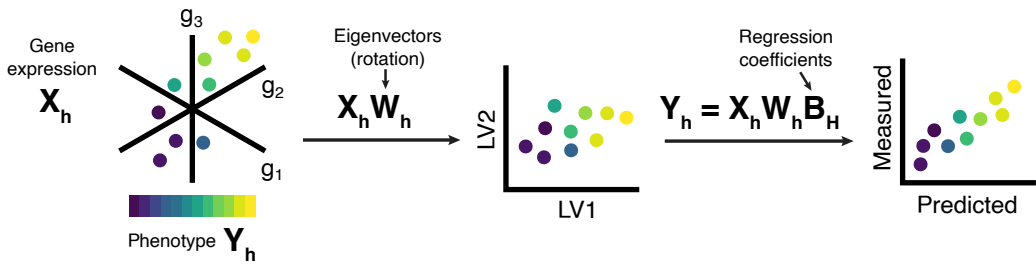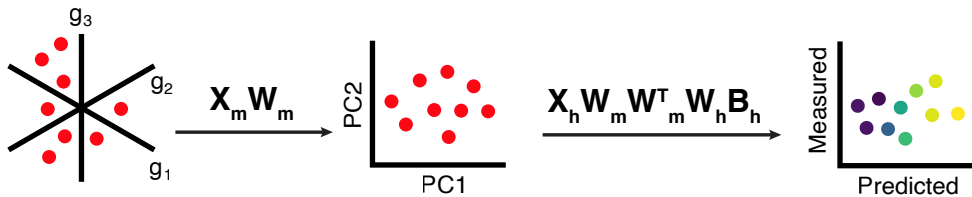Preprocessing: Filter **human** and **MPS** datasets for common features (e.g. measured genes)
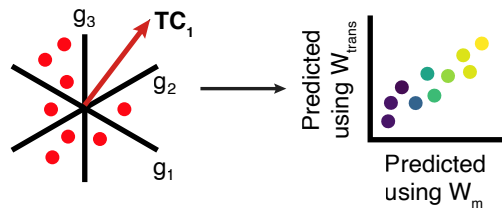
Step 1: Train a model (e.g. PLSR) to predict human phenotypes from human molecular features.

Gene expression X_h

g_3 g_2 g_1

Phenotype Y_h

Eigenvectors (rotation)

$X_h W_h$

LV2 / LV1

Regression coefficients

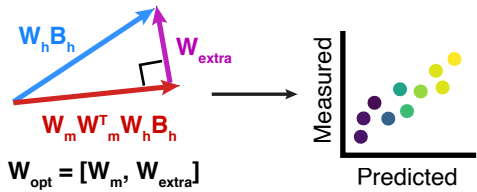$Y_h = X_h W_h B_H$

Measured / Predicted

Step 2: Build a representation of the MPS gene expression space by using PCA. Project and backproject human gene expression through MPS space to remove human variance not captured in MPS space.

g_3 g_2 g_1

$X_m W_m$

PC2 / PC1

$X_h W_m W_m^T W_h B_h$

Measured / Predicted

Step 3a: Calculate translatable components from MPS dataset.

Step 3b: Calculate extra components required for MPS to capture more relevant human variance.

g_3 **TC$_1$** g_2 g_1

Predicted using W_{trans} / Predicted using W_m

$W_h B_h$ W_{extra} $W_m W_m^T W_h B_h$

$W_{opt} = [W_m, W_{extra}]$

Measured / Predicted

Step 4: Interpret translatable and extra components to prioritize experimental stimuli and to propose new perturbations.