Supplemental Material

Cluster Analysis

Data Normalization

A z-score algorithm was applied to each analyte which returns the deviation of each data element from its mean, normalized by its standard deviation. Thus two vectors of analytes, r and s, can be plotted on the same scale and their values would correspond to each sample's deviation from the mean, zero.

The z-score transformation, z_i, of the i-th element of a vector of analytes, r, is:

$$z_i = \frac{r_i - \bar{r}}{\mathrm{std}(r)}$$

$\bar{r}$ and std(r) are the arithmetic mean and sample standard deviation of the vector of analytes, r, respectively.

Distance Measure

The distance, d_{ij}, calculation for two vectors of analytes, i and j, is computed:

$$d_{ij} = 1 - \frac{(i - \bar{i}) \cdot (j - \bar{j})^T}{\left[(i - \bar{i}) \cdot (i - \bar{i})^T\right]^{1/2} \left[(j - \bar{j}) \cdot (j - \bar{j})^T\right]^{1/2}}$$

^{T} is the matrix transposition operator, and $\bar{i}$ and $\bar{j}$ are the arithmetic mean of the vector of analytes i and j, respectively.

Cluster Linkage

Similar analytes were organized into a binary hierarchical cluster tree using an average linkage algorithm based on their distances, d_{ij}, calculated as above. At the first iteration, each cluster consisted of only one or two analytes. At the second and subsequent iterations of the algorithm two clusters, a and b, with the smallest average distance, l_{ab}, formed a cluster by comparing the unweighted average distance between all pairs of analytes in two clusters.

$$l_{ab} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} d_{ij}$$

where N is the number of objects in cluster a and M is the number of objects in cluster b, and a_i is analyte i in cluster a and b_j is analyte j in cluster b, and the average linkage distance between a and b is l_{ab}.

Cluster Visualization

The hierarchical, binary cluster tree defined by the distance measures were most easily understood when viewed graphically. At the first level of linkage, x-axis values between analytes were pairwise Pearson correlation coefficients. Comparisons between three or more analytes were more complex, as the linkage distances on the x-axis reflect the unweighted average Pearson correlation coefficient between all members.

Figure A1. Heat map (red: positive *z*-score; blue: negative *z*-score) of the 5+9 data set, displayed by distance (vertical axis: sampling day / subject identification; horizontal axis: endpoint).