

Application of Bioinformatics for Biomedical Sciences



By Ravin Poudel

True Zen of Bioinformatics

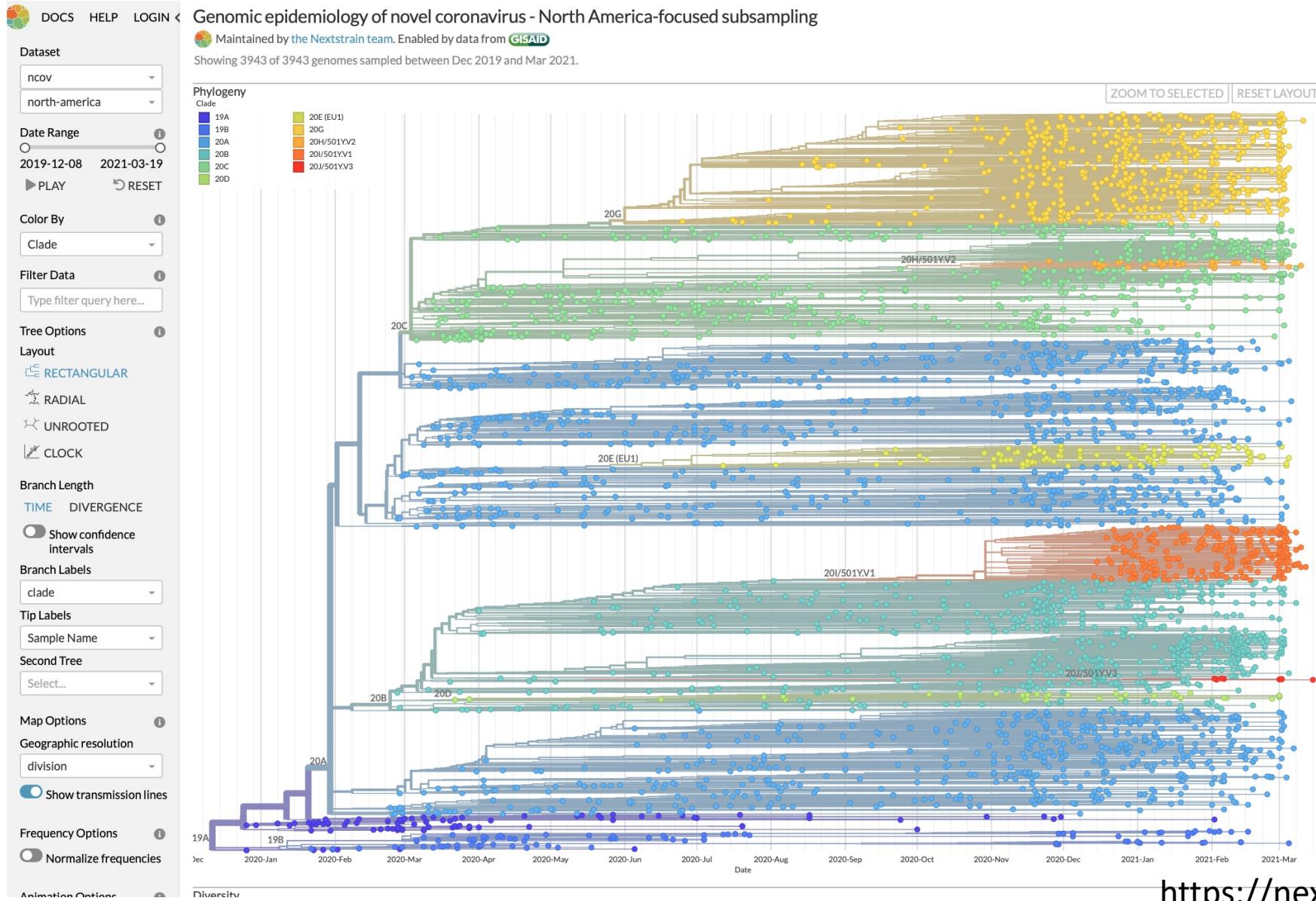
Nextstrain

Real-time tracking of pathogen evolution

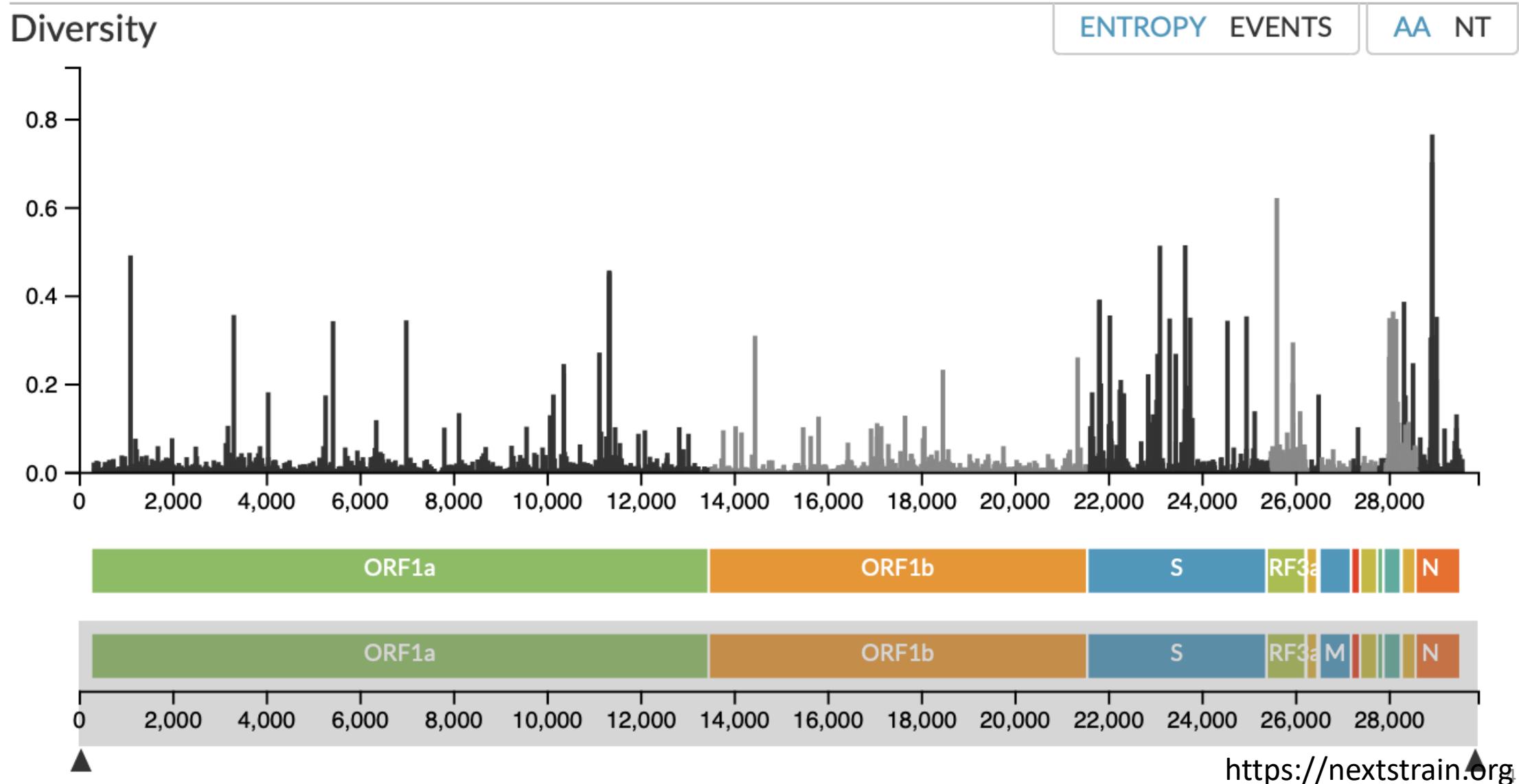
Nextstrain is an open-source project to harness the scientific and public health potential of pathogen genome data. We provide a continually-updated view of publicly available data alongside powerful analytic and visualization tools for use by the community. Our goal is to aid epidemiological understanding and improve outbreak response. If you have any questions, or simply want to say hi, please give us a shout at hello@nextstrain.org.

