Problem #1: Brain Region and Sex Effects on RNA Expression in Rats

The data set, **rat\_brain\_regions.txt**, contains normalized/transformed RNA expression estimates for 16,569 Ensembl genes expressed above background. There are 45 sequencing libraries from 10 rats (5 female and 5 male) across 5 different brain regions (nucleus accumbens core [Acbc], lateral habenula [LHB], prelimbic cortex [PL], infralimbic cortex [IL], orbital frontal cortex [VoLo]). The library information is contained in **library\_info.txt**.

Using these data:

1. Determine the relationship among the 45 samples using hierarchical clustering. What is the major factor that contributes to clustering?
2. Examine and identify (when possible) the sources of the 5 top principal components
3. Examine the distribution of expression values within a region for P2rx4 (ENSRNOG00000001300). Make two comments about these distributions.
4. Using a paired t-test, identify genes that are differentially expressed between the prelimbic cortex and the infralimbic cortex. Plot the p-values and comment on their distribution compared to the distribution that would be expected by chance.
5. Calculate a false discovery rate and identify the number of genes that are differentially expressed at an FDR threshold of 0.05.
6. Plot the expression data for the 50 most significant (i.e., lowest p-values) genes in the paired t-test in a heatmap that includes all 45 samples. What general patterns can be detected?