

FLU FIGHTERS:

Prediction challenge
- Final team task -



adrianatomic



@TomicAdriana



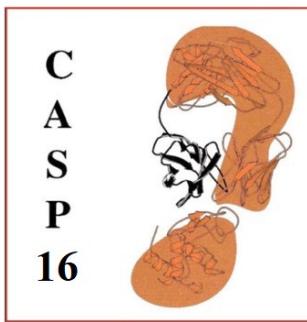
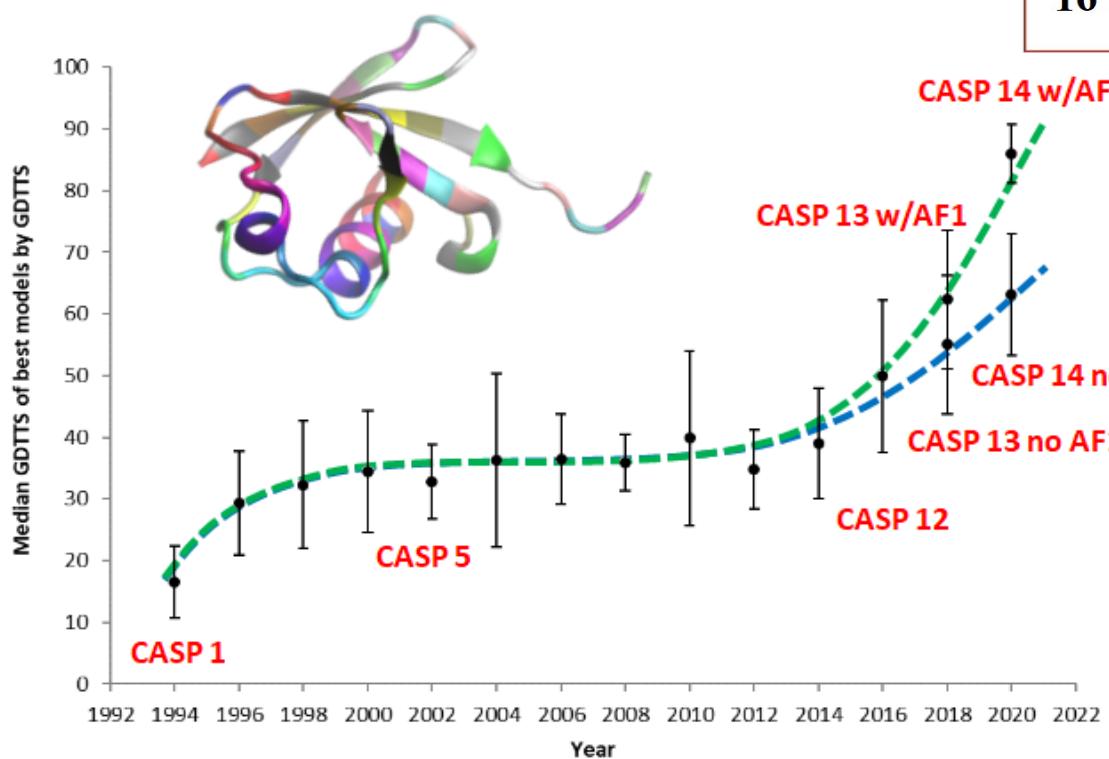
atomic@bu.edu



atomiclaboratory

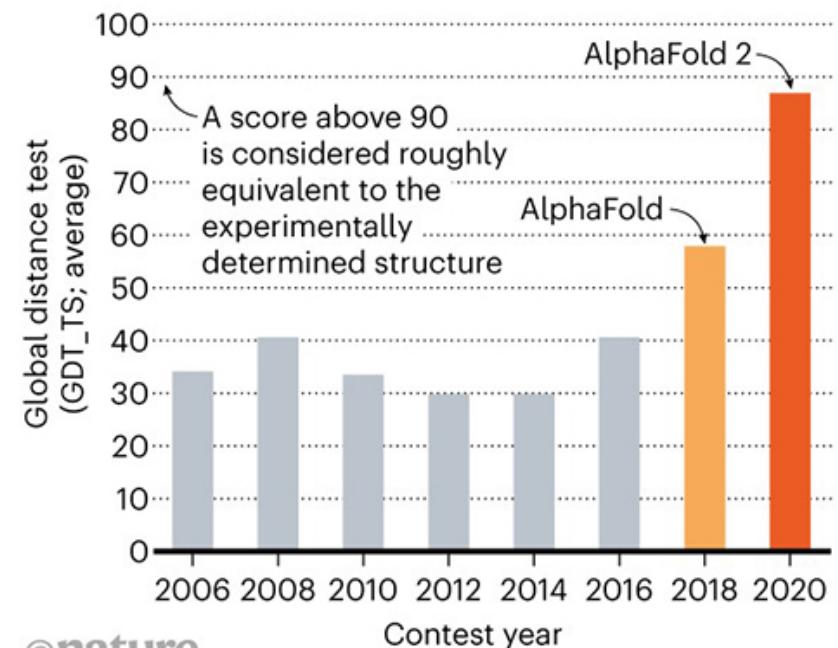
atomic-lab.org

Community-driven Critical Assessment of Structure Prediction (CASP)



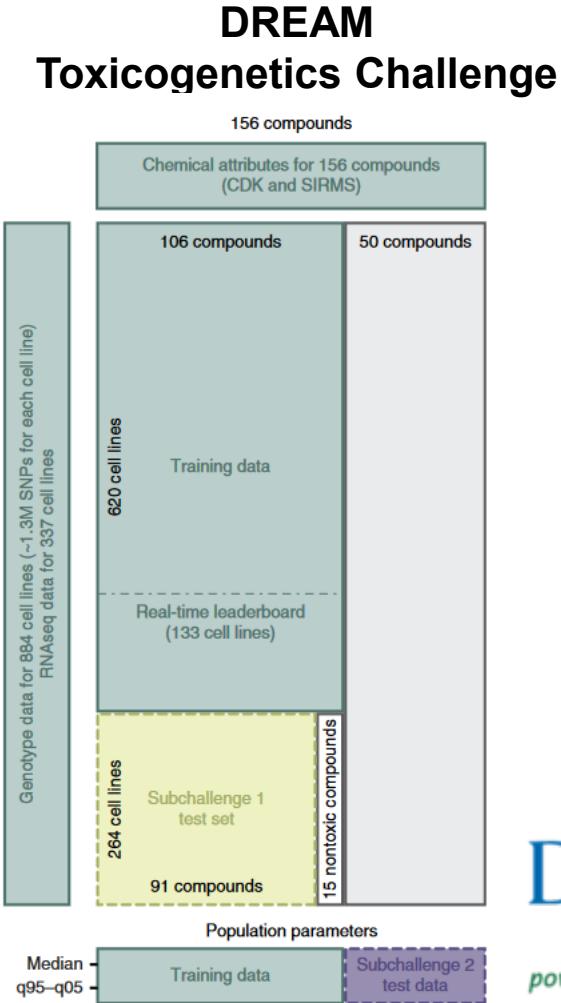
STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



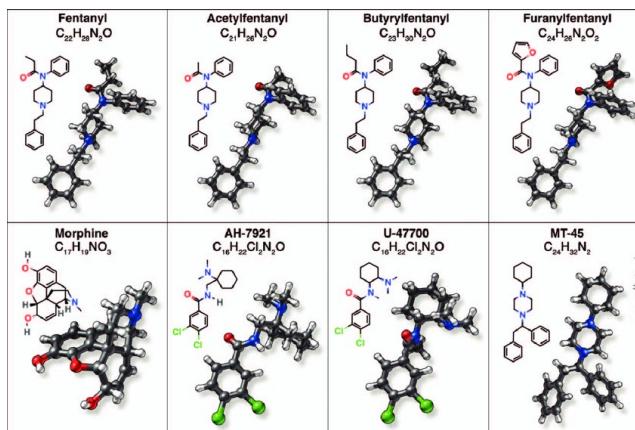
©nature

Community-based prediction challenge to predict toxicities of environmental compounds with potential adverse health effects for human populations



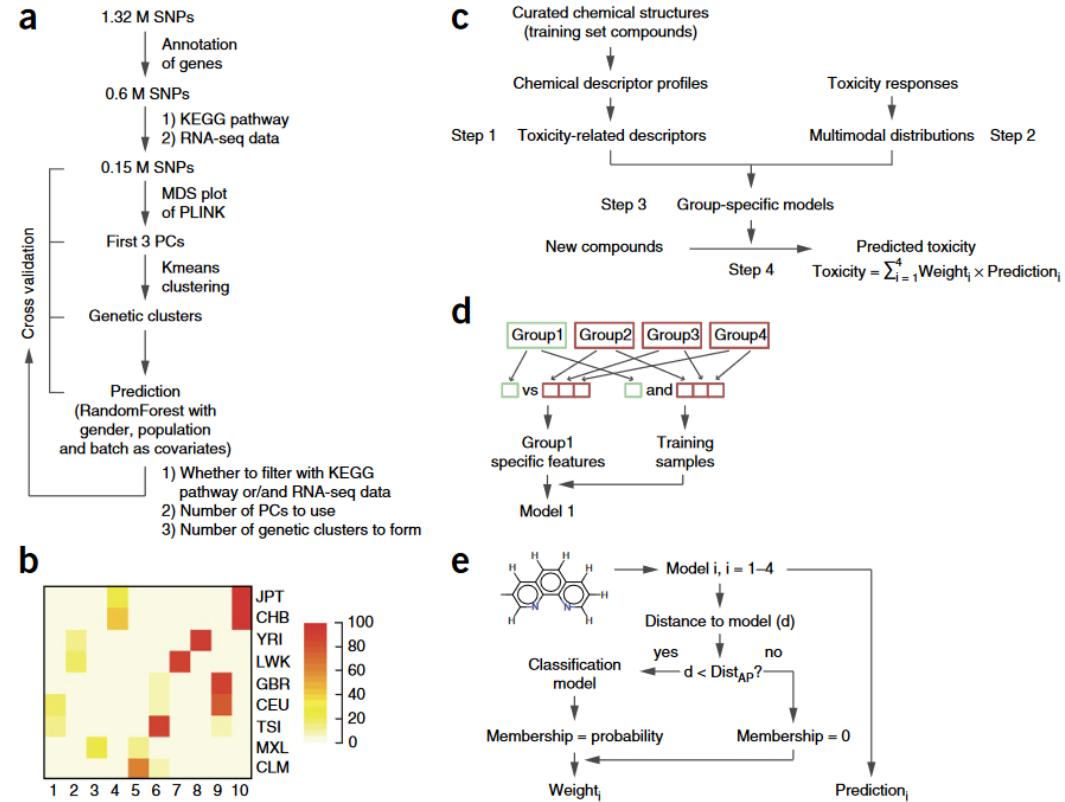
The challenge participants developed algorithms to predict interindividual variability of toxic response from genomic profiles and population-level cytotoxicity data from structural attributes of the compounds.

