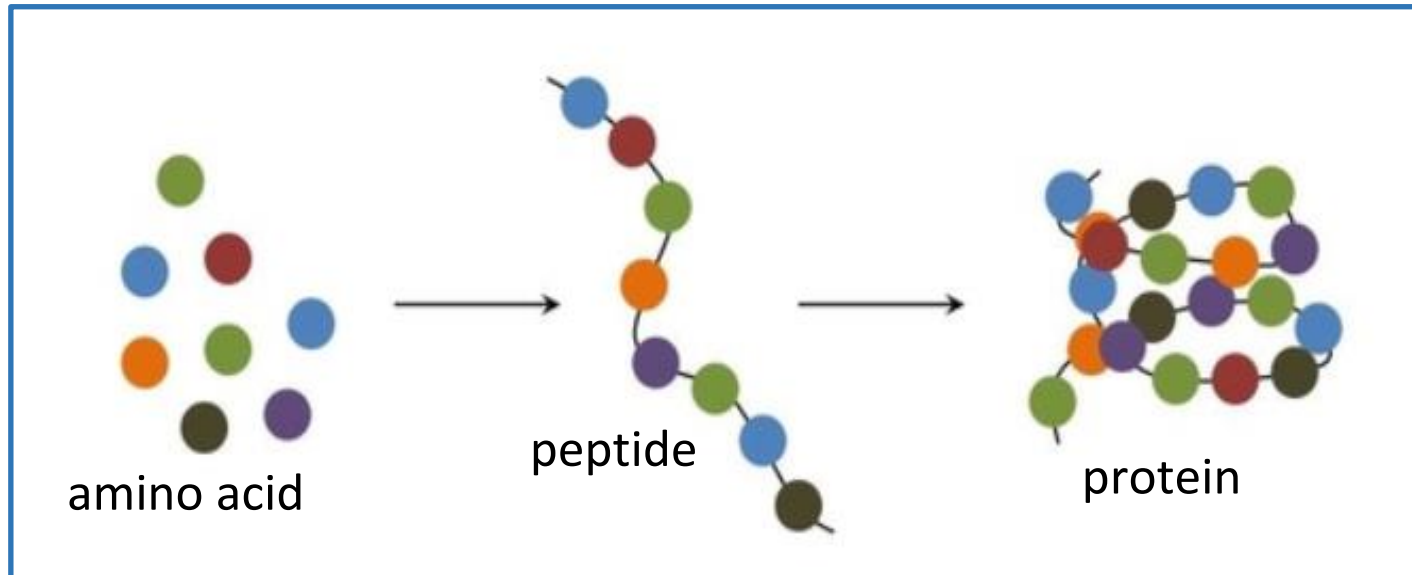


The multiple layers of the neural network are composed of nonlinear cascade units, which have the task of extracting the information to build the classification model. For this reason, this approach to learning is also called non-linear learning.

AI: AlphaFold e RoseTTAFold

Milestones for protein structure

- Proteins are complex molecules that interact with each other and perform the most varied functions (digestion, fighting infections, transmitting messages, perceiving the surrounding environment, etc.). A single cell contains about 42 million proteins (Proteomics)



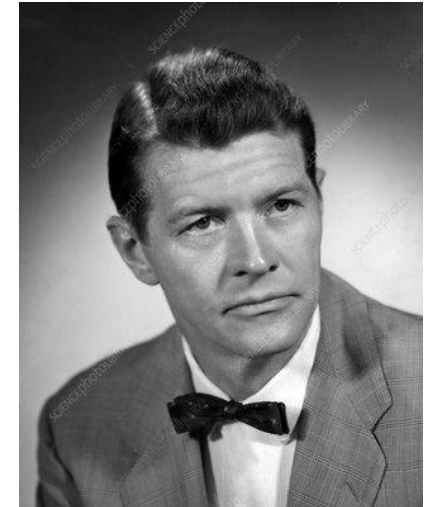
- Several diseases are caused by incorrect protein folding or sequence mutations (Huntington's Korea, Parkinson's, Alzheimer's, Cystic Fibrosis, hemoglobinopathies, etc.). Understanding how protein folding works allows us to study different diseases and design new drugs that can block or replace the wrong proteins.

AI: AlphaFold e RoseTTAFold

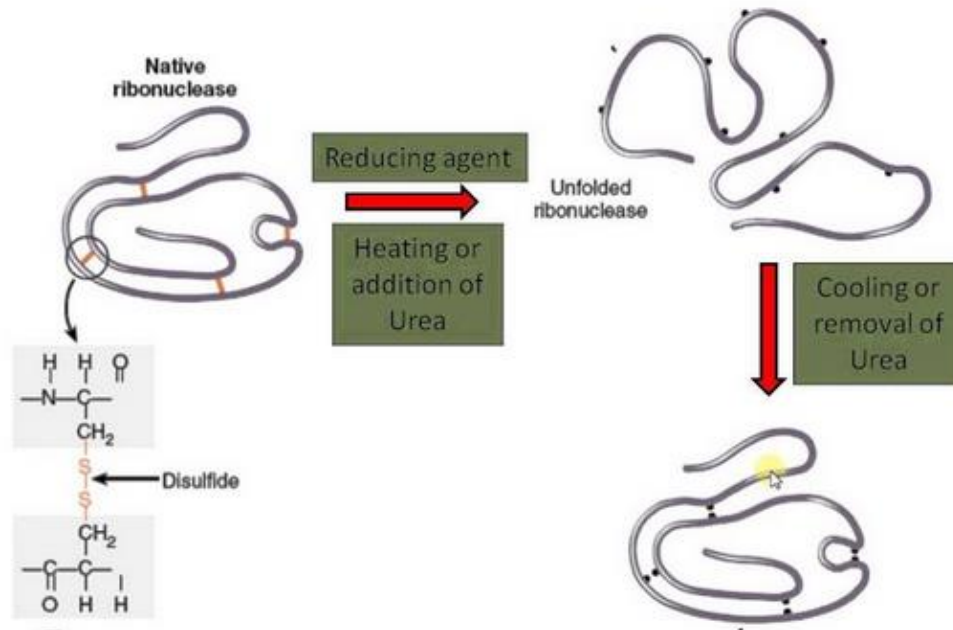
Milestones for protein structure

During his acceptance speech for the Nobel Prize in Chemistry, Christian Anfinsen postulated that, in theory, the amino acid sequence of a protein should completely determine its structure.

This triggered a 50-year quest to predict the structure of proteins from the amino acid sequence.



Christian Anfinsen
Nobel Prize in Chemistry in 1972



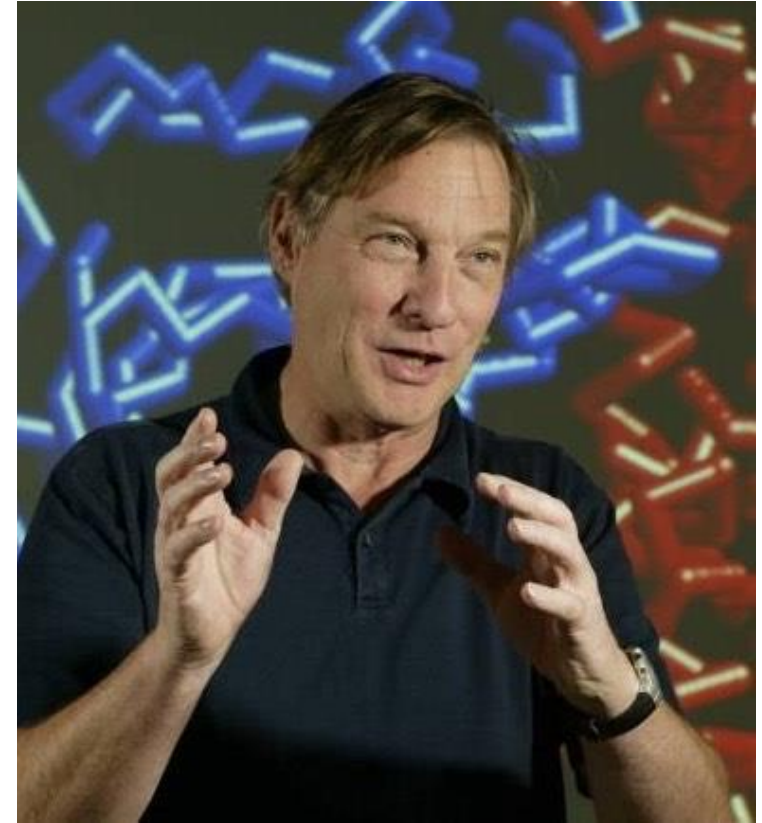
AI: AlphaFold e RoseTTAFold

Milestones for protein structure

- Traditional techniques for defining the 3D structure of proteins, such as nuclear magnetic resonance, X-ray crystallography or the latest cryo-electron microscopy, after numerous trials and errors, required years of painstaking and laborious work for each protein structure and were extremely expensive.
- In 1994, Prof. John Moult and Prof. Krzysztof Fidelis of the University of Maryland launched a competition called CASP (Critical Assessment of Structure Prediction).



Sending to Laboratories of gene sequences of proteins with resolved but unpublished structure.



Prof. John Moult

AI: AlphaFold e RoseTTAFold

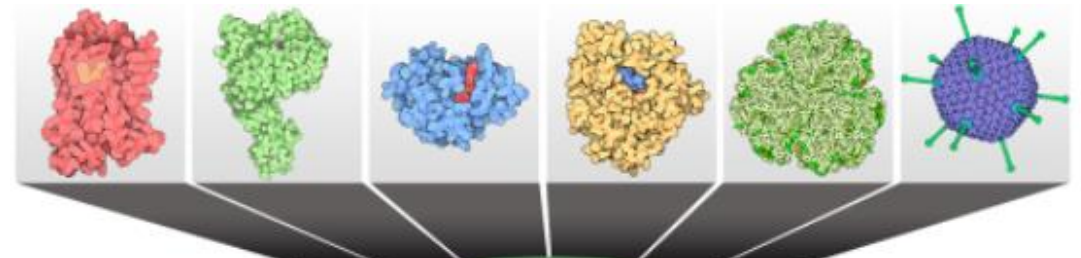
Milestones for protein structure

- In 2018, in CASP13 the best groups achieved a decent result, but were outclassed by a new participant: Google's Deep Mind AI system called AlphaFold achieved an average score of 57% (in second place, Baker's group, who had been working for more than 20 years, achieved a score of 50%).

