CMI-PB Prediction Challenge

1st Informational Session

Oct 6th, 2023

La Jolla Institute for Immunology

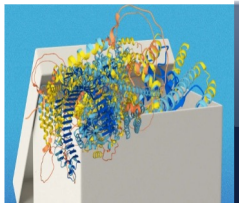
CMI-PB
COMPUTATIONAL MODELS OF IMMUNITY
PERTUSSIS BOOST

Examples of prediction challenges

Benefits:

Quantitative and unbiased evaluation of model performance
Drives participants to strive for accuracy and outperform their other participants
Encourages a wide range of participants to join
Contributes to a deeper understanding of the topic at hand

The Netflix Prize: How a $1 Million Contest Changed Binge-Watching Forever



15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

DeepMind has predicted the structure of almost every protein known to science

The Food and Drug Administration (FDA) and National Cancer Institute (NCI) call on the scientific community to develop and evaluate computational algorithms that can accurately detect and correct mislabeled samples using rich multi-omics datasets

kaggle

DREAM CHALLENGES

Obstacles in developing computational models for biological applications

CMI-PB
COMPUTATIONAL MODELS OF IMMUNITY
PERTUSSIS BOOST

- Systems vaccinology studies have:
  - Varying study designs
  - Multi-dimensional and heterogeneous data
  - Many parameters and few study participants
  - A lack of methods to test the generalizability and predictive performance of models objectively
- Accounting for variability between individuals is necessary to ensure the robustness and accuracy of computational models
- It is important to capture temporal dynamics in system vaccinology models

Our Approach:

- To establish a community platform to develop and test computational models of immunity in vaccination
- To better understand vaccine-induced immunity to *B. pertussis*

Agenda for Today's Session

1. Introduction - Project Overview
2. The CMI-PB Challenge
3. Example Model Presentation
4. Resources Overview
5. Q & A

Presentation Agenda

1. **Introduction- Project Overview**
2. The CMI-PB Challenge
3. Example Model Presentation
4. Resources Overview
5. Q & A

PERTUSSIS

a.k.a. WHOOPING COUGH

fb.com/dawnsbrain

CAUSE

Bordetella pertussis
bacteria in upper respiratory system

SYMPTOMS

weeks 1-2

low fever
runny nose
mild cough
apnea (in infants)

weeks 3-8

coughing fits
ending in a "whoop"
vomiting
exhaustion

PREVENTION

DTaP vaccine
injections plus
Tdap adult booster

COMPLICATIONS

earache

pneumonia

encephalopathy
seizures
cerebral hypoxia

fractured ribs
failure to thrive

death

EPIDEMIOLOGY

spread via **coughing**, **sneezing**, and other close contact

3+ months
avg. length of infection

5.5
avg. number of people an infected person will infect while contagious (R_0)

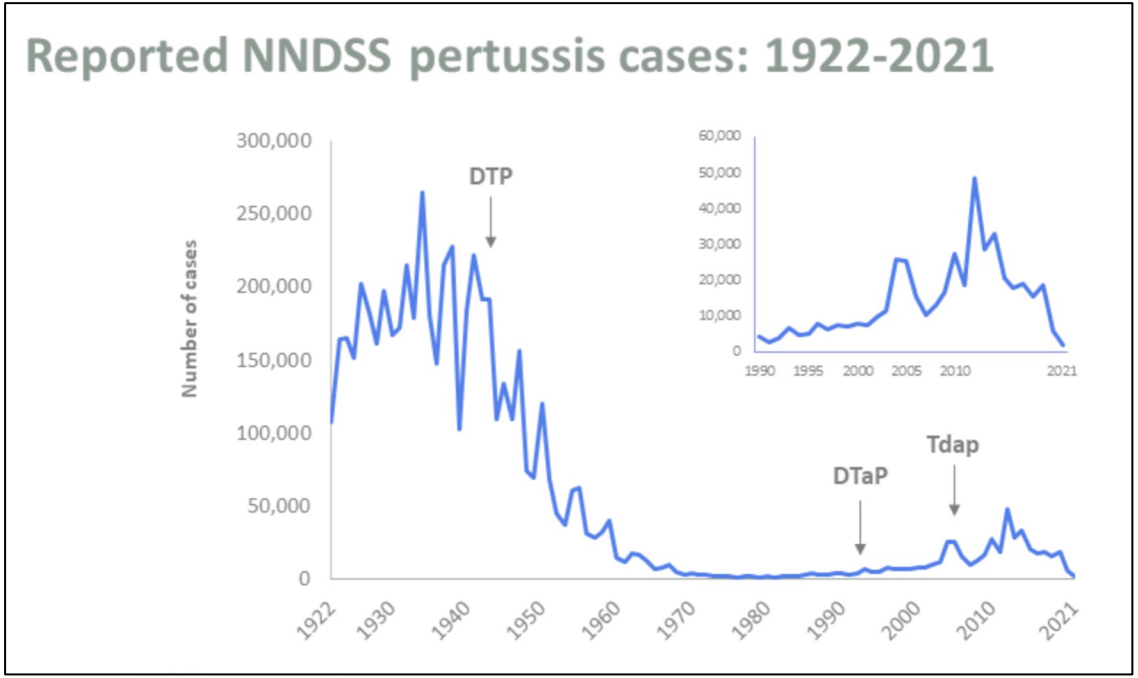
48,500,000
people infected annually worldwide

195,000
annual **deaths** worldwide

SOURCES: Centers for Disease Control, World Health Organization, PLOS Medicine, PubMed

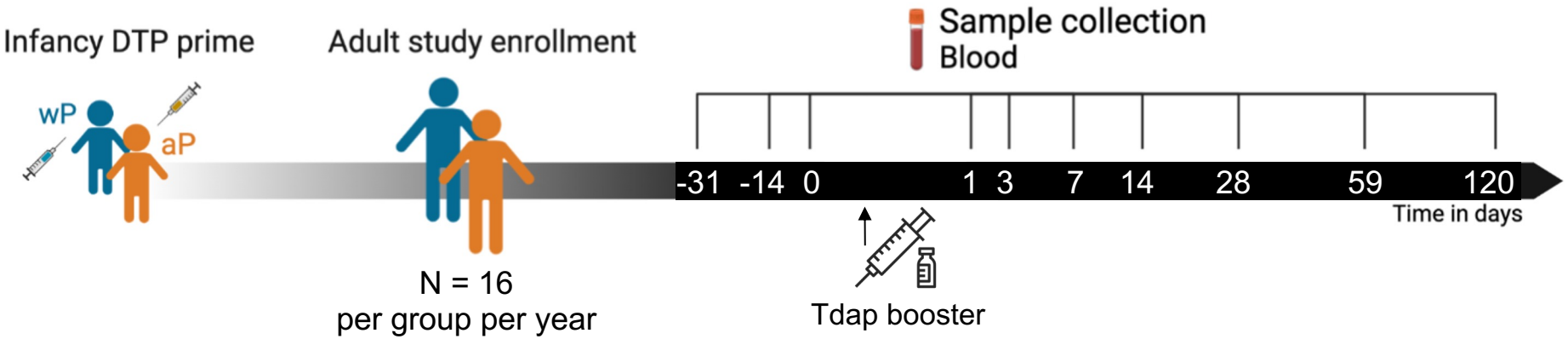
Waning Immunity from aP pertussis vaccination

- 1940s: Introduction of an inactivated whole bacteria PT vaccine (**wP**) dramatically decreased cases
- 1995: Vaccine-related side effects led to a replacement with the acellular PT vaccine (**aP**) in the USA
- aP induced protection wanes faster than wP → **Why?**



