

# Andrew B. Nobel

Professor of Statistics and Biostatistics  
Department of Statistics and Operations Research  
University of North Carolina  
Chapel Hill, NC 27599-3260

Fax: 919.962.1279  
Telephone: 919.962.1352  
E-mail: nobel@email.unc.edu

## Academic Appointments

Professor, Department of Statistics and Operations Research, UNC Chapel Hill, 2006 -  
Professor, Department of Biostatistics, UNC Chapel Hill, 2007 -  
Full Member, Lineberger Comprehensive Cancer Center, UNC Chapel Hill  
Faculty Fellow, Statistics and Applied Mathematical Sciences Institute, 2004/5, 2013/14  
Adjunct Professor, Department of Computer Science, UNC Chapel Hill, 2006 - 2010  
Assistant/Associate Professor, Department of Statistics, UNC Chapel Hill, 1994 - 2006  
Visiting Associate Professor of Statistics, University of Chicago, Fall Quarter, 2000  
Visiting Assistant Professor of Mathematics, Université de Paris-Sud, Orsay, May-June 2000  
Postdoctoral Fellow, Beckman Institute for Advanced Science and Technology,  
University of Illinois Urbana-Champaign, 1992 - 1994

## Education

Ph.D., Electrical Engineering, Stanford University, 1992  
Masters, Electrical Engineering, Stanford University, 1988  
Certificate of Study, Mathematics, Cambridge University, England, 1986  
Bachelor of Science (with honors), Electrical Engineering, Cornell University, 1985

## Honors

Fellow of the Institute of Mathematical Statistics (IMS), 2008  
Lucent Distinguished Lecturer, Department of Electrical Engineering and  
Computer Science, University of Michigan, Ann Arbor, March, 2002  
National Science Foundation CAREER Grant, 1995  
Beckman Institute Fellow, University of Illinois U-C, 1992-1995  
IBM Doctoral Fellow, 1988  
National Science Foundation Graduate Fellow, 1985-1988  
Churchill Scholar, Cambridge University, 1985-1986

## External Professional Activities

**Associate Editor**, The Annals of Statistics, 2013-present

**Member and UNC PI**, NIH Genotype-Tissue Expression (GTEx) Consortium, 2010-present

**Associate Editor**, Electronic Journal of Statistics, 2007-2009

### Institute of Mathematical Statistics (IMS)

Program Secretary and Member of the Executive Committee, 2003 - 2006

Committees on Special Lectures ('06-'09), Nominations ('08-'09), Selection of Officers ('11-'13), Funding Advisory ('12-'13), Publications ('15-present).

**Associate Editor**, IEEE Transactions on Information Theory, 2002-2004

### Winston Churchill Scholarship

National Selection Committee, Chicago, IL, December, 2002

UNC Selection Committee, 1998-present

### Conferences:

Scientific Advisory Committee for *Probability Theory and Combinatorial Optimization*, Duke University, 2015.

Program Committee for the *Joint World Congress of the Bernoulli Society and Annual Meeting of the Institute for Mathematical Statistics*, National University of Singapore, 2008.

Program Committee for the *Third IMS-ISBA Joint Meeting*, Bormio, Italy, 2007.

Organizing Committee for the *IMS Special Topics Meeting, 50<sup>th</sup> Anniversary Commemorative Conference*, Department of Statistics, University of North Carolina, Chapel Hill, 1996.

### Statistics and Applied Mathematical Sciences Institute (SAMSI)

Local Scientific Coordinator and Faculty Fellow, Program on Low Dimensional Structure in High Dimensional Systems, 2013-2014

Advisory Committee and Faculty Fellow, Program on Computational Biology of Infectious Diseases, 2004-2005

Local Development Committee, 2002-2003

### Reviewing:

**Grant applications:** NSF and NSA

**National Science Foundation:** Grant review panel 2012, 2014, 2016.

**Journals:** Annals of Statistics; Probability Theory and Related Fields; Proceedings of the U.S. National Academy of Sciences; IEEE Transactions on Information Theory, Pattern Analysis and Machine Intelligence, and Computational Biology and Bioinformatics; Machine Learning; Journal of Multivariate Analysis; Annales de l'Institut Henri Poincaré; Electronic Journal of Statistics; Applied and Computational Harmonic Analysis; Statistics and Probability Letters; Annals of Applied Probability; Annals of the Institute of Statistical Mathematics; Bernoulli; Bioinformatics; Dynamical Systems; Statistics; Statistica Neerlandica; Computational Statistics; Journal of Theoretical Probability; Journal of Theoretical Statistics; Scandinavian Journal of Statistics; Journal of Nonparametric Statistics.

