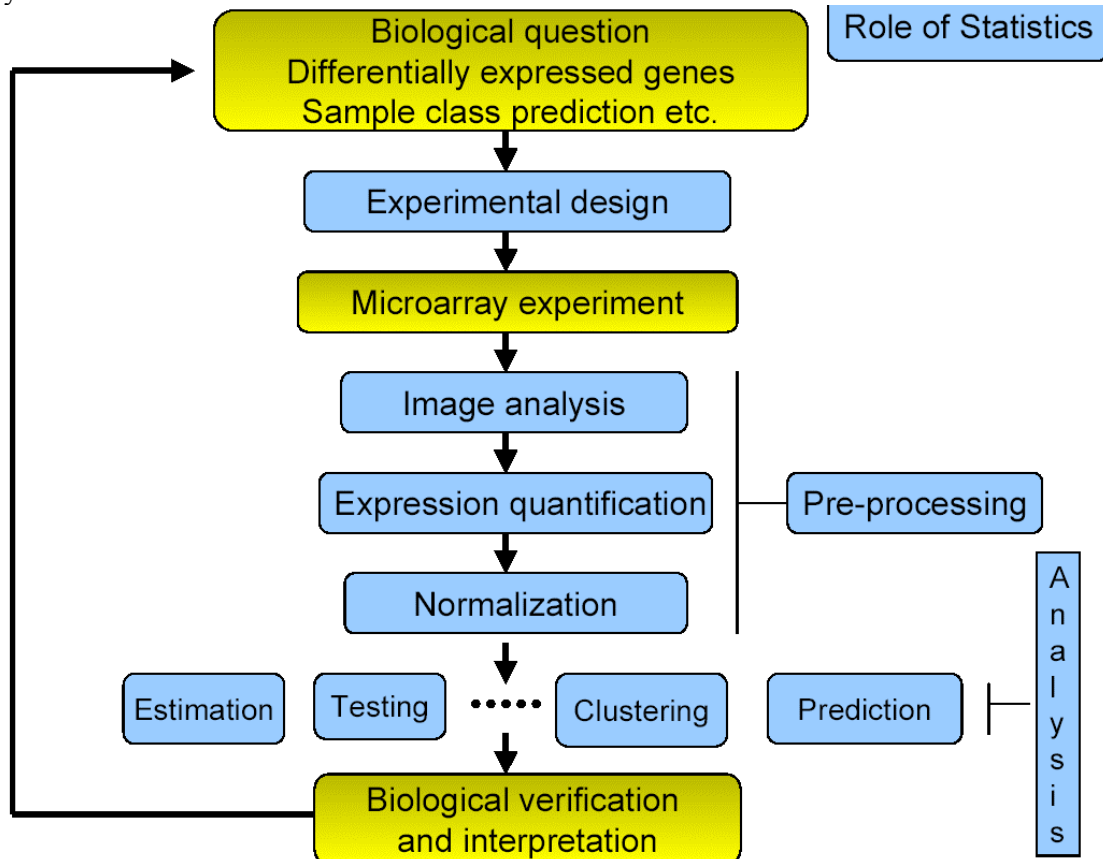# Role of Statistics in Microarray Research

Kellie J. Archer, Department of Biostatistics

Statistics plays a fundamental role in microarray research from experimental design to pre-processing issues, and including statistical analysis. The figure below demonstrates various areas in which statistical thinking is necessary.

Proper experimental design is needed to ensure that biological questions of interest can be answered and answered accurately, given the experimental constraints such as cost of reagents and availability of mRNA. Image analysis is required to translate the raw data from a microarray experiment (i.e., the set of images) into a numerical form. General issues of image analysis include addressing, the process of identifying the location or coordinates of the probes; segmentation, the process of classifying the pixels as either signal or background; and intensity extraction, the process of calculating the intensity measures for each probe. Expression quantification is the process of combining intensity values for a probe set to provide a measure of gene expression. The purpose of normalization is to remove experimental artifacts of no direct interest, that is, the removal of systematic effects other than differential expression. Normalization procedures often include background subtraction, detection of outliers, and removal of variation due to differences in sample preparation, array differences, and scanning differences. Analysis of microarray data fall into two broad classes, unsupervised and supervised learning methods. Using unsupervised learning methods, the statistical analysis seeks to find structure in the data without knowledge of class labels. Alternatively when using supervised learning methods the class or group labels are known a priori and the goal of the statistical analysis pertains to identifying differentially expressed genes (AKA feature selection) or identifying combinations of genes that are predictive of class or group membership. My research interests include (i) development of novel methods for performing image analysis, normalization, and expression summaries for oligonucleotide microarray data; (ii) development of unsupervised and supervised learning methods for microarray data analysis; (iii) development of software libraries for preprocessing steps and for microarray data analysis; and (iv) participation in the analyses of microarray data from ongoing projects.