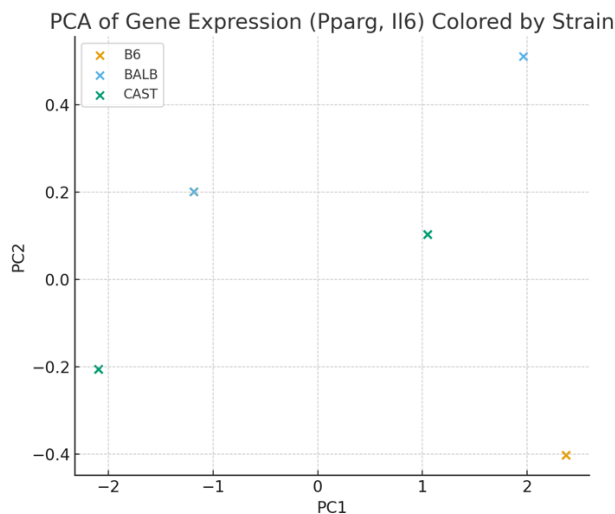


### Module 5C Questions:

1. You are studying the gene expression profiles of different mouse tissues (e.g. liver, brain, kidney) to find patterns in how genes behave across tissues. The dataset is large, with thousands of genes and multiple conditions. You decide to use Principal Component Analysis to reduce the complexity of this data set and visualize it. Explain (at a high level) how PCA helps you identify patterns in the gene expression data, and what does it mean when a particular tissue type appears close to others in a PCA plot? Explain what the principal components are?
2. Can we use dimensionality reduction to uncover patterns in gene expression, and do these patterns relate to mouse strain or diet? Examine the PCA plots below and interpret what you see (Note: with only 12 data points, this is much less obvious than it would be with many more points)
  - Do PCA clusters appear?
  - Are the clusters separated by **strain**, **diet**, both, or neither?
  - Which gene (Pparg or Il6) seems most influential in separating points?
  - Do the points overlap heavily or form distinct groups?

### PCA of Gene Expression Colored by Strain

(Based on *Pparg* and *Il6* from the 12-mouse dataset) This plot shows whether gene-expression variation naturally groups according to **strain**.



### PCA of Gene Expression Colored by Diet

This plot shows whether gene-expression variation naturally groups according to **diet** (Chow vs. HFD).