179 submitted predictions



DREAM CHALLENGES
powered by Sage Bionetworks

Prediction of human population responses to toxic compounds by a collaborative competition



Prediction challenges



powered by Sage Bionetworks

DREAM Challenges use crowd-sourcing to solve complex biomedical research questions

<https://dreamchallenges.org/>



Home About DREAM 101 Challenges Publications Latest News



Explore Past DREAM Challenges

Categories

ALS(1)	Alzheimer's Disease(1)
Breast Cancer(2)	Cancer(10)
Cell Lineage(2)	
Clinical Prediction(15)	
Clinical Prediction-Progression(1)	
Clinical Prediction-Survival(1)	
Diagnosis(1)	
Diagnosis and Outcome(3)	
Disease Modules(1)	DNA(2)
Drug Efficacy(1)	EHR(2)
Flow Cytometry(1)	
Gene Expression(1)	General(3)
Immuno-Oncology(1)	Kinetics(1)
Lung Cancer(1)	
Medical Imaging(3)	
Metadata Automation(1)	
Method Optimization(19)	
Motif-Binding(1)	
Network Inference(10)	
Olfaction(2)	Open Science(1)
Outcome(3)	
Parkinson's Disease(2)	
Pathology(1)	
Peptide-antibody binding(1)	
Peptide binding(1)	
Pharmacology(5)	
Phosphoproteins(1)	
Predictive Modeling(7)	
Protein Concentration(1)	



Placental Clock Challenge

June 3, 2024 | Categories: Challenges

[Read More >](#)



Personal Environment and Genes Study (PEGS) Challenge

May 7, 2024 | Categories: Challenges, Geospatial, Predictive Modeling

[Read More >](#)



Olfactory Mixtures Prediction Challenge

April 19, 2024 | Categories: Challenges | Tags: Olfaction, Predictive Modeling

[Read More >](#)



Xcelerate RARE: A Rare Disease Open Science Data Challenge

May 8, 2023 | Categories: Challenges | Tags: Diagnosis and Outcome, Open Science, Rare Disease

[Read More >](#)



COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge

September 28, 2022 | Categories: Challenges, Infectious Disease

[Read More >](#)



Heart Failure Prediction: Microbiome

September 21, 2022 | Categories: Challenges

[Read More >](#)



BraTS Continuous



Preterm Birth



scRNA-seq and

kaggle

<https://www.kaggle.com/>

Prediction challenges

⌚ Active Competitions

Hotness ▾



AI Mathematical Olympiad - Progress Prize 2

Solve national-level math challenges using AI...
Featured · Code Competition
506 Teams

\$2,117,152

4 months to go



Google - Unlock Global Communication with...

Create Gemma model variants for a specific language...
Analytics

\$150,000

2 months to go



Jane Street Real-Time Market Data Forecasting

Predict financial market responders using AI...
Featured · Code Competition
1636 Teams

\$120,000

2 months to go



Google - Gemini Long Context

Demonstrate interesting use cases for Gemini...
Analytics

\$100,000

10 days to go

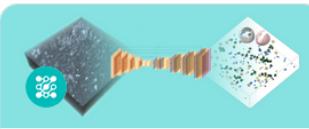


NFL Big Data Bowl 2025

Help use pre-snap behavior to predict a...
Analytics

\$100,000

2 months to go

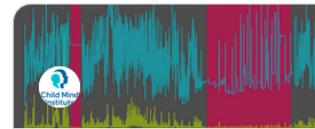


CZII - CryoET Object Identification

Find small biological structures in large ...
Featured · Code Competition
174 Teams

\$75,000

3 months to go

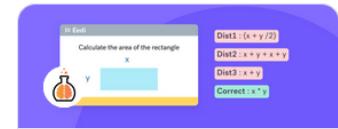


Child Mind Institute — Problematic Internet Use

Relating Physical Activity to Problematic...
Featured · Code Competition
2458 Teams

\$60,000

A month to go



Eedi - Mining Misconceptions in Science

Predict affinity between misconceptions...
Featured · Code Competition
1289 Teams

\$55,000

21 days to go



UM - Game-Playing Strength of MCTS Variants

Predict which variants of Monte-Carlo Tree Search...
Research · Code Competition
1482 Teams

\$50,000

11 days to go



FIDE & Google Efficient Chess AI Challenge

Create agents to play chess with resources...
Featured · Simulation Competition
222 Teams

\$50,000

3 months to go



WSDM Cup - Multilingual Chatbot Arena

Predict human preference across multiple languages...
Featured · Code Competition
76 Teams