<https://nextstrain.org>

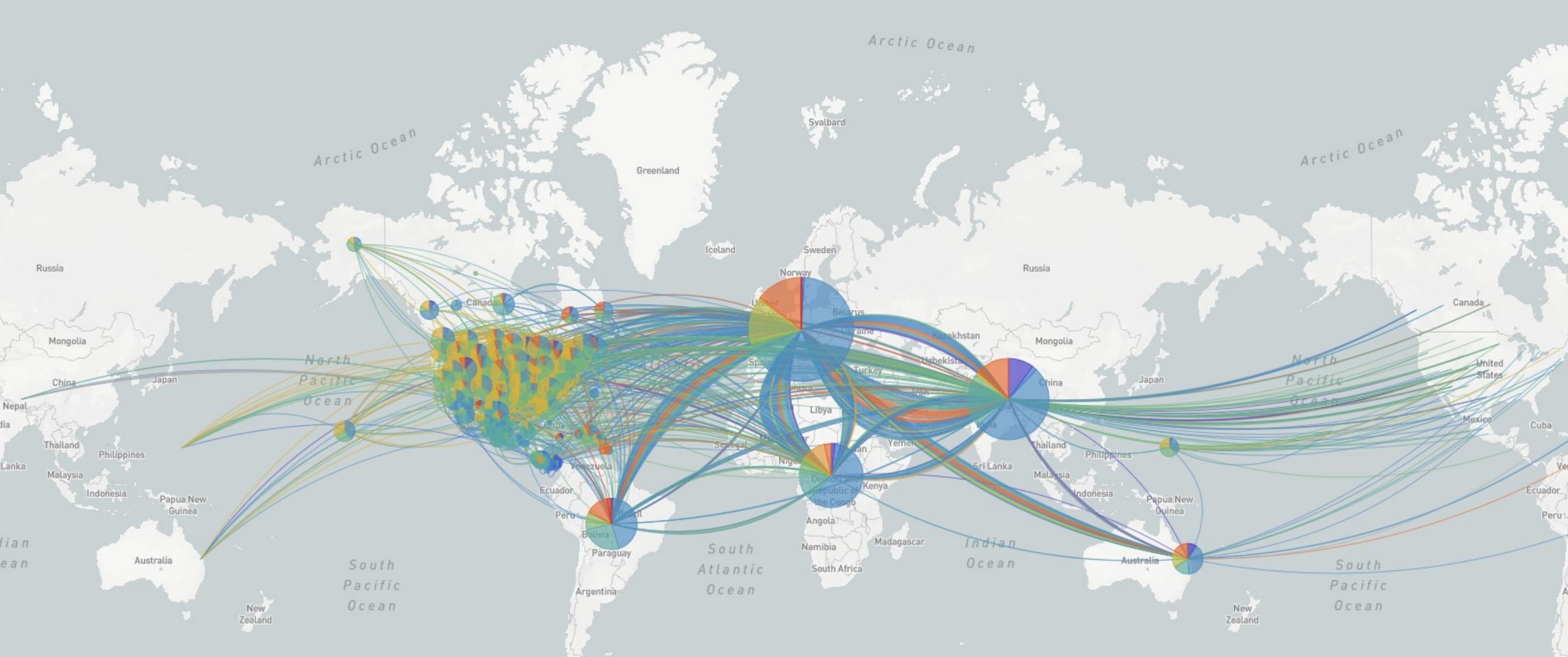
True Zen of Bioinformatics



True Zen of Bioinformatics



True Zen of Bioinformatics



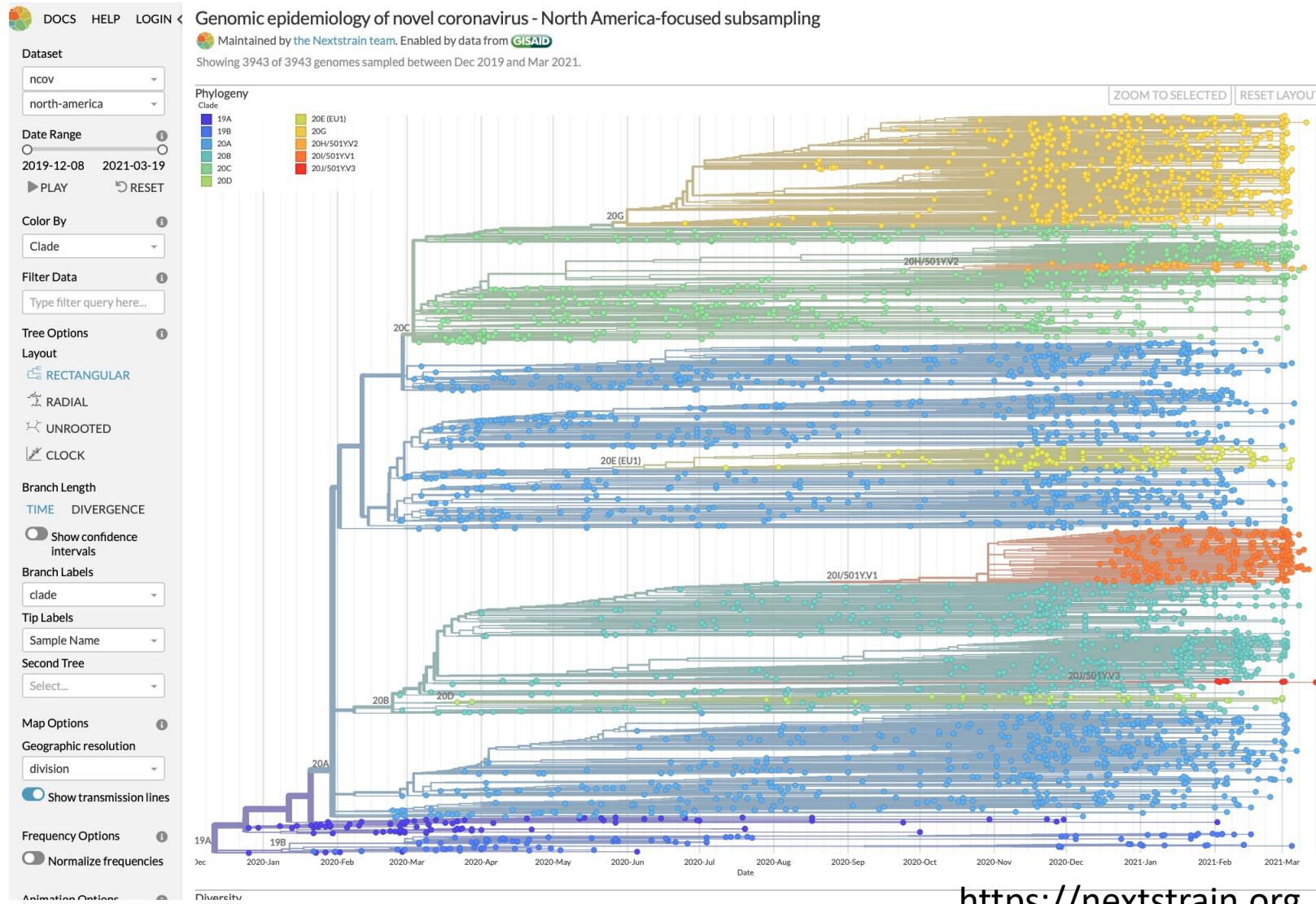
<https://nextstrain.org>

