Statistical Bioinformatics Spring 2014 Final Exam

Assigned: Mon. 7 Apr.
Due: Wed. 30 Apr. (due by 8:00 P.M. - no late work will be accepted)

Take-Home Final Exam Scenario:

Suppose that you are a biostatistician working for a hypothetical client called ACME Labs. ACME Labs is interested in the individual genes and biological processes involved in growth factor response of neural stem cells in mice. The application of growth factor to a culture of neural stem cells is typically done daily, but ACME wanted to learn what would change if the application were done every two days or every four days instead. They conducted a gene expression study using three control (C) samples where growth factor was applied daily, three samples where application was every two days (T2), and three samples where application was every four days (T4). The nine resulting .CEL files are available as the GSE37516.zip data on the course website. You are to perform an appropriate statistical analysis to identify and characterize the genes of interest to ACME Labs.

General Directions:

Present your analysis in a 7-10 page summary report and accompanying interactive file (.html, summarizing all the significant candidate genes; recall slides 10-11 of Notes 4.2 – you will need to use `"mouse4302.db"` instead of `"hgu95av2.db"` in the `aafTableAnn` function, and don't just include the top 25) for the client. Include brief discussions of quality checks, choice of preprocessing method, filtering, method of identifying candidate genes, multiple testing adjustments, effective visualization and communication of results, and gene set testing (for biological process GO terms) and graph visualization. Also provide your R code in a separate file. (Note that the .zip file includes a .pdf version of the published paper corresponding to these data, but you do not need to read or understand that paper for this exam.)

The main purpose of this take-home exam is to help students organize and finalize the main concepts from this course. Note that this exam leaves considerable room for individual adaptation (including the specific comparison or contrast to make in the test for differential expression, the threshold for significance [or whether to even perform a formal test of significance], and the exact visualization tools to employ). Use appropriate grammar, spelling, and format. Submit your three final report files (7-10 page summary report file, interactive .html file, and R code file), via Canvas before the due date.

Individual Completion

In completing this exam, feel free to make use of class notes, the course text, and your previous homeworks. However, you may not discuss this exam with other students in the class until after the due date. Assistance from the instructor will be limited to clarifying the intent of the exam (and very limited R debugging), and will only be available before the last day of classes. To ensure fairness on this exam, you must include the following statement, along with your name and the date, on the cover page of your report:

"By submitting this report, I affirm that I have completed this exam independently of others and that the work contained herein is my own. I also affirm that I have not assisted any other student in this course with their final exam. I understand that any such assistance, either given or received, would result in a score of zero for this exam."

Caution:

While it is not intended to overwhelm you, this assignment should not be taken lightly. Every point of your final score must be earned by clearly addressing the necessary topics and including the necessary components of the report (see rubric below). You must use a writing style appropriate for your hypothetical client.

Statistical Bioinformatics Final Exam Grading Rubric:

Points	Portion
55	Identify (10) quality checks (15) preprocessing / filtering / DE / multiple testing / identification of candidate genes (10) effective visualization of candidate genes (20) effective communication of candidate genes (accompanying .html report file)
20	Characterize (15) gene set testing (5) visualization (graph(s))
5	Accompanying R code file
20	Grammar / spelling / format (including cover page statement)
100	