# Data integration for predictive modelling

## CMI-PB model submission

Nicky Thrupp

La Jolla Institute FOR IMMUNOLOGY | Life Without Disease.®

# Model concept

Stochasticity in individual measures

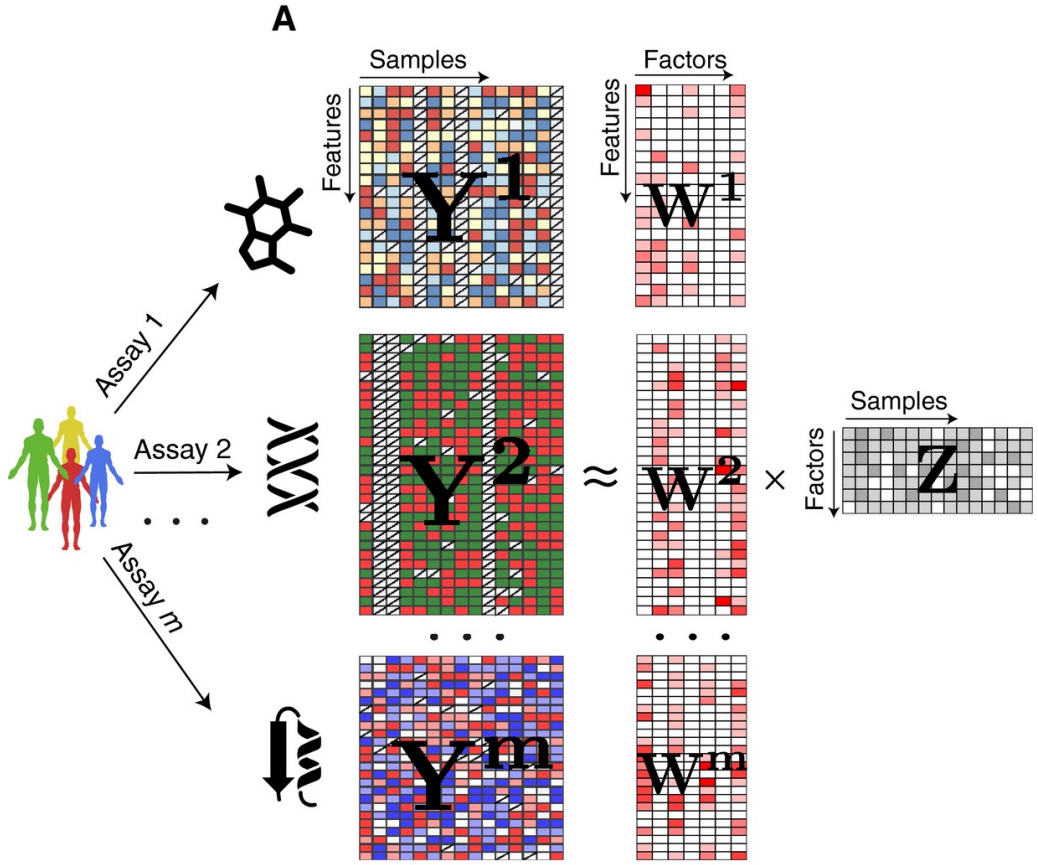
Measures between experiments may suffer from technical bias

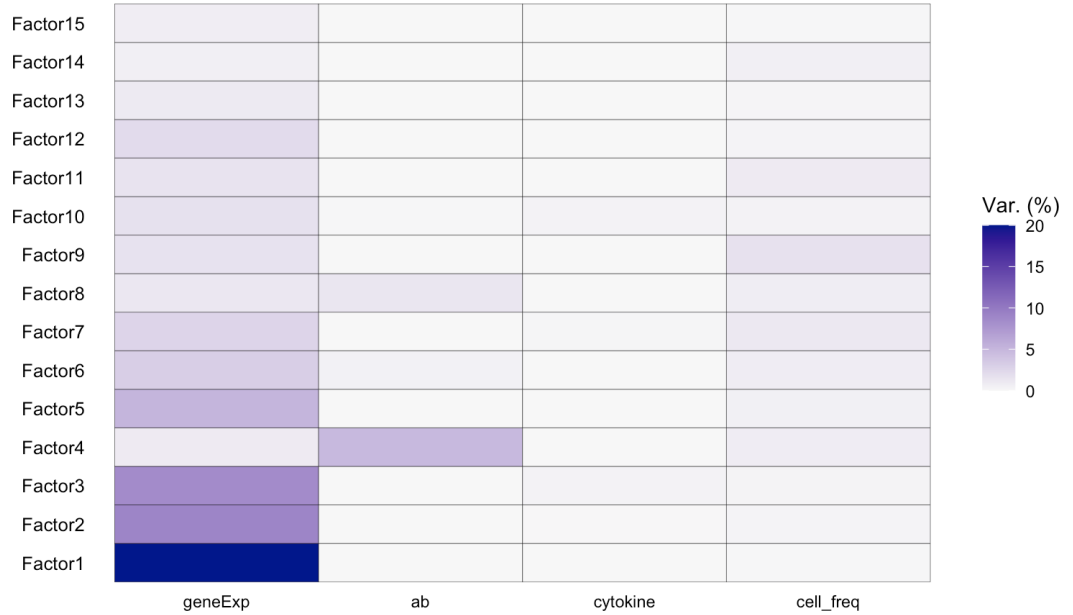
How to maximise signals coming from multiple sources