True Zen of Bioinformatics

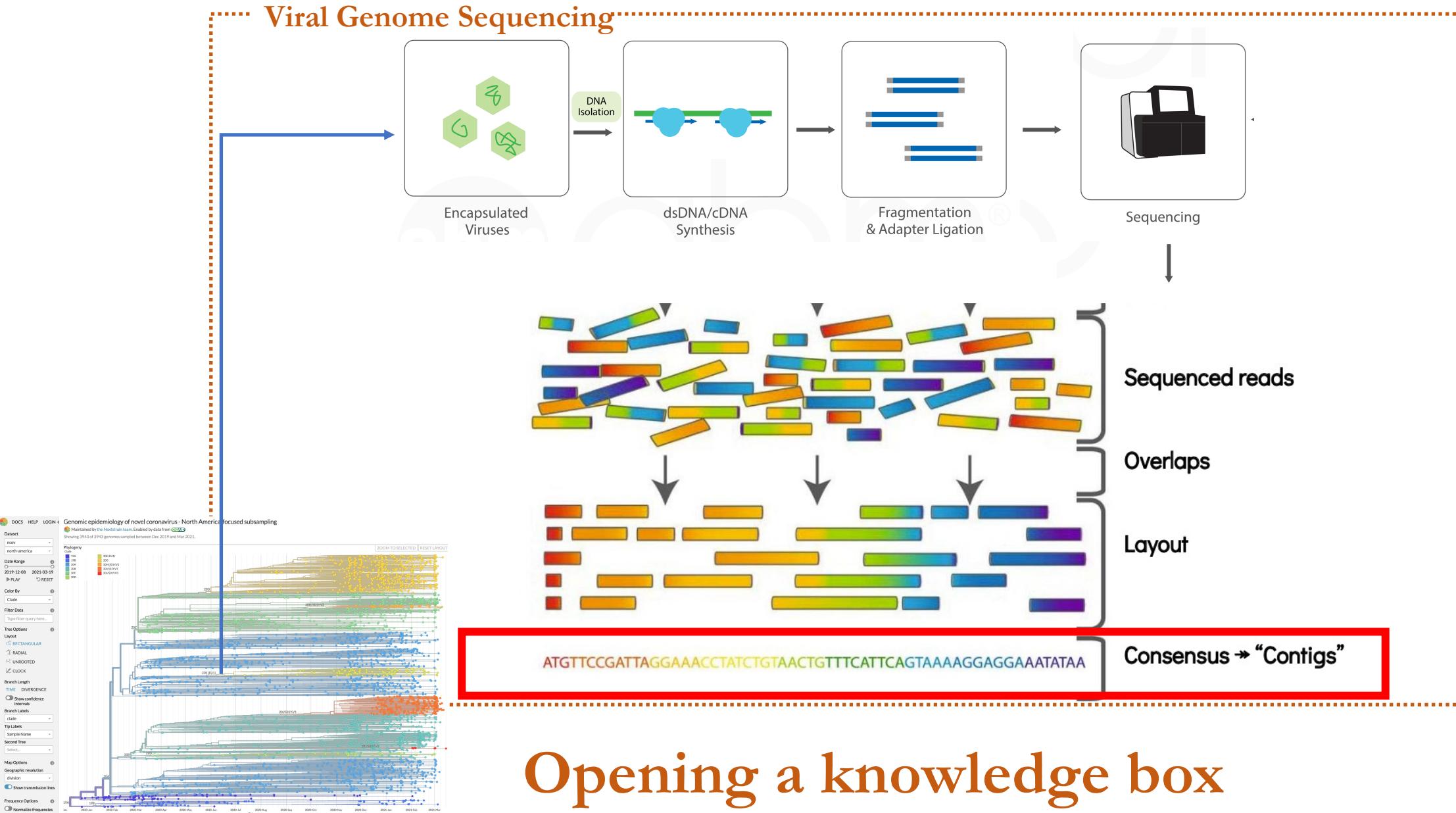
- What does it take to add a single dot in this picture? Or phylogenetic tree?
- What is the probability one can generate this analyses without the bioinformatician in the team?

