San Francisco Bay Area Chapter of the American Statistical Association Symposium

Statistical Genetics, Genomics, and Proteomics

California State University East Bay, Hayward Hills
Valley Business Technology (VBT) Center, Room 124
Saturday, May 14, 2011 Registration: 12:30pm – 1:00pm Symposium: 1:00pm – 5:40pm

1: 00 - 1:15 Welcome and Program Announcement

Session I 1:15 - 3:15

1: 15 - 1:55

David M. Rocke, Distinguished Professor of Biostatistics, UC Davis
Serum Glycan Biomarkers for Diagnosis and Prognosis

Biography

David M. Rocke is Distinguished Professor of Biostatistics and Biomedical Engineering, University of California, Davis. He is Director of Biostatistics in the Clincal and Translational Science Center, and Director of the Center for Biomarker Discovery. He directs a research group aimed at bioinformatics and data analysis of gene expression arrays, proteomics, metabolomics and other high-throughput biological assays. He also has a funded experimental program in radiation biology, directed at molecular mechanisms of response in human skin to low- and moderate-dose ionizing radation. He is the author or co-author of two scholarly books and over 150 scientific papers. His work has been recognized with several honors including Fellowship, the Award for Interlaboratory Testing, and the Statistics in Chemistry Award from the American Statistical Association; the Youden Prize and the Shewell Award from the Chemical Division of the American Society for Quality Control; and Fellowship in the American Association for the Advancement of Science. He is also an elected member of the International Statistical Institute. His work is currently supported by several sources including the National Institutes of Health and the US Department of Energy.

Abstract

Serum biomarkers are potentially important because they do not require highly invasive procedures. Protein assays have been one important route to discovery of biomarkers for diagnosis and prognosis. Many human proteins (probably over half) are glycosylated, and many biomarkers of cancer are glycosylated proteins. In the glycomics approach described in this presentation, we examine the glycans (sugars) adducted to proteins in serum to discover those that are differentially present in patients with cancer compared to normal controls, or to patients with other conditions. A particularly strong study design is to analyze samples from patients presenting with possible tumors and attempt to discover predictive patterns. This can be done for example for women presenting with suspicious ovarian masses, or for men with high PSA levels. In each case, both malignant and benign diseases are possible.

The samples are analyzed by extracting the glycans and measuring the relative concentrations using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), which has potentially high mass accuracy. An R package, FTICRMS, has been provided to process the spectra and identify glycans that differ between conditions. Beginning with the raw spectra, this involves mass calibration, baseline correction, data transformation, normalization, peak identification, peak quantification, and data analysis. Applications of this methodology to ovarian, breast, and prostate cancer are described.

1:55 - 2:35

Michael R. Crager, Fellow, Biostatistics, Genomic Health, Inc.

Gene Identification Using True Discovery Rate Degree of Association (TDRDA) Sets and Estimates Corrected for Regression to the Mean.

Biography

Michael Crager received his B.A. in Mathematics from the University of Denver in 1978 and his Ph.D. in Statistics from Stanford University in 1982. He worked as a statistician in the pharmaceutical industry for 27 years in positions of increasing responsibility at G.D. Searle in Chicago, Syntex Research, Roche Palo Alto, InterMune, and CV Therapeutics. He is currently Fellow, Biostatistics at Genomic Health, Inc., where he consults with staff and develops innovative statistical methodologies to address the many technical challenges of developing genomic markers from tumor tissue for assessing risk of cancer recurrence and prediction of efficacy of prophylactic therapies. Research interests include false discovery rate methodologies, multivariate prediction and predictive model selection.

Abstract

Analyses for identifying genes with expression associated with clinical outcome or state are often based on ranked *p*-values from tests of point null hypotheses of no association. Van de Wiel and Kim (2007) take the innovative approach of testing the interval null hypotheses that the degree of association for a gene is less than some value of interest against the alternative that it is greater. Combining this idea with the false discovery rate (FDR) controlling methods of Storey, Taylor and Siegmund (2004), I will show a computationally simple way to identify true discovery rate degree of association (TDRDA) sets of genes among which a specified proportion are expected to have an absolute association of a specified degree or more. From this analysis, genes can be ranked using the maximum lower bound (MLB) degree of association for which each gene belongs to a TDRDA set. Estimates of each gene's degree of association with approximate correction for "selection bias" due to regression to the mean (RM) are derived using simple bivariate normal theory and Efron and Tibshirani's (2002) empirical Bayes approach. TDRDA sets, the gene ranking and the RM-corrected estimates of degree of association can be displayed graphically. I will illustrate these methods using clinical-genomic studies of of the association of gene expression with time to cancer recurrence analyzed using standardized hazard ratios from Cox proportional hazards regression.

