Cross-Platform Omics Prediction (CPOP)

Kevin YX Wang[1], Gullietta Pupo[2], Varsha Tembe[2], Graham Mann[2,3], Garth Tarr[1], Samuel Mueller[1], Jean Yang[1]

1 School of Mathematics and Statistics, The University of Sydney

2 Melanoma Institute Australia, The University of Sydney, North Sydney, Australia

3 The John Curtin School of Medical Research, Australian National University

THE UNIVERSITY OF SYDNEY

Motivation

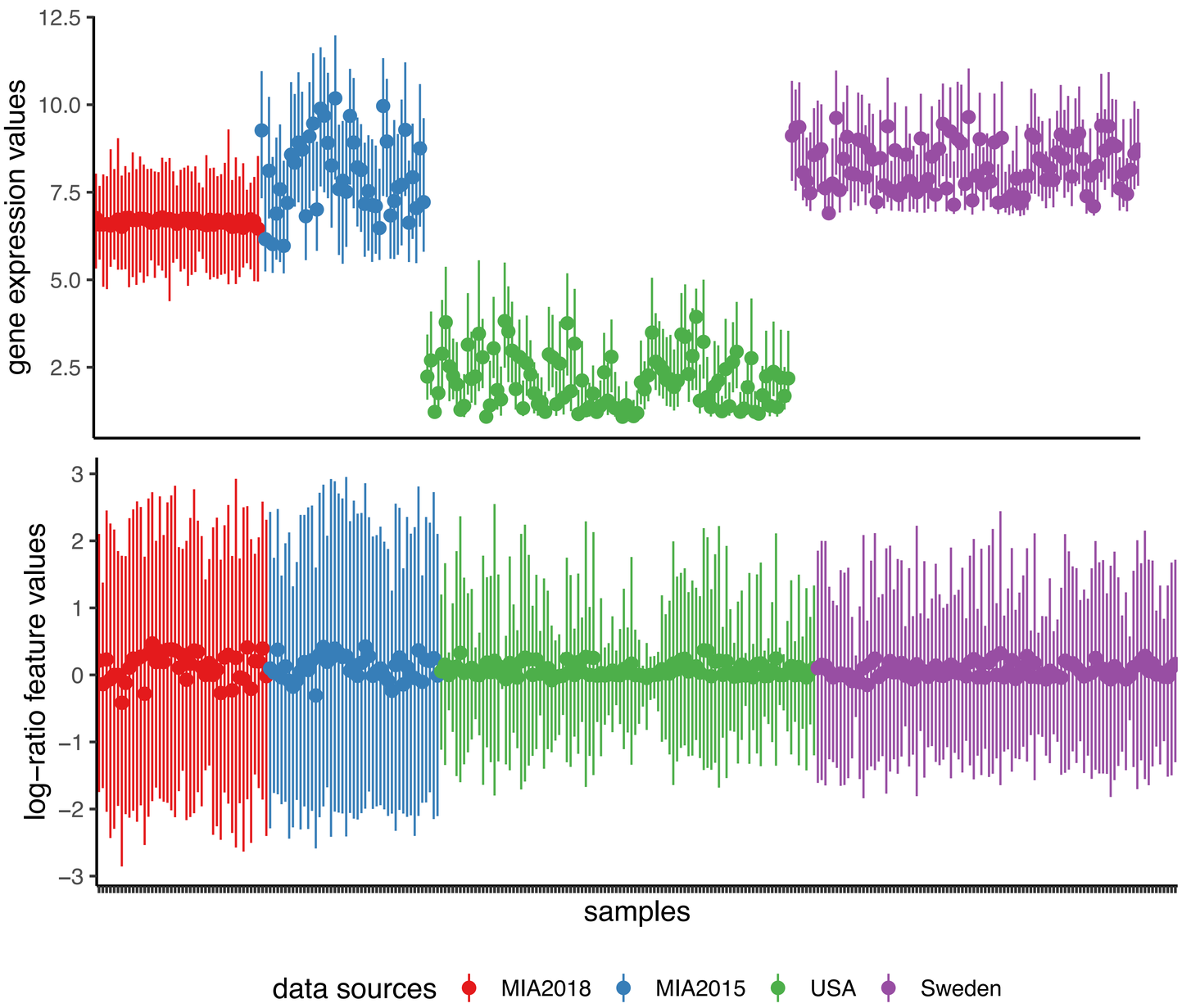
When implementing a prediction model in the clinics, what is more important?

Classification accuracy or stability of prediction?

Which one is more widely reported in publications?

