UNIVERSITÀ DEGLI STUDI DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Intelligenza **A**rtificiale per le **S**cienze della **V**ita

Elena Buscaroli

Cancer Data Science Laboratory
University of Trieste, Italy.

Presentazione modificata di Giulio Caravagna

Contenuti

1. IA e Scienze della Vita, perché?
2. Esempi di Successo
3. Considerazioni

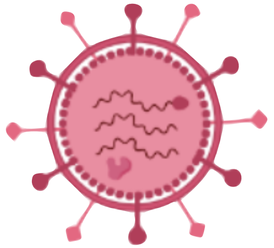
IA e Scienze della Vita, perché?

Cosa sono le Scienze della Vita

Lo studio della vita **in ogni sua forma**, passata e presente (piante, animali, virus, batteri, uomini, …)

- *Biologia*
- *Ecologia*
- *Zoologia*
- *Medicina*
- *Genetica*
- *Epidemilogia*
- ….

Cellule



Virus

Batteri

La “modernizzazione”

Sviluppo tecnologico (IA) **funzionale** alla società

desiderio di comprendere la “natura”
desiderio di vivere meglio, più a lungo

Here Come the Artificial Intelligence Nutritionists

Artificial intelligence application for detecting diseases and pests in horticultural crops

Bart De Witte a Wired Health: “Dati e AI per una sanità più inclusiva e open source”

Five ways AI is saving wildlife – from counting chimps to locating whales

Trial begins of AI scan that could reduce risk of stillbirth and other conditions

Researchers predict population trends of birds worldwide

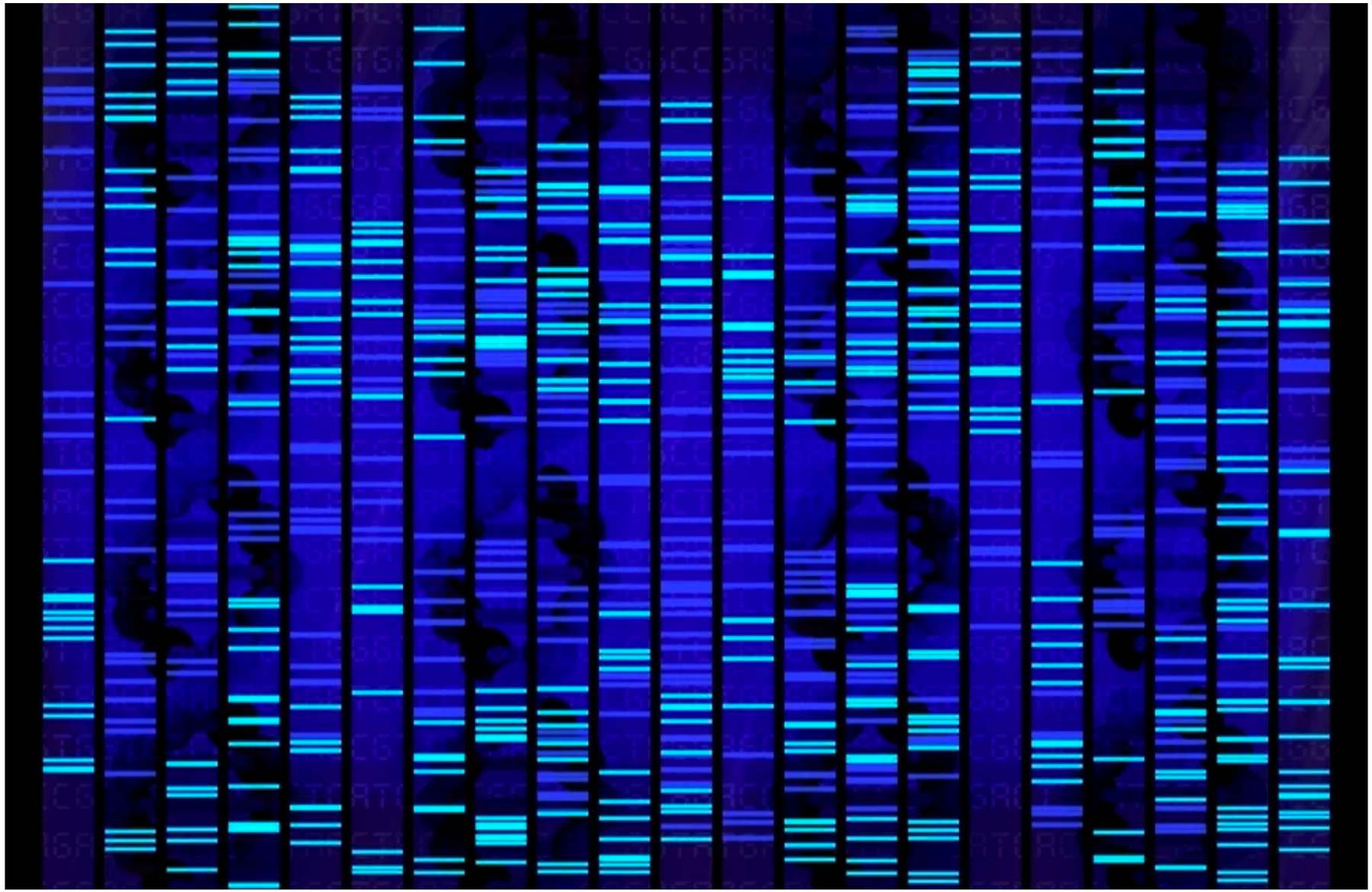
Sanità digitale, un viaggio alla scoperta della medicina del futuro

Using AI to detect antibiotic resistance

'Fingerprint' machine learning technique identifies different bacteria in seconds

How artificial intelligence can help us figure out how life began

La “modernizzazione”: tecnologia

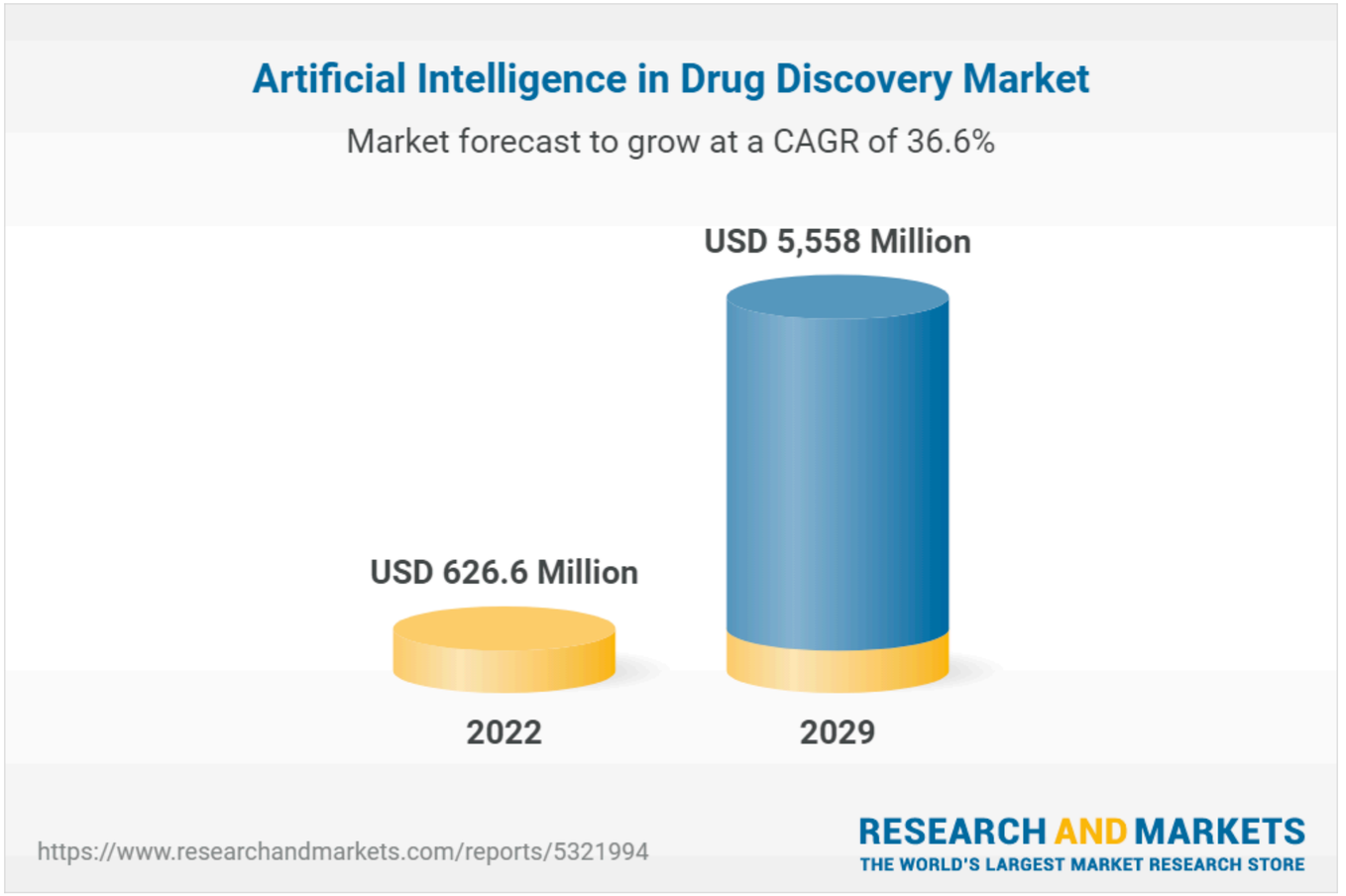


La “modernizzazione”: digitalizzazione

https://www.aamc.org/news-insights/electronic-health-records-what-will-it-take-make-them-work

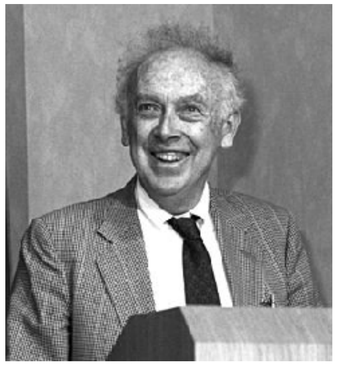
L’opportunità

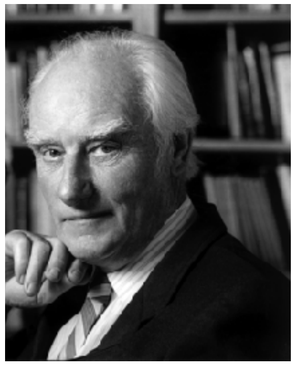
Artificial Intelligence in Retail Market 2021:	2,938.20 USD Million
Artificial Intelligence in Retail Projected Market 2028:	17,086.54 USD Million

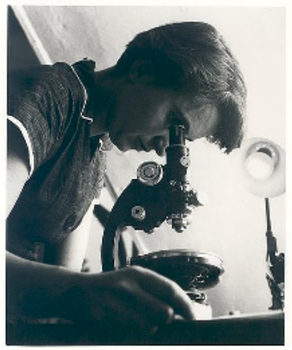


https://www.vantagemarketresearch.com/industry-report/artificial-intelligence-in-retail-market-1362

Che cosa misuriamo?



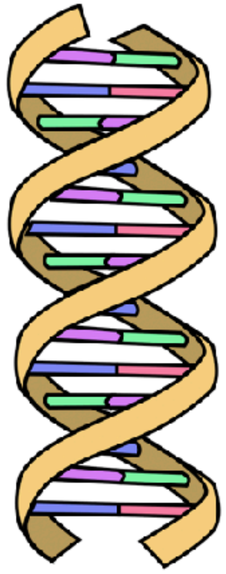




1953

Watson, Crick, Franklin

Struttura del DNA



= Adenina

= Timina

= Citosina

= Guanina

DNA

3.200.000.000 di “nucleotidi” (umano)

ACCTAAGGAGAATACCTAAGAAT …….

- **1 cellula :** ~1.8m (srotolato)
- **1 umano:** 150.000 viaggi Terra-Luna
- **99.6%** uguale fra individui (0.4% sono 12.000.000 di nucleotidi)
- **98,5%** “non-codificante”

