

# INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

University of Minnesota

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455

FAX (612) 626-7370

telephone (612) 624-6066

e-mail: [ima-staff@ima.umn.edu](mailto:ima-staff@ima.umn.edu)

IMA Schedules via finger: [finger\\_seminar@ima.umn.edu](mailto:finger_seminar@ima.umn.edu)

Newsletters, Updates and preprints are available via

anonymous ftp: <ftp://www.ima.umn.edu>, [www: http://www.ima.umn.edu/](http://www.ima.umn.edu/)

The IMA was founded by and receives major support from the National Science Foundation.

---

## IMA NEWSLETTER # 323

1 September–3 October 2003

2003–2004 Program

### PROBABILITY AND STATISTICS IN COMPLEX SYSTEMS

See <http://www.ima.umn.edu/complex/> for a full description of the 2003–2004 program on Probability and Statistics in Complex Systems: Genomics, Networks, and Financial Engineering

IMA schedules are subject to revision, particularly during workshops. See

<http://www.ima.umn.edu/~seminar/sched> and

<http://www.ima.umn.edu/newsletters/> for the latest scheduling information.

\*\*\*\*\*

<b>PART I: NEWS AND NOTES</b>
-------------------------------

New IMA Participating Institutions Council members

David Hoff is the new chair at Indiana University, and IMA Participating Institution. Welcome! He replaces Daniel Maki, whom we thank for his service on the council.

Welcome, also, to Nigel Higson, the new chair at Pennsylvania State University, and thanks to Gary Mullen for his help to the IMA in that post.

#### IMA Website

Comments or suggestions concerning the IMA website may be addressed to

[webmaster@ima.umn.edu](mailto:webmaster@ima.umn.edu).

In particular, we appreciate any information about World-Wide Web links appropriate to current and upcoming IMA programs.

---

PARTICIPATING INSTITUTIONS: Centrum voor Wiskunde en Informatica (CWI), Consiglio Nazionale delle Ricerche, Georgia Institute of Technology, Indiana University, Iowa State University, Kent State University, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Michigan State University, Mississippi State University, Northern Illinois University, Ohio State University, Pennsylvania State University, Purdue University, Sandia National Laboratories, Seoul National University (BK21 Math-SNU), Seoul National University (SRCCS), Texas A&M University, University of Chicago, University of Cincinnati, University of Delaware, University of Houston, University of Illinois (Urbana), University of Iowa, University of Kentucky, University of Maryland, University of Michigan, University of Minnesota, University of Notre Dame, University of Pittsburgh, University of Wisconsin, University of Wyoming, Wayne State University.

PARTICIPATING CORPORATIONS: Boeing, ExxonMobil, Ford, General Motors, Honeywell, IBM, Lockheed Martin, Lucent, Motorola, Schlumberger, Siemens, Telcordia Technologies, 3M.

Version of September 11, 2003

\*\*\*\*\*

**PART II: SCHEDULE FOR 1 SEPTEMBER–3 OCTOBER 2003**

**Monday, September 1**

Labor Day, a University of Minnesota holiday. IMA offices will be closed.

**Tuesday, September 2**

The 10:30 IMA break will be in Lind Hall 400.

**Wednesday, September 3**

2:30-4 pm      **Math 8660: Random Matrices**      Course meets in Lind Hall 409  
G. Anderson and O. Zeitouni

The 10:30 IMA break will be in Lind Hall 400.

**Thursday, September 4**

The 10:30 IMA break will be in Lind Hall 400.

**Friday, September 5**

The 10:30 IMA break will be in Lind Hall 400.

**Monday, September 8**

The 10:30 IMA break will be in Lind Hall 400.

8:30-10 am      **Math 8660: Random Matrices**      Course meets in Lind Hall 409  
G. Anderson and O. Zeitouni

**Tuesday, September 9**

The 10:30 IMA break will be in Lind Hall 400.  
Start of year meeting (release faculty, IMA staff, postdocs, and visitors)

**Wednesday, September 10**

The 10:30 IMA break will be in Lind Hall 400.

**Thursday, September 11**

The 10:30 IMA break will be in Lind Hall 400.

1-2:30 pm      **Math 8660: Random Matrices**      Course meets in Lind Hall 409  
G. Anderson and O. Zeitouni

**Friday, September 12**

The 10:30 IMA break will be in Lind Hall 400.

**IMA/MCIM INDUSTRIAL PROBLEM SEMINAR, 570 Vincent Hall:**

1:25pm      **Robert Crone**      Applied Mathematics and Disk Drive Design  
Seagate Technology, Minneapolis, MN

*Abstract:* Applied mathematics is used throughout the design process, from deriving governing equations to analyzing manufacturing data. This presentation will give an overview of the steps involved in software development, focusing attention on the applied mathematics algorithms commonly used today. Potential problems with the current formulation will also be discussed. The presentation will conclude by discussing some additional applications of applied mathematics (e.g., level sets or active contours) within the design process.

**Monday, September 15**

**IMA Tutorial:  
Tools for Model and Data Integration in the Genome  
Sciences/Information Integration Technologies for  
Complex Systems**

15–19 September 2003

Organizers: Sallie Keller-McNulty (Los Alamos National Laboratory), Michael A. Newton (University of Wisconsin-Madison), Robert Gentleman (Harvard University) Simon Tavaré (University of Southern California), Alyson Wilson (Los Alamos National Laboratory)

See <http://www.ima.umn.edu/complex/fall/t1.html>

The first four days of the tutorial, Tools for Model and Data Integration in the Genome Sciences, will provide an introduction to molecular biology and molecular technology for mathematicians and statisticians. The aim is to give a broad overview of the problems being addressed in the post-genomic era, such as array technology in human disease, mass spectrometry for protein sequencing and high throughput variation detection, regulatory networks, signal detection and comparative genomics and proteomics. The final day, on Information Integration Technologies for Complex Systems, will develop, present, and discuss a framework for the integration of experience, models and information in the highly dynamic context of contemporary scientific discovery driven by vast amounts and rapidly changing acquired information.