Stable prediction
=
stable features
+
stable modelling

Due to experimentation constrains, we might not be able to normalise our data and use the standard prediction model like Lasso and Random Forest.



Methods

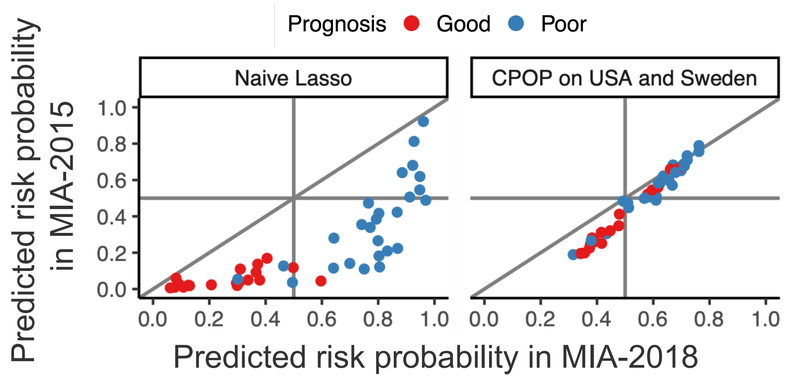
- CPOP is a statistical procedure that enables prediction using omics data in a clinical settings.
- CPOP avoids re-normalisation of additional data through the use of log-ratio features and thus also enable prediction for single omics samples.
- The novelty of the CPOP procedure lies in its ability to construct a *transferable* model across gene expression platforms and for prospective experiments.
- Such a transferable model can be trained to make predictions on independent validation data with an accuracy that is similar to a re-substituted model.

Results

We tested CPOP on four late-stage melanoma datasets.

- MIA-2015 (Jayawardana et al. 2015)
- MIA-2018 (Unpublished)
- USA (TCGA 2015)
- Sweden ((Cirenajwis et al. 2015)

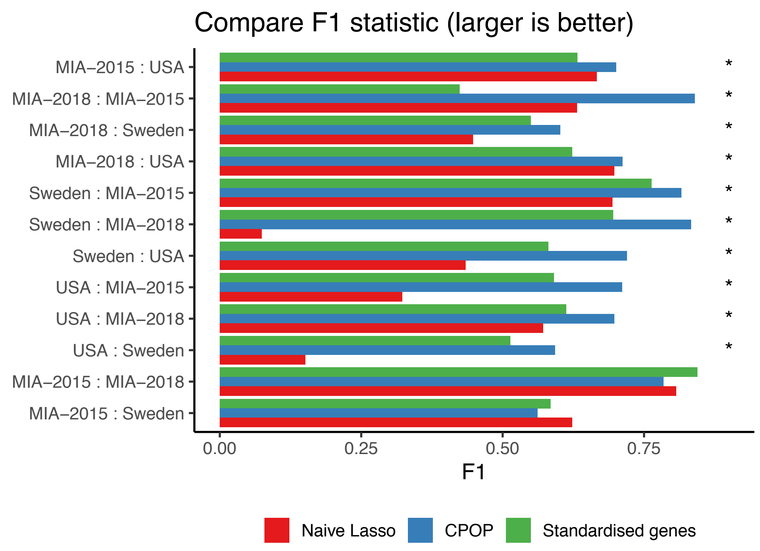
Equivalent predictions



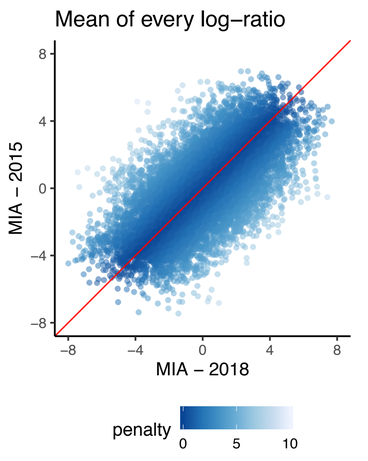
Accuracy: 0.63
Avg. deviation: 15%

Accuracy: 0.81
Avg. deviation: 5%

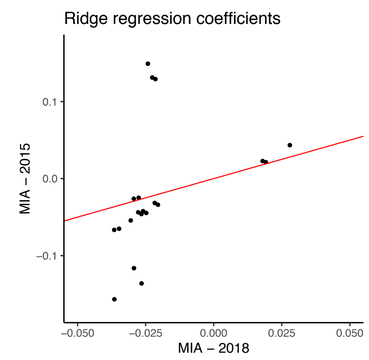
High prediction accuracy



How does CPOP work?



Weight each feature by its stability across datasets



Select only features with similar estimated coefficients across datasets