BOSTON UNIVERSITY
Virology, Immunology & Microbiology | Boston University Chobanian & Avedisian School of Medicine
Biomedical Engineering | College of Engineering
aTOMIClab
NEIDL

SIMON says: Advancing Human Immunology using AI

From Theory to Practice: AI in Immunological Cases

Module 4 – Day 1

Adriana Tomic

adrianatomic
@TomicAdriana
atomic@bu.edu
atomiclaboratory

atomic-lab.org

Module 4 – Day 1 - overview

Part I – Lecture plus Q&A

- Lecture on *"SIMON says: Advancing Human Immunology using AI"*
- Discussion

Lecture (3:30-4:15pm) ~45min

Discussion (4:15-4:30pm) ~15min

Part II – Installation of PANDORA software

- Installation & Troubleshooting

Hands-on (4:30-5:15pm) ~45min

BOSTON UNIVERSITY
Virology, Immunology & Microbiology | Boston University Chobanian & Avedisian School of Medicine
Biomedical Engineering | College of Engineering
aTOMIClab
NEIDL

SIMON says:

Advancing Human Immunology using AI

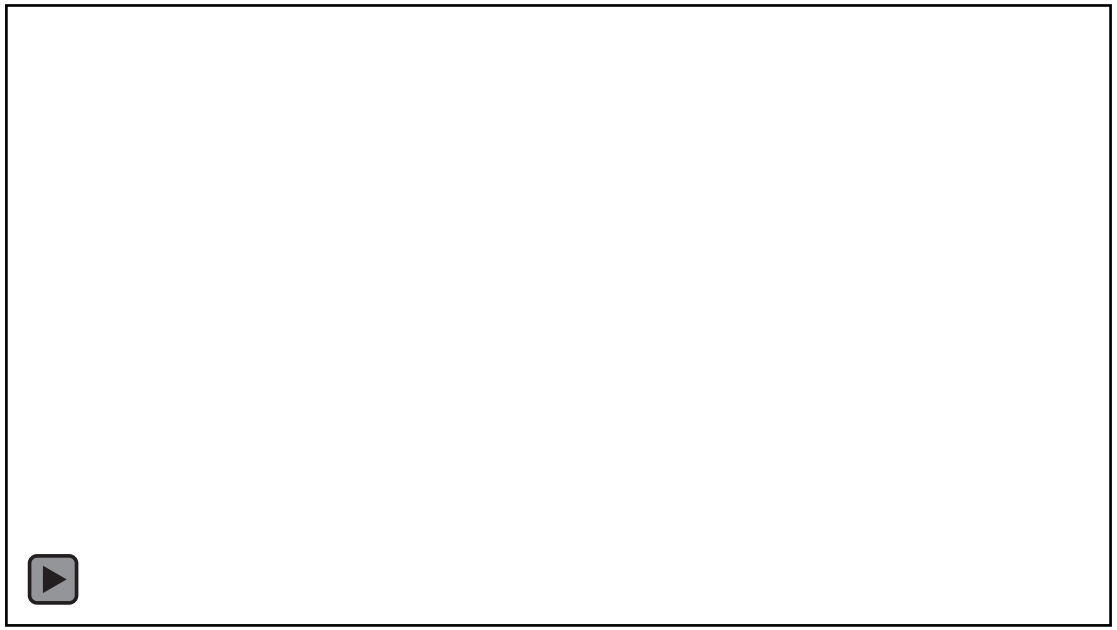
Adriana Tomic

adrianatomic
@TomicAdriana
atomic@bu.edu
atomiclaboratory

atomic-lab.org

Unifying biological scales to understand the biological function

Heart cells beating

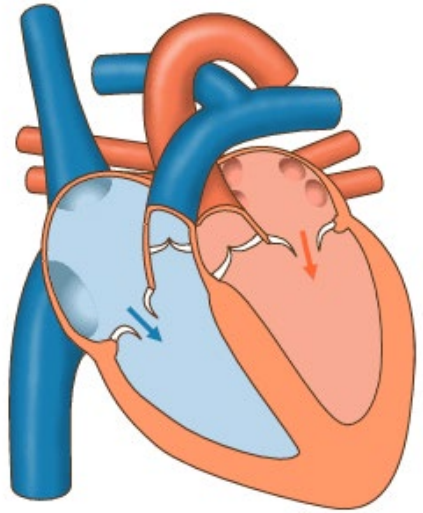


https://www.popularmechanics.com/science/health/a15071/artificial-heart-cells-beating-video-wake-forest/

Organ-on-a-chip project;
Anthony Atala group at the Wake Forest Institute for Regenerative Medicine, School of Medicine, NC, USA

Skardal A et al, Sci reports, 2017

Blood flow through the heart valves



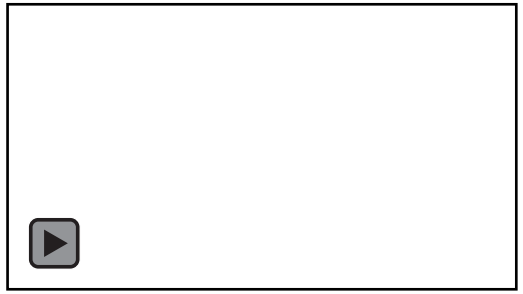
Josiño
https://en.wikipedia.org/wiki/Heart

Running motion

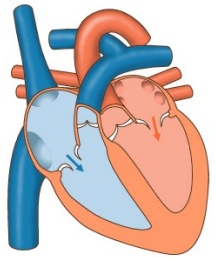
https://design.tutsplus.com/tutorials/animation-for-beginners-how-to-animate-a-character-running--cms-25730