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

8:30      **Coffee and Registration**      Reception Room EE/CS 3-176

9:15      **Douglas N. Arnold, Scot Adams, and Organizers**      Welcome and Introduction

9:30	<b>Simon Tavaré</b> University of Southern California	Intro to Molecular Biology 1
10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Robert Gentleman</b> Harvard University	Basic Introduction to R and Its Philosophy
11:50	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Arkady Khodursky</b> University of Minnesota	Microarrays: Biological, Technological and Analytical Issues
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Robert Gentleman</b> Harvard University	Basic Introduction to R and Its Philosophy
3:50	<b>Discussion</b>	
4:00	<b>Group Photo</b>	
4:10–5:30	<b>IMA Tea and more</b>	IMA East, 400 Lind Hall
7–8 pm	<b>Public Lecture</b> <b>Leroy Hood</b> , President, Institute for Systems Biology, Seattle, Washington	After the Human Genome Project: Systems Biology and Predictive, Preventive and Personalized Medicine Smith Hall 100

*Abstract:* The Human Genome Project has catalyzed the emergence of a new approach to biology termed systems biology. Systems biology analyzes all the interrelationships of the elements in a biological system, rather than studying them one at a time, as has been the modus operandi in biology for the past 30 years. This systems approach has also emerged in the context of the view biology is an informational science and the development of high- throughput tools for capturing biological information and powerful new computational tools for analyzing it. The application of systems approaches to medicine will lead to the rise of predictive, preventive, and personalized medicine over the next 15-20 years, giving rise to a total transformation in how medicine is practiced.

I will discuss these issues and note along the way several mathematical and/or computational challenges they pose. For the speaker's biographical sketch, see <http://www.ima.umn.edu/public-lecture/2003-04/hood/index.html>.

**Tuesday, September 16**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

8:30	<b>Coffee</b>	Reception Room EE/CS 3-176
9:30	<b>Charles Geyer</b> University of Minnesota	Introduction to Markov Chain Monte Carlo
10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Simon Tavaré</b> University of Southern California	Intro to Molecular Biology 2
11:50	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Robert Gentleman</b> Harvard University	Bioconductor, with Array Analysis
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Simon Tavaré</b> University of Southern California	Genetic Variation and the HapMap Project 1
3:50	<b>Discussion</b>	
4:00	<b>Walk along the Mississippi</b>	

---

12:45-2:15 pm	<b>Math 8660: Random Matrices</b> G. Anderson and O. Zeitouni	Course meets in Physics 133
---------------	--	-----------------------------

---

Wednesday, September 17
-------------------------

All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.

8:30	<b>Coffee</b>	Reception Room EE/CS 3-176
9:30	<b>Michael A. Newton</b> University of Wisconsin-Madison	Stochastic Computation Case Study 1: Detecting Motifs, Lawrence-Liu Model
10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Robert Gentleman</b> Harvard University	Bioconductor, with Array Analysis
11:50	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Michael A. Newton</b> University of Wisconsin-Madison	Stochastic Computation Case Study 2: Phylogenetics, Bayesian Analysis of Trees
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Simon Tavaré</b> University of Southern California	Genetic Variation and the HapMap Project 2
3:50	<b>Discussion</b>	

Thursday, September 18
------------------------

All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.

8:30	<b>Coffee</b>	Reception Room EE/CS 3-176
------	---------------	----------------------------

9:30	<b>Michael A. Newton</b> University of Wisconsin-Madison	Stochastic Computation Case Study 3: Gene Expression Microarrays, Heierarchial Mixture Models
10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Robert Gentleman</b> Harvard University	Where R is Headed
11:50	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Michael A. Newton</b> University of Wisconsin-Madison	Stochastic Computation Case Study 4: Cancer Genomics
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Simon Tavaré</b> University of Southern California	Approximate Bayesian Computation
3:50	<b>Discussion</b>	
6:00	<b>Workshop Dinner</b>	Kikugawa Restaurant in Riverplace, 43 Main St. SE.

---

12:45-2:15 pm    **Math 8660: Random Matrices**                      Course meets in Physics 133  
G. Anderson and O. Zeitouni

---

**Friday, September 19**

**Special Mini-Workshop with Sallie Keller-McNulty and Alyson Wilson**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

8:30                      **Coffee**    Reception Room EE/CS 3-176

9:00	<b>Sallie Keller-McNulty</b> Los Alamos National Laboratory	Complex Multidisciplinary R&D: Characterization of the Scientific Method and Decision Processes in this Context
	<b>Greg Wilson</b> Los Alamos National Laboratory	Mapping Decision Spaces: Role of Ethnographic Methods
10:00	<b>Discussion</b>	
10:10	<b>Coffee Break</b>	Reception Room EE/CS 3-176
10:40	<b>Andrew Koehler</b> Los Alamos National Laboratory	System Representations: Graphical Models and Metaphors
11:40	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Alyson Wilson</b> Los Alamos National Laboratory	Quantification/Information Integration: Integration of Heterogeneous Information Sources through Bayesian Methods
2:30	<b>Short Break</b>	
2:40	<b>Alyson Wilson</b> Los Alamos National Laboratory	Quantification/Information Integration: Integration of Heterogeneous Information Sources through Bayesian Methods
3:40	<b>Coffee Break</b>	Reception Room EE/CS 3-176
4:10	<b>Sallie Keller-McNulty</b> Los Alamos National Laboratory	Assessment: Optimization and Resource (e.g., R&D) Allocation, including Evolutionary Methods
4:40	<b>Discussion</b>	
4:50	<b>Concluding Remarks by Organizers</b>	
5:00	<b>End of Tutorial</b>	

**Monday, September 22**

The 10:30 IMA break will be in Lind Hall 400.

**Tuesday, September 23**

The 10:30 IMA break will be in Lind Hall 400.

12:45-2:15 pm    **Math 8660: Random Matrices**                      Course meets in Physics 133  
G. Anderson and O. Zeitouni

**Wednesday, September 24**

The 10:30 IMA break will be in Lind Hall 400.

**Thursday, September 25**

12:45-2:15 pm    **Math 8660: Random Matrices**                      Course meets in Physics 133  
G. Anderson and O. Zeitouni

The 10:30 IMA break will be in Lind Hall 400.

**Friday, September 26**

The 10:30 IMA break will be in Lind Hall 400.

