

Description of FunctionScore package

Manhong Dai

July 25, 2007

Contents

1 Introduction	1
-----------------------	----------

1 Introduction

The *FunctionScore* package introduced an approach to measure the pathway/geneset activity in individual samples.

1. Class 'functionDb' represents functionally related genes and biological functions. One unique feature is that it can combine biological functions that have the same gene list into one function. This is especially helpful when many similar functions would have the same gene list after their gene list is filtered by the expression gene list.
2. Function 'read.gmt' creates a 'functionDb' object by reading a gmt file. Gmt file format detail is at http://www.broad.mit.edu/cancer/software/gsea/wiki/index.php/Data_formats#GMT:_Gene_Matrix_Transposed_file_format_.28.2A.gmt.29
3. Function 'write.gmt' writes a 'functionDb' object to a gmt file
4. Function 'calcfunctionscore', Knowledge-based gene set analysis is quickly gaining popularity due to its ability to detect subtle but coordinated expression changes in functionally related genes as well as its ability to link gene expression changes to higher level biological functions. However, prevailing gene set analysis methods are only applicable to two-class sample comparison situations. Multiple sample class analysis is not possible nor the investigation of potential heterogeneity of pathway/gene set activity among samples in the same class. In addition, existing methods mainly deal with the additive effects of multiple genes without considering higher order gene interaction effects. It is not possible to consider pathway topology in current gene set analysis methods, either. We propose a simple and

flexible approach called FunctionScore to measure the pathway/gene set activity in individual samples based on the z-score of gene expression levels. The Function score approach can easily incorporate higher order gene interactions and pathway topology. We expect the FunctionScore method can extend the knowledge-based expression analysis paradigm to meta expression data analysis, genetic genomic analysis, functional correlation between different pathways/gene sets, pathway/gene set level functional heterogeneity as well as roles of gene interactions in various pathophysiological processes.