

Syllabus
COMPUTATIONAL FUNCTIONAL GENOMICS
Class meets Tuesday and Thursday
3 Credit Hours

<u>Lectures</u>	<u>Topic</u>
Lecture 1	Basic molecular genetics, the central dogma, microarray technology, stochasticity of biological data and biological processes, examples of statistical models and inference
Lecture 2	Introduction to R, statistical inference and statistical significance, p-value, two-sample t-test, t-distribution, differential expression
Lecture 3	Identifying differentially expressed genes on a genome-wide scale, <i>alternatives to simple t-test, Welch's t-test, non-parametric rank-based test, paired t-test, multiple comparisons and the Bonferroni adjustment</i> , , effects of loess-based normalization
Lecture 4	Single-channel vs Two-channel microarrays, Paired t-test
Lecture 5	limma package, effects of subtracting the background intensities, paired t-test using limma
Lecture 6	Multiple comparisons revisited, extreme value distribution, False Discover Rates vs the traditional multiple comparison adjustments,
Lecture 7	The analysis of Affymetrics microarray data
Lecture 8	Multi-factorial experiments, experimental design, statistical power, sample size, biological vs technical replicates, generalizability of the conclusions, to pool or not to pool biological samples, paired t-test as a two-way ANOVA, extracting gene-specific dye effect using three-way ANOVA, maximum likelihood, least squared means, checking model assumptions
Lecture 9	Cluster analysis of microarray data, hierarchical clustering, partitioning methods, similarity/distance measures, linkage principles
Lecture 10	The functional annotation of differentially expressed genes, Gene Ontologies, hypergeometric distribution and the Fisher's exact test, establishing significance of the clustering results through external validation

- Lecture 11 Classifying biological samples based on the expression data, identifying informative genes, constructing a classifier, nearest neighbor classification, validation of the classifier, establishing statistical significance vs predictive accuracy
- Lecture 12 **Mid-term exam**
- Lecture 13 Conditional probabilities, Bayes theorem, Bayesian inference, empirical Bayes, pooling information from different genes
- Lecture 14 Conditional probabilities, Bayes theorem, Bayesian inference, empirical Bayes, pooling information from different genes to identify differentially expressed genes
- Lecture 15 Mixture model for clustering functional genomics data, EM algorithm, mclust, assessing the model fit
- Lecture 16 Bayesian mixtures, Gibbs sampler, the problem of choosing the “correct” model, statistical significance in cluster analysis
- Lecture 17 The product-multinomial model for identifying TF binding motifs, Gibbs sampler
- Lecture 18 The product-multinomial model for identifying TF binding motifs, Gibbs sampler
- Lecture 19 Bayesian networks and general graphical models in functional genomics

In-class mid-term exam

Final project

Instructor: Mario Medvedovic, Ph.D. (558-8564)
108.1 Kettering Laboratory
email: Mario.Medvedovic@uc.edu

Text:

Introductory Statistics with R. Peter Delgraad. Springer-Verlag, NY, 2002.

A Primer of Genome Science. Gibson, G., Muse, S.V.

Additional Textbook References:

Statistical methods in bioinformatics: an introduction / Warren J. Ewens, Gregory R. Grant. New York : Springer, 2001.

Bioinformatics : the machine learning approach / Pierre Baldi, Søren Brunak. Cambridge, Mass. : MIT Press, 2001.

Grading:

Mid-term exam:	50
Final project	<u>50</u>
Total	100

Letter grades will be assigned according to the standard procedure (90-100%=A, 80-90%=B, etc). Instructor reserve rights to adjust and/or scale scores from various the projects and assignment.

Projects

Projects will be assigned by the instructor while taking into account individual backgrounds and interests.

Office Hours: By appointment.