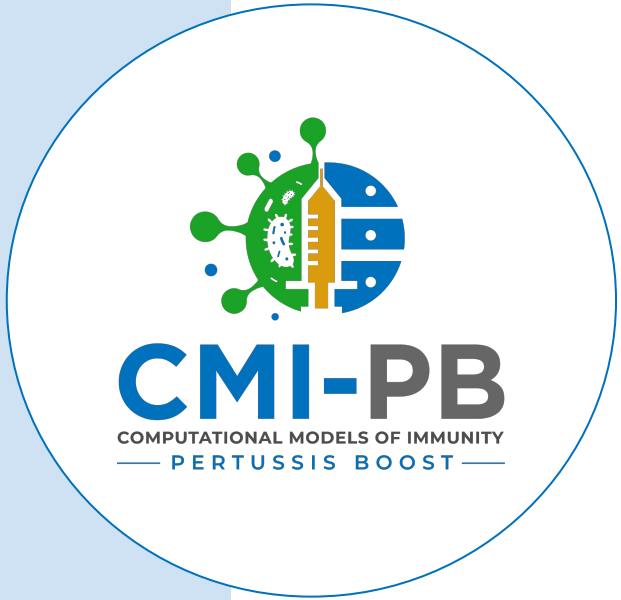
3rd CMI-PB Prediction Challenge

1st Informational Session
September 6th, 2024

La Jolla Institute for Immunology



Background and Benefits of Prediction Challenges

- ✔ 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

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 (Whooping Cough)

1 Cause

Bordetella Pertussis

bacteria in the upper respiratory system

2 Symptoms

Weeks 1-2

- low fever
- runny nose
- mild cough
- apnea (in infants)

Weeks 3-8

- coughing fits ending in a 'whoop'
- vomiting
- exhaustion

3 Prevention

DTaP vaccine injection during childhood plus Tdap adult booster

4 Epidemiology

spread via coughing, sneezing, and other close contact

195,000 annual deaths worldwide	**48,500,000** people infected annually worldwide
3+ Months avg. length of infection	**5.5** avg. number of people an infected person will infect when contagious

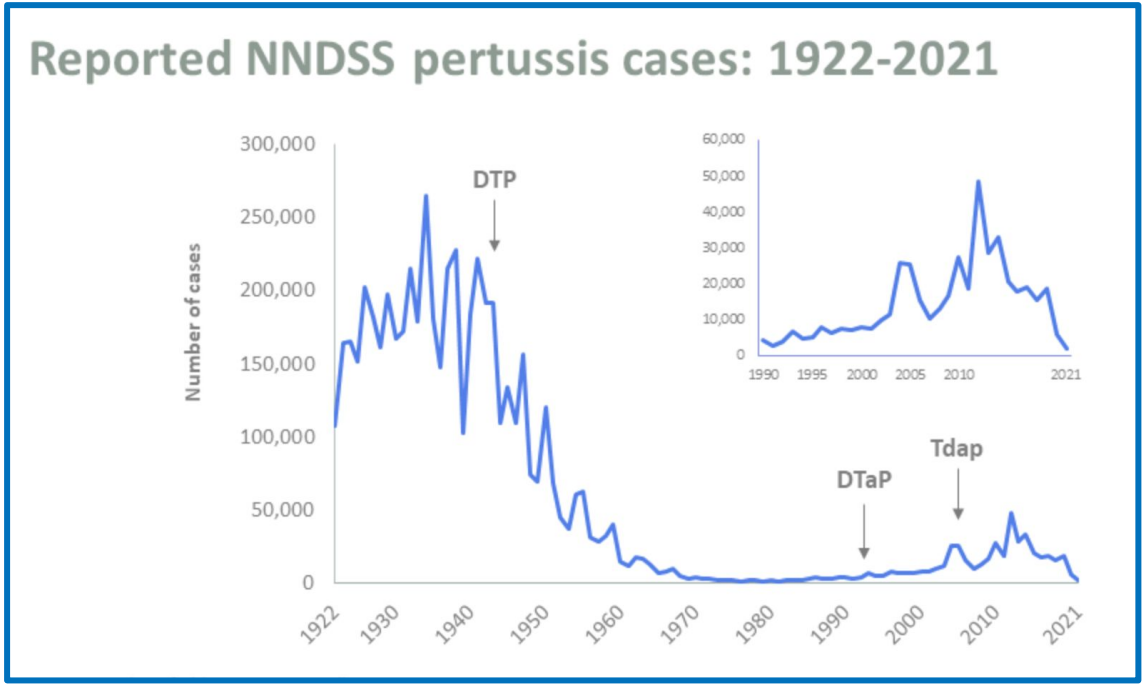
5 Complications

- earache
- pneumonia
- encephalopathy
- seizures
- cerebral hypoxia
- fractured ribs
- failure to thrive
- death

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

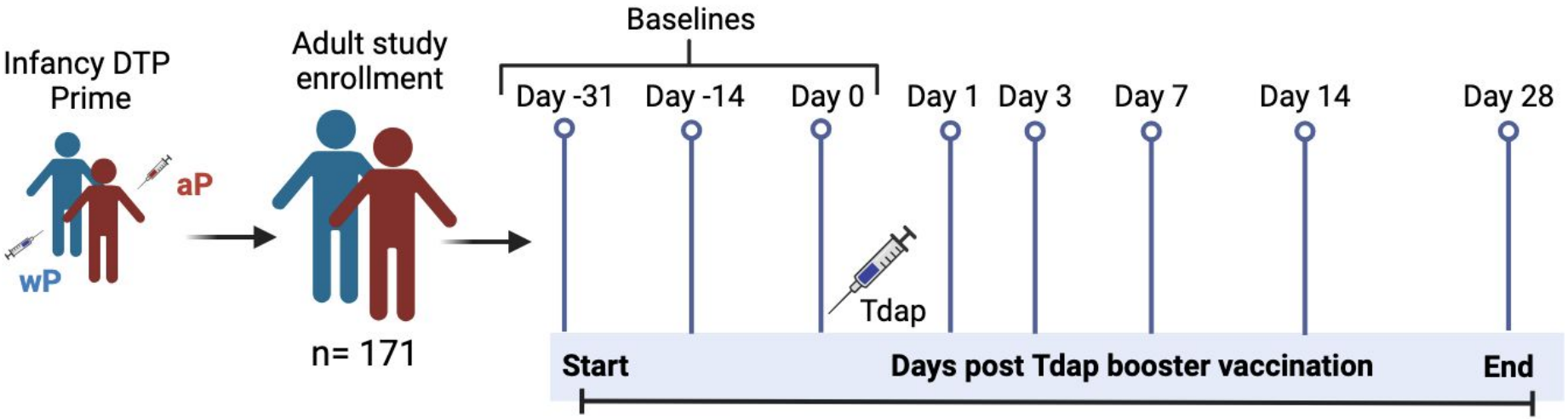
Waning Immunity from aP pertussis vaccination