References

Crager MR (2010). Gene identification using true discovery rate degree of association sets and estimates corrected for regression to the mean. Statistics in Medicine **29:** 33–45. DOI: 10.1002/sim.3789

Efron B and Tibshirani R (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology* **23**:70–86.

Storey JD, Taylor JE, Siegmund D (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B* **66**, 187–205.

Van de Wiel MA and Kim KI (2007). Estimating the false discovery rate using nonparametric deconvolution. *Biometrics* **63**: 806–815.

2:35 - 3:15

Jane Fridlyand, Senior Statistical Scientist, Genentech

Realities of Drug-diagnostic Co-development - The Notes from the Battlefield

Biography

Jane got her PhD from UC Berkeley under Dr. Terry Speed where she worked on questions in high throughput genomic analysis (gene expression, SNP and genome sequence data). She took part in Human Genome Project and did some very early work on cancer gene expression in human and mice. After completed a post-doc training at UCSF cancer center working on methods for the analysis of the copy number data, she became a faculty in the department of Biostatistics at UCSF where she continued to work on the question of data integration of high throughput genomic data in cancer patients. For the past 4 years she has been working on oncology drugs at Genentech in Early Clinical Development.

Abstract:
In the past decade, our understanding of biology underlying cancer development has dramatically increased. Many new pathways and pathway nodes have been identified and are being studied as possible targets for new cancer therapies. We no longer think of cancer only in terms of its primary site (e.g. lung, breast, or colon) or in terms of histology (e.g. squamous, adeno or small cell), but also in terms of the tumor's specific genetic alterations (e.g. K-Ras mutation, EGFR over-expression or mutation, HER2 amplification). At the same time new technologies have advanced new assay methodologies, allowing for interrogation of smaller amounts of tumor tissue and/or cancer cells circulating in the blood. These new advances in basic research raise hopes for new, efficacious, less toxic targeted cancer treatments. Development of these new treatments however, cannot follow the usual established path; new clinical strategies and approaches are required for the co-development of the novel therapeutics and the diagnostic.

In this talk we will review the approaches to, and challenges of drug development strategies incorporating predictive biomarkers into clinical programs. We will discuss the impact of the "biomarker boom" on all phases of drug development: from cell lines and xenograft models, Phases I, II and III, and eventual label; and address the statistical, strategic, regulatory and operational challenges. We will conclude with some open philosophical questions that drug makers are facing.

3:15 - 3:40 Break and Refreshment

Session II 3:40 - 5:40

3:40 - 4:20
Trevor Hastie, Professor, Statistics, Stanford University
Learning with Sparsity

Biography
Trevor Hastie was born in South Africa in 1953. He received his university education from Rhodes University, South Africa (BS), University of Cape Town (MS), and Stanford University (Ph.D Statistics 1984).

After graduating he returned to South Africa for a year, and then returned in March 1986 and joined the statistics and data analysis research group at what was then AT&T Bell Laboratories. After 9 enjoyable years at Bell Labs, he returned to Stanford University in 1994 as Professor in Statistics and Biostatistics.

His main research contributions have been in the field of applied nonparametric regression and classification, and he has written two books in this area: "Generalized Additive Models" (with R. Tibshirani, Chapman and Hall, 1991), and "Elements of Statistical Learning" (with R. Tibshirani and J. Friedman, Springer 2001). He has also made contributions in statistical computing, co-editing (with J. Chambers) a large software library on modeling tools in the S language ("Statistical Models in S", Wadsworth, 1992), which form the basis for much of the statistical modeling in R and S-plus. His current research focuses on applied problems in biology and genomics, medicine and industry, in particular data mining, prediction and classification problems.

