CS2220 Introduction to Computational Biology
Student Presentations on 20/10/11, 27/10/11, and 3/11/11

This presentation contributes 15% to the course grade

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 your 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

[bolstad-2003] B. M. Bolstad, R. A. Irizarry, M. Astrand, T. P. Speed. **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics*, 19(2):185-193, 2003.

[troyanskaya-2001] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, et al. **Missing value estimation methods for DNA microarrays**. *Bioinformatics*, 17(6):520-525, 2001.

[zhang-2009] M. Zhang, L. Zhang, J. Zou, C. Yao, et al. **Evaluating reproducibility of differential expression discoveries in microarray studies by considering correlated molecular changes**. *Bioinformatics*, 25(13):1662-1668, 2009.

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

[doniger-2003, **MAPPFinder**] S. W. Doniger, N. Salomonis, K. D. Dahlquist, K. Vranizan, et al. **MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data**. *Genome Biology*, 4(1):R7, 2003.

[khatri-2005] P. Khatri, S. Draghici. **Ontological analysis of gene expression data: Current tools, limitations, and open problems**. *Bioinformatics*, 21(18):3587-3595, 2005.

[zeeberg-2003] B. R. Zeeberg, W. Feng, G. Wang, M. D. Wang, et al. **GoMiner: A resource for biological interpretation of genomic and proteomic data**. *Genome Biology*, 4(4):R28, 2003.

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

[goeman-2004, **FCS**] J. J. Goeman, S. A. van de Geer, F. de Kort, H. C. van Houwelingen. **A global test for groups of genes: Testing association with a clinical outcome**. *Bioinformatics*, 20(1):93-99, 2004.

[kim-2005, **PAGE**] S. Y. Kim, D. J. Volsky. **PAGE: Parametric analysis of gene set enrichment**. *BMC Bioinformatics*, 8(6):144, 2005.

[subramanian-2005,**GSEA**] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, et al. **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proc. Nat. Acad. Sci. USA*, 102(43):15545-15550, 2005.

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

[chuang-2007] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, T. Ideker. **Network-based classification of breast cancer metastasis**. *Molecular Systems Biology*, 3:140, 2007.

[sivachenko-2007, **NEA**] A. Y. Sivachenko, A. Yuryev, N. Daraselia, I. Mazo. **Molecular networks in microarray analysis**. *Journal of Bioinformatics and Computational Biology*, 5(2b):429-546, 2007.

[sohler-1004, **ToPNet**] F. Sohler, D. Hanisch, R. Zimmer. **New methods for joint analysis of biological networks and expression data**. *Bioinformatics*, 20(10):1517-1521, 2004.

[soh-2012.pdf, **SNet**] D. Soh, D. Dong, Y. Guo, L. Wong. **Finding consistent disease subnetworks across microarray datasets**. Manuscript, July 2011.

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:

	poor	ok	super	remarks
quality of ppt	10	20	30	
quality of presentation	10	20	30	
level of understanding	10	20	30	
gone beyond the paper assigned	0	0	10	

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.

Presenter: _____

Grader:_____

	poor	ok	super	remarks
quality of ppt	10	20	30	
quality of presentation	10	20	30	
level of understanding	10	20	30	
gone beyond the paper assigned	0	0	10	

Presenter: _____

Grader: _____

	poor	ok	super	remarks
quality of ppt	10	20	30	
quality of presentation	10	20	30	
level of understanding	10	20	30	
gone beyond the paper assigned	0	0	10	

Presenter:_____

Grader:_____

	poor	ok	super	remarks
quality of ppt	10	20	30	
quality of presentation	10	20	30	
level of understanding	10	20	30	
gone beyond the paper assigned	0	0	10	