CMI-PB Third Challenge Debrief Session

14.03.2025 Philipp Sven Lars Schäfer

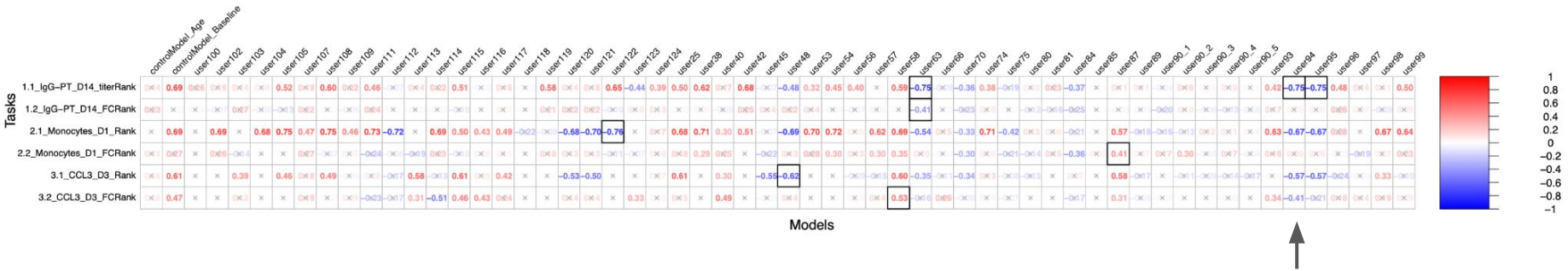
https://psl-schaefer.github.io/cmi_pb_third_challenge/

Thanks for Organizing the Challenge!

General Thoughts on the Challenge

- Key Challenges
  - Large-P-Small-N Problem
  - Seven different prediction tasks
  - Six different data modalities
  - Cohort-specific “batch” effects
-
- Key Priorities
  - Solid QC and data preprocessing of the experimental data
  - Solid Model evaluation framework, because it is easy to overfit on the training data in the Large-P-Small-N setting

General Thoughts on the Challenge



- Key Results
  - Very difficult to beat the baseline!
  - I cannot read

“We expect contestants to generate computational models, and upon making predictions, these values are ranked from highest to lowest (i.e. highest = 1, lowest = N)”

QC & Preprocessing - Removing Measurements

- Manually subset features for certain assays:
  - PBMC frequencies: Selected a subset of cell types based on the gating info
  - Olink: Only use proteins with NA fraction below 50%.
- Olink QC Qarning:
  - -> Removed measurements with QC warning
- Different Units in Assay
  - Plasma Antibody Titers:
    - In the 2020 dataset we have `IU/ML` whereas for 2021-2023 we have MFI (fluorescence intensity).
    - `IU/ML` is obtained using a serum standard, thus there is not trivial conversion from MFI to IU/ML.
    - -> Removed the 2020 Ab titer data
  - Olink:
    - In the 2020 dataset we have `Normalized Protein eXpression`, whereas for 2021-2023 we have `PG/ML`
    - -> Removed all measurements with different units.
- Lower Limit of Detection in Plasma Antibody Titers:
  - -> Removed specimen with more than 50% of measurements below LOD
- Lower Limit of Quantification in Olink assay
  - -> Removed specimen with more than 50% of measurements below LOD
- Outlier Removal in Legendplex Assay
  - -> Removed 8 samples based on PCA plots