- 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 & Legendplex

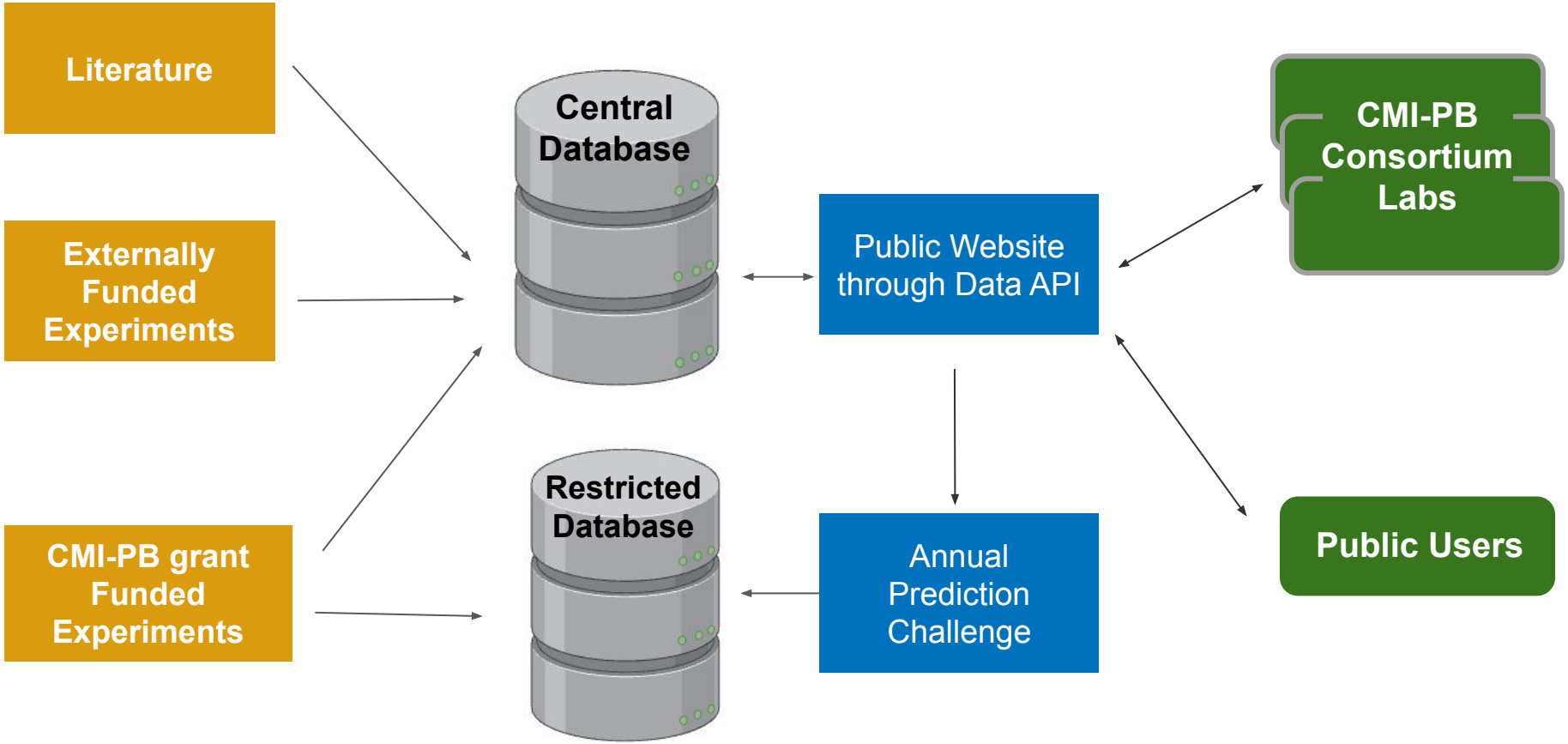
- Concentration of 48 cytokines

PBMC Transcriptomics by bulk RNA-Seq

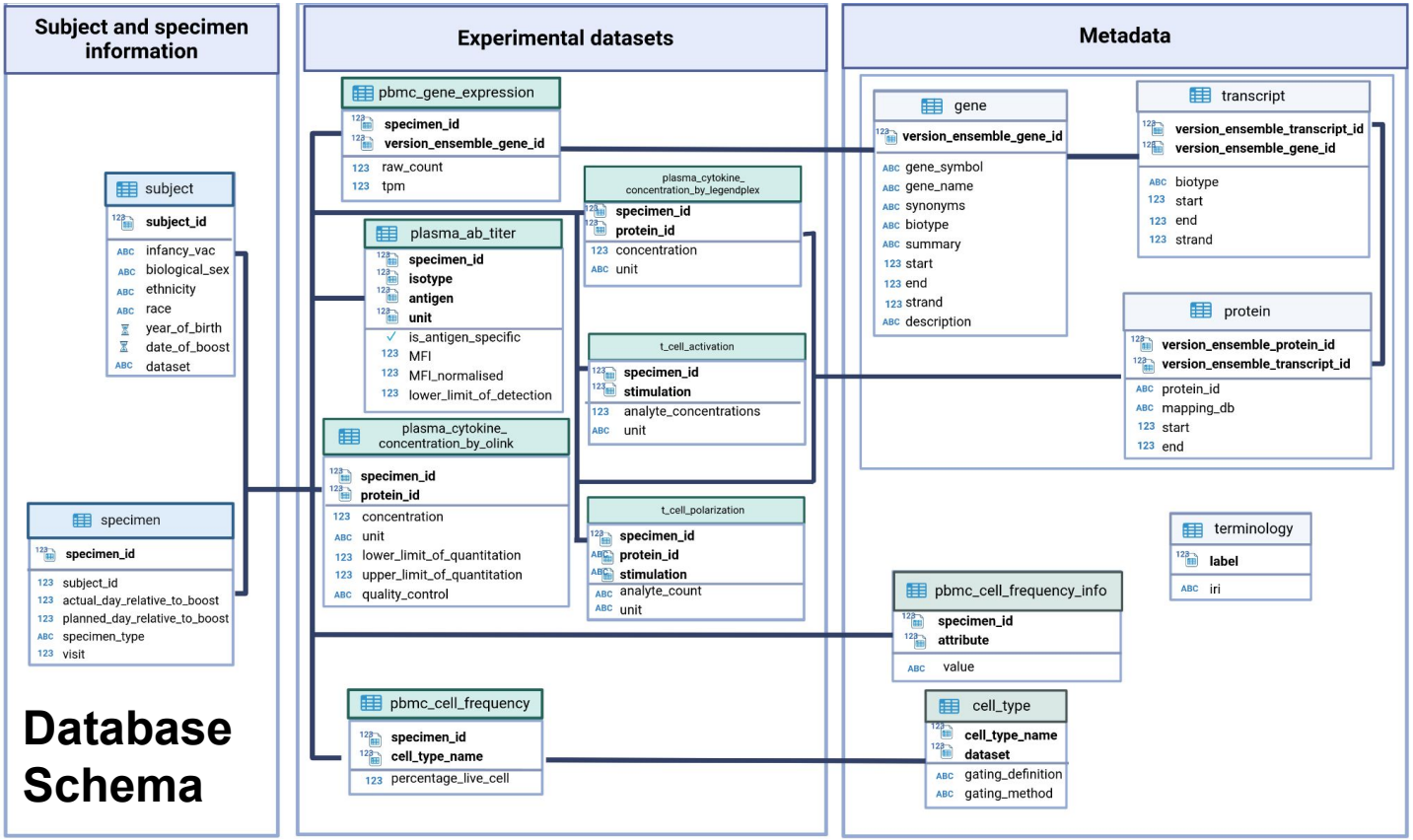
Antigen-specific T cell responses by AIM assay & FluoroSpot

- Magnitude of $CD4^+$ T cell activation (% $OX40^+CD25^+$ $CD4^+$ T cells)
- Cytokine polarization (IFN-γ, IL-5 and IL-17)

Databases, Model building, and Prediction Challenge



Providing access to experimental data in a standardized format



Find more information on the website [here]