Abstract
Many problems in machine learning have to deal with wide data - many more features than observations. Most of the features are of no use, and even the useful ones are often too sparse. For these problems L1 regularization and its variants have proven to be useful for both feature selection and complexity control. This talk is a review of a number of topics in this area, with a focus on computational aspects. This is joint work with Jerome Friedman, Rob Tibshirani, and our past and present students

4:20 - 5:00
Andrew Kasarskis, Senior Director, Pacific Biosciences
Applications and Peculiarities of Third Generation Sequence Data

Biography

Andrew Kasarskis currently serves as Senior Director at Pacific Biosciences and manages a portfolio of initiatives that exploit the company's Molecule Real Time (SMRT^{TM}) sequencing technology to gain new insight into biological systems. Prior to Pacific Biosciences, Andrew was Scientific Director of Genetics at Merck Research Laboratories. The Merck Genetics Department he built and ultimately led worked across clinical, preclinical, and basic research to deliver novel drug targets and biomarkers and was grounded in a systems approach to reconstructing the gene networks underpinning human biology and disease. Andrew also exported this same approach through his contributions to the launch of Sage Bionetworks, a not-for-profit medical research organization dedicated to advancing human health by open application of systems biology to biomedical research.

Andrew has focused his career on addressing urgent biological problems by developing and applying new technology. From 1998-2000 he contributed to the development of the Gene Ontology, Stanford Microarray Database, and Saccharomyces Genome Database while a member of the Genome Databases Group at Stanford University. In 2000, he moved to DoubleTwist, Inc., where he served as Senior Product Manager and was responsible for the commercial success and scientific validity of the Prophecy Genome Database and software suite. In 2002, he joined Merck Research Laboratories at the Rosetta site in Seattle, WA as a Senior Scientist in Informatics. During his seven-year tenure at Merck he had responsibility for a variety of research and software initiatives, all focused on the generation, ntegration, and effective use of complex orthogonal data sets to create increasingly predictive models of human disease.

Andrew holds a Ph.D. in Molecular and Cellular Biology from UC Berkeley as well as a B.S. in Biology and a B.A. in Chemistry
from the University of Kentucky.

Abstract:

Pacific Biosciences' single molecule real time (SMRT) sequencing fully exploits the high catalytic rates and processivity of DNA polymerase to radically increase the rate of synthesis and read length of sequencing studies. SMRT synthesis rates are more than 10,000 times faster than leading second-generation technologies and SMRT reads are routinely thousands of base pairs. Samples can be prepared for SMRT sequencing and sequenced within the same typical workday that the sequence analysis is completed, all at moderate cost. SMRT technology therefore lends itself to applications that leverage granular, prompt sample processing and long sequence reads, including pathogen identification, genome assembly, transcriptome characterization, and targeted resequencing, all of which present interesting computational and statistical problems. Central to the utility of Pacific Biosciences sequence data is that each sequence read is a measurement of a single DNA molecule. This means that there is an opportunity to perform statistical analysis of not only the base sequence, but also the kinetics of polymerization and modification status across a population of individual molecules.

5:00 - 5:40

Zhaoshi Jiang, Computational biologist, Department of Bioinformatics and Computational Biology, Genentech

The Effects of Hepatitis B Virus Integration in the Genome of Hepatocellular Carcinoma Patients

Biography

Zhaoshi holds a PhD in Genomic Sciences from University of Washington, Seattle. Before that he completed 6 years of resident training in clinical pathology at Peking Union Medical College Hospital after receiving his Bachelor of Medicine from ZheJiang University in HangZhou China. He now works as a computational biologist at Genentech. His research interest includes but not limited to Genomics, Bioinformatics, Cancer Biology and anti-cancer drug discovery.

Abstract

Zhaoshi Jiang[1*], Suchit Jhunjhunwala[1*], Jinfeng Liu[1], Peter Haverty[1], William Lee[1], Krishna Pant[2], Michael I. Kennemer[2], Paolo Carnevali[2], Yinghui Guan[3], Jeremy Stinson[3], Peter Dijkgraaf[3], Julie Rae[4], Stephanie

Johnson[4], Colin Watanabe[1], Jingyu Diao[5], Sharookh Kapadia[5], Fred de Sauvage[3], Robert Gentleman[1], Howard Stern[4], Sekar Seshagiri[3], Zora Modrusan[3], Dennis Ballinger[2], Zemin Zhang1[1§]

1. Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, California 94080, USA
2. Complete Genomics Inc., Mountain View, California 94043, USA
3. Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080, USA
4. Department of Pathology, Genentech Inc., South San Francisco, California 94080, USA
5. Department of Microbial Pathogenesis, Genentech Inc., South San Francisco, California 94080, USA

More than 350 million people are infected by hepatitis B virus (HBV) worldwide, causing an estimated 320,000 deaths annually. Approximately 30-50% HBV-related deaths are due to hepatocellular carcinoma (HCC). The association of HBV infection with HCC development is evident by significantly increased incidence of HCC among HBV chronically infected populations. Despite clear evidence supporting the involvement of HBV in HCC carcinogenesis, the underlying mechanisms remain elusive. Recent advances in sequencing technologies provide an opportunity to systematically investigate impact of viral integration in host genomes. Here we sequenced the entire genomes (>80-fold coverage) and transcriptomes of HCC and matched non-neoplastic samples (adjacent liver and/or peripheral blood cells) from four HCC patients. The data that we report here provide the most comprehensive picture of the somatic alteration landscape of HCC genomes and reveal the widespread impact of HBV integration during HCC development.