AlphaFold



DEEP LEARNING



Trained on thousands of protein structures present in the PDB.

It looks for similar existing sequences and estimates the distances of amino acid pairs in unknown sequences (technique called Attention).

AI: AlphaFold e RoseTTAFold

Milestones for protein structure

- In 2020, in CASP14 Google's DeepMind AI system, **AlphaFold2** achieved amazing results with a score of 94%.
- Dr. Baker and his group, inspired by AlphaFold2, developed a new competitive system, **RoseTTAFold** (a software that predicts the structure of proteins quickly and accurately with limited information).



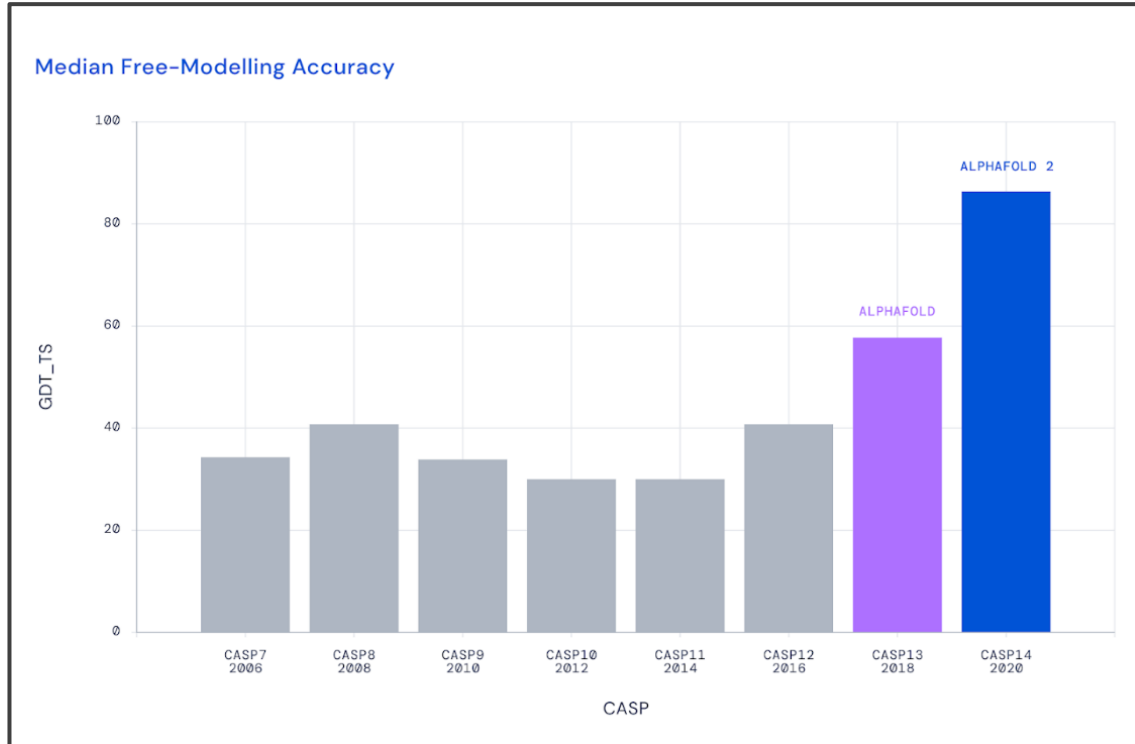
AlphaFold2 and **RoseTTAfold** work on simple and complex proteins allowing to solve the structure of all proteins.



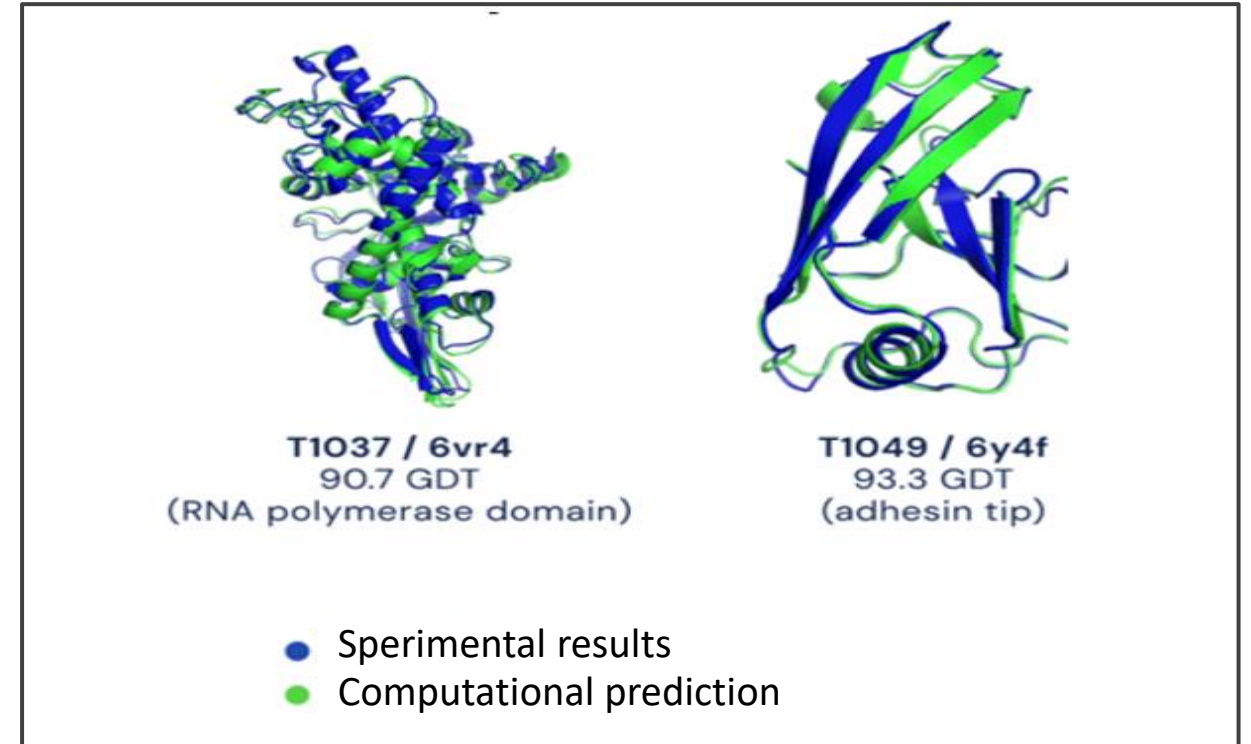
Prof. David Baker-University of Washington

AI: AlphaFold e RoseTTAFold

Milestones for protein structure



AlphaFold accuracy (purple and blue)

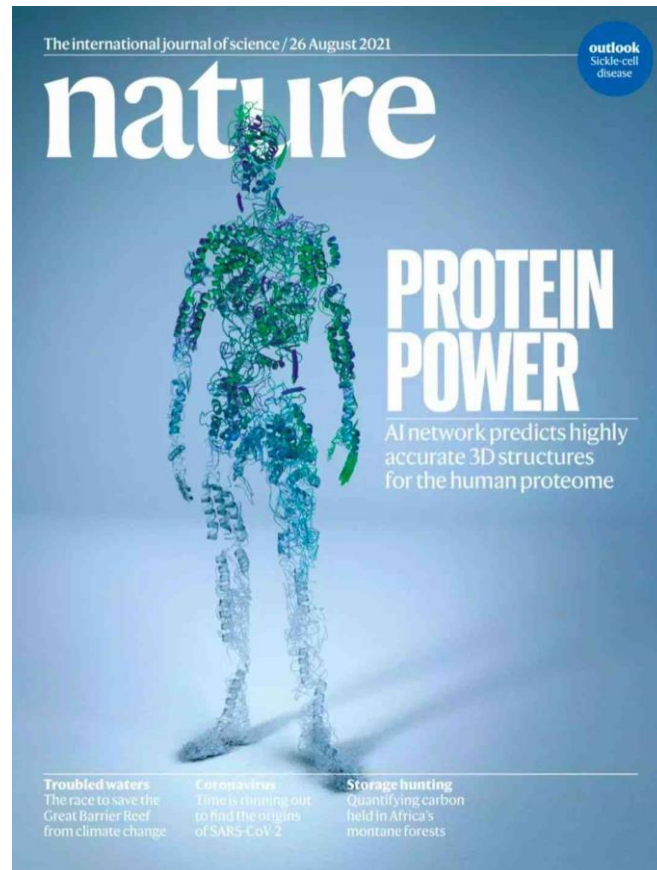


AlphaFold prediction

AI: AlphaFold e RoseTTAFold

Milestones for protein structure

- On July 15 2021, Baker's group and Google's DeepMind published the results of the developed computational systems in Nature and Science, respectively.