## University Activities

Administrative Board of the Graduate School, 2015-Present

UNC Lineberger Comprehensive Cancer Center, Full Member

Advisory Board of the Odum Institute for Research in Social Science, 2014-present

Applied Physical Sciences Strategic Planning Advisory Committee, 2013

Director of Graduate Studies, Statistics Program, 2005 - present

## External Funding

### Select Current Funding

*Multi-tissue and network models for next-generation eQTL studies*

NIH R01 HG009125-01

A.B. Nobel and F.A. Wright, PIs

Grant period 2016-2019, total award \$1.2 million

*Iterative testing procedures and high-dimensional scaling limits of extremal random structures*

National Science Foundation, DMS-1613072

A.B. Nobel and S. Bhamidi, PIs

Grant period 2016-2019, total award \$375,000

*Random dynamical systems and limit theorems for optimal tracking*

National Science Foundation, DMS-1613261

K. McGoff (UNC-C) PI, A.B. Nobel and S. Mukherjee (Duke), co-PIs

Grant period 2016-2019, total award \$225,000

*Systems Approaches to Link Tissue-Specific Expression to Disease*

NIH R01 MH101819-01

A.B. Nobel and F.A. Wright, PIs

Grant period 2013-2017, total award \$1.3 million

*Optimality Landscapes and Exploratory Data Analysis*

National Science Foundation, DMS-1310002

A.B. Nobel and S. Bhamidi, PIs

Grant period 2013-2017, total award \$210,000

## Select Past Funding

*Significance Based Procedures for Mining and Prediction of Large Data Sets*

National Science Foundation, DMS-0907177

A.B. Nobel, PI

Grant period 2009-2013, total award \$210,000

*Facilitating GTEx, Disease, and GxE Analysis via Fast Expression (e)QTL Mapping*

NIH/NIMH 1R01MH090936-01

A.B. Nobel, I. Rusyn, F.A. Wright, PIs

Grant period 2010 - 2013, total award \$860,000

*Analysis of High Dimensional Data Using Subspace Clustering*

National Science Foundation, DMS-0406361

A.B. Nobel, PI, Wei Wang, Co-PI

Grant period 2004-2008, total award \$253,000

*UNC Kenan Faculty Retention Fund for Arts and Sciences, 2004-2005*

*The Carolina Center for Exploratory Genetic Analysis*

National Institutes of Health, P20 RR020751-01

Dan Reed, PI

Grant period 2004-2007, total award \$400,000

*Discovering New Functional Relationships from Gene Expression Data with Structured Permutation Tests*

Cystic Fibrosis Foundation Therapeutics, Inc., NOBEL04V0

A.B. Nobel, PI

Grant period 2004-2007, total award \$74,478

*Generation and Validation of Synthetic Internet Traffic*

National Science Foundation, ANI-0323648

K. Jeffay, PI, A.B. Nobel and F.D. Smith, Co-PI

Grant period 2003-2005, total award \$470,000

*Estimation from dynamical systems and individual sequences*

National Science Foundation, DMS-9971964.

A.B. Nobel, PI

Grant period 1999-2003, total award \$75,000

*Greedy Growing and its Applications*

Faculty Early Career Development (CAREER) Award

National Science Foundation, DMS-9501926.

A.B. Nobel, PI

Grant period 1995-1998, total award \$72,000

## Teaching: New and Updated Courses

### **STOR 89.2 (New): Data Analysis and Common Sense**

First Year (Freshman) Seminar introducing students to the fundamentals of data science, including the scientific method, reasoning about (and in the presence of) uncertainty, assessing risk, and the elements of data analysis.

### **STOR 565 (New): Machine Learning**

An introductory course on machine learning intended for upper level undergraduates and masters students. Material includes unsupervised (exploratory) data analysis as well as classification and regression. Lectures are supplemented by computer assignments in R.