Source: National Notifiable Diseases Surveillance System, CDC

Recruitment Strategy



Characterizing immune responses - Multiomics approach

PBMC cell frequencies by flow cytometry

- Total of 37 distinct cell populations

Plasma antigen-specific antibody titers by Luminex

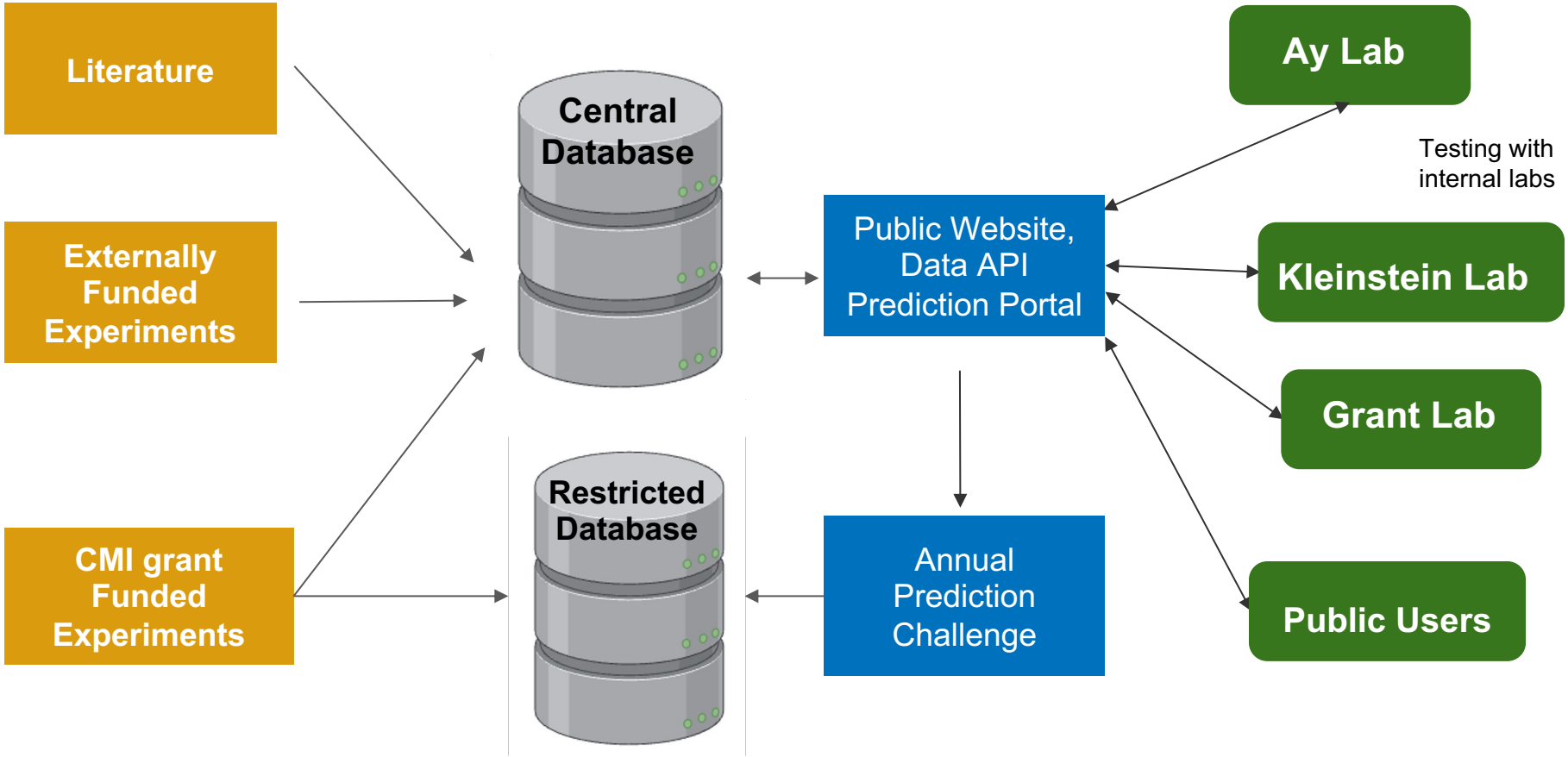
- Antibody Isotypes: IgG, IgG1, IgG2, IgG3, IgG4
- Vaccine Antigens
  - Pertussis Toxin (PT), PRN, FHA, FIM2/3
  - Tetanus Toxoids (TT)
  - Diphtheria Toxoids (DT)
  - OVA (irrelevant control)