AI: AlphaFold e RoseTTAFold Milestones for protein structure

[nature](#) > [news feature](#) > [article](#)

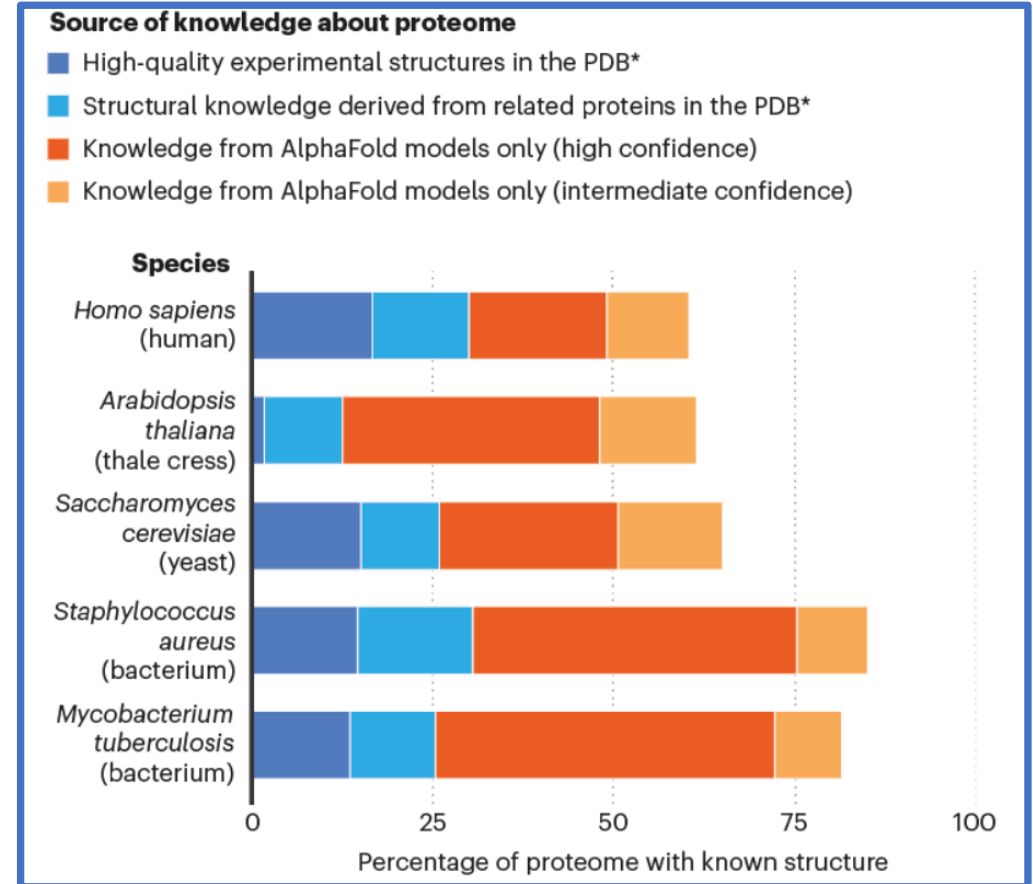
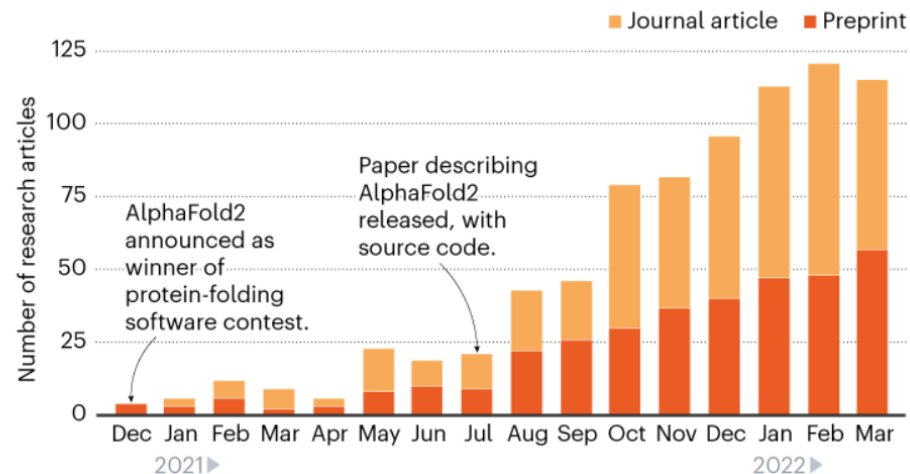
NEWS FEATURE | 13 April 2022 | Correction [25 April 2022](#)

What's next for AlphaFold and the AI protein-folding revolution

DeepMind software that can predict the 3D shape of proteins is already changing biology.

ALPHAFOLD MANIA

The number of research papers and preprints citing the AlphaFold2 AI software has shot up since its source code was released in July 2021*.



More than 200 million predicted protein structures in July 2022

AI: AlphaFill

Enriching AlphaFold models with ligands and cofactors

nature methods

Explore content ▾ About the journal ▾ Publish with us ▾

nature > nature methods > resources > article

Resource | Open Access | Published: 24 November 2022

AlphaFill: enriching AlphaFold models with ligands and cofactors

Maarten L. Hekkelman, Ida de Vries, Robbie P. Joosten & Anastassis Perrakis

Nature Methods (2022) | Cite this article

4171 Accesses | 87 Altmetric | Metrics

Alphafill helps scientists to easily generate new hypotheses for protein function.

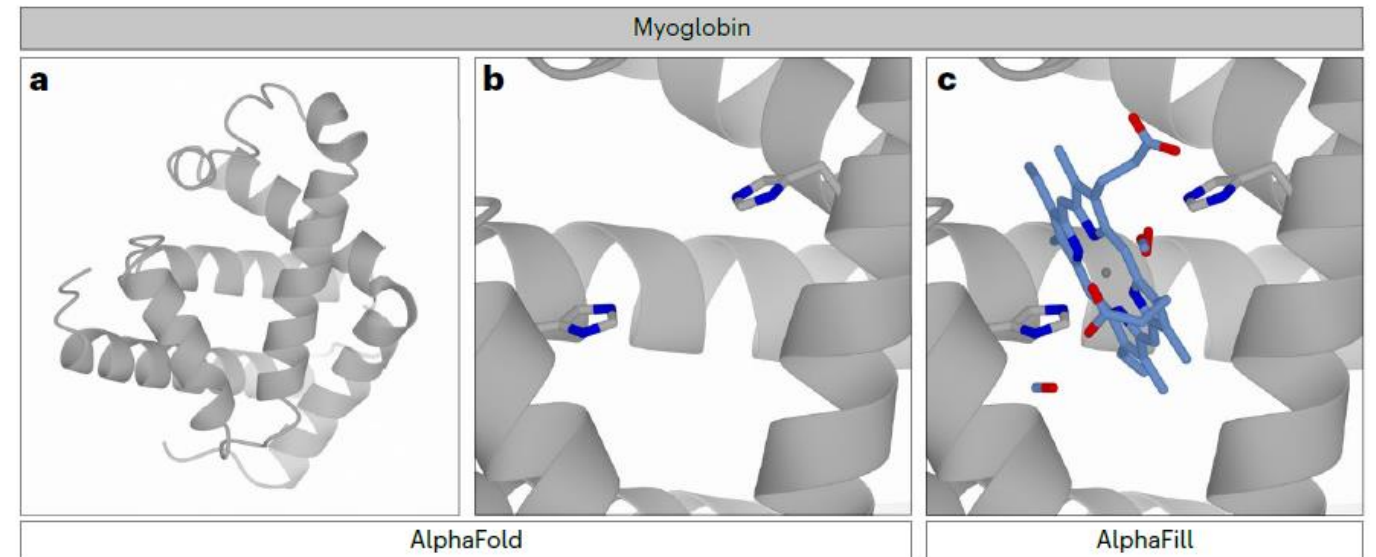
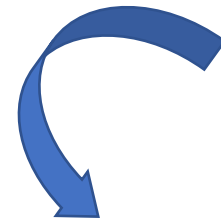


Fig. 3 | Human myoglobin structures in AlphaFold and AlphaFill. **a**, The ribbon diagram of the AlphaFold model of human myoglobin. **b**, The heme-shaped cavity in the AlphaFold model, wherein the histidine side chains (gray cylinders colored by atom type) are ready to facilitate the heme binding. **c**, The heme-

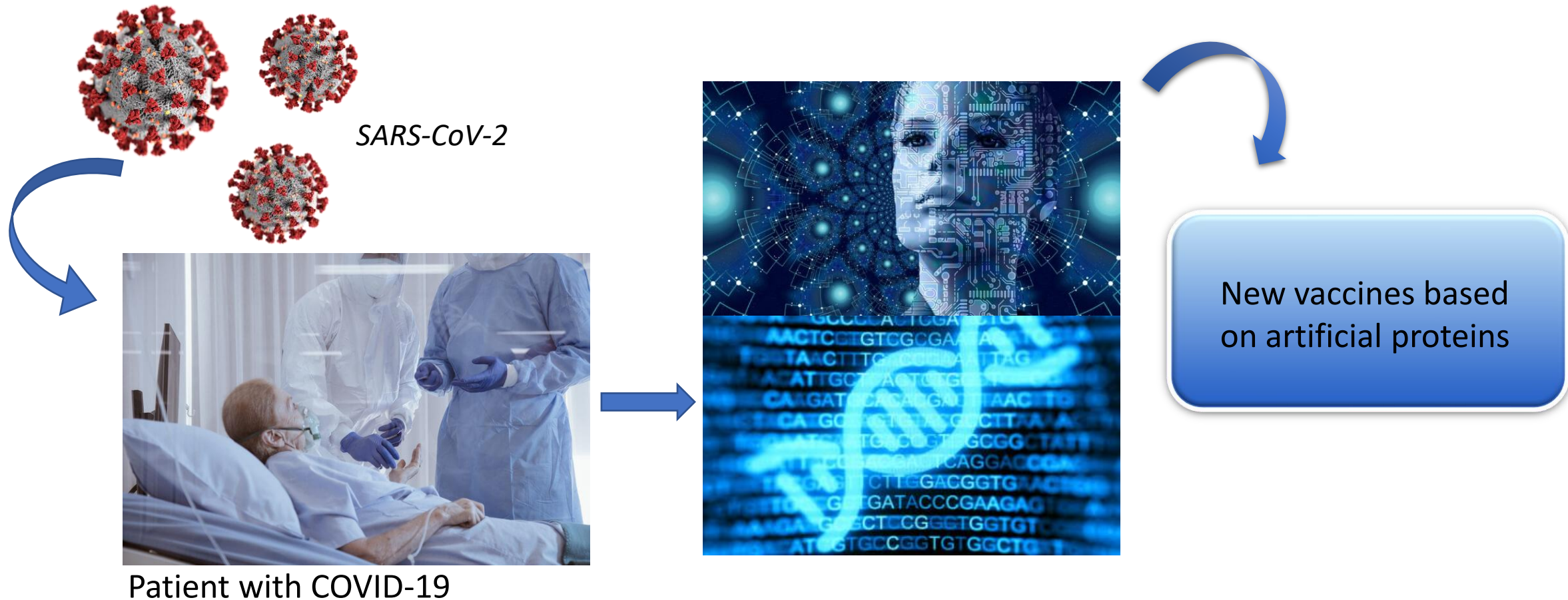
shaped cavity in the AlphaFill model, wherein the binding site is 'filled' with the transplanted heme group and the CO and O₂ ligands; ligands are shown in stick-mode colored by atom type (heme) with the heme iron as a gray sphere.