### **STOR 654 and 655 (Updated): Statistical Theory I and II**

These courses constitute the first year graduate theoretical statistics sequence. Updates include Azuma-Hoeffding and Bernstein inequalities, sub-gaussian random variables, concentration inequalities for bounded and gaussian random variables, high-dimensional estimation and convex geometry, gaussian comparison theorems, and principles of symmetrization and contraction.

## PRODUCTS AND PRESENTATIONS

### **A. Refereed Publications and Proceedings**

Evaluating the performance of a simple inductive procedure in the presence of overfitting error, A.B. Nobel. In *Proceedings of the Fourth Annual Conference on Computational Learning Theory*, pp.267-274, Santa Cruz, CA, 1991.

A Recurrence theorem for dependent processes with applications to data compression, A.B. Nobel and A.D. Wyner, *IEEE Transactions on Information Theory*, vol. 38, pp. 1561-1564, 1992.

A note on uniform laws of averages for dependent processes, A.B. Nobel and A. Dembo, *Statistics and Probability Letters*, vol. 17, pp.169-172, 1993.

A counterexample concerning uniform ergodic theorems for a class of functions, A.B. Nobel, *Statistics and Probability Letters*, vol. 24, pp.165-168, 1995.

Termination and continuity of greedy growing for tree-structured vector quantizers, A.B. Nobel and R.A. Olshen, *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp.191-205, 1996.

Consistency of data-driven histogram methods for density estimation and classification, G. Lugosi and A.B. Nobel, *Annals of Statistics*, vol. 24, no. 2, pp.687-706, 1996.

Vanishing distortion and shrinking cells, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp.1303-1305, 1996.

Histogram regression estimation using data-dependent partitions, A.B. Nobel, *Annals of Statistics*, vol. 24, pp.1084-1105, 1996.

Recursive partitioning to reduce distortion, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 43, no. 4, pp.1122-1133, 1997.

Density estimation from an individual numerical sequence, A.B. Nobel, G. Morvai and S. Kulkarni, *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp.537-541, 1998.

On density estimation from an ergodic process, T.M. Adams and A.B. Nobel, *Annals of Probability*, vol. 26, no. 2, pp.794-804, 1998.

Limits to classification and regression estimation from ergodic processes, A.B. Nobel, *Annals of Statistics*, vol. 27, pp.262-273, 1999.

Adaptive model selection using empirical complexities, G. Lugosi and A.B. Nobel, *Annals of Statistics*, vol. 27, pp.1830-1864, 1999.

Regression estimation from an individual stable sequence, G. Morvai, S.R. Kulkarni, and A.B. Nobel, *Statistics*, vol. 33, pp.99-118, 1999.

Finitary reconstruction of a measure preserving transformation, T.M. Adams and A.B. Nobel, *Israel Journal of Mathematics*, vol. 126, pp.309-326, 2001.

Estimating a function from ergodic samples with additive noise, A.B. Nobel and T.M. Adams, *IEEE Transactions on Information Theory*, vol. 47, pp.2895-2902, 2001.

Analysis of a complexity based pruning method for classification trees, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 48, pp.2362-2368, 2002.

On optimal sequential prediction schemes for general processes, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 49, pp.83-98, 2003.

Indistinguishability of absolutely continuous and singular distributions, S.P. Lalley and A.B. Nobel, *Statistics and Probability Letters*, vol. 62, pp.145-154, 2003.

Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets, T. Sørli, R. Tibshirani, J. Parker, T. Hastie, J.S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, C.M. Perou, P.E. Lønning, P.O. Brown, A-L. Børresen-Dale and D. Botstein, *Proceedings of the U.S. National Academy of Sciences*, vol. 100, pp.8418-8423, 2003.

Some statistical properties of memoryless individual sequences, A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 50, pp.1497-1505, 2004.

Understanding Patterns of TCP Connection Usage with Statistical Clustering, F. Hernandez Campos, A.B. Nobel, F.D. Smith, K. Jeffay. *Proceedings of the Thirteenth MASCOTS Symposium*, Atlanta, GA, September 2005.

Significance analysis of functional categories in gene expression studies: a structured permutation approach, W.T. Barry, A.B. Nobel and F.A. Wright, *Bioinformatics*, vol. 21, pp.1943-1949, 2005.

ChIPOTle: A user-friendly tool for the analysis of ChIP-chip data, M.J. Buck, A.B. Nobel and J.D. Lieb, *Genome Biology*, vol. 6, R97, 2005.