**IMA/MCIM INDUSTRIAL PROBLEM SEMINAR, 570 Vincent Hall:**

1:25pm            **Dipak Chowdhury**                      Title TBA  
Corning Inc, Corning, NY

Monday, September 29

**IMA Workshop:  
Statistical Methods for Gene Expression: Microarrays and  
Proteomics**

29 September – 3 October 2003

Organizers: Michael A. Newton (University of Wisconsin-Madison), Giovanni Parmigiani  
(Johns Hopkins University),

See <http://www.ima.umn.edu/complex/fall/c1.html>

Technological advances and resources created by genome sequencing projects have enabled biomedical scientists to measure precisely and simultaneously the abundance of thousands of molecular targets in living systems. The effect has been dramatic, not only for biology, where now the cellular role for all genes may be investigated, or for medicine, where new drug targets may be found and new approaches discovered for characterizing and treating complex diseases; the effect has also been dramatic for the mathematical and computational sciences. To assess a major component of this activity, the IMA workshop will focus on statistical issues in the study of gene expression both at the level of RNA transcripts and proteomics. Leading investigators will describe current research and future challenges in this area. The workshop format entails three lectures per day and extensive opportunity for fruitful discussion and interaction.

Various microarray technologies for measuring RNA transcript abundance have created some challenging statistical problems. There are many sources of variation in a typical experiment and these can be accounted for using statistical design and analysis-of-variance methodology; but careful attention has to be given to the high-dimensionality and complicated interactions. Statistical methods invoked early in the data analysis pipeline can remove systematic errors and improve subsequent inferences. Robust statistical methods are important to account for influential observations that may be hidden in massive data sets. A wide range of supervised and unsupervised learning methods have been considered to better organize data, be it to infer coordinated patterns of gene expression, to discover molecular signatures of disease subtypes, or to derive various predictions. Related efforts aim to reconstruct regulatory networks from large sets of expression data. Theoretical problems arise in statistical inference when attempting to address thousands of gene-specific hypotheses at once, such as the problem to bound the rate of false detections of differential expression. Further, research in statistical computing concerns infrastructure to enable efficient and flexible calculations with large expression data sets. The workshop will consider these and other pressing problems generated in current research.

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

8:30	<b>Coffee and Registration</b>	Reception Room EE/CS 3-176
9:15	<b>Douglas N. Arnold, Scot Adams, and Organizers</b>	Welcome and Introduction
9:30	<b>Terry Speed</b> WEHI, Melbourne and UC Berkeley	Mining a Tandem Mass Spectrometry Database to Determine the Trends and Global Factors Influencing Peptide Fragmentation

*Abstract:* A statistical and non-statistical method have been used to analyse the gas phase fragmentation behavior of protonated peptides that involves mining a database of several thousand unique product ion spectra derived from tryptic digestion and low-energy collision induced dissociation in a quadrupole ion trap mass spectrometer. This bioinformatic approach has resulted in the derivation of a relative proton mobility scale that takes into account both the charge state and the amino acid composition of a peptide, and provides an effective classification system for categorizing peptide



novel subtypes of disease.

Given the over 6000 arrays with data publicly available and the surfeit of microarray facilities, the rate-limiting steps is no longer the sample collection, hybridization, scanning, or even the analysis. Instead, the new challenge is in taking findings, such as the traditional “list of genes” resulting from a microarray analysis, and ascertaining the meaning of the results, such as the biological relationships between the genes. However, tools that link these genes back to known biological pathways, as well as discovering new pathways, are in their infancy. Tools that automatically suggest the importance of particular findings have yet to be invented.

During this presentation, I will describe four packages we have made freely available to the academic genomics community. I will present examples of and would like to discuss (1) not all pathways will be reverse engineered using microarrays, (2) looking for simultaneous gene associations ignores the fact that biology takes time, (3) a discovered diagnostic model doesn’t imply the underlying molecular physiology, (4) due to rapidly changing information about the genes already measured, one is never truly finished analyzing a microarray dataset, and (5) the real bottleneck in microarray analysis is not the analysis, but the interpretation of the findings.

2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Second Chances</b>	Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions.
3:30	<b>Group photo</b>	
3:40	<b>IMA Tea and more (with Poster Session)</b>	IMA East, 400 Lind Hall
poster session	<b>Shilpi Arora</b> Princess Margaret Hospital	Gene Expression Profiling of Human Oral Cancer Using CDNA Microarrays

*Abstract:* Oral Squamous Cell Carcinoma (OSCC) is a clinically heterogeneous disease. Patients with stage-matched tumors show differences in treatment response and outcome, suggesting that a sub classification system may be possible. In the present study, we used cDNA microarrays and a novel method of analysis (Binary Tree-Structured Vector Quantization - BTSVQ) to classify 20 OSCC samples based on their gene expression profiles. BTSVQ analysis combines k-means clustering and self-organizing maps in a complementary fashion. In our study, the binary tree generated by BTSVQ revealed groups of patients that significantly correlated with male gender (P=0.035), T III-IV disease stage (P=0.035), and nodal metastasis (P=0.035). Further data mining revealed a subset of genes present in the sample cluster that enriched for node positive tumors, and thus may represent potential biomarkers for metastasis. The differential expression of these genes was validated by quantitative real-time PCR. We conclude that molecular sub typing of OSCC can identify distinct patterns of gene expression that correlate with clinical- pathological parameters. The genes identified may influence tumor growth, development and metastasis due to the over expression of normal gene products, gene amplification or mutation. They may therefore represent potential biomarkers for oral carcinomas. Our findings may help to form the basis for a molecular classification of OSCC, thus improving diagnosis, therapeutic decisions and outcome for patients with this lethal disease.

poster session	<b>David B. Dahl</b> University of Wisconsin, Madison	Modeling Differential Gene Expression using a Dirichlet Process Mixture Model
----------------	--	---

*Abstract:* The literature has given considerable attention to the task of identifying differentially-expressed genes using data from DNA microarrays. This poster proposes a conjugate Dirichlet Process mixture model which naturally incorporates any number of treatment conditions, clusters genes based on their treatment effects and variance, and readily makes general

inference on the treatment effects and variance. As a consequence of the model, probabilities of co-regulation are available and there is no need to estimate the correct number of clusters. Any number of hypotheses concerning the parameters can be performed and false discovery rates are easily computed. The proposed methods are applied to a dataset of 10,043 genes measured at 10 treatments conditions with 3 replicates per treatment.

poster session     **Shmuel Friedland**  
                          IMA

A Theoretical Framework For Reconstructing Missing  
Data in Genome - Wide Matrix

*Abstract:* Joint with Amir Niknejad.

Since last decade, molecular biologist have been using DNA microarray(chip) as a tool for analyzing information embedded in gene expression data. During the laboratory process,some spots on the array may be missed and probing genes might fail . It is still very costly making chips to probe genes(DNA microarray).