\$50,000

2 months to go



Exploring Mental Health Data

Playground Series - Season 4, Episode 11
Playground
1826 Teams

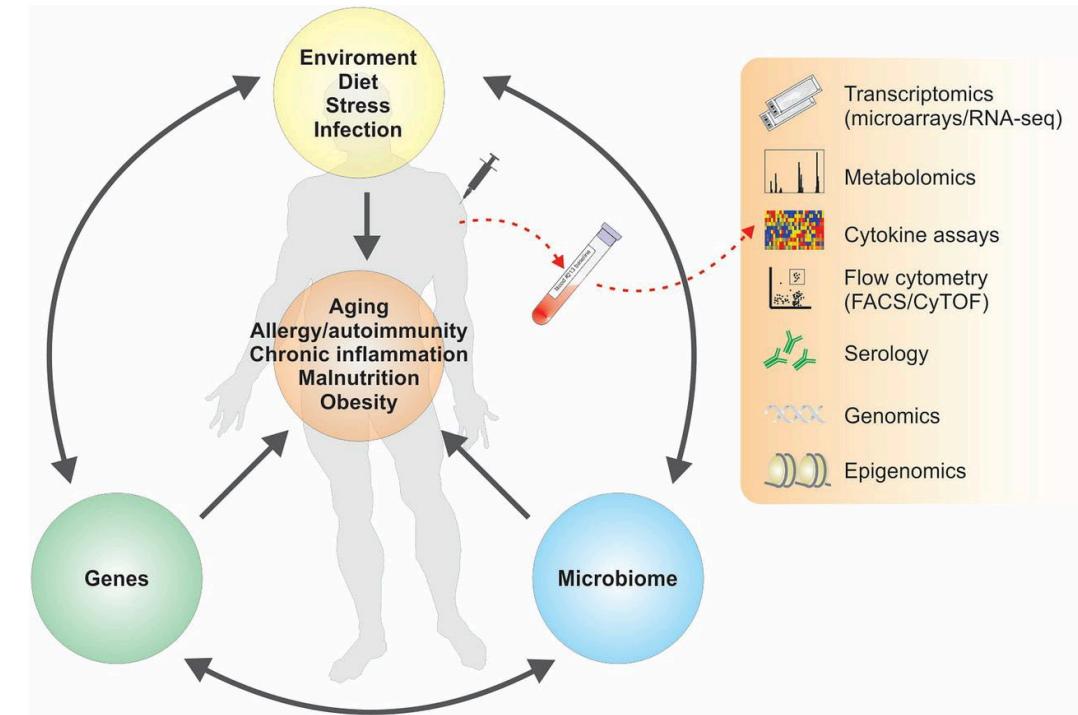
\$50,000

9 days to go

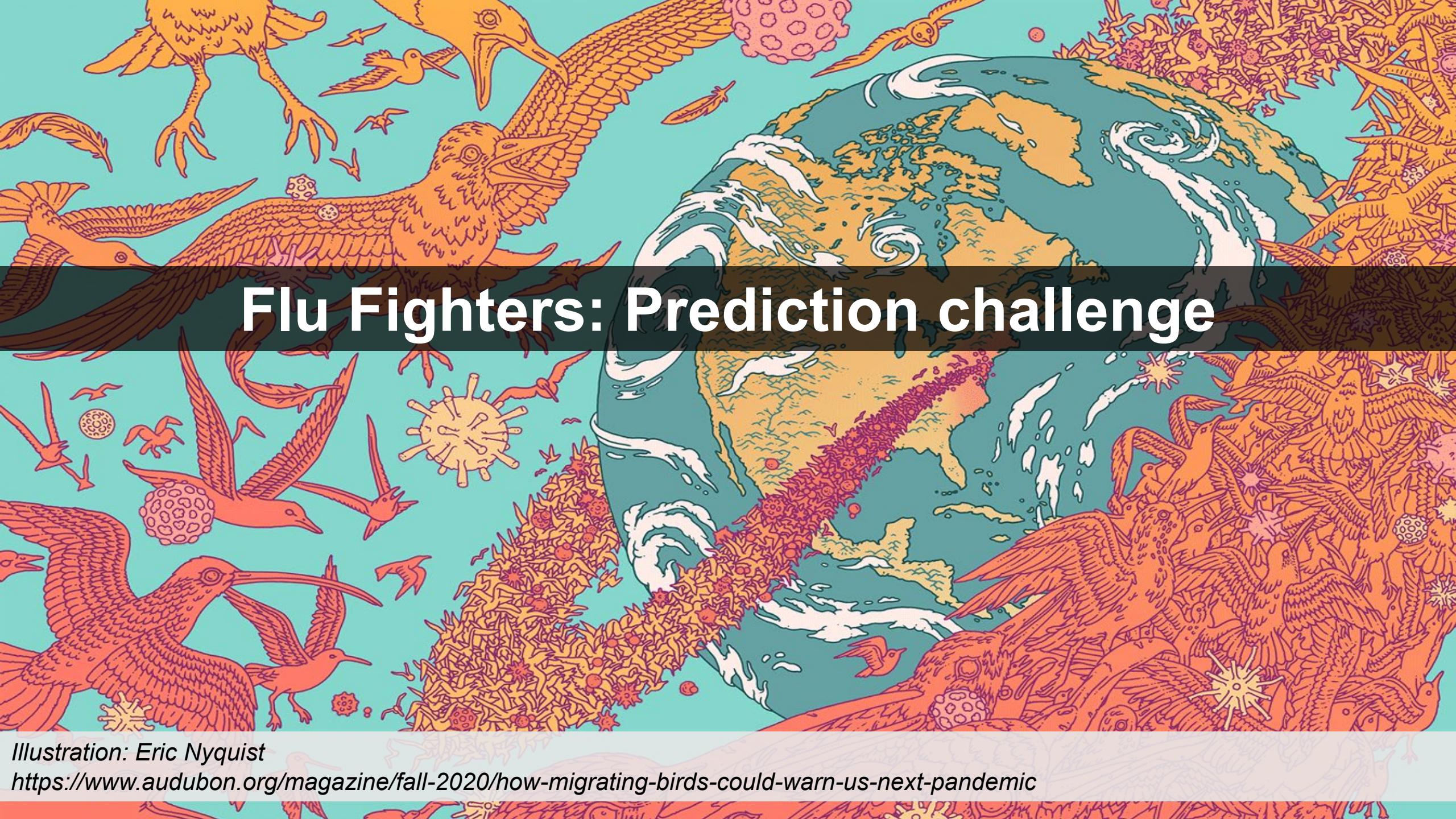
Obstacles in developing predictive models for biomedical applications

Systems vaccinology studies are complex:

- variable study design
- multi-dimensional and heterogeneous data
- many parameters and few study participants
- lack of methods to test the generalizability and predictive performance of models objectively
- Biological outliers of interest
- Accounting for variability between individuals is necessary to ensure robustness and accuracy of computational models



Bali Pulendran. Systems vaccinology: Probing humanity's diverse immune systems with vaccines. PNAS. 2014



Flu Fighters: Prediction challenge

Illustration: Eric Nyquist

<https://www.audubon.org/magazine/fall-2020/how-migrating-birds-could-warn-us-next-pandemic>

Influenza pandemics in 20th century



1918 Spanish Flu



1957 Asian Flu

1968 Hong Kong Flu



H1N1

50 to 100 million deaths

H2N2

2 million deaths

H3N2

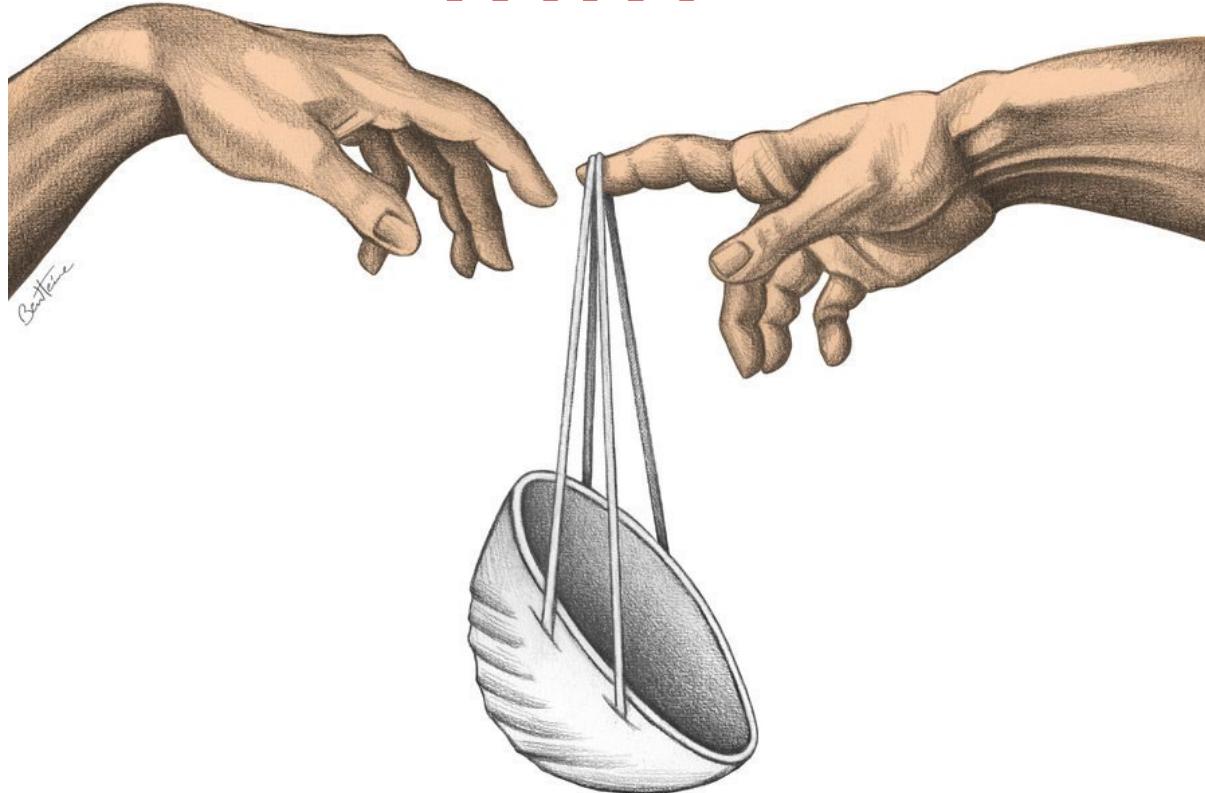
1 - 4 million deaths

COVID-19: 7.5M deaths worldwide

2009 Swine Flu

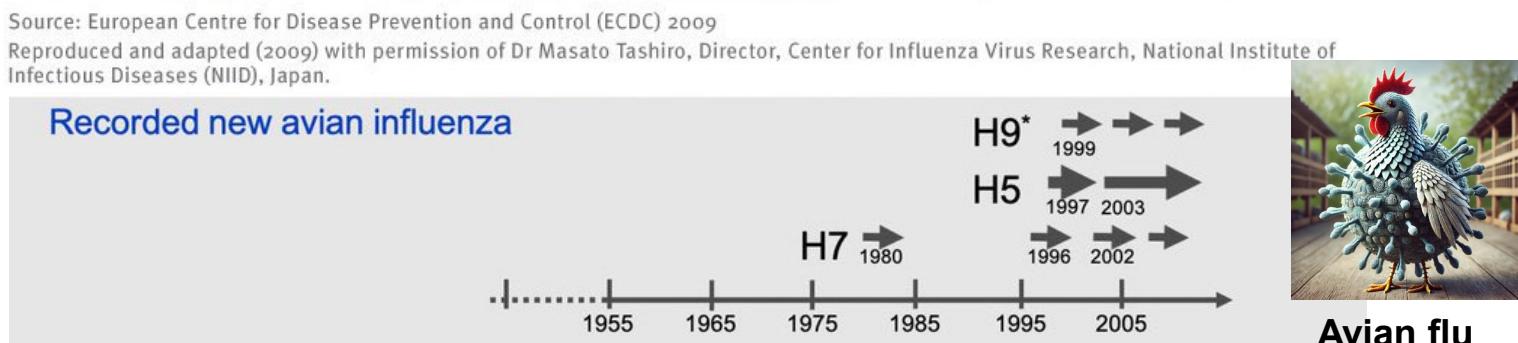
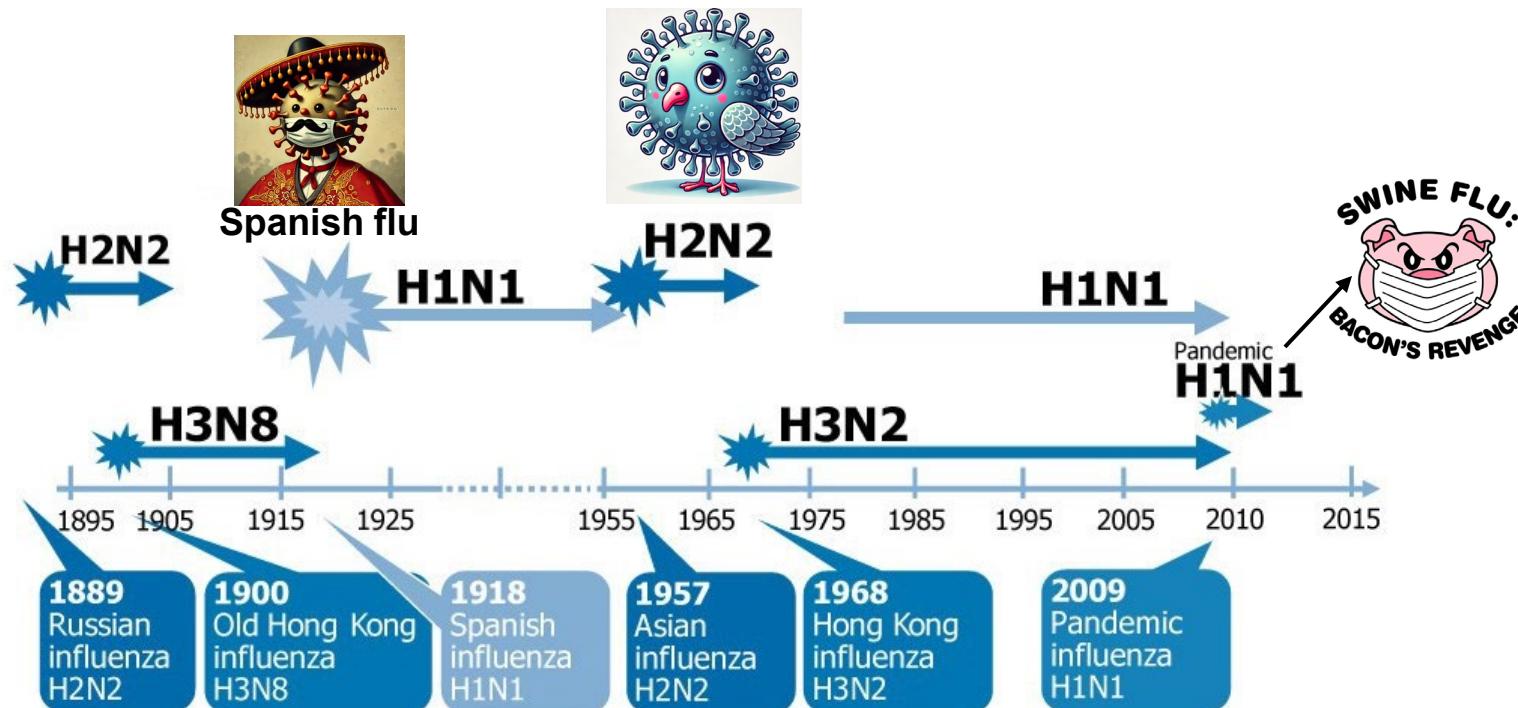


