BST/STA 226 **Winter 2014**	**Statistical Methods for Bioinformatics**	**David M. Rocke** **January 8, 2014**

Course Information

Class Meetings:	Mon/Wed 1:10pm–2:30pm, 1143 MSB
Lab:	Mon/Wed 2:40pm–3:00pm, 1143 MSB
Office Hours:	By appointment.
Office:	140B Med Sci 1C (530-752-6999) FAX: 530-752-3239 e-mail: dmrocke@ucdavis.edu web site: http://dmrocke.ucdavis.edu/
Text:	None. Lectures, notes, papers, and data sets will be posted on line.
Computer	Please bring a laptop to class, OS X, Windows, or Linux should all be fine. A few of the data sets will be large (up to several gigabytes). We will mostly use R for analysis.
Course Grading:	Letter Grades based on 　　Homework 　　Class Attendance 　　Class Participation
Class Dates	Mondays and Wednesdays, January 8 to March 17
No Class	January 6 (would have been the first class) March 3 and 5 (I will be at a conference)
Prerequisites	Statistics 131C or the equivalent. Some experience with programming and statistical software is helpful.

　　This course is an introduction to statistical methods for biological measurement data, including of gene expression, protein abundance, and metabolite abundance. We will cover methods for one-variable analysis, as in Western blots or PCR, as well as high-throughput techniques such as gene expression arrays, multiplexed protein assays, metabolomics, and RNA-Seq. We will us a variety of statistical methods and computational approaches as demanded by the data and the biological issues.

Tentative Topical Outline

I.   What are we measuring and why?
      The Biology of Bioinformatics
      A.   DNA Sequences
      B.   RNA Transcripts
      C.   Proteins
      D.   Metabolites
      E.   Epigenetics and Gene Regulation
II.   Assay Measurement
      A.   Calibration
      B.   Standardization
      C.   Normalization
      D.   Detection limits
      E.   Error models
      F.   Interlaboratory studies
      G.   Intralaboratory studies
III. Survey of Instrumental Methods
      A.   Mass Spectrometry
      B.   Arrays
      C.   RNA-Seq
      D.   NMR Spectroscopy
      E.   ELISA and similar assays (Luminex)
      F.   PCR
      G.   Westerns, etc.
IV.   Experimental Design
      A.   Principle of Experimental Design
      B.   Factorial Designs
V.   Expression arrays
VI.   Array analysis with R
VII. Random and Mixed Effects Models
VIII. GO and gene/protein annotation
IX.   Gene groups, pathways, and inference
X.   Proteomics and Metabolomics
XI.   Significance, Prediction, and Classification
XII. Generalized linear models and RNA-Seq
XIII. Multivariate analysis
XIV. Reproducible Research