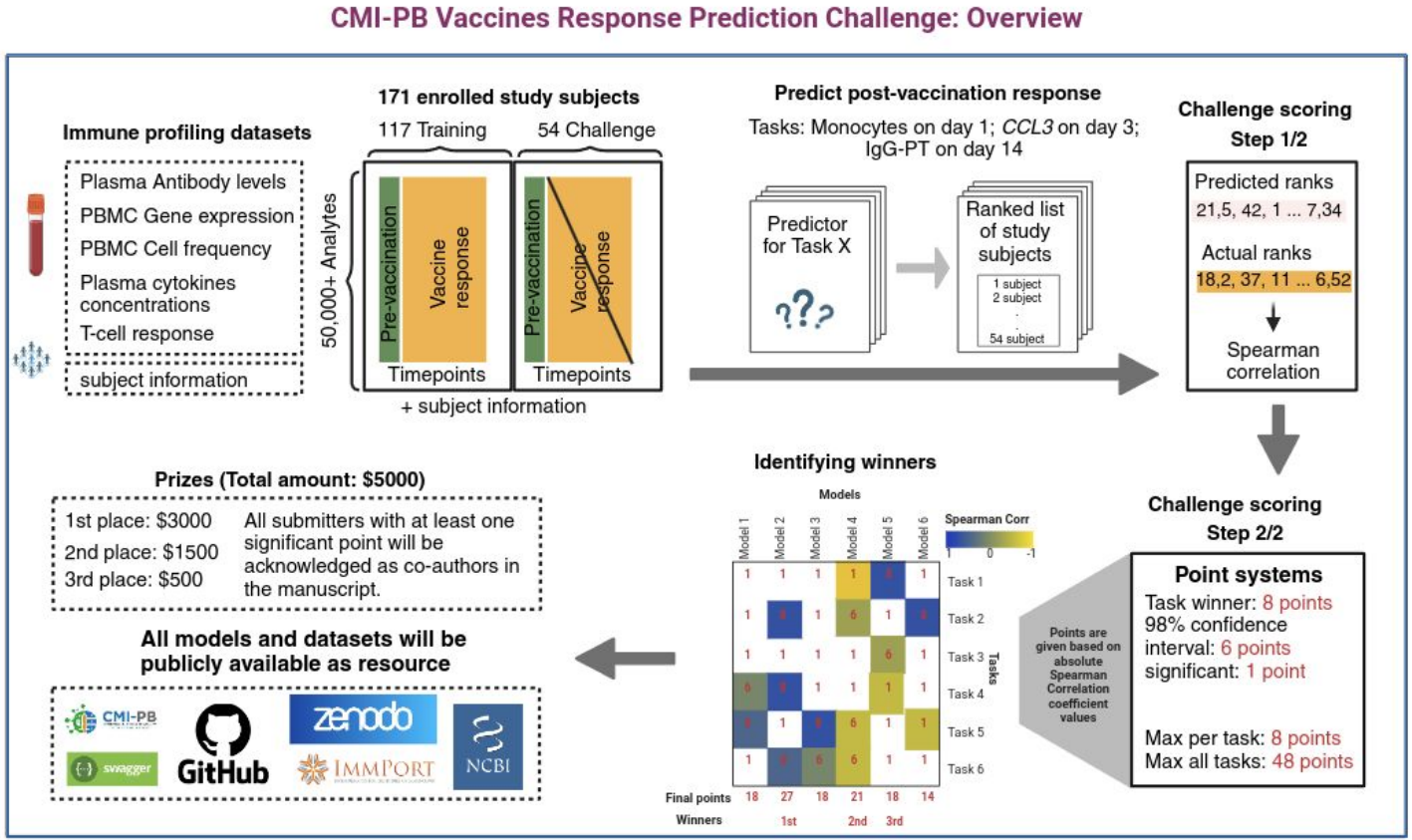
Presentation Agenda

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

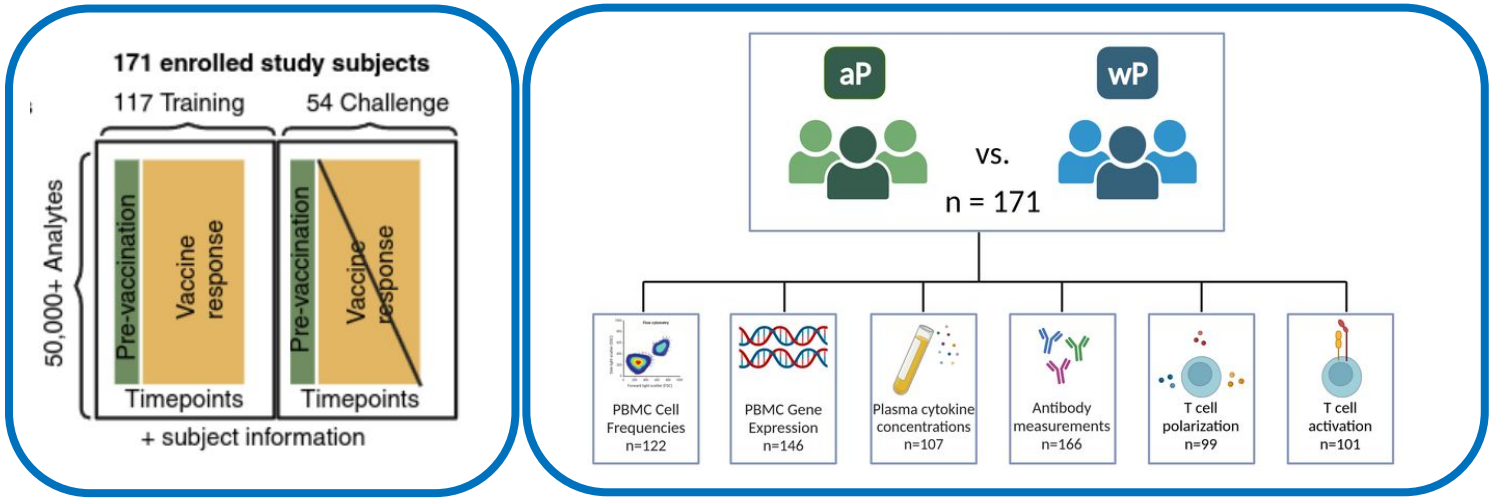
A. Past and current CMI-PB annual prediction challenges

	Annual prediction challenge title	Contestants	Number of subjects		Current status
			Training dataset	Challenge 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)	21 (11 aP + 10 wP)	Concluded in January 2024
3	**Third Challenge: Open Challenge**	**Public**	**117 (58 aP + 59 wP)**	**54 (27 aP + 27 wP)**	**Announced on August 27 2024**

B. Prediction challenge outline



C. Overview of the CMI-PB Challenge data



The data is split into two groups:

- **Training dataset (2020, 2021, 2022)**: Used to build models, including known outcomes ("ground truth"). Features are based on multi-omics readouts and demographic data, with potential for feature engineering.
- **Challenge dataset (2023)**: Used to evaluate model performance on unseen data. The task is to predict vaccine response outcomes without provided ground truth.

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

C. CMI-PB Challenge Data: Raw and Processed Data

- The dataset comprises three multi-omics datasets (from 2020, 2021, and 2022) and the challenge dataset (2023).
- The data require careful processing and normalization to generate computable matrices suitable for model development.
- While data processing and normalization approaches can vary depending on user preferences, the CMI-PB team has provided a standardized data processing method inspired by the approach used in the 2nd CMI-PB challenge.
- Pipeline: RPub and GitHub

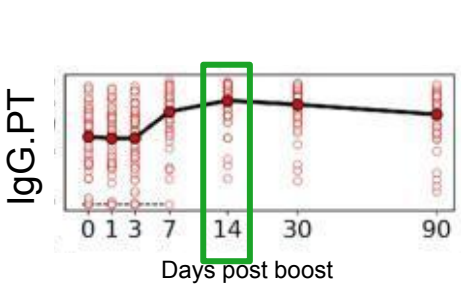
Data Access: https://www.cmi-pb.org/downloads/cmipb_challenge_datasets/current/3rd_challenge/

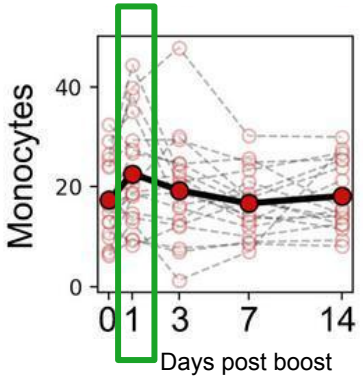
D. Prediction tasks for CMI-PB Challenge

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

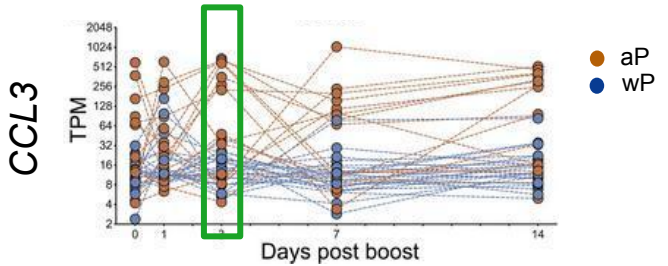
> **A system-view of *Bordetella pertussis* booster vaccine responses in adults primed with whole-cell versus acellular vaccine in infancy**
>
> Ricardo da Silva Antunes, … , Alessandro Sette, Bjoern Peters
>
> *JCI Insight*. 2021;6(7):e141023. https://doi.org/10.1172/jci.insight.141023.

- **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



D. 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
vaccination compared to gene expression values at

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

SubjectID	BaselineVal	Post-VacVal	FoldChangeVal	RankPost-VacVa	RankFoldChangeVal
119	2.9	6.7	2.31	2	3
120	5.1	8.7	1.71	**1**	4
121	1.2	3.5	2.92	6	**1**
122	4.5	5.1	1.13	4	5
123	4.7	4.8	1.02	5	6
124	2.7	6.5	2.41	3	2

E. Submission File

SubjectID	Age	BiologicalSexAtBirth	VaccinePrimingStatus	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
119	23	Female	aP	32	43	10	52	9	28
120	27	Female	wP	33	37	13	21	44	25
121	22	Female	aP	14	2	32	15	7	34
122	23	Female	aP	26	19	9	45	37	6
123	26	Female	wP	22	54	46	49	2	15
124	22	Male	aP	35	25	53	44	16	10
125	29	Male	wP	36	17	38	23	36	43
126	29	Male	wP	7	23	23	20	6	21
127	26	Female	aP	28	50	2	47	26	24
128	28	Female	wP	48	46	21	2	20	50
129	31	Male	wP	53	52	4	31	3	14
130	26	Male	wP	29	8	39	8	45	33
131	24	Female	aP	43	44	1	29	34	49
132	27	Male	wP	45	5	19	4	40	17
133	25	Female	aP	17	33	37	10	47	20
134	32	Male	wP	31	41	41	16	22	46
135	27	Male	wP	50	51	16	19	28	13
136	27	Female	wP	30	7	47	1	50	54
137	24	Female	aP	39	9	28	38	52	37
138	22	Male	aP	16	10	24	51	25	8
139	29	Female	wP	24	21	3	28	38	22

A submission file with entries for **all tasks**

A submission file with entries for **two tasks**

SubjectID	Age	BiologicalSexAtBirth	VaccinePrimingStatus	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
119	23	Female	aP	17				19	
120	27	Female	wP	46				43	
121	22	Female	aP	10				11	
122	23	Female	aP	30				6	
123	26	Female	wP	44				30	
124	22	Male	aP	20				54	
125	29	Male	wP	13				41	
126	29	Male	wP	11				10	
127	26	Female	aP	25				51	
128	28	Female	wP	53				4	
129	31	Male	wP	36				8	
130	26	Male	wP	41				29	
131	24	Female	aP	34				39	
132	27	Male	wP	22				9	
133	25	Female	aP	21				40	
134	32	Male	wP	7				53	
135	27	Male	wP	51				24	
136	27	Female	wP	4				52	
137	24	Female	aP	32				18	
138	22	Male	aP	23				13	
139	29	Female	wP	50				37	

Populated with random numbers.
Submission file template is available on website.