H1N1



PANDEMIC

Influenza pandemics

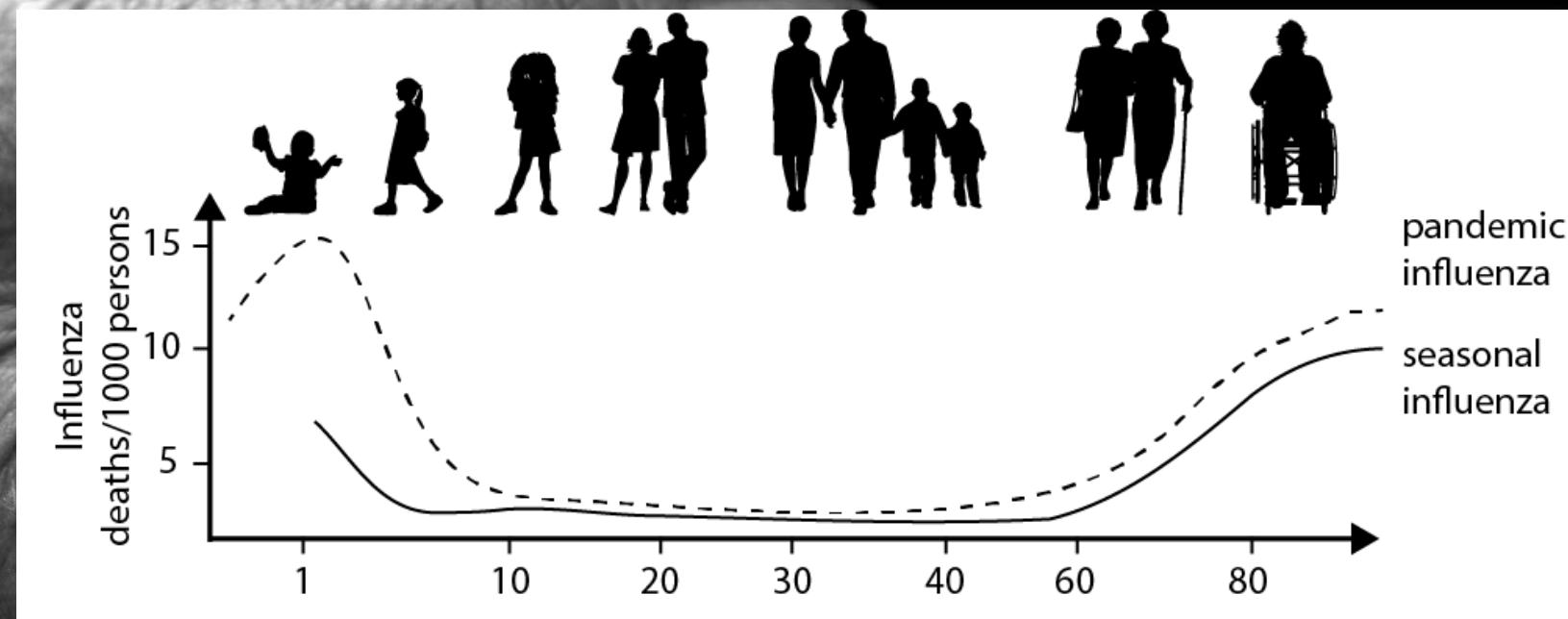


**H5N1 & H7N9 – newly emerging strains
HIGH MORTALITY RATES
30-60%**

**20x higher
than Spanish flu**

Every year

**1 in 1000 children and elderly are hospitalized
due to Influenza infection**



Influenza vaccines

INACTIVATED INFLUENZA VACCINE (IIV)



40-60%

*matched strain

LIVE ATTENUATED INFLUENZA VACCINE (LAIV)



?

85%

*mismatched strain

Rhorer et al, Vaccine, 2009

- “detergent”- split vaccine
- excellent safety profile
- low immunogenicity (adjuvants)
- Ab wane over time, no CD8+ T cells

CHILDREN FROM 6 MONTHS OF AGE

- temperature-sensitive phenotype – grows at 25°C (nasal passage)
- heterosubtypic immunity
- Ab and T cell responses maintained (IFN γ responses, $\gamma\delta$ T cells)

CHILDREN FROM 2 YEARS OF AGE

How Vaccines Work

1. A vaccine activates various immune cells because it contains part of the germ, called an antigen, that stimulates the body's immune response. An antigen by itself or in a vaccine has little to no disease-causing ability.

Immune Response to a Vaccine

Vaccines help the body acquire immunity against many disease-causing germs and cancers. A vaccine contains a killed or weakened form or derivative of an infectious germ. The vaccine has little or no disease-causing ability, but its presence in the body still provokes an immune response. This activates various immune cells that learn from the vaccine to recognize and destroy the germ.

3. When antigen-specific helper T cells encounter an antigen-presenting cell, they become activated and send a chemical messenger to other immune cells — e.g., B cells and killer T cells. The chemical messenger helps these immune cells become activated.

Antigen-presenting cell (APC)

2. The first immune cells that encounter the vaccine are called antigen-presenting cells. Each antigen-presenting cell digests an antigen, then displays on its surface a small piece of that antigen that can be recognized by T cells.

Germ activate B cell

Antigen

Helper T cell

Killer T cell

Antigen-presenting cells activate killer T cell

Plasma cell

Memory B cell

Memory Helper T cell

Memory Killer T cell

Active Killer T cell

Cells Contain

Antibodies

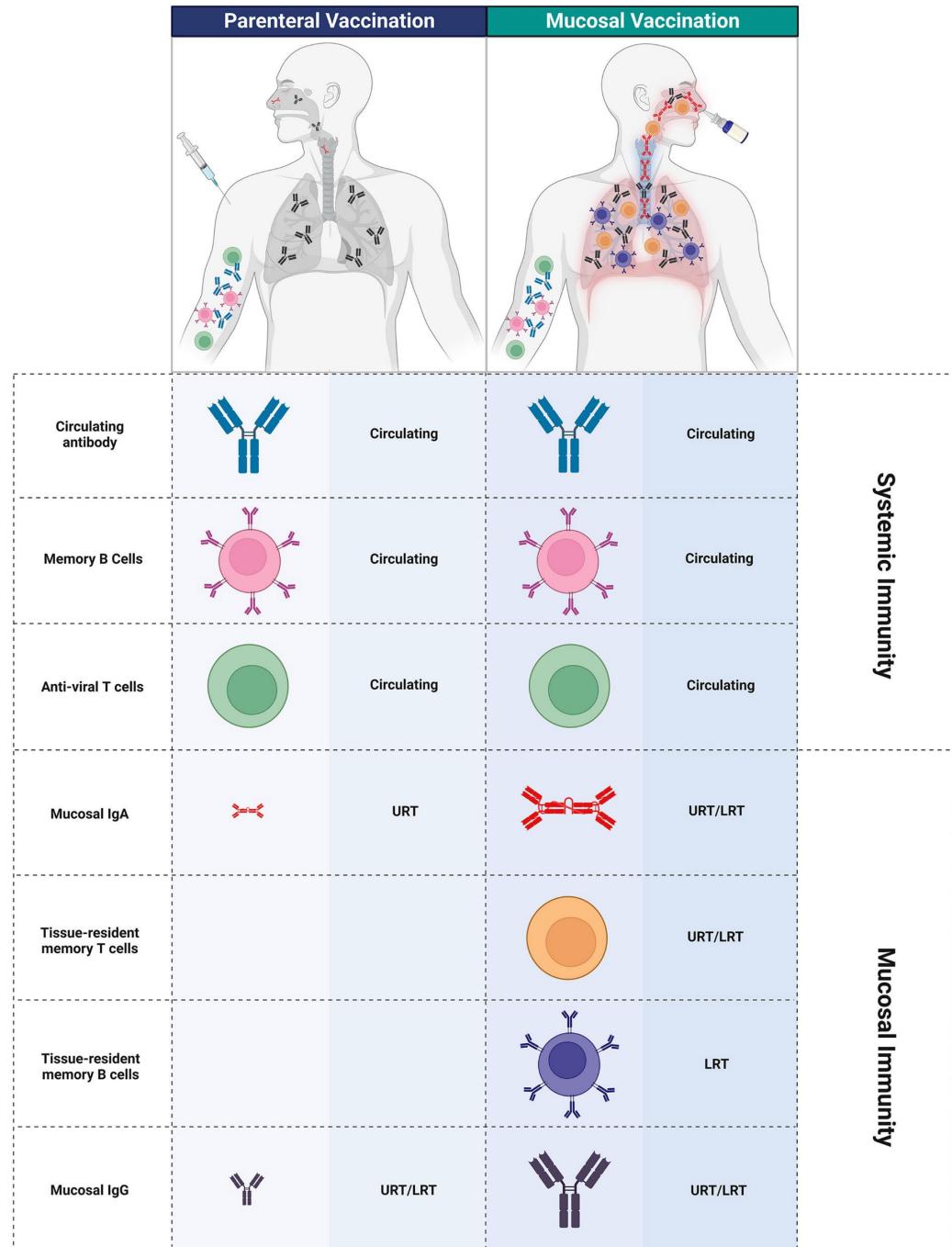
4. Once stimulated by the antigen and the chemical message from the helper T cells, the B and killer T cells divide and transform into specialized immune cells that fight back against that specific antigen. Also, a small but important fraction of the B and T cells transform into memory cells that react quicker when they encounter the same antigen again.

