**Hierarchical PCA**

The goal of HPCA (hierarchical principal component analysis) is to improve the interpretability and mechanics of a standard PCA (principal component analysis) by incorporating knowledge of biologically relevant groupings of variables. Standard PCA on a correlation matrix gives equal weighting to all included variables when calculating the principal components (PCs). Thus, if a metabolic process has a greater number of components in its pathway, then this metabolic process will have a greater influence on how the variance is captured across the PCs. HPCA mitigates this problem by first performing a group-wise PCA on the variables one user-defined grouping at a time, then another overall PCA is performed across the groups’ selected PCs. Group-wise PCs selected for overall PCA are first subject to quality control, by requiring a minimum coverage (>10%) of variance across its group variables. To improve the interpretation of the resultant PCs at each of the 2 hierarchical PCA steps, each PC is annotated with its closest underlying variables (by correlation), in order to identify a single underlying variable as a potential proxy for that PC. The greater the correlation the PC is with one of its underlying variables, the better the variable can be represented by it. This annotation step provides a biologically meaningful name to each PC when the correlation is sufficiently high. As a means of variable selection, the variables used to annotate each PC are also good candidates for downstream regression analysis. Prior to the group-wise PCA, all variables are tested with the Shipiro-Wilks test for lack of normality (p-value < 0.2) to determine if a log transformation would improve its normality. If the transformed variable is more statistically significantly non-normal, then the original values of the variables are retained. For diagnostics, dendrograms based on a distance matrix calculated with the metric: 1-abs(cor(x,y)) are generated to visualize the relationship between selected PCs and the other variables in the group.