Mining Approximate Frequent Itemsets from Noisy Data, J. Liu, S. Paulsen, W. Wang, A.B. Nobel and J. Prins (short paper). *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM)*, Houston, TX, November, 2005.

Mining approximate frequent itemsets in the presence of noise: Algorithms and analysis, J. Liu, S. Paulsen, X. Sun, W. Wang, A.B. Nobel and J. Prins. *Proceedings of the 2006 SIAM Conference on Data Mining (SDM)*, Bethesda, MD, April 2006.

Hypothesis testing for families of dependent processes, A.B. Nobel, *Bernoulli*, vol. 12, pp.251-269, 2006.

The Molecular Portraits of Breast Tumors Are Conserved Across Microarray Platforms, Z. Hu, C. Fan, D.S. Oh, J.S. Marron, X. He, B.F. Qaqish, C. Livasy, L.A. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M.G. Ewend, L.R. Sawyer, D. Xiang, J. Wu, Y. Liu, R. Nanda, M. Tretiakova, A.R. Orrico, D. Dreher, J.P. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J.F. Quackenbush, O.I. Olopade, P.S. Bernard and C.M. Perou, *BMC Genomics*, 7:96, 2006.

Different gene expression-based predictors for breast cancer patients are concordant, C. Fan, D.S. Oh, L. Wessels, A. Nobel, L.J. van't Veer, and C.M. Perou, *The New England Journal of Medicine*, vol. 355, pp.560-569, 2006.

Significance and Recovery of Block Structures in Binary Matrices with Noise, X. Sun and A.B. Nobel, *Proceedings of the 19th Annual Conference on Learning Theory (COLT)*, pp.109-122, H.U. Simon and G. Lugosi editors, Springer, 2006.

Denosing deterministic time series, S.P. Lalley and A.B. Nobel, *Dynamics of Partial Differential Equations*, vol. 3, pp.259-279, 2006.

Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer, T. Sørlie, C.M. Perou, C. Fan, S. Geisler, T. Aas, A. Nobel, G. Anker, L.A. Akslen, D. Botstein, A-L. Børresen-Dale, and P.E. Lønning, *Molecular Cancer Therapeutics*, vol. 5, pp.2914-2918, 2006.

A statistical framework for testing functional categories in microarray data, W.T. Barry, A.B. Nobel and F.A. Wright, *The Annals of Applied Statistics*, vol. 2, pp.286-315, 2008.

Sequential procedures for aggregating arbitrary estimators of a conditional mean, F. Bunea and A.B. Nobel, *IEEE Transactions on Information Theory*, vol. 54, pp.1725-1735, 2008.

Merging two gene expression studies via cross platform normalization, A.A. Shabalín, H. Tjelmeland, C. Fan, C.M. Perou and A.B. Nobel, *Bioinformatics*, vol. 24, pp.1154-1160, 2008.

Statistical significance of clustering for high dimension low sample size data, Y. Liu, D.N. Hayes, A.B. Nobel and J.S. Marron, *Journal of the American Statistical Association*, vol. 103, pp.1281-1293, 2008.

On the Size and Recovery of Submatrices of Ones in a Random Binary Matrix, X. Sun and A.B. Nobel, *Journal of Machine Learning Research*, vol. 9, pp.2431-2453, 2008.

FastMap: Fast eQTL Mapping in homozygous populations, D.M. Gatti, A.A. Shabalín, T-C. Lam, F.A. Wright, I. Rusyn and A.B. Nobel, *Bioinformatics*, vol. 25, pp.482-489, 2008.

Supervised risk predictor of breast cancer based on intrinsic subtypes, J.S. Parker, M. Mullins, M.C.U. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J.F. Quackenbush, I.J. Stijleman, J. Palazzo, J.S. Marron, A.B. Nobel, E. Mardis, T.O. Nielsen, M.J. Ellis, C.M. Perou, P.S. Bernard, *Journal of Clinical Oncology*, JCO.2008.18.1370, 2009.

The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency, C.R. Lickwar, B. Rao, A.A. Shabalín, A.B. Nobel, B.D. Strahl, J.D. Lieb, *PLoS One*, 4(3):e4886, 2009.

Finding large average submatrices in high dimensional data, A.A. Shabalín, V.J. Weigman, C.M. Perou and A.B. Nobel, *Annals of Applied Statistics*, vol.3, pp.985-1012, 2009.