There have been several attempts by molecular biologists,statistician, and computer scientists to recover the missing gene expressions by some ad-hoc methods.

Most recently, microarray gene expression has been formulated as a gene-array matrix .In this setting, the analysis of missing gene expression on the array would translate to recovering some missing entries in gene - expression matrix.

The most common methods for recovery are: (a) Various clustering analysis methods such as K - nearest neighbor clustering , hierarchical clustering. (b) SVD - Singular Value Decomposition.

In these methods, the recovery of missing data is done independently, i.e. the completion of each missing entry does not influence the completion of other entries.

We suggest here a new method in which the completion of missing entries is done simultaneously, i.e. the completion of one missing entry influence the completion of other entries. Our method is closely related to the methods and techniques for solving inverse eigenvalue problems.

poster session     **Boris and Dmitry Khots**  
                          Compressor Controls Corp. and U of  
                          Iowa

Why Infinite-dimensional topological groups may work  
for Genetics

*Abstract:* A significant role in mathematical modeling and algorithms for applications to processing of Genetic data (for example, Gene expression data) may play infinite-dimensional P-spaces and connected with them infinite-dimensional P-groups and P-algebras (B.S. Khots, Groups of local analytical homeomorphisms of line and P-groups, Russian Math Surveys, v.XXXIII, 3 (201), Moscow, London, 1978, 189-190). Investigation of the topological-algebraic properties of P-spaces, P-groups and P-algebras is connected with the solution of the infinite-dimensional fifth David Hilbert problem. In Genetic data processing the utilization of the topological-algebraic properties of P-spaces, P-groups and P-algebras may permit to find "Gene functionality". We applied these methods to Yeast Rosetta and Lee-Hood Gene expression data, leukemia ALL-AML Gene expression data and found the sets of Gene-Gene dependencies, Gene-Trait dependencies. In particular, accuracy of leukemia diagnosis is 0.97. On the other hand, Genetics requires a solution of new mathematical problems. For example, what are the topology-algebraic properties of a P-group (subgroups, normal subgroups, normal serieses,P-algebras, subalgebras, ideals, etc) that is finitely generated by local homeomorphisms of some manifold onto itself?

Presented by Boris Khots and Dmitriy Khots.

poster session     **Pim Kuurman**  
                          ID-Lelystad

Procedure for standardisation and normalisation of cDNA  
microarrays

*Abstract:* Expression levels for large numbers of genes under different conditions can be measured by using microarrays. In livestock species often cDNA-arrays are used for this purpose, because the complete genome sequences are not yet available to engineer oligo-arrays, and use of cDNA arrays allows the direct use of available cDNA libraries. However, cDNA-arrays exhibit larger variability than oligo arrays and therefore require more care in order to reduce noise, standardise and normalise the data, and require some different statistical approaches for analysis because two samples are measured on the same slide, unlike in oligo-array technology. This poster describes procedures developed to treat such data consisting of: (1) correction for background using special blank spots; (2) automatic outlier treatment using iteratively reweighted analysis to allow for a robust fit, similar to using medians; (3) a lowess fit to allow for dye-bias on the ratio's with varying

intensity; (4) a procedure to identify poor duplicated values (1 duplicate is made within slide) fitting a heterogeneous variance contour to allow for increasing repeatability with increasing intensity; (5) fitting of a heterogeneous variance contour for sample values to allow for decreasing variance with increasing intensity, used to provide weights for a weighted analysis. The procedure is illustrated on a data set showing differences in gene expression levels between malabsorption syndrome infected and control chickens.

M.H. Pool, W.W. Kuurman\*, B. Hulsegge, L.L.G Janss, J.M.J. Rebel, S. van Hemert. Animal Sciences Group, Wageningen UR, P.O. Box 65, 8200 AB Lelystad, The Netherlands.

poster session     **Rebecka Jornsten**  
Rutgers University

Data Depth Based Clustering and Classification

*Abstract:* Clustering and classification are important tasks for the analysis of microarray gene expression data. Classification of tissue samples can be a valuable diagnostic tool for diseases such as cancer. Clustering samples or experiments may lead to the discovery of subclasses of diseases. Clustering can also help identify groups of genes that respond similarly to a set of experimental conditions. In addition to these two tasks it is useful to have validation tools for clustering and classification. Here we focus on the identification of outliers - units that may have been misallocated, or mislabeled, or are not representative of the classes or clusters. We present two new methods: DDclust and DDclass, for clustering and classification. These non-parametric methods are based on the intuitively simple concept of data depth. We apply the methods to several gene expression and simulated data sets. We also discuss a convenient visualization and validation tool - the Relative Data Depth (ReD) plot.

poster session     **Adriana Lopez**  
University of Pittsburgh

Cancer tumor classification using gene expression data

*Abstract:* At the end of the 90's, biotechnologies such as microarrays have been developed and their use in the research of cancer has increased because they can lead to a more precise and reliable classification of cancer tumors. This research was concerned with discriminant analysis or classification of cancer tumors using expression genetic data from microarrays in previously known classes, using kernel density estimation and combination of classifiers based on this methodology. This technique was compared to other well known discriminant analysis techniques using the misclassification proportion, estimated using training and test sets: fixed and obtained by the 2:1 sampling scheme. An equally efficient performance of the fixed kernel classifiers and the adaptative kernel classifiers was observed for the three data sets that were studied and generally, the kernel classifier was the best nonparametric classifier.

poster session     **Hae-Hiang Song**  
The Catholic University of Korea

Statistical Inference Methods for Detecting Altered Gene Associations

*Abstract:* In many gene expression studies, the assumption is that knowledge of where and when a gene is expressed carries important information about what the gene does. We consider the problem of understanding the gene functions with microarray expression data of histological progressive grades, starting from dysplastic nodule in cirrhotic liver to hepatocellular carcinoma Edmonson grade III. The statistical procedures are divided into two parts: First, microarray data are suitably normalized including a method of analysis of variance (ANOVA). Much diverse comments are found for the currently used normalization methods. In order to proceed to the second part of statistical analyses of gene-pair associations, these normalization methods need first to be compared. Based on the assumption that a union set of significant genes from these normalization methods includes sufficiently general and well defined differentially expressed genes, the second part of statistical analyses of searching evidence of altered gene-gene relationships with progression of disease is carried out. Significantly altered gene-pair associations are identified with the ratio of gene-pair correlations. When we use the phrase of "difference between normal and tumor expression patterns," in a broad sense it contains not only the information summarized by the first moment of average expression levels, but also imply correlation changes between two stages, and this kind of exploration goes on to a higher order moments. The need to study association changes naturally arises when analyzing gene expression levels of multiple arrays obtained in different stages of progression. We identify altered gene-gene relationships with replicated microarray expression data.