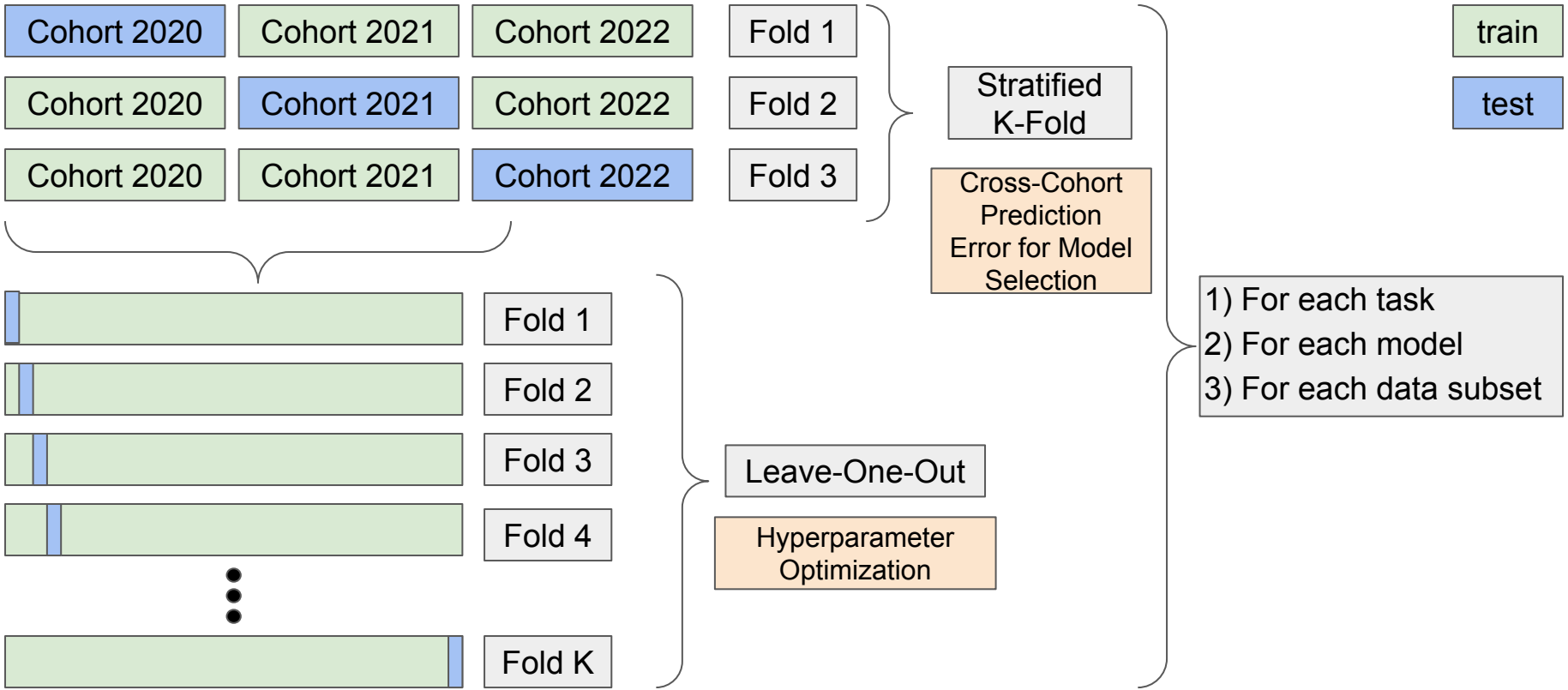
QC & Preprocessing - Normalization

- pbmc_cell_frequency:
  - Median Baseline Normalization (i.e. divide each feature by the median of that feature in the measurement from specimen from day 0)
- pbmc_gene_expression:
  - VST (from DESeq2)
- plasma_ab_titer:
  - Median Baseline Normalization
- plasma_cytokine_concentration_by_legendplex:
  - Median Baseline Normalization
- plasma_cytokine_concentration_by_olink:
  - No Normalization
- t_cell_activation:
  - No Normalization
- t_cell_polarization:
  - No Normalization

QC & Preprocessing - Integration / Batch Removal

- pbmc_cell_frequency:
  - No Integration
- pbmc_gene_expression:
  - ComBat-seq
- plasma_ab_titer:
  - No Integration
- plasma_cytokine_concentration_by_legendplex:
  - ComBat
- Plasma_cytokine_concentration_by_olink:
  - No Integration
- t_cell_activation:
  - No Integration
- t_cell_polarization:
  - No Integration

Model Evaluation Framework - Nested CV Setup



Model Evaluation Framework

- To evaluate and select models for each task, I used nested cross validation (CV):
- Folds in the outer loop:
  - Each cohort is a fold (i.e. Group k-fold)
  - Outer loop is used to estimate cross-cohort model performance
- Fold in the inner loop:
  - Each subject is a fold (i.e. LOOCV)
  - Inner loop is used to select the best set of hyperparameters for any model.
  - Using this set of hyperparameters I then estimate model performance on the hold-out fold from the outer loop
- I used this model evaluation framework to test all combinations of models and features:
  - Models:
    - LASSO, Elastic Net, Random Forest
  - Features:
    - Power set of all assays (+/- PCA per assay)