A Bayesian model for cross-study analysis of differential gene expression (with discussion and rejoinder), R.B. Scharpf, H. Tjelmeland, G. Parmigiani and A.B. Nobel, *Journal of the American Statistical Association*, vol.104, pp.1295-1310, 2009.

Uniform convergence of Vapnik-Chervonenkis classes under ergodic sampling, T.M. Adams and A.B. Nobel, *Annals of Probability*, vol.38, pp.1345-1367, 2010.

Heading down the wrong pathway: on the influence of correlation within gene sets, D.M. Gatti, W.T. Barry, A.B. Nobel, I. Rusyn, F.A. Wright, *BMC Genomics*, vol.11:574, 10 pages, 2010.

DiNAMIC: A Method for assessing the statistical significance of DNA copy number aberrations, V.A. Walter, A.B. Nobel, and F.A. Wright, *Bioinformatics*, vol.27, pp.678-685, 2011.

Uniform approximation of Vapnik-Chervonenkis classes, T.M. Adams and A.B. Nobel, *Bernoulli*, vol.18, pp.1310-1319, 2012.

Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival, V. Weigman, H.H. Chao, A.A. Shabalín, X. He, J.S. Parker, S.H. Nordgard, T. Grushko, D. Huo, C. Nwachukwu, A.B. Nobel, V.N. Kristensen, A.L. Børresen-Dale, O.I. Olopade, C.M. Perou, *Breast Cancer Research and Treatment*, vol.133, pp.865-880, 2012.



Reprioritizing genetic associations in hit regions using LASSO-based resample model averaging, W. Valdar, J. Sabourin, A.B. Nobel, and C.C. Holmes, *Genetic Epidemiology*, vol.36, pp.451-462, 2012.

On the size of large-average and ANOVA-fit submatrices in Gaussian random matrices, X. Sun and A.B. Nobel, *Bernoulli*, vol.19, pp.275-294, 2013.

Joint and Individual Variation Explained (JIVE) for Integrated Analysis of Multiple Datatypes, E.F. Lock, K.A. Hoadley, J.S. Marron, A.B. Nobel, *The Annals of Applied Statistics*, vol.7, pp.523-542, 2013.

Reconstruction of a low-rank matrix in the presence of Gaussian noise, A.A. Shabalin and A.B. Nobel, *Journal of Multivariate Analysis*, vol.118, pp.67-76, 2013.

A testing based extraction algorithm for identifying significant communities in networks, J.D. Wilson, S. Wang, P.J. Mucha, S. Bhamidi, A.B. Nobel, *The Annals of Applied Statistics*, vol.8, pp.1853-1891, 2014.

Consistency of maximum likelihood estimation for some dynamical systems, K. McGoff, S. Mukherjee, A.B. Nobel, and N. Pillai, *The Annals of Statistics*, vol.43, pp.1-29, 2015.

Fine-mapping additive and dominant SNP effects using group-LASSO and fractional resample model averaging, J. Sabourin, A.B. Nobel, and W. Valdar. *Genetic Epidemiology*, vol.39, pp.77-88, 2015.

Entropy and the uniform mean ergodic theorem for a family of sets, T.M. Adams and A.B. Nobel. *Transactions of the American Mathematical Society*, vol.369, pp.605-622, 2017.

Supervised singular value decomposition and its asymptotic properties, G. Li, D. Yang, A.B. Nobel, and H. Shen, 2015. *Journal of Multivariate Analysis*, vol.146, pp.7-17, 2016.

The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans, The GTEx Consortium, *Science*, vol.348, pp.648-660, 2015.

A permutation approach for selecting the penalty parameter in penalized model selection, J. Sabourin, W. Valdar, and A.B. Nobel. *Biometrics*, vol.71, pp.1185-1194, 2015.

Consistent testing for recurrent genomic aberrations, V. Walter, F.A. Wright, and A.B. Nobel. *Biometrika*, vol.102, pp.783-796, 2015.

## B. Submitted Papers and Preprints

Energy landscape for large average submatrix detection problems in Gaussian random matrices, S. Bhamidi, P.S. Dey and A.B. Nobel. Under revision at *Probability Theory and Related Fields*. Preprint available at arXiv:1211.2284.