→ **can we address this with data integration**

- Leverage measures that are similar across modalities
- Remove noise
- Biologically interpretable

# MOFA for data integration



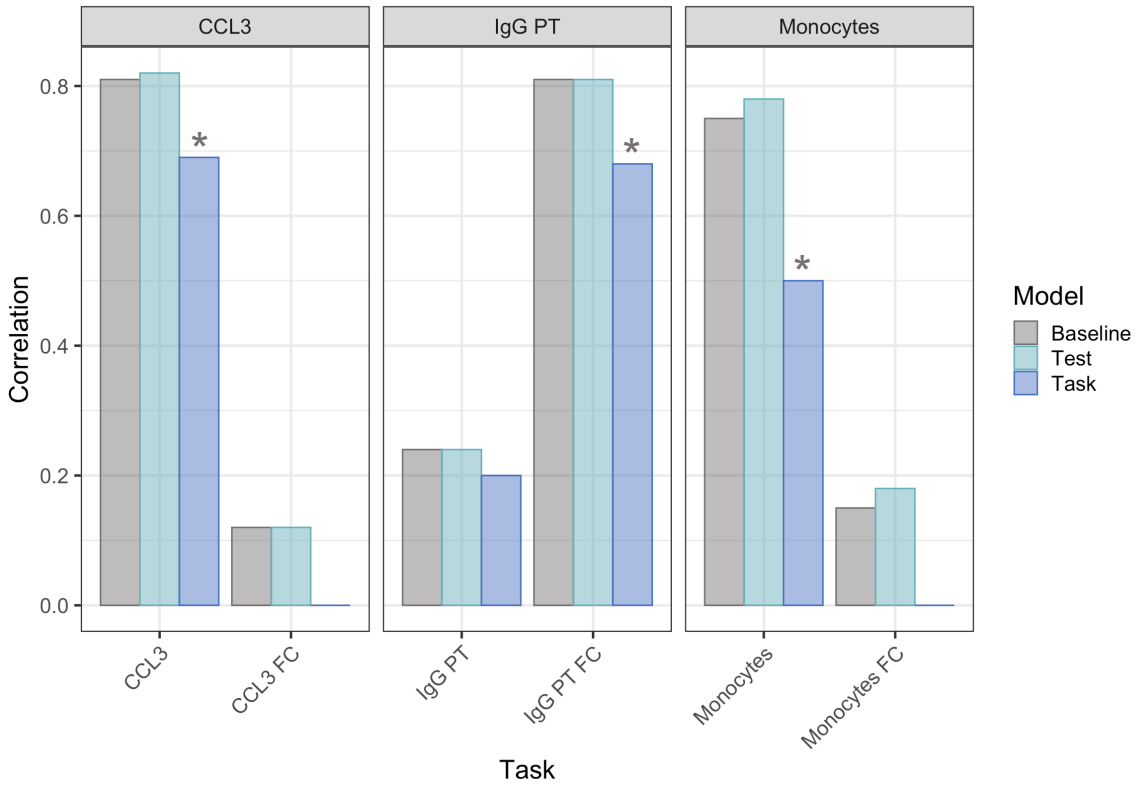

Variance decomposition

Matrix factorization : Y = observed measures
Z = inferred latent variables
W = feature weights