* These authors contributed equally to this study § Corresponding author

Click here for a pdf of abstracts and biographies.

Valley Business Technology (VBT) Center, Room 124,
California State University East Bay, Hayward
25800 Carlos Bee Boulevard, Hayward, CA94542

Chair: Kit Fun Lau, Ph.D., i3 Statprobe; Ruixiao Lu, Ph.D., Novartis
Local Hosts: Eric Suess, Ph.D.; Lynn Eudey, Ph.D.; CSUEB Stat/Biostat

Cost: Free

To reserve a seat, registration by May 12 is highly recommended. Please register by replying to asagenomics@yahoo.com with your name, affiliation, and email address using subject line = "register asagenomics". Registration will close by May 12th or when the conference room capacity is reached.

Driving directions:

From San Francisco Bay Bridge
Cross the Bay Bridge and get on 880 South. Exit the Jackson St. East turnoff in Hayward. As you come off the freeway go to the first signal and make a right turn on Santa Clara. Santa Clara will turn into Harder Road. Follow Harder Road 1-1/2 miles to the University.
From the San Mateo Bridge (Highway 92)
Heading east on the San Mateo bridge, highway 92 turns into Jackson St. As you come off the freeway go to the first signal and make a right turn on Santa Clara. Santa Clara will turn into Harder Road. Follow Harder Road 1-1/2 miles to the University.
From Oakland Highway 880 south
Follow 880 to the Jackson St. East turnoff in Hayward. As you come off the freeway go to the first signal and make a right turn on Santa Clara. Santa Clara will turn into Harder Road. Follow Harder Road 1-1/2 miles to the University.
From Oakland Highway 580 south

Follow Highway 580 to Hayward exiting at the 238 / South Hayward turnoff. This brings you onto Foothill Blvd. Follow Foothill, saying in the left lane. You will reach a major intersection, follow signs that say Mission Blvd. Follow Mission Blvd. to Carlos Bee Blvd. Make a left turn there and stay in the right lane. Cal State is at the top of the hill.

From San Jose, Fremont, Union City and Surrounding Areas via Highway 880

From 880 North, take the Jackson St. east turnoff in Hayward. As you come off the freeway go to the first signal and make a right turn on Santa Clara. Santa Clara will turn into Harder Road. Follow Harder Road 1-1/2 miles up the hill to the University.

From Palo Alto and the Surrounding Areas via the Dumbarton Bridge

Get on the Dumbarton Bridge heading east and then take 880 north. Follow 880 north and get off at the Jackson St. east turnoff in Hayward. As you come off the freeway go to the first signal and make a right turn on Santa Clara. Santa Clara will turn into Harder Road. Follow Harder Road 1-1/2 miles to the University.

From Walnut Creek and San Ramon Areas via Highway 680 South

Take 680 south to 580 west. Take the Castro Valley turnoff. As you come off the freeway make 3 immediate left turns (following the Hayward signs) this will bring you heading down Center Street. At the bottom of the hill, at the next light make a right turn onto "B" Street. Follow "B" Street to Mission Blvd., turn left on Mission Blvd. Follow Mission Blvd. to Carlos Bee Blvd. Make a left turn there and stay in the right lane. Cal State is at the top of the hill.

Campus Map:

Click here (pdf) or go to

http://www20.csueastbay.edu/about/visitor-information/maps-campus-locations/hayward-campus-map/index.html

Please look for the sign “Genomics Symposium” to go to the conference room in Valley Business Technology Center.

Parking: Parking permits are required.

- Daily/hourly permits on Saturday are $5 daily or $2 hourly. These permits, obtainable from machines located in Lots E, G, K, N, and the Harder Road kiosk, authorize parking privileges in "General" parking lots when properly displayed. All dispensers take coins and bills. Dispensers in Lots E2 & G accept credit cards as well. Parking is authorized in "Faculty/Staff" parking lots as well on weekends.
- More information on campus parking can be found at

http://www20.csueastbay.edu/af/departments/parking/parking-trans.html

Return to Bay Area ASA Homepage