A testing-based approach to the discovery of differentially correlated variable sets, K. Bodwin, K. Zhang, and A.B. Nobel, 2016. Under revision at *The Annals of Applied Statistics*. Preprint available at arXiv:1509.08124v2.

Change point detection in network models: preferential attachment and long range dependence, S. Bhamidi, J. Jin, and A.B. Nobel. Under revision at *The Annals of Applied Probability*. Preprint available at arXiv:1508.02043v1.

A log-of-linear model for eQTL effect size, J. Palowitch, A. Shabalin, Y. Zhou, A.B. Nobel, and F.A. Wright. Under revision at *Biometrics*. Preprint available at arXiv: .

An empirical Bayes approach for multiple tissue eQTL analysis, G. Li, A.A. Shabalin, I. Rusyn, F.A. Wright, and A.B. Nobel. Submitted. Preprint available at arXiv:1311.2948.

The continuous configuration model: a null model for community detection on weighted networks, J. Palowitch, S. Bhamidi, and A.B. Nobel, 2016. Submitted. Preprint available at arXiv:1601.05630.

Optimal tracking and estimation for dynamical systems, K. McGoff and A.B. Nobel. Submitted. Preprint available at arXiv:1601.05033.

Community extraction in multilayer networks with heterogeneous community structure, J.D. Wilson, J. Palowitch, S. Bhamidi, A.B. Nobel. Submitted. Preprint available at arXiv:1610.06511.

Empirical risk minimization and complexity of dynamical models, K. McGoff and A.B. Nobel. Submitted. Preprint available at arXiv:1611.06173.

HT-eQTL: Integrative eQTL analysis in a large number of human tissues, G. Li, D.D. Jima, F.A. Wright, A.B. Nobel. Submitted. Preprint available at arXiv:1701.05426.

Large subgraphs in pseudo-random graphs A. Basak, S. Bhamidi, S. Chakraborty, A.B. Nobel. Preprint available at arXiv:1610.03762.

### C. Lightly Refereed Publications

Histogram density estimation using data-dependent partitions, A.B. Nobel and G. Lugosi. In *Proceedings of the 1994 Conference on Information Science and Systems*, pp.775-780, Princeton University, Princeton, NJ, 1994.

Consistent estimation of a dynamical map, A.B. Nobel. In *Nonlinear Dynamics and Statistics*, pp.267-280, A.I. Mees editor, Birkhauser, Boston, 2001.

Discussion of “Adaptive confidence intervals for the test error in classification” (by E.B. Laber and S.A. Murphy), S. Wei and A.B. Nobel, *Journal of the American Statistical Association*, vol.106, pp.931-936, 2011.

Discussion of “Population value decomposition, a framework for the analysis of image populations” (by C.M. Crainiceanu, B.S. Caffo, S. Luo, V. Zipunnikov and N.M. Punjabi), E.F. Lock, A.B. Nobel, and J.S. Marron, *Journal of the American Statistical Association*, vol.106, pp.798-802, 2011.

Identification of recurrent DNA copy number aberrations in tumors, V. Walter, A.B. Nobel, D.N. Hayes, and F.A. Wright, in *Statistical Diagnostics for Cancer: Analyzing High-Dimensional Data*, M. Dehmer and F. Emmert-Streib editors, pp.241-260, Wiley-Blackwell, 2013.

A Counterexample concerning the extension of uniform strong laws to ergodic processes, T.M. Adams and A.B. Nobel, in *A Festschrift in Honor of Jon A. Wellner*, IMS collections vol.9, pp.1-4, 2013.

Measuring the statistical significance of local connections in directed networks, J.D. Wilson, S. Bhamidi, and A.B. Nobel, *NIPS 2013, Workshop on Frontiers of Network Analysis*, December, 2013.

Exploratory methods to integrate multi-source data, E.F. Lock and A.B. Nobel, in *Integrating Omics Data*, G. Tseng, D. Ghosh, and X.J. Zhou editors, Cambridge University Press. To appear 2015.

## D. Technical Reports

A.B. Nobel and A.D. Wyner, *On the Asymptotic Behavior of Block Codes*, Bell Labs Technical Memorandum, 1990.

First order predictive sequences and induced transformations, A.B. Nobel, Technical Report #2367, Department of Statistics, UNC Chapel Hill, 1999.

The gap dimension and uniform convergence under ergodic sampling, T.M. Adams and A.B. Nobel, arXiv:1007.2964, 2010.