What defines response to LAIV?

How LAIV works?

Why is this challenge especially difficult?!

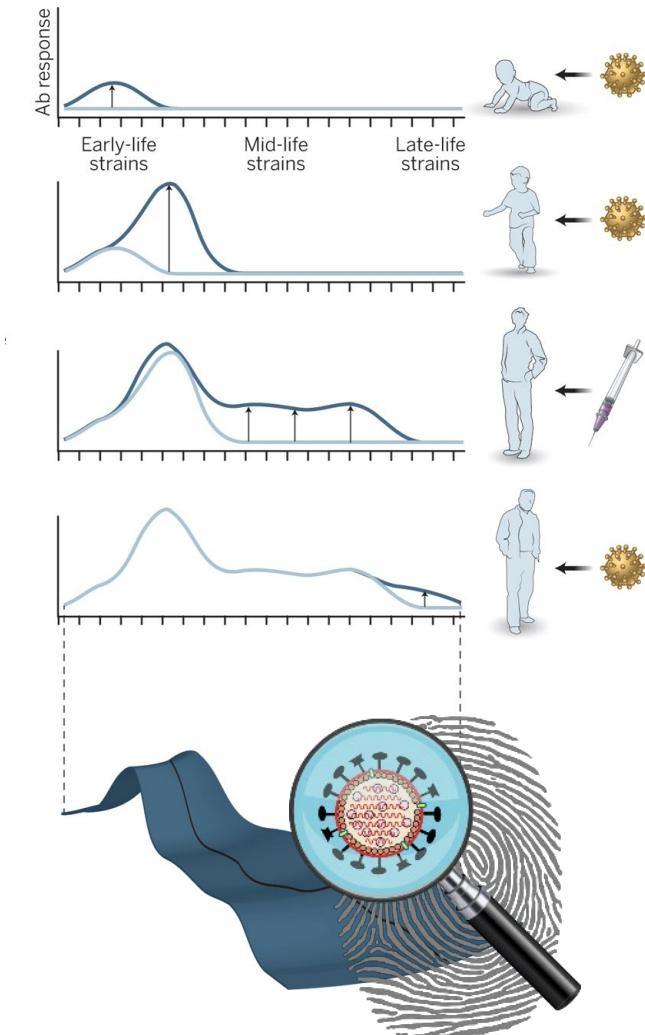
1. Immune complexity: Vaccine-induced immunity involves interactions between multiple immune compartments—humoral, cellular, and mucosal



Why is this challenge especially difficult?!

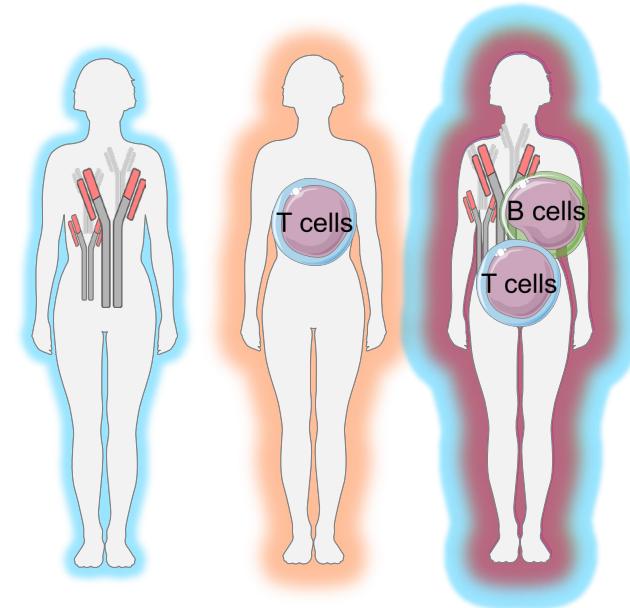
- 1. Immune complexity:** Vaccine-induced immunity involves interactions between multiple immune compartments—humoral, cellular, and mucosal
- 2. FluPRINT: pre-existing influenza immunity that shapes responses to flu vaccines** (baseline immune states across both mucosal and systemic sites).

FLUPRINT: lifetime of exposure to influenza



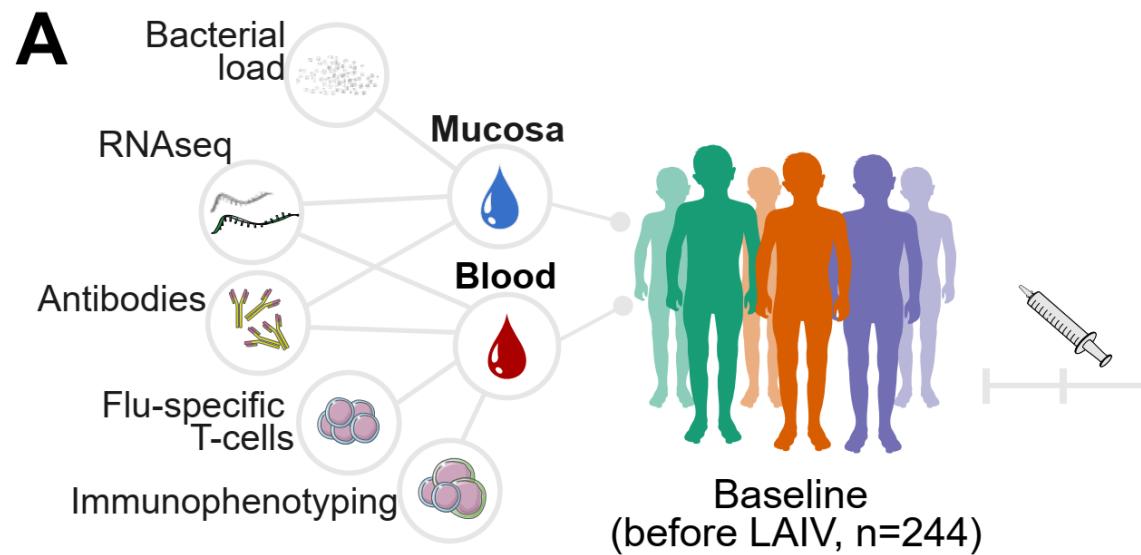
Why is this challenge especially difficult?!

- 1. Immune complexity:** Vaccine-induced immunity involves interactions between multiple immune compartments—humoral, cellular, and mucosal
- 2. FluPRINT: pre-existing influenza immunity that shapes responses to flu vaccines** (baseline immune states across both mucosal and systemic sites).
- 3. Interindividual variability:** No two individuals respond to a vaccine in exactly the same way, and biological outliers make predictions even harder.

