

Bioinf 301 Assignment #5 – GeneChip Analysis

Nov. 15, 2011

Retin-A or Trentinoin is a retinoic acid derivative that is used in various creams to treat acne. It is also used as a hair loss treatment and is a component of many commercial products that claim to slow skin aging or remove wrinkles and reduce the appearance of stretch marks. Given the impressive claims for Retin-A, a number of scientists have suggested that it may be worthwhile to use microarray experiments to figure out how Retin-A really works. At the Bioinformatics 301 website (<http://redpoll.pharmacy.ualberta.ca/bioinfo301/>), under assignment #5, you will find a zipped file of 12 Affymetrix GeneChip experiments done to investigate Retin-A activity on skin cells using a HG-U133A 2.0 array. You will also find a list of 100 potentially important genes (with their Affymetrix gene identifiers) that you will need to complete your analysis. The data were collected from 6 people (age 15-28) who were treated with Retin-A. Specifically, skin biopsies were obtained from 6 patients at baseline and at one-week after Retin-A treatment. The baseline group is called “control1, control2, etc.” while the treated group is called “treated1, treated2, etc.” This type of study is called a “two condition” study.

From this data you are expected to:

- 1) Download the Zipped file (retin_a.zip) of the 12 GeneChips from the Bioinformatics 301 website. Use the GEPAS website (<http://gepas.bioinfo.cipf.es/>) to normalize your GeneChip data. Use a text editor to look at your data to make sure it is in the correct format. You should see about 22,000 Affy gene IDs with 12 data values each. Each number corresponds to an expression value given in Log_2 units
- 2) Using the assignment #5 gene list data (genelist.txt) which is also at the Bioinformatics 301 website (~100 genes of interest with their Affymetrix IDs), extract the expression data for these 100 genes from your normalized data set (obtained in step #1). Format this 100-gene data file (using the #CLASS identifier described in GEPAS) so that you can run the data through GEPAS's Differential Expression module. Be sure to read the tutorial at the GEPAS website! This extraction step is necessary so that you won't have to wait 6 hours to get an answer using the whole set of 22,000 genes. *The 100 gene data set should only take ~10 seconds if you formatted things correctly.*
- 3) From GEPAS Differential Expression module, identify the most significant genes (of the 100 normalized gene expression values obtained in step #2) that have been upregulated or downregulated. Use tables, heat maps or other display tools as well as appropriate fold-change criteria to identify these “important” genes. Include these lists/tables/figures in your answers and explain your criteria for selection. This should be no more than 3 pages.
- 4) From your “significant gene” list obtained in step #3 use the DAVID website (<http://david.abcc.ncifcrf.gov/>) or other websites you can find to help you determine how Retin-A works and what pathways it seems to activate. Keep your explanation or supporting information to <1 page.

Assignment due on Nov. 22, 2011 by 11:00 am (10% off for each day late)