Unifying biological scales to understand the biological function

Heart cells beating

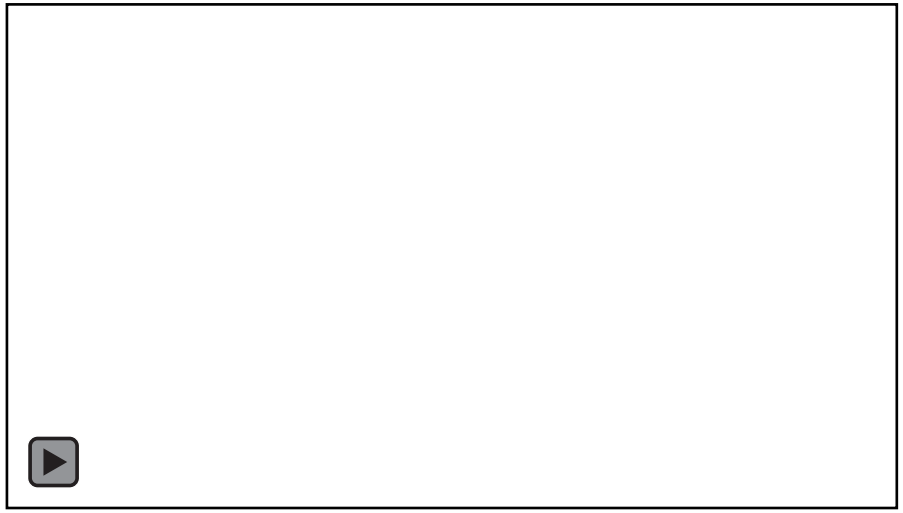


Blood flow through the heart valves

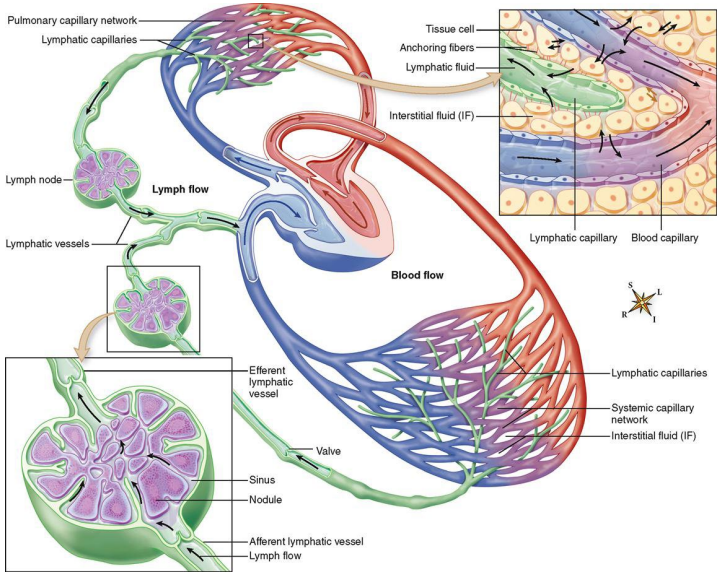


Running motion

T cells killing infected cells

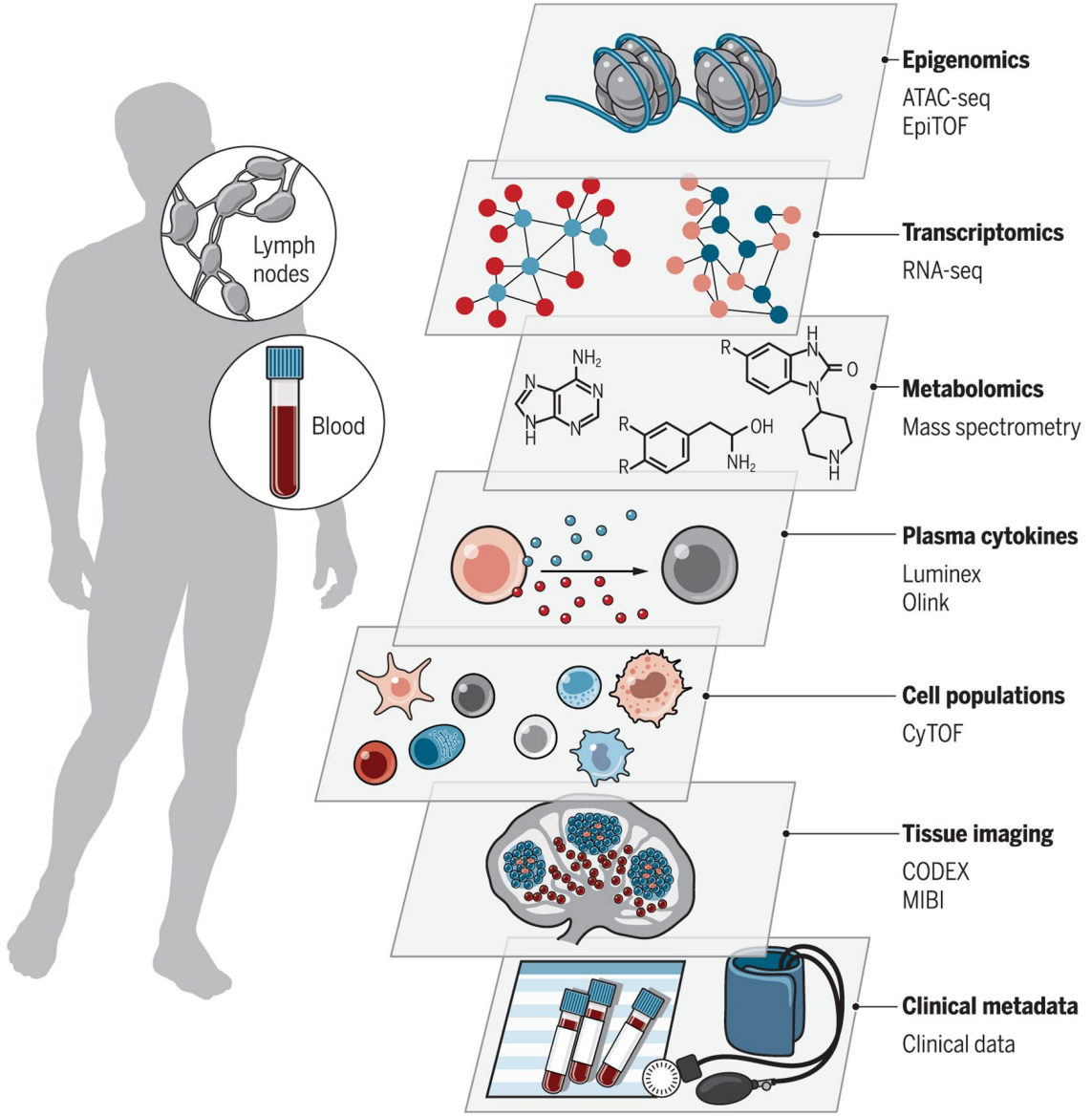


Lymph flow



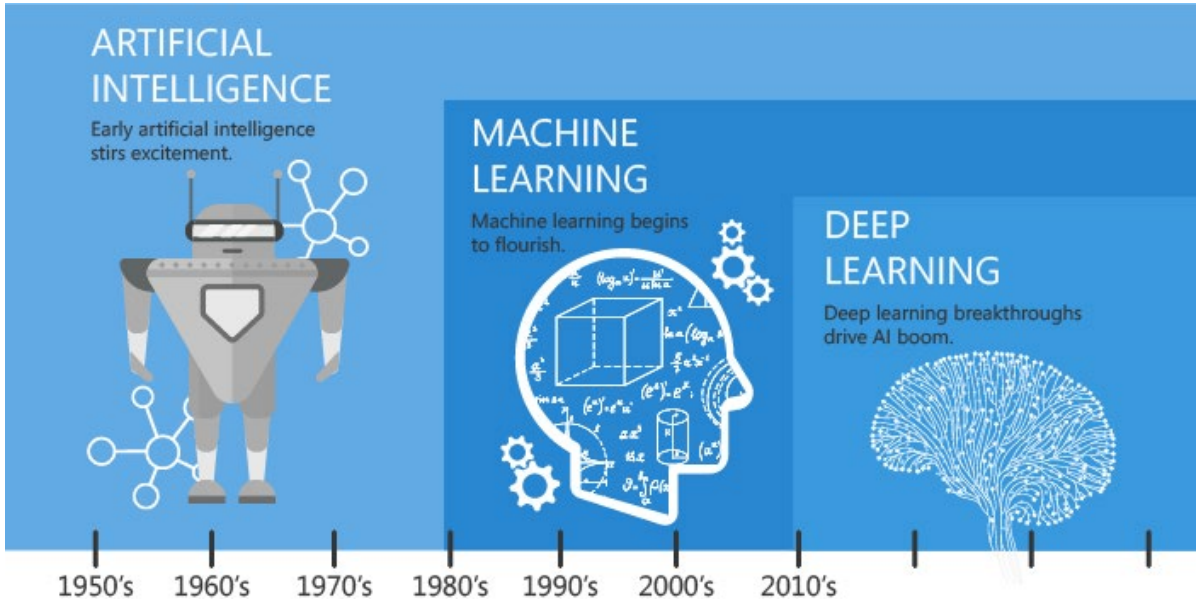
Protection = Immunity

HUMAN IMMUNOLOGY 2.0



Pulendran B and Davis M. Science, 2020

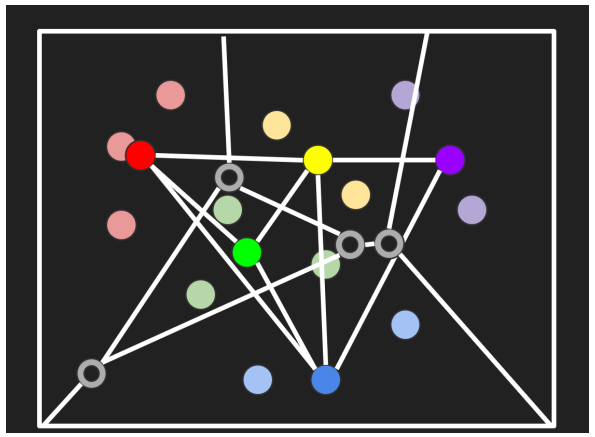
AI to the rescue!



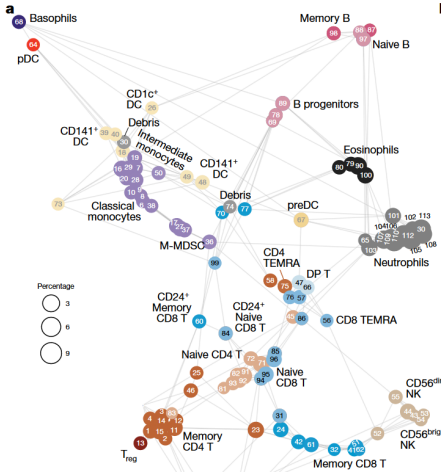
Kapil Tandon. AI & Machine Learning: The evolution, differences and connections.

Have you used ML algorithms?

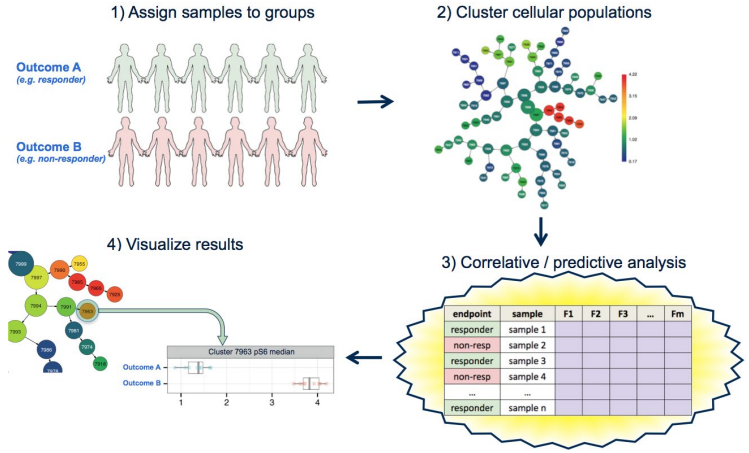
TEAM 1: JONAS SALK



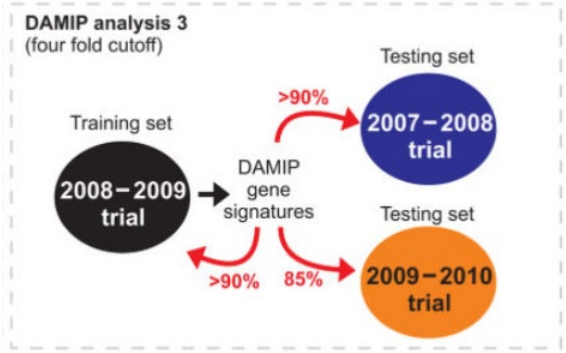
TEAM 3: LYNN MARGULIS



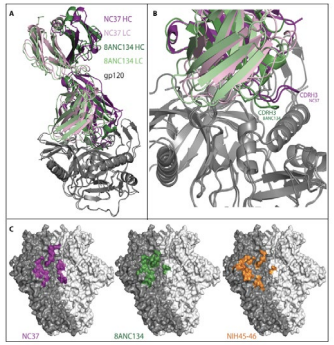
TEAM 5: POLLY MATZINGER



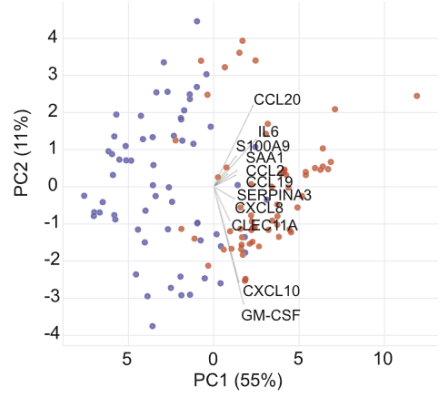
TEAM 2: SYDNEY BRENNER



TEAM 4: IRUN COHEN

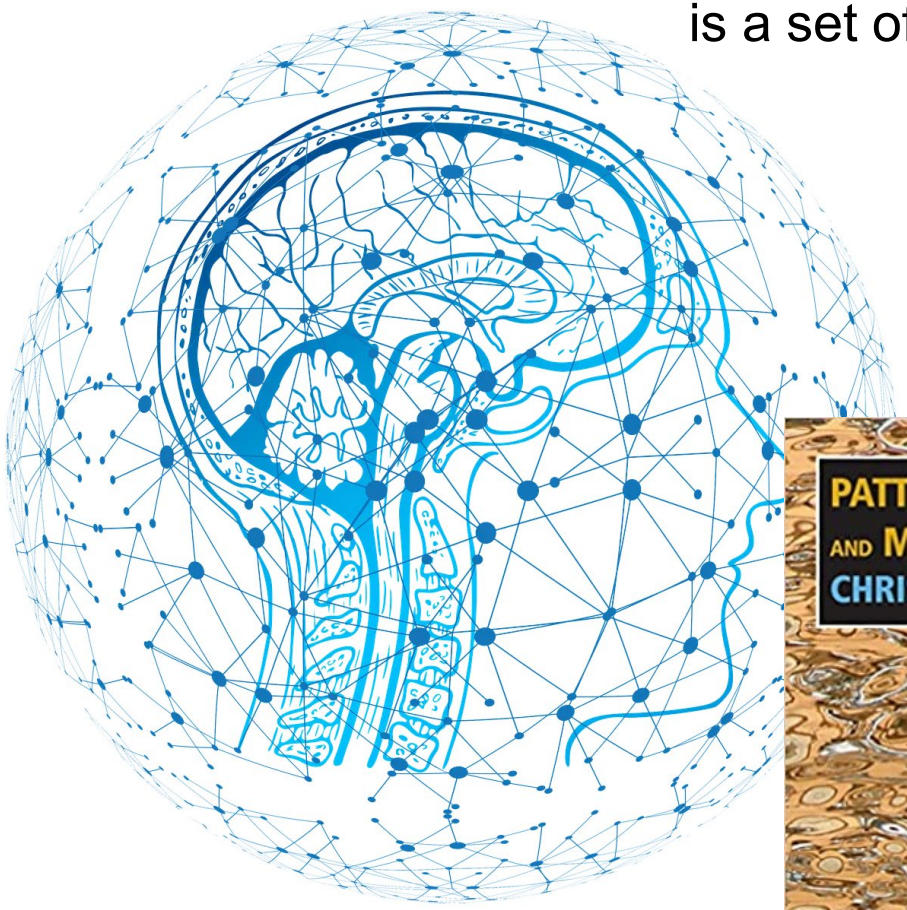


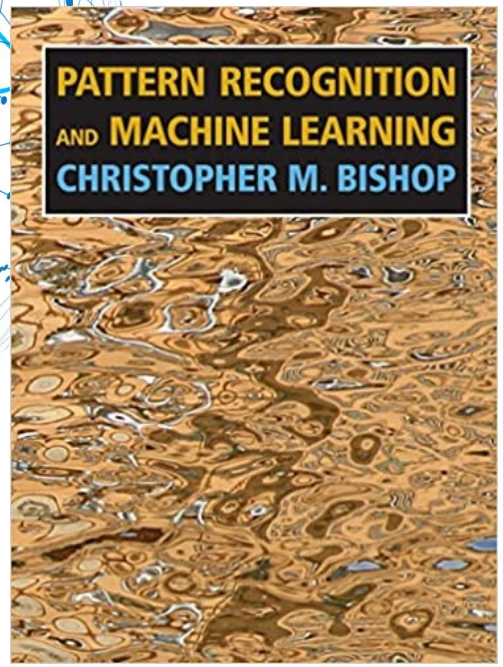
TEAM 6: DENIS NOBLE



Machine learning (ML), also known as **data mining** or **pattern recognition** is a set of methods (algorithms) that can identify patterns based on the data* and use those patterns to make predictions on new data

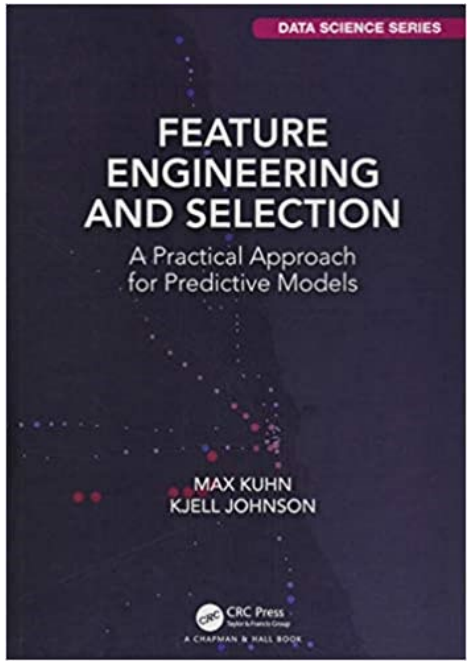
**even when the expert knowledge is incomplete*





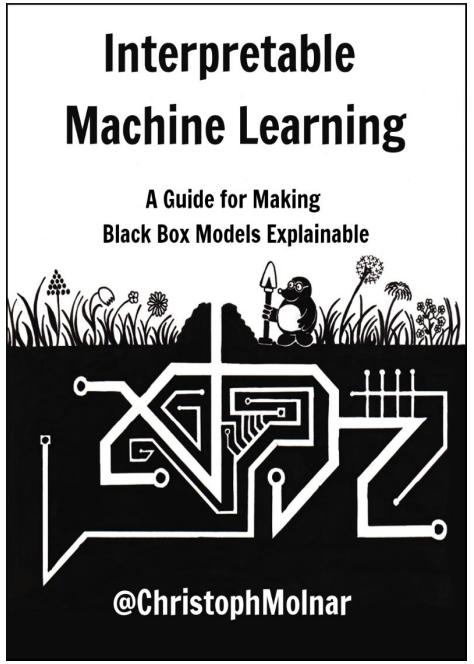
Christopher Bishop; Springer-Verlag New York: 2006

Free book online:
https://bookdown.org/max/FES/



Max Kuhn and Kjell Johnson; Chapman & Hall/CRC Data Science Series: 2019

Free book online:
https://christophm.github.io/interpretable-ml-book/index.html



Christopher Molnar; 2021

Machine learning (ML)

Supervised ML

- Classification
- Regression
- Image recognition

Unsupervised ML

- Clustering
- Dimensionality reduction

cat

cat

cat

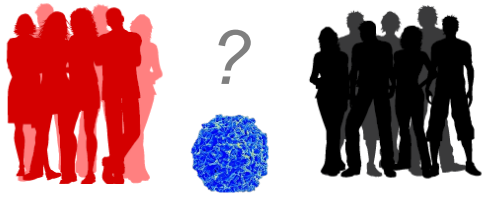
dog

What is this?