F. Model Evaluation and Scoring

Challenge scoring
Step 1/2

Predicted ranks*
21,5, 42, 1 ... 7,34

Actual ranks
18,2, 37, 11 ... 6,52

↓

Spearman correlation

* If you have a blank cell (or put NA) in your submission, we will replace that cell with the median rank of that list

Challenge scoring
Step 2/2

Point systems

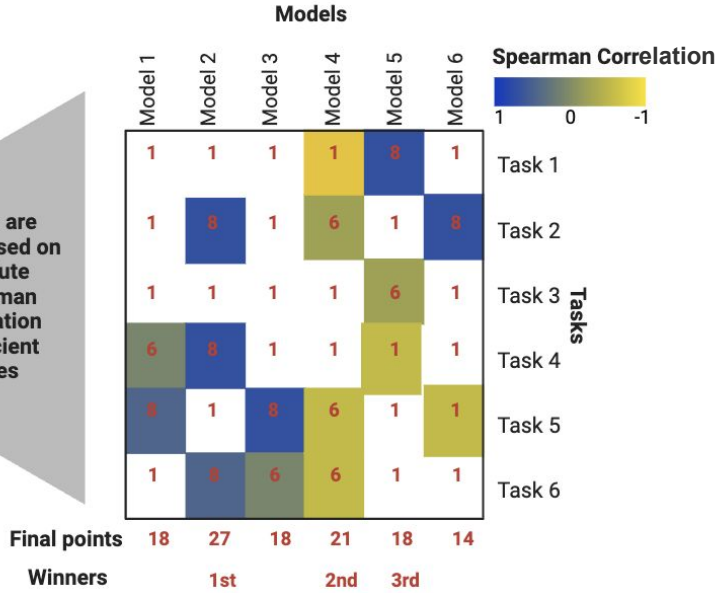
Task winner: 8 points
98% confidence interval: 6 points
significant: 1 point

Max per task: 8 points
Max all tasks: 48 points

* You cannot receive a combination of these points (ie. 8 + 1)

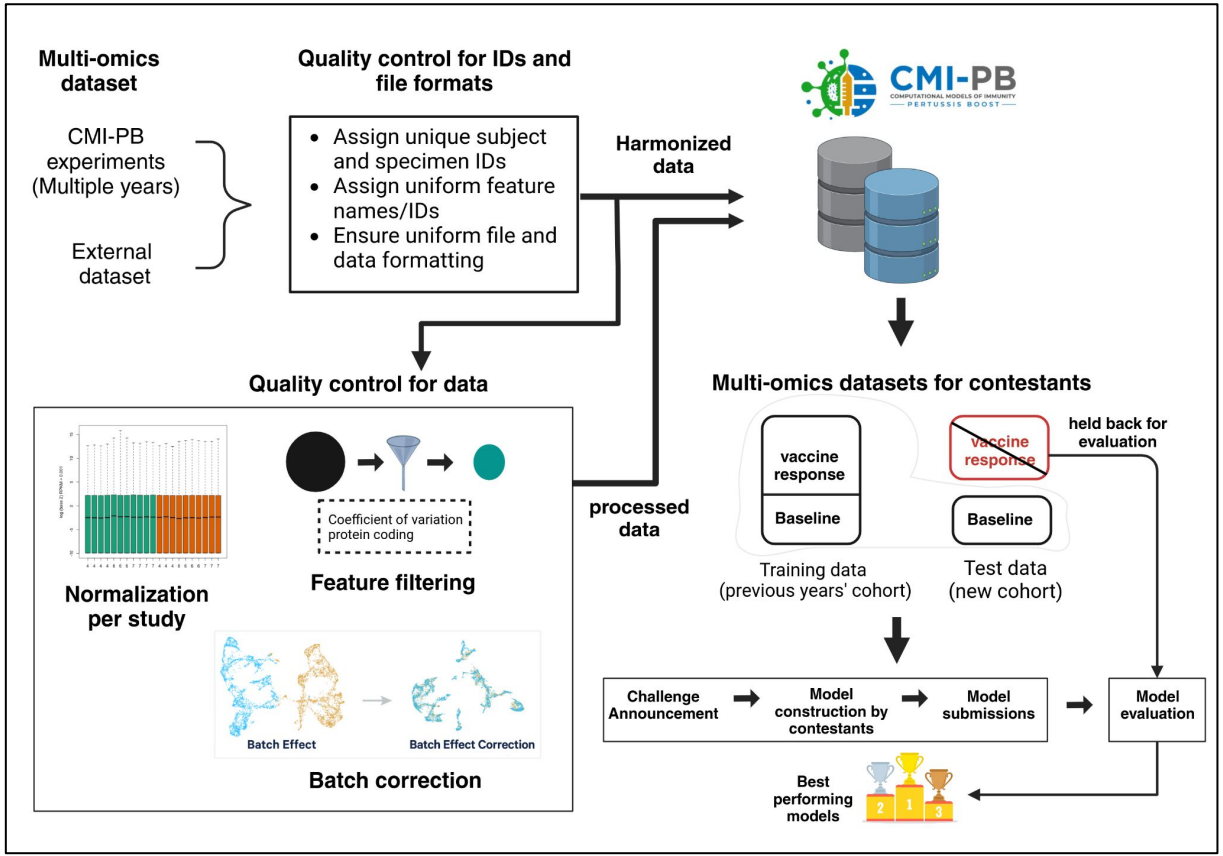
Identifying winners

Points are given based on absolute Spearman Correlation coefficient values



Find the detailed information regarding Prizes on the website here.