poster session     **Mahlet Tadesse**  
Texas A & M University

A Bayesian Method for Class Discovery and Gene Selection

*Abstract:* The analysis of the high-dimensional data ( $p \gg n$ ) generated by DNA microarrays poses challenge to standard statistical methods. This has revived a strong interest in clustering algorithms. A typical goal in these analyses is the discovery of new classes of disease and the identification of relevant genes. Currently, investigators first resort to data filtering procedures or dimension reduction techniques before clustering the data. In addition, the clustering algorithms that are widely used do not provide an objective way to assess the number of classes. We propose a Bayesian method, which simultaneously identifies the number of clusters in the data and selects genes that best discriminate the different groups.

poster session	<b>Achim Tresch</b> Fraunhofer Institute for Algorithms and SC	Using Text Mining Networks for the Context Specific Interpretation of Expression Data
----------------	--	---

*Abstract:* Gene expression data are most often analysed without utilizing biomedical a-priori knowledge. The inclusion of metabolic, regulatory or protein-protein interaction networks into the analysis process itself provides a way to put results of expression experiments into a biological context. Unfortunately, network information stored in databases is oftentimes incomplete or not specific enough with respect to certain species or cell types. For this reason, we developed text mining methods for the construction of interaction networks based on biomedical free text. These methods were applied to the complete set of MEDLINE abstracts and resulted in a substantial network of protein relations. In this method, we used an automatically generated and curated gene/protein dictionary together with a biomedical grammar which defines rules to extract concepts describing relevant relations between genes/proteins and other biological entities. The resulting text mining network can be used for explorative data analysis by mapping the results of gene expression experiments onto the network. For this purpose, the ToPNet application was developed. Besides its visualization capabilities, it is able to identify sub-networks relevant according to observed expression patterns by applying a new method called =93Significant Area Search=94. Our approach was successfully applied to data from two sets of gene expression experiments in the context of epilepsy and brain cancer research.

poster session	<b>Kenny Q. Ye</b> SUNY at Stony Brook	Pooling or Not Pooling in Microarray Experiments - an Experimental Design Point of View
----------------	---	---

*Abstract:* Microarray experiments are often used to detect differences in gene expression between two populations of cells; a test population versus a control population. However in many cases, such as individuals in a population, the biological variability can present changes that are irrelevant to the question of interest and it then becomes important to assay many individual samples to collect statistically meaningful results. Unfortunately the cost of performing some types of microarray experiments can be prohibitive. A potentially effective but not well publicized alternative is to pool individual RNA samples together for hybridization on a single microarray. This method can dramatically reduce the experimental costs while maintaining high power in detecting the changes in expression levels that relate to the specific question of interest. In this talk, we will discuss why this technique works and the optimal design strategy for pooling. This idea will also be illustrated by a synthetic experiment and a real experiment that studies Afib (cardiac atrial fibrillation), a condition that is a serious health condition that affects a large percent of the population but mechanistically remains not well understood.

**Tuesday, September 30**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

9:00	<b>Coffee</b>	Reception Room EE/CS 3-176
9:30	<b>S. Dudoit</b> University of California, Berkeley	Loss-Based Estimation Methodology with Cross-Validation: Prediction of Clinical Outcomes Using Microarray Data

*Abstract:* We propose a unified loss-based methodology for estimator construction, selection, and performance assessment with cross-validation. In this approach, the parameter of interest is defined as the risk minimizer for a suitable loss function and candidate estimators are generated using this (or possibly another) loss function. Cross-validation is applied to select



2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Second Chances</b>	Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions.
4-6 pm	<b>Soccer match</b> (weather permitting)	Van Cleve Park, 15th Ave and Como Ave

---

12:45-2:15 pm	<b>Math 8660: Random Matrices</b> G. Anderson and O. Zeitouni	Course meets in Physics 133
---------------	--	-----------------------------

---

**Wednesday, October 1**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

9:00	<b>Coffee</b>	Reception Room EE/CS 3-176
9:30	<b>Keith Baggerly</b> M. D. Anderson Cancer Center	The Analysis of Proteomics Spectra from Serum Samples

*Abstract:* Mass spectrometry profiles can provide quick summaries of the relative levels of hundreds of proteins. By surveying profiles from a large number of samples, we can hopefully zoom in on proteins that are linked with a difference of interest such as the presence or absence of cancer. Using examples from two case studies, we will address issues of experimental design, data cleaning and processing, discriminating subsets, and protecting against spurious structure.

10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Yee Hwa (Jean) Yang</b> University of California, San Francisco	Comparing Normalization Methods Based on Splice Array Experiments

*Abstract:* There are many sources of systematic variation in microarray experiments that affect measured gene expression levels. Normalization is the term used to describe the process of removing such variations. In this talk, I will describe a set of experiments based on splice-specific microarrays. These arrays provide a basis to investigate the effect of mutations and other factors on splicing events in the creation of mature mRNA. In particular, the design of these arrays provides a platform for comparing the performance of different normalization methods.

11:25	<b>Discussion</b>	
11:30	<b>Wolfgang Huber</b> German Cancer Research Center	Interpretation and Transformation of Microarray Data

*Abstract:* Data from microarray experiments is often reported in the form of logarithmic ratios or logarithm-transformed intensities. This amounts to the assumption that an increase from, say, 1000 units to 2000 units has the same biological significance as one from 10000 to 20000. While this approach is useful for large intensities, it fails when the true level of expression of a gene in one of the conditions is small or zero. However, these genes may be biologically relevant, perhaps even the most relevant ones.

We derive a measure of differential expression that has comparable resolution across the whole dynamic range of expression. Mathematically, it can be expressed in terms of a variance stabilizing transformation. The measure coincides with the log-ratio in those cases where the latter is well-defined, and is a meaningful extrapolation in those cases where the log-ratio is unstable. The measure is closely related to the standardized log-ratio (“moving-window z-score”), but has more preferable mathematical and computational properties.

We present a parametric statistical model that leads to a robust estimator for the transformation parameters, as well as the between-array normalization parameters. In applications to several benchmark datasets, this approach compares favorably to other normalization algorithms.