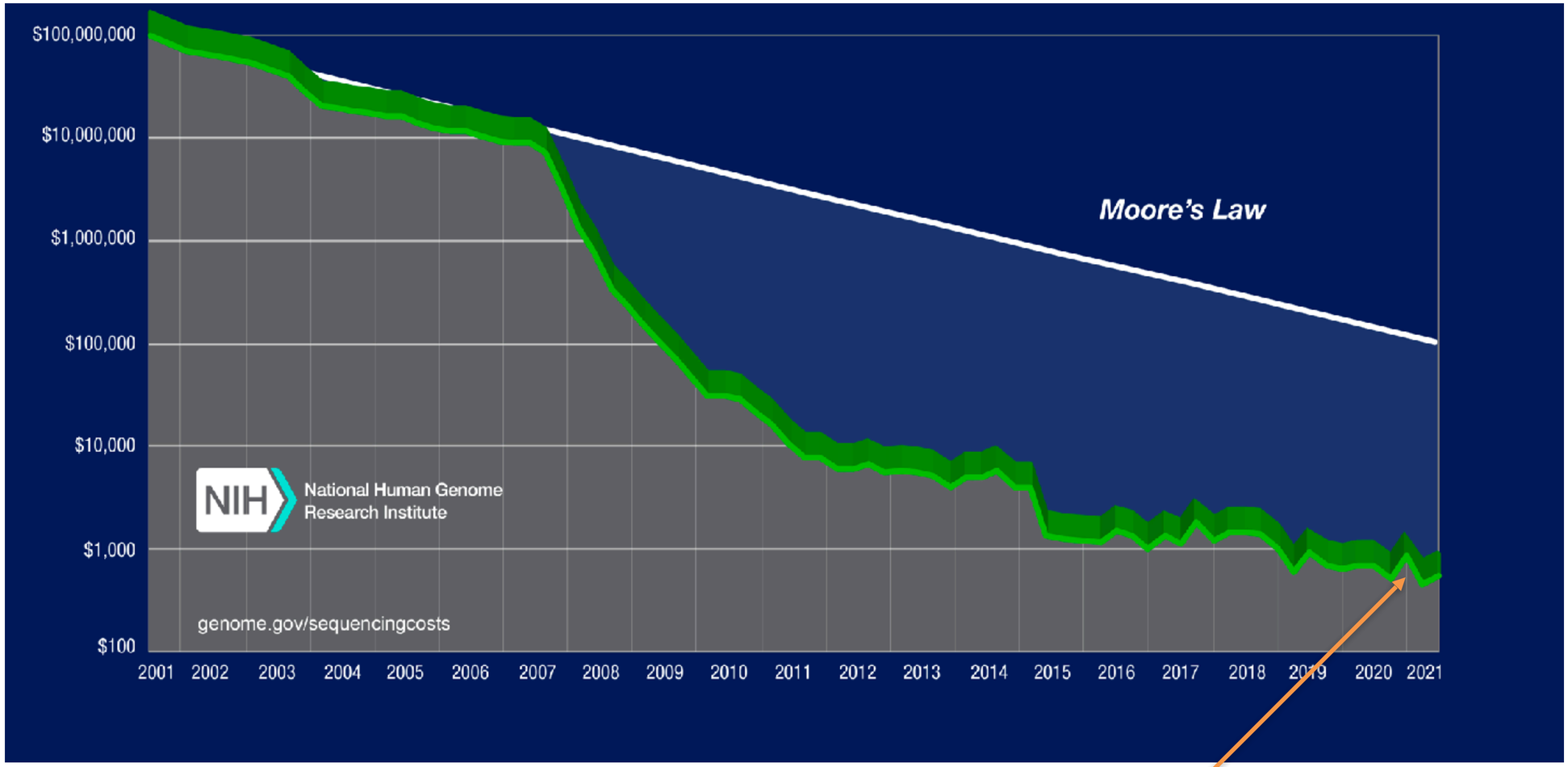
Wikipedia

Human Genome Project

1990-2003

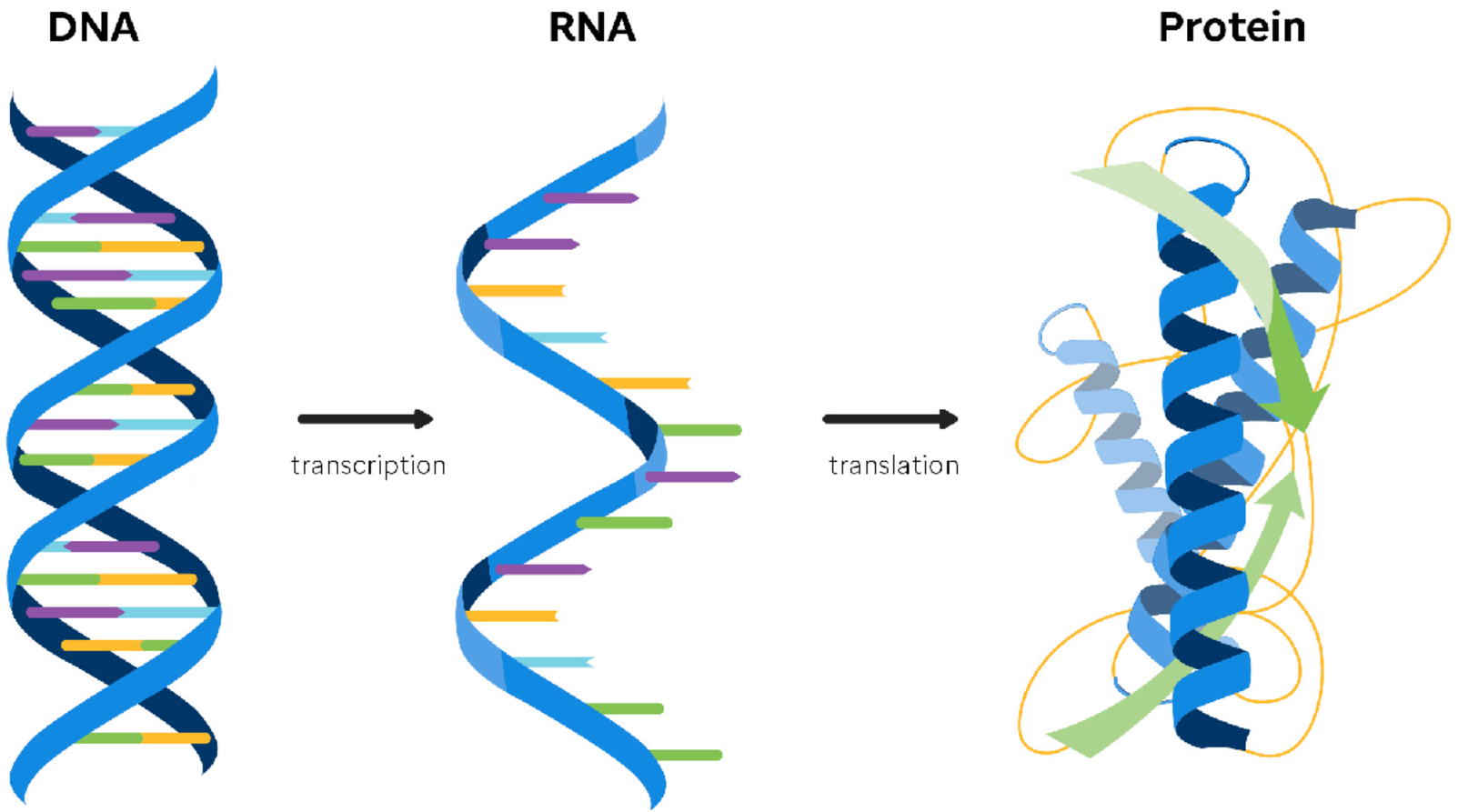
Costo dei dati

“Lettura” di un DNA umano con tecnologia di sequenziamento

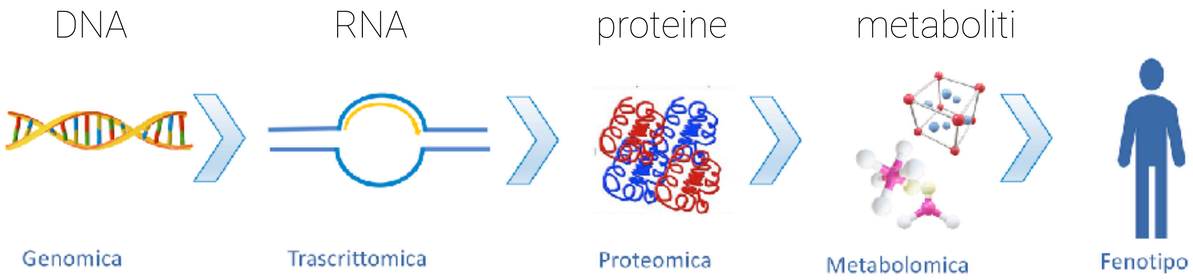


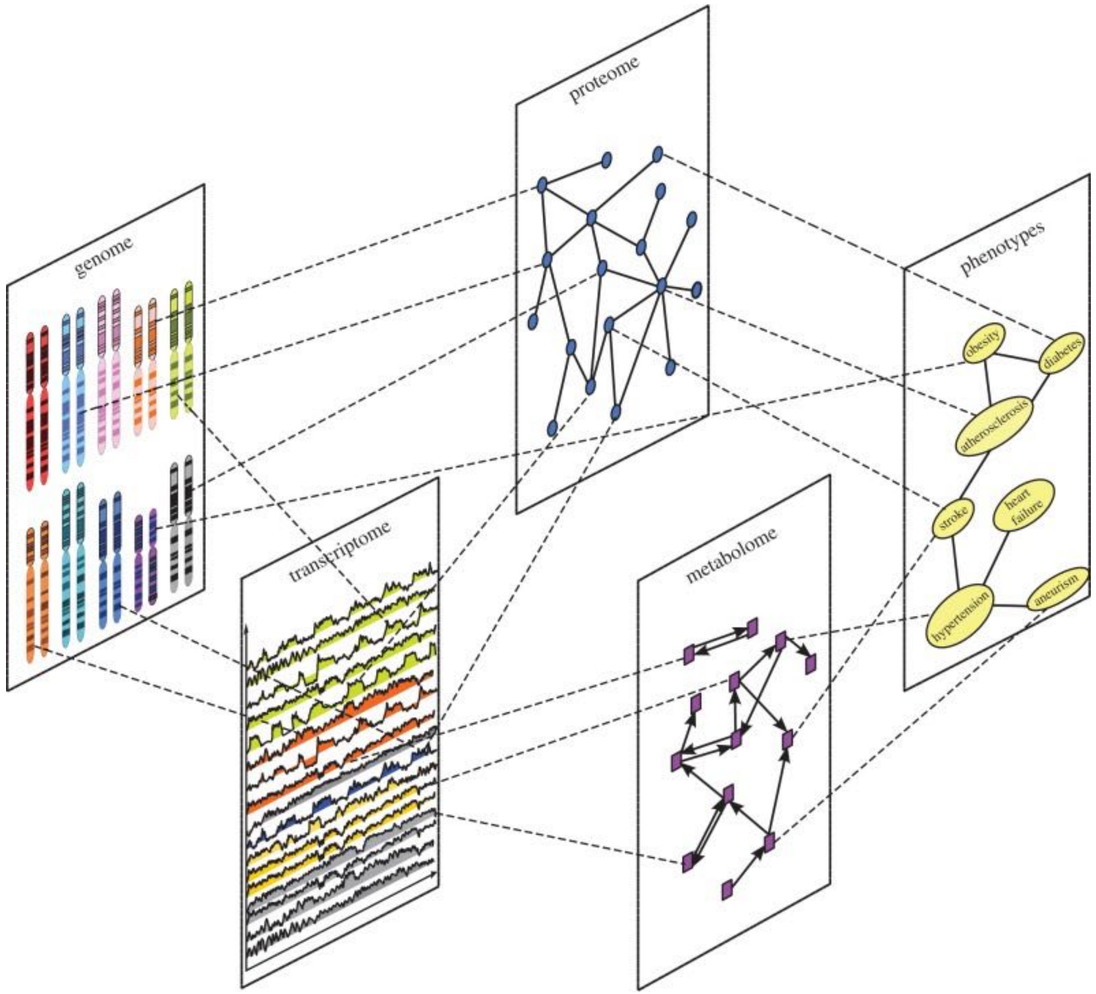
~1.000 USD

Central Dogma of Biology



Multi-omica


DNA
RNA
proteine
metaboliti
Genomica
Trascrittomica
Proteomica
Metabolomica
Fenotipo
Complessità

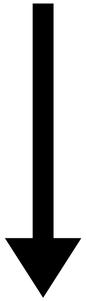

genome
proteome
transcriptome
metabolome
phenotypes
obesity
diabetes
atherosclerosis
stroke
heart failure
hypertension
aneurism

I dati sono “complicati”

Profilo “vero” della persona

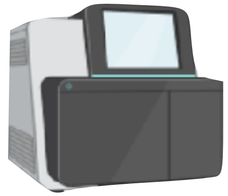
ACCTATGGAGAATACATAAGAATACCTAAAATGA

~3.000.000.0000 di “letture”
non necessariamente perfette



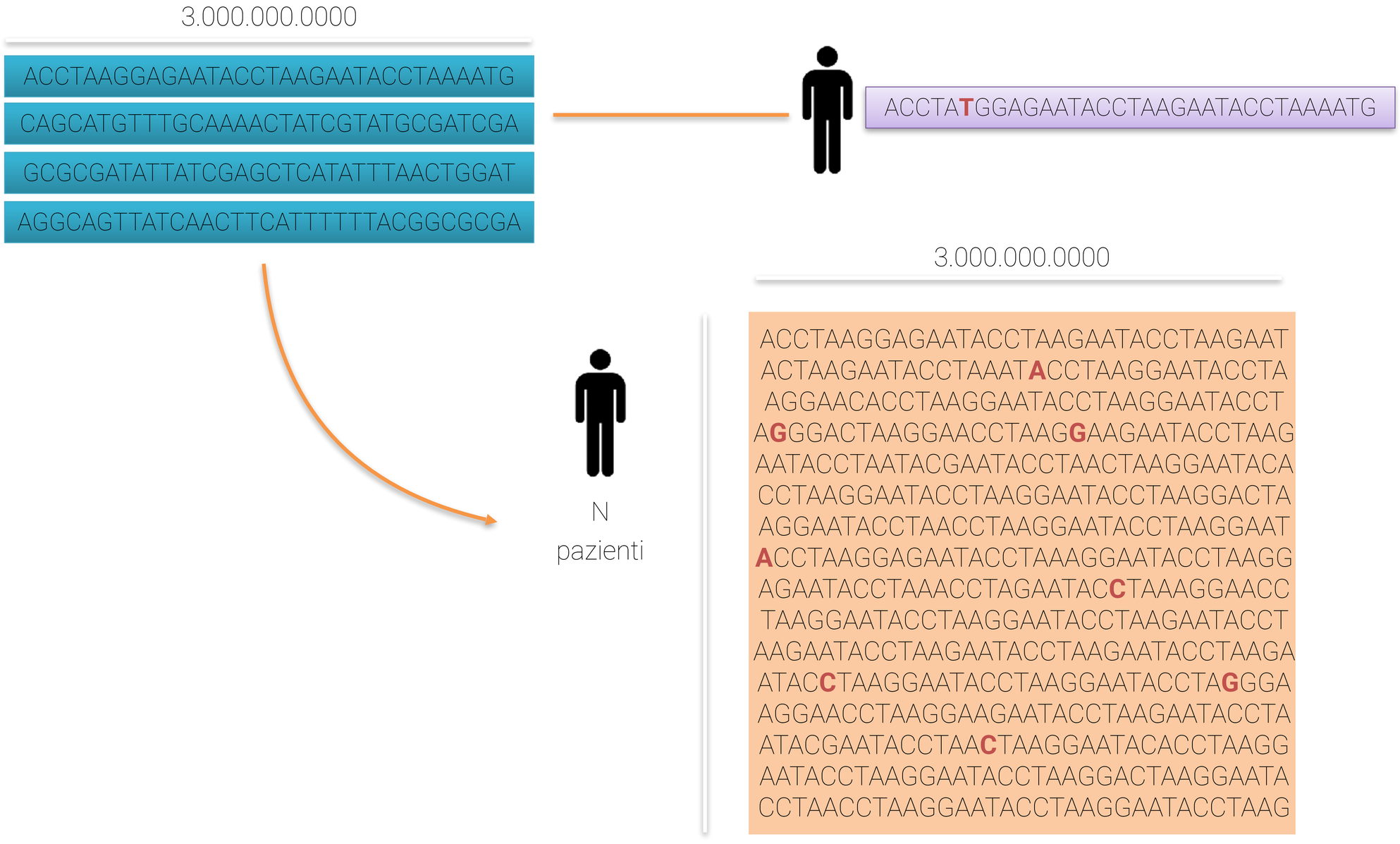
ACCTAAGGAGAATACCTAAGAATACCTAAAATGA

ACCTATGGAGAATACATAAGAATACCTAAAATGA



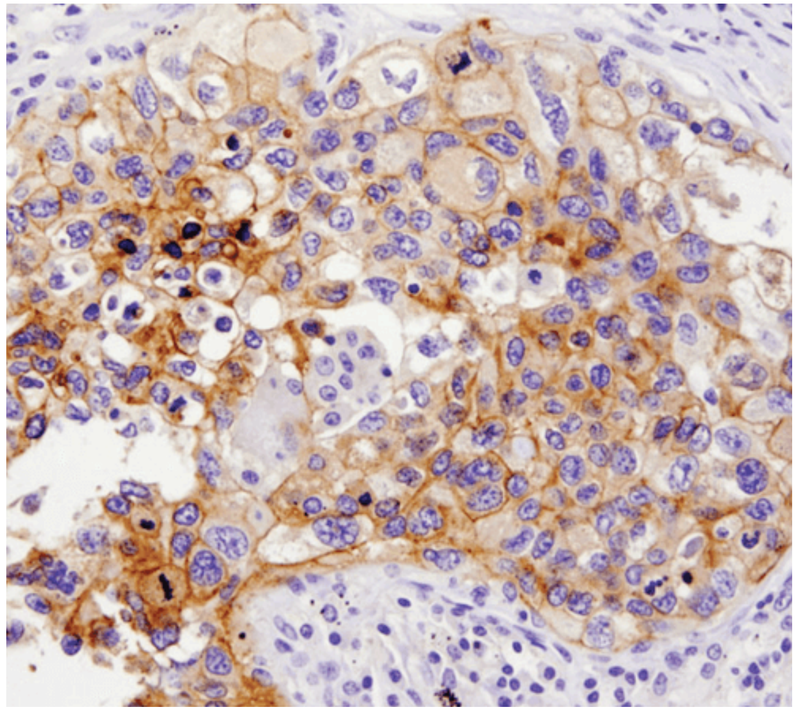
Mutazione o errore del macchinario?

I dati sono “complicati”

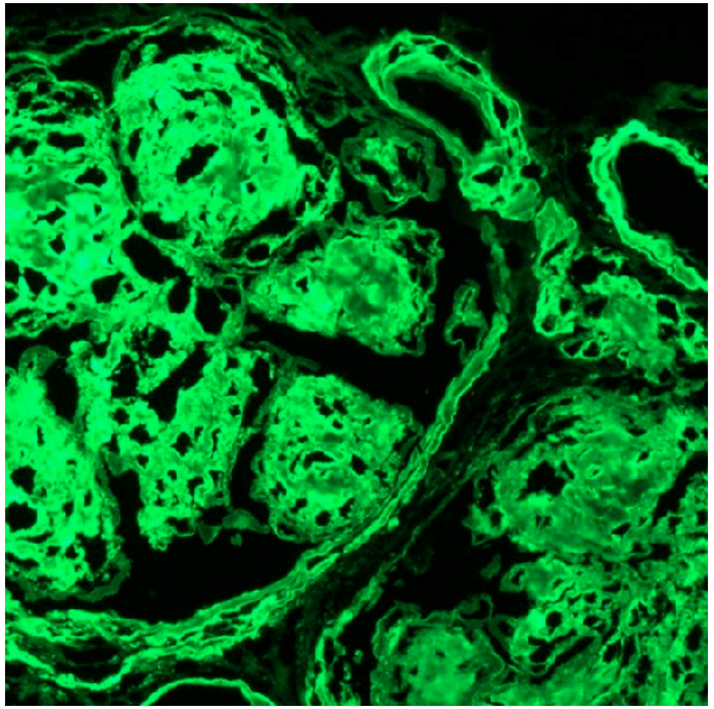


I dati sono “complicati”

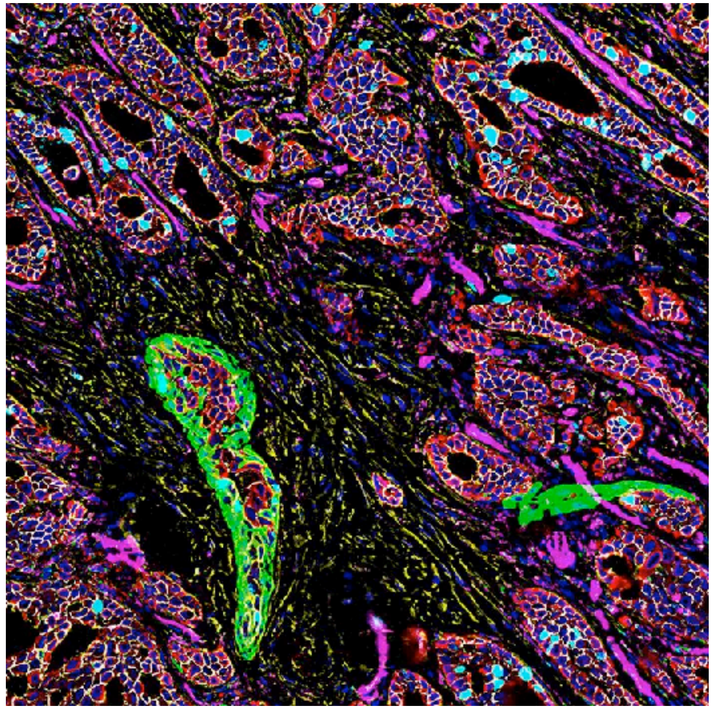
Immunoistochimica



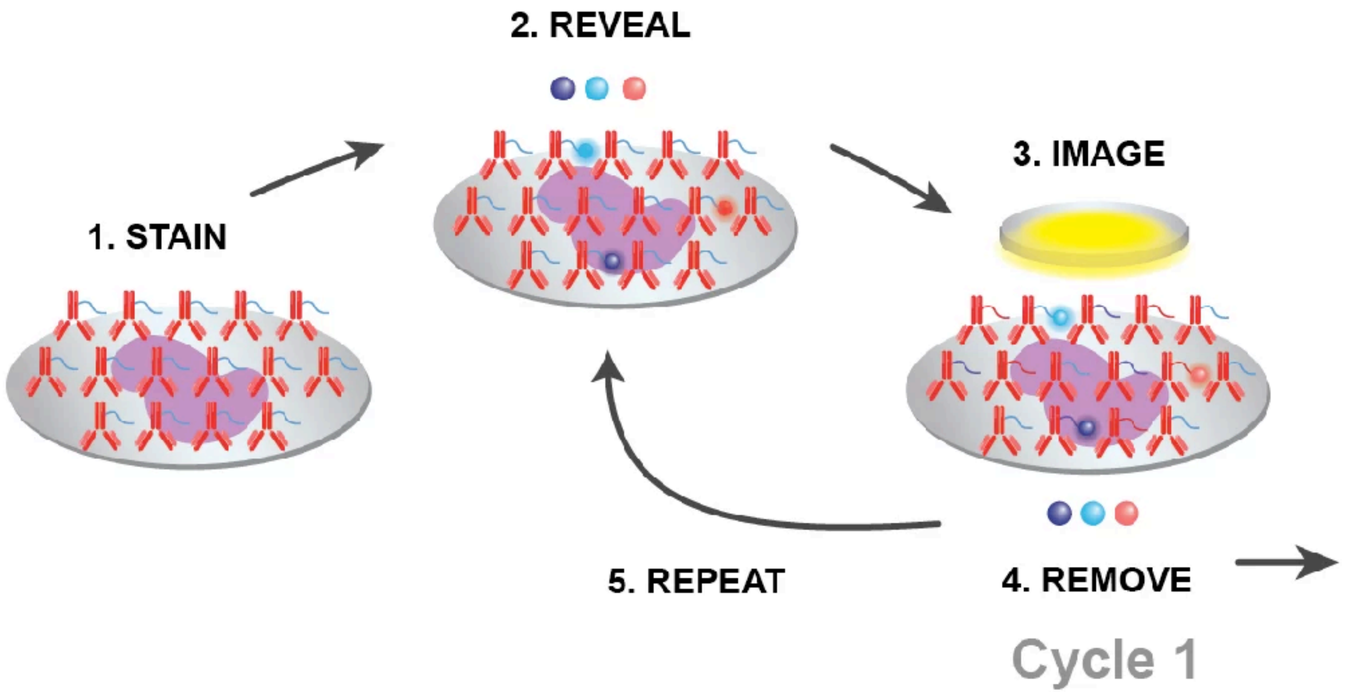
Immunofluorescenza

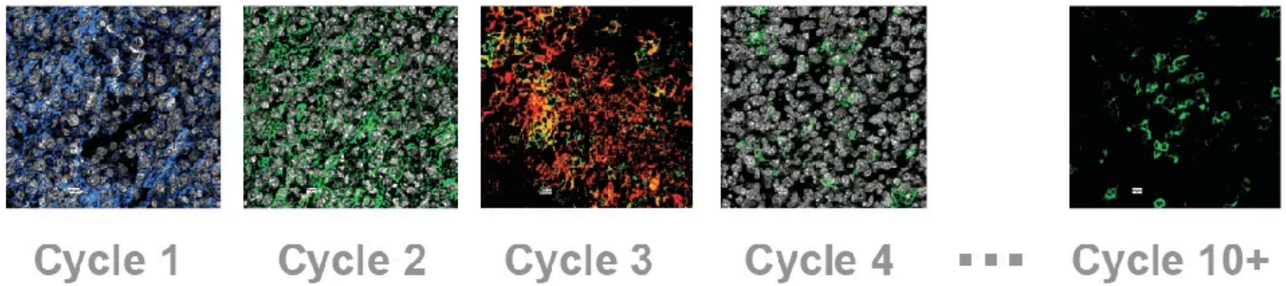


Imaging



I dati sono “complicati”





