

BBSI: Fundamentals of Microarrays

Michael F. Miles, M.D., Ph.D.

Professor

Depts. of Pharmacology/Toxicology and Neurology and

The Center for Study of Biological Complexity

Virginia Commonwealth University

e-mail: mfmiles@vcu.edu

phone: 804-225-4054

Problem Set/Questions 2: 7/29/04

A t-test is a simple statistical measure of significance. It basically asks how likely a given experimental result is expected to occur just by chance. The datasheet contains a column where t-test values have been calculated for human vs. wookie PDNN results. Filter the genes for a t-test value (“p-value”) ≤ 0.01 (1/100 likelihood of occurring by chance). This is a set of genes that have significant expression differences between wookies and humans. Using the database resources provided (see URL links for 7/28 lecture), explore whether these genes contain information about a coherent biological function:

1) You want to first determine whether this list of “wookie genes” might contain an unexpected number of genes related to a specific biological function (e.g. genes involved in glucose metabolism). Copy the “probeset ids” for the wookie genes and paste these into the EASE gene list window. (go to the “DAVID/EASE” URL, click on “EASEonline” on the sidebar under “DAVID tools”). Be sure to select the “AFFYID” button and then click on “Submit Text”. This will bring up the following options:

[Annotation Tool](#) - An automated method for the functional annotation of genome-scale datasets.

[GoCharts](#) - A visualization tool that graphically displays the distribution of differentially expressed genes among functional categories.

[KeggCharts](#) - A visualization tool that graphically displays the distribution of differentially expressed genes among metabolic pathways.

[DomainCharts](#) - A visualization tool that graphically displays the distribution of differentially expressed genes among functional protein domains.

[EASEonline](#) - Provides statistical methods for discovering enriched biological themes within gene lists. For detailed information, please see the **[EASE FAQ](#)**.

Each of these will take you to various annotation resources for the genes you entered in your list. Some databases may provide information that others don’t contain. Explore the annotations of the genes on your list. Anything interesting? Take the “EASEonline” link. This performs a statistical analysis on your list of genes in terms of how many are in which biological category. As a “background”, you must choose the type of chip used for the experiment, click the “MGU74AV2” button and then “submit”. The EASE Results will list the number of genes from your list that are in each category (“LH” column)

compared to the number of genes on the whole chip that are in the category (PH). The EASE score is like a “p-value” in that it estimates the statistical significance of the finding (score ≤ 0.05 deemed significant). Anything interesting about the categories that are significant for the wookie genes? Can you formulate a hypothesis to explain the result? How would you test your hypothesis that this biological function category has something to do with hair growth? How might you design a drug to alter how growth, given these results?

2) Take the 5 most positive genes (lowest p-values on the t-test) from the wookie gene list and explore their relationships using the “ACC link” to derive quick info on chromosomal location, function. Also submit the probeset Ids or ACC numbers for these genes to PubGene, the UCSC genome browser and WebQTL (to be demonstrated in class – but explore on your own first!!). Are there any new conclusions to be drawn from your list of wookie genes?