Project

**General description**: The purpose of the term project is for you to gain familiarity with the methods presented in class as you apply the methods to your specific project within a field that interests you. Additionally, the project should help solidify your understanding of such methods, and will enable you to compare/contrast key course concepts.

For your project, you are required to do a case study on a selected microarray dataset. Your project should be a complete and thorough analysis of your data, which includes application of normalization, expression summary, quality assessment, and analytical methods presented in this course to your dataset.

Prior to beginning your project, you must obtain approval for your proposed dataset from the instructor. If you need help in identifying a microarray gene expression dataset, there are several public data repositories that may be browsed including those listed in the table below.

<table>
<thead>
<tr>
<th>Data Repository</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 Critical Assessment of Microarray Data</td>
<td><a href="http://www.camda.duke.edu/camda06/datasets/">http://www.camda.duke.edu/camda06/datasets/</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.camda.duke.edu/camda04/datasets/">http://www.camda.duke.edu/camda04/datasets/</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.camda.duke.edu/camda03/datasets/">http://www.camda.duke.edu/camda03/datasets/</a></td>
</tr>
</tbody>
</table>

In selecting a dataset, you should pay critical attention to issues concerning data quality, since a sloppy dataset may leave you little time to complete the project. Please take time to talk with me about your project either during my scheduled office hours or by making an appointment.

There will be a few due dates (to be announced) throughout the semester so progress on your project can be monitored. Your paper should follow the customary format to include an Introduction, Methods, Results, Discussion and Conclusion section. Your project content will be submitted in the following phases:

**Phase I Content Components**
1. A simple and clear exposition of the dataset you will be using and the question you are addressing. This should be written in the form of an Introduction/Background section(s).

**Phase II Content Components**
1. A description of how image analysis was performed on your dataset should be included in the Methods section.
2. A detailed description of the normalization methods used to process your raw microarray data should be presented in the Methods section.
3. A discussion of the results of the normalization method applied with applicable
graphs demonstrating pre- and post-normalized data should be presented in the
Results section.
4. A description of expression summaries applied to your dataset should be
presented in the Methods section.

Phase III Content Components
1. A description of quality assessment measures should be reported in the Methods
section.
2. The Results section should include the details of the quality evaluation.

Phase IV Content Components
1. A description of the unsupervised and/or supervised learning methods used should
be described in the Methods section.
2. The results of the unsupervised and/or supervised learning methods should appear
in the Results section.
3. Summarize your findings from your analyses in the Discussion and Conclusion
sections.
4. Computational component: source and sink files for your R code as well as the
accompanying dataset must be provided to the instructor so it can be tested. Make
sure your code is readable and commented, so I can do a little troubleshooting if
necessary.

Note: Computer output should be incorporated into your report in the usual way, i.e., put
tables in the text and use graphics, but do not paste in a pile of unedited computer output.
Tables and figures should be numbered and captioned.

Final Presentation
Each student will make a 15 minute presentation of their final project, to be
followed by a 5 minute question/answer period. Final presentations are scored by
peer evaluation - any student failing to attend all student presentations will result
in a one letter grade reduction on their final presentation. Any student interested
in making a one-hour Biostatistics seminar lecture can see the instructor to get on
the calendar, and can be made in place of the in-class presentation.