In the AlphaFill model the heme-shaped cavity is 'filled' with the transplanted heme group and the CO and O₂ ligands.

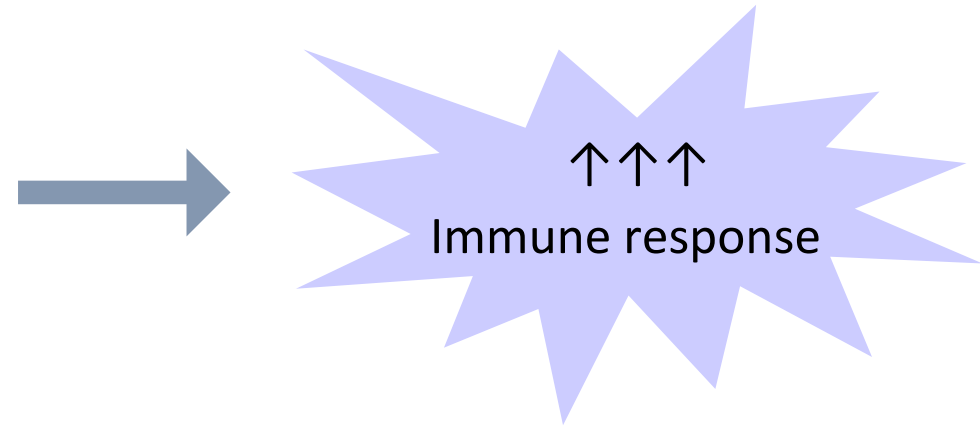
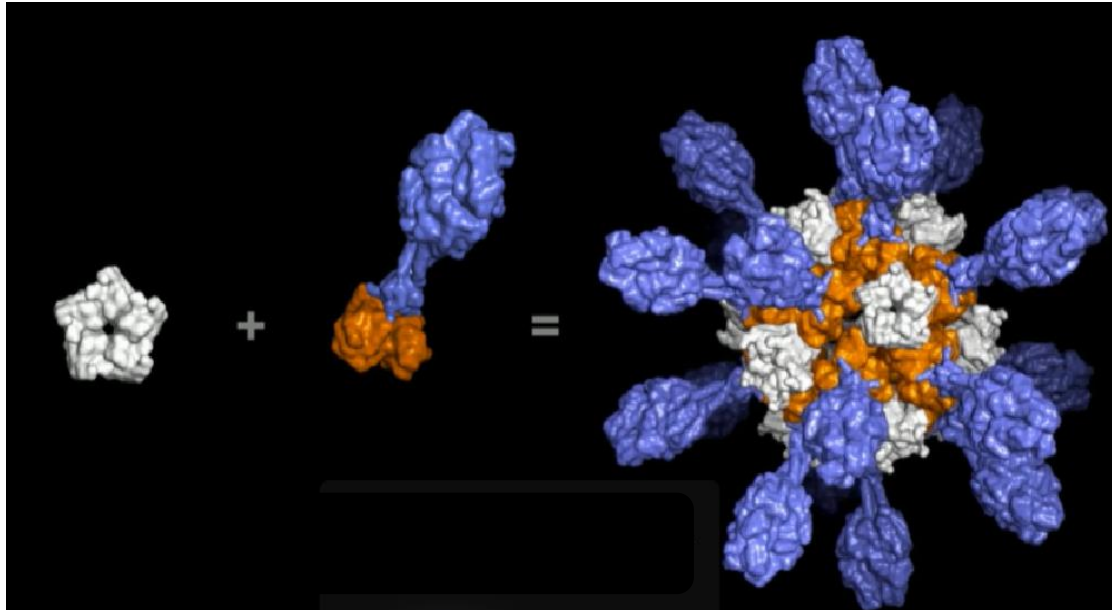


Artificial Intelligence: Innovative solutions to fight COVID-19

New advances in Genomics and Artificial Intelligence create the basis for the development of innovative solutions to fight COVID-19.



Innovative vaccines with proteins obtained from Artificial Intelligence



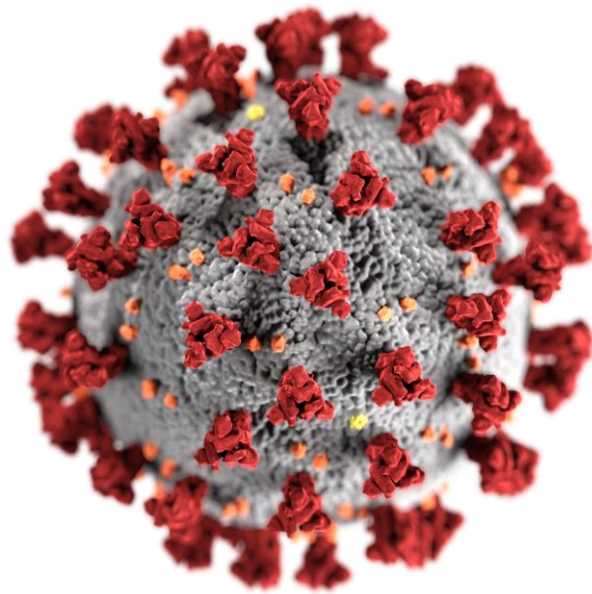
Respiratory Syncytial Virus (RSV) vaccine.
Marcandalli *et al.*, Cell 2019

The researchers created small proteins designed with complementary interfaces that, when mixed in a solution, joined and assembled autonomously into nanoparticles and acquired a spherical shape to resemble a virus.

They coated the nanoparticles with 20 respiratory syncytial virus (RSV) spike proteins.

Innovative vaccines with proteins obtained from Artificial Intelligence

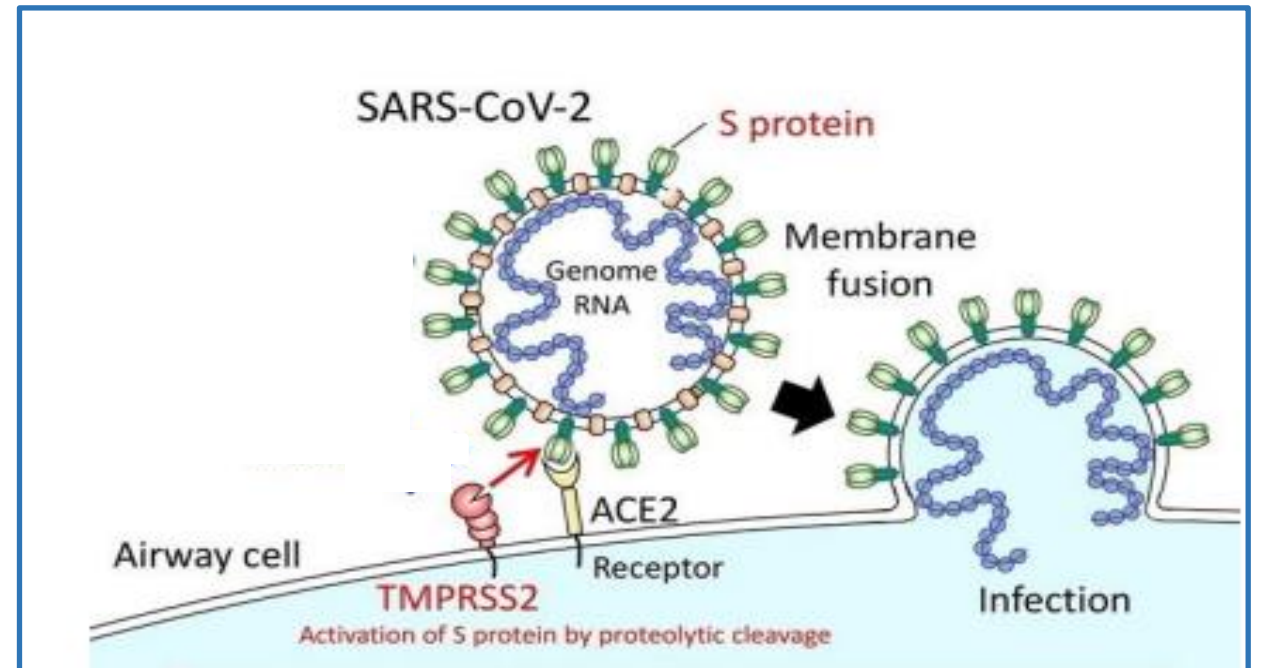
At the end of 2019, mysterious pneumonia occurred in the Chinese city of Wuhan.
Soon the SARS-CoV-2 virus was isolated and the sequence of its viral genome was published.



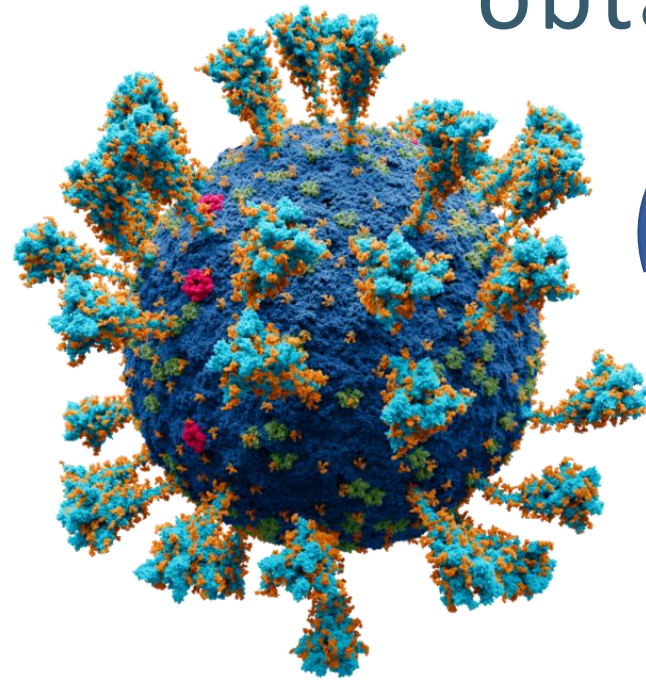
SARS-CoV-2



Studies began to counter the epidemic, which led to the development of innovative vaccines.