Probably **less** possible

We need to have a **bioinformatician** in the room



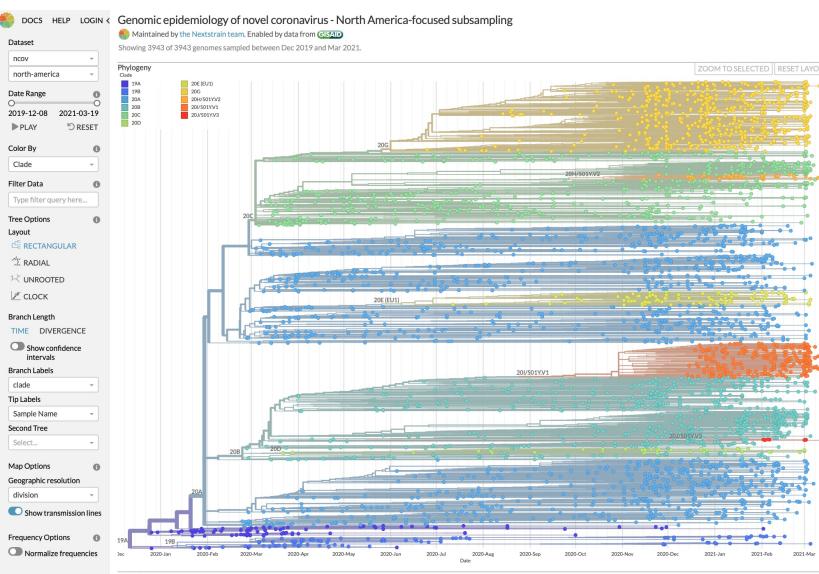
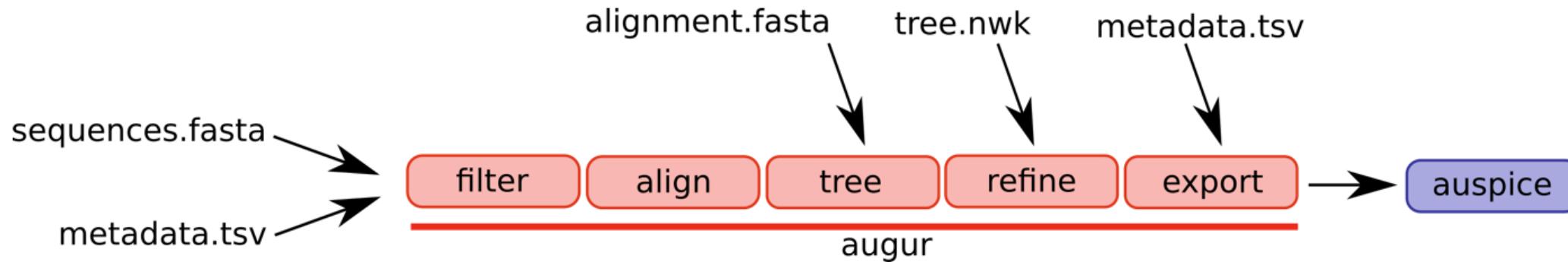
Process involved to add a dot to the figure



Assembled genome: Opening a knowledge box

- Produce a reference for new genome
 - Human reference genome
- Allows for comparative studies against the reference
 - Variant analyses
- Coding sequences vs non-coding
- High level genome structure (number, arrangement of genes and repeats)
- A wide variety of **molecular biological tools** become available
 - CRISPR gRNA search
 - Early diagnosis - biomarkers
 - Prognosis- biomarkers

Current process in next strain is good, but not efficient method



What is the problem here?

Time!

What is time?

- Patient's life
- Possibility of preventing the next pandemic
- Financial burden

How many dots are being collected each day?

200,000 and counting: how the UK has sequenced so many cases of coronavirus

“ **E**ach time the virus infects somebody, it has an opportunity to make a mistake.

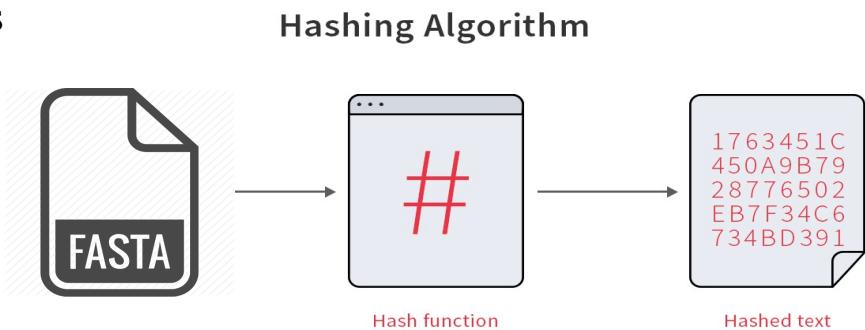
A White House document circulating in Washington, D.C., last week calls for spending \$340 million to sequence 5% of samples from confirmed COVID-19 cases, which would be about 39,000 genomes per week at current infection rates. But some scientists support a \$2 billion request to sequence

Source: <HTTP://www.sciencemag.org>

“The Biden administration is doing a lot, but ... it is not yet sufficient,” says **Rob Knight**

The counting Problem

- Cardinality :
 - How many unique genomes are there in SRA archive? For covid virus
- Who else might have already solved such problem ?
 - Facebook : Counting new users
 - Google: Counting new visitor of any www page