11:55	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Peter Munson</b> National Institutes of Health	Title TBA
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Second Chances</b>	Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions.
3:30	<b>Mississippi walk</b> (weather permitting)	Begins and ends at EE/CS 3-180

**Thursday, October 2**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

9:00	<b>Coffee</b>	Reception Room EE/CS 3-176
9:30	<b>Geoff McLachlan</b> University of Queensland	Classification of Microarray Gene-Expression Data

*Abstract:* In the context of cancer diagnosis and treatment, we consider the problem of classifying a relatively small number of tumour tissue samples containing the expression data on very many (possibly thousands) of genes from microarray experiments. For the supervised problem where there are tumour samples of known classification, we discuss the need to correct for the selection bias in assessing the error rate of a prediction rule formed from a small subset of selected genes. We also consider the unsupervised problem where the aim is to cluster the tumour samples on the basis of the gene expressions. The associated problem of assessing the number of clusters is addressed. Attention is concentrated on the

mixture model-based approach called EMMIX-GENE. Its performance is demonstrated on various microarray data sets available in the bioinformatics literature.

10:20	<b>Discussion</b>	
10:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
11:00	<b>Christina Kendzierski</b> University of Wisconsin - Madison	Title TBA
11:25	<b>Discussion</b>	
11:30	<b>Michael Ochs</b> Fox-Chase Cancer Center, Philadelphia	Encoding Prior Biological Knowledge in Functional Genomics Analysis

*Abstract:* Cancer is a leading cause of death throughout the world. The fundamental cellular biology underlying the development of cancer is extremely complex, since cancer arises from a myriad of different cellular malfunctions. It is clear, however, that cellular signaling pathways that control cell growth, differentiation, apoptosis, and motility play a critical role in many cancers. New technologies such as microarrays and protein arrays offer the possibility of elucidating key pathways involved in cancer and of monitoring the effect of targeted therapeutics on those pathways. However, because of the limited nature of our knowledge of signaling pathways in humans and high noise levels in the data, difficulties arise during analysis. The inclusion of prior knowledge can enhance probabilistic reasoning in such a case. Analysis of functional genomics data is especially suitable for the inclusion of prior information, since a vast framework of biological knowledge exists.

Bayesian Decomposition is a Markov chain Monte Carlo method that uses Bayesian statistics to encode prior knowledge. The inclusion of biological information both during the analysis and when interpreting patterns identified in the data has greatly increased the power of the algorithm. This is demonstrated with three separate data sets. First, the recovery of a pattern related to the yeast mating pathway is accomplished by use of annotations from the Yeast Proteome Database. Second, tissue identification in Black6 mice is used to isolate tissue specific expression patterns that can be interpreted using gene ontology. Third, links between genes known to be coregulated in yeast is used to demonstrate the effect of such prior knowledge on the analysis.

11:55	<b>Discussion</b>	
12:00	<b>Lunch Break</b>	
1:30	<b>Marco Ramoni</b> Children's Hospital Informatics Program	Title TBA
2:20	<b>Discussion</b>	
2:30	<b>Coffee Break</b>	Reception Room EE/CS 3-176
3:00	<b>Second Chances</b>	Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions.



11:30	<b>Second Chances</b>	Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions.
11:50	<b>Concluding Remarks by Organizers</b>	
12:00	<b>End of Workshop</b>	

\*\*\*\*\*

<b>PART III: CURRENT IMA PARTICIPANTS</b>
---

FIRST YEAR POSTDOCTORAL MEMBERS

NAME	PREVIOUS INSTITUTION
Gerard Awanou	University of Georgia
Karen Ball	Indiana University
Antar Bandyopadhyay	UC Berkeley
Tim Garoni	University of Melbourne
Chuan-Hsiang Han	North Carolina State University
Lea Popovic	UC Berkeley

SECOND YEAR POSTDOCTORAL MEMBERS

NAME	PREVIOUS INSTITUTION
Olga Brezhneva	Russian Academy of Sci.
Lisa Evans	Georgia Tech
Balaji Gopalakrishnan	Georgia Tech
Herve Kerivin	University Blaise Pascal-France
Tamon Stephen	University of Michigan
Jing Wang	University of Minnesota

POSTDOCTORAL MEMBERS IN INDUSTRIAL MATHEMATICS

NAME	PREVIOUS INSTITUTION	INDUSTRIAL AFFILIATION
Yusuf Altundas	University of Pittsburgh	Schlumberger
Lili Ju	Iowa State University	VA Hospital
Haewon Nam	Texas A & M University	GE
Jun Zhao	Texas A & M University	Schlumberger

LONG TERM VISITORS

NAME	HOME INSTITUTION
Ahn, Soohan	AT&T Labs - Research
Ali, Montaz	Witwatersrand University
Anderson, Greg	University of Minnesota
Bramson, Maury	University of Minnesota
Chihara, Laura	Carleton College
Hunt, Fern	National Institute of Standards and Technology
Jain, Naresh	University of Minnesota
Kendziorski, Christina	University of Wisconsin - Madison
Khan, Mohammad Kazim	Kent State University
Kim, Dohyun	Seoul National University
Kurtz, Thomas G.	University of Wisconsin - Madison
Korea National Open University	trlee@knou.ac.kr
Mcgehee, Richard P.	University of Minnesota
Newton, Michael	University of Wisconsin - Madison
Niknejad, Amir	University of Illinois - Chicago
Rempala, Greg	University of Louisville
Scheel, Arndt	University of Minnesota
Willson, Stephen J.	Iowa State University
Zeitouni, Ofer	University of Minnesota

VISITORS IN RESIDENCE (as of 28 September 2003)