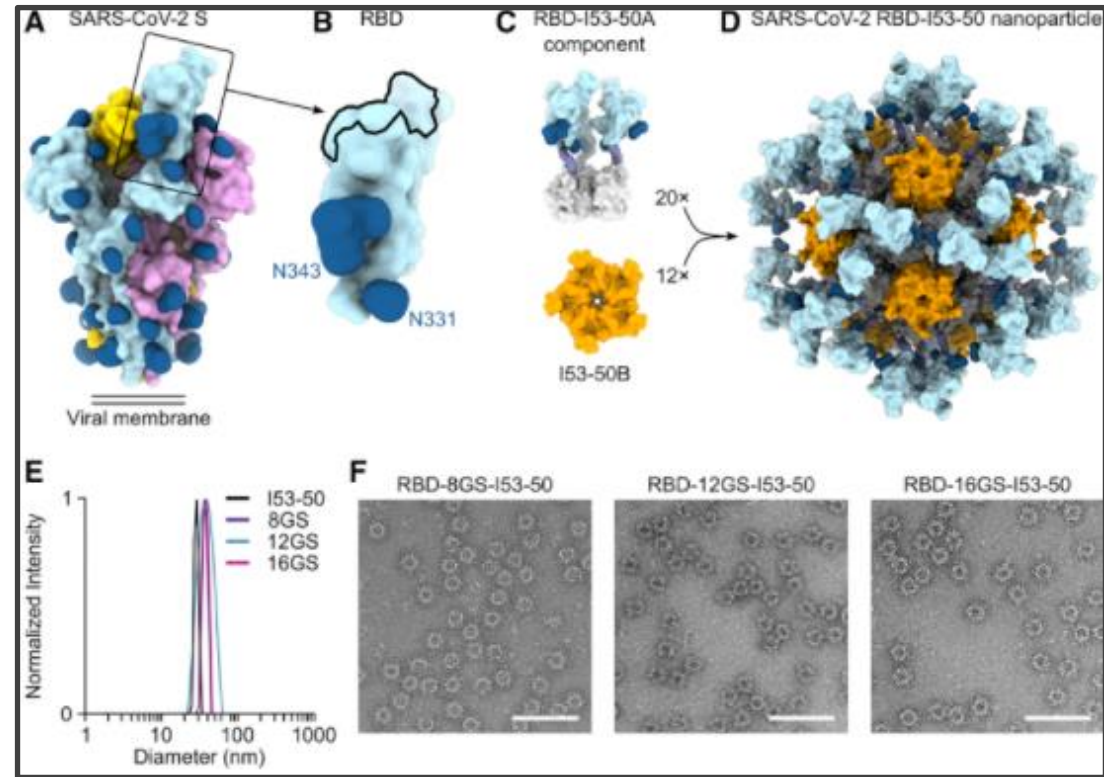


Innovative vaccines with proteins obtained from Artificial Intelligence



SARS-CoV-2

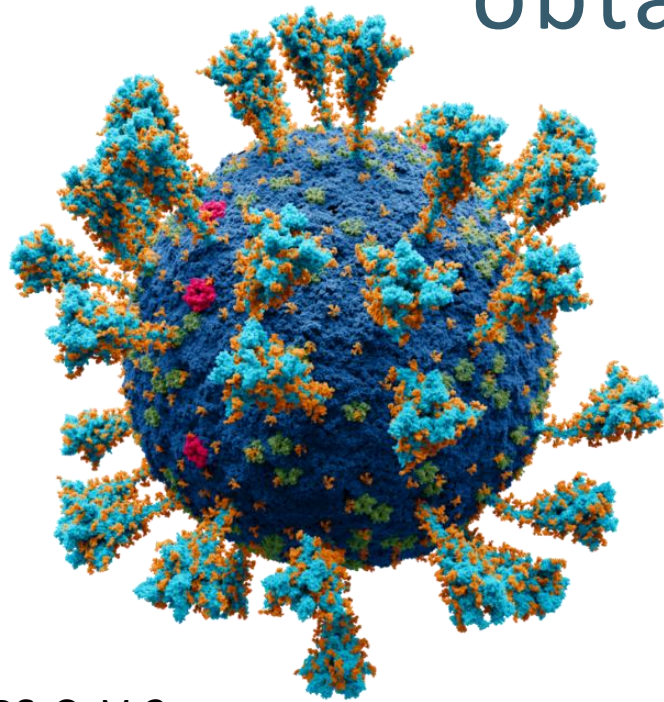
The RDB portion of Spike is inserted into the nanoparticles with a peduncle and takes on a kite-like shape.



Vaccine for SARS-CoV-2, trial NCT05007951, SK Bioscience Co., Ltd
Walls *et al.*, Cell 2020.

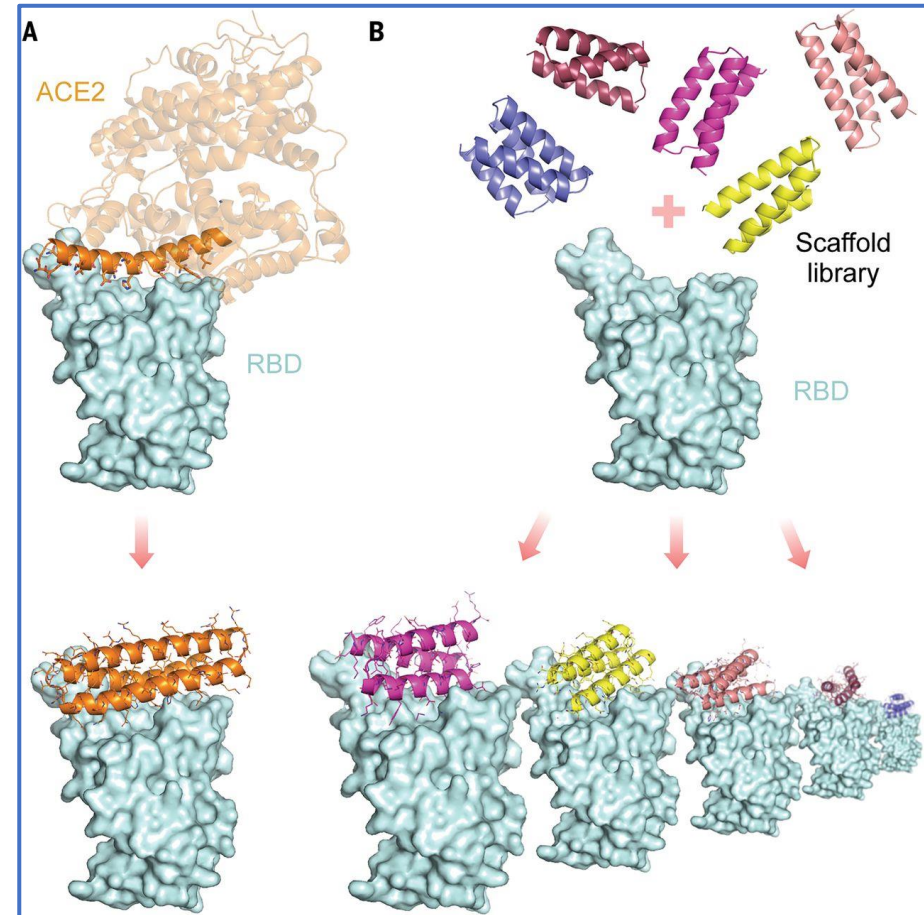
Icosavax, fase1/2 trial, 24 November 2022

Innovative vaccines with proteins obtained from Artificial Intelligence



SARS-CoV-2

Minibinders proteins created on the computer that can bind the RDB portion of the Spike protein

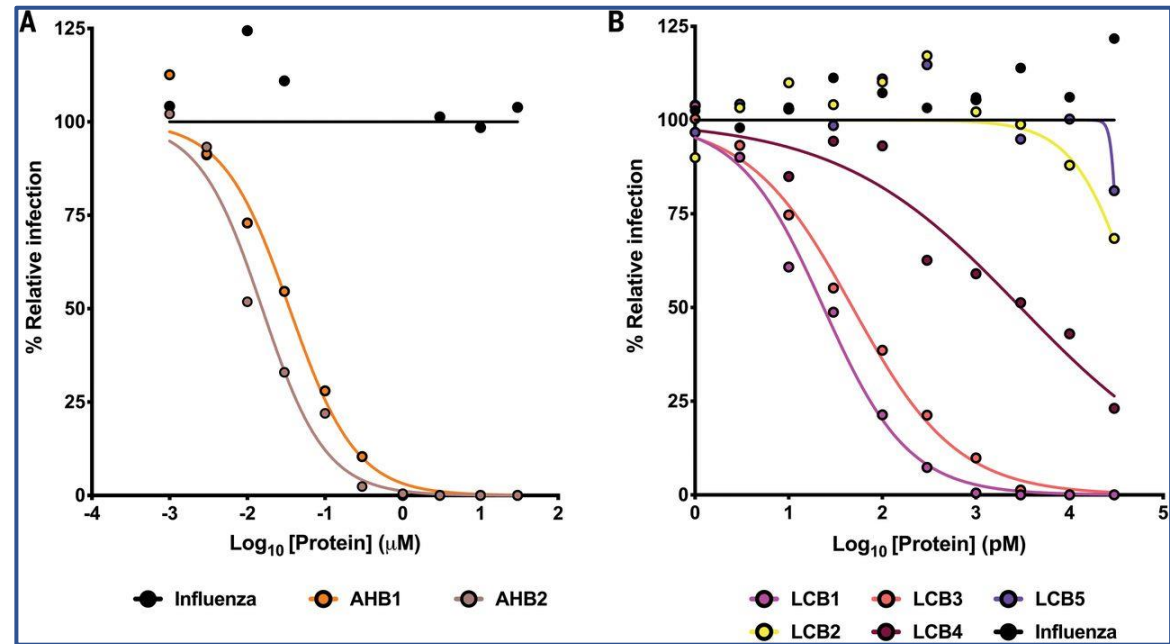


Artificial Minibinders Proteins bound to Spike Protein RBD: A) best protein incorporating ACE2 helix, B) Large-scale de novo design of small helical scaffolds and binding analyses

Innovative vaccines with proteins obtained from Artificial Intelligence

Virus neutralization: Minibinders prevent infection of the virus *in vitro*.