Plasma proteomics by Olink

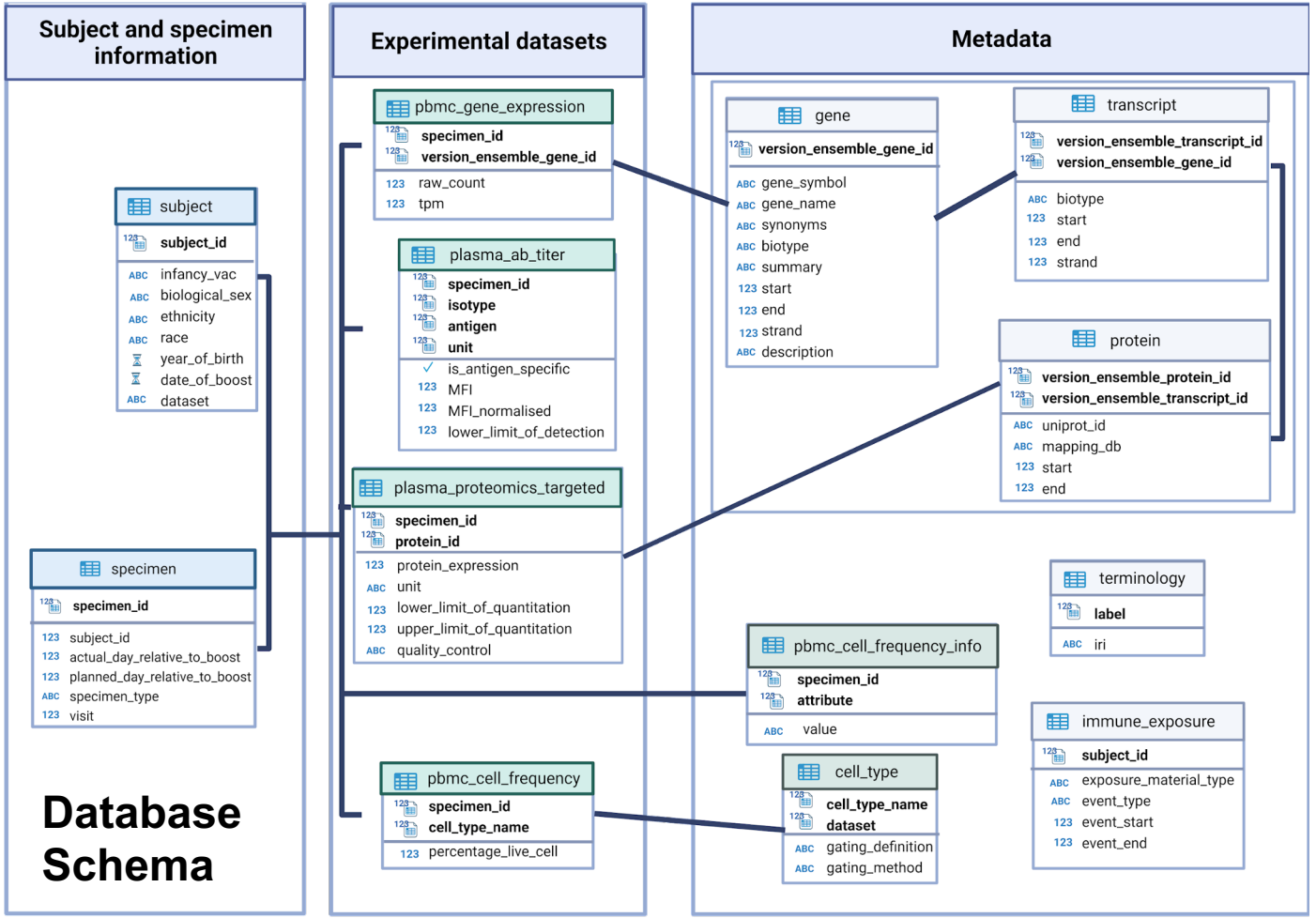
- Concentration of 45 cytokines

Transcriptomics by bulk RNA-Seq

d. Databases, Model building, and Prediction Challenge



e. Providing access to experimental data in a standardized format



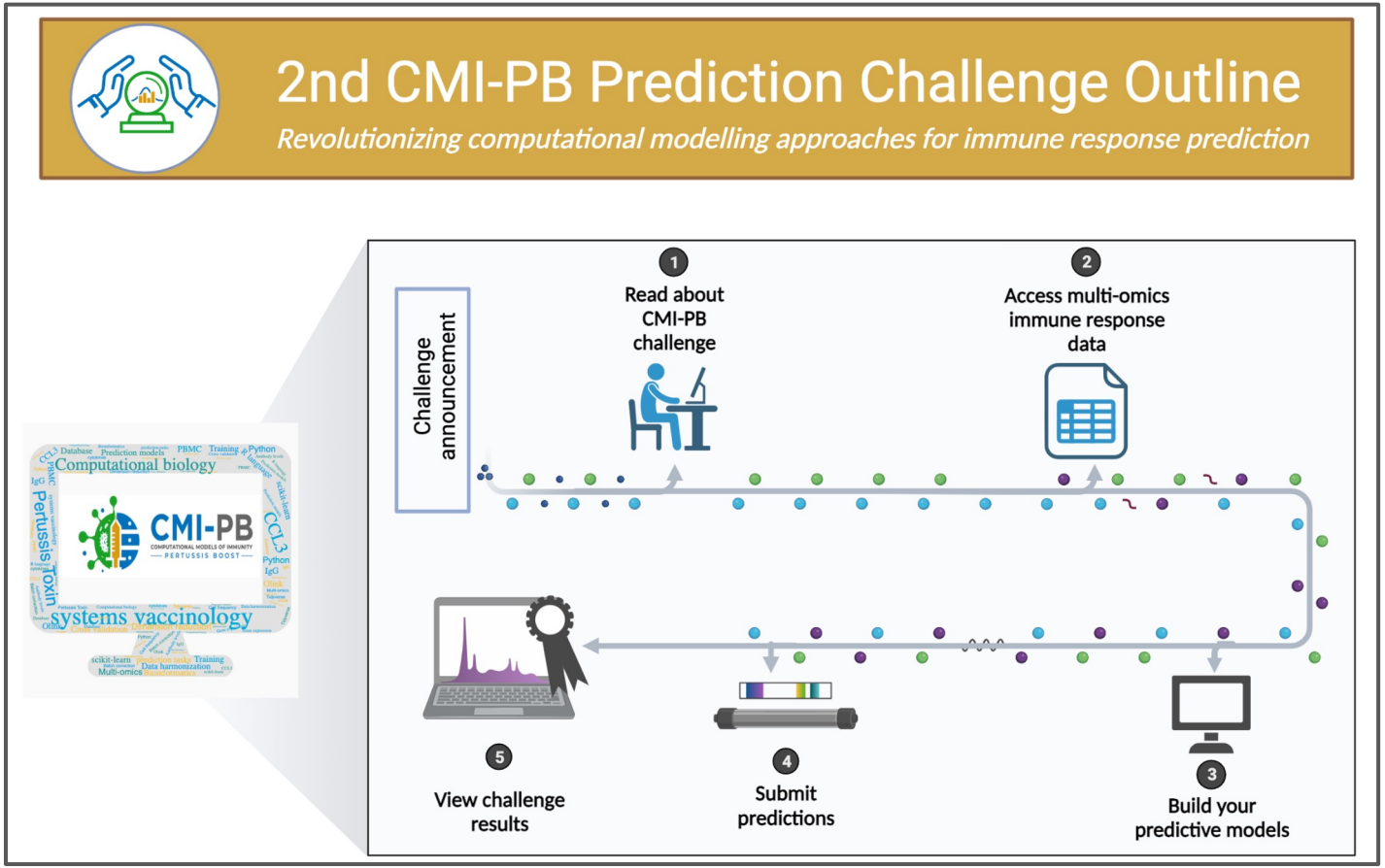
Presentation Agenda

1. Introduction - Project Overview
2. **The CMI-PB Challenge**
3. Example Model Presentation
4. Resources Overview
5. Q & A

a. Past and future CMI-PB annual prediction challenges

	Annual prediction challenge title	Contestants	Number of subjects		Current status
			Training dataset	Test dataset	
1	**First Challenge:** Internal dry run	CMI-PB consortium	60 (28 aP + 32 wP)	36 (19 aP + 17 wP)	Concluded in May 2022
2	**Second Challenge:** Invited challenge	Invited contestants	96 (47 aP + 49 wP)	22 (13 aP + 9 wP)	Announced on September 12, 2023
3	**Third Challenge:** Open Challenge 1	Public	118 (60 aP + 58 wP)	32 (16 aP + 16 wP)	Will be announced in April 2024
4	**Fourth Challenge:** Open Challenge 2	Public	150 (76 aP + 74 wP)	32 (16 aP + 16 wP)*	Will be announced in December 2024

b. Prediction challenge outline

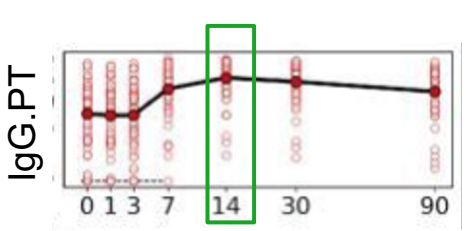


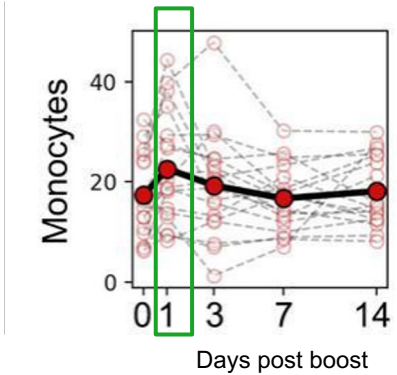
c. Formulating prediction tasks for CMI-PB Challenge

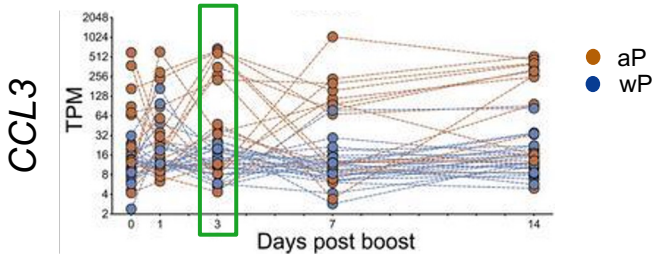
- Previously identified **vaccine responses** are formulated as prediction tasks*

- **General vaccine responses**:
  - Plasma IgG levels increased at day 14 post-booster vaccination compared to baseline
  - Increase in the percentage of monocytes on day 1 post-booster than baseline

- **aP/wP specific vaccine responses:**
  - A subset of aP-primed individuals showed an increased expression of proinflammatory genes, including *CCL3* at day 3 post-booster vaccination







* **A system-view of *Bordetella pertussis* booster vaccine responses in adults primed with whole-cell versus acellular vaccine in infancy**

Ricardo da Silva Antunes, … , Alessandro Sette, Bjoern Peters

JCI Insight. 2021;6(7):e141023. https://doi.org/10.1172/jci.insight.141023.

c. Formulating prediction tasks for CMI-PB Challenge

List of tasks

1) Antibody titer tasks

1.1) Rank the individuals by IgG antibody titers against pertussis toxin (PT) that we detect in plasma 14 days post booster vaccinations.