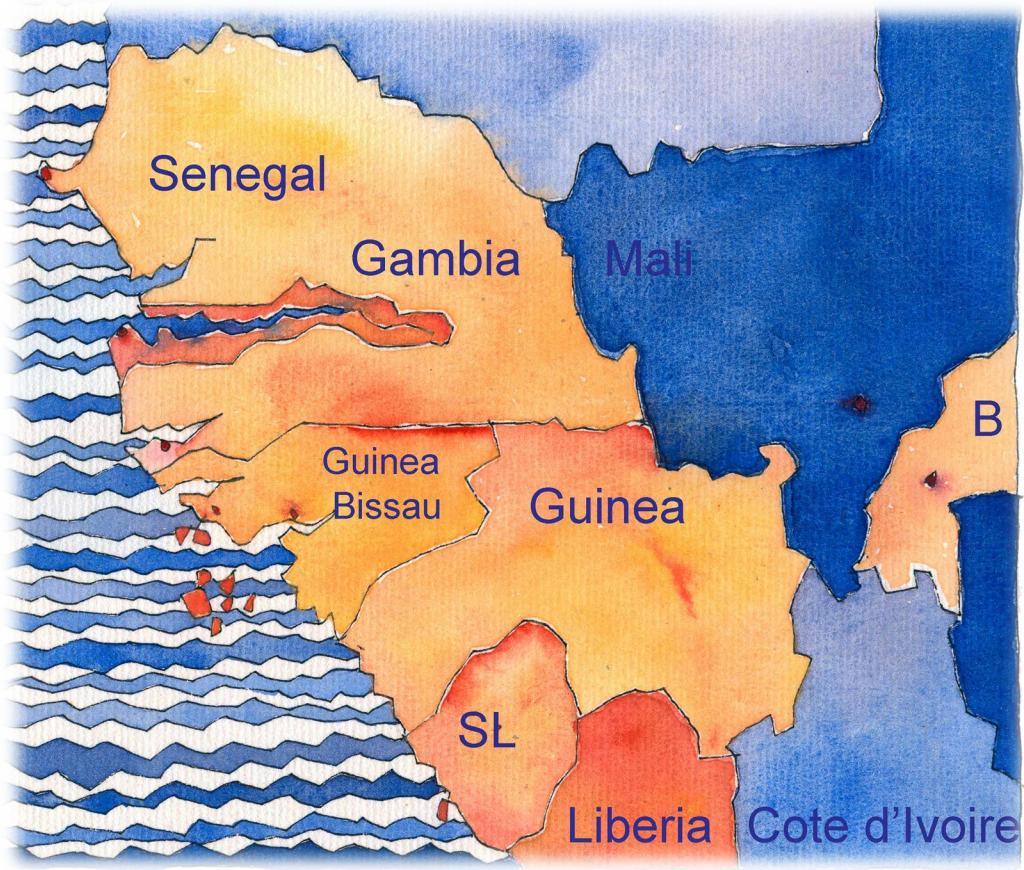
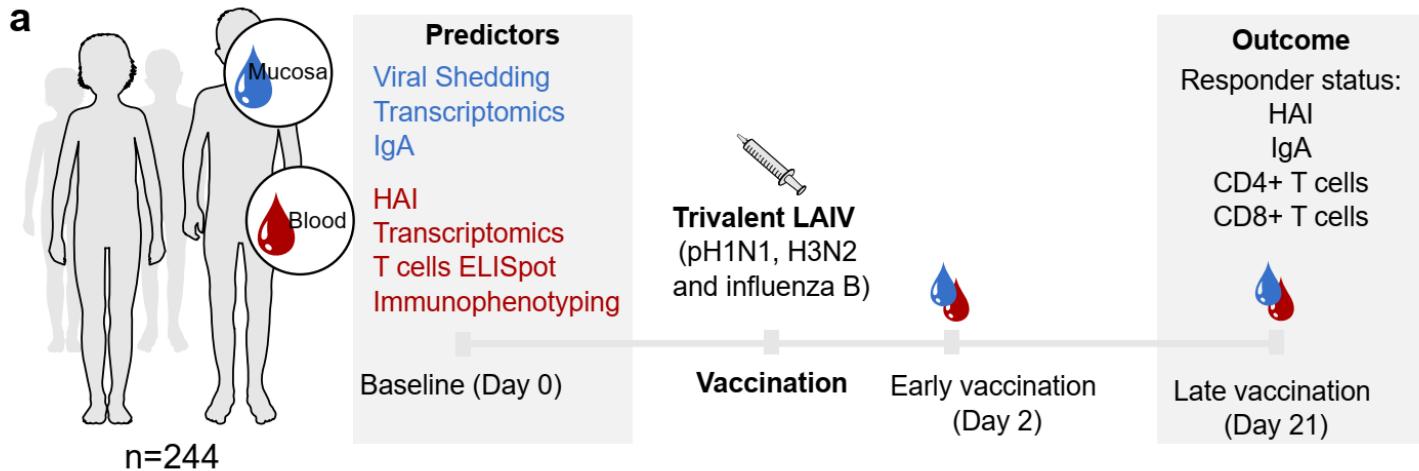


Why is this challenge especially difficult?!

- 1. Immune complexity:** Vaccine-induced immunity involves interactions between multiple immune compartments—humoral, cellular, and mucosal
- 2. FluPRINT: pre-existing influenza immunity that shapes responses to flu vaccines** (baseline immune states across both mucosal and systemic sites).
- 3. Interindividual variability:** No two individuals respond to a vaccine in exactly the same way, and biological outliers make predictions even harder.
- 4. High-dimensional data:** You'll work with transcriptomics, cellular data, and serological markers, all integrated into a single dataset.
- 5. Missing Data and Noise:** Real-world datasets like this one come with challenges like missing values and measurement noise.



The dataset

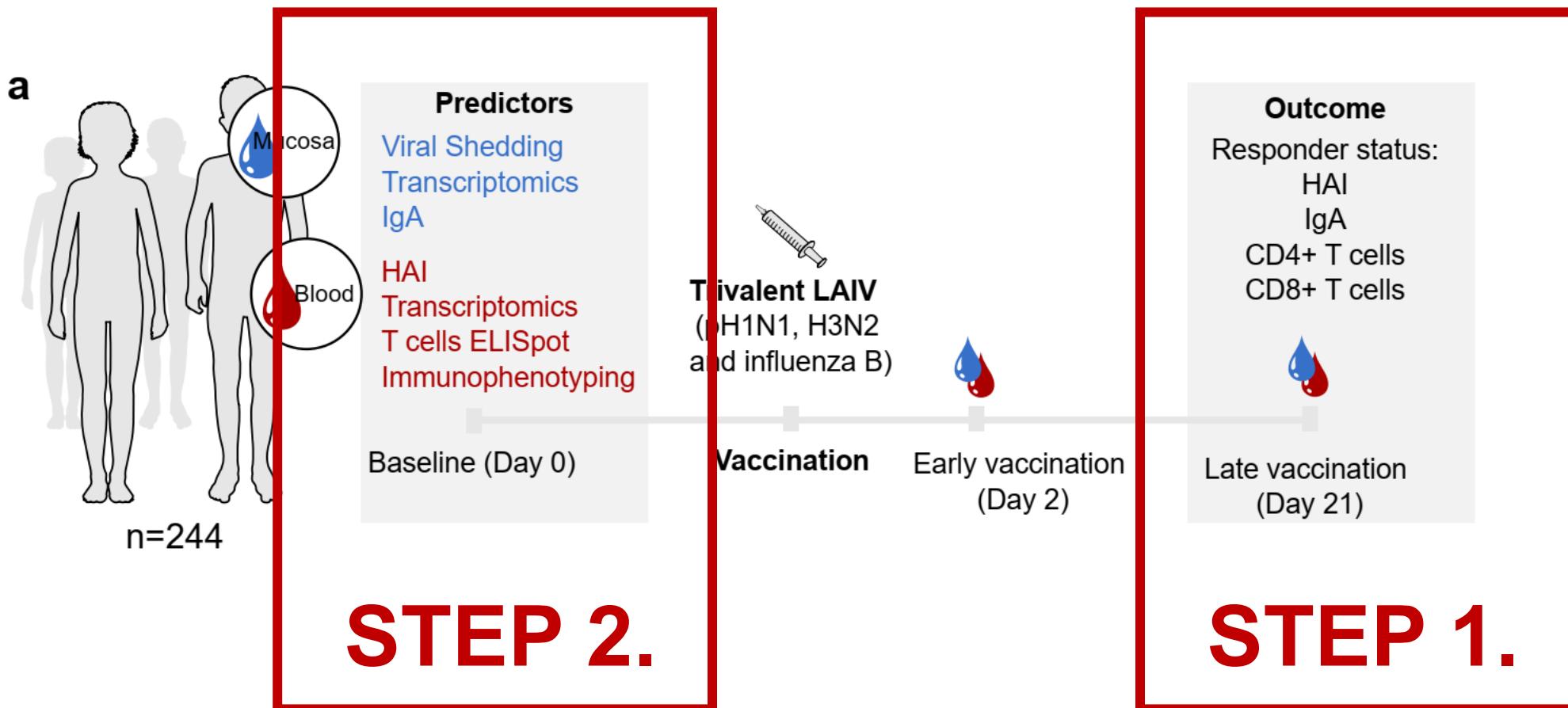


Effect of a Russian-backbone live-attenuated influenza vaccine with an updated pandemic H1N1 strain on shedding and immunogenicity among children in The Gambia: an open-label, observational, phase 4 study

Benjamin B Lindsey*, Ya Jankey Jagne*, Edwin P Armitage*, Anika Singanayagam, Hadjatou J Sallah, Sainabou Drammeh, Elina Senghore, Nuredin I Mohammed, David Jeffries, Katja Höschler, John S Tregoning, Adam Meijer, Ed Clarke, Tao Dong, Wendy Barclay, Beate Kampmann, Thushan I de Silva

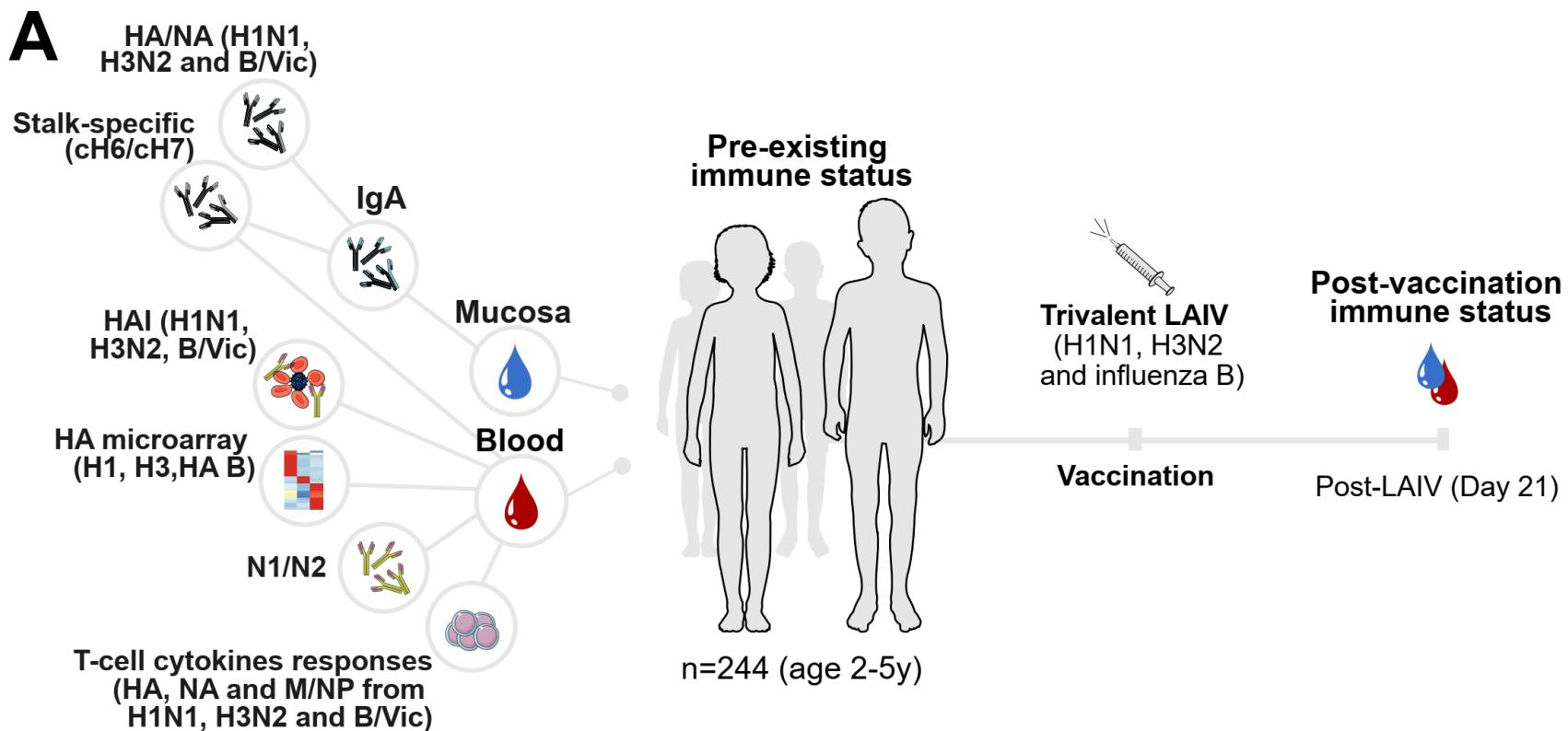
The task

To predict **vaccine responsiveness**
(immunogenicity) based on **baseline**
immune features.

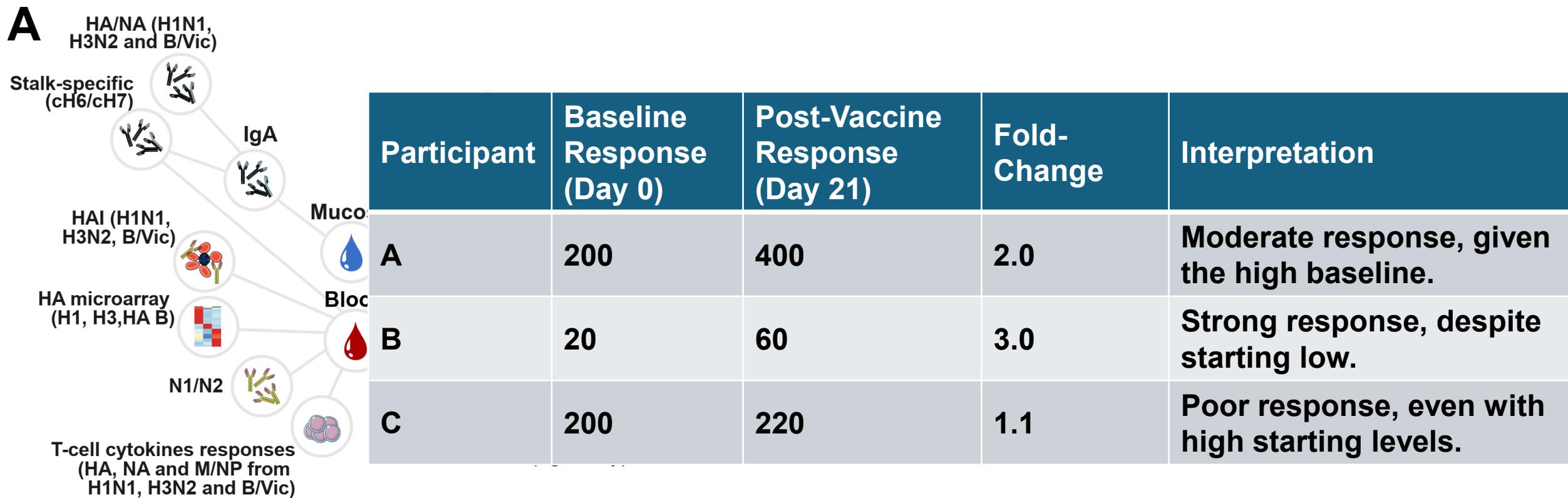


Step 1. Vaccine immunogenicity (responsiveness)

Responsiveness reflects how well a vaccine elicits an immune response, spanning humoral, cellular, and mucosal immunity.

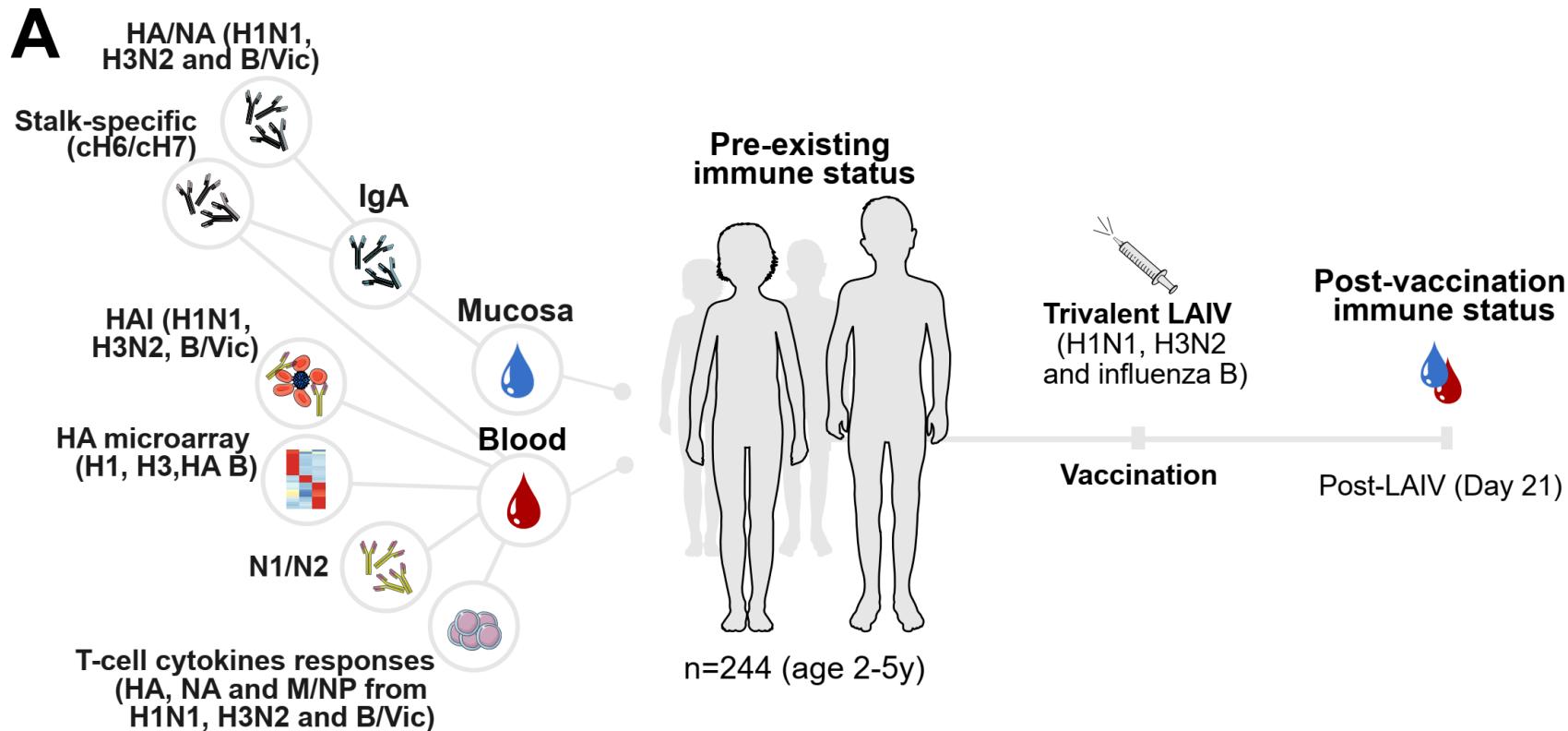


Step 1. Vaccine immunogenicity (fold-change)



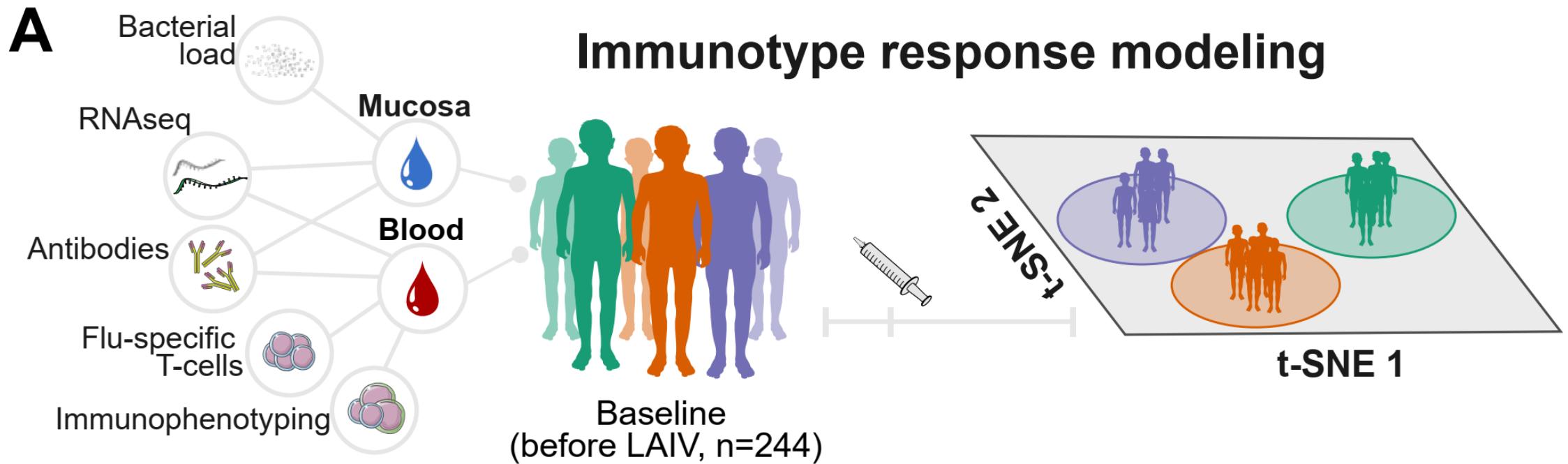
Fold-change (V_{21}/V_0) → This approach ensures that **pre-existing immune states** are accounted for when assessing vaccine-induced changes. Using the fold-change, we can compare individuals starting at very different baseline levels.

Step 1. Defining research question – what are you predicting?



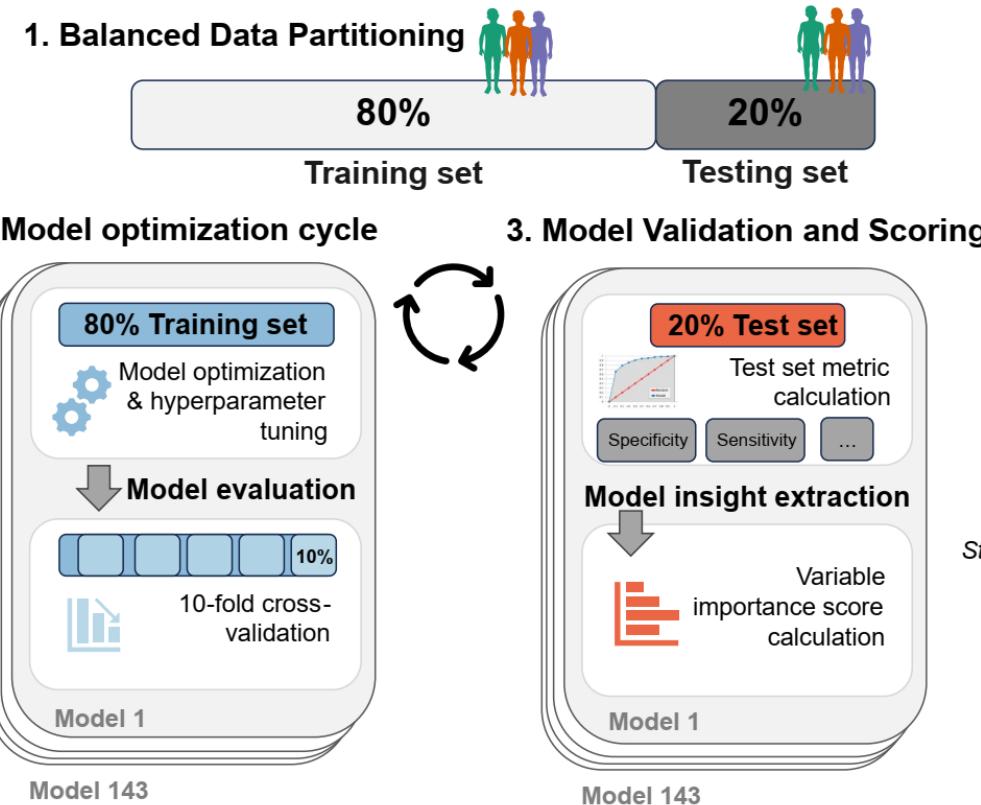
1. Are you predicting a specific **outcome**: such as Humoral responses (HAI titers), cellular responses (T-cell cytokine production), or Antibody binding (IVPM)?
High vs low HAI responder?
2. Are you exploring clusters of individuals based on shared **vaccine responsiveness patterns**? Group 1 individuals with High IgA and high T-cell responses vs Group 2 (strong T-cell responses and high systemic antibodies).
Clustering-based exploration with distinct groups.

Step 2. Baseline predictors



1. Which predictors (features) will you use to build your model?
2. Will you include only **baseline (Day 0)** data, or also early post-vaccination data (Days 2, 7)?
3. Do you need **Day 21 (V21)** data?

Step 3. How will you model it?



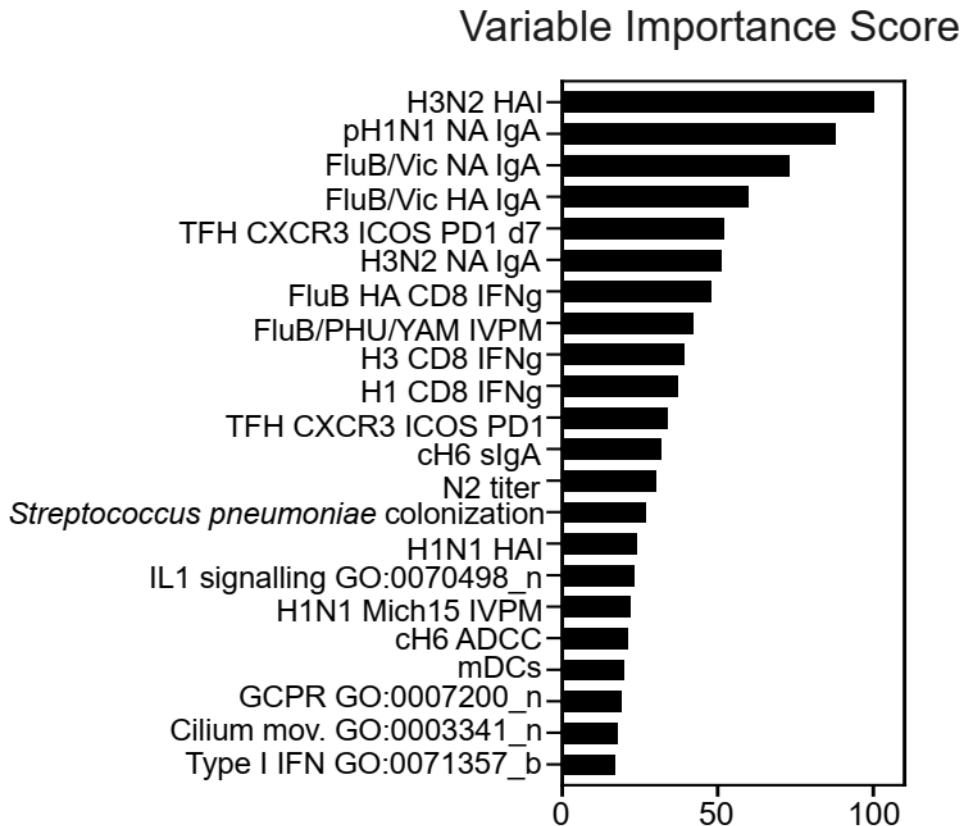
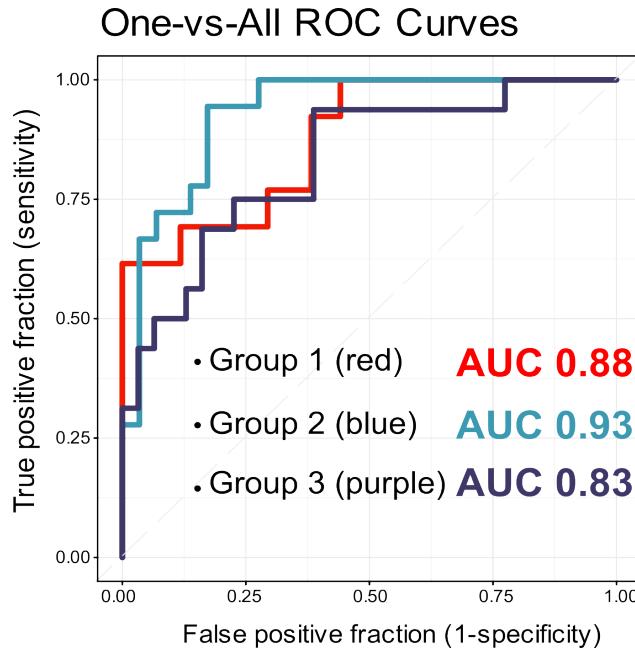
1. How will you preprocess your data?

- Will you handle missing values with imputation or use **Mulset** for noise reduction?
- Will you remove highly correlated features and drop near-zero variance features?

2. How will you partition your data into training and test sets?

3. What algorithms will you use?

Step 4. How will you know if your model is good?



1. How will you measure model **performance**?
2. How will you ensure your model is **not overfitting**?
3. What **biological insights** can you draw from your model?

Timeline and Guidelines

Nov 26th: Self-paced

Teams will work independently to develop a **strategy** for predictive modeling.

Dec 3rd: Teamwork in the class with the instructor.

Showcase your strategy to the instructor and discuss any final questions or ideas to **improve your approach**.

Dec 4th: Submission Deadline

Submit your final models and analysis by **end of the day**. Each team should email their final model to Dr. Tomic.

Dec 5th: Team Presentations

Each team will present their **strategy for machine learning predictive modeling**.
A winner will be announced based on creativity, execution, and predictive performance

Flu Fighters of the Year announcement