d. Overview of the CMI-PB Challenge data



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

F. Evaluation: Prizes & Awards

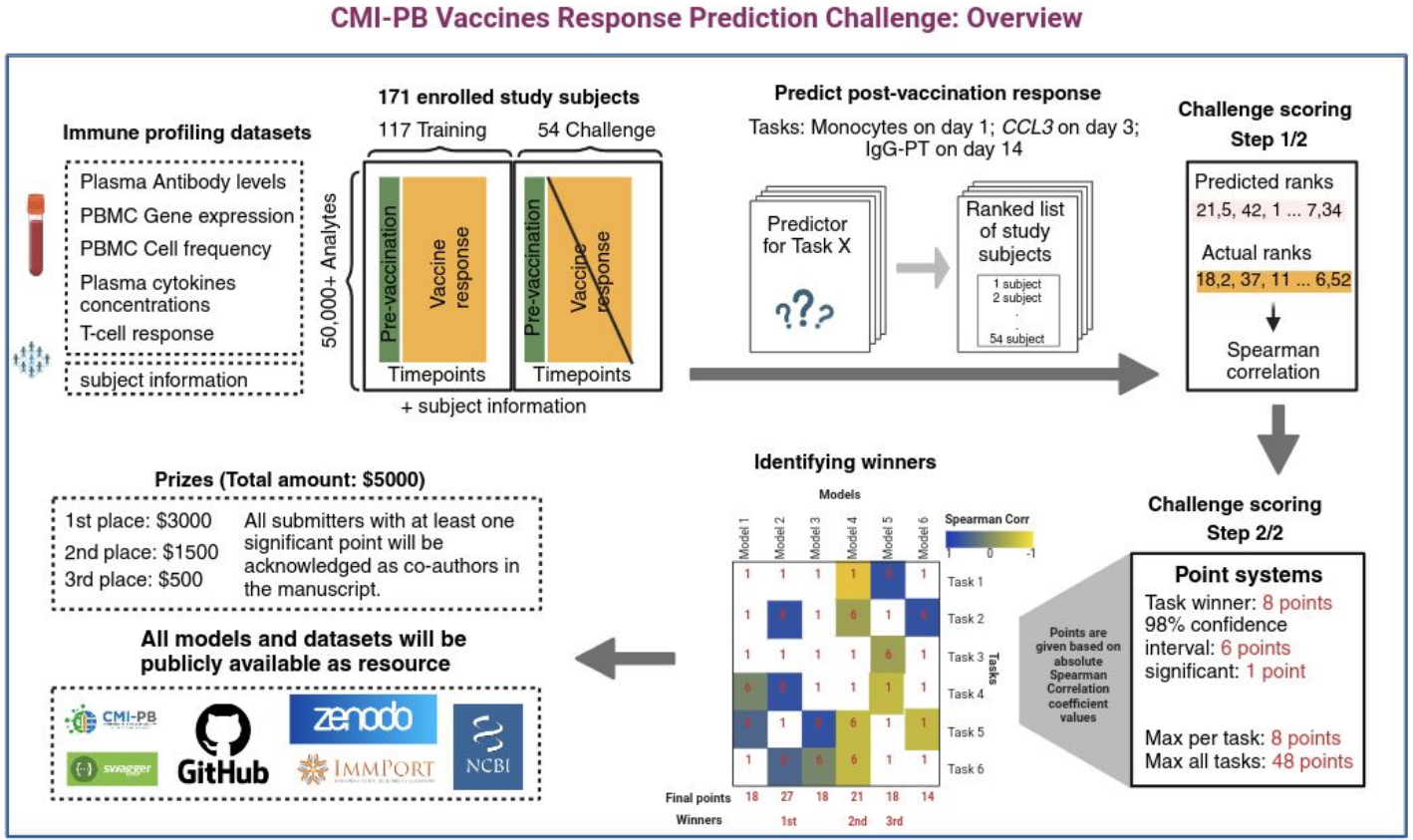
- All submitters with at least one significant score will be acknowledged as co-authors in the manuscript written following this challenge.
- Additionally, prizes will be awarded to the top three teams with the best-performing models. The total money prize is $5000.
  - **1st place: $3000**
  - **2nd place: $1500**
  - **3rd place: $500**

Prize Details:

- If there is a tie for first place, the prizes for first and second place will be combined and split equally among the winning teams. There will be no separate second-place prize awarded, and the team originally in second-place will receive the third place prize. The team originally in third place will not receive a prize.
- In the event of multiple teams tying for first place, the prizes for first, second, and third place will be combined and split equally among all the winning first-place teams. No separate second or third-place prizes will be awarded.
- If there is a tie for second place, the prizes for second and third place will be combined and split equally between the tied teams. The original third-place team will not receive a prize. This applies if there are multiple teams tied for second place.
- In the case of a tie for third place, only the prize for third place will be shared equally between the tied teams.

Find more information on the website here.

G. Prediction challenge outline



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

Cell Reports Methods Open access Submit

This journal Journals Publish News & events About Search for...

RESOURCE · Volume 4, Issue 3, 100731, March 25, 2024 · *Open Access*

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

Pramod Shinde [1,10] · Ferran Soldevila [1,10] · Joaquin Reyna [1,2,10] · … · Ferhat Ay [1,6] · Steven H. Kleinstein [5,8] · Bjoern Peters [1,6,11] … Show more

Affiliations & Notes Article Info

Download Full Issue

Download PDF Cite Share Set Alert Get Rights Reprints

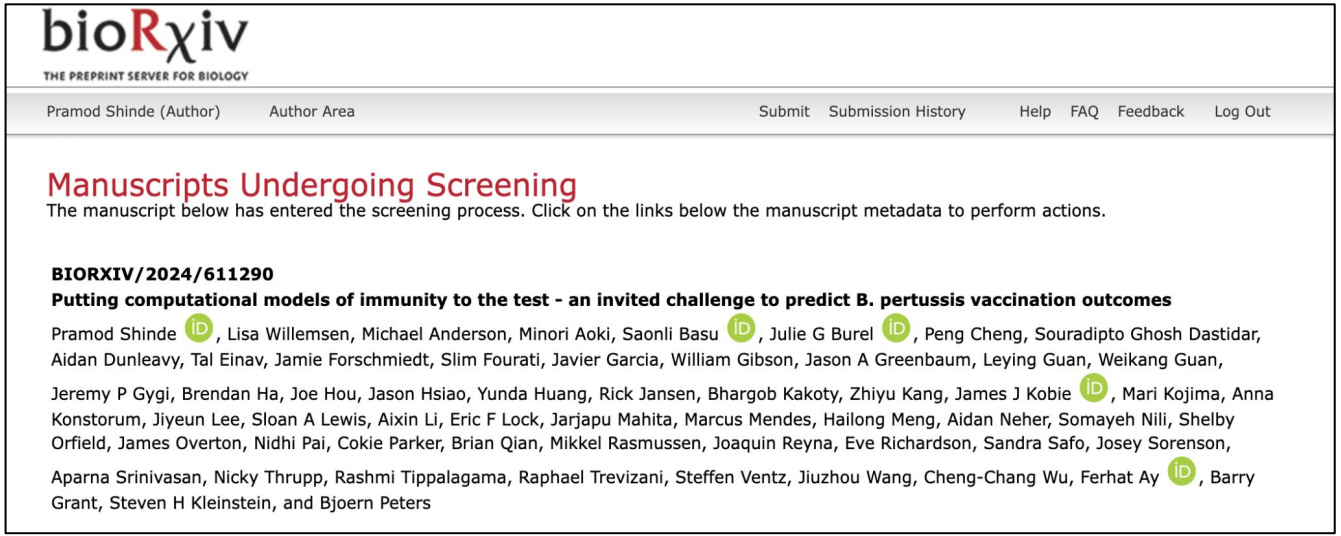
Previous article Next article

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)

Link to paper here

B. Models from Second (invited) challenge



bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

Pramod Shinde (Author) Author Area Submit Submission History Help FAQ Feedback Log Out

Manuscripts Undergoing Screening

The manuscript below has entered the screening process. Click on the links below the manuscript metadata to perform actions.

BIORXIV/2024/611290

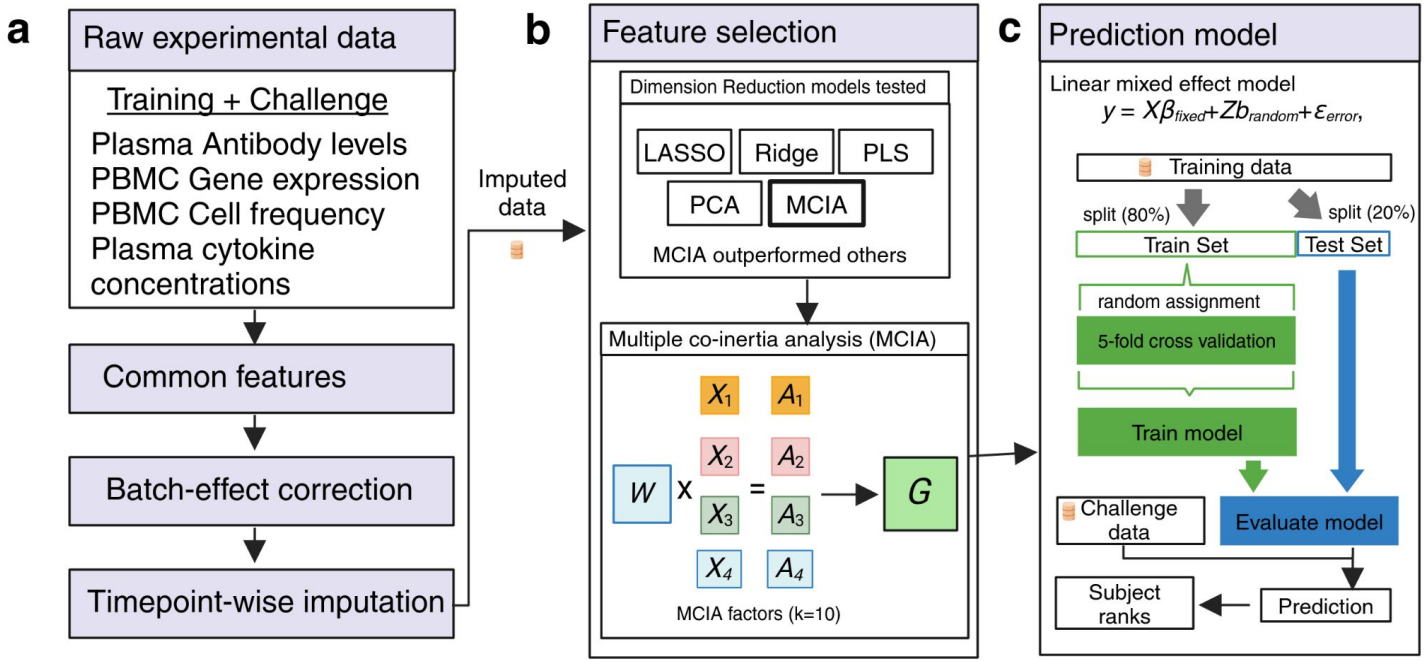
Putting computational models of immunity to the test - an invited challenge to predict B. pertussis vaccination outcomes

