# Microarray Myths

- That the greatest challenge is managing the mass of microarray data.

- That pattern recognition or data mining are the most appropriate paradigms for the analysis of microarray data.

- That cluster analysis is the generally appropriate method of data analysis.

- That comparing tissues or experimental conditions is based on looking for red or green spots on a single array.

- That reference RNA for two-channel arrays must be biologically relevant.

- That multiple testing issues can be ignored without filling the literature with spurious results.

- That complex classification algorithms such a neural networks perform better than simpler methods for class prediction.

- That prepackaged analysis tools are a good substitute for collaboration with statistical scientists in complex problems.

---Richard Simon

**Microarray Truths**

- The greatest challenge is organizing and training for a more multi-disciplinary approach to systems biology. The greatest specific challenge is good practice in design and analysis of microarray based experiments.

- Pattern recognition and data mining are often what you do when you don't know what your objectives are. Effective microarray based research requires clear objectives.

- Cluster analysis is useful for some types of studies such as finding potentially co-regulated genes. For most microarray studies, however, supervised methods of analysis are much more powerful.

- Comparing expression in two RNA samples tells you only about those samples, and may relate more to sample handling and assay artifacts than to biology. Robust knowledge requires multiple samples that reflect biological variability.

- The reference generally serves only to control variation in the size of corresponding spots on different arrays and variation in sample distribution over the slide.

- Comparing two classes of samples with regard to expression of 20,000 genes, one expects 1000 erroneous findings of genes that appear differentially expressed at the 5% significance level. This is true regardless of the correlation pattern of the genes. Eyeball analysis of multi-colored image plots for genes that appear differentially expressed is no more reliable.

- "Artificial intelligence" sells to journal reviewers and institute leaders who cannot distinguish hype from substance when it comes to data analysis. But comparative studies have shown that simpler methods work better for microarray problems where the number of candidate predictors greatly exceeds the number of samples.

- Biologists need both good analysis tools and good statistical collaborators. Both are in short supply.

---Richard Simon