Acuna, Edgar	U of Puerto Rico at Mayaguez	9/28/03 – 10/04/03
Adams, Scot	U of Minnesota	9/01/02 – 8/31/04
Agarwal, Pujya	U of Minnesota	9/29/03 – 10/03/03
Ahn, Soohan	AT&T Labs - Research	9/01/03 – 2/29/04
Ali, Montaz	Witwatersrand U	11/01/02 – 10/31/03
Allison, David B.	U of Alabama - Birmingham	9/28/03 – 10/03/03
Altundas, Yusuf Bilgin	IMA	9/03/02 – 9/02/04
Ambrosi, Alessandro	U of Padua - Italy	9/28/03 – 10/03/03
Anderson, Greg	U of Minnesota	9/01/03 – 6/30/04
Arnold, Douglas N.	U of Minnesota	9/01/02 – 8/31/04
Aronson, Donald G.	U of Minnesota	9/01/02 – 8/31/04
Arora, Shilpi	Princess Margaret Hospital	9/28/03 – 10/04/03
Awanou, Gerard	IMA	9/02/03 – 8/31/04
Baddorf, Mary	U of Minnesota	9/29/03 – 10/03/03
Baggerly, Keith A.	U of Texas	9/28/03 – 10/03/03
Ball, Karen	IMA	9/02/03 – 8/31/04
Ballman, Karla	Mayo Clinic	9/29/03 – 10/03/03
Bandyopadhyay, Antar	IMA	9/03/03 – 8/31/04
Beyene, Joseph	Hospital for Sick Children	9/27/03 – 10/03/03
Bowman, Christopher	National Research Council, Canada	9/28/03 – 10/03/03
Bramson, Maury	U of Minnesota	9/01/03 – 6/30/04
Brezhneva, Olga	IMA	9/01/03 – 8/31/04
Brimacombe, Michael	New Jersey Medical School	9/28/03 – 10/04/03
Broet, Philippe	Universite Paris XI and INSERM U472	9/28/03 – 10/03/03
Butte, Atul	Children's Hospital Informatics Program	9/28/03 – 10/03/03
Capobianco, Enrico	[None]	9/28/03 – 10/05/03
Chandran, Uma	U of Pittsburgh	9/29/03 – 10/03/03
Chen, Meng	U of Wisconsin - Madison	9/28/03 – 10/03/03
Chen, Yidong	National Institutes of Health	9/28/03 – 10/03/03
Chihara, Laura	Carleton College	9/01/03 – 12/31/03
Chowdhury, Dipak	Corning Inc.	9/25/03 – 9/26/03
Church, Timothy R.	U of Minnesota	9/29/03 – 10/03/03
Crone, Bob	Seagate Technology	9/12/03 – 9/12/03
Cui, Yan	U of Tennessee	9/27/03 – 10/03/03
Dahl, David B.	U of Wisconsin - Madison	9/28/03 – 10/04/03
Dawson, Deborah V.	U of Iowa	9/26/03 – 10/05/03
Dayananda, P.	U of St. Thomas	9/29/03 – 10/03/03
Dewitt, Ann	3M	9/15/03 – 9/19/03
Dewitt, Ann	3M	9/29/03 – 10/03/03
Dong, Shen	U of Minnesota	9/29/03 – 10/03/03
Dudoit, Sandrine	U of California - Berkeley	9/28/03 – 10/03/03
Dunbar, Donald	Organon Laboratories LTD	9/28/03 – 10/03/03
Ein-Dor, Liat	Weizmann Institute of Science	9/29/03 – 10/03/03
El, Fayek Salah Khwsky	Alexandria U	9/28/03 – 10/03/03
Fan, Flora	U of Minnesota	9/15/03 – 9/19/03
Fan, Flora	U of Minnesota	9/29/03 – 10/03/03
Forche, Anja	U of Minnesota	9/29/03 – 10/03/03
Friedland, Shmuel	U of Illinois - Urbana-Champaign	9/01/03 – 6/30/04
Gaffney, Patrick M.	U of Minnesota	9/29/03 – 10/03/03
Garoni, Tim	IMA	8/25/03 – 8/31/04
Garrett, Elizabeth S.	Johns Hopkins U	9/28/03 – 10/03/03
Gatsonis, Constantine	Brown U	9/29/03 – 10/03/03
Gentleman, Robert	Harvard School of Public Health	9/14/03 – 9/19/03
Gieger, Christian	Fraunhofer Institute	9/28/03 – 10/04/03

Glenn, Nancy	U of South Carolina	9/28/03 – 10/03/03
Gopalakrishnan, Balaji	U of Minnesota	9/03/02 – 9/02/04
Grindle, Suzanne M.	U of Minnesota	9/29/03 – 10/03/03
Hajek, Bruce	U of Illinois - Urbana-Champaign	9/01/03 – 8/31/04
Hall, Jennifer	U of Minnesota	9/29/03 – 10/03/03
Han, Chuan-Hsiang	IMA	9/02/03 – 8/31/04
Hanin, Leonid	Idaho State U	9/28/03 – 10/03/03
Hawkins, Douglas M.	U of Minnesota	9/29/03 – 10/03/03
Hebbel, Robert	U of Minnesota	9/29/03 – 10/03/03
Hood, Leroy E.	Institute for Systems Biology	9/15/03 – 9/16/03
Huber, Wolfgang	National Cancer Research Inst, Germany	9/27/03 – 10/03/03
Huebner, Marianne	Michigan State U	9/27/03 – 10/03/03
Hunt, Fern	National Inst of Standards and Technology	9/14/03 – 10/03/03
Irizarry, Rafael	Johns Hopkins U	9/28/03 – 10/03/03
Jain, Naresh	U of Minnesota	9/01/03 – 6/30/04
Jain, Sonia	U of California - San Diego	9/28/03 – 10/04/03
Ji, Yuan	U of Wisconsin - Madison	9/28/03 – 10/03/03
Jiang, Aixiang	U of Wisconsin	9/29/03 – 10/03/03
Joo, Kisee	U of Kentucky	9/28/03 – 10/04/03
Jornsten, Rebecka	Rutgers	9/27/03 – 10/03/03
Ju, Lili	IMA	9/03/02 – 9/02/04
Kandil, Mahrous	U of Minnesota	9/29/03 – 10/03/03
Kazi, Sabera	Honeywell	9/15/03 – 9/19/03
Keller-Mcnulty, Sallie	Los Alamos National Laboratory	9/18/03 – 9/19/03
Kendziorski, Christina	U of Wisconsin - Madison	9/02/03 – 12/31/03
Kerivin, Herve	U of Minnesota	9/03/02 – 9/02/04
Kerr, Kathleen	U of Washington	9/27/03 – 10/01/03
Khan, Mohammad Kazim	Kent State U	9/02/03 – 6/30/04
Khots, Boris	Compressor Controls Corporation	9/28/03 – 10/04/03
Khots, Dmitriy	U of Iowa	9/28/03 – 10/04/03
Kim, Dohyun	Seoul National U	9/01/03 – 2/29/04
Kim, Seongjai	U of Kentucky	9/28/03 – 10/04/03
Kumar, Ashish	U of Minnesota	9/29/03 – 10/03/03
Kurtz, Thomas G.	U of Wisconsin - Madison	9/01/03 – 5/31/04
Kuurman, Pim	ID-Lelystad	9/28/03 – 10/04/03
Lahiri, Soumendra Nath	Iowa State U	9/14/03 – 9/19/03
Lahiri, Soumendra Nath	Iowa State U	9/28/03 – 10/03/03
Lande, Jeff	U of Minnesota	9/29/03 – 10/03/03
Lazzeroni, Laura	Stanford U	9/28/03 – 10/03/03
Lee, Mei-Ling Ting	Harvard U	10/01/03 – 10/03/03
Lee, Taerim	Korea National Open U	9/12/03 – 10/12/03
Li, Hongzhe	U of California - Davis	9/28/03 – 10/03/03
Li, Qingsheng	U of Minnesota	9/15/03 – 9/19/03
Li, Wentian	North Shore - LIJ Research Institute	9/28/03 – 10/03/03
Lopez, Adriana	U of Pittsburgh	9/27/03 – 10/04/03
Lynch, Jessica	U of Minnesota	9/29/03 – 10/03/03
Ma, Changqing	U of Pittsburgh	9/28/03 – 10/03/03
Maki, Daniel	Indiana U	9/14/03 – 9/19/03
Mallick, Bani K.	Texas A & M U	9/29/03 – 10/03/03
Mcgehee, Richard P.	U of Minnesota	9/01/03 – 6/30/04
Mclachlan, Geoff	U of Queensland	9/27/03 – 10/04/03
Medvedovic, Mario	U of Cincinnati Medical Center	9/28/03 – 10/03/03
Morlan, Bruce W.	Mayo Clinic	9/28/03 – 10/04/03
Munson, Peter	National Institutes of Health	9/28/03 – 10/03/03
Nam, Haewon	IMA	9/03/02 – 9/02/04
Newton, Michael	U of Wisconsin - Madison	9/01/03 – 12/31/03