“Colorazione” delle molecole

Lettura della fluorescenza

Bassa qualità di scansione

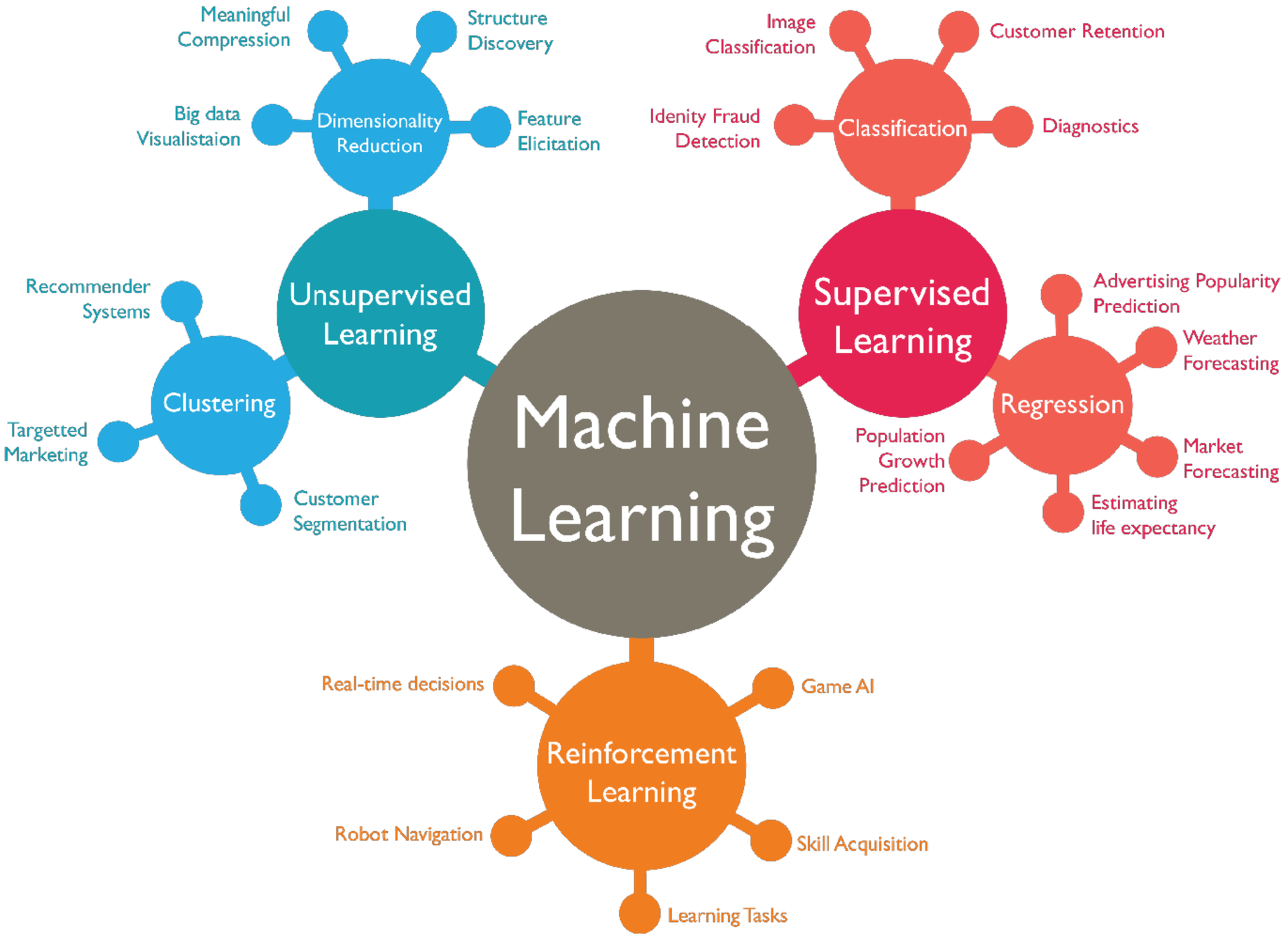
...

Quindi

- Capacità di elaborazione di dati molto “grandi”
- Capacità di integrare dati molto diversi
- Modelli resistenti al “rumore” delle misurazioni
- Automatizzazione dell’elaborazione

https://lucidworks.com/post/ai-trends-life-sciences/

Ma quale IA in questo contesto?

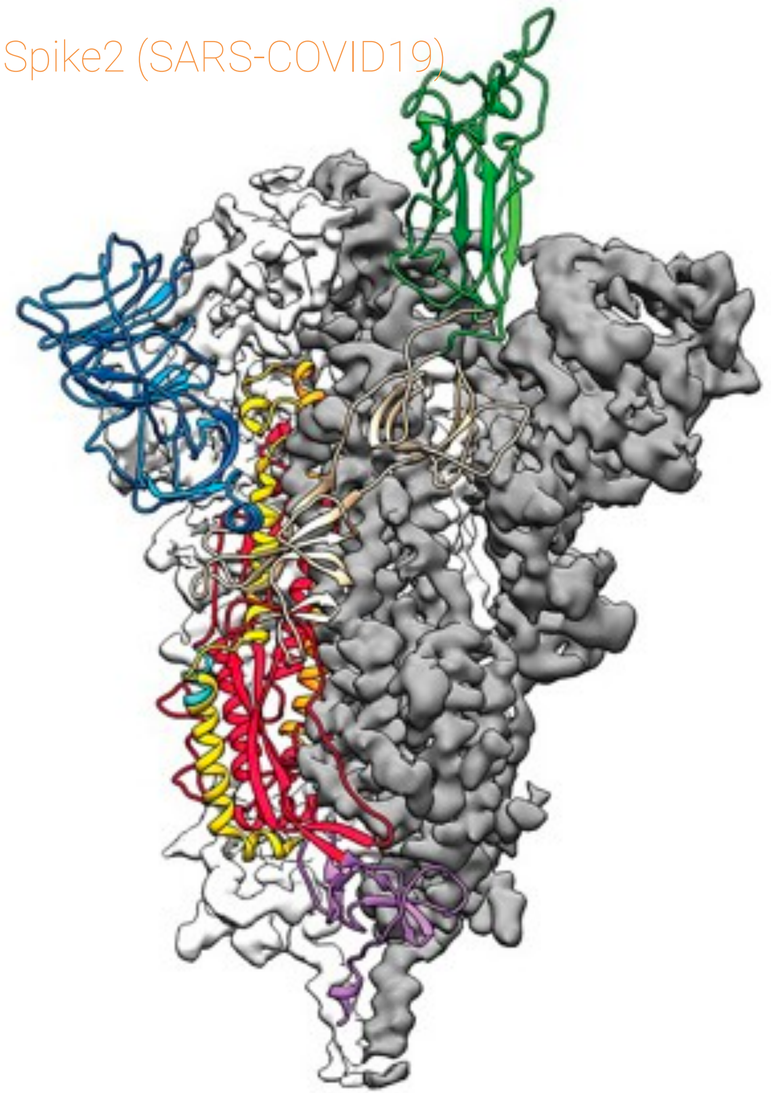

Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Proteine

AI per il ripiegamento di proteine

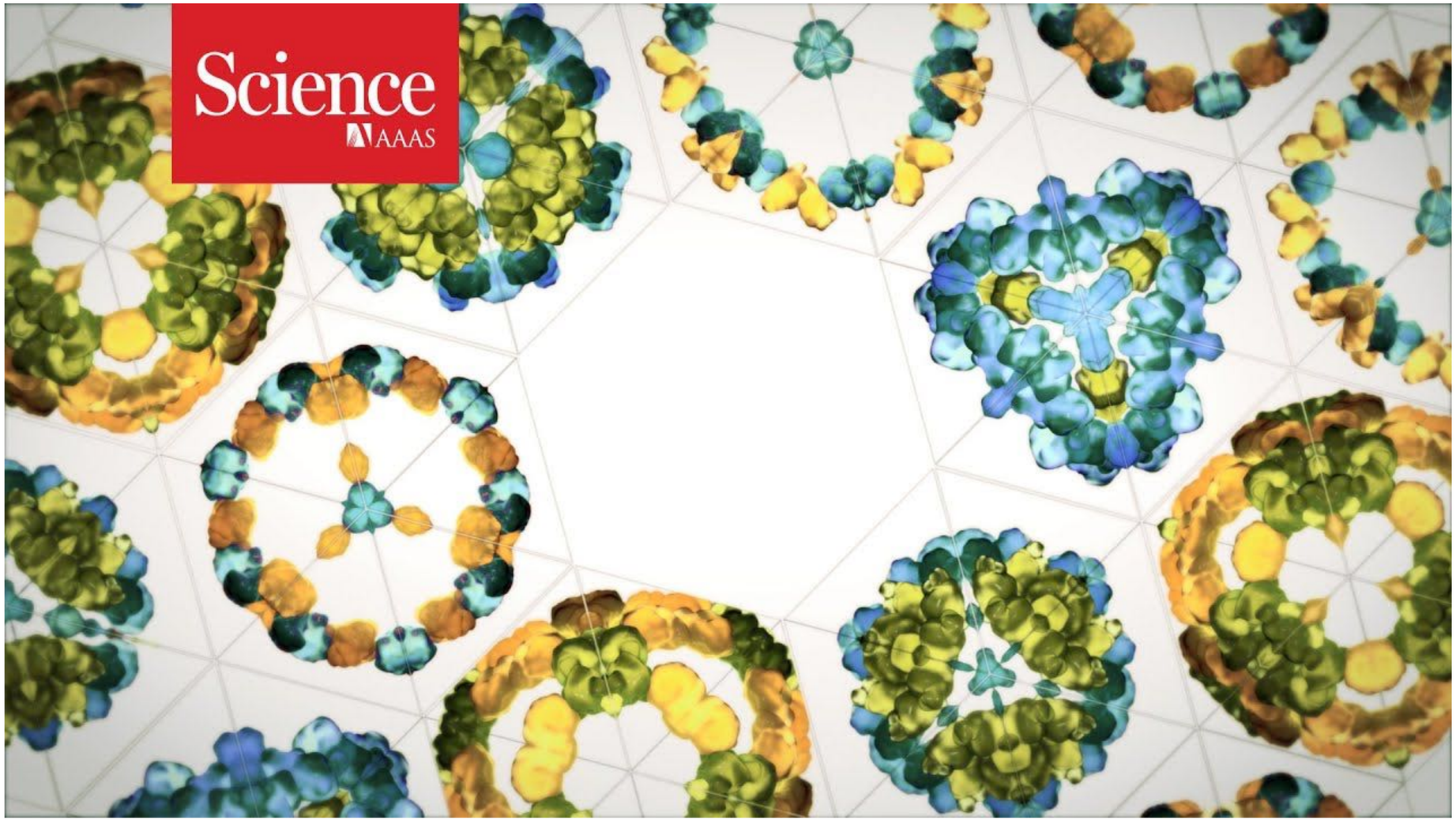
- unità fondamentale della vita
- costruite a partire dal DNA
- **importantissime per le malattie**

- La **forma 3D** determina le funzionalità



https://cen.acs.org/analytical-chemistry/structural-biology/Structure-novel-coronavirus-spike-protein/98/i8

Come nascono le proteine?



Sequenza
(DNA)

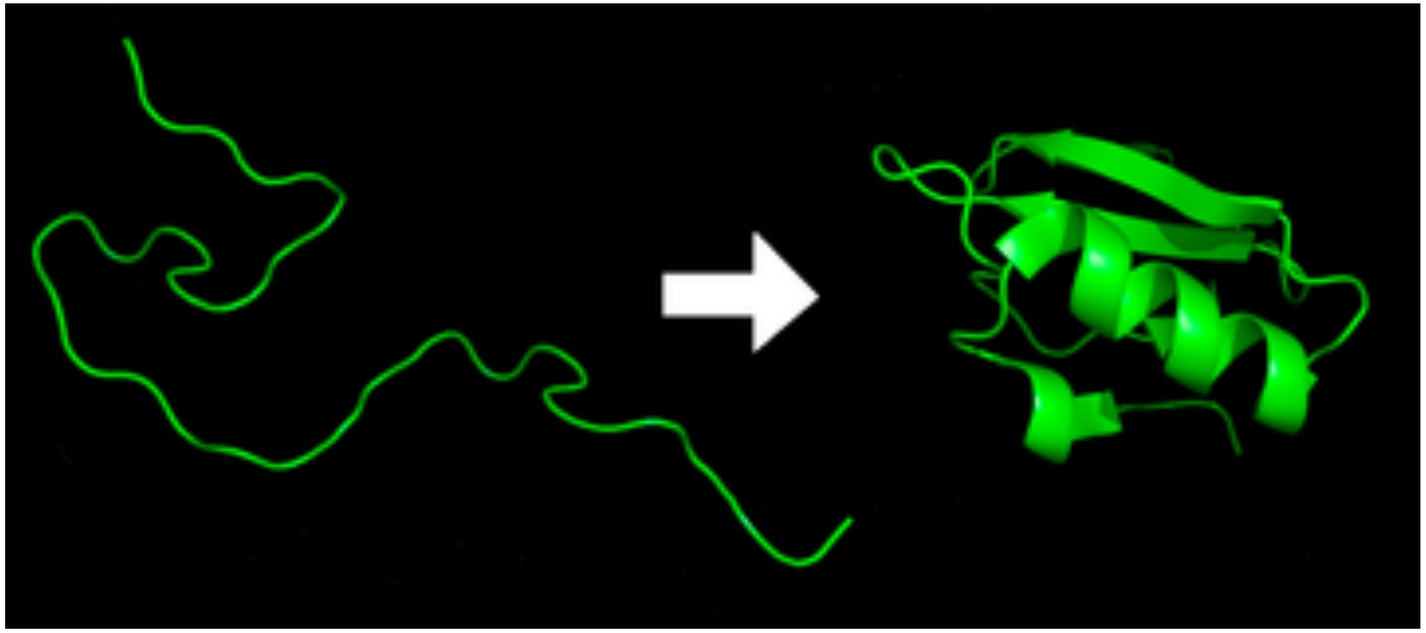
Trascritto
(RNA)

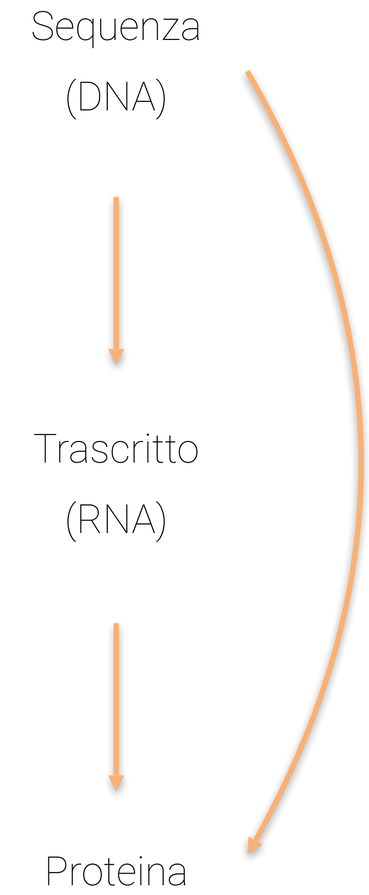
Proteina

https://www.youtube.com/watch?v=cAJQbSLIonI

Protein Folding problem

Determinare la struttura dalla sequenza è un problema noto da 50 anni, ed irrisolto.

