

CURRICULUM VITAE

Ingo Ruczinski

Part I

PERSONAL DATA

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205-2179. Phone: (410) 614-7840, Fax: (410) 955-0958, Email: ingo@jhu.edu

EDUCATION AND TRAINING

Degrees

- PhD, University of Washington, Seattle, WA, Statistics (2000).
- MA, University of Würzburg, Germany, Mathematics and Computer Science (1995).
- MA, State University of New York, Albany, NY, Mathematics and Statistics (1994).
- BA, University of Würzburg, Germany, Mathematics and Computer Science (1991).

Postdoctoral Training

- Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA (2000-2001).
- Department of Biochemistry, University of Washington, Seattle, WA (2000-2001).

Visiting

- Department of Chemistry, University of California at Santa Barbara (2003-2008, ~5mo).
- Centre National de Genotypage, Paris-Evry, France (2005, 2wks).
- Department of Statistics, University of Umeå, Sweden (1994, 2wks).
- Department of Mathematics, University of Caen, France (1993, 2wks).

PROFESSIONAL EXPERIENCE

- Associate Professor, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University (2007-present).
- Member, Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins (2006-present).
- Assistant Professor, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University (2001-2007).
- Postdoctoral Fellow, Department of Biochemistry, University of Washington (2000-2001).
- Research Associate, Public Health Sciences, Fred Hutchinson Cancer Research Center (2000-2001).

- Consultant, Department of Statistics, University of Washington (1998-1999).
- Research Assistant, Public Health Sciences, Fred Hutchinson Cancer Research Center (1998-2000).
- Research Assistant, Department of Biochemistry, University of Washington (1996-2000).
- Teaching Assistant, Department of Statistics, University of Washington (1995-1996).

PROFESSIONAL ACTIVITIES

Professional Memberships

American Statistical Association (Sections on Statistical Computing and Graphics, Risk Analysis, Epidemiology), International Biometric Society (ENAR), Institute of Mathematical Statistics, R Foundation for Statistical Computing.

Participation on Advisory Panels

- Member of the Protein Folding Consortium, a group of experimental protein folding investigators, spanning 18 laboratories in eight countries. The goal of the consortium is to set uniform standards for the experimental community and to compile a dataset of the folding properties of a large number of reversibly folding proteins that will aid ongoing efforts to understand the protein folding process (2004-present).
- Member of the CLUE serology advisory committee, which oversees issues related to the specimen and data from the CLUE cohort study. These issues concern genotyping, genomic amplification, combining results from different laboratories and different assay methods, and preserving specimens. The committee also reviews collaborative proposals from students and investigators outside Johns Hopkins University (2005-present).

Program Development

- Program Chair, American Statistical Association, Section on Risk Analysis (2005-2006).
- Program Committee, Joint Statistical Meetings (2006).
- Program Committee, Eastern North Atlantic Region (2006).
- Session Chair, Joint Statistical Meetings (2001,2004,2005).
- Session Organizer, Joint Statistical Meetings (2002,2005,2006).
- Session Chair, Eastern North Atlantic Region Meeting (2003).
- Session Organizer, Eastern North Atlantic Region Meeting (2006,2007).

EDITORIAL ACTIVITIES

Peer Review Activities

Algorithms for Molecular Biology, American Journal of Human Genetics, AMS-Dimacs Volume on Epidemiology and Data Mining, Bioinformatics, Biophysical Reviews and Letters, Biostatistics, BioMed Central Bioinformatics, BioMed Central Genomics, Biotechnology Progress, Breast Cancer Research, Computational Statistics and Data Analysis, Human Mutation, International Journal of Biostatistics, Journal of the American Statistical Association, Journal of Computational and Graphical Statistics, Journal of Heuristics, Journal of the Royal Statistical Society (Series B), Journal of Statistical Software,

Mammalian Genome, Protein Science, Proteins: Structure, Function and Genetics, Proteins: Structure, Function and Bioinformatics, Public Library of Science (PLOS), Springer Verlag (New York), Statistical Applications in Genetics and Molecular Biology, Statistics in Medicine.