Niknejad, Amir	U of Illinois - Chicago	9/01/03 – 5/30/04
Oberg, Ann L.	The Mayo Clinic and Foundation	9/28/03 – 10/03/03
Ochs, Michael	Fox-Chase Cancer Center	9/28/03 – 10/05/03
Pan, Wei	U of Minnesota	9/28/03 – 10/03/03
Parmigiani, Giovanni	Johns Hopkins U	9/28/03 – 10/03/03
Popovic, Lea	IMA	9/02/03 – 8/31/04
Quackenbush, John	The Institute for Genomic Research	9/29/03 – 10/03/03
Raghavan, Arvind	U of Minnesota	9/15/03 – 9/19/03
Raghavan, Arvind	U of Minnesota	9/29/03 – 10/03/03
Raghavan, Shalini	3M	9/15/03 – 9/19/03
Ramoni, Marco	Children's Hospital Informatics Program	9/28/03 – 10/03/03
Rao, J. Sunil	Case Western Reserve U	9/28/03 – 10/03/03
Rejto, Lidia	U of Delaware	9/27/03 – 10/03/03
Rempala, Greg	U of Louisville	9/02/03 – 6/30/04
Rocke, David M.	U of California - Davis	9/28/03 – 10/03/03
Rosa, Guilherme J. M.	Michigan State U	9/28/03 – 10/03/03
Rosen, Michal Zvi	U of California - Berkeley	9/29/03 – 10/03/03
Rosenberger, James L.	Pennsylvania State U	9/28/03 – 10/03/03
Saha, Suhas	U of Minnesota	9/15/03 – 9/19/03
Salcedo, Mauricio	Hospital de Oncologia	9/28/03 – 10/04/03
Santosa, Fadil	U of Minnesota	9/01/02 – 8/31/04
Scheel, Arndt	U of Minnesota	9/01/03 – 6/30/04
Sebastiani, Paola	U of Massachusetts - Amherst	9/28/03 – 10/03/03
Segal, Mark R.	U of California - San Francisco	9/28/03 – 10/03/03
Shaffer, Juliet Popper	U of California - Berkeley	9/28/03 – 10/03/03
Shannon, William D.	Washington U School of Medicine	9/28/03 – 10/03/03
Song, Hae-Hiang	The Catholic U of Korea	9/28/03 – 10/04/03
Speed, Terence P.	U of California - Berkeley	9/27/03 – 10/02/03
Srivastav, Sudesh K.	Tulane U	9/28/03 – 10/03/03
Stephen, Tamon	U of Minnesota	9/03/02 – 9/02/04
Storey, John D.	U of California - Berkeley	9/28/03 – 10/03/03
Tadesse, Mahlet G.	Texas A & M U	9/28/03 – 10/05/03
Tavare, Simon	U of Southern California	9/01/03 – 8/31/04
Tetrick, Kenneth	[None]	9/14/03 – 9/19/03
Theilhaber, Joachim	Aventis Pharmaceuticals	9/28/03 – 10/03/03
Van, Mark J. Laan	U of California - Berkeley	9/28/03 – 10/03/03
Vannucci, Marina	Texas A & M U	9/28/03 – 10/04/03
Wang, Jing	U of Minnesota	9/03/02 – 9/02/04
Wang, Lin	U of Minnesota	9/29/03 – 10/03/03
West, Mike	Duke U	9/28/03 – 10/03/03
Willson, Stephen J.	Iowa State U	9/02/03 – 12/31/03
Wilson, Alyson	Los Alamos National Laboratory	9/14/03 – 9/20/03
Wright, Fred	U of North Carolina	9/27/03 – 10/03/03
Xu, Jie	U of Minnesota	9/15/03 – 9/19/03
Yakovlev, Andrei	U of Rochester	9/28/03 – 10/03/03
Yang, Jean Yee Hwa	U of California - San Francisco	9/28/03 – 10/03/03
Yang, Yuhong	Iowa State U	9/01/03 – 6/30/04
Ye, Kenny	State U of New York - Stony Brook	9/28/03 – 10/03/03
Yu, Haoyu	U of Minnesota	9/29/03 – 10/03/03
Yuan, Ming	U of Wisconsin - Madison	9/28/03 – 10/03/03
Zeitouni, Ofer	U of Minnesota	9/01/03 – 6/30/04
Zhang, Yongqing	U of Minnesota	9/14/03 – 9/19/03
Zhao, Jun	IMA	9/03/02 – 9/02/04
Zhao, Lue Ping	Fred Hutchison Cancer Research Center	9/28/03 – 10/03/03

See also URL: <http://www.ima.umn.edu/people/>