1.2) Rank the individuals by fold change of IgG antibody titers against pertussis toxin (PT) that we detect in plasma 14 days post booster vaccinations compared to titer values at day 0.

2) Cell frequencies tasks

2.1) Rank the individuals by predicted frequency of Monocytes on day 1 post boost after vaccination.

2.2) Rank the individuals by fold change of predicted frequency of Monocytes on day 1 post booster vaccination compared to cell frequency values at day 0.

3) Gene expression tasks

3.1) Rank the individuals by predicted gene expression of CCL3 on day 3 post-booster vaccination.

3.2) Rank the individuals by fold change of predicted gene expression of CCL3 on day 3 post booster vaccination compared to gene expression values at day 0.

predicted values

predicted fold-change values

The ultimate goal is to model as many of the tasks as possible. However, contestants are not required to submit answers for all tasks.

Example of Rankings

Subject ID	Predicted value	Rank
101	2.9	4
102	**9.1**	**1**
103	**1.2**	**5**
104	4.5	3
105	4.7	2

Submission file preparation

	A	B	C	D	E	F	G	H	I	J
1	Subject ID	Age	Biological Sex at Birth	Vaccine Priming Status	1.1) IgG-PT-D14-titer-Rank	1.2) IgG-PT-D14-FC-Rank	2.1) Monocytes-D1-Rank	2.2) Monocytes-D1-FC-Rank	3.1) CCL3-D3-Rank	3.2) CCL3-D3-FC-Rank
2	97	35	Male	wP	14	6	15	21	11	17
3	98	28	Female	wP	3	7	13	15	4	1
4	99	22	Female	aP	7	2	18	12	15	13
5	100	20	Female	aP	20	20	10	11	9	6
6	101	18	Male	aP	19	5	1	8	1	19
7	102	18	Male	aP	5	21	11	17	2	9
8	103	27	Female	wP	21	4	4	7	10	12
9	104	32	Female	wP	15	12	14	16	7	14
10	105	27	Female	wP	9	15	17	13	20	11
11	106	25	Female	aP	1	11	3	18	19	3
12	107	23	Female	aP	2	16	19	10	5	4
13	108	26	Female	wP	10	17	7	1	21	5
14	109	32	Female	wP	8	18	12	20	8	15
15	110	24	Female	aP	12	13	16	19	12	2
16	111	25	Male	wP	11	8	20	5	14	16
17	112	25	Male	aP	4	10	5	9	3	20
18	114	31	Male	wP	13	1	21	14	13	8
19	115	19	Female	aP	18	3	8	2	18	21
20	116	21	Male	aP	8	19	8	4	8	7
21	117	27	Female	aP	17	9	9	8	17	18
22	118	23	Male	aP	16	14	2	3	16	10

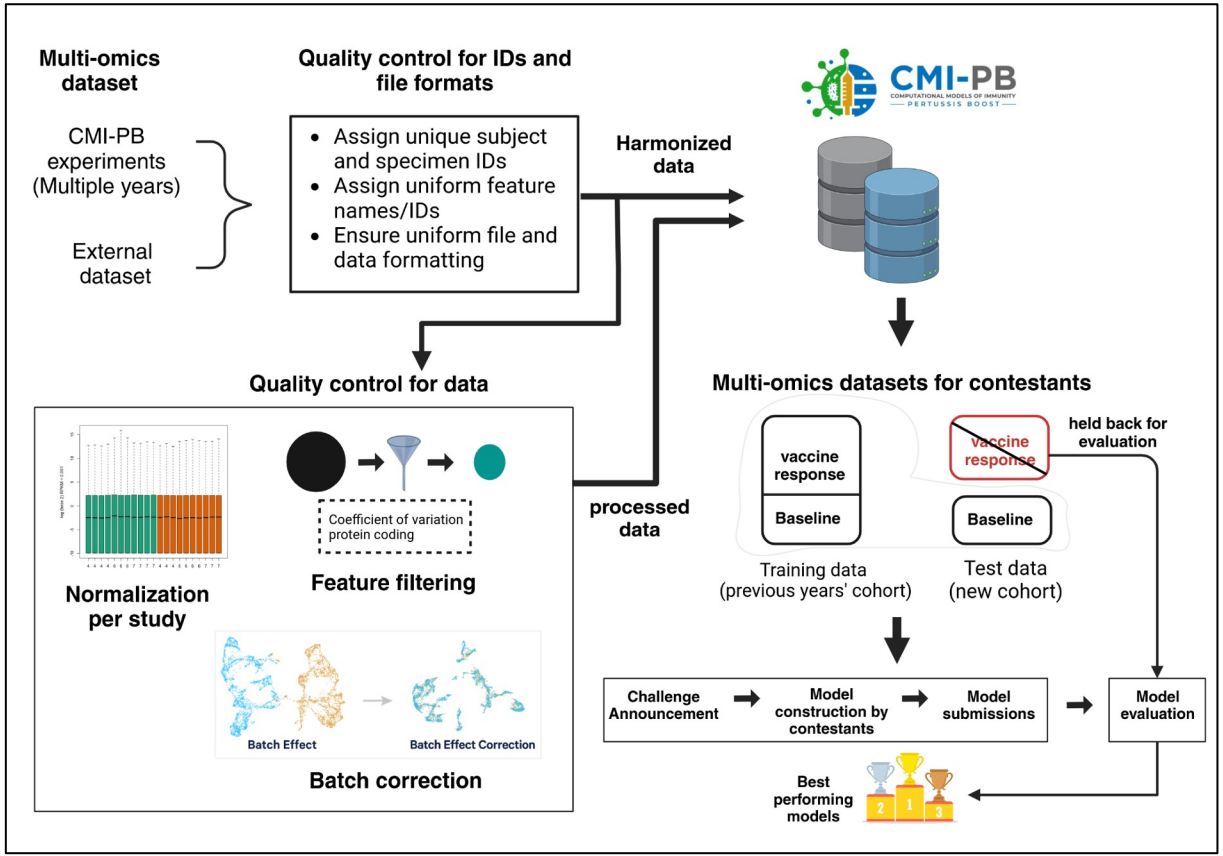
A complete submission

A submission file with entries for two tasks

	Subject ID	Age	Biological Sex at Birth	Vaccine Priming	1.1) IgG-PT-D14-tit	1.2) IgG-PT-D14-FC-Rank	2.1) Monocytes-D1-R	2.2) Monocytes-D1-P	3.1) CCL3-D3	3.2) CCL3-D3-FC
2	97	35	Male	wP	14				11	
3	98	28	Female	wP	3				4	
4	99	22	Female	aP	7				15	
5	100	20	Female	aP	20				9	
6	101	18	Male	aP	19				1	
7	102	18	Male	aP	5				2	
8	103	27	Female	wP	21				10	
9	104	32	Female	wP	15				7	
10	105	27	Female	wP	9				20	
11	106	25	Female	aP	1				19	
12	107	23	Female	aP	2				5	
13	108	26	Female	wP	10				21	
14	109	32	Female	wP	8				8	
15	110	24	Female	aP	12				12	
16	111	25	Male	wP	11				14	
17	112	25	Male	aP	4				3	
18	114	31	Male	wP	13				13	
19	115	19	Female	aP	18				18	
20	116	21	Male	aP	8				8	
21	117	27	Female	aP	17				17	
22	118	23	Male	aP	16				16	

Populated with random numbers.