Editorial Board Membership

- Editorial Board, *Cancer Prevention Research*, American Association for Cancer Research (2008-present).
- Associate Editor, *Biometrics*, International Biometric Society (2008-present).

Ad Hoc Review of Proposals

- Research and Development Program, Louisiana Board of Regents (2004).
- Cancer Research UK Bioinformatics Programme (2007).
- National Science Foundation, Biological Database and Informatics Program (2007).
- Johns Hopkins NIEHS Center in Urban Environmental Health (2008).
- Science Unbound Foundation (2008).
- National Institutes of Health (NIH), Center for Scientific Review, Cancer Biomarker Study Section (February 2006, October 2006, February 2007, October 2007, October 2008).

HONORS AND AWARDS

- ZW Birnbaum Award for Excellence in Teaching, University of Washington (1996).
- Faculty Innovation Award, Johns Hopkins University (2002).
- Research Grant, Maryland Cigarette Restitution Fund to the Johns Hopkins Medical Institutions (2002).
- Young Researchers Travel Award, Int. Federation of Classification Societies, Cracow, Poland (2002).
- Best Contributed Paper Award, Joint Statistical Meetings from the ASA Section on Risk Analysis (2003).
- NSF Young Researchers Travel Award, University of Florida (2003).
- NSF Young Researchers Travel Award (University of California at Los Angeles (2004).
- Johns Hopkins Bloomberg School of Public Health Faculty Research Initiative Award (2006).
- Advising, Mentoring, and Teaching Recognition Award, Johns Hopkins University (2007).

PUBLICATIONS

Peer-Reviewed Articles

- **Ruczinski I**, Plaxco KW (2008). *Some recommendations for the practitioner to improve the precision of experimentally determined protein folding rates and Φ values*. Proteins: Structure, Function and Bioinformatics (to appear).
- Schulze KJ, Christian P, **Ruczinski I**, Ray AL, Nath A, Wu LS, Semba RD (2008). *Hepcidin and iron status among pregnant women in Bangladesh*. Asia Pacific Journal of Clinical Nutrition 17(3): 451-6.
- Pattaro C, **Ruczinski I**, Fallin DM, Parmigiani G (2008). *Haplotype block partitioning as a tool for dimensionality reduction in SNP association studies*. BMC Genomics 9(1): 405.