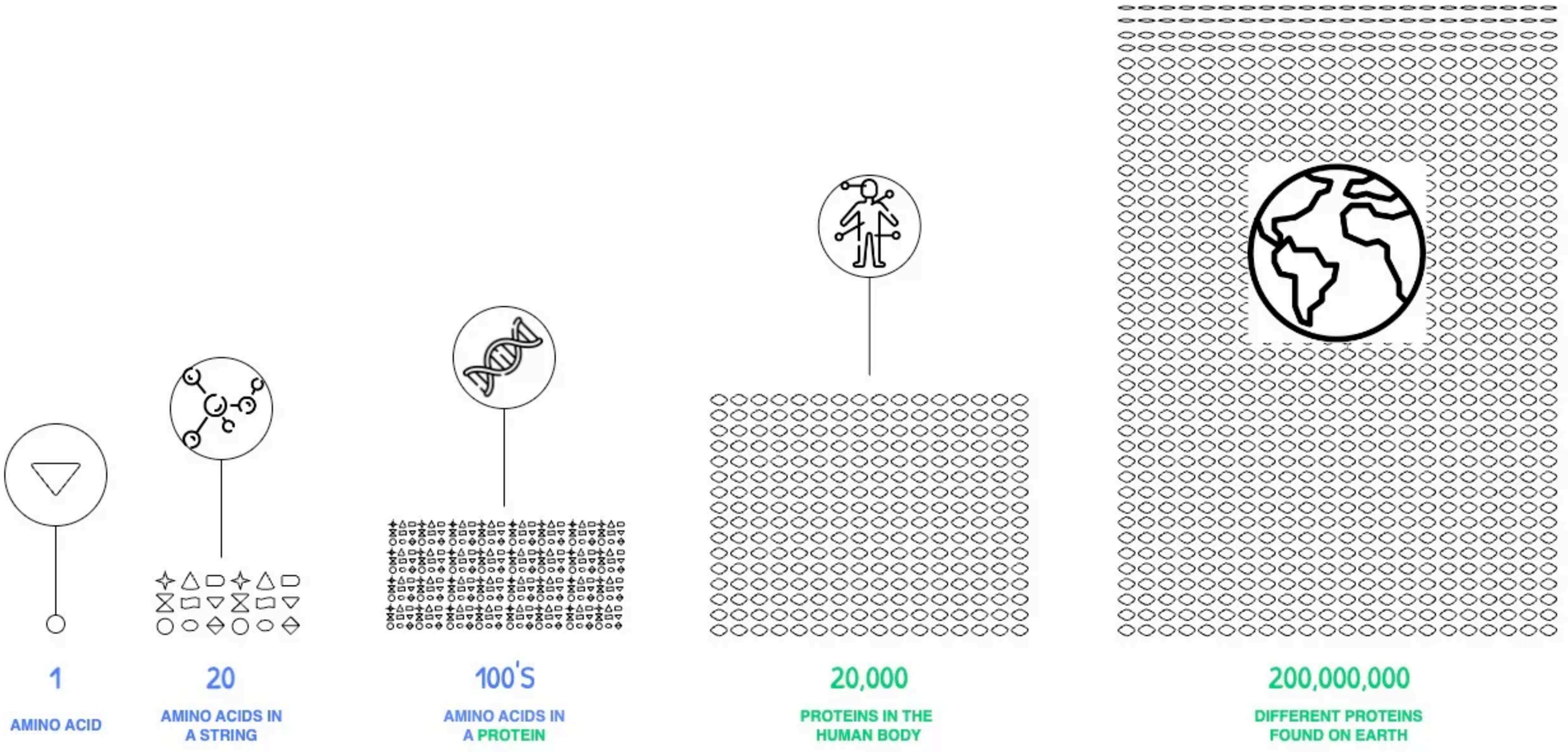




Sperimentalmente (1 proteina)

- 1 anno di lavoro e 100.000+ USD
- conosciamo le strutture di appena lo 0,1 percento di circa 200.000.000

Protein Folding problem



https://towardsdatascience.com/what-does-alphafold-do-60b6370dafe4

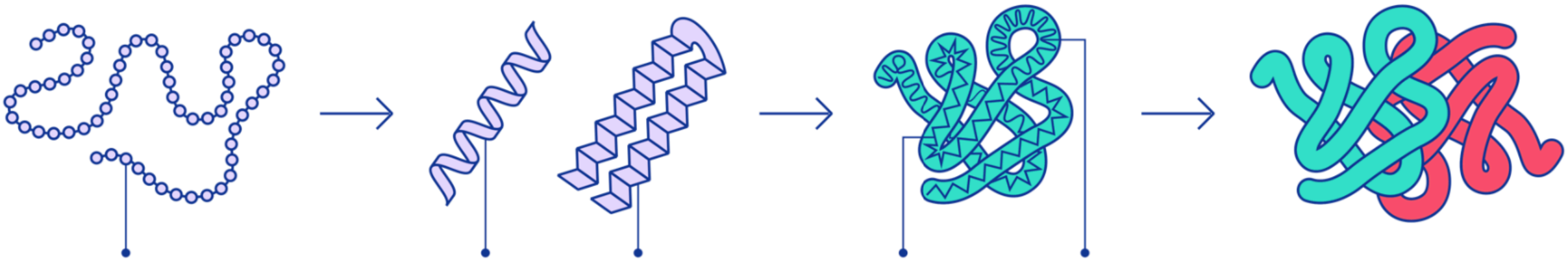
Calcolare il ripiegamento

struttura primaria

struttura secondaria

struttura terziaria

struttura quaternaria



Amminoacidi

Elica

Foglietto

Elica

Foglietto

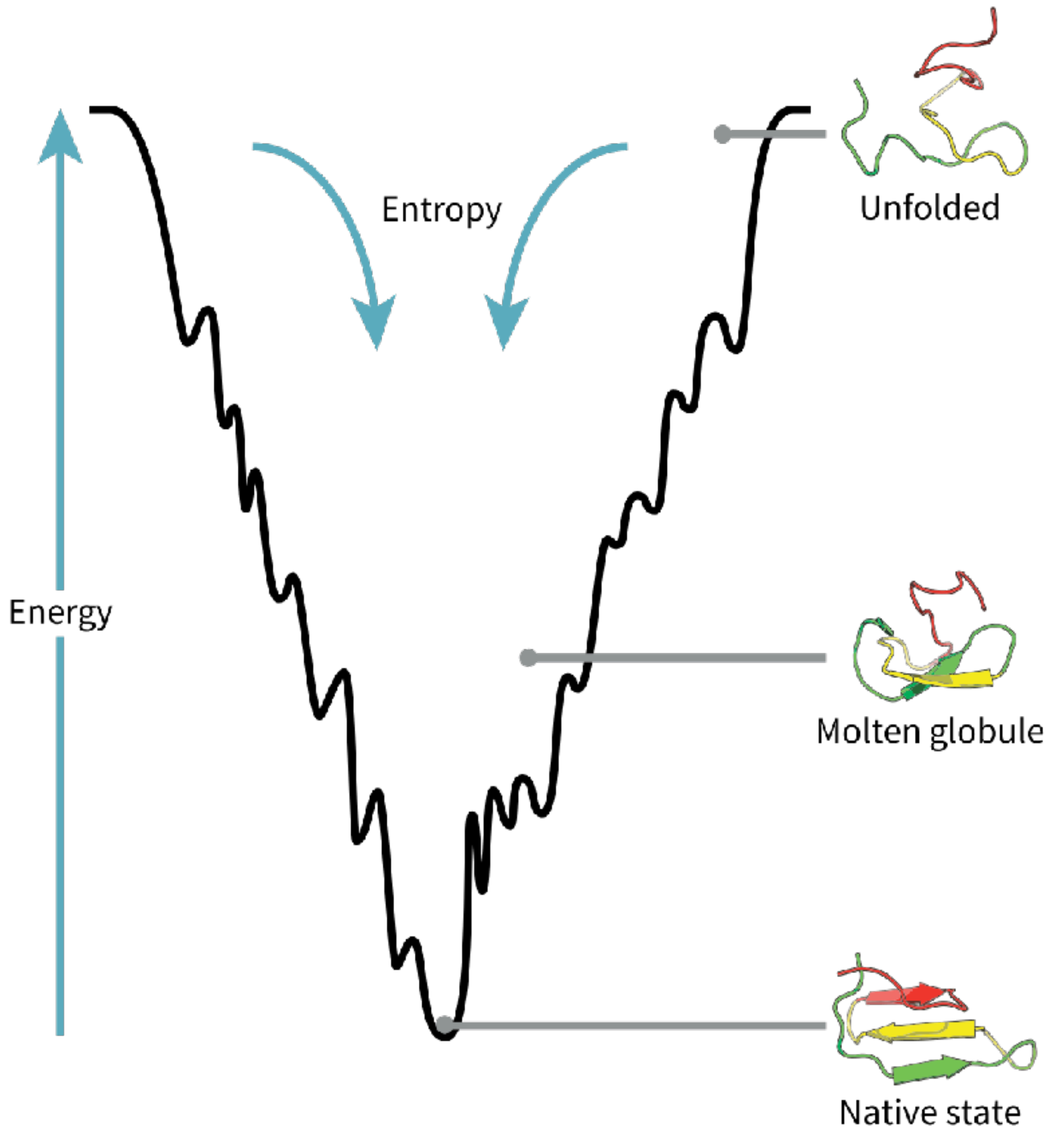
Interazione tra proteine

Interagiscono localmente
per formare sotto-forme

Si ripiegano
globalmente e formano
la struttura 3D

https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

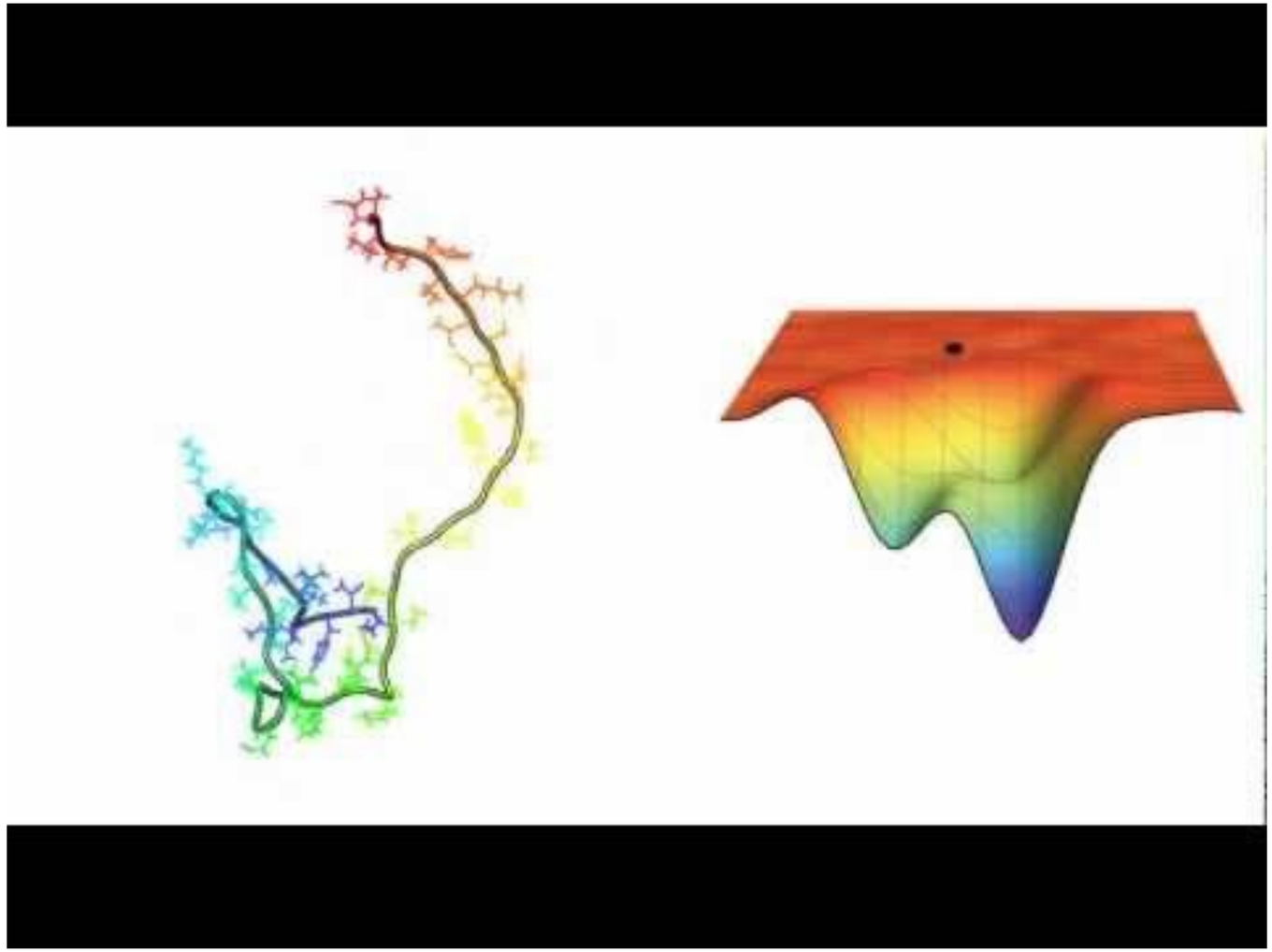
Protein Folding Funnel



Wikipedia

Problema computazionale

- Possiamo usare tecniche di predizione molecolare

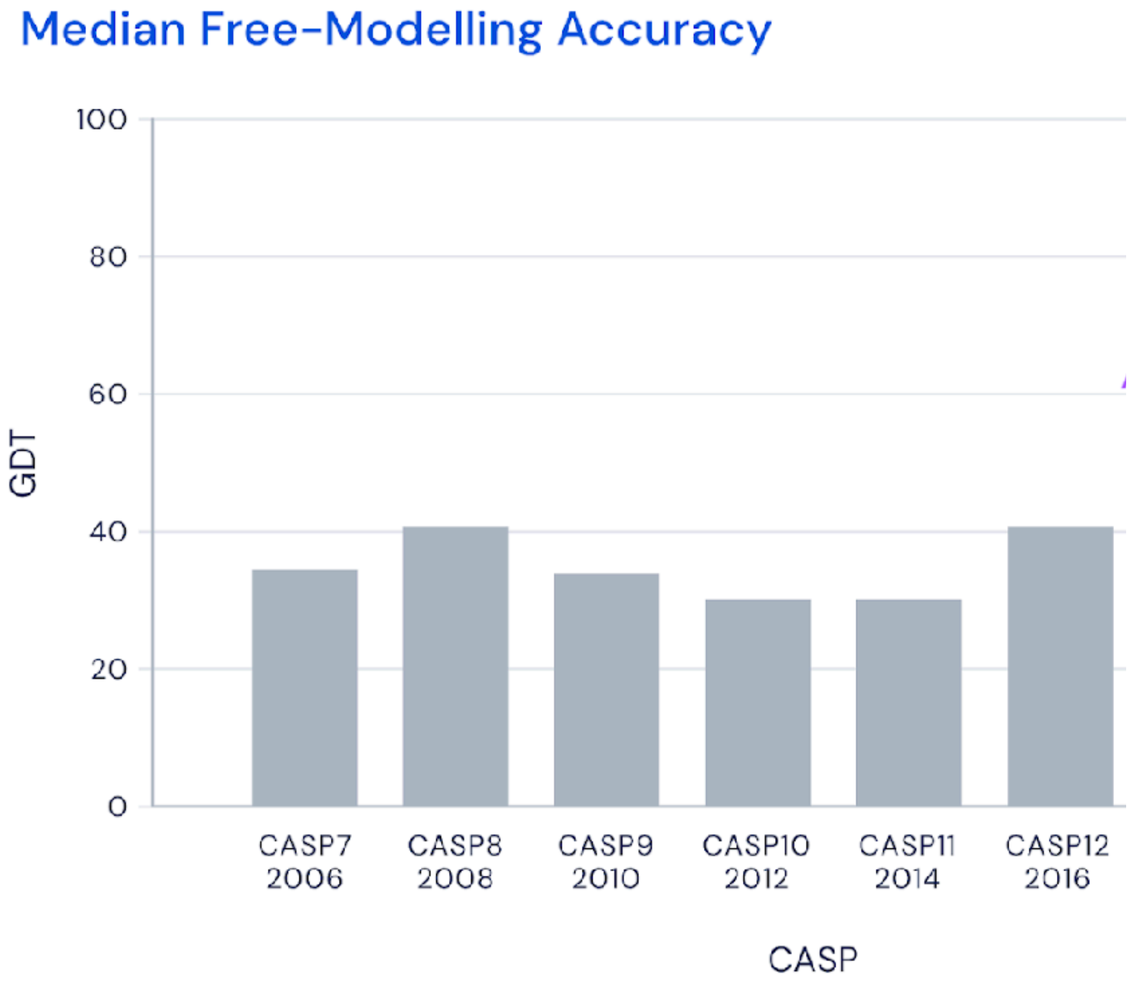


https://www.youtube.com/watch?v=DDHYIReVPr0

CASP

- CASP Critical Assessment of protein Structure Prediction

- Si svolge ogni due anni dal 1994.

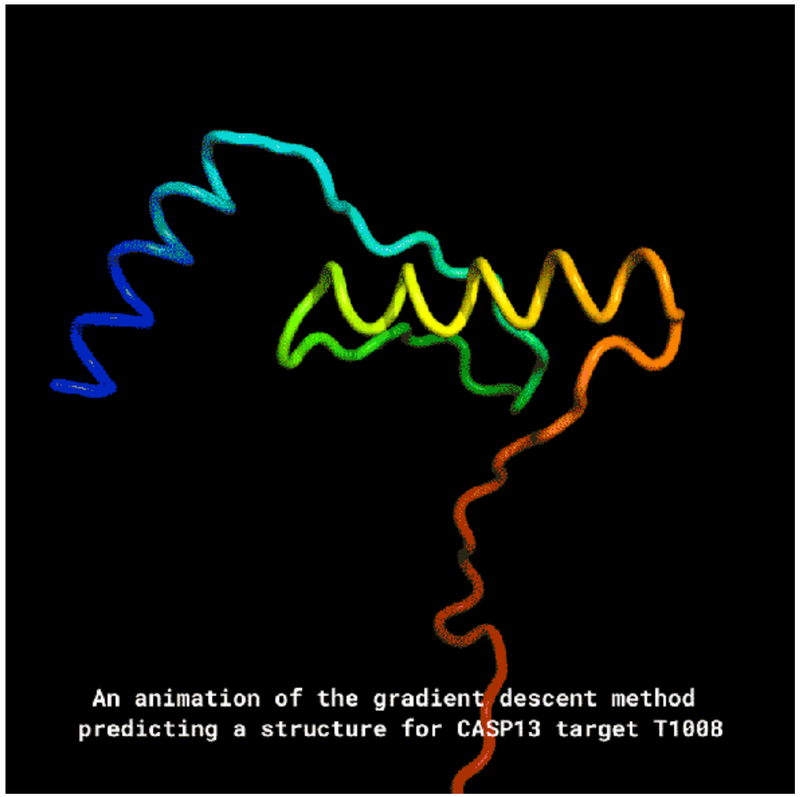


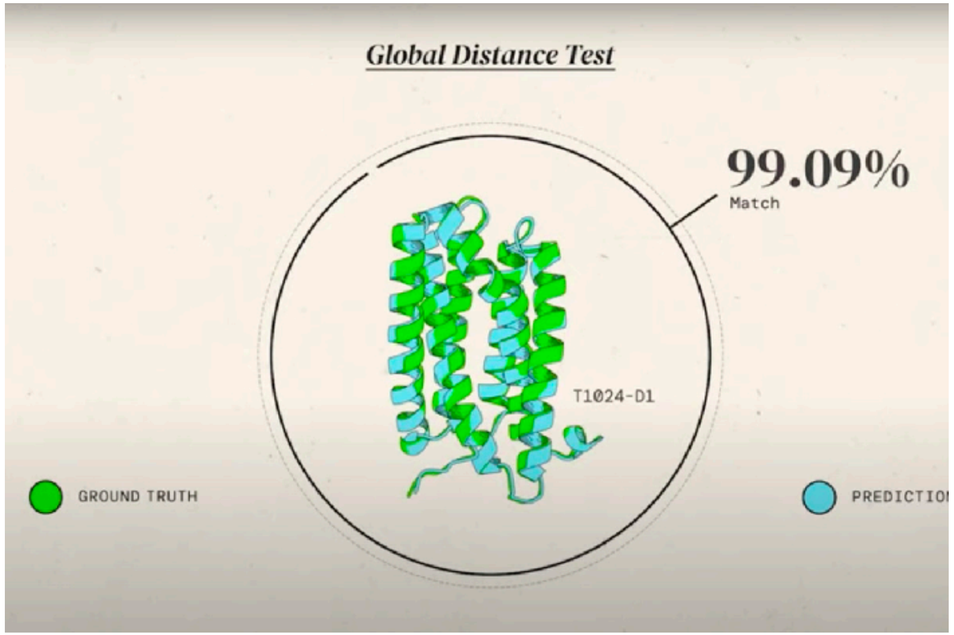
https://it.wikipedia.org/wiki/CASP

AlphaFold

DeepMind

- AI supervisionata che impara da ~100.000 proteine note

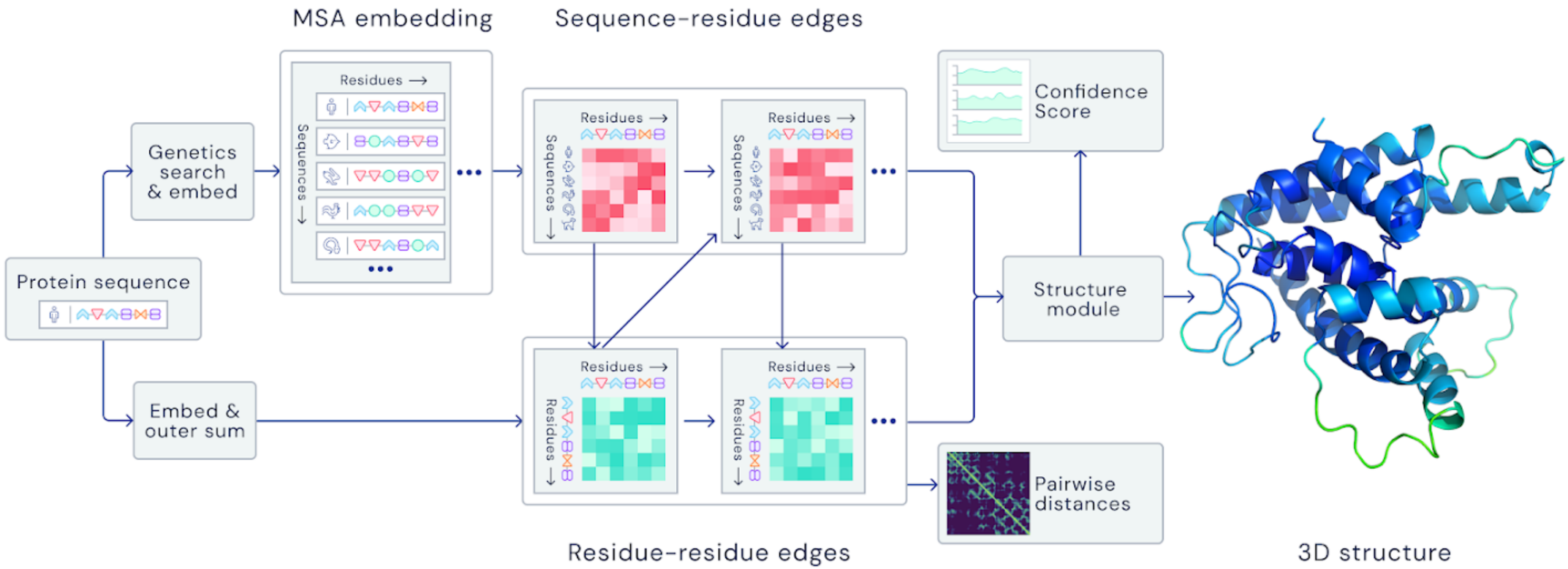




https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

AlphaFold

DeepMind

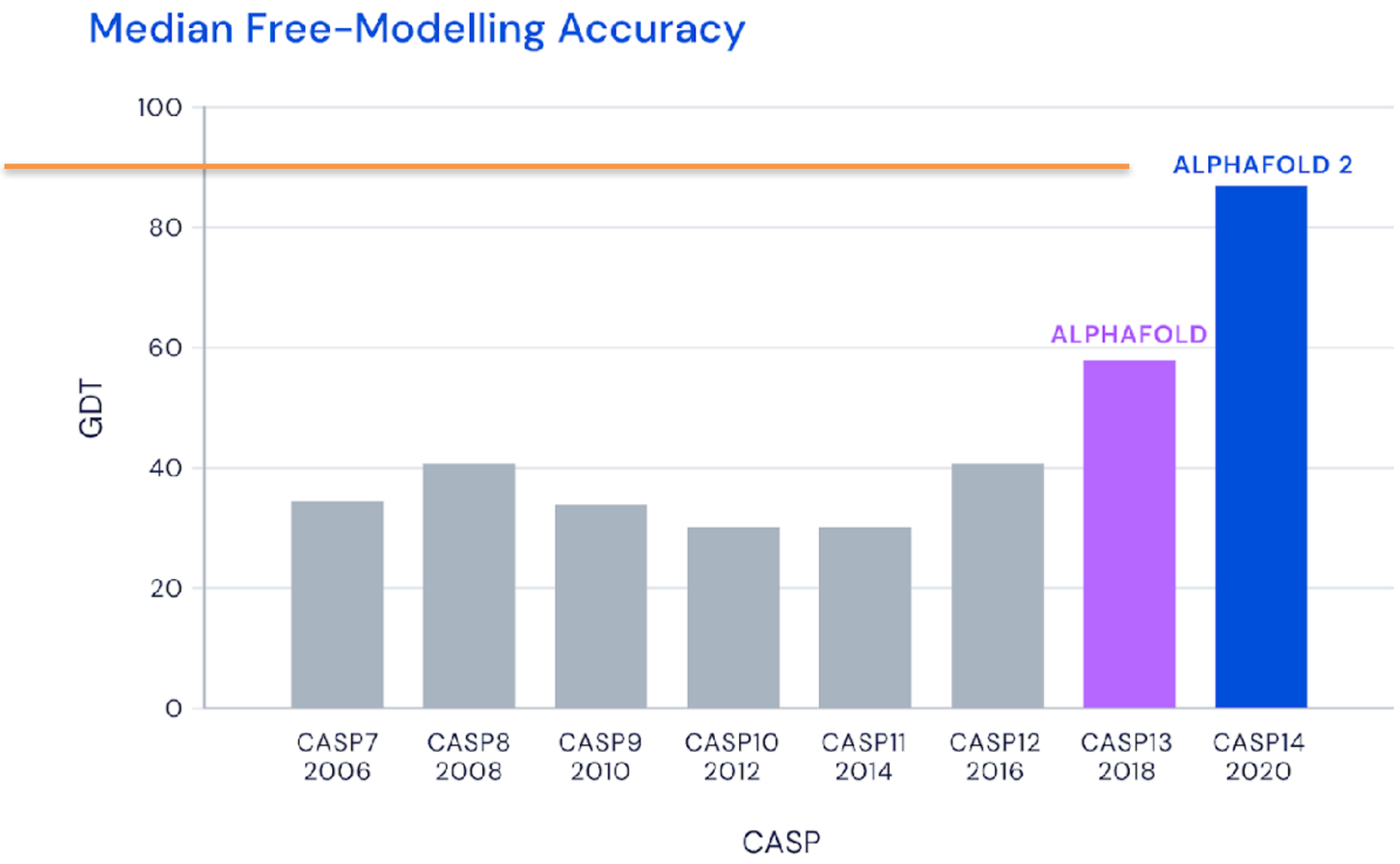


- Transformers: usati inizialmente per NLP
- Multiple Sequence Alignments (MSA)
- …

https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

AlphaFold a CASP

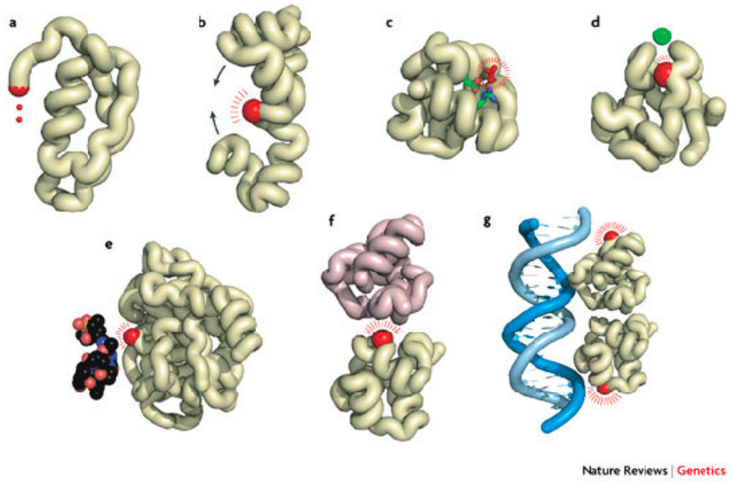
Precisione sperimentale (90%)



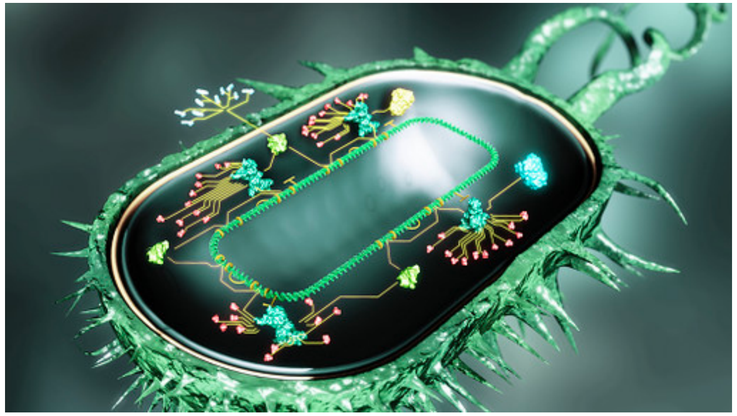
Ancora qualche limitazione per alcuni tipi di proteine (e.g. strutture con più monomeri)

https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

AlphaFold - cosa ci posso fare?



Mutazioni genetiche



Biologia sintetica

Medicine and Pharmaceuticals

- Drug-discovery
- Antibody production
- Vaccine innovation (mRNA vaccines)

Biofuels and Sustainable Energies

- Fermentation processes
- Ethanol made by thermophilic organisms
- Biodiesel from yeast cultured on sugarcane

Environmental Bioremediation

- Biodegradation of methylphenols
- Biosensors that alert the presence of arsenic
- Using microbes to clean up chemical waste

Food and Agriculture

- Improved vitamin content in yogurts, cheese, cereals
- Bolstered crop resistance to devastating fungus
- Stevia sweeteners and other dietary supplements

Space Systems and Exploration

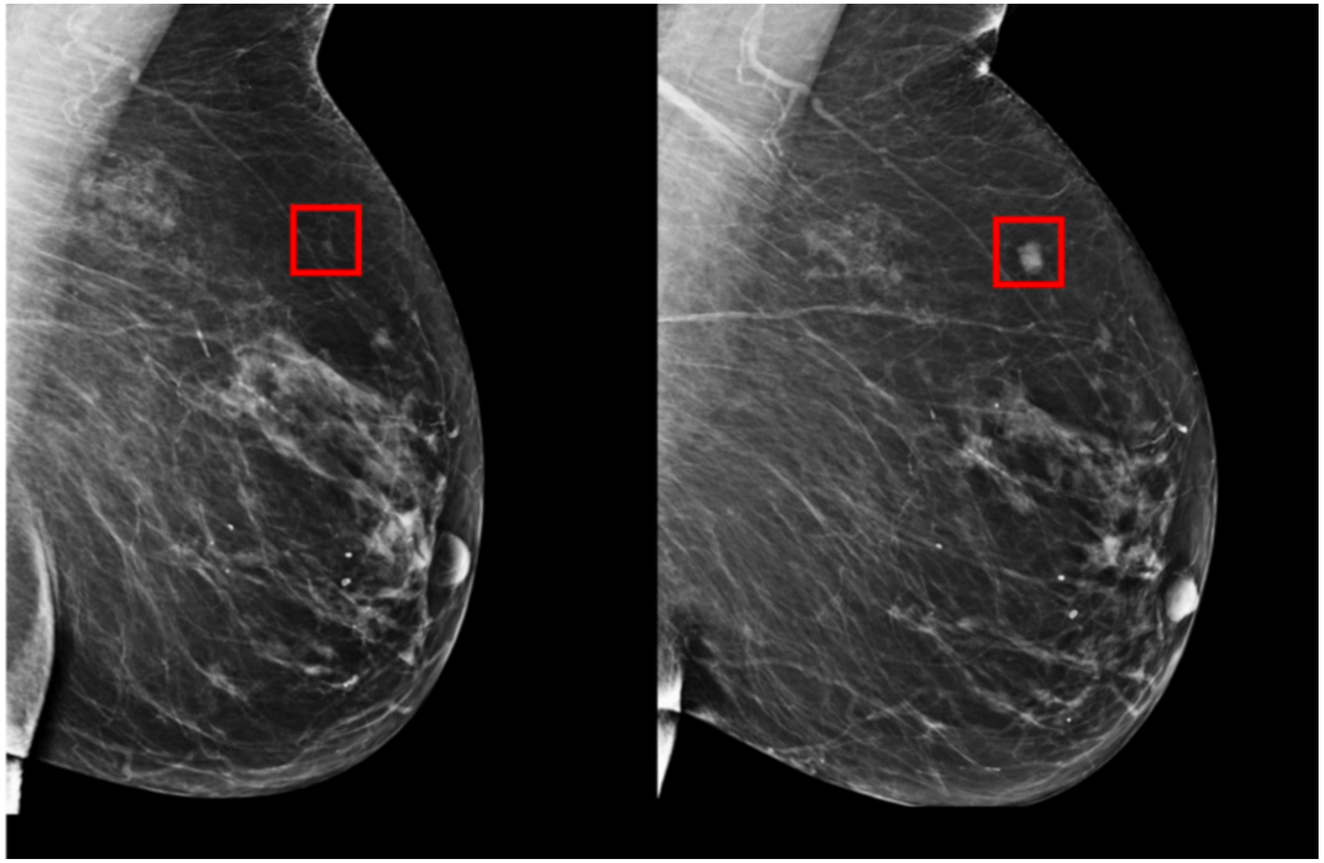
- Medicine
- Food
- Plastics and other substances

https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

Diagnostica per immagini

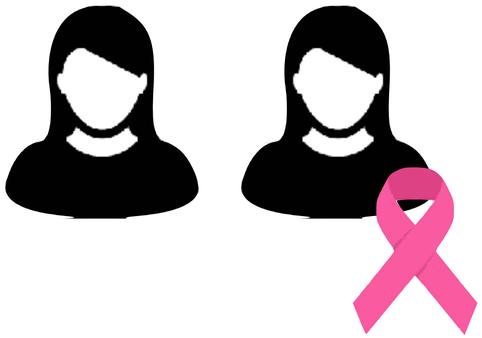
La patologia per immagini

- MIRAI (MIT): sistema creato nel 2019 che predice, da una mammografia, se una paziente ha una certa probabilità di sviluppare un tumore al seno entro i prossimi 5 anni.



60.000 pazienti

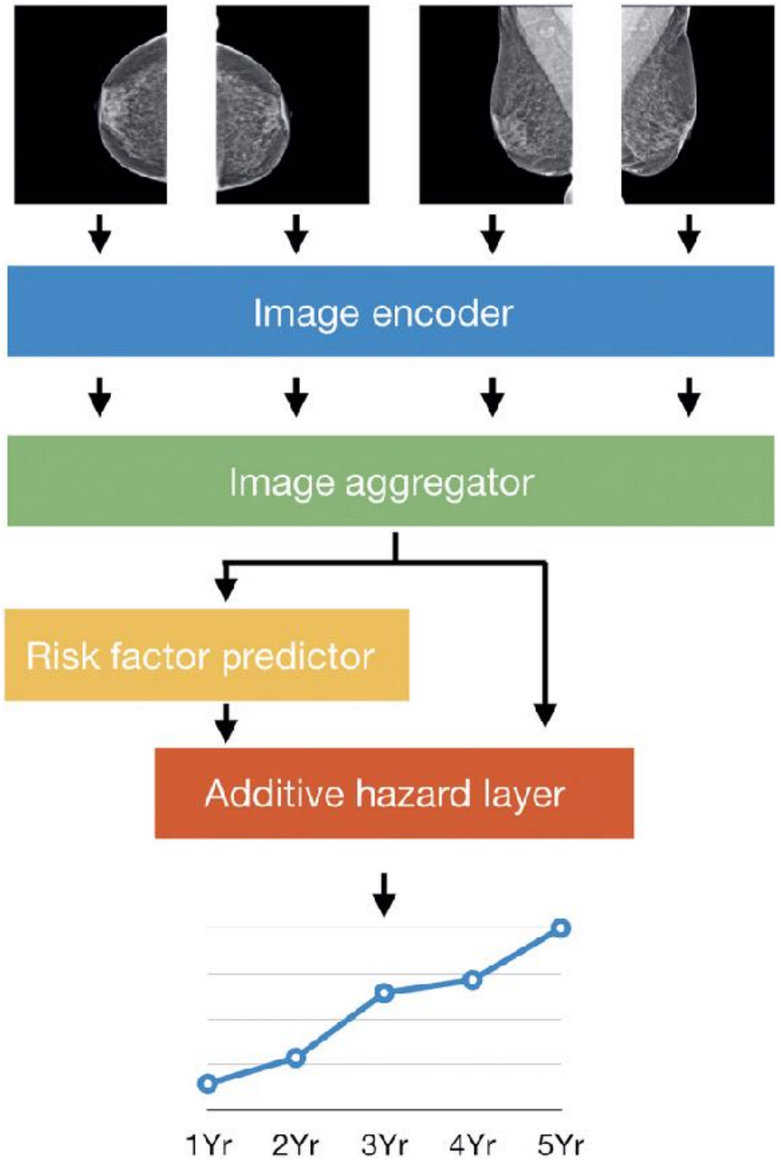
90.000 mammografie



Diagnostica oltre fattori di rischio “classici” quali età, familiarità, ormoni, dieta, etc.

https://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507

La patologia per immagini



Input

Codifica delle imagini in vettori

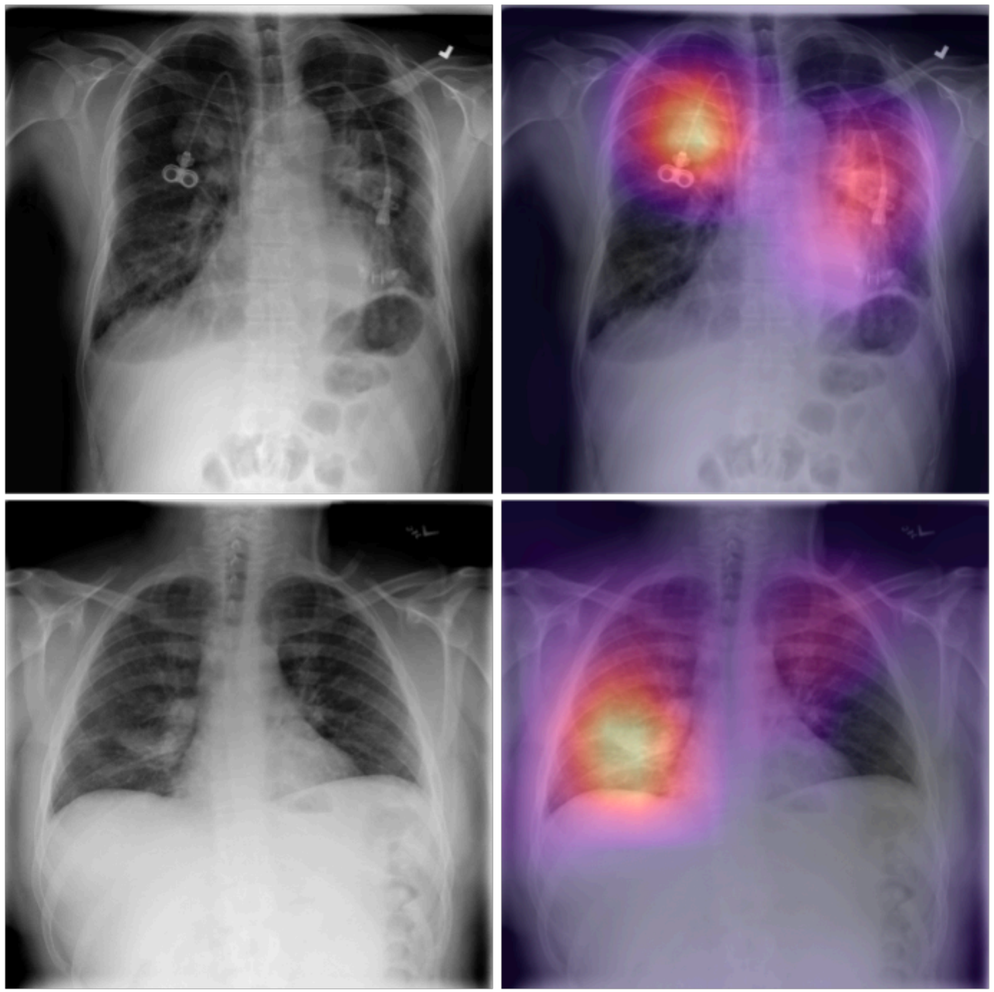
Combinazione dei quattro input

Predizione del rischio

DOI: 10.1126/scitranslmed.aba4373

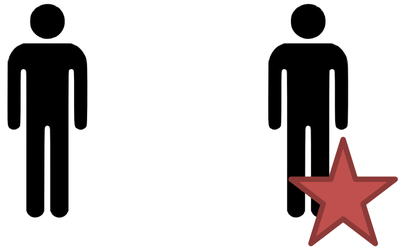
La patologia per immagini

- CheXNeXt (Stanford): diagnostica automatica da radiografie, per 14 malattie toraciche acute (incluso tubercolosi e tumore al polmone)



~30.000 pazienti

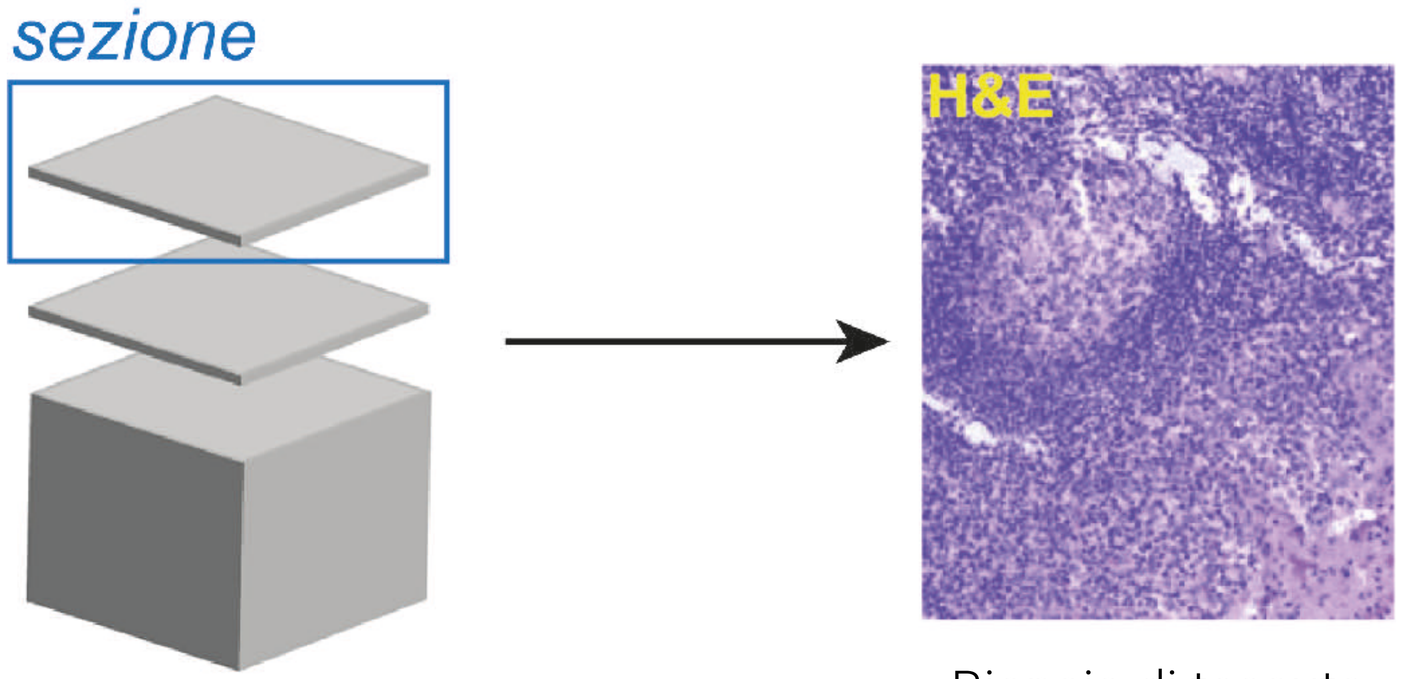
>100.000 radiografie



Rete neurale convoluzionale

https://stanfordmlgroup.github.io/projects/chexnext/

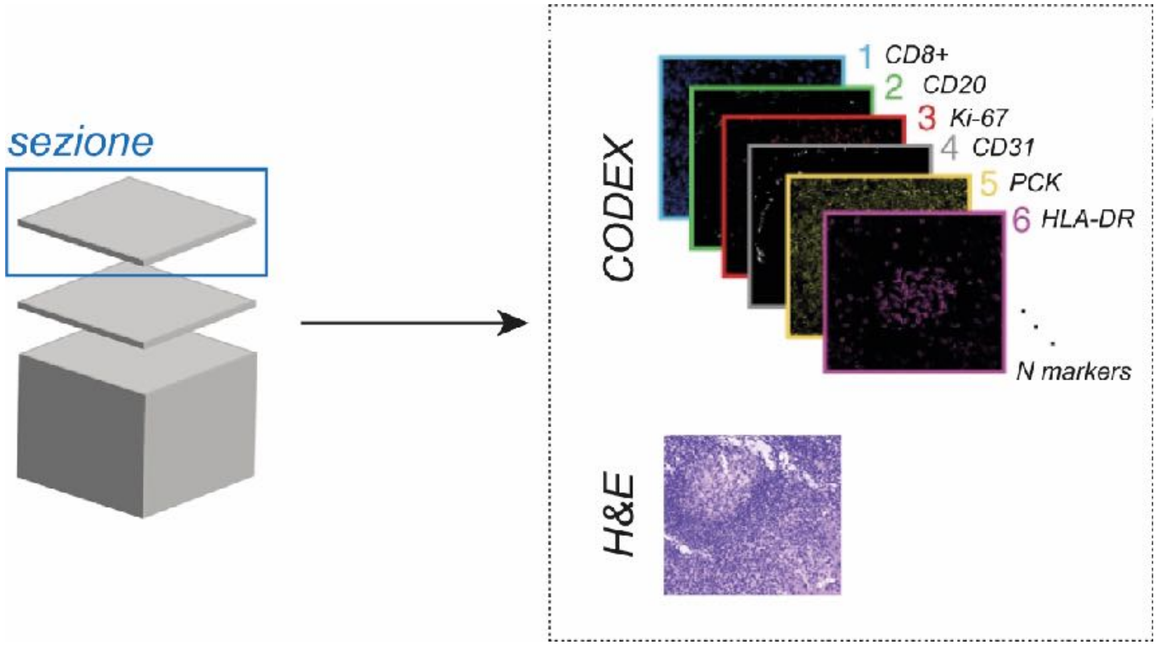
La patologia per immagini



Biopsia di tessuto

La patologia per immagini

- CODEX: identificazione delle cellule e localizzazione di marcatori

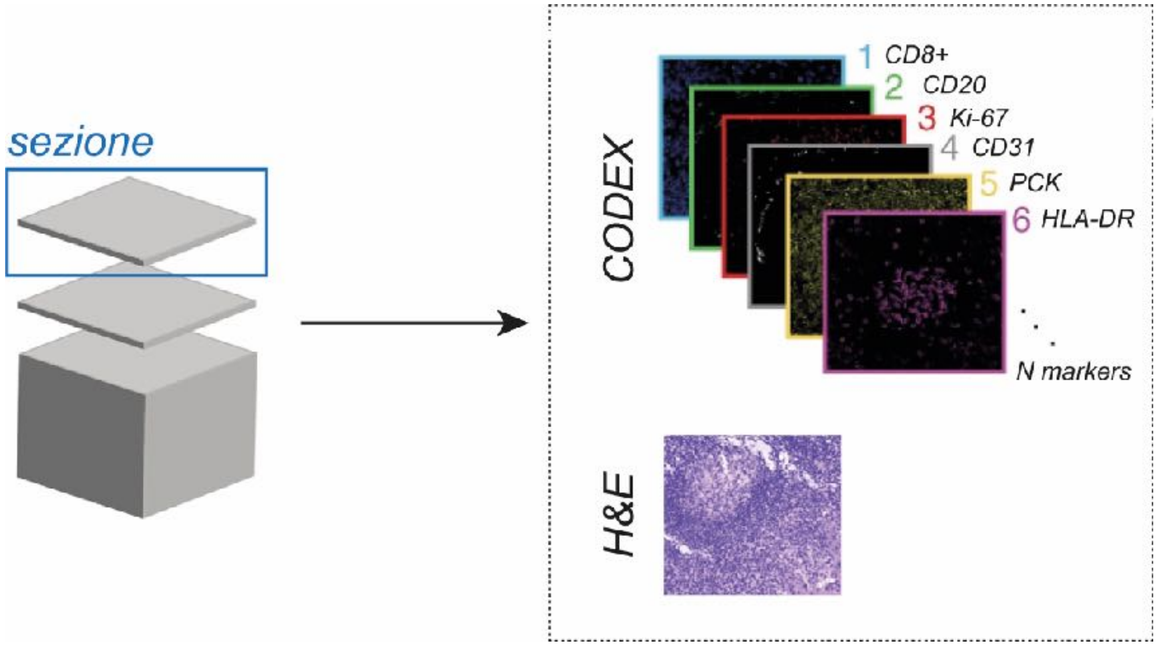


N marcatori → immagine con N canali

https://www.kitware.com/cell-nuclei-detection-on-whole-slide-histopathology-images-using-histomicstk-and-faster-r-cnn-deep-learning-models/

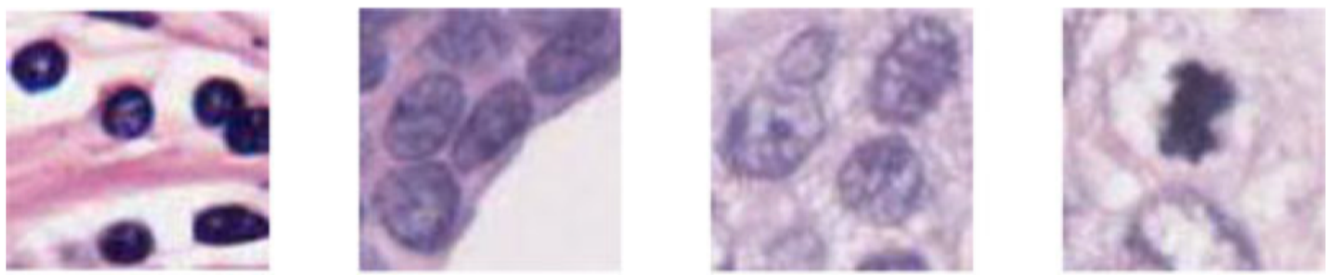
La patologia per immagini

- CODEX: identificazione delle cellule e localizzazione di marcatori



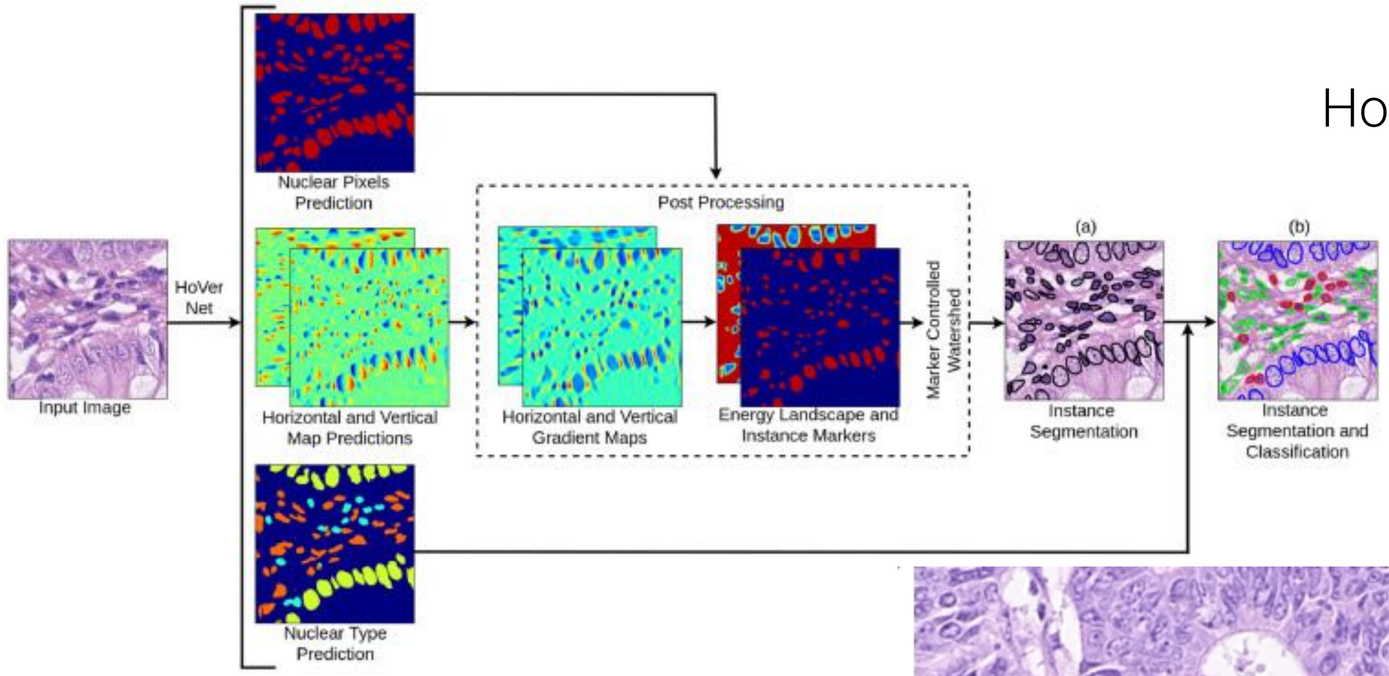
N marcatori → immagine con N canali

Nuclei di diversi tipi cellulari (~200 tipi cellulari)

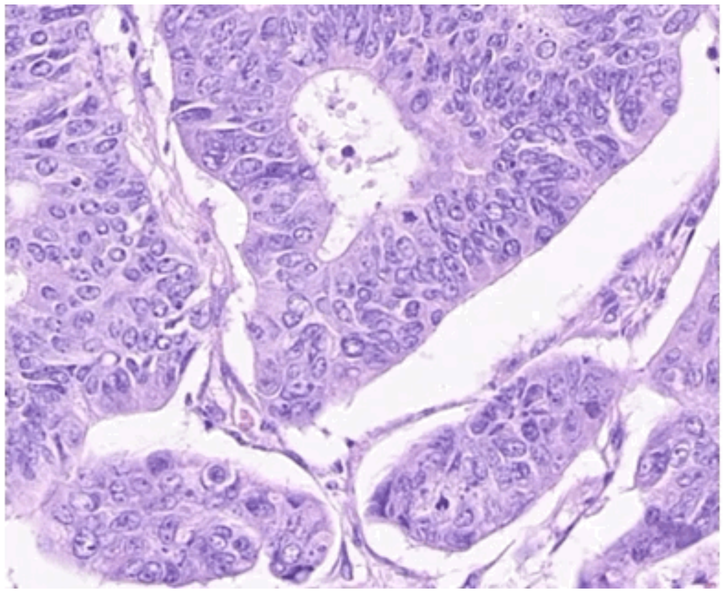


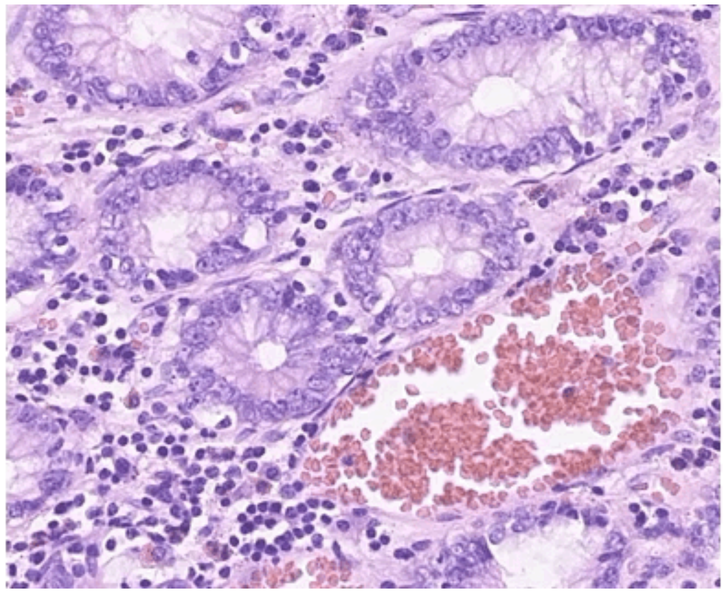
https://www.kitware.com/cell-nuclei-detection-on-whole-slide-histopathology-images-using-histomicstk-and-faster-r-cnn-deep-learning-models/

Classificazione di tipi cellulari



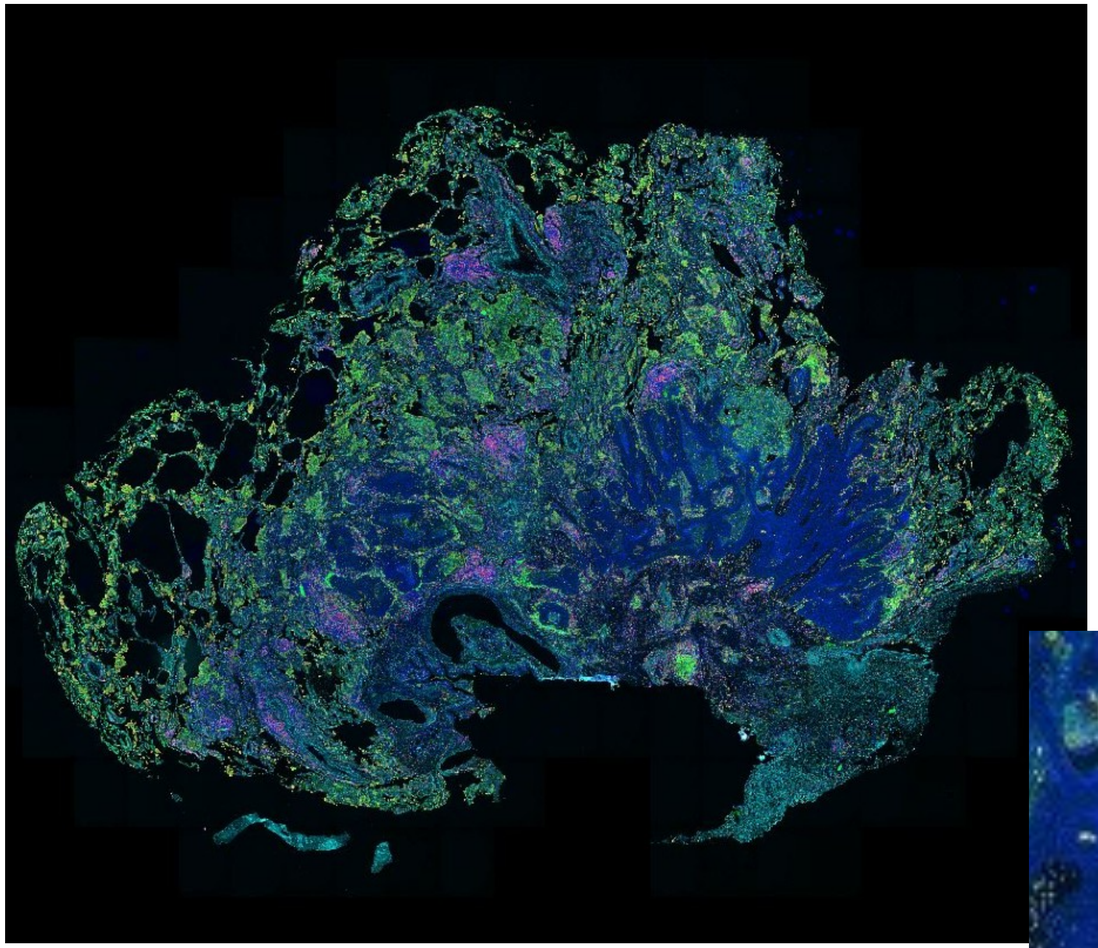
HoVer-Net: segmentazione e classificazione



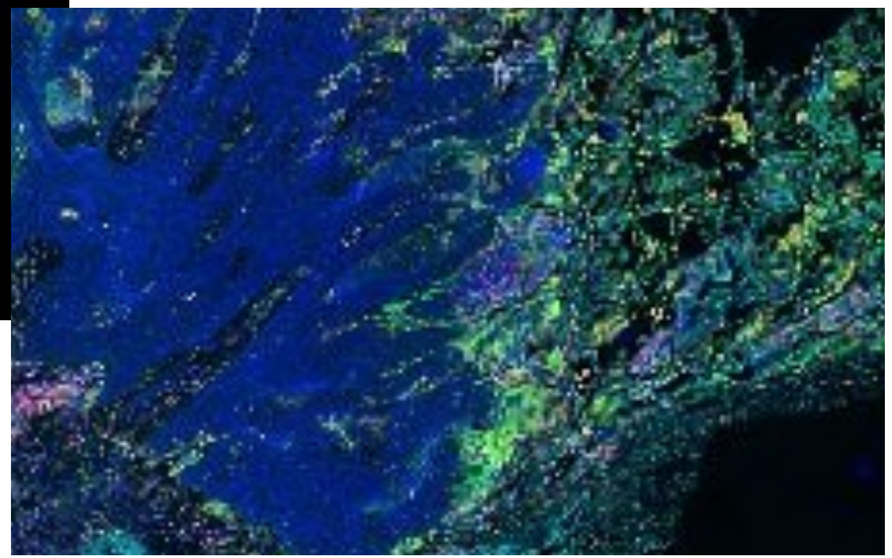


https://doi.org/10.48550/arXiv.1812.06499

La patologia per immagini



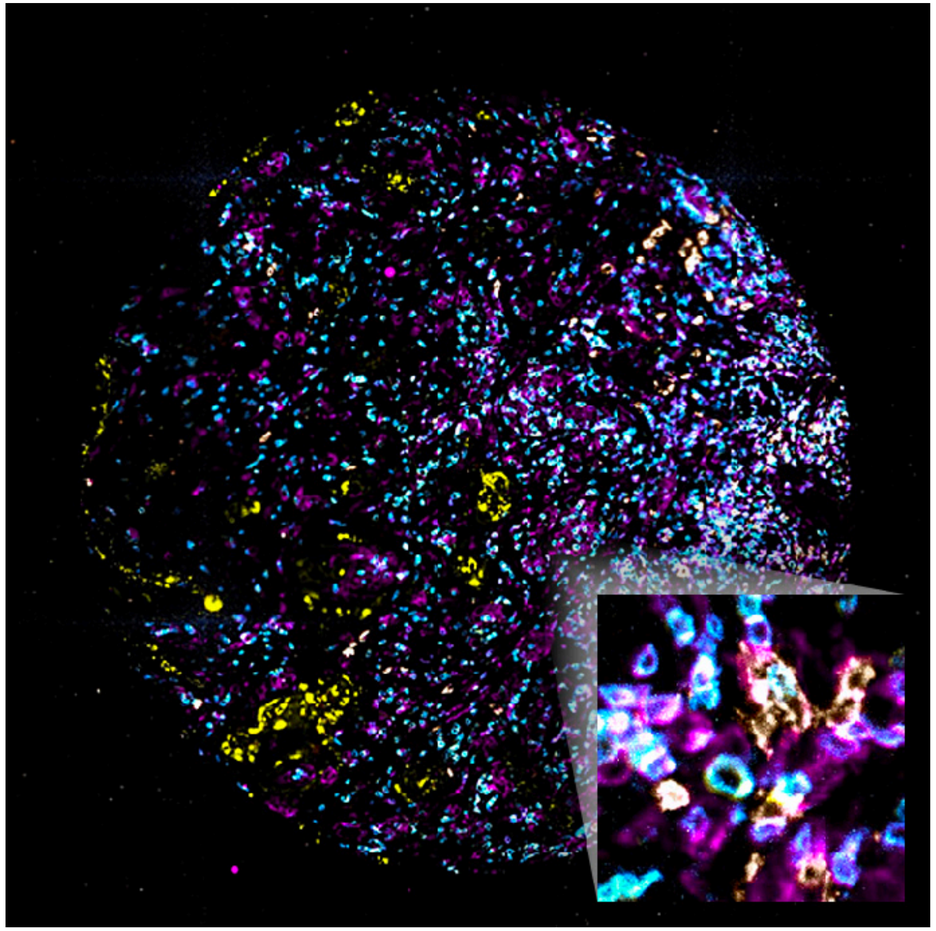
- Tumore (tessuto)
- Immagine per immunofluorescenza
- Composizione cellulare

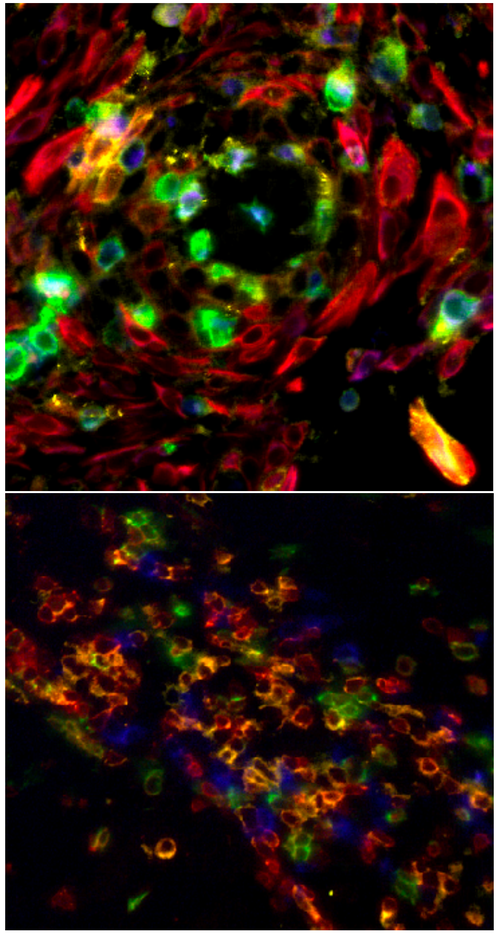


https://www.nature.com/articles/d42473-020-00071-6

La patologia per immagini

- Impossibile da analizzare senza IA - troppo complicato!

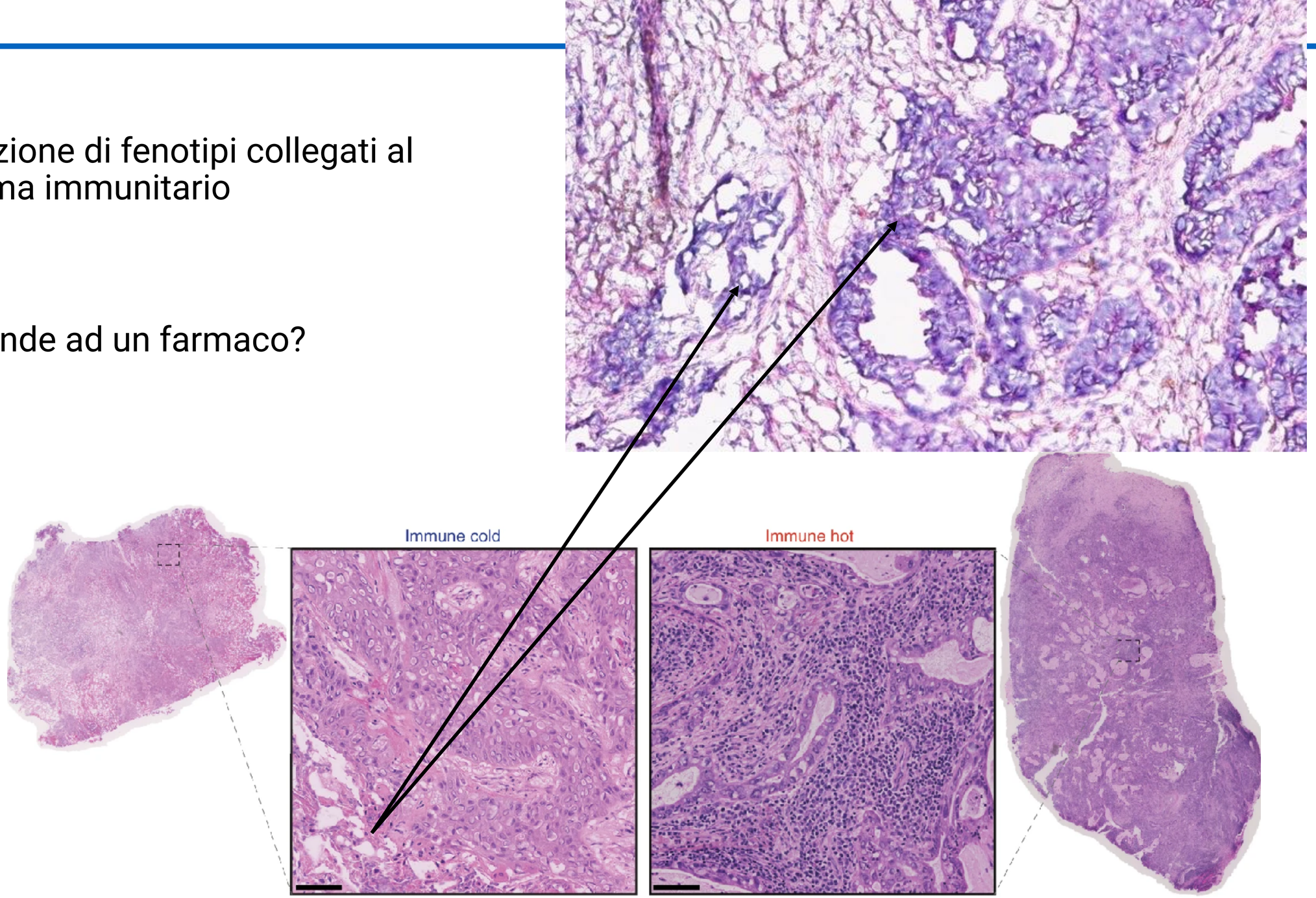




https://cellidx.com/gallery/ultraplex-gallery-human-immune-subsets

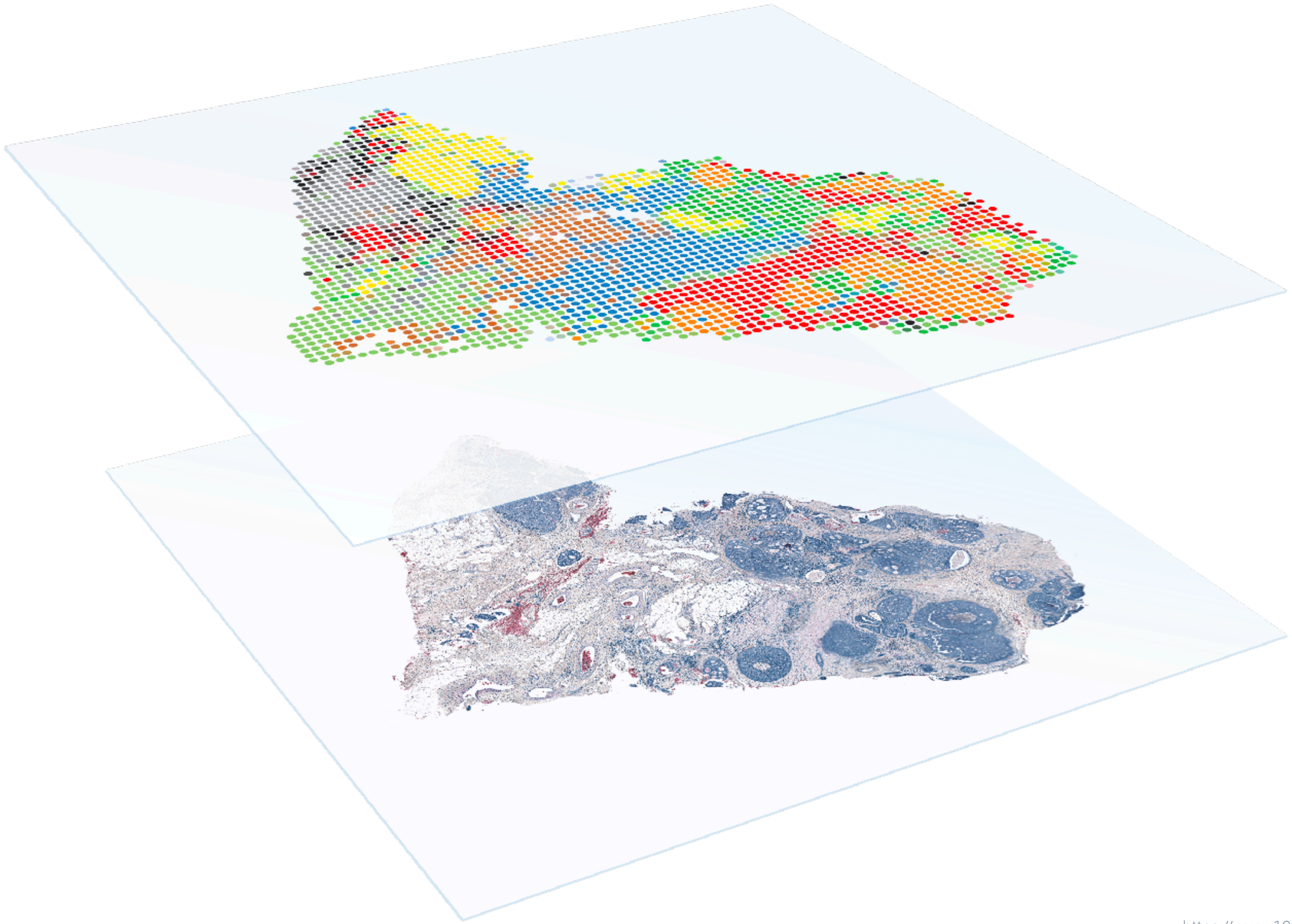
- Predizione di fenotipi collegati al sistema immunitario

- Risponde ad un farmaco?



https://www.nature.com/articles/s41591-020-0900-x/figures/2

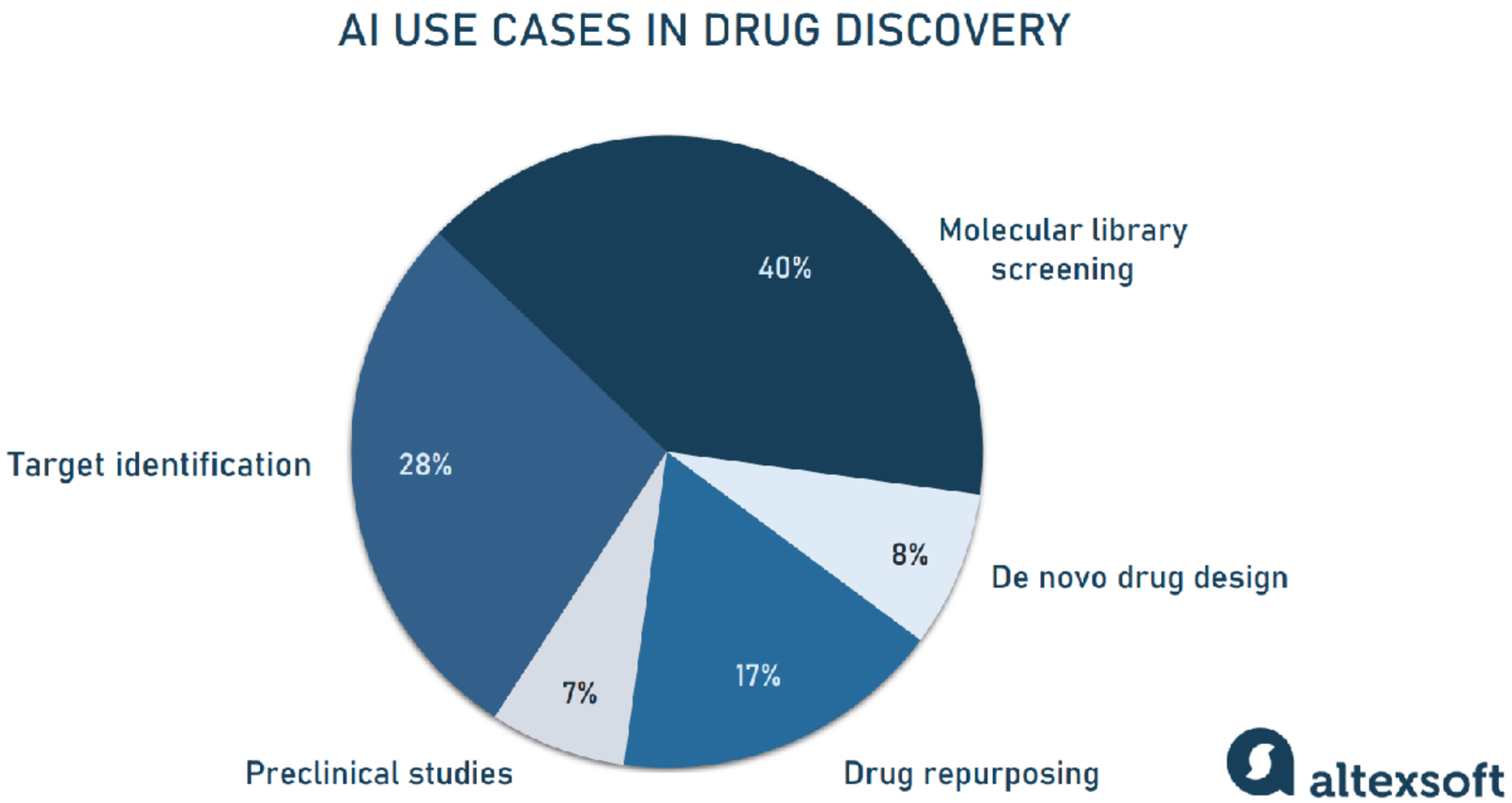
Immagini + dati molecolari



https://www.10xgenomics.com/products/spatial-gene-expression

Drug discovery

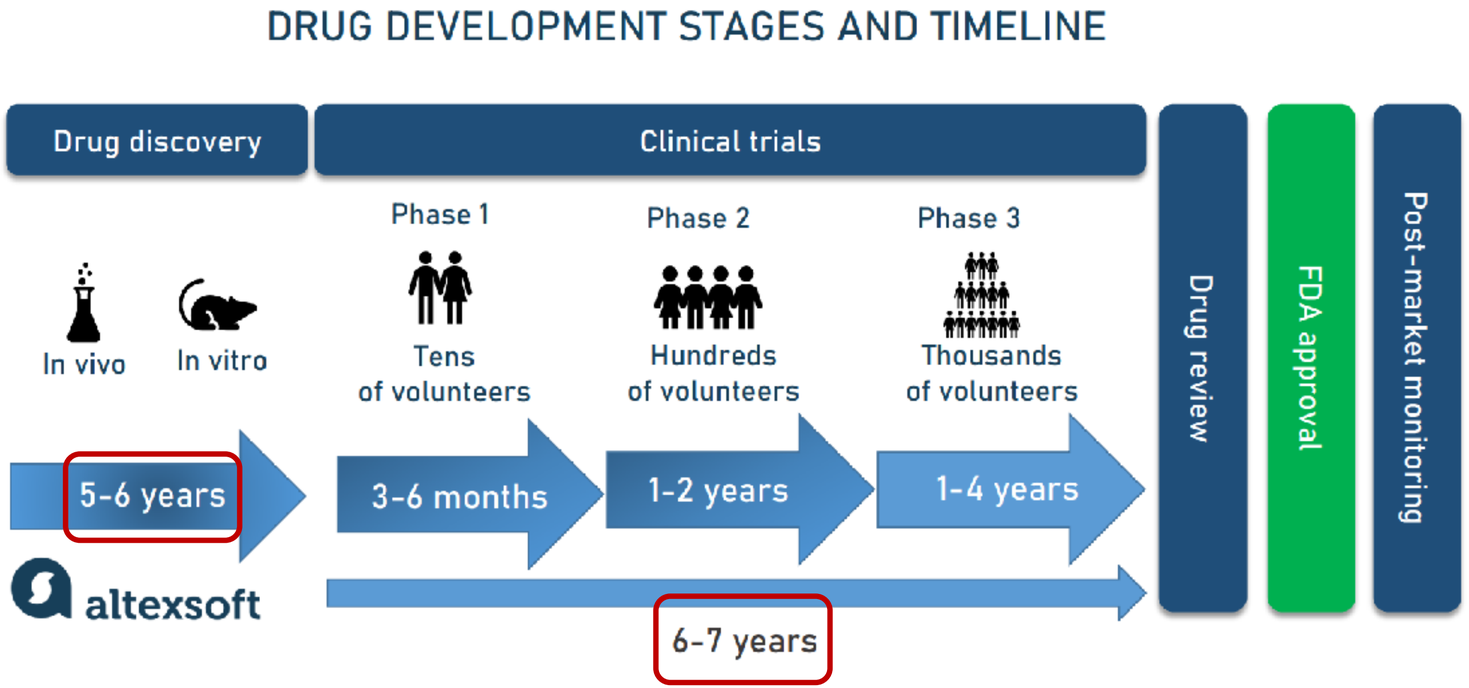
Drug discovery



https://www.altexsoft.com/blog/ai-drug-discovery-repurposing/

Drug repurposing

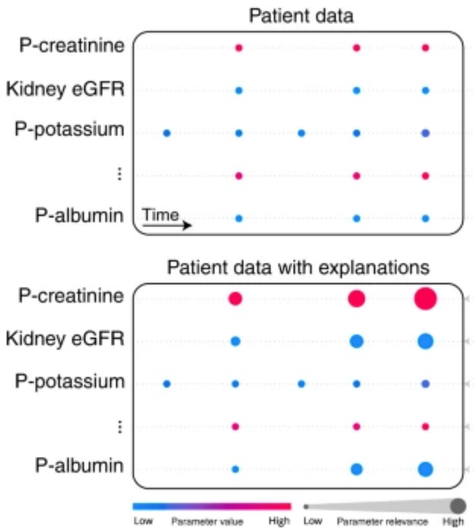
- Esplorazione di grandi librerie di molecole
- Esempio: farmaco per artrite reumatoide utilizzato per casi gravi di COVID-19 (scoperto in 48 ore)



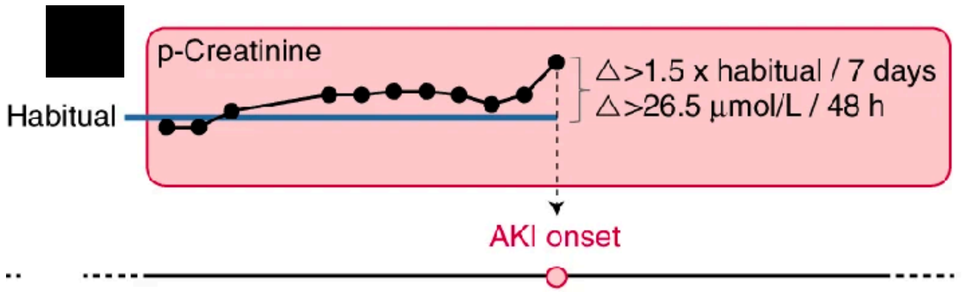
https://www.altexsoft.com/blog/ai-drug-discovery-repurposing/

Altre applicazioni

- Diagnosi **parzialmente** automatizzata
- Diagnosi **precoce**
- Sviluppo di farmaci (**in silico**)
- **Automazione** di compiti ripetitivi



La **traiettoria** del paziente ….



https://www.thirdway.org/report/a-lifetime-electronic-health-record-for-every-american

https://www.nature.com/articles/s41467-020-17431-x

Un ragionamento finale

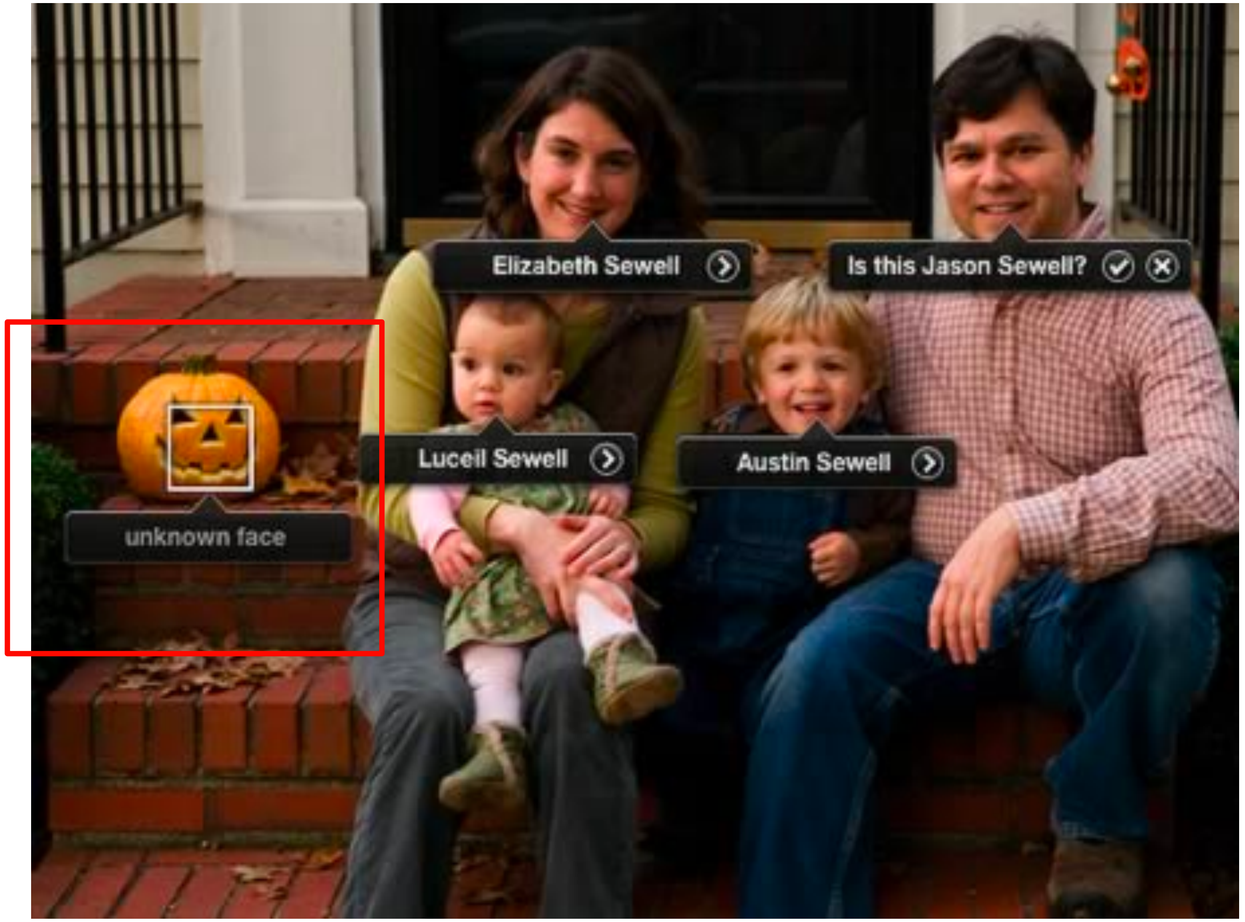
Un ragionamento più critico

- Possiamo automatizzare la diagnostica ed il trattamento usando tecnologie di IA?

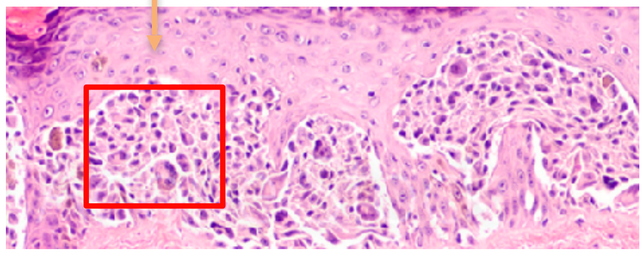
Il costo di un errore

- ERRORE

Taggo un volto sbagliato



Sbaglio una diagnosi



Il bias (apprendimento supervisionato)

Il campione su cui “apprendo” è “rappresentativo”?

Prof. Giulio Caravagna
(PI)

Dr. Riccardo Bergamin
(PostDoc)

Dr. Nicola Calonaci
(PostDoc)

Dr. Alice Antonello
(PhD)

Dr. Salvatore Milite
(PhD)

Dr. Arianna Tasciotti
(PhD)

Dr. Elena Buscaroli
(PhD)

Dr. Giovanni Santacatterina
(PhD)

Dr. Lucrezia Valeriani
(PhD)