Separate into 2 clusters!

Hypothesis

Is there a difference in the frequency of T cells between healthy and infected person?

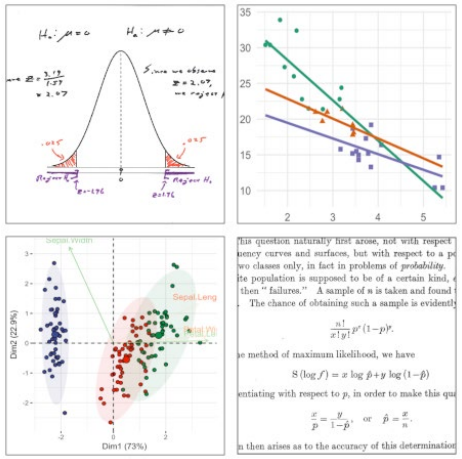


T cells

Hypothesis-driven research

Data analysis

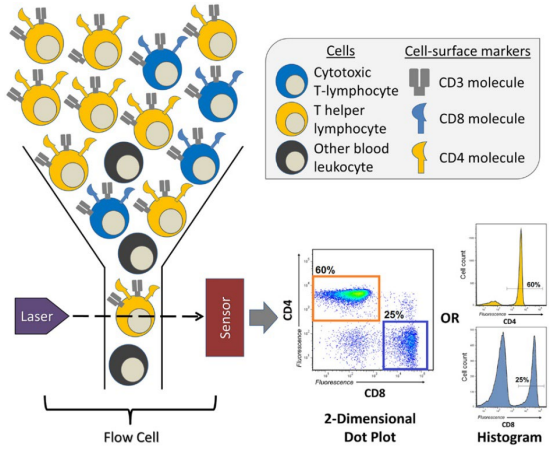
Comparison, statistics



https://statistics.rutgers.edu/

Experiments

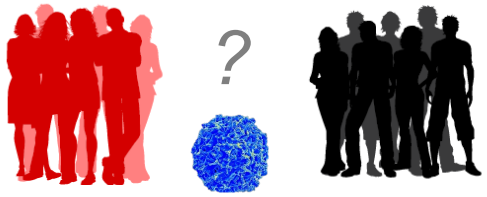
Assays to evaluate frequency, phenotype and functionality



Verschoor C et al, Front Immunol, 2015

Hypothesis

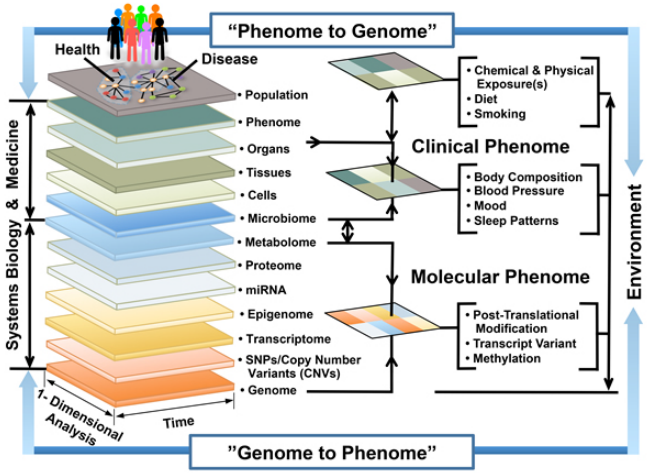
Why is frequency of T cells increased among healthy vs infected person?



T cells

Data analysis

Which cells are present at different frequencies between heathy and infected person?

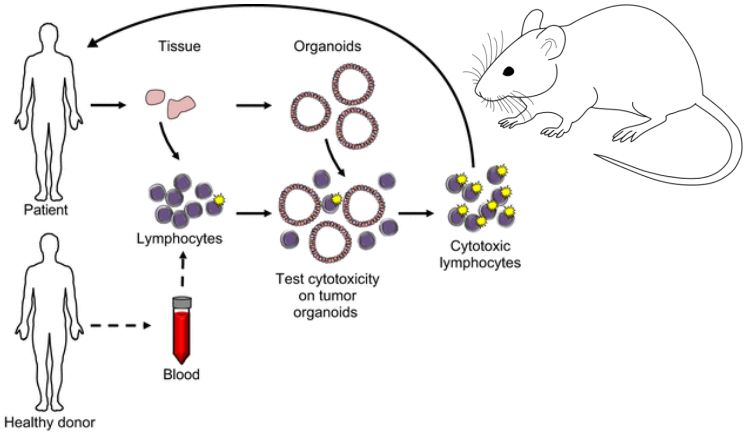


https://medicine.umich.edu/dept/dcmb/research

Data-driven research

Experiments

Assays to confirm phenotype and reveal new mechanisms



Drost and Celevers, Development, 2017

Understanding the Immunity to Influenza virus

Illustration: Eric Nyquist
https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic

Every year

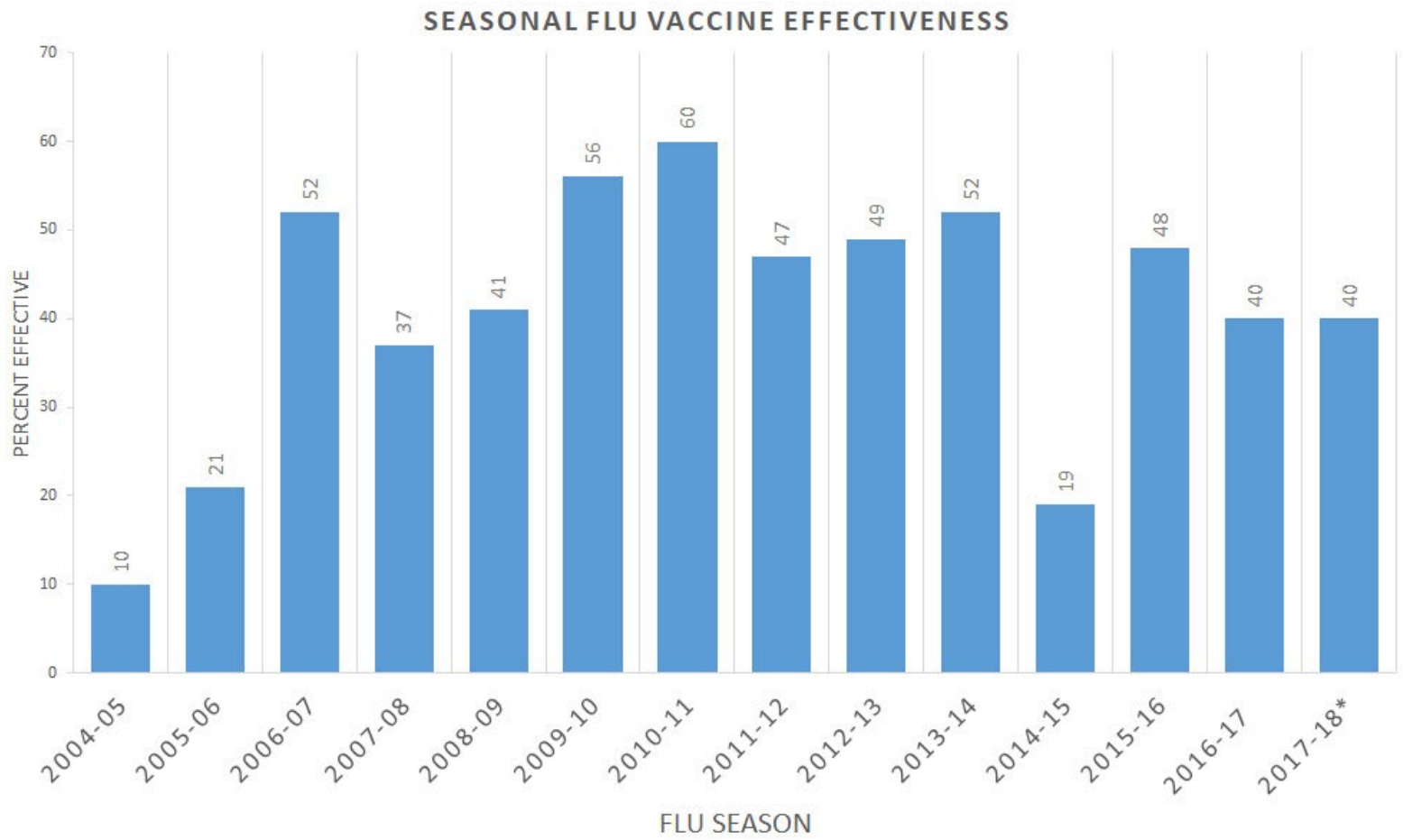
1 in 1000 children and elderly are hospitalized

due to influenza infection

Thompson et al, JAMA, 2004

Seasonal influenza vaccine: the least protective vaccine?

Influenza vaccine effectiveness in the US during the 2004 – 2019 Flu Seasons



2018/2019: **29%**

40%

***Complete match with the circulating strains!**

https://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.ht

FOR HOW LONG IMMUNE SYSTEM REMEMBERS?

For a **LONG TIME**

Chickenpox *(Varicella zoster)*

FOR HOW LONG IMMUNE SYSTEM REMEMBERS?

