

Supplementary Figure 1

Pseudo-code for the cross-validation:

Outer CV loop (Leave one out)

Selection of the $(X_{(n-1) \times p}, Y_{n-1})$ training set and the test set $(X_{1 \times p}, Y_1)$

Training:

PLS model for $(X_{(n-1) \times p}, Y_{n-1})$

Number of PLS components h chosen by F-tests

VIPs for each variable calculated for each X_p

Sort the variables according to the VIPs $X_1, \dots, X_p \rightarrow X_{1:p}, \dots, X_{p:p}$

$X_{select} \leftarrow X_1$

for(i in $1:p$)

$F\text{-test}(X_{select}, X_{i:p})$

if(p value < 0.05)

$X_{select} \leftarrow (X_{select} \vee X_{i:p})$

end

Estimate linear model (OLS, or PLS) on $(X_{select}, Y_{n-1}) \rightarrow \beta$

Test:

Predict the response of the test set using β : $\hat{Y} = X_{select} \beta$

end CV loop

Predictive power = $R = \text{cor}(Y, \hat{Y})$