Model Evaluation Framework

A	B	C	D	E	F	G	H	I	J	K
task	model	srho_mean	srho_sd	srho_min	2021_dataset	2022_dataset	2020_dataset	dim_red	n_modalities	feature_set
task_21	rf	0,793	0,064	0,748	0,838	0,748		pca	5	pbmc_cell_frequency_pca+plasma_cytokine_concentration_by_legendplex_pca+t_cell_activation_pca+m
task_21	elnet	0,79	0,039	0,763	0,817	0,763		none	5	pbmc_cell_frequency+plasma_cytokine_concentration_by_legendplex+t_cell_activation+metadata+base
task_21	rf	0,775	0,015	0,764	0,785	0,764		pca	6	pbmc_cell_frequency_pca+plasma_cytokine_concentration_by_legendplex_pca+t_cell_activation_pca+t
task_21	elnet	0,766	0,08	0,71	0,823	0,71		pca	6	pbmc_cell_frequency_pca+plasma_cytokine_concentration_by_legendplex_pca+t_cell_activation_pca+t
task_21	lasso	0,763	0,06	0,721	0,806	0,721		none	5	pbmc_cell_frequency+plasma_cytokine_concentration_by_legendplex+t_cell_activation+metadata+base
task_21	lasso	0,759	0,15	0,59	0,876	0,59	0,811	pca	4	pbmc_cell_frequency_pca+pbmc_gene_expression_pca+metadata+baseline
task_21	elnet	0,755	0,161	0,571	0,87	0,571	0,825	none	4	pbmc_cell_frequency+plasma_cytokine_concentration_by_legendplex+metadata+baseline
task_21	rf	0,755	0,162	0,64	0,87	0,64		none	5	pbmc_cell_frequency+plasma_cytokine_concentration_by_legendplex+t_cell_activation+metadata+base
task_21	elnet	0,754	0,192	0,544	0,801	0,544	0,918	pca	4	pbmc_gene_expression_pca+t_cell_activation_pca+metadata+baseline
task_21	elnet	0,753	0,055	0,715	0,792	0,715		pca	6	pbmc_cell_frequency_pca+plasma_cytokine_concentration_by_legendplex_pca+plasma_cytokine_conce
task_21	lasso	0,751	0,195	0,531	0,821	0,531	0,9	pca	3	pbmc_gene_expression_pca+metadata+baseline
task_21	lasso	0,751	0,195	0,531	0,821	0,531	0,9	pca	4	pbmc_gene_expression_pca+plasma_cytokine_concentration_by_legendplex_pca+metadata+baseline
task_21	lasso	0,751	0,195	0,531	0,821	0,531	0,9	pca	4	pbmc_gene_expression_pca+t_cell_activation_pca+metadata+baseline
task_21	lasso	0,748	0,02	0,735	0,762	0,735		pca	7	pbmc_cell_frequency_pca+plasma_cytokine_concentration_by_legendplex_pca+plasma_cytokine_conce
task_21	baseline	0,746	0,172	0,552	0,876	0,552	0,811	none	1	baseline

Choosing the Final Model + Data for Task

- Top ranking performance
- Low variance between the test sets
- If several models with top performance and low variance, choose the more regularized model
- If several models with top performance and low variance, choose model that is not using any assay that is missing often in test data (specifically Olink!)

