ComparativeMarkerSelection Documentation

Module name:	ComparativeMarkerSelection
Description:	Compare different approaches to marker selection
Author:	Stefano Monti, Gaddy Getz, Joshua Gould (Broad Institute) gp-help@broad.mit.edu
Date:	December 10, 2004
Release:	1.0

The ComparativeMarkerSelection module includes several different approaches to determine the features that are most closely correlated with a class template and the significance of that correlation. The module outputs a file containing the following columns.

1. Rank - The rank of the feature within the dataset based on the value of the test statistic.
2. Feature - The feature name.
3. Score - The value of the test statistic.
4. Feature P - The feature-specific p value from permutation testing.
5. FPR - The false positive rate is the rate at which features that are not significant are called significant.
6. FWER – The family wise error rate is the probability that at least one feature will be falsely called significant.
7. Rank P - The rank-based p value from permutation testing.
8. FDR (BH) - An estimate of the false discovery rate by the Benjamini and Hochberg procedure. The FDR is the proportion of true null hypotheses rejected out of the total number of null hypotheses rejected.
9. Bonferroni - The value of the Bonferroni correction applied to the feature specific p value.
10. Q Value - An estimate of the FDR using the procedure developed by Storey and Tibshirani. See the definition of the FDR in (7).

The results from the ComparativeMarkerSelection algorithm can be viewed with the ComparativeMarkerSelectionViewer.

Parameters:

Name	Description	Choices
input.filename	The input file - .res, .gct, .odf	
cls.filename	The class file - .cls	
test.direction	The test to perform	Class 0;Class 1; 2 Sided
test.statistic	The statistic to use	SNR;T-Test;SNR (median);T-Test (median);T-Test (min std)
min.std	The minimum standard deviation if test statistic is T-Test (min std)	
number.of.permutations	The number of permutations to perform	100
complete	Whether to perform all possible permutations	yes;no
balanced	Whether to perform balanced	yes, no

	permutations	
fix.standard.deviation	Whether to adjust the standard deviation, as is done in GeneCluster	yes;no
output.file	The name of the output file	

Return Value: An R list with components:
1. An odf file containing the results

References:

- Golub T.R., Slonim D.K., et al. “Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring,” Science, 531-537 (1999). and the supplemental information on the website http://www-genome.wi.mit.edu/cgi-bin/cancer/publications/pub_menu.cgi for a more complete description of marker permutation testing.

- Slonim, D.K., Tamayo, P., Mesirov, J.P., Golub, T.R., Lander, E.S. (2000) Class prediction and discovery using gene expression data. In Proceedings of the Fourth Annual International Conference on Computational Molecular Biology (RECOMB) 2000. ACM Press, New York, pp. 263–272.

- Benjamini, Y., Hochberg, Y. (1995). " Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing ", *Journal of the Royal Statistical Society B,* 57 289-300.

- Storey JD and Tibshirani R. (2003) Statistical significance for genome-wide experiments. Proceedings of the National Academy of Sciences, 100: 9440-9445.

Platform dependencies:

Task type:	GeneListSelection
CPU type:	any
OS:	any
Java JVM level:	1.4
Language:	Java, R
Support files:	Jama-1.0.1.jar, broad-cg.jar, colt.jar, MarkerSelection.jar

Native command line: <java> -DR\=<R_HOME> -Dlibdir\=<libdir> <java_flags> -cp <libdir>Jama-1.0.1.jar<path.separator><libdir>broad-cg.jar<path.separator><libdir>colt.jar<path.separator><libdir>MarkerSelection.jar edu.mit.broad.marker.MarkerSelection <input.filename> <cls.filename> <number.of.permutations> <test.direction> <output.file> <balanced> <complete> <fix.standard.deviation> <test.statistic> <min.std>