d. Overview of the CMI-PB Challenge data

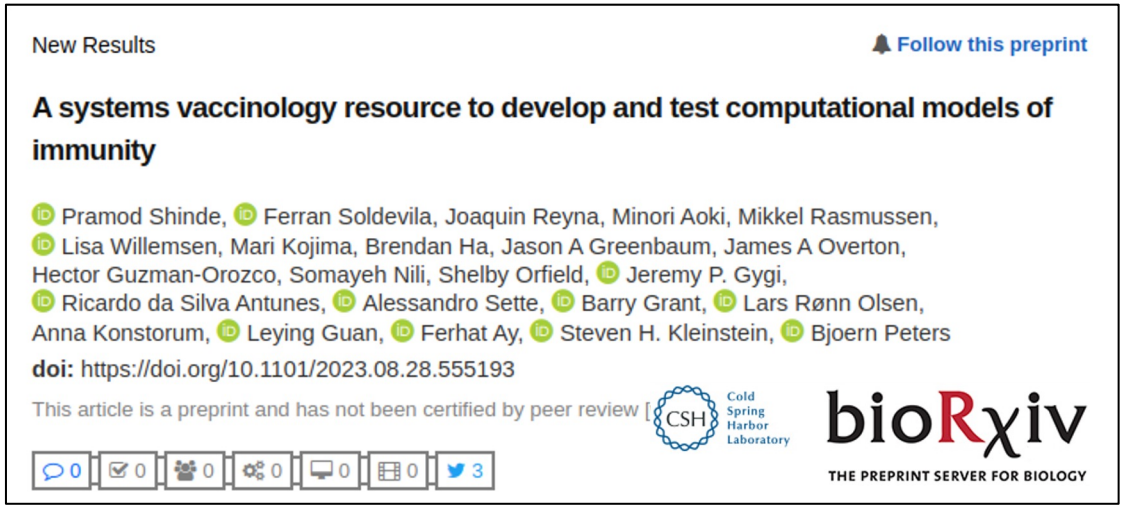


Challenge related information and Data access is provided via the CMI-PB website

Presentation Agenda

1. Introduction-Project Overview
2. The CMI-PB Challenge
3. **Example Model Presentation**
4. Resources Overview
5. Q & A

a. Models from First (internal) challenge



New Results

Follow this preprint

A systems vaccinology resource to develop and test computational models of immunity

Pramod Shinde, Ferran Soldevila, Joaquin Reyna, Minori Aoki, Mikkel Rasmussen, Lisa Willemsen, Mari Kojima, Brendan Ha, Jason A Greenbaum, James A Overton, Hector Guzman-Orozco, Somayeh Nili, Shelby Orfield, Jeremy P. Gygi, Ricardo da Silva Antunes, Alessandro Sette, Barry Grant, Lars Rønn Olsen, Anna Konstorum, Leying Guan, Ferhat Ay, Steven H. Kleinstein, Bjoern Peters

doi: https://doi.org/10.1101/2023.08.28.555193

This article is a preprint and has not been certified by peer review [

Cold Spring Harbor Laboratory

bioRxiv

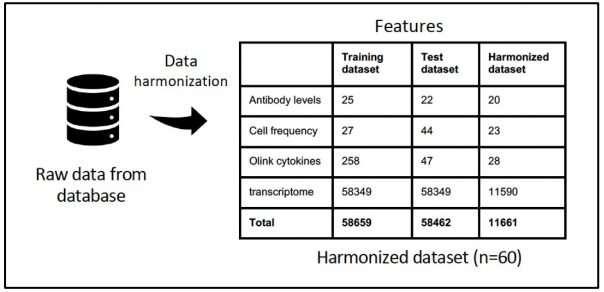
THE PREPRINT SERVER FOR BIOLOGY

0 0 0 0 0 0 3

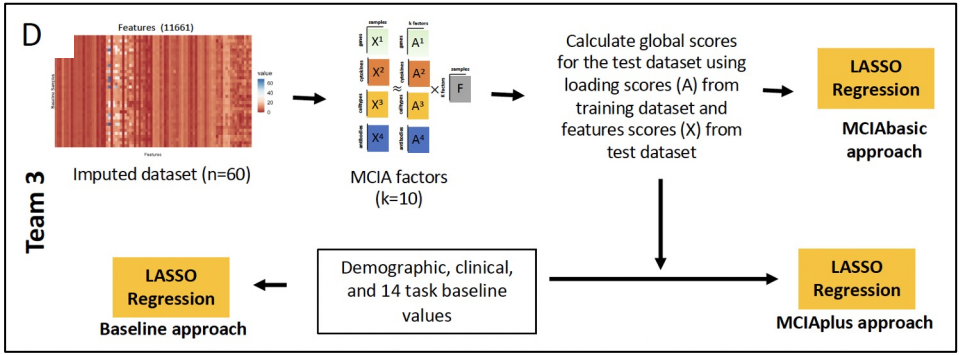
- **32 models** were developed during the first challenge by three teams
- **Team 1:** Establishing baseline prediction models from the systems vaccinology literature
- **Team 2:** Establishing purpose-built models using Joint and Individual Variation Explained (JIVE)
- **Team 3:** Establishing purpose-built models using Multiple Co-Inertia Analysis (MCIA)

a. Establishing models using the MCIA approach (First challenge)

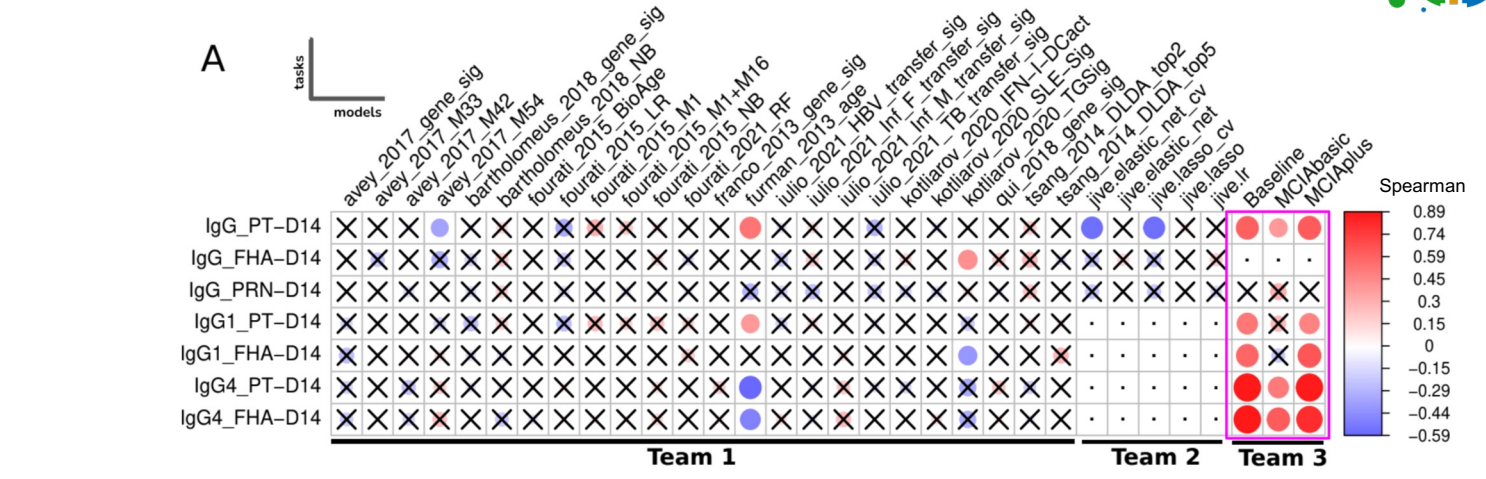
- **Step 1:** Data imputation is key step in building MCIA models
  - Creation of harmonized dataset
  - Imputation of missing data in the baseline training set using Multiple Imputation by Chained Equations (MICE) algorithm
- **Step 2:** Build three models using MCIA approaches:
  - **Baseline approach** - Clinical features (age, infancy vaccination, biological sex) and baseline task values
  - **MCIAbasic approach** - Only MCIA factors
  - **MCIAplus approach -** Features from first two models