The test allows to identify and select the most efficient Minibinders.

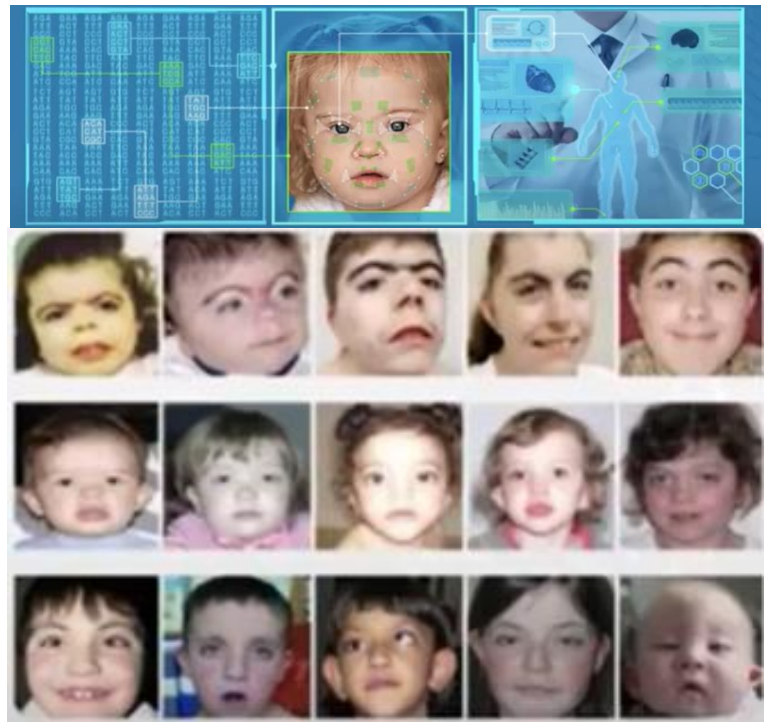
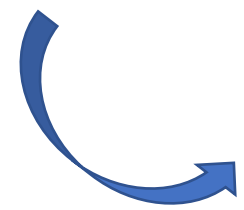


- Minibinders, unlike antibodies, do not require expression in mammalian cells.
- Their small size and high stability allow a formulation for direct administration to the nose or respiratory system (as a spray).
- Effective multivalent minibinders have been developed (*in vitro* and *in vivo* in mice against different variants, (Hunt *et al.*, Sci Trans Med 2022).

Artificial Intelligence: Identification of genetic diseases by facial recognition



Face2Gene



MACHINE LEARNING

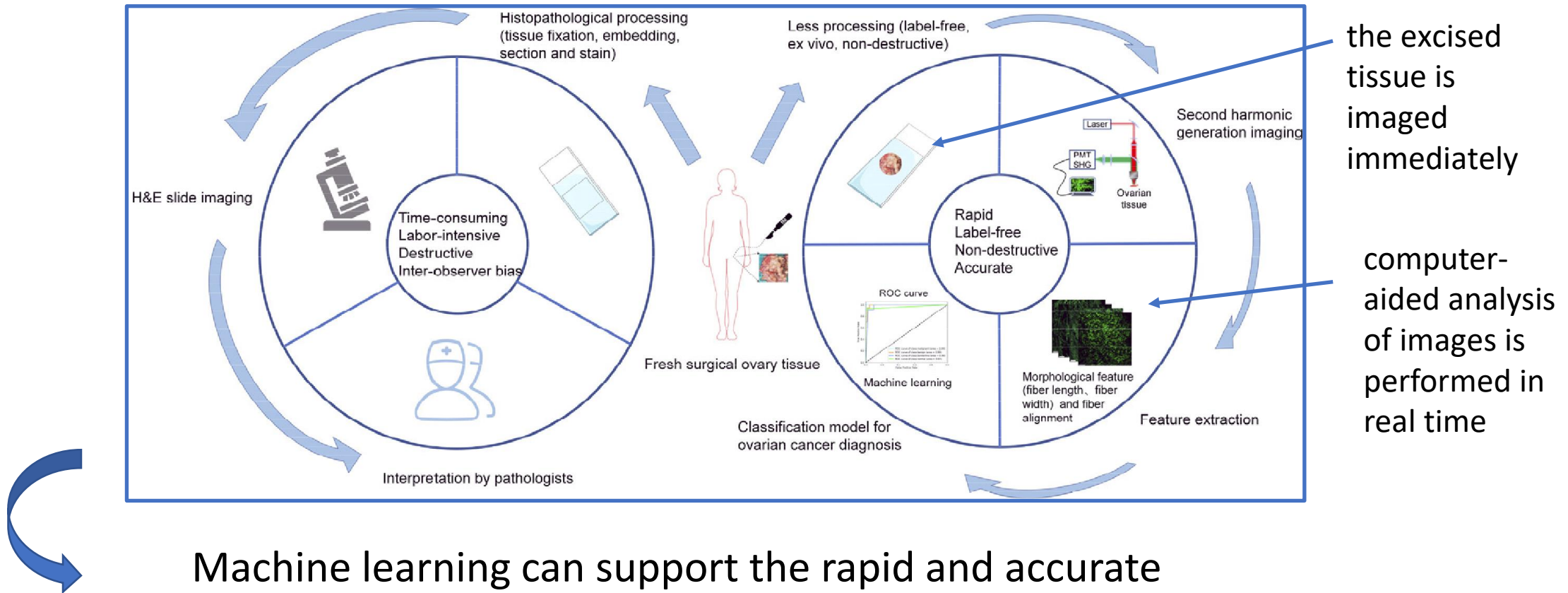


Genetic
syndrome
identification

The researchers provided the algorithm with more than 17,000 images of diagnosed cases of 216 different syndromes.

Artificial Intelligence: Diagnosis of ovarian cancer

They developed a computer-aided diagnosis method to classify ovarian tissues as being malignant, benign, borderline and normal.

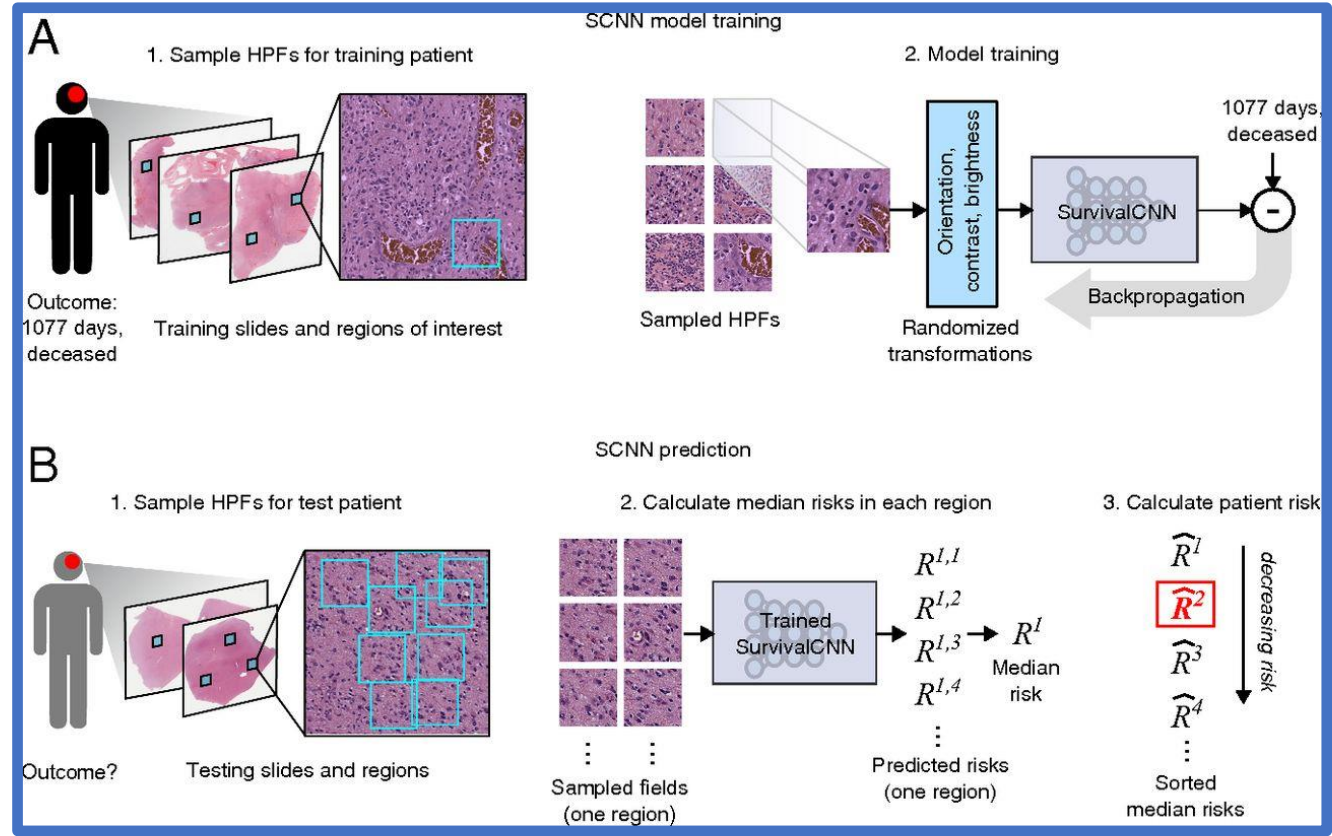


Machine learning can support the rapid and accurate detection of ovarian cancer in clinical practice.