Choosing the Final Model + Data for Task

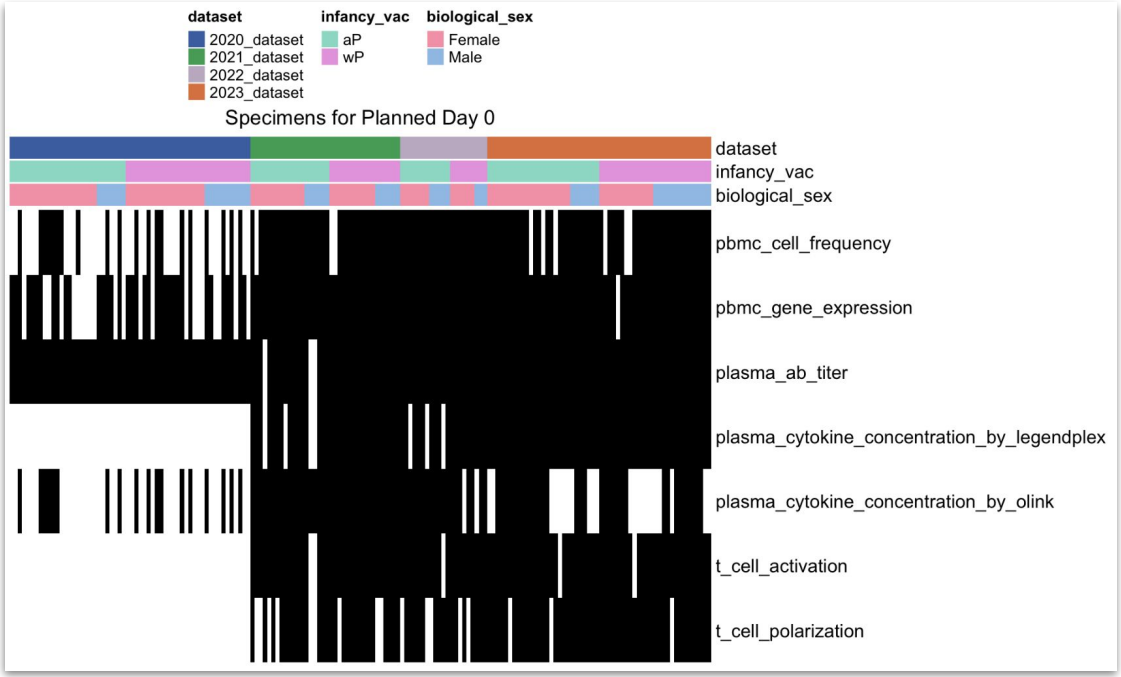
- Other notes:
  - I tested Boruta Algorithm for feature selection but results were not convincing (and it costs quite some compute)
  - I wanted to test multi-omics integration methods, such as MOFA+, but didn’t have the time to do so…

Thank you for you attention! Questions?

Appendix

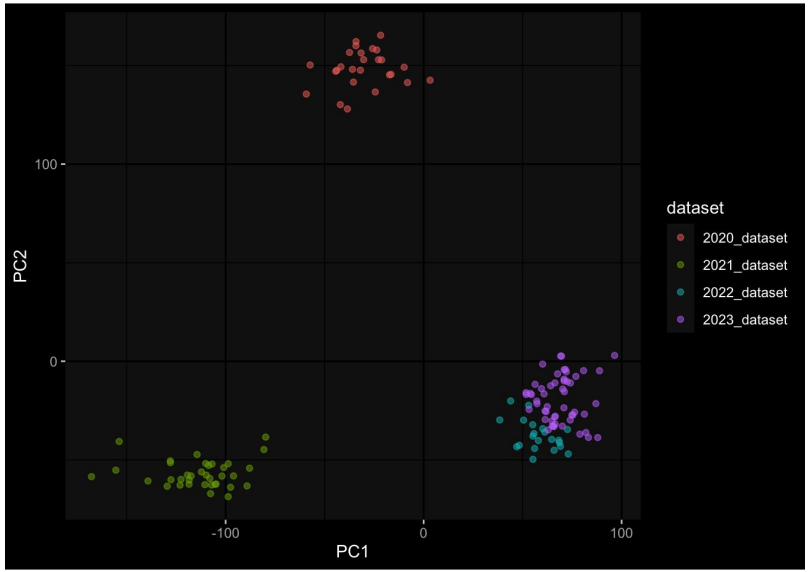
Issues with the 2020 Cohort

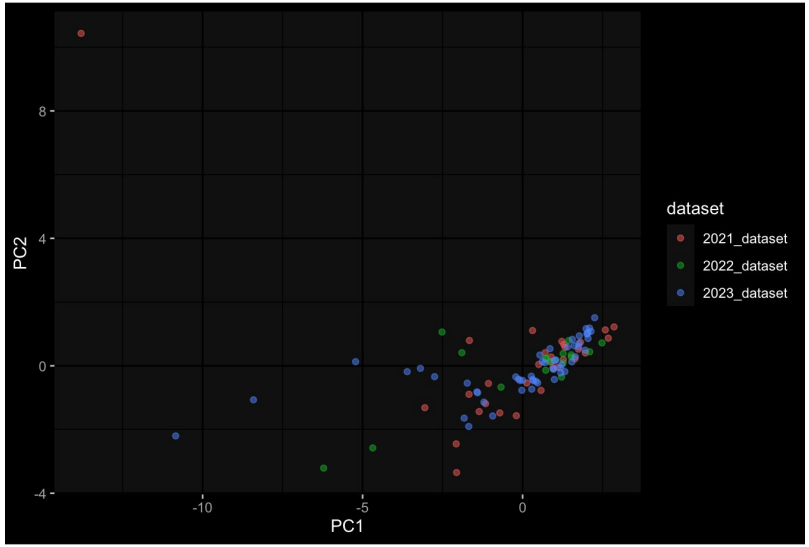
- Many assays are missing, so I actually ended up not using the 2020 Cohort for most prediction tasks



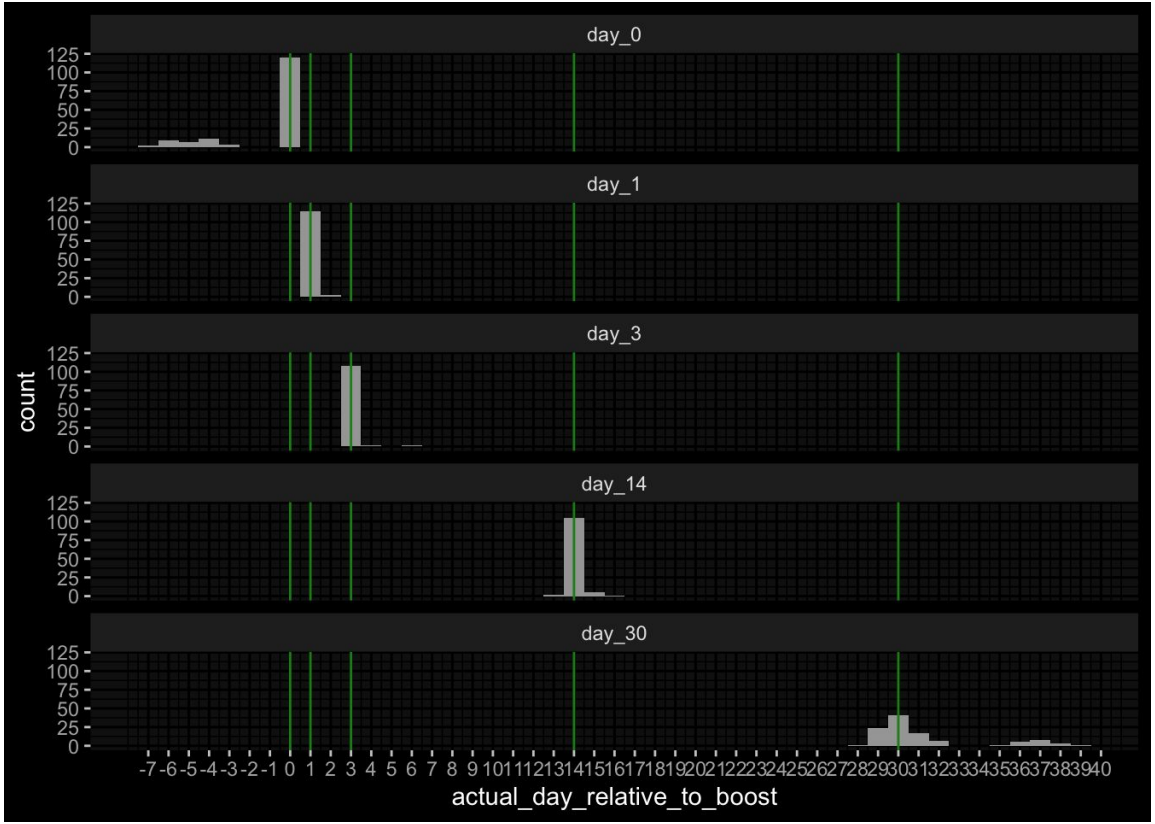
Batch Effects after Normalization

- Strong batch effect for PBMC Gene Expression
- Small batch effect for Lengendplex (driven by outliers?)





What mismatch do we allow for when generating target tracks?



Differences in the Marginal Distributions of the Targets between Cohorts

- Ab Titer Tasks (1.1 and 1.2), raw data