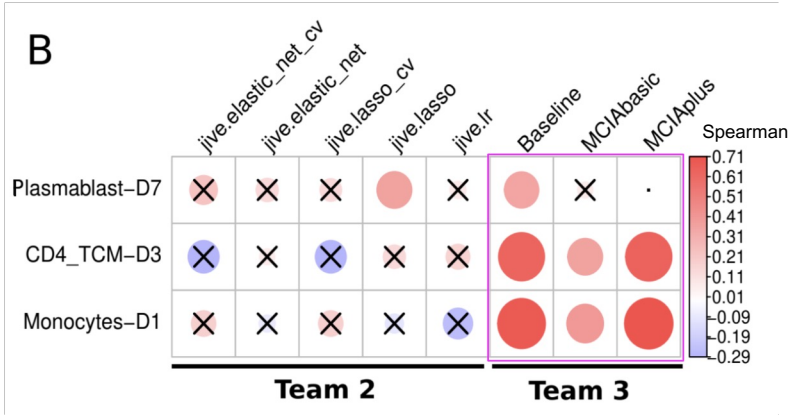


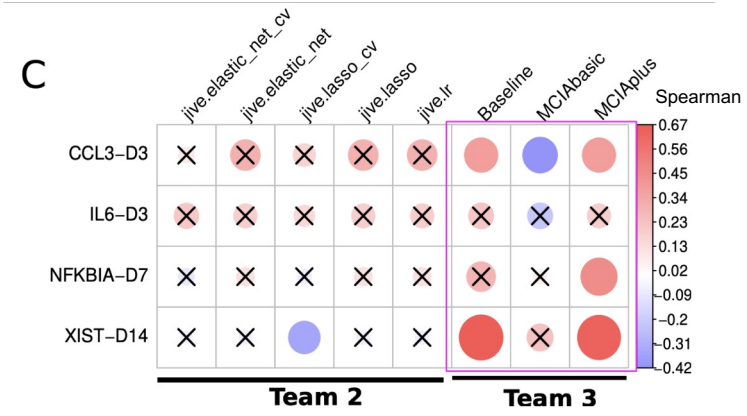
	Training dataset	Test dataset	Harmonized dataset
Antibody levels	25	22	20
Cell frequency	27	44	23
Olink cytokines	258	47	28
transcriptome	58349	58349	11590
Total	**58659**	**58462**	**11661**



Evaluation of the MCIA models for the first CMI-PB challenge





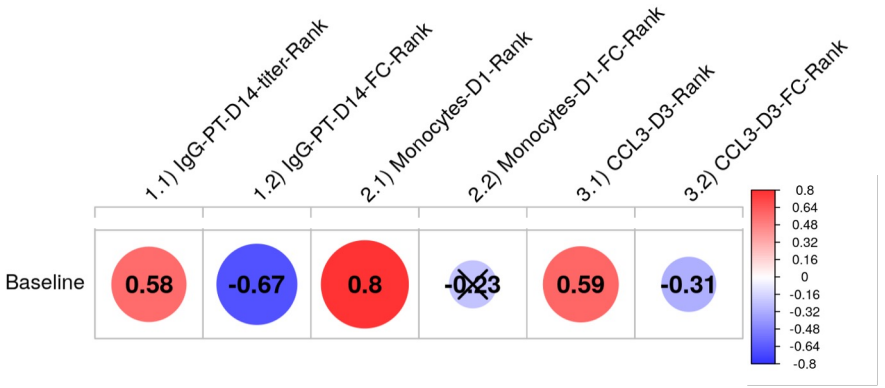


Note: The first challenge consisted of 14 tasks: 7 Ab titer, 3 cell frequency and 4 gene expression tasks.

b. Establishing baseline response of task variables to predict post-vaccination response

- **Goal:** Illustrate the processes of data access, model building, and model submissions for users.
- **Datasets used:**
  - Training dataset (2020 + 2021)
- We calculated the Spearman correlation using the baseline and response values of the task variable.
  - For instance, we used the values of CCL3 from Day 0 to predict the response on Day 3.

Corr(CCL3_D0, CCL3_D3)



The correlation plot is generated by the training dataset.

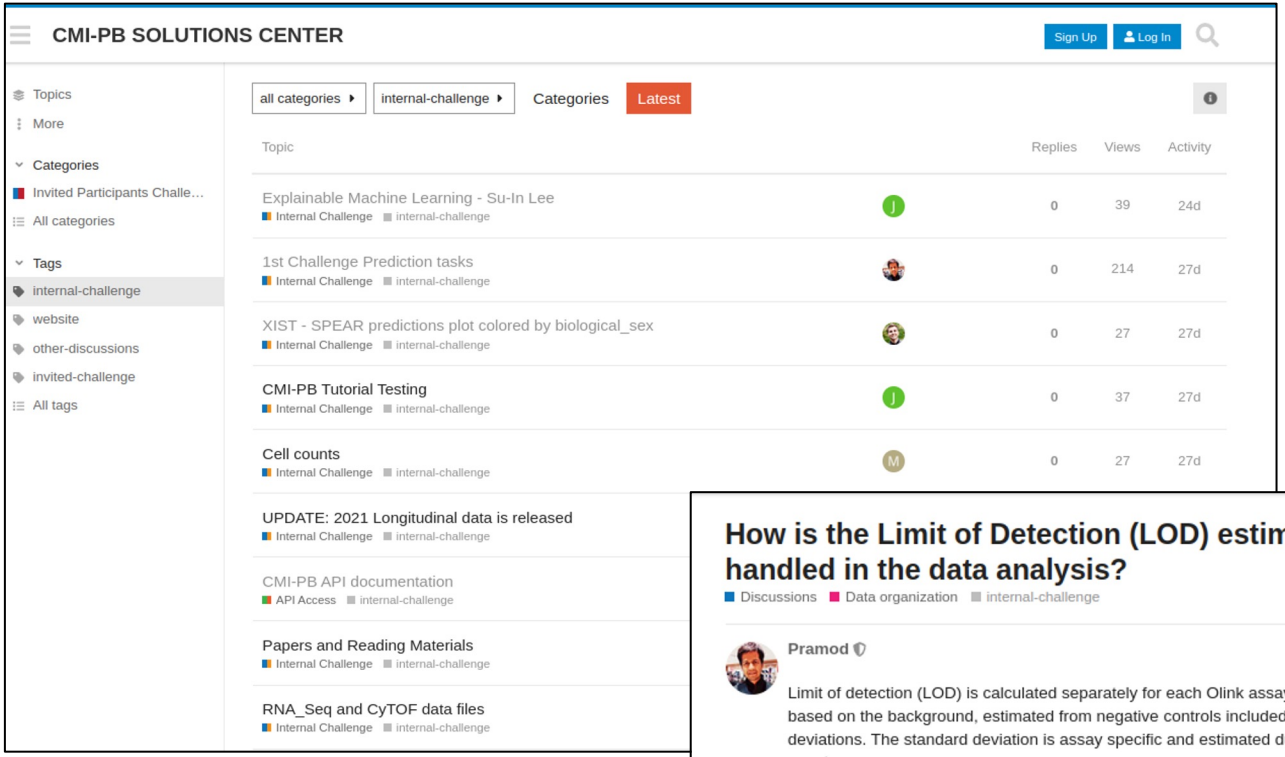
Codebase: https://rpubs.com/Rasteh/1080854