G. Wang et al. *Biom.Op. Ex.* 2021

Artificial Intelligence: Cancer prognosis using images

Training



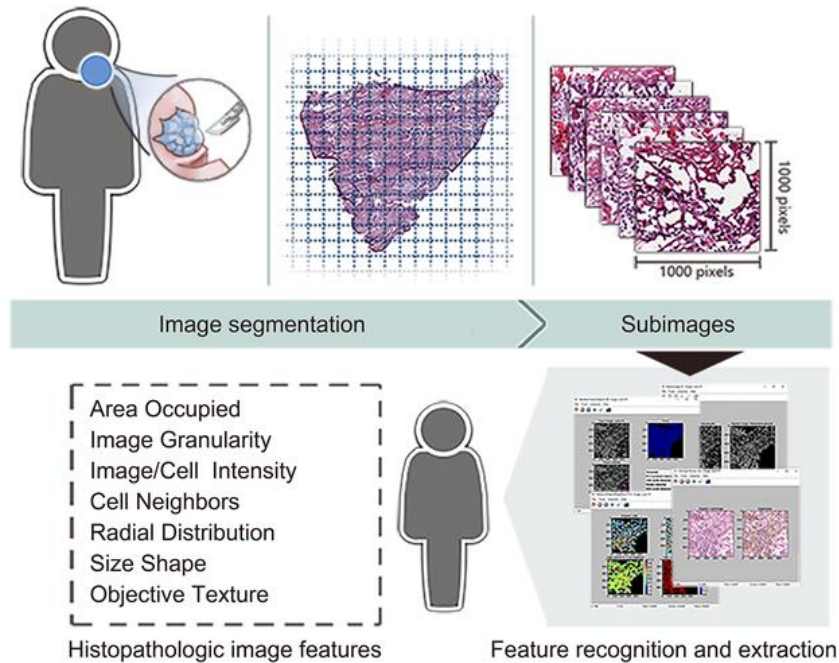
Prognosis



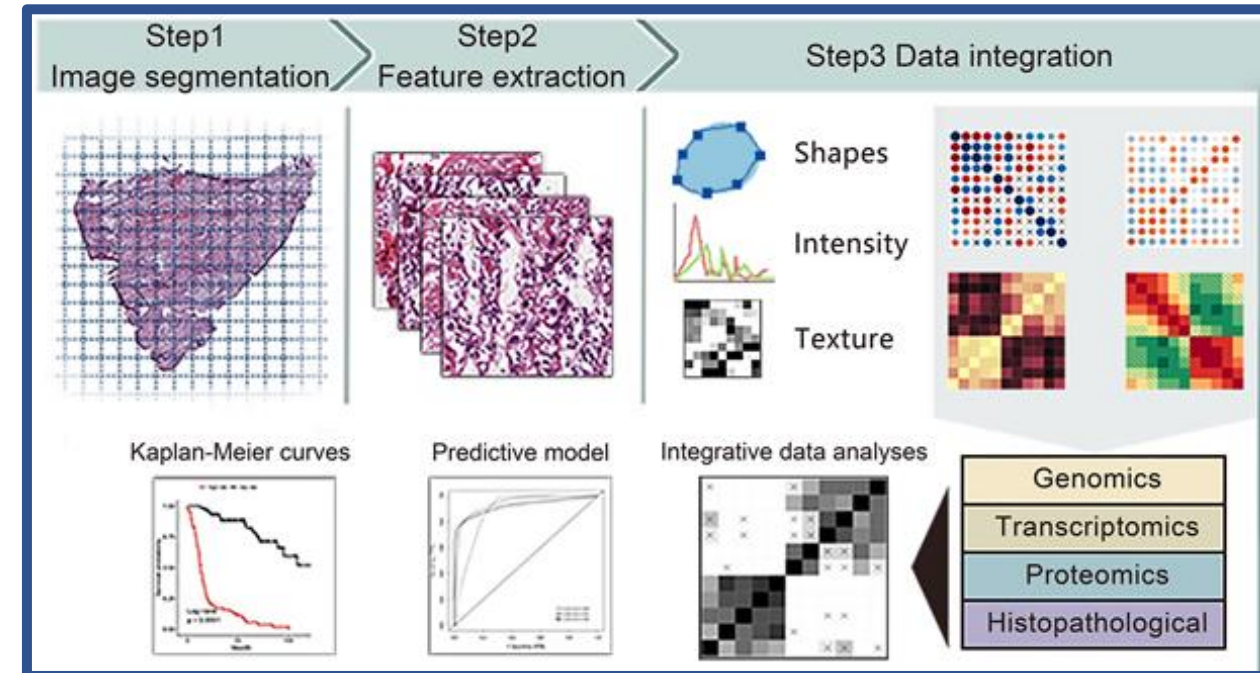
They used an approach called survival convolutional neural networks (SCNNs).

High-power fields (HPFs) images were collected and used to train a deep convolutional network integrated with an hazard model to predict the patient outcome.

Artificial Intelligence: Cancer prognosis with data integration



(A) The histopathological images were analyzed with Next, CellProfiler to obtain mean value of image features related to shape, intensity and texture.



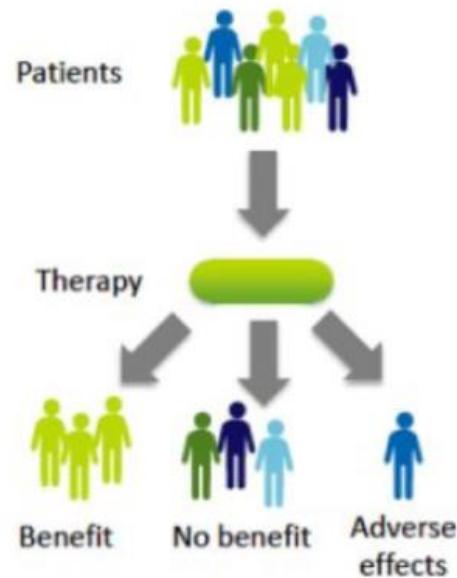
(B) Data integration: genomics, transcriptomics and proteomics.

Improved prognostic models

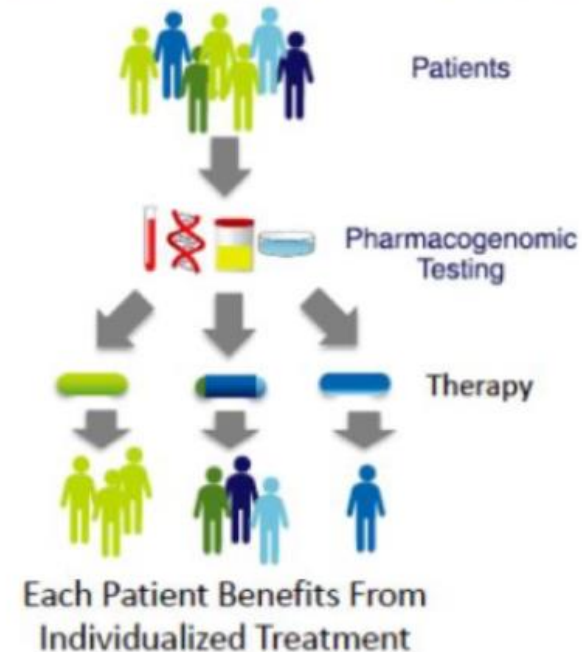
Artificial Intelligence in Precision Medicine

It refers to the application of technologies and processes for the diagnosis, prevention and treatment of diseases in consideration of the individual variability of the genome, the environment and the lifestyle of each person. This approach allows a more accurate prediction of which clinical strategies will be most effective for a given pathology in a specific group of people.

Without Personalized Medicine:
Some Benefit, Some Do Not

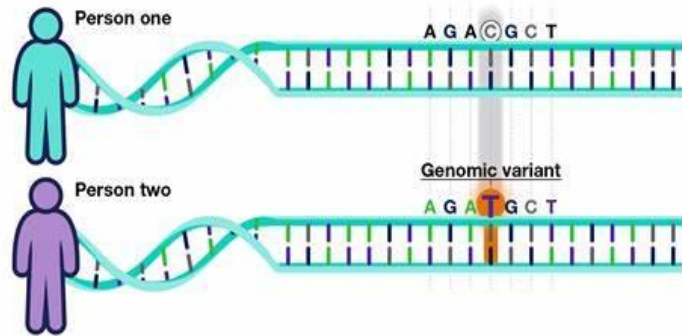


With Personalized Medicine:
Each Patient Receives the Right Medicine For Them



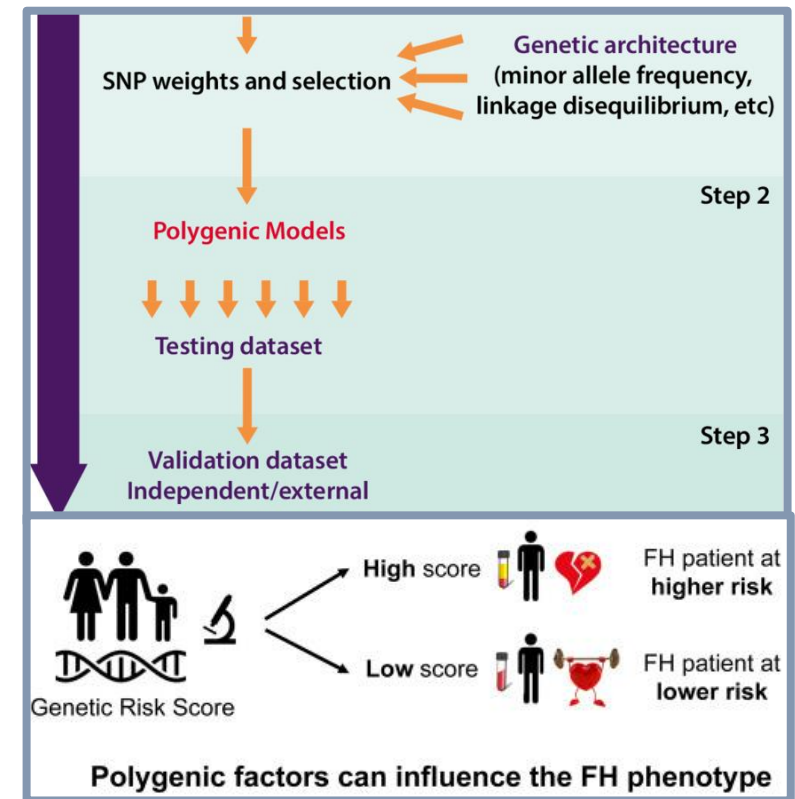
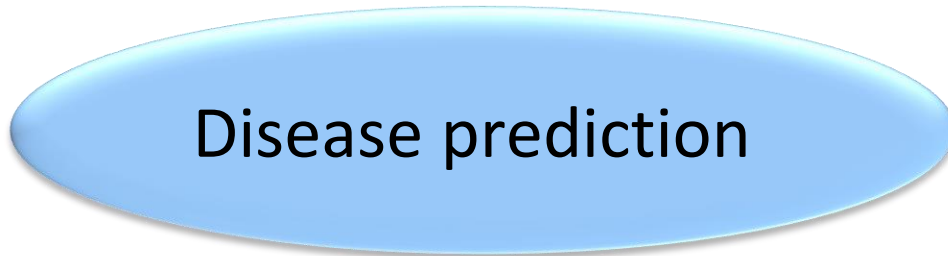
Artificial Intelligence: Polygenic risk score Familial hypercholesterolemia

Familial hypercholesterolemia (FH) is an inherited disease characterized by an excess of cholesterol in the blood. It can be of monogenic origin (LDLR) or polygenic.



