Why we do not have an omics-based clinical risk score?

Most ML methods apply standardisation on both training and validation data

 But in clinical prediction, you might only have single samples to predict on

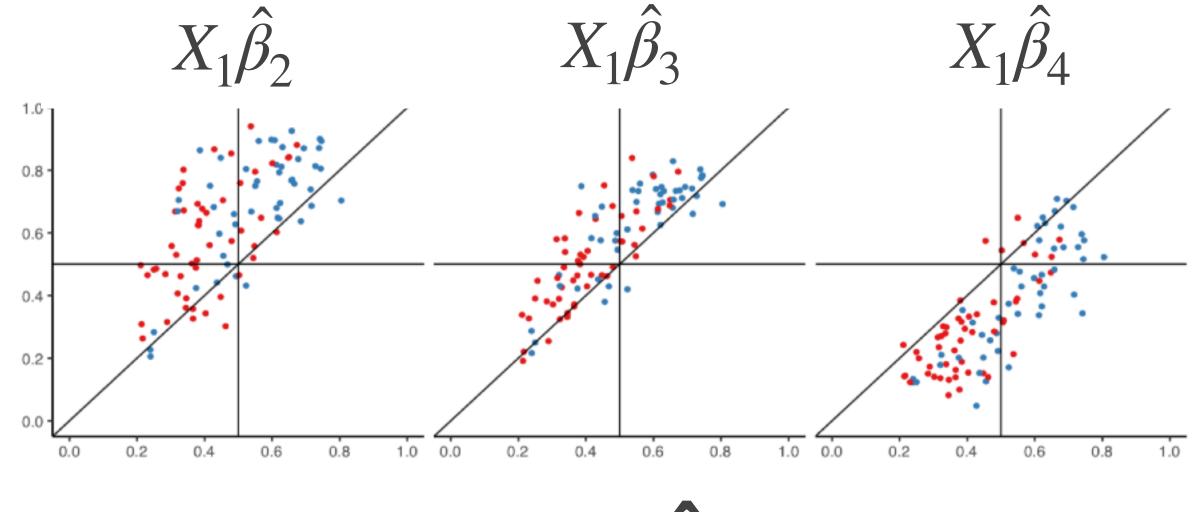
New X	Sample 4
Gene 1	4.2
Gene 2	3.8
Gene 3	8.4
Gene 4	3.1

Statistical challenges

- 1. Different scaling on genes leads to differences in predicted values
- 2. Single-patient prediction
- 3. Variable selection/coefficient estimates due to differences in correlation

Transferability

The prediction on one gene expression platform should be equivalent to another platform



$$X_1\hat{\beta}_1$$