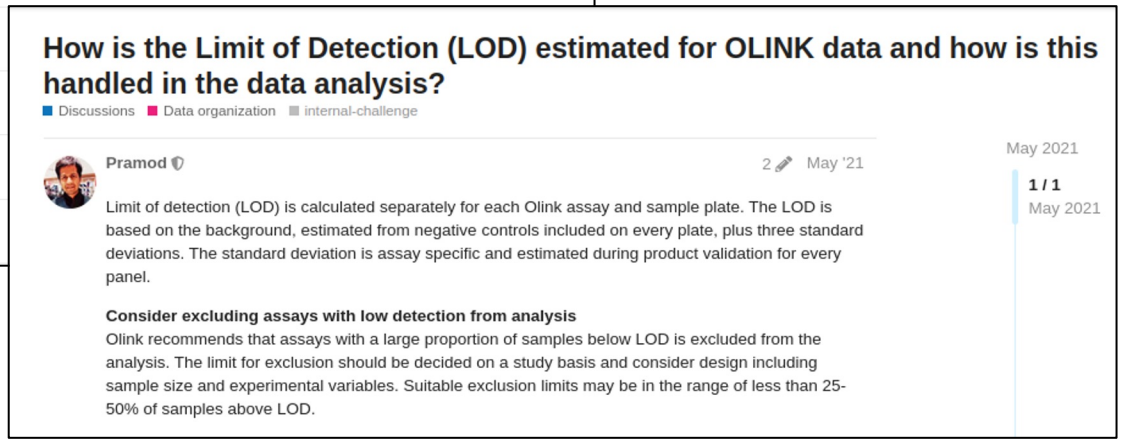
Presentation Agenda

1. Introduction-Project Overview
2. The CMI-PB Challenge
3. Example Model Presentation
4. **Resources Overview**
5. Q & A

4. Summary of other resources on the site

CMI-PB Solutions center

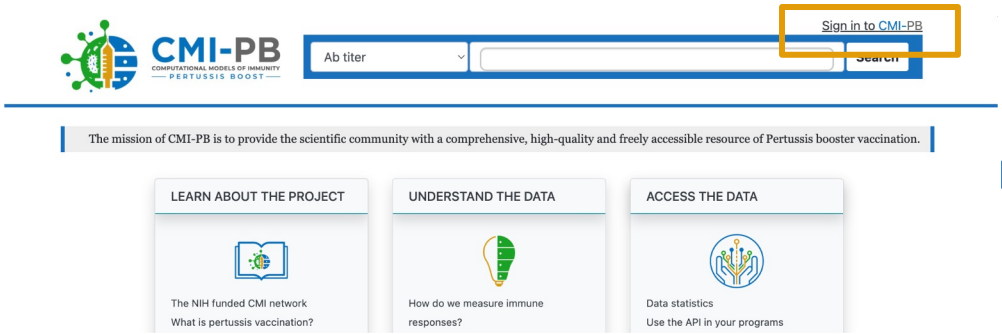




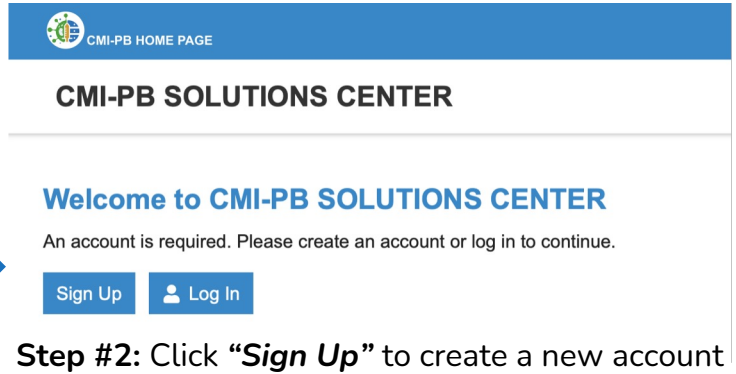
- Platform for knowledge sharing and discussion
- An account is required to post or respond to threads

https://discuss.cmi-pb.org/

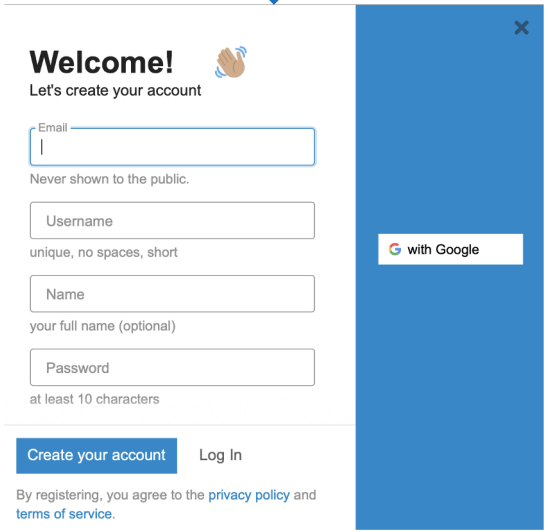
Creating an account



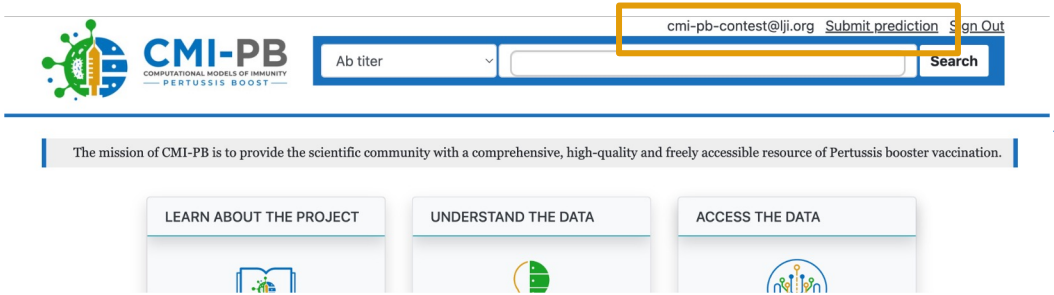
Step #1: Click ***"Sign in to CMI-PB"*** in the upper right hand corner



Step #2: Click ***"Sign Up"*** to create a new account

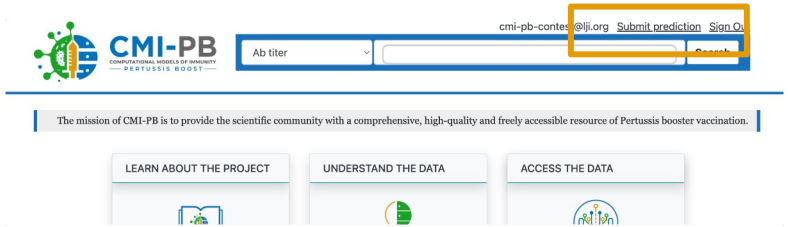


Step #3: Fill out registration form or use SSO with Google

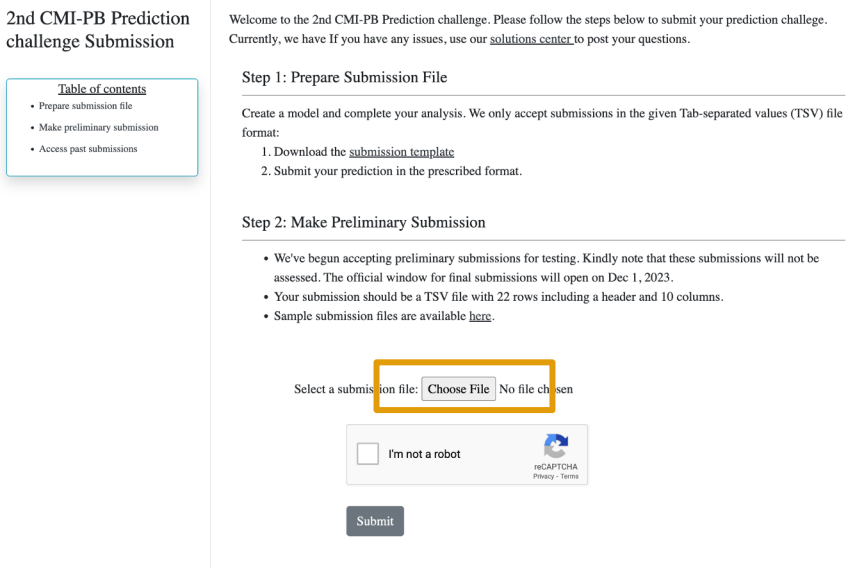


Step #4: Confirm that when you are signed in, your email is shown in the upper right hand corner