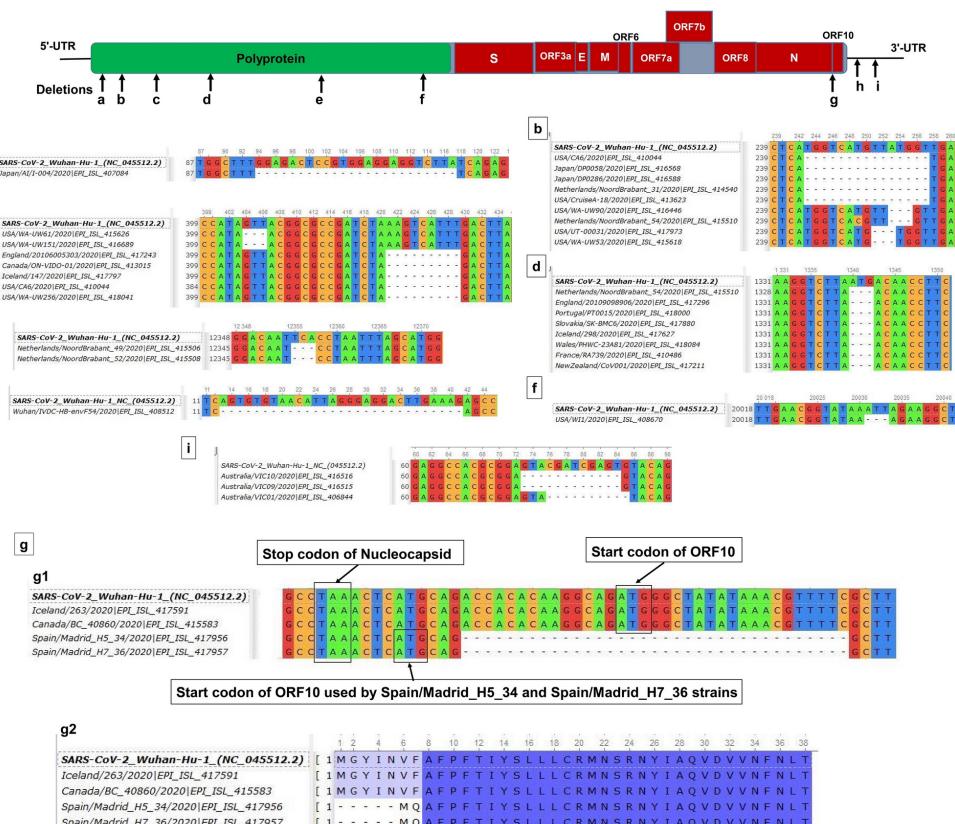
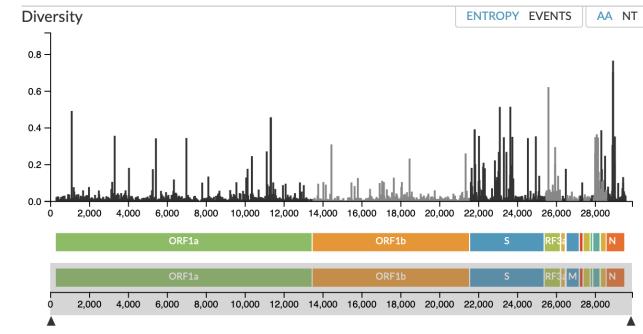
- Improvements: Data Sketches (HyperLogLog and MinHash)

- MinHash - small "signatures" of data sets that can be used to compare data sets quickly
- HyperLogLog algorithm (Probabilistic data structure)
 - Fast
 - 10^9 with a relative accuracy
 - Standard error of 2% while only using 1.5kb of memory
- Run a surveillance server – 24 hours a days, 7 days a week
- Fetch data from SRA archive > Data Sketches > HLL > Jaccard/mash distance > Tree > unusual/new suspect **ALERT**
- Can also be used for – tracking or early detection of super bugs that are resistance to antibiotics
- Finding samples of interest e.g. With similar expression patterns :
 - Search SRA collection using signatures from own RNAseq experiments

Drug Design: Arms race; Red Queen Hypothesis

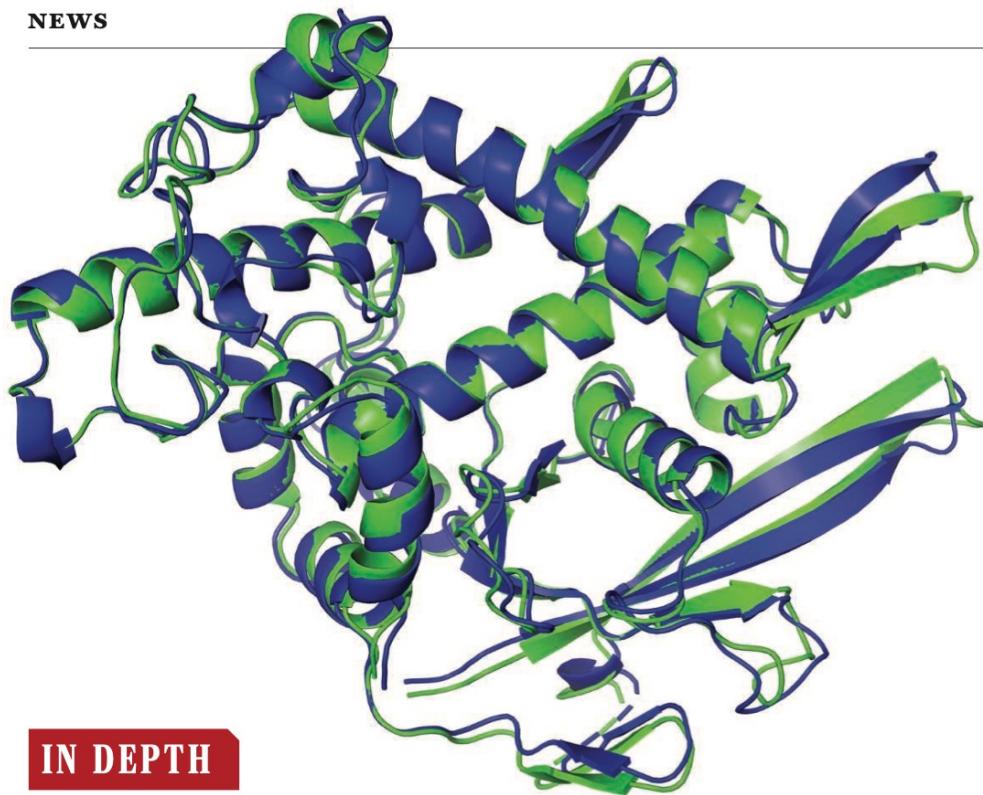
What does it mean to have these differences in the multiple sequence alignments?

- How does this effect the efficacy of vaccines or drugs?
- Does this make the pathogen more virulent?
- How should we modify the composition or the structure of our vaccines or drugs?
- How does those changes impact the binding or docking of drugs on the targets?
- Selection of markers/targets for drug design
- In-silico studies of drug binding