Pramod Shinde, Lisa Willemsen, Michael Anderson, Minori Aoki, Saonli Basu, Julie G Burel, Peng Cheng, Souradipto Ghosh Dastidar, Aidan Dunleavy, Tal Einav, Jamie Forschmiedt, Slim Fourati, Javier Garcia, William Gibson, Jason A Greenbaum, Leying Guan, Weikang Guan, Jeremy P Gygi, Brendan Ha, Joe Hou, Jason Hsiao, Yunda Huang, Rick Jansen, Bhargob Kakoty, Zhiyu Kang, James J Kobie, Mari Kojima, Anna Konstorum, Jiyeun Lee, Sloan A Lewis, Aixin Li, Eric F Lock, Jarjapu Mahita, Marcus Mendes, Hailong Meng, Aidan Neher, Somayeh Nili, Shelby Orfield, James Overton, Nidhi Pai, Cokie Parker, Brian Qian, Mikkel Rasmussen, Joaquin Reyna, Eve Richardson, Sandra Safo, Josey Sorenson, Aparna Srinivasan, Nicky Thrupp, Rashmi Tippalagama, Raphael Trevizani, Steffen Ventz, Jiuzhou Wang, Cheng-Chang Wu, Ferhat Ay, Barry Grant, Steven H Kleinstein, and Bjoern Peters

25 submissions were made total for the second challenge

- **1st Place:** Team from University of Minnesota, achieving superior predictions in tasks 1.2 (r = 0.7, p-value = 0.001) and 2.1 (r = 0.81, p-value = 0.0031)
- **2nd Place:** Team from LJI (Teams 54 and 38) ranked second overall.
- **3rd Place:** Team from National Institutes of Health (Team 51) ranked third overall and achieved the top rank for task 2.2

C. Winning model from 2nd Challenge

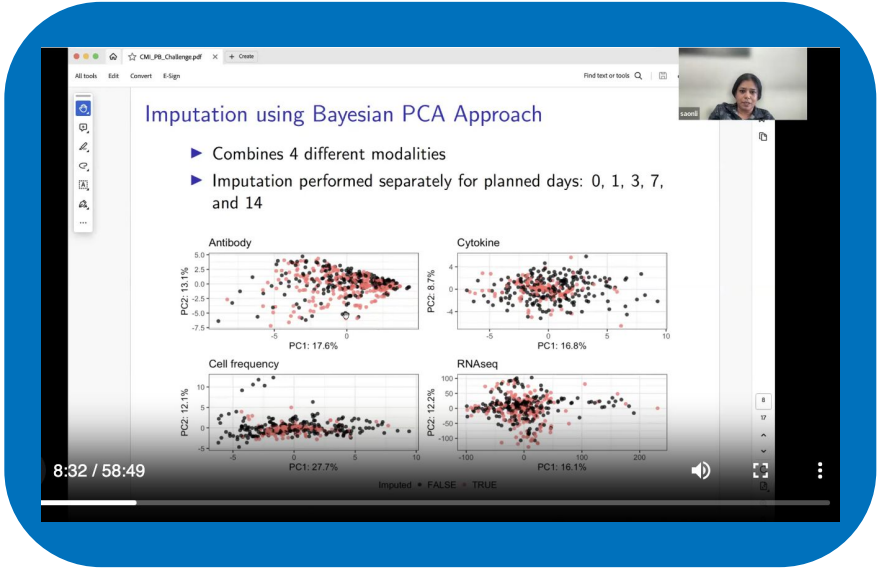


From University of Minnesota Group (Saonli Basu & Team)

Dimension reduction through Multiple Co-inertia analysis and modeled with Linear mixed effect

D. Models from Second (invited) challenge

Debrief session where top 3 winners from the 2nd Challenge presented on Zoom:

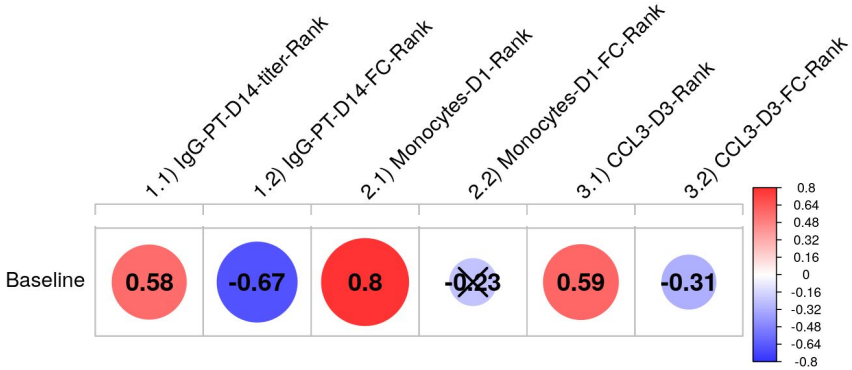


Recording and slides Can be found on our discourse here:
https://discuss.cmi-pb.org/c/2nd-challenge/20

e. 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+ 2022)
- 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)



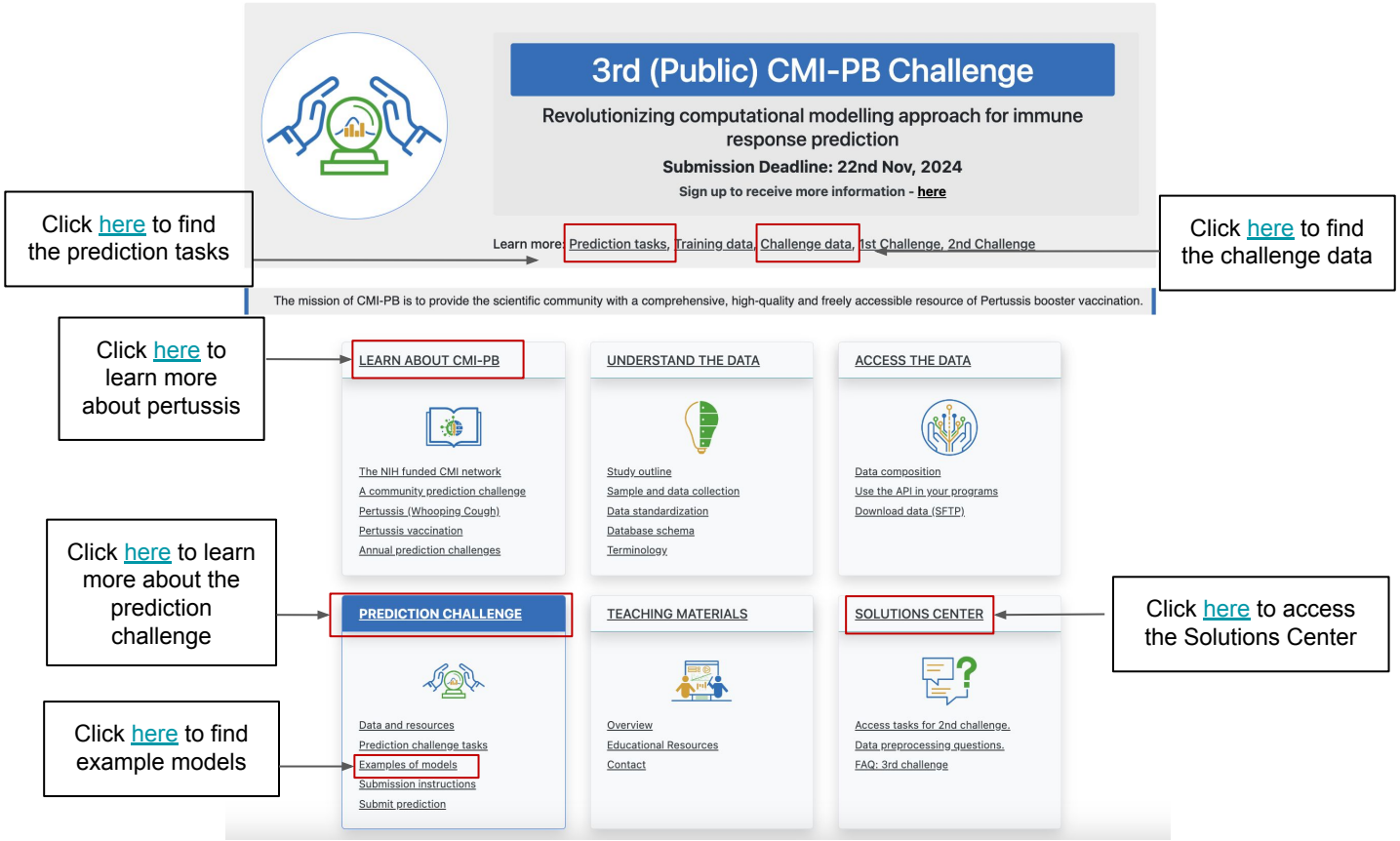
The correlation plot is generated by the training dataset.

Codebase: Rpubs

Presentation Agenda

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

Summary of other resources on the site



Click here to find the prediction tasks

Click here to find the challenge data

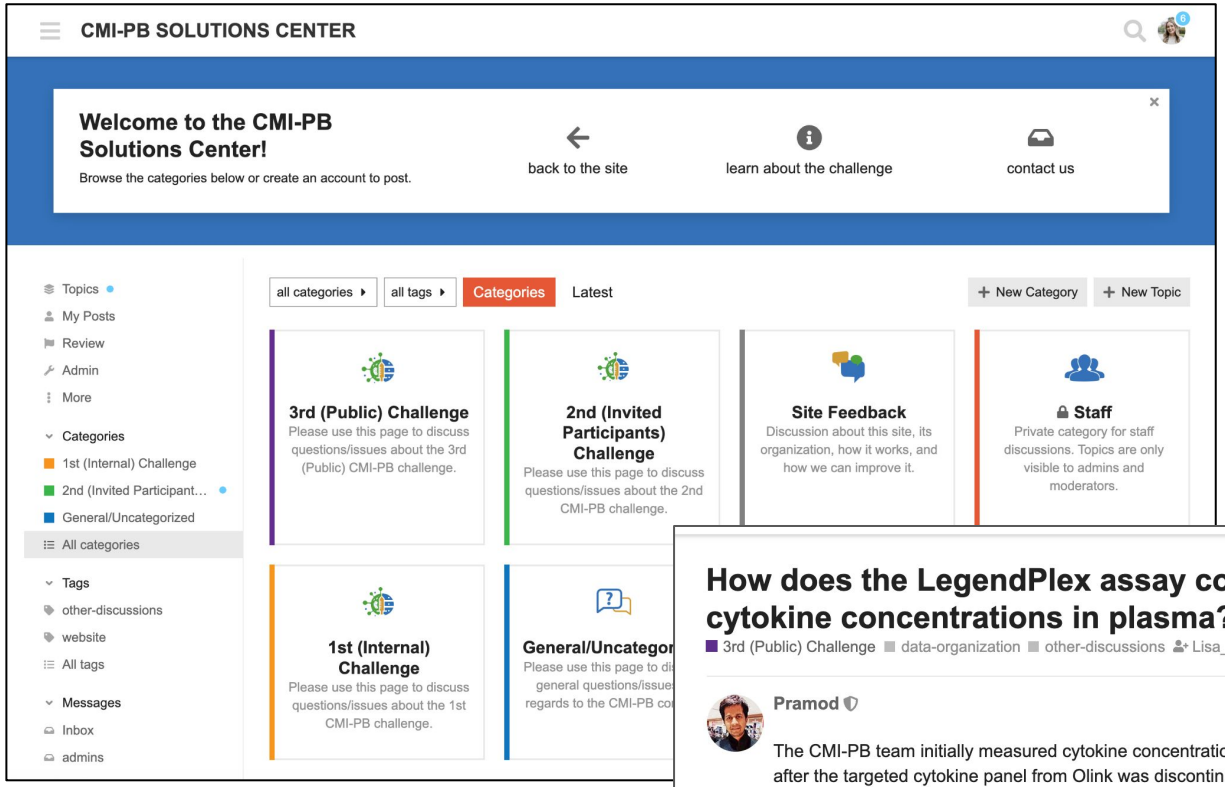
Click here to learn more about pertussis

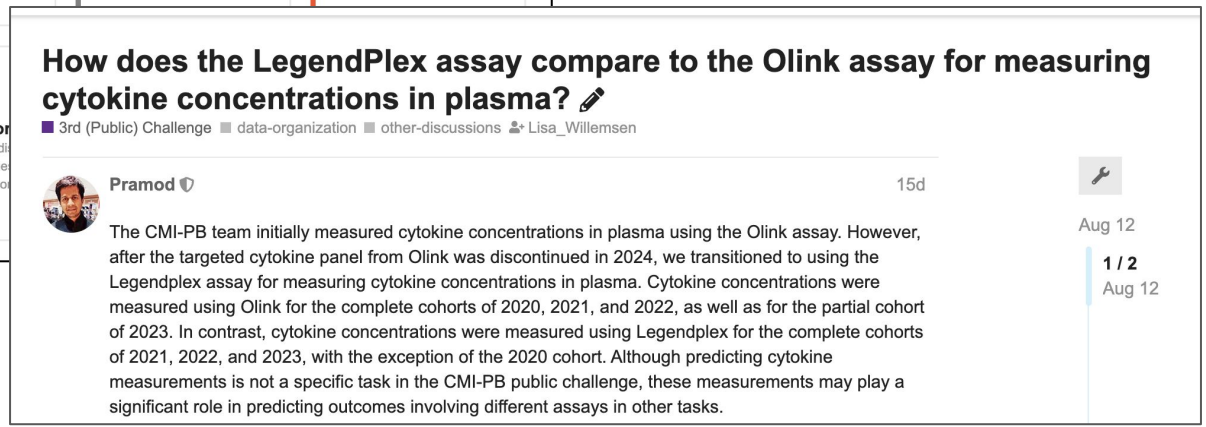
Click here to learn more about the prediction challenge

Click here to access the Solutions Center

Click here to find example models

CMI-PB Solutions Center

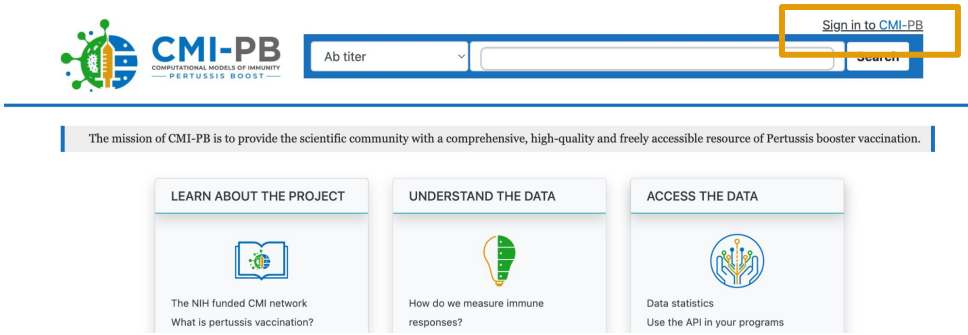




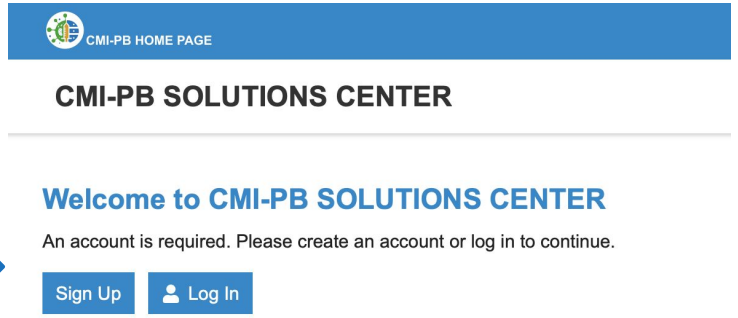
- Platform for knowledge sharing and discussion
- An account is required to post or respond to threads

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

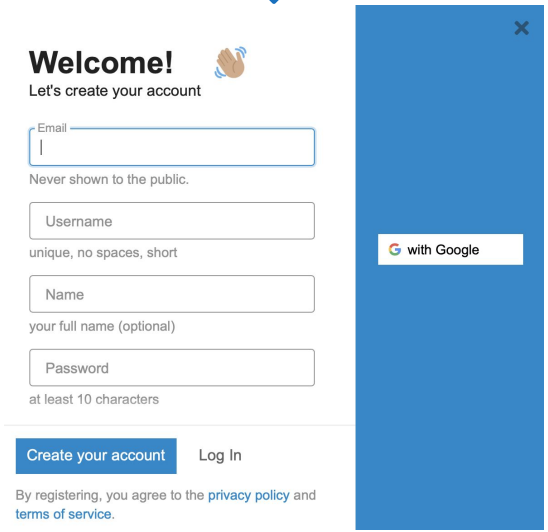
Creating an account



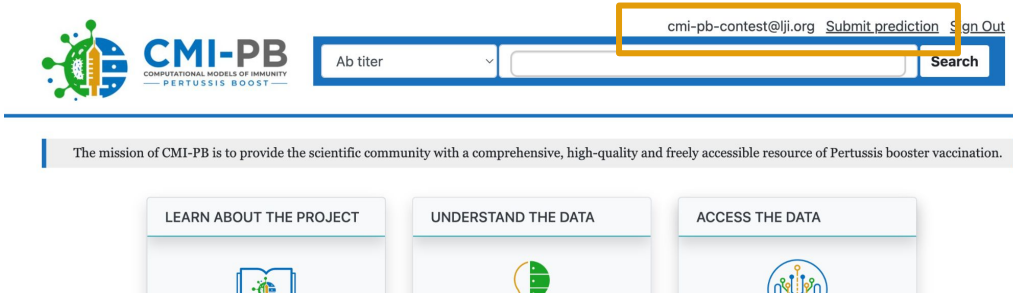
Step #1: Click ***"Sign in to CMI-PB"*** in the upper right hand corner



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

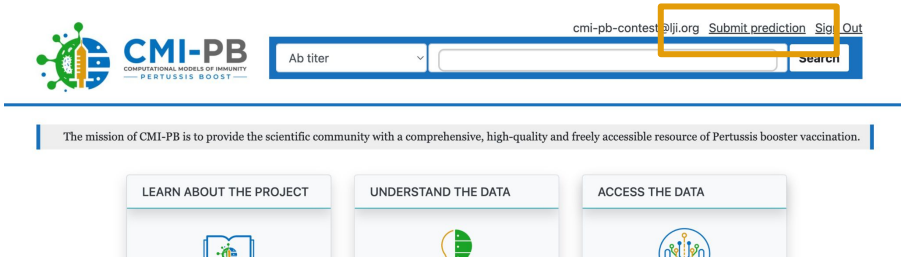


Step #3: Fill out registration form or use SSO with Google

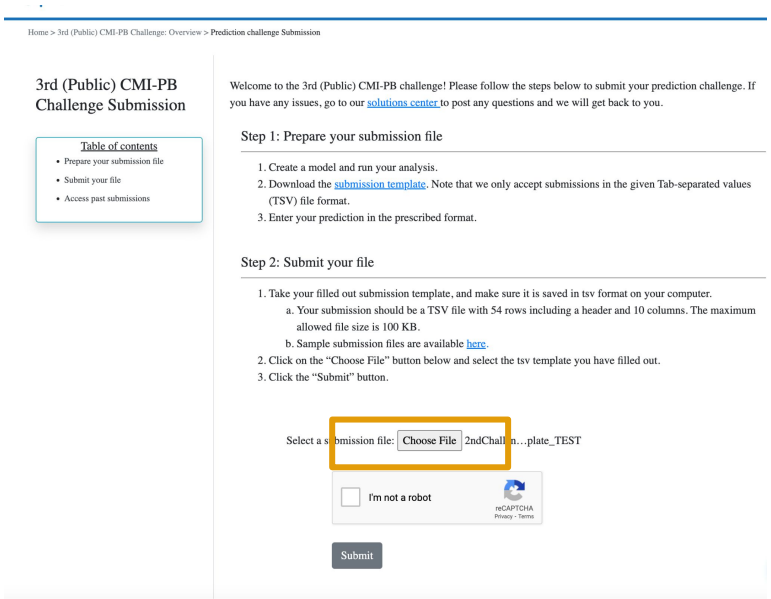


Step #4: Confirm that when you are signed in, your email is shown in the upper right hand corner

Creating a submission

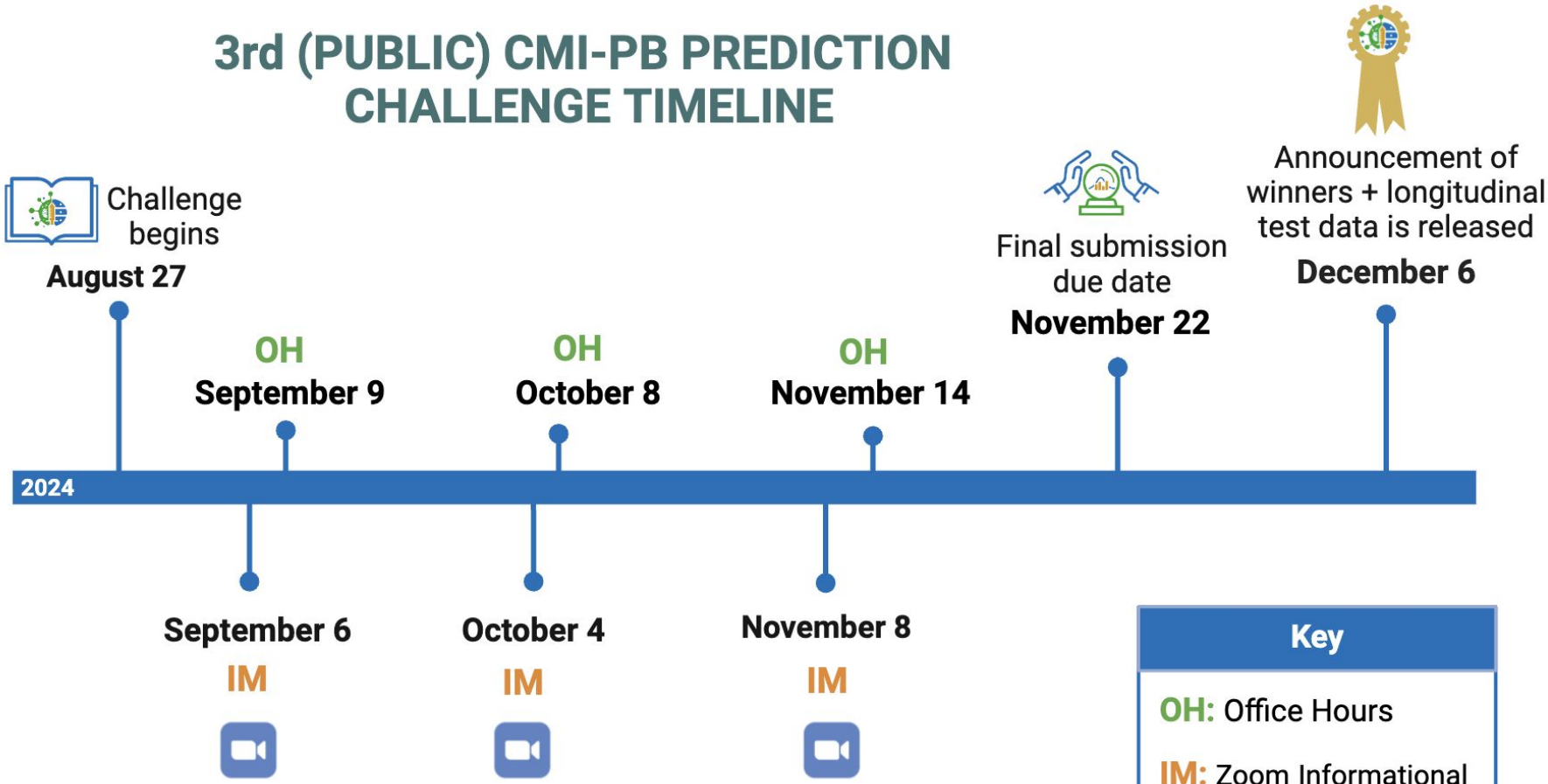


Step #1: Once logged in, click ***"Submit Prediction"*** in the upper right hand corner



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

3rd (PUBLIC) 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
Jian Xing
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
Aaron Ren
Lisa Willemsen
Shelby Orfield
Nick Chan

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

The CMI-PB team members

CMI-PB
COMPUTATIONAL MODELS OF IMMUNITY
PERTUSSIS BOOST

Bjoern Peters

Steven Kleinstein

Ferhat Ay

Barry Grant

Shane Crotty

Alessandro Sette

Pramod Shinde

Shelby Orfield

Lisa Willemsen

Leying Guan

Joaquin Reyna

Mari Kojima

Jason Greenbaum

Brendan Ha

Aaron Ren

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 post your questions on https://discuss.cmi-pb.org/
under the 3rd Public Challenge

We will be hosting an open office hour session via Zoom on **Monday, September 9th 11:30am-12:30pm PT/2:30pm-3:30pm ET**.

Feel free to drop by if you have any questions!
Zoom information is available on the Solutions Center here.