For a **LONG TIME**

… even **LIFETIME!**

Poliovirus

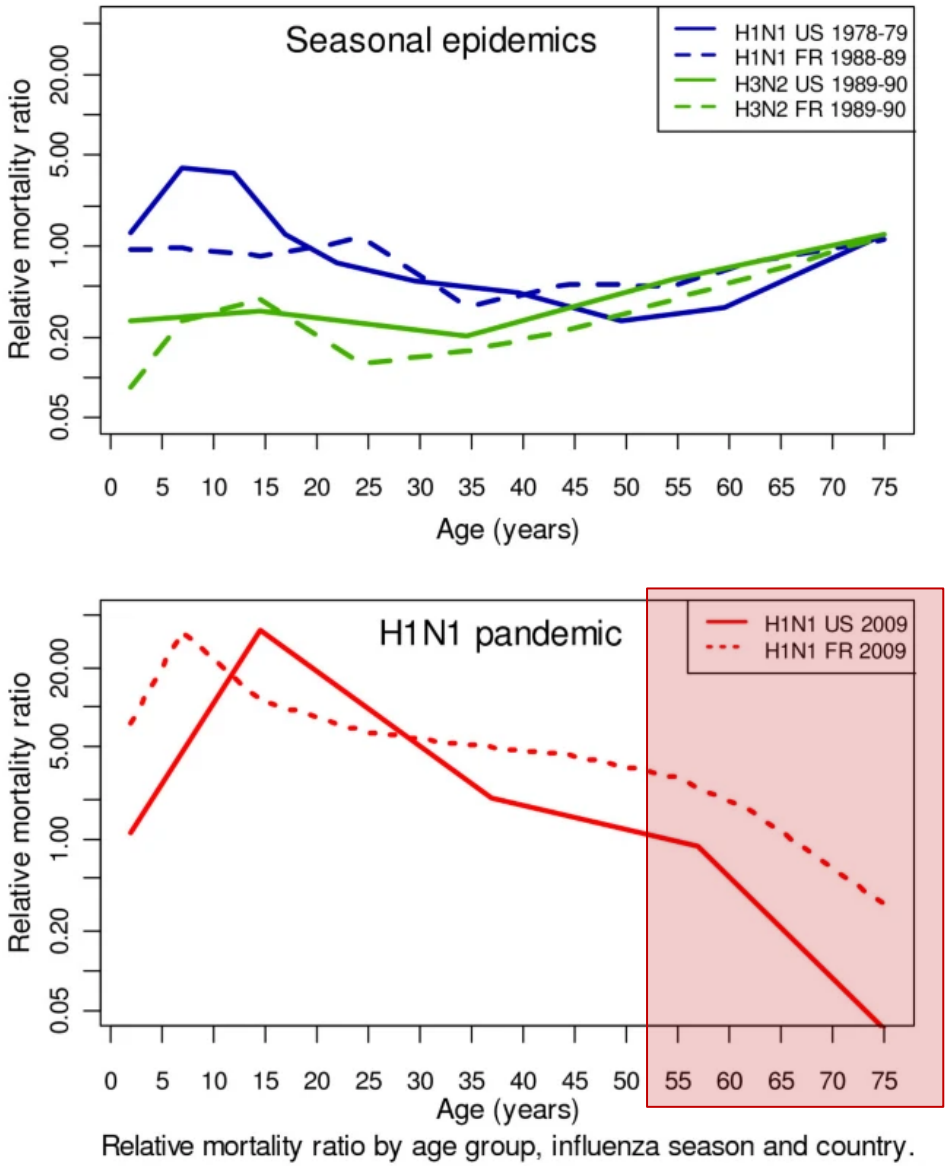
Chickenpox *(Varicella zoster)*

Yellow fever virus

Measles *(Measles morbillivirus)*

Understanding influenza immunity

Illustration: Eric Nyquist
https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic

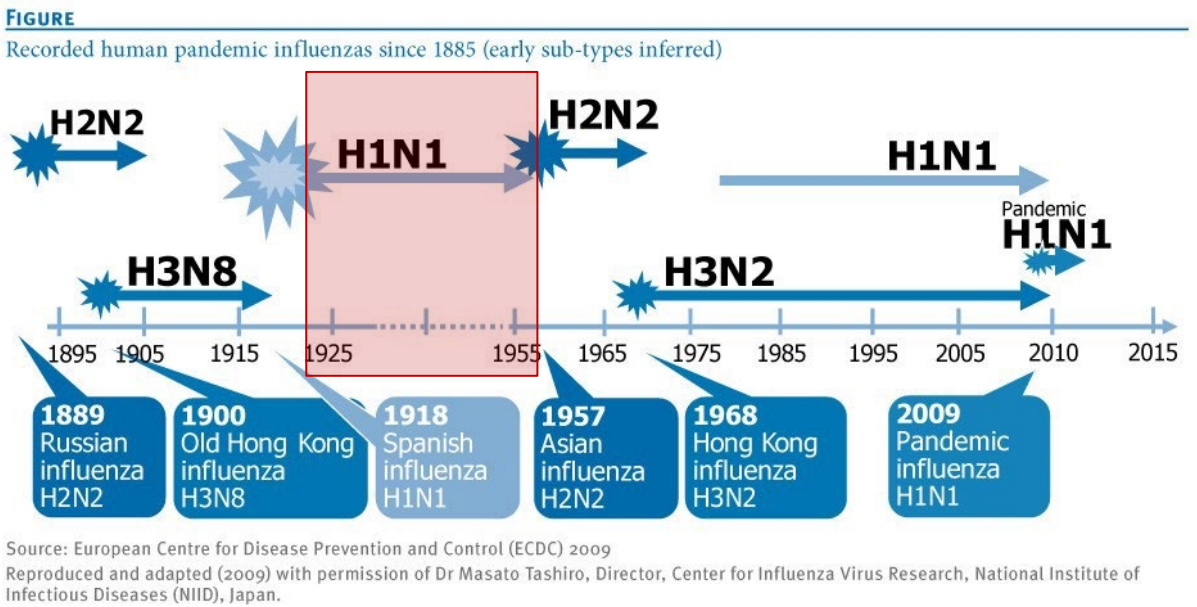


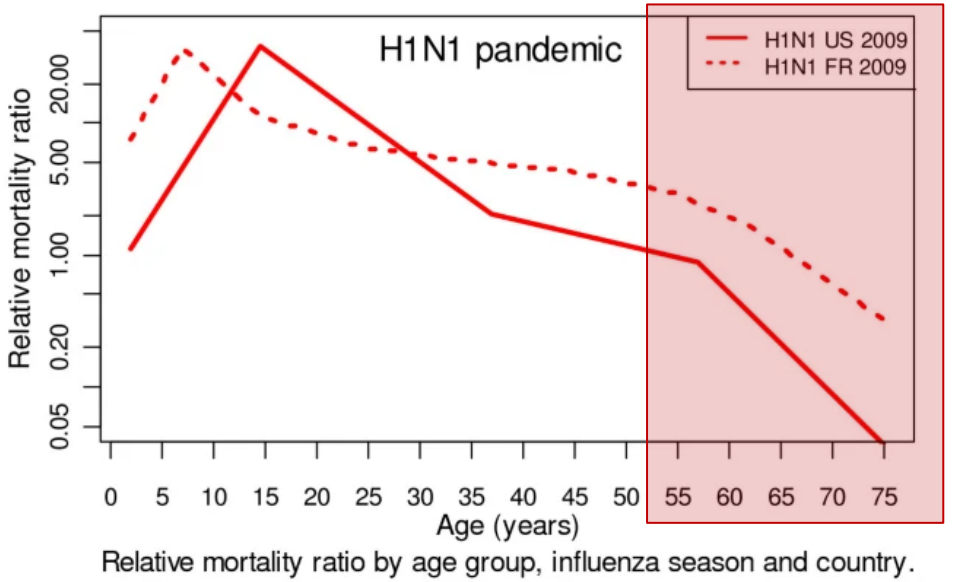
Relative mortality ratio by age group, influenza season and country.

Lemaitre, M., Carrat, F. BMC Infect Dis, 2010.

Understanding influenza immunity

Illustration: Eric Nyquist
https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic



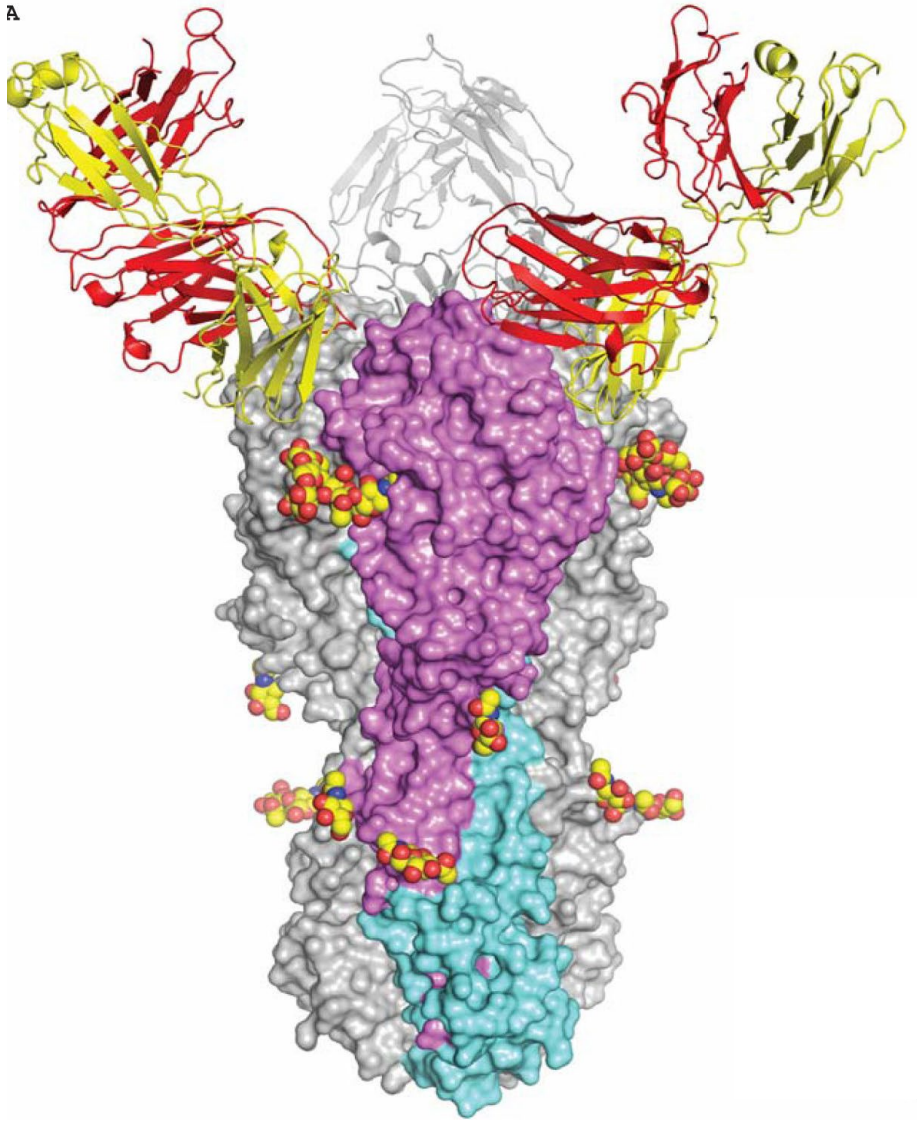


Relative mortality ratio by age group, influenza season and country.

Lemaitre, M., Carrat, F. BMC Infect Dis, 2010.

Understanding influenza immunity

Illustration: Eric Nyquist
https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic

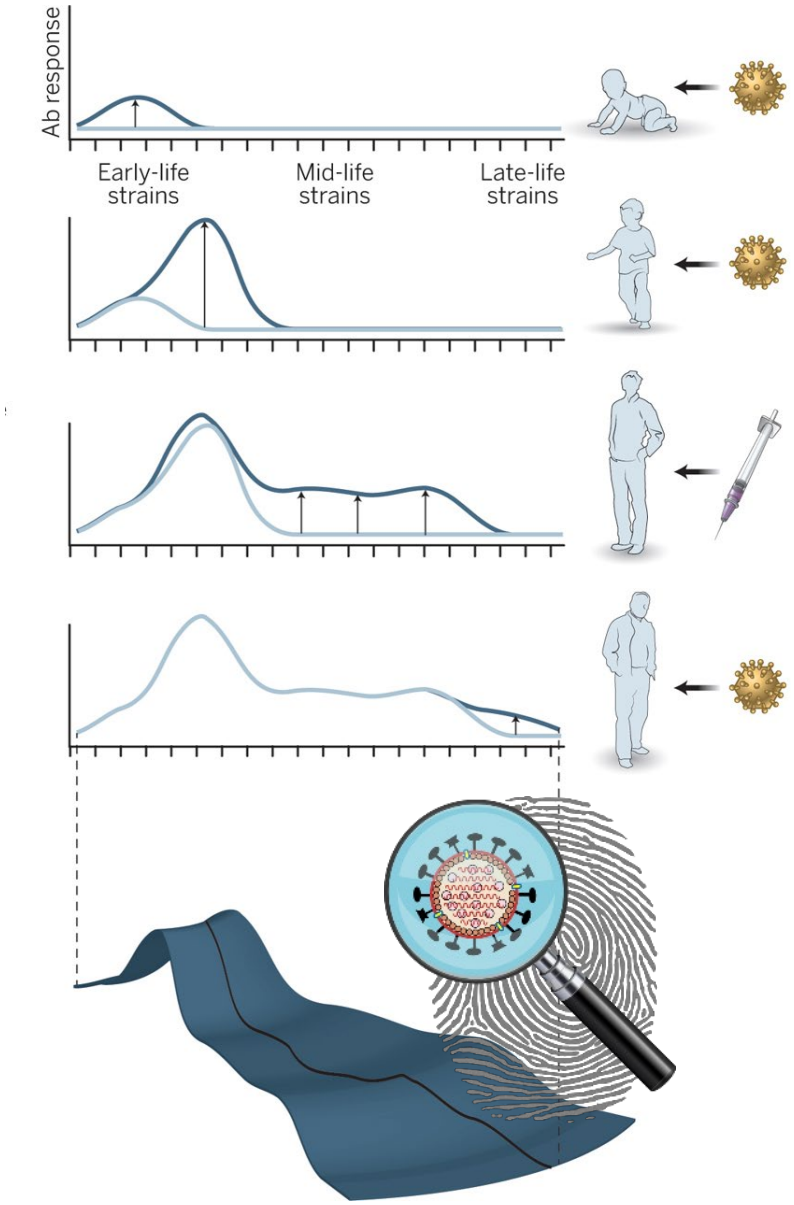


Xu, R. et al, Science, 2010

Understanding influenza immunity

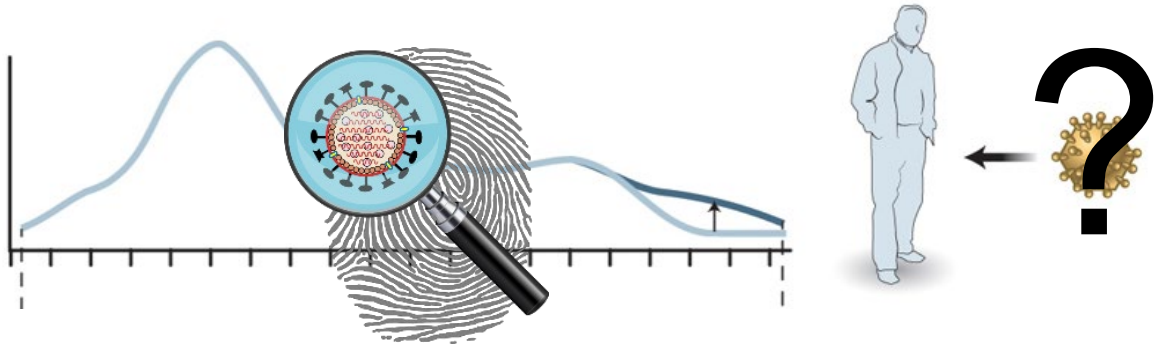
Illustration: Eric Nyquist
https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic

FLUPRINT: lifetime of exposure to influenza



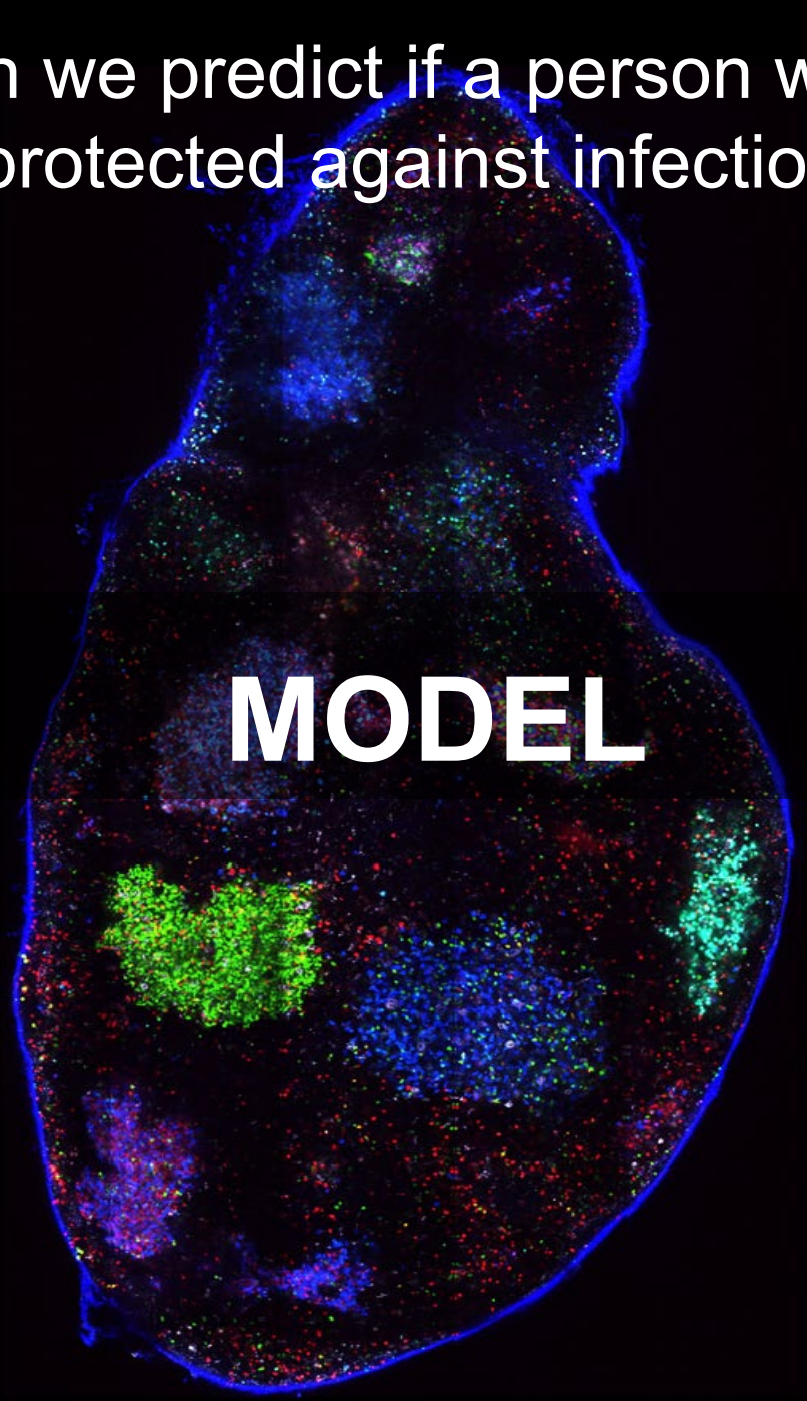
Justin Lessler, Science, 2013

Can we predict if a person will be protected against infection?

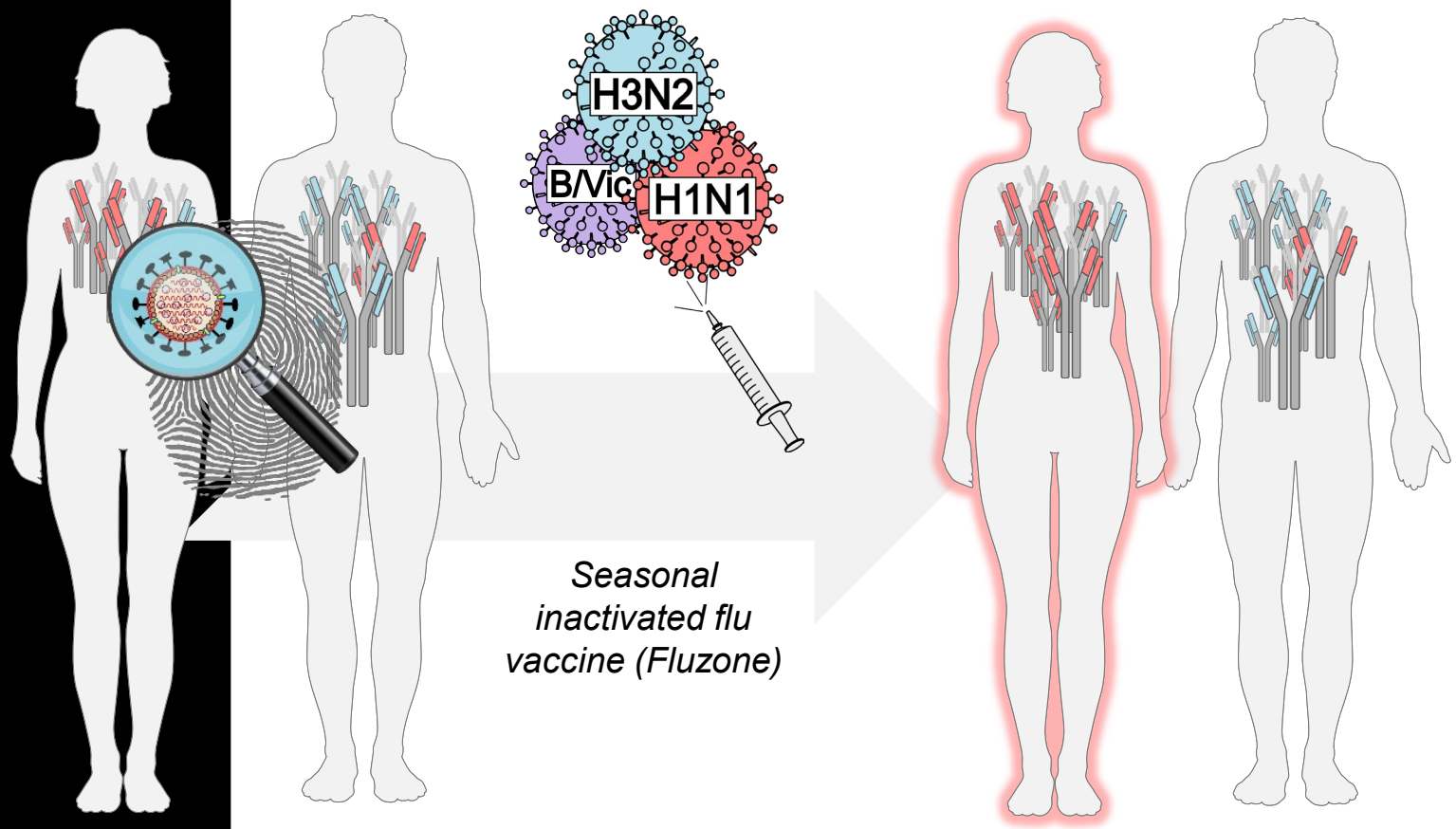


Justin Lessler, Science, 2013

Can we predict if a person will be protected against infection?



INDUCTION OF HAI ANTIBODIES
(HAI titer >=40) associated with
reduction in influenza disease



'HIGH RESPONDER'

Hemagglutination inhibition (HAI) antibodies correlate with protection against influenza-like disease

Hobson, D., Curry, R. L., Beare, A. S. & Ward- Gardner, A. J. Hyg. 1972.

Can we predict if a person will be protected against infection?

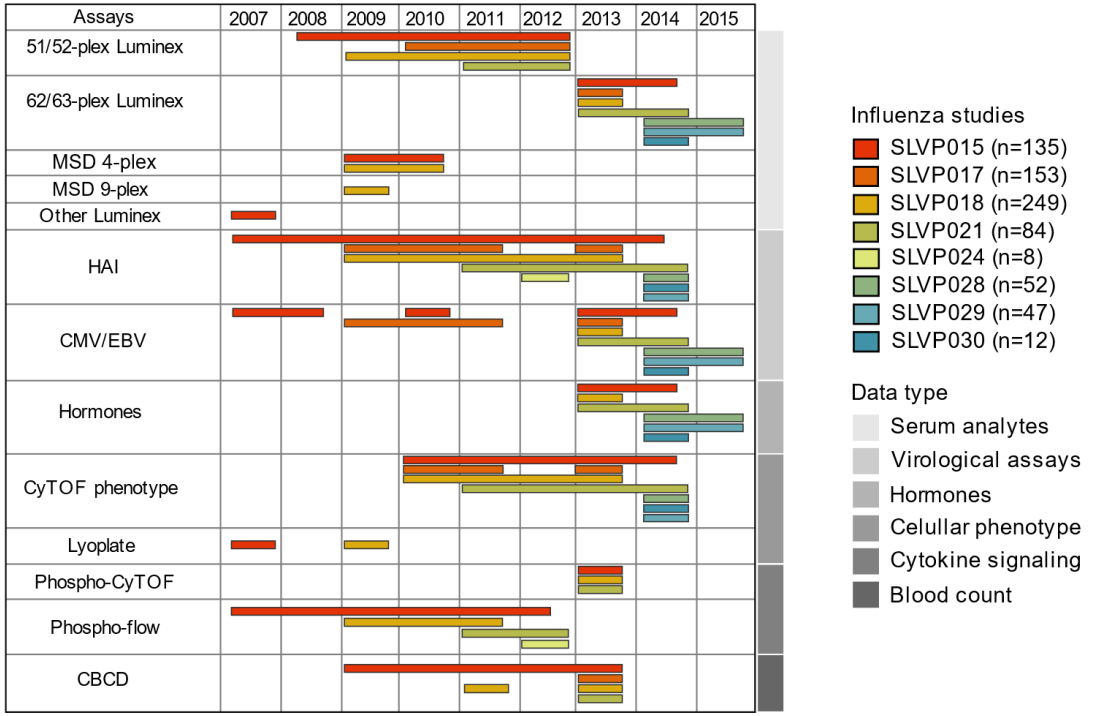
MODEL:

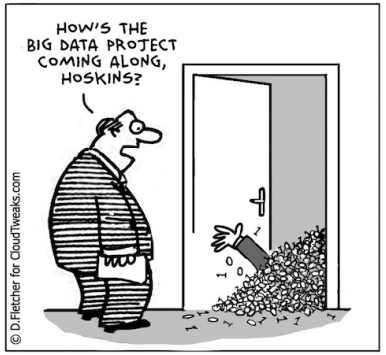
Inactivated flu vaccine

APPROACH

Systems immunology

FluPRINT open-access database





Tomic et al, Sci Data, 2019

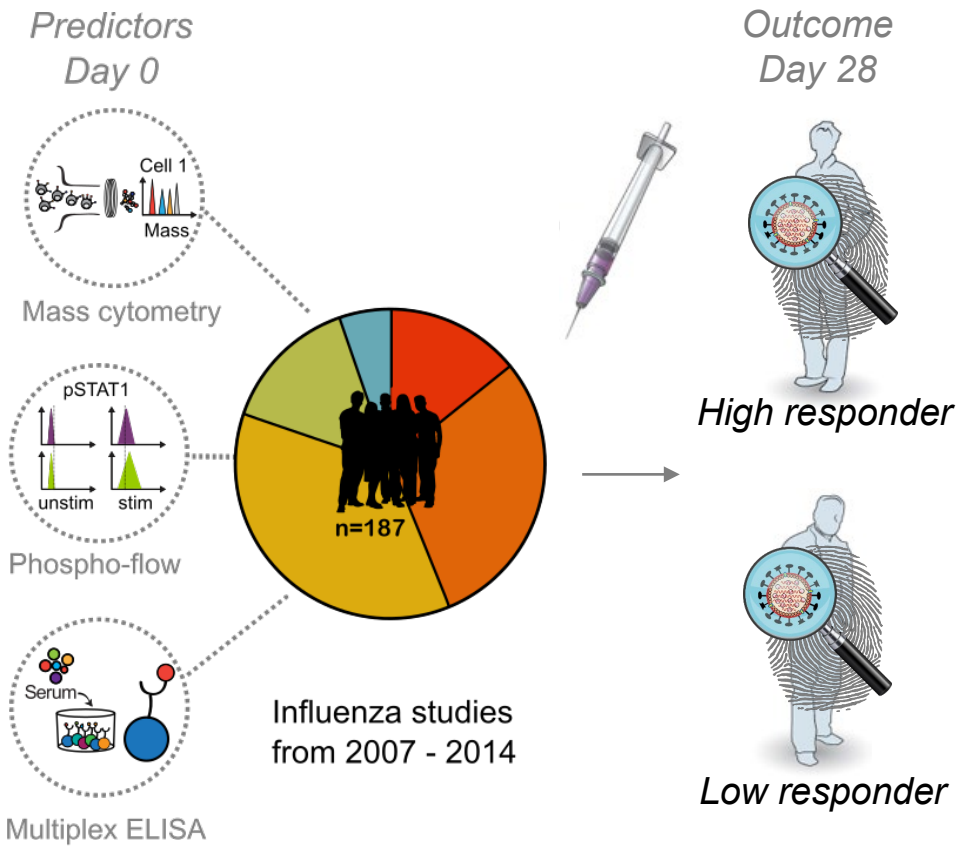
Can we predict if a person will be protected against infection?

MODEL:

Inactivated flu vaccine

APPROACH:

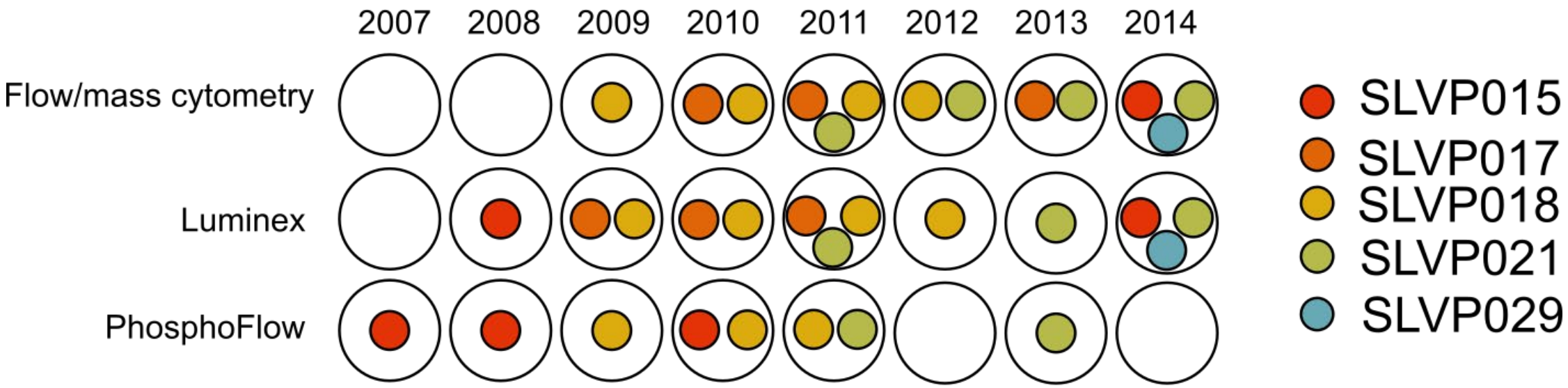
Systems immunology



Tomic et al, JI, 2019; Tomic et al, Sci Data, 2019

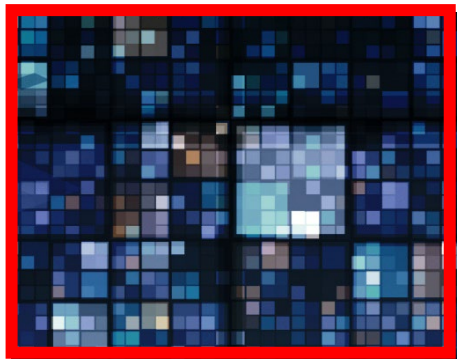
THE PROBLEM 1

Not all data are available

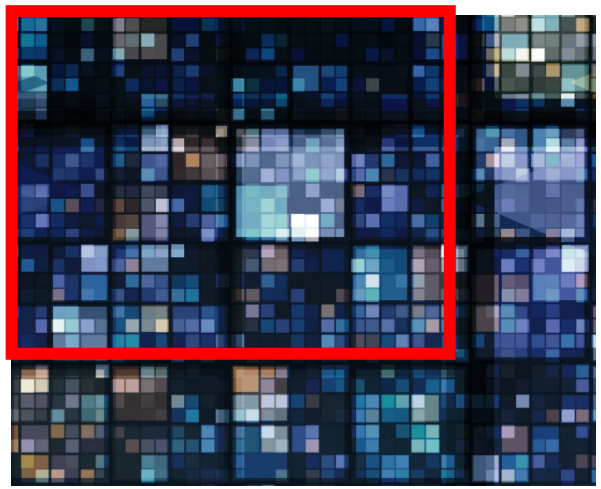


The biomedicine “BIG” problem: High percentage of missing data

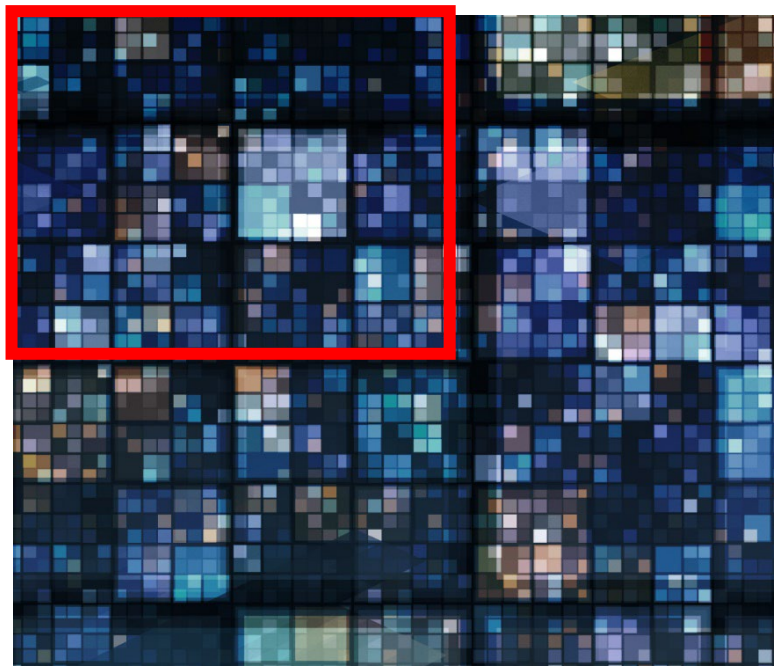
2007



2010



2014



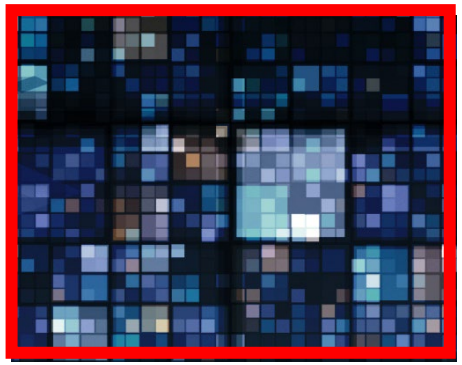
How to select optimal number of donors and optimal number of features?

SUBSAMPLING

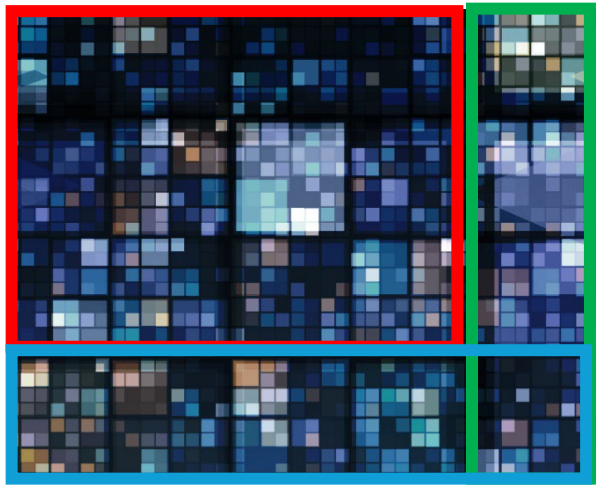
→ loosing a lot of information

THE SOLUTION

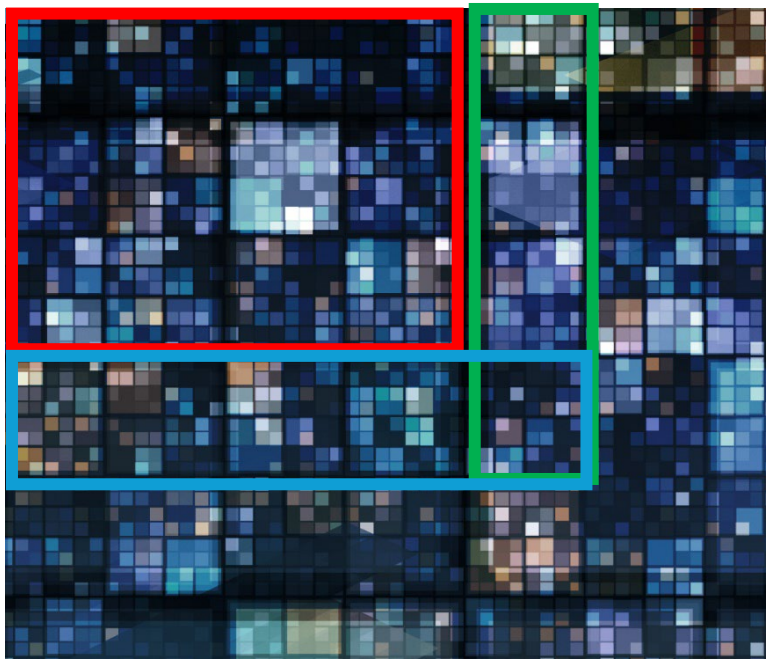
2007



2010

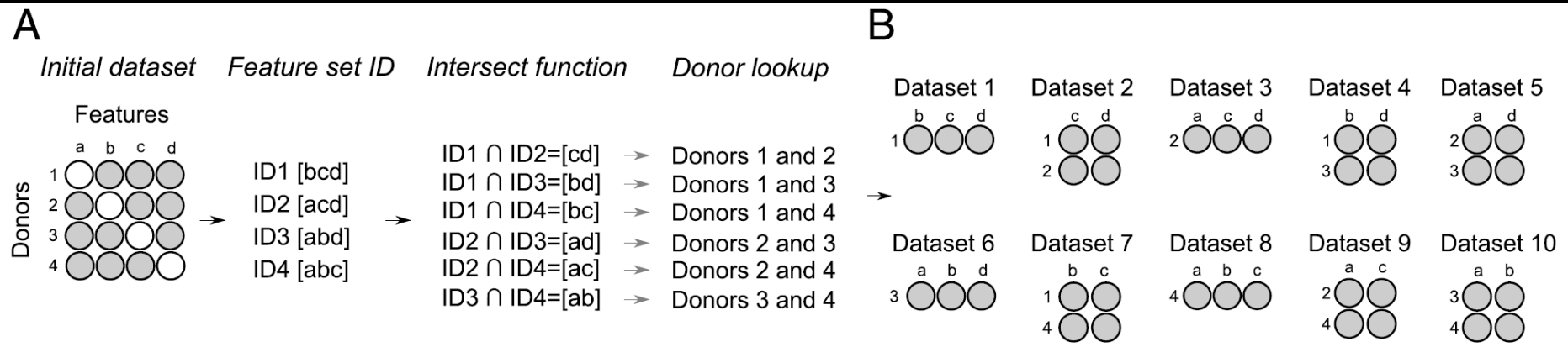


2014



A fully automated script for feature subset selection, dimensionality reduction and data sampling