Red Queen Hypothesis: Maybe we can win

NEWS

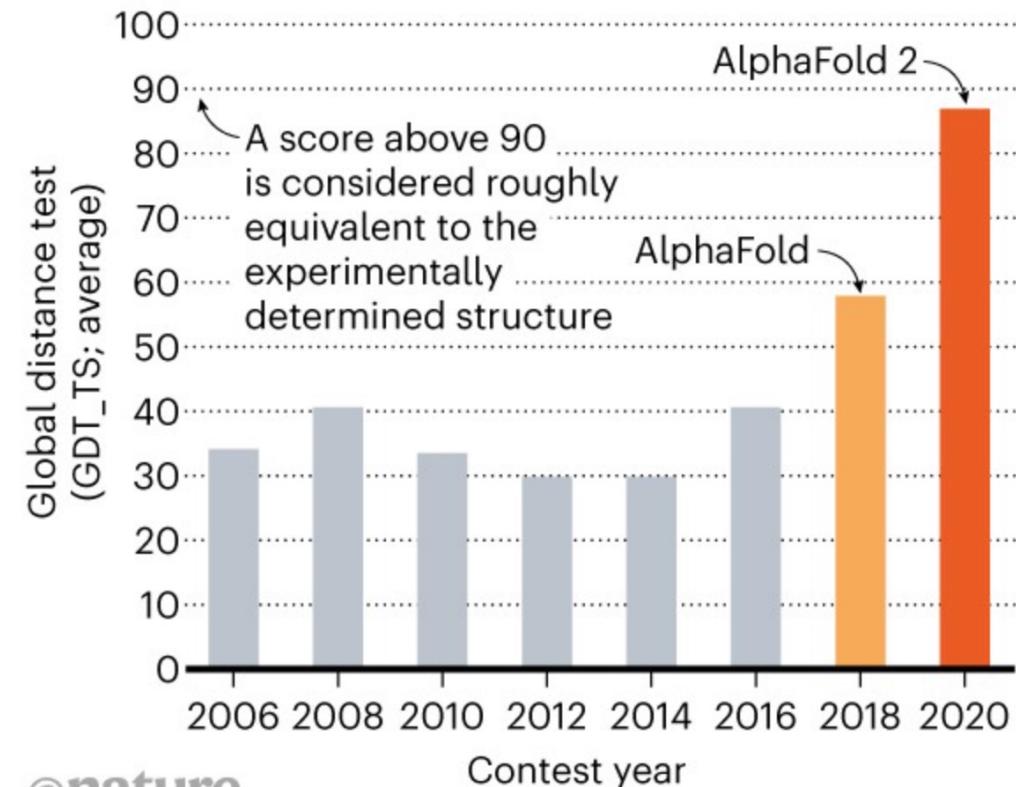


IN DEPTH

STRUCTURAL BIOLOGY

'The game has changed.' AI triumphs at protein folding

In milestone, software predictions finally match structures calculated from experimental data



Source: DeepMind

We have a Vaccine: Will it work? How to test?

- Are there side effects? How does vaccine work?
 - RNAseq experiments
- Will drug or vaccine work in patients / individual expressing certain molecular marker?
 - Personalized medicine
 - **Imprecision medicine**
- Expanding the reach of drugs or therapeutics to other similar systems with similar clinical manifestation
- In some cases, only subpopulation express the targets or molecular marker upon which the drugs are designed.
 - Clinical recruitment
 - Stratification of samples
 - Experimental design
 - Effective sample size

Imprecision Medicine is a Reality

Every day, millions of people are taking medications that will not help them. The top ten highest-grossing drugs in the United States help between 1 in 25 and 1 in 4 of the people who take them (see ‘Imprecision medicine’). For some drugs, such as statins — routinely used to lower cholesterol — as few as 1 in 50 may benefit¹.

“We can do better, especially by applying bioinformatics”.

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (ariPIPRAZOLE)
Schizophrenia



2. NEXIUM (esomeprazole)
Heartburn



3. HUMIRA (adalimumab)
Arthritis



4. CRESTOR (rosuvastatin)
High cholesterol



5. CYMBALTA (duloxetine)
Depression



6. ADVAIR DISKUS (fluticasone propionate)
Asthma



7. ENBREL (etanercept)
Psoriasis



8. REMICADE (infliximab)
Crohn's disease



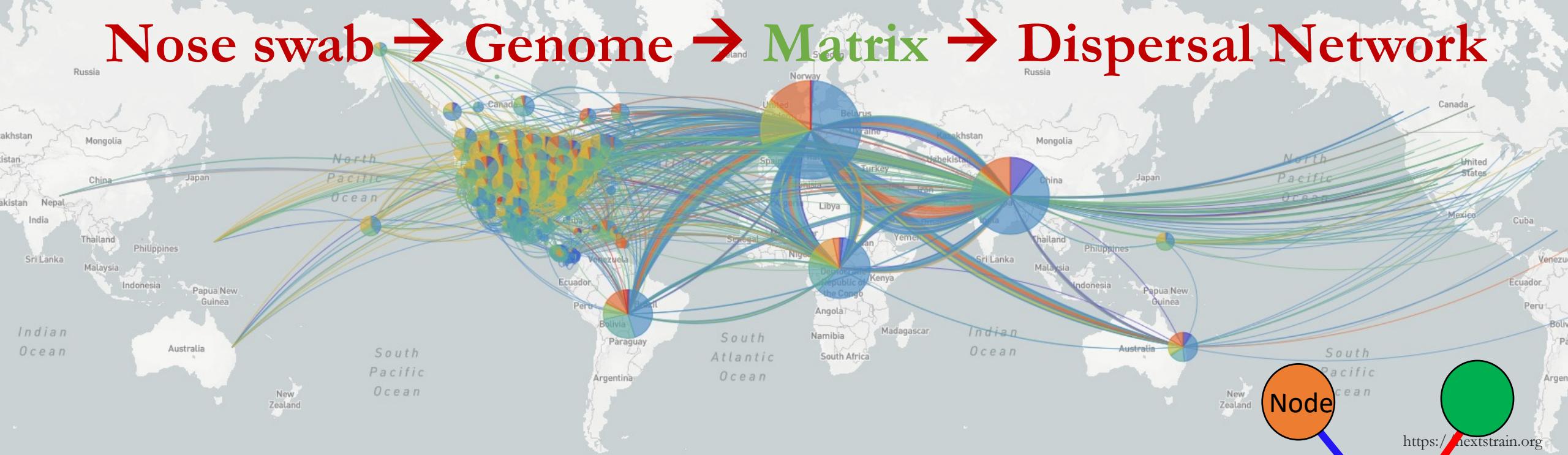
9. COPAXONE (glatiramer acetate)
Multiple sclerosis



10. NEULASTA (pegfilgrastim)
Neutropenia

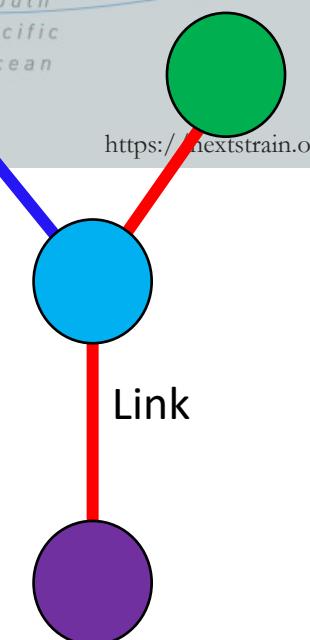


Nose swab → Genome → Matrix → Dispersal Network



Network – system perspectives ~ informed decision

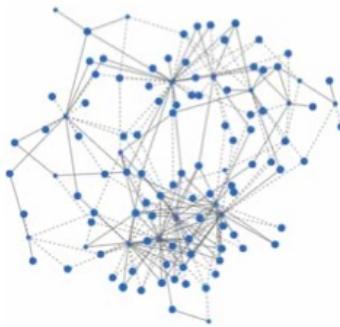
- How the various elements interact in the systems?
- Identifies key nodes or group of nodes that could be useful
 - Where to sample / hub ? Catch the “Super-Spreaders” early
 - Disease dispersal network- which node to manage such that spread will be minimal
 - Vaccination network- how to vaccinate individuals such that we can get effective coverage
 - Resource optimization ~ communication grid management



