You may choose to earn up to 15% of the course grade by picking a paper below and making a presentation on 20/10/11, 27/10/11, or 3/11/11.

You will be graded according to:
- the quality of your ppt (readability, organization, attractiveness)
- the quality of your presentation (organization, delivery, Q&A)
- the level of understanding of what you are presenting
- inputs from your fellow students

Background

The possibility of using gene expression profiling by microarrays for diagnostic and prognostic purposes has also generated much excitement and research in the last ten years. Nevertheless, a number of issues persist such as how to rectify batch effects (i.e., non-biological variations) [bolstad-2003], how to handle missing values [troyanskaya-2001] and, most importantly, how to identify genes that are meaningful in explaining the difference in disease phenotypes [zhang-2009]. There are three main groups of approaches, that make use of biological pathways (e.g., enzymatic pathways, gene regulatory pathways, and protein interaction networks), for improving gene selection and for transitioning from the selected genes to the understanding of the sequences of causative molecular events. The first group are the overlap analysis methods [doniger-2003, khatri-2005, zeeberg-2003], which test the significance of the intersection of differentially expressed genes with a biological pathway. The second group are the direct group analysis methods [goeman-2004, kim-2005, subramanian-2005], which test whether a biological pathway is differentially expressed as a whole. The third group are the network-based analysis methods [chuang-2007, sivachenko-2007, sohler-2004, soh-2012], which zoom into a subnetwork of a biological pathway and test whether the subnetwork is differentially expressed. All of these approaches have their basis on the fact that every disease phenotype has some underlying biological causes. Therefore, it is reasonable to analyse the gene expression profiles of disease phenotype with respect to the biological contexts provided by biological pathways and protein interaction networks.

To be presented on 20/10/11: Issues in Microarray Analysis


To be presented on 20/10/11 and 27/10/11: Overlap-Based Approaches


To be presented on 27/10/11 and 3/11/11: Direct Group Approaches


To be presented on 3/11/11: Network-Based Approaches


**Grading Scheme**

You will be graded according to:
- the quality of your ppt (readability, organization, attractiveness)
- the quality of your presentation (organization, delivery, Q&A)
- the level of understanding of what your are presenting

Your marks for the presentation will be the average of the inputs from your classmates and myself using the distribution scheme below:

<table>
<thead>
<tr>
<th></th>
<th>poor</th>
<th>ok</th>
<th>super</th>
<th>remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality of ppt</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>quality of presentation</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>level of understanding</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>gone beyond the paper assigned</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

A computational biologist often has to communicate with biologists or computer scientists who either do not have sufficient background in computing, mathematics, or biology. The inputs from your classmates are especially important for assessing whether your presentation is sufficiently clear and easily understood by such non-experts.
<table>
<thead>
<tr>
<th></th>
<th>poor</th>
<th>ok</th>
<th>super</th>
<th>remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality of ppt</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>quality of presentation</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>level of understanding</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>gone beyond the paper assigned</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>