Novel approach to cope with high sparsity in biomedical data using a multi-set intersection function



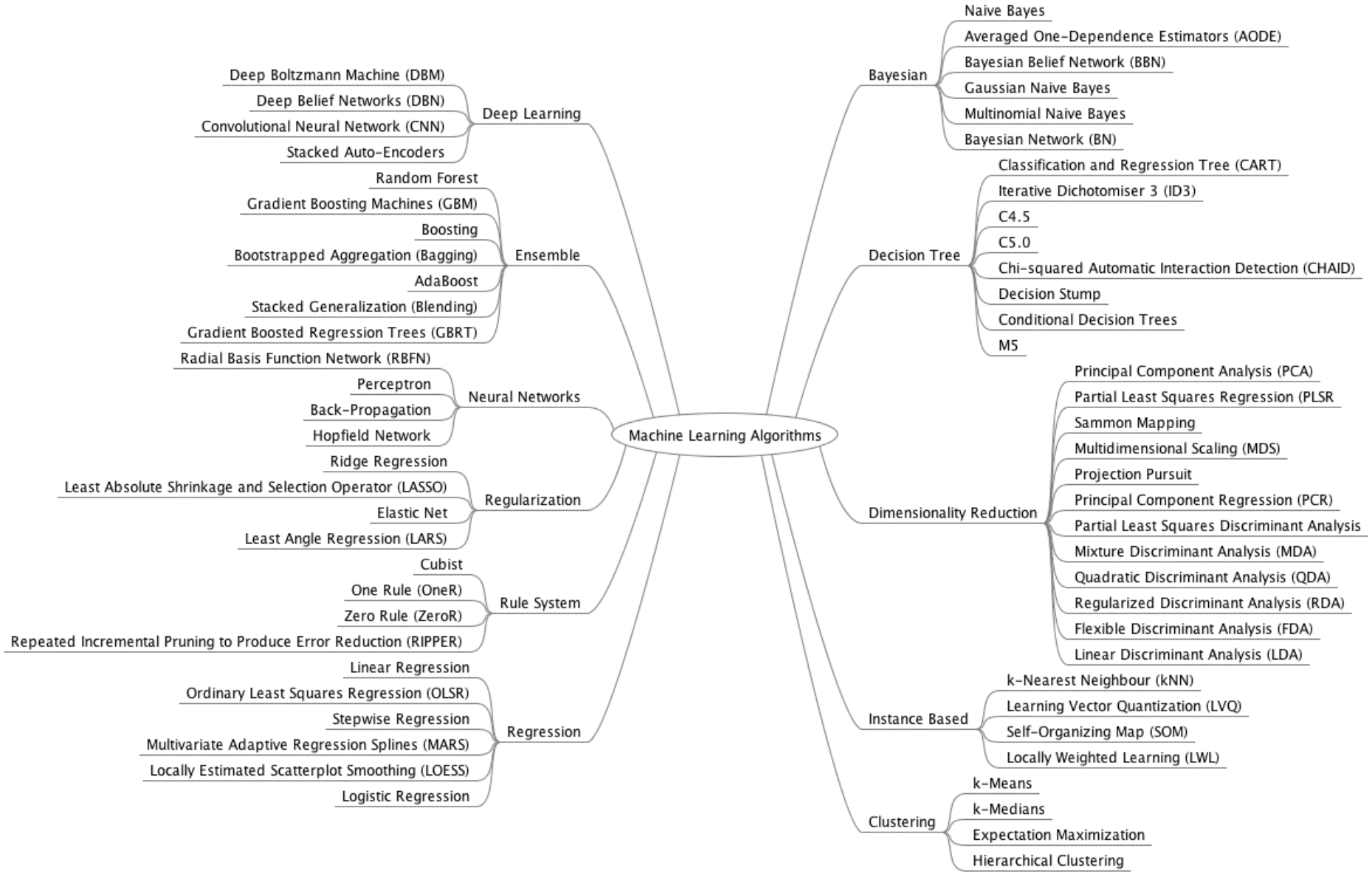
Tomic et al. SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses, JI, 2019

mulset

https://github.com/LogIN-/mulset

THE PROBLEM 2

Machine learning models: Which one to use? Use all of them!



Jason Brownlee, machinelearningmastery.com

Can we predict if a person will be protected against infection?

MODEL:

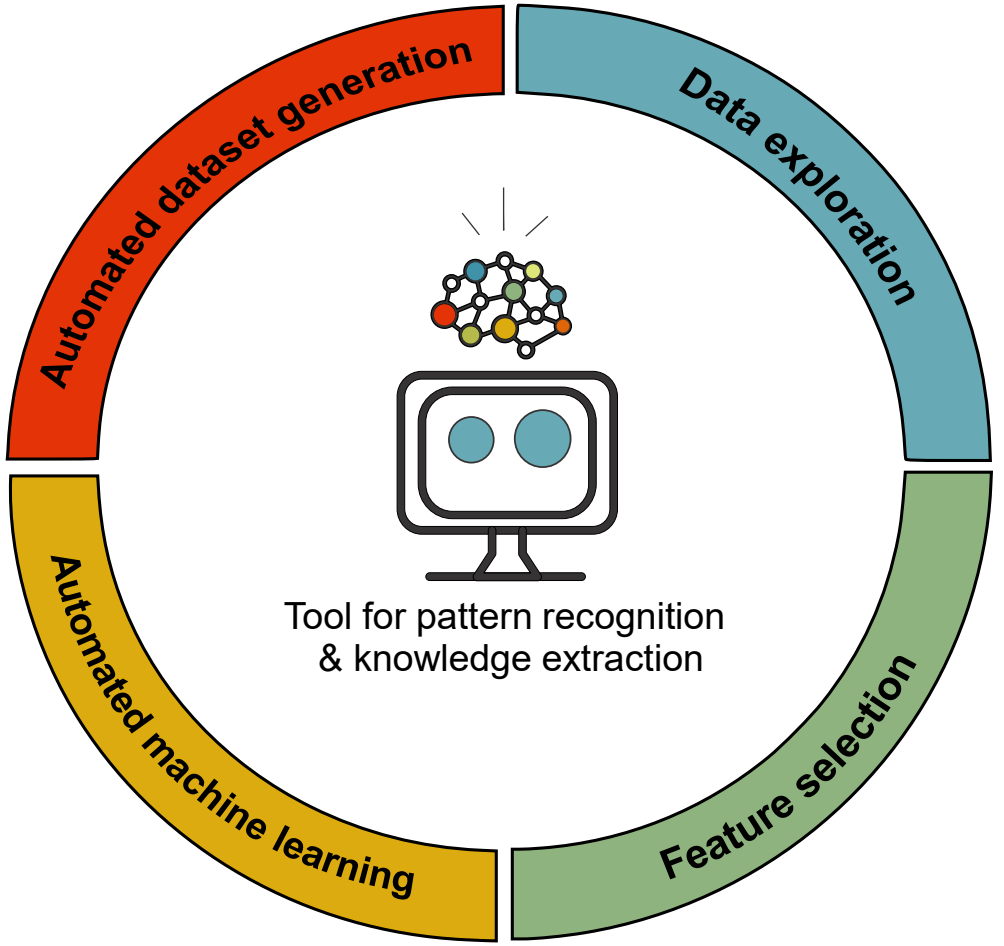
Inactivated flu vaccine

APPROACH:

Systems immunology

TOOL

AI to the rescue!



Tomic et al, JI, 2019; Tomic et al, Patterns, 2021

Can we predict if a person will be protected against infection?

MODEL:

Inactivated flu vaccine

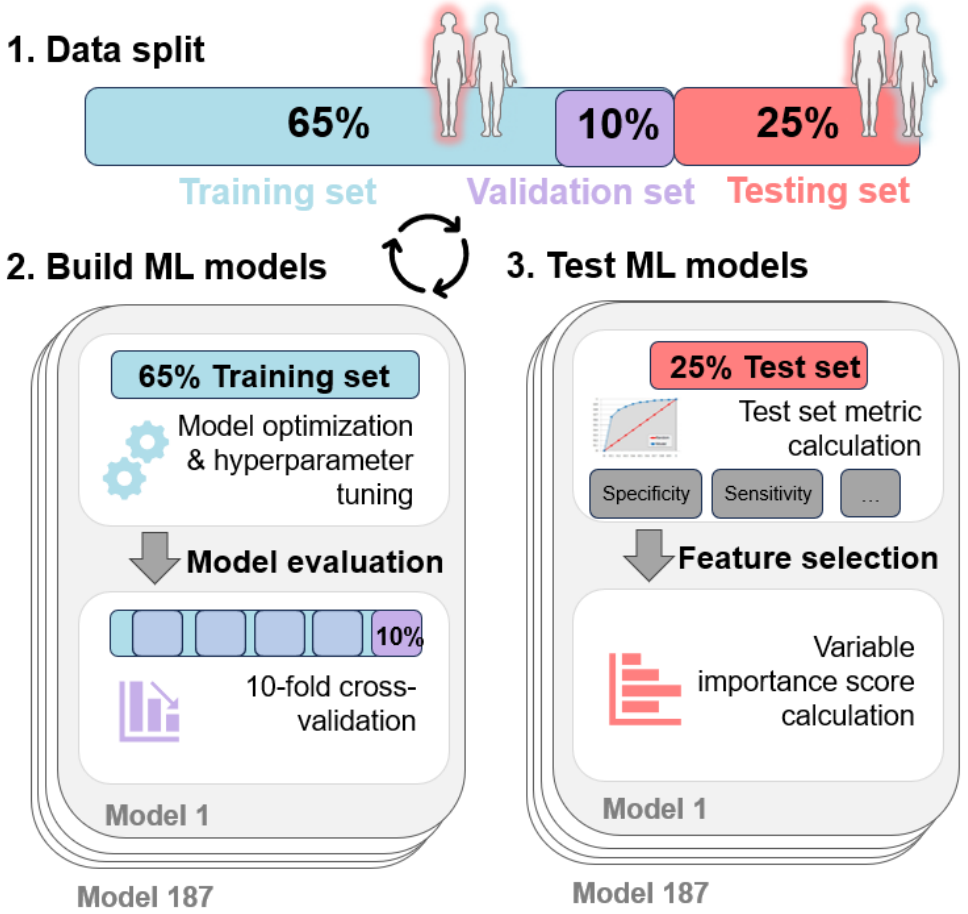
APPROACH:

Systems immunology

TOOL:

SIMON

SIMON: Sequential Iterative Modelling Over Night



Tomic et al, JI, 2019; Tomic et al, Patterns, 2021

2,400 machine learning analysis run on **34** datasets

Models were built for **19** datasets, with an average of **54** models built per dataset

After model selection, **2** datasets with average of **3** models built per dataset

SIMON facilitates exploratory analysis and discovery of high-performing models

Can we predict if a person will be protected against infection?

Yes, SIMON can!

- **Increases model performance**

Automated ML process
Testing 128 algorithms

Filtering step
Selection of high-performing models

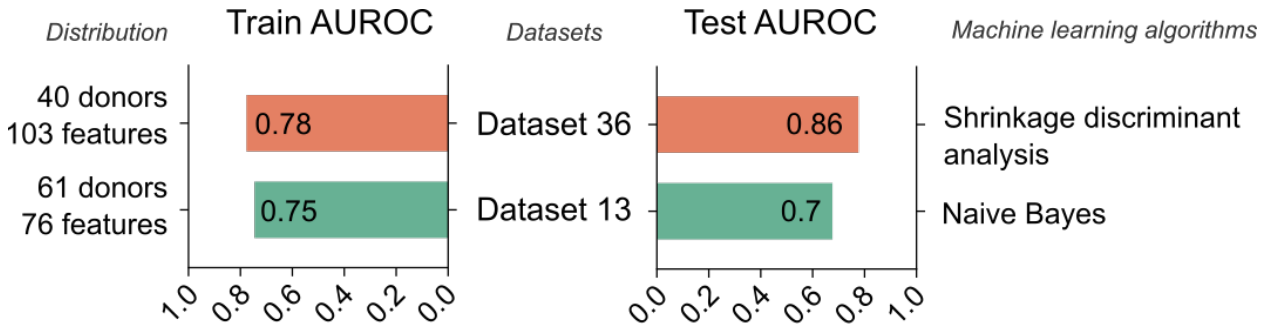
Dataset ID	Number of models	Train AUROC (min)	Train AUROC (max)	Improvement (%)
205	35	0.08	0.92	91.3
4	57	0.29	0.87	66.7
5	58	0.39	0.82	52.4
10	55	0.43	0.8	46.3
36	55	0.43	0.79	45.6
34	43	0.28	0.78	64.1
35	61	0.29	0.75	61.3
13	52	0.32	0.75	57.3

Tomic et al, JI, 2019

Can we predict if a person will be protected against infection?

Yes, SIMON can!

- **Increases model performance**
- **Selection of optimal algorithms**



SDA
Works the best with the small sample size, but high-dimensional setting

Bankruptcy prediction

Image recognition

Marketing

Mkhadri A, Pattern Recognition Letter 1995

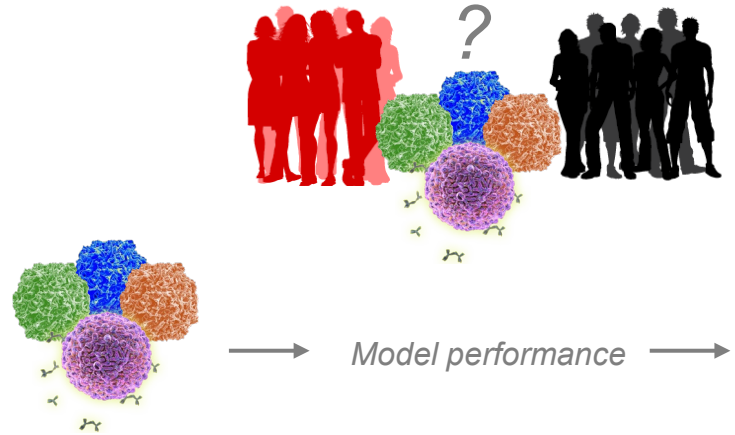
Tomic et al, JI, 2019

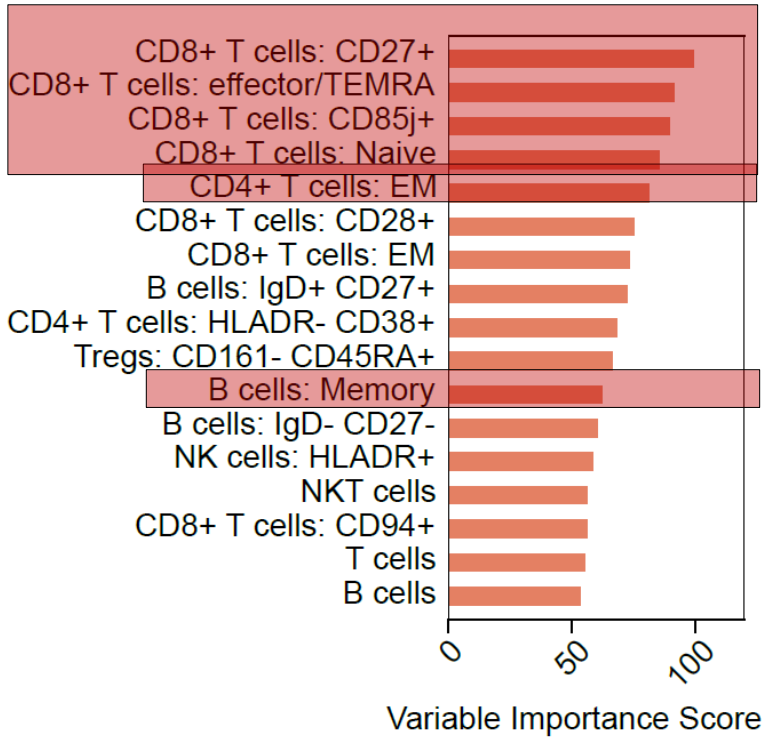
Can we predict if a person will be protected against infection?

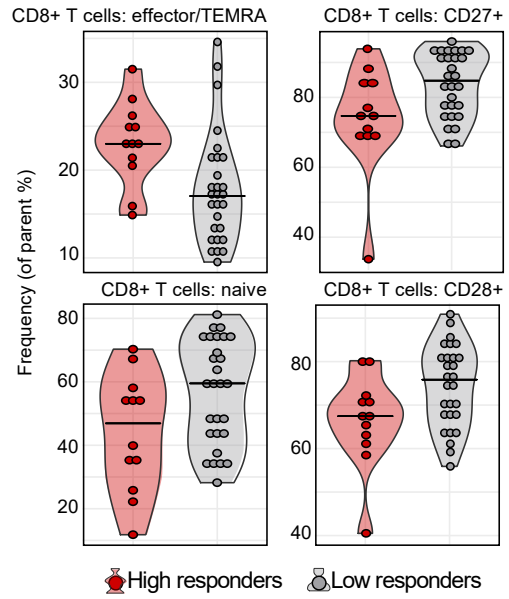
Yes, SIMON can!

- Increases model performance
- Selection of optimal algorithms
- Feature selection

Is there a difference in the frequency of immune cells between high and low responders?

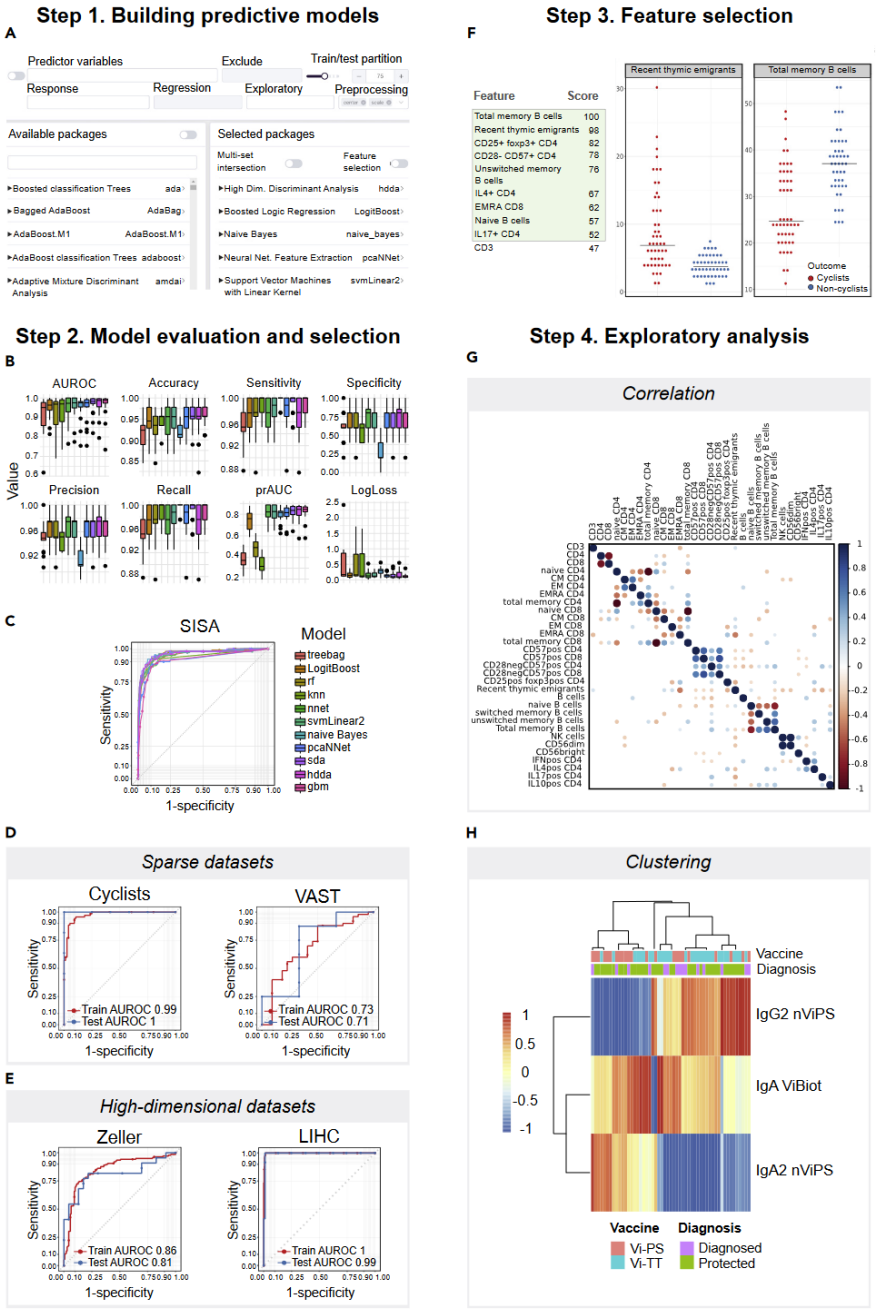






Tomic et al, JI, 2019

SIMON: Open-source knowledge discovery platform

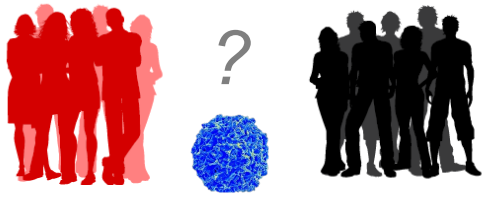


Tomic et al, Patterns, 2021

SIMON says: Time for Data-Driven Research!

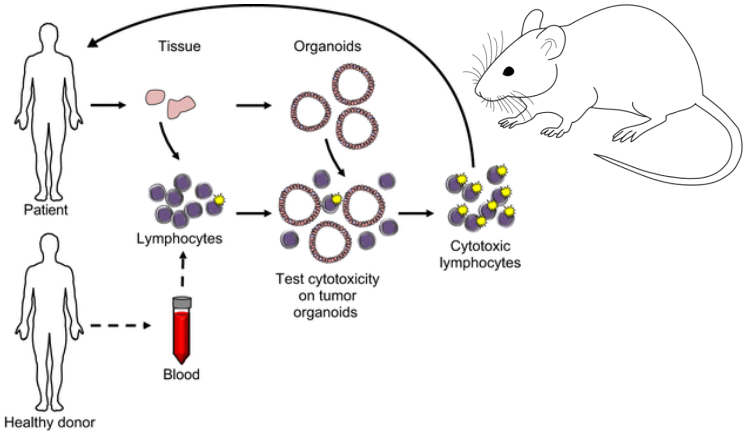
Hypothesis

Why is frequency of T cells increased among healthy vs infected individuals?



Experiments

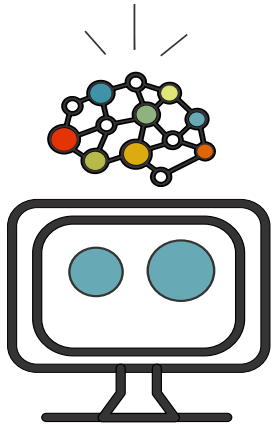
Assays to confirm phenotype and reveal new mechanisms of T cells In both groups



Drost and Celevers, Development, 2017

Data-driven research

Data analysis



*SIMON**

- Feed dataset
- 180+ machine learning algorithms
- Build 1000s of models in one click
- Explore top models
- Identify top important variables