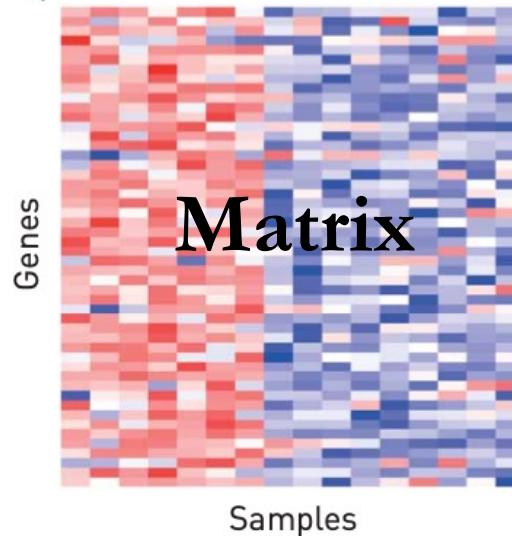
Network analyses with omics data

a)

Pearson/Spearman coexpression network



Defining Links



b)

Coexpression matrix

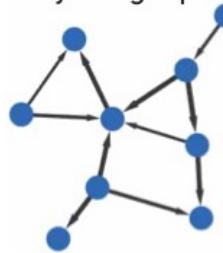


c) | Prior knowledge

x_1	x_2	...	x_n
x_1^1	x_2^1	...	x_n^1
x_1^2	x_2^2	...	x_n^2
x_1^3	x_2^3	...	x_n^3
...
x_1^m	x_2^m	...	x_n^m

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Directed probabilistic acyclic graph



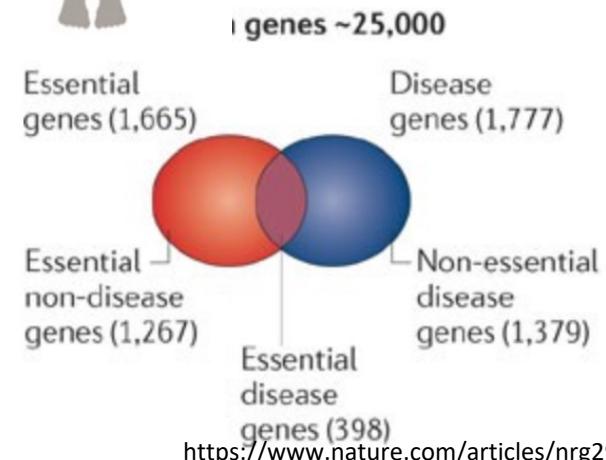
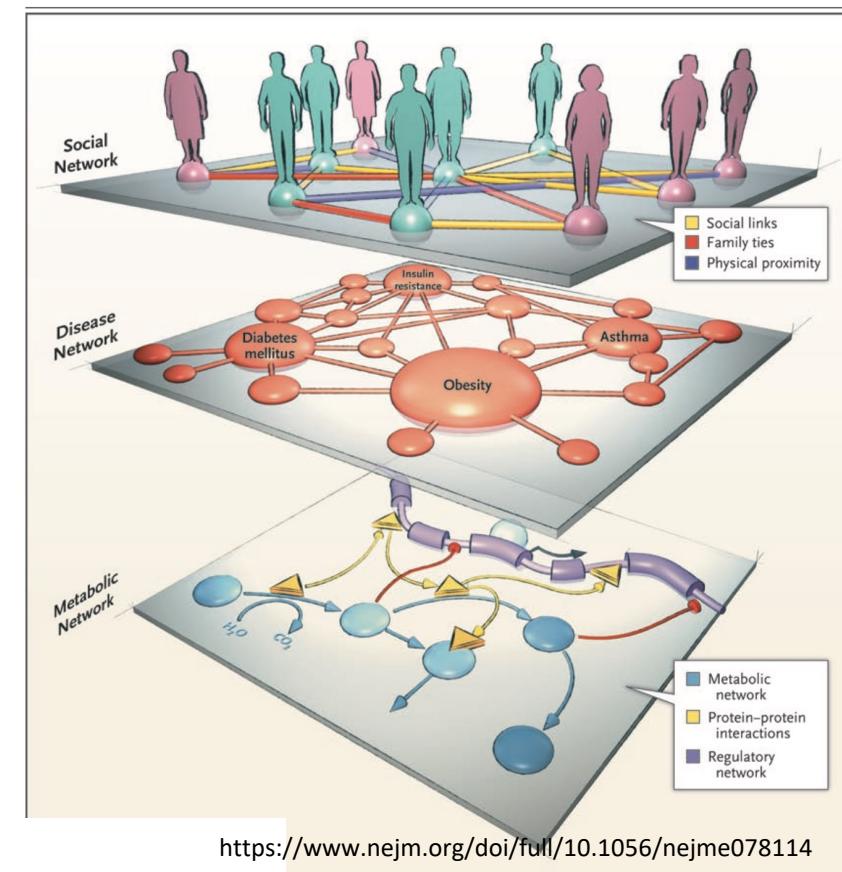
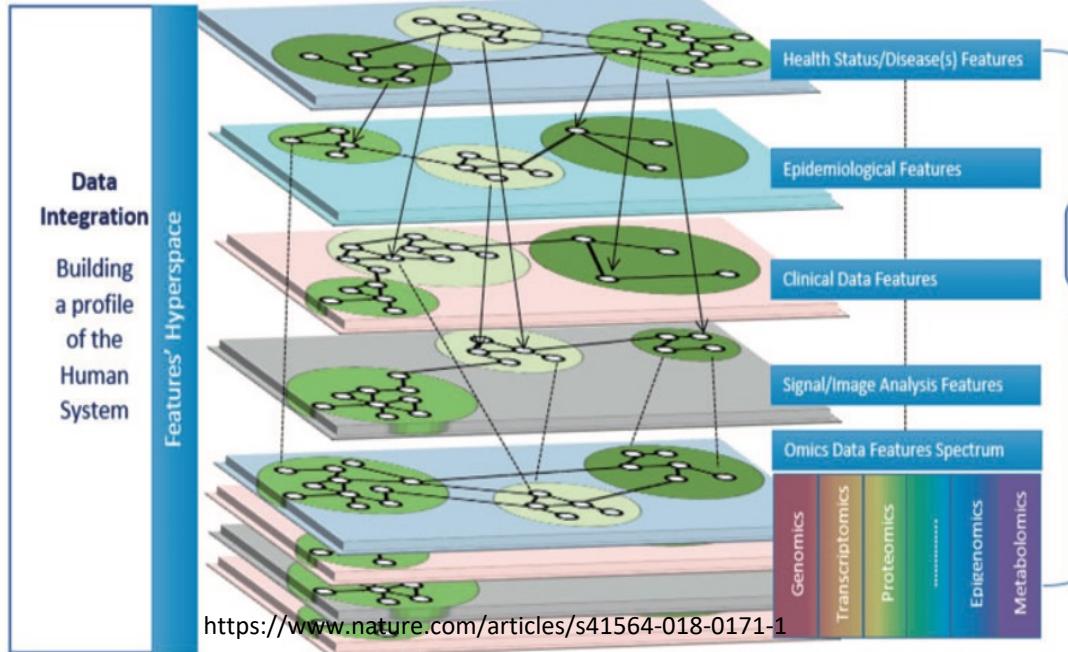
Untapped areas that could be useful