## E. Software (all publicly available)

MT-eQTL and HT-eQTL: Empirical Bayes method for multi-tissue expression quantitative trait loci (eQTL) analysis.

DCM: Method for finding sets of differentially correlated variables in case-control data.

JIVE: Expresses multiple data matrices on a common set of samples into a sum of low-moderate rank matrices capturing joint and individual variation.

SWITCHdna: A sequential method, based on change-point detection, for segmenting copy number data and identifying regions of copy number aberration.

LAS: Biclustering algorithm that finds statistically significant large average submatrices in high dimensional data.

SigClust: Permutation based method for assessing the significance of clustering results.

SAFE: Permutation based procedure for assessing the significance of functional categories in gene expression data.

XPN: Cross-platform normalization method for combining gene expression data from different studies.

XDE: Bayesian method for assessing the differential expression of genes using data from multiple studies.

FastMap: Fast method for performing eQTL analyses in homozygous populations, including permutation based assessment of significance.

ChIPOTle: Peak finding algorithm for ChIP-chip microarray data.

## F. Patents and Patent Applications

Methods, systems and computer program products for modeling and simulating application-level traffic characteristics in a network based on transport and network layer header information, F. Hernandez-Campos, K. Jeffay, F.D. Smith, and A.B. Nobel. U.S. Letters Patent No. 7,447,209.

Gene expression profiles to predict breast cancer outcomes, P. Bernard, M. Cheang, M. Ellis, E. Mardis, J.S. Marron, T. Nielsen, A.B. Nobel, J. Parker, and C. Perou. Patent pending.

## G. Select Presentations

“Histogram density estimation using data-dependent partitions”, *Conference on Information Science and Systems*, Princeton University, Princeton, NJ, 1994.

“Greedy growing of tree-structured classification rules using a composite splitting criterion”, *IEEE-IMS Workshop on Information Theory and Statistics*, Alexandria, VA, 1994.

“Analysis of several recursive growing algorithms for classification quantization trees”, *IMS/WNAR Meeting*, Stanford University, 1995.

“Adaptive model selection using empirical complexities”, presented by G. Lugosi, *4th World Congress of the Bernoulli Society*, Vienna, Austria, 1996.

“Adaptive model selection using empirical complexities”, *Non-Gaussian Signal Processing Workshop*, Fort Meade, MD, 1996.

“Limits to nonparametric estimation from ergodic processes, and some positive results”, *Workshop on Non-Linear Time Series for Prediction and Control*, Technion, Haifa, Israel, June 1998.

“Estimating a measure preserving transformation from a single trajectory”, *Workshop on Nonlinear Dynamics and Statistics*, Isaac Newton Institute, Cambridge University, England, 1998, and *Joint Statistical Meetings*, Baltimore, MD, August 1999.

“Transformations and singular sequences”, *936th AMS Sectional Meeting*, Wake Forest University, Winston-Salem, NC, 1998.

“Classification from ergodic samples and individual sequences”, *IEEE Information Theory Workshop on Detection, Estimation, Classification, and Imaging*, Santa Fe, NM, 1999.

“Estimating a measure preserving transformation from one of its trajectories”. Presented in the session on Inference for Chaotic Dynamical Systems, *Joint Statistical Meetings*, Baltimore, MD, August 1999.

“Denoising deterministic time series”, *Gertrude M. Cox Statistics Conference*, Research Triangle Institute, Research Triangle Park, NC, September 1999, and *International Conference on Stochastic Optimization and Adaptation*, Cochin, India, 2000.

“The role of cluster analysis in the study of gene expression arrays”, *National Science Foundation Pattern Recognition Workshop*, University of Michigan, Ann Arbor, MI, 2002.

“Denoising deterministic time series and some related questions of statistical inference”, *Lucent Distinguished Lecture Series in Communications and Networks*, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, 2002.

“Distinguishing families of dependent processes”, presented at the *Nonparametric Statistics Research Conference*, Florida State University, Tallahassee, FL, 2003, and the *Ergodic Theory Workshop*, UNC Chapel Hill, NC, 2003, and the *2003 Joint Statistical Meetings*, San Francisco, CA, 2003.

“Statistical Clustering of Internet Communication Patterns”, presented by F. Hernandez-Campos, *35th Symposium on the Interface: Computing Science and Statistics*, Salt Lake City, Utah, 2003.