➔ Several SNPs

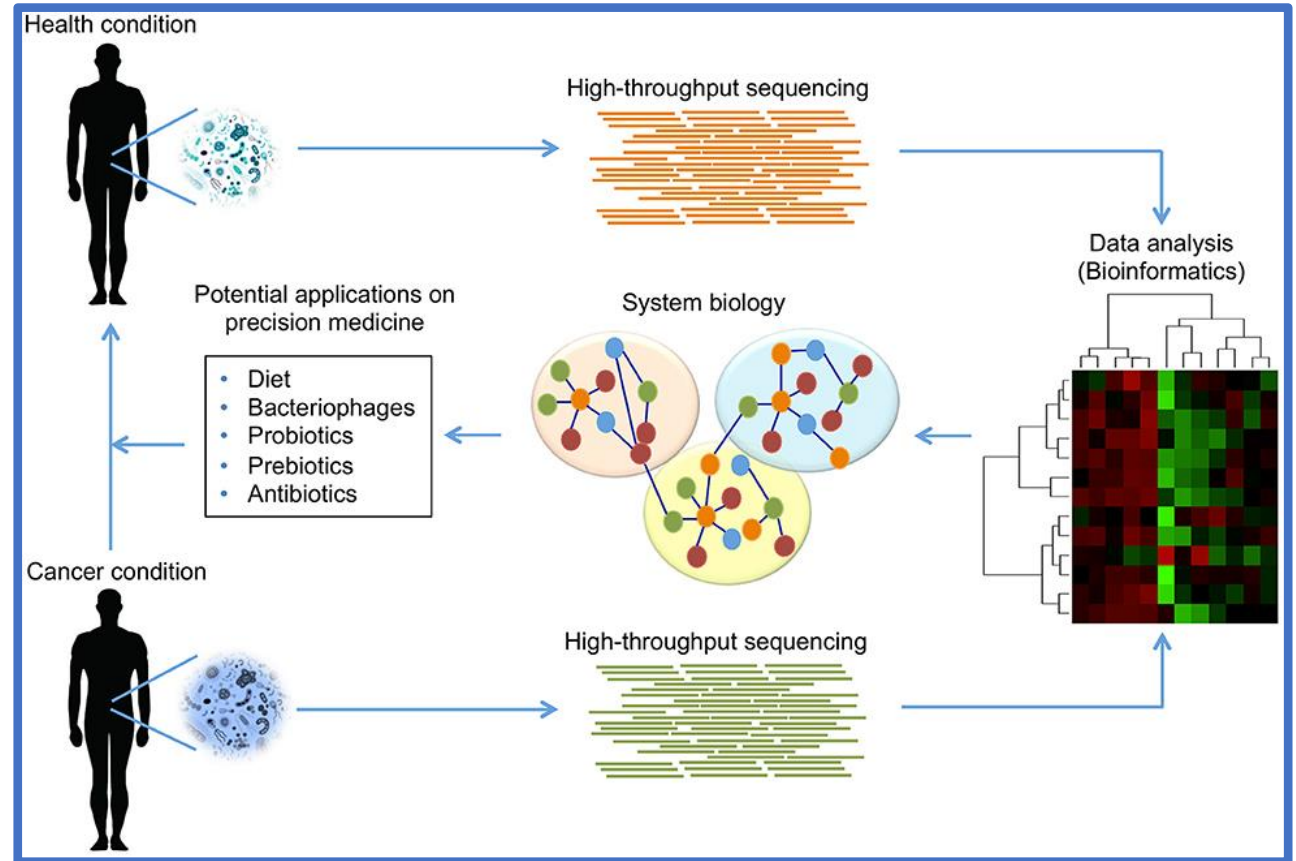
Polygenic risk score (PRS) is an expression of an individual's likelihood of developing a particular medical condition. It is a value obtained by considering the cumulative effect of single polymorphisms (SNPs) each of which has a small effect on the considered trait.



Artificial Intelligence: Microbiome studies

Microbiome studies look at all genetic material within the microbiota.

Machine learning methods are used to classify specific microbial sequences in a sample and investigate the link between dynamic changes in the microbiome and the phenotype of the host or disease.



The evidence suggests that diet, bacteriophages, probiotics, prebiotics and antibiotics can modulate human microbiome to reduce microbial dysbiosis, eliminate pathogenicity in cancer condition, and promote beneficial effects leading a health condition

Artificial Intelligence: Automatic analysis of genetic data

In pediatric intensive care units, a rapid diagnosis that could give access to life-saving care is important. A recent study implemented artificial intelligence techniques to develop a highly automated analysis pipeline to accelerate diagnoses of suspected genetic diseases in seriously ill infants.

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

GENETIC DIAGNOSIS

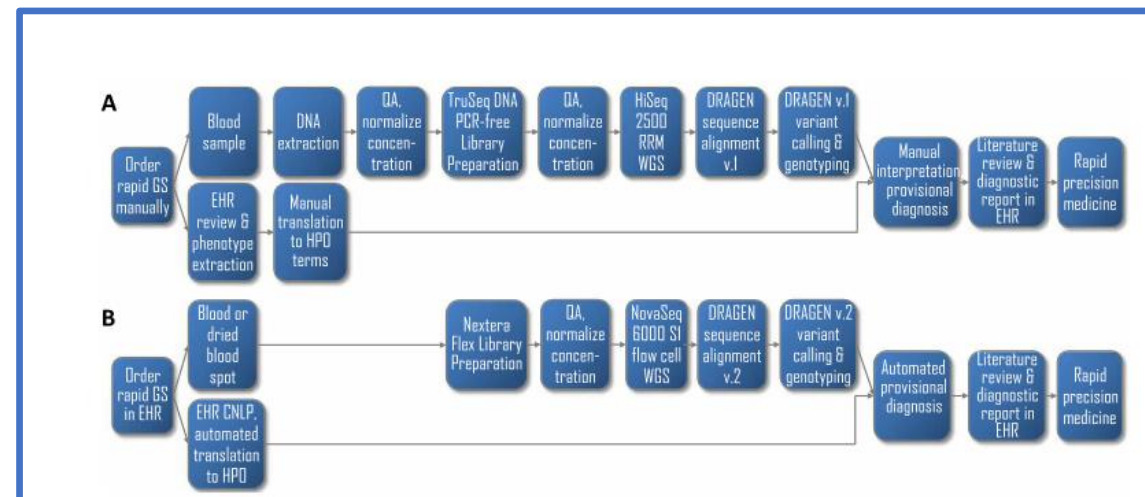
Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation

Michelle M. Clark¹, Amber Hildreth^{1,2,3}, Sergey Batalov¹, Yan Ding¹, Shimul Chowdhury¹, Kelly Watkins¹, Katarzyna Ellsworth¹, Brandon Camp¹, Cyrielle I. Kint⁴, Calum Yacoubian⁵, Lauge Farnaes^{1,2}, Matthew N. Bainbridge^{1,6}, Curtis Beebe⁷, Joshua J. A. Braun¹, Margaret Bray⁸, Jeanne Carroll^{1,2}, Julie A. Cakici¹, Sara A. Caylor¹, Christina Clarke¹, Mitchell P. Creed⁹, Jennifer Friedman^{1,10}, Alison Frith⁵, Richard Gain⁵, Mary Gaughran¹, Shauna George⁷, Sheldon Gilmer⁷, Joseph Gleeson^{1,10}, Jeremy Gore¹¹, Halying Grunenwald^{1,2}, Raymond L. Hovey¹, Marie L. James¹, Kejia Lin⁷, Paul D. McDonagh⁸, Kyle McBride⁷, Patrick Mulrooney¹, Charles Nabel¹, Barbara Oh¹, Albert Orial⁷, Laura Ruskott¹, Zia Sadqi¹, Martin G. Reese¹³

> *Genome Med.* 2021 Oct 14;13(1):153. doi: 10.1186/s13073-021-00965-0.

Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases

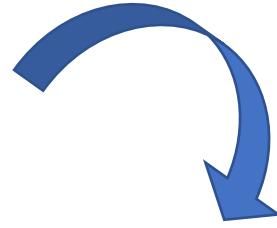
Francisco M De La Vega^{1 2 3}, Shimul Chowdhury⁴, Barry Moore⁵, Erwin Frise¹, Jeanette McCarthy¹, Edgar Javier Hernandez⁵, Terence Wong⁴, Kiely James⁴, Lucia Guidugli⁴, Pankaj B Agrawal^{6 7}, Casie A Genetti⁶, Catherine A Brownstein⁶, Alan H Beggs⁶, Britt-Sabina Löscher⁸, Andre Franke⁸, Braden Boone⁹, Shawn E Levy⁹, Katrin Öunap^{10 11}, Sander Pajusalu^{10 11}, Matt Huentelman¹², Keri Ramsey¹², Marcus Naymik¹², Vinodh Narayanan¹², Narayanan Veeraraghavan⁴, Paul Billings¹, Martin G Reese¹³, Mark Yandell^{14 15}, Stephen F Kingsmore⁴



A Traditional Pipeline (on average 16 days but also 26 hours for 1 sample).

B Fast Pipeline: analysis time 20.10 hours for 6 probandi/run Machine learning incorporating natural clinical language from EHR (electronic health record) extracts phenotypic data and identifies phenotypes associated with genetic disease (retrospective study: 97% sensitivity, 99% accuracy in 95 children with 97 genetic diseases)

Clark *et.al*, *Sci Trans Med*, 2019



Artificial Intelligence is set to play an increasingly important role in precision medicine and clinical diagnostics



Thank you for your attention!!