- ERGM: Exponential Random Graph Models-family of statistical models for analyzing networks

- Null models

- Identify key nodes modules
- Association of key genes across multiple diseases

Network medicine: a network-based approach to human disease



Robust database management: Semantic analyses and knowledge graph- even extremely useful for rare diseases

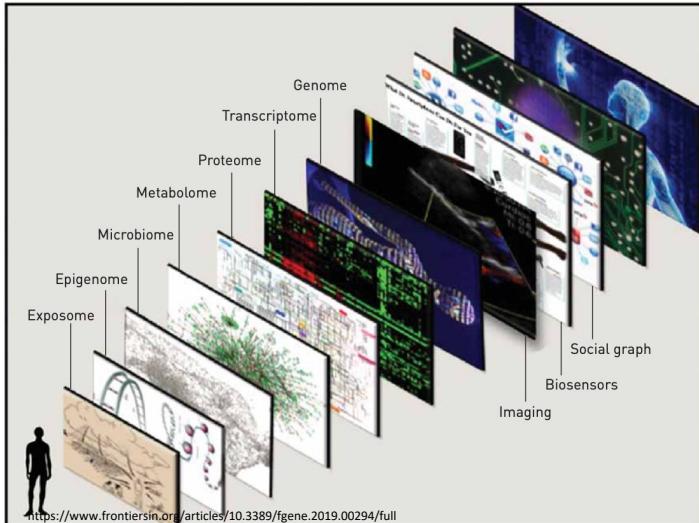
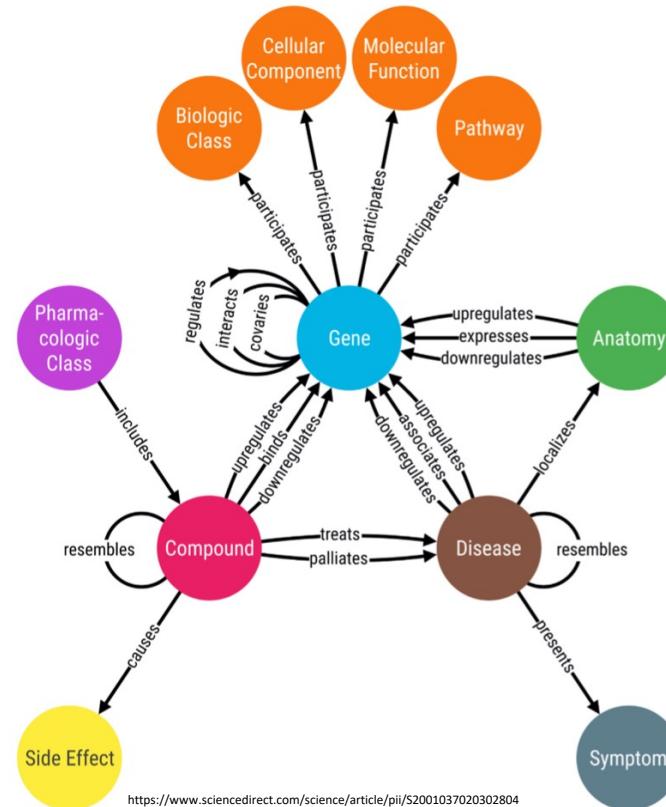


FIGURE 1 Multilevel layers of biological, environmental and social information ideally integrated in systems biomedicine approaches. For further explanations, see text. Reproduced and modified from [2] with permission.



<https://www.sciencedirect.com/science/article/pii/S2001037020302804>

Benjamin Franklin



Founding Father of the United States

Benjamin Franklin FRS FRSA FRSE was one of the Founding Fathers of the United States. A polymath, he was a leading writer, printer, political philosopher, politician, Freemason, postmaster, scientist, inventor, humorist, civic activist, statesman, and diplomat.

[Wikipedia](#)

Born: January 17, 1706, Milk Street, Boston, MA

Died: April 17, 1790, Philadelphia, PA

Vice President: Charles Biddle; Peter Muhlenberg; David Redick

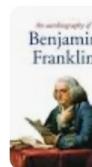
Spouse: Deborah Read (m. 1730–1774)

Instruments: Glass Armonica [fi.edu](#)

Books



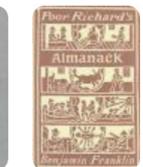
The Autobiog...
of Benja...
1791



Benjamin Franklin's
1789



Quotations of
Benjam...
2003



Poor Richard's
Almanack
1732

[View 45+ more](#)

People also search for



Thomas Jefferson



George Washingt...



John Adams



Alexander Hamilton

Application of bioinformatics in biomedical research

- Basic and fundamental biology
 - Protein
 - Metabolites
 - Many kinds of omics....
- Find the right biomarker and targets for drugs development
- Identification of molecular, genetical, or metadata level features that allows for **resource optimization**.
 - Sample size optimization
 - Finding right group or subgroup of populations for clinical trials.
 - Personalize medicine – one drug does not fit all the patients of same disease type
- Finding features predictive of future disease or early diagnosis of diseases.
- Preventing outbreak of diseases and supporting clinical epidemiology.
 - Surveillance
 - Evolution of pathogens
- System or interactions-based understanding of layers of information
 - Network medicine
 - Graphical models
 - Knowledge graphs

