Statistical Methods

Generating Gene Lists with Permutation F and t Tests

Michael D. Radmacher
Richard Simon

mdradmac@helix.nih.gov
Rsimon@nih.gov

A strategy for establishing differences in gene expression patterns between tumor groups is to compute lists of genes differentially expressed between known biological groups and compare the results to what would be expected under the assumption of random gene expression (i.e., no true difference between groups). Permutation F and t tests are two methods we used to create gene lists. In the permutation F test, the standard F statistic was computed on the log-expression ratios of each gene in order to analyze the variation between *BRCA1*, *BRCA2* and sporadic tumors. Next, labels (i.e*., BRCA1, BRCA2* and sporadic) were randomly permuted among the specimens and the F statistic for each gene in the permuted data set was computed. This process was repeated 10,000 times. Finally, a 0.999 critical value of the F statistic was determined for each gene based on the empirical distribution of F from permuted data sets for that gene. If the F statistic for a gene in the original labeling of specimens was larger than its critical value, the gene is differentially expressed between the three groups and is included in the gene list. The permutation t test is similar and is used when only two groups (e.g., *BRCA1* and *BRCA2*) are being tested.