“Significance Analysis of Bi-clustering for Gene Expression Data”, *New Inference Concepts for Analysing Complex Data*, Mathematisches Forschungsinstitut Oberwolfach, 2004, and *Second IMS-ISBA Joint Meeting*, Bormio, Italy, 2005.

“Significance and Recovery of Block Structures in Binary Matrices with Noise”, workshop on *Statistische und Probabilistische Methoden der Modellwahl*, Mathematisches Forschungsinstitut Oberwolfach, 2005.

“A Bayesian Model for Cross-Study Differential Expression”, ASA Session on Statistical Methods for Integrative Genomics, *Joint Statistical Meetings*, Seattle, WA, 2006.

“Finding significant large-average blocks in high dimensional data”, *39th Symposium on the Interface: Computing Science and Statistics*, Philadelphia, PA, 2007.

Invited participant, *NSF Workshop on Discovery in Massive or Complex Data Sets*, Washington, DC, October, 2007.

“Statistically-based mining of copy number and correlation matrices”, *Joint Statistical Meetings*, Washington, DC, 2009.

“Uniform laws of large numbers for ergodic processes”, *From Probability to Statistics and Back: High-Dimensional Models and Processes*, Seattle, WA, 2010.

“Finite Approximation of VC Classes”, workshop on *Very High Dimensional Semiparametric Models*, Mathematisches Forschungsinstitut Oberwolfach, October, 2011.

“On the maximal size of large average submatrices in a Gaussian random matrix”, *New Trends in Mathematical Statistics*, Centre International de Rencontres Mathématiques, Luminy, November, 2011.

“Joint and Individual Variation Explained (JIVE) for the Integrated Analysis of Multiple Datatypes”, *ENAR*, March, 2012.

“Large Average Submatrices of a Gaussian Random Matrix: Landscapes and Local Optima”, *Structural Inference in Statistics*, Potsdam, Germany, September 2013.

“False Discovery Rate Control and Group Testing for Complex Omics Data”, *ENAR*, March 2014.

“Data Integration and Iterative Testing”, *Workshop on Scientific Inference for Big Data*, National Academy of Sciences, June 2016.

“Mining Differential Correlation”, *Workshop on Theoretical Foundations for Statistical Network Analysis*, Isaac Newton Institute, Cambridge, December 2016.

“Least Squares Estimation for Dynamical Models”, *NIPS Time Series Workshop*, December 2016.

### Graduate Students and Post-Docs

Ms. Soyoun Park, adviser, Statistics, 2003-2005, MS.

Mr. Felix Hernandez-Campos, co-adviser, Computer Science, 2001-2006, Ph.D.

Mr. William T. Barry, co-adviser, Biostatistics, 2004-2006, Ph.D.

Mr. Xing Sun, adviser, Statistics and O.R., 2003-2007, Ph.D.

Mr. Andrey A. Shabalin, adviser, Statistics and O.R., 2006-2010, Ph.D.

Mr. Vonn A. Walter, co-adviser, Biostatistics, 2008-2010, Ph.D.

Mr. Eric Lock, co-adviser, Statistics and O.R., 2008-2013, Ph.D.

Mr. Jeremy Sabourin, co-adviser, Statistics and O.R., 2009-2014, Ph.D.

Dr. Andrey A. Shabalin, co-adviser, Biostatistics, 2010-2012, Post-Doc

Mr. James Wilson, co-adviser, Statistics and O.R., 2011-2015, Ph.D.

Mr. Pratyay Rudra, co-adviser, Biostatistics, 2013-2015, Ph.D.

Dr. Kevin McGoff, Duke Mathematics, co-adviser, 2013-2015, Post-doc

Mr. Gen Li, co-adviser, Statistics and O.R., 2012-2015, Ph.D.

Mr. Bryan Davis, co-adviser, Statistics and O.R., 2015, M.S.

Ms. Kelly Bodwin, co-adviser, Statistics and O.R., 2013-present, Ph.D.

Mr. Jimmy Jin, co-adviser, Statistics and O.R., 2014-present, Ph.D.

Mr. John Palowich, adviser, Statistics and O.R., 2013-present, Ph.D.

Mr. Suman Sen, co-adviser, Statistics and O.R., 2015-present, Ph.D.