*Argelaguet et al. (2018) , Molecular Systems Biology*

# Results

- 2 models:
  - 2020 train, 2021 test
  - 2020+21 train / test
- Lasso regression
- Prediction features:
  - Baseline values
  - Demographic information
  - Top MOFA features



# Conclusions

- Models predicted CCL3 expression and IgG PT fold change well
- MOFA factors improved predictions by small margins
- MOFA model could still be improved:
  - Signals dominated by gene expression modality
  - Too many factors
  - Newer implementation takes groupings (including time courses) into account
- Very similar results when training 2020 and 2020+2021

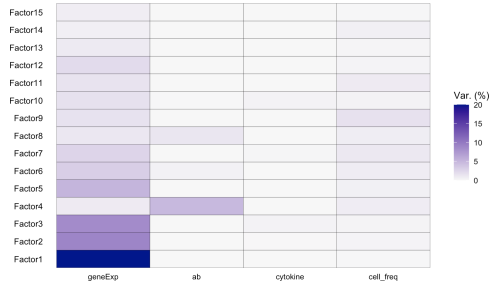
https://github.com/nixstix/CMI-PB2
https://github.com/topics/2nd-cmipb-challenge

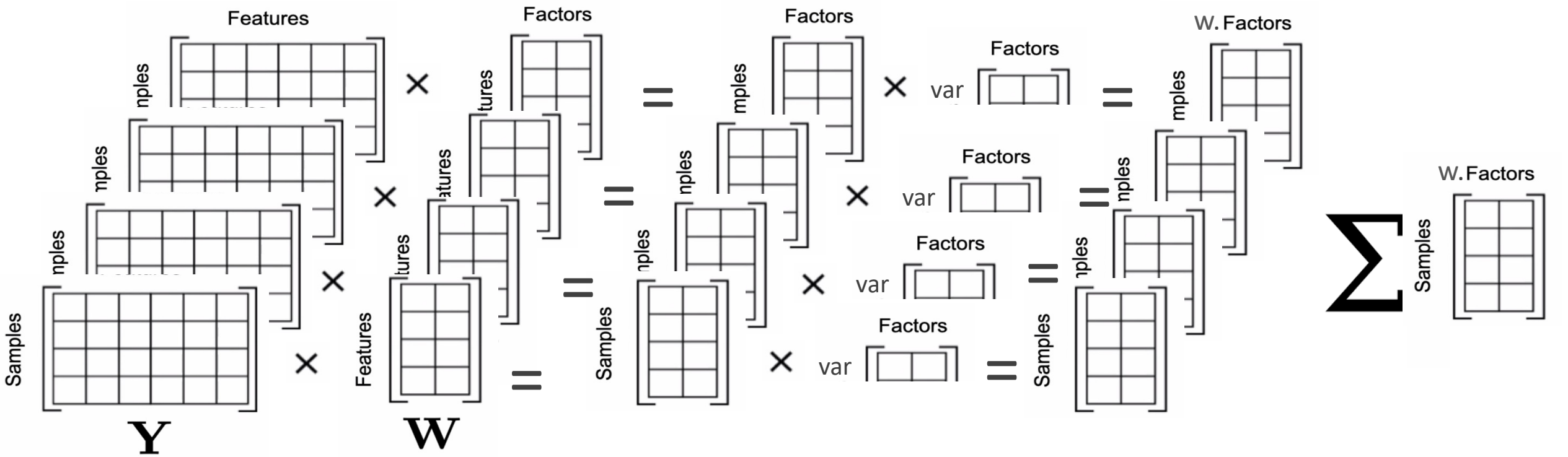
# Thanks !

La Jolla Institute FOR IMMUNOLOGY | Life Without Disease.®

# Training data:

- clinical data
- baseline assay data (ab, cytokine, cell freq, gene expr)
- MOFA factors:




Test data: Baseline data

Train data: MOFA weights

Train data: MOFA variances

Test data

# LASSO regression in R

```
library(glmnet)
```

```
model<-cv.glmnet(x=as.matrix(predictors.rmNA),
                 lambda = NULL,
                 task, family='gaussian',
                 alpha=alpha,
                 nfolds=nrow(predictors.rmNA)  ,
                 type.measure="mse")
```

Cross-validation using leave-one-out
Hyperparameter tuning of lambda parameter
Additional tuning of alpha parameter (lasso -> elastic -> ridge)