- Chen J, **Ruczinski I**, Jorgensen TJ, Yenokyan G, Yao Y, Alani R, Liegeois NJ, Hoffman SC, Hoffman-Bolton J, Strickland PT, Helzlsouer KJ, Alberg AJ (2008). *Nonmelanoma skin cancer as a marker of risk for subsequent malignancy: a community-based prospective cohort study*. Journal of the National Cancer Institute 100(17): 1215-22.
- Sull JW, Liang KY, Hetmanski JB, Fallin MD, Ingersoll RG, Park J, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Park BY, Jee SH, Jabs EW, Redett R, Jung E, **Ruczinski I**, Scott AF, Beaty TH (2008). *Differential parental transmission of markers in RUNX2 among cleft case-parent trios from four populations*. Genetic Epidemiology 32(6): 505-12.
- Scharpf R, Parmigiani G, Pevsner J, **Ruczinski I** (2008). *Hidden Markov models for the assessment of chromosomal alterations using high-throughput SNP arrays*. Annals of Applied Statistics 2(2): 687-713.
- Gallicchio L, Chang H, Christo DK, Thuita L, Huang HY, Strickland P, **Ruczinski I**, Hoffman SC, Helzlsouer KJ (2008). *Single nucleotide polymorphisms in inflammation-related genes and mortality in a community-based cohort in Washington county, Maryland*. American Journal of Epidemiology 167(7): 807-13.
- Ting JC, Roberson ED, Miller N, Lysholm-Bernacchi A, Stephan DA, Capone GT, **Ruczinski I**, Thomas GH, Pevsner J (2007). *Visualization of uniparental inheritance, Mendelian inconsistencies, deletions and parent of origin effects in single nucleotide polymorphism trio data with SNP trio*. Human Mutation 28(12): 1225-35.
- Scharpf RB, Ting JC, Pevsner J, **Ruczinski I** (2007). *SNPchip: R classes and methods for SNP array data*. Bioinformatics 23(5): 627-8.
- Visvanathan K, Crum RM, Strickland PT, You X, **Ruczinski I**, Berndt S, Alberg AJ, Hoffman SC, Comstock GW, Bell DA, Helzlsouer KJ (2007). *Alcohol dehydrogenase genetic polymorphisms, low-to-moderate alcohol consumption and risk of breast cancer*. Alcoholism: Clinical and Experimental Research 31(3): 467-76.
- Jorgensen J, Visvanathan K, **Ruczinski I**, Thuita L, Helzlsouer KJ (2007). *Breast cancer risk is not associated with polymorphic forms of Xeroderma Pigmentosum genes in a cohort of women from Washington County, Maryland*. Breast Cancer Research and Treatment 101(1): 65-71.
- Dai J*, **Ruczinski I***, LeBlanc M, Kooperberg C (2006). *Imputation methods to improve inference in SNP association studies*. Genetic Epidemiology 30(8): 690-702 * contributed equally.
- **Ruczinski I**, Sosnick TR, Plaxco KW (2006). *Methods for the accurate estimation of confidence intervals on protein folding Φ -values*. Protein Science 15(10): 2257-64.
- Brown EE, Fallin D, **Ruczinski I**, Hutchinson A, Staats B, Vitale F, Lauria C, Serraino D, Rezza G, Mbisa G, Whitby D, Messina A, Goedert JJ, Chanock SJ, and the Kaposi Sarcoma Working Group (2006). *Associations of classic kaposi sarcoma with common variants in genes that modulate host immunity*. Cancer Epidemiology, Biomarkers and Prevention 15(5): 926-34.
- De Los Rios MA, Muralidhara BK, Wildes D, Sosnick TR, Marqusee S, Wittung-Stafshede P, Plaxco KW, **Ruczinski I** (2006). *On the precision of experimentally determined protein folding rates and Φ -values*. Protein Science 15(3): 553-63.
- Brewster AM, Jorgensen TJ, **Ruczinski I**, Huang HY, Hoffman S, Thuita L, Newschaffer C, Lunn RM, Bell D, Helzlsouer KJ (2006). *Polymorphisms of the DNA repair genes XPD (Lys751Gln) and XRCC1 (Arg399Gln and Arg194Trp): relationship to breast cancer risk and familial predisposition to breast cancer*. Breast Cancer Research and Treatment 95(1): 73-80.

- Ting JC, Ye Y, Thomas GH, **Ruczinski I**, Pevsner J (2006). *Analysis and visualization of chromosomal abnormalities in SNP data with SNPscan*. BMC Bioinformatics, 18:7(1): 25.
- Mao R, Wang X, Spitznagel Jr EL, Frelin LP, Ting JC, Ding H, Kim J, **Ruczinski I**, Downey TJ, Pevsner J (2005). *Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart*. Genome Biology, 6(13): R107.
- McCarney ER, Werner JH, Bernstein SL, **Ruczinski I**, Makarov DE, Goodwin PM, Plaxco KW (2005). *Site-specific dimensions across a highly denatured protein: a single molecule study*. Journal of Molecular Biology, 352(3): 672-82.
- Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L, Horng JC, Bona D, Miller EJ, Vallee-Belisle A, Main ER, Bemporad F, Qiu L, Teilum K, Vu ND, Edwards AM, **Ruczinski I**, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y, Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE, Sosnick TR, Marqusee S, Davidson AR, Plaxco KW (2005). *Protein folding: defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins*. Protein Science 14(3): 602-16.
- Kooperberg C, **Ruczinski I** (2005). *Identifying interacting SNPs using monte carlo logic regression*. Genetic Epidemiology 28(2): 157-70.
- **Ruczinski I**, Kooperberg C, LeBlanc M (2004). *Exploring interactions in high dimensional genomic data: an overview of logic regression, with applications*. Journal of Multivariate Analysis 90: 178-95.
- Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, **Ruczinski I**, Doniach S, Plaxco KW (2004). *Random-coil behavior and the dimensions of chemically unfolded proteins*. Proceedings of the National Academy of Sciences 101(34): 12491-6.
- **Ruczinski I**, Kooperberg C, LeBlