

Predicting Th1/Th2 Polarization Ratio Using MOFA

By Aanya Gupta

CMI-PB 3rd Vaccine Challenge

Methodology

Data Preprocessing



Feature Selection - MOFA



Prediction - LASSO

Methodology

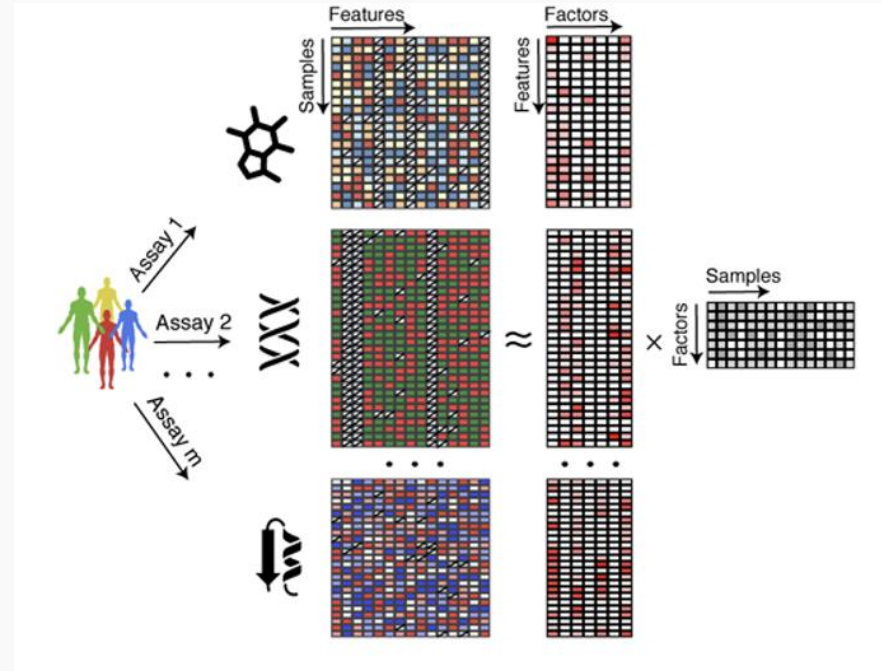
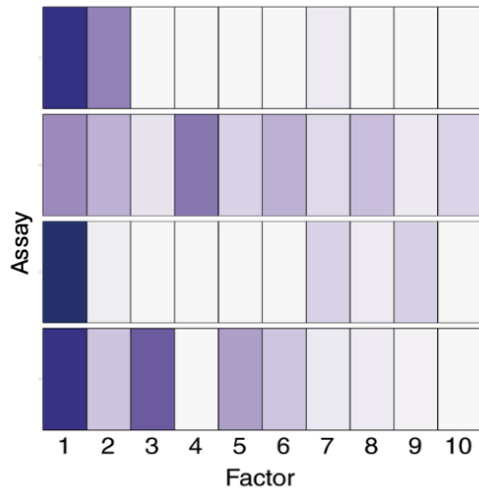
Pt. 1 – Data Preprocessing

- Selected timepoint 0 of batch-corrected data for all assays and years (2020, 2021, 2022)
- Outcome was binned into 8 equal range bins since MOFA requires discrete values for group, which is how outcome was represented.

Methodology

Pt. 2 – Feature Selection (MOFA)

- PCA-like analysis through latent factors
- ARD to promote sparsity in weights & factors
- Factors capture modality and multi-omics variance



Why MOFA?

Problems

- Large number of missing values
- High-dimensional data can cause irrelevant features to be chosen
- Connectivity and heterogeneity of relating assays

MOFA's Solutions

- Unsupervised learning
- Can deal with missing values
- Can work with highly dimensional data through dimensionality reduction
- Easy interpretability due to simple structure

Methodology

Pt. 3 – Prediction (LASSO)

- Linear regression model
- ElasticNet penalty to find sparse but adequate amount of features
- Grid search with five-fold cross validation was used to find the best parameters for the model.

Why LASSO?

Problems

- Overfitting due to complex structure and noise of features
- Sensitive to hyperparameter changes
- Suffers from 'black box' problem

Lasso's Solutions

- Easy to interpret
- Promote sparsity by reducing coefficients to zero
- ElasticNet penalty = balances L1 and L2 regularization

Results

- One factor was taken for MOFA, to explain the most variance.
- For each assay, the top four features captured by the factor were taken.

- | | | |
|---------------------|------------|-------------|
| • TcmCD8 | • IgG3 PRN | • DPEP2 |
| • CD4+ CD8+ T cells | • IgG1 OVA | • SEC24C |
| • Basophils | • IgG1 PT | • GTF3C2 |
| • Naive CD4 | • IgG3 FHA | • PT Q16552 |

- IgG1 PT -> antibody levels against pertussis toxin
- IgG3 FHA -> antibody levels against FHA, virulence factor of pertussis-causing bacteria
- Achieved correlation of 0.371 on 2023 dataset

Thanks!