Creating a submission



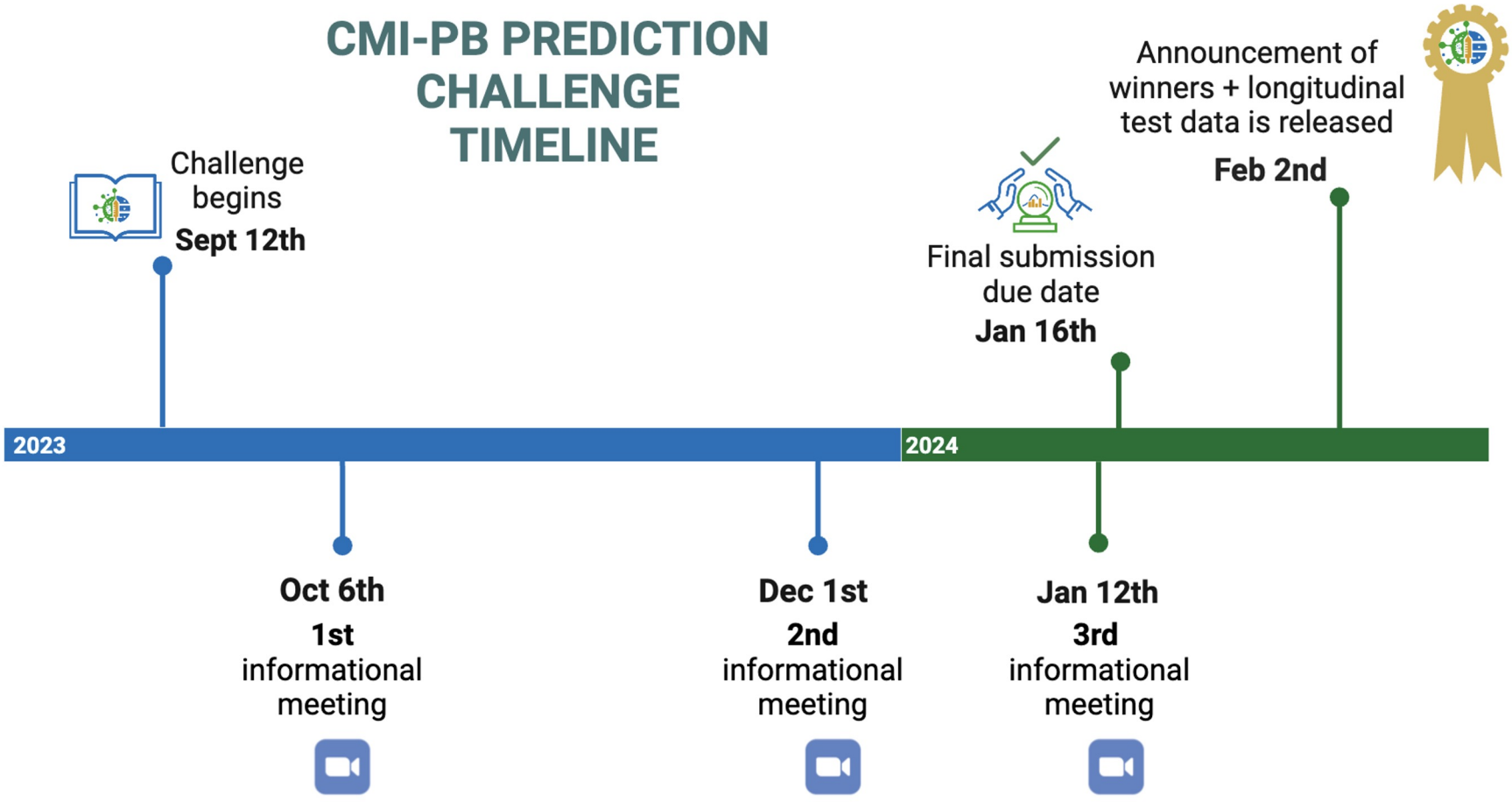
Step #1: Once logged in, click ***"Submit Prediction"*** in the upper right hand corner



Step #2: Follow all steps, click the ***"Choose File"*** button, and make your submission

CMI-PB
COMPUTATIONAL MODELS OF IMMUNITY
PERTUSSIS BOOST

CMI-PB PREDICTION CHALLENGE TIMELINE



The CMI-PB team

CMI-PB
COMPUTATIONAL MODELS OF IMMUNITY
PERTUSSIS BOOST

Kleinstein Lab (Yale)

- Expertise: A combination of "big data" analysis and immunology domain.
- Collaborating on data and models being released to the community to support reproducibility and the prediction contest, and also participate in the prediction challenge.

Steven Kleinstein
Jeremy Gygi
Leying Guan
Anna Konstorum

Grant Lab (UCSD)

- Expertise: the use of computational approaches, based on both biophysics and bioinformatics, to study the structure, function and evolution of key biological macromolecules.
- Dr. Grant will engage and advise over 40 biology graduate students in the CMI-PB Prediction Challenge.

Barry Grant

Ay Lab (LJI)

- Expertise: Development of bioinformatics tools that utilize high-dimensional and high-throughput datasets to deduce insights into chromatin conformation, genetic variation, and the regulation of gene expression.
- The Ay lab is focused on developing predictive machine learning models, which will serve as examples and baselines for participants in the CMI-PB challenge.

Ferhat Ay
Joaquin Reyna

Peters Lab (LJI)

- Expertise: Both experimental and computational studies to better understand human immune responses in the context of infectious diseases, allergy, cancer and vaccines.
- The team is responsible for the generation of experimental data, making it accessible in a central and standardized fashion, and coordinating the creation and coordination of the prediction contest.

Bjoern Peters
Jason Greenbaum
James Overton
Brendan Ha
Pramod Shinde
Mari Kojima
Rasteh Haji Kazem Nili
Jiyeun Lee
Lisa Willemsen
Shelby Orfield

And thank you to the Sette Lab, Crotty lab, LJI Clinical Core, LJI Bioinformatics Core

The CMI-PB team members


Bjoern Peters


Steven Kleinstein


Ferhat Ay


Barry Grant


Shane Crotty


Alessandro Sette


Pramod Shinde


Shelby Orfield


Lisa Willemsen


Leying Guan


Joaquin Reyna


Mari Kojima


Ferran Soldevila


Rasteh Nili


Jason Greenbaum


Brendan Ha


Jiyeun Lee

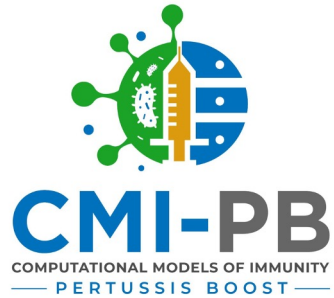

Ricardo De Silva Antunes


Jeremy Gygi


Anna Konstorum

Presentation Agenda

1. Introduction- Project Overview
2. The CMI-PB Challenge
3. Example Model Presentation
4. Resources Overview
5. **Q & A**



Questions?

Please email cmi-pb-challenge@lji.org at any time throughout the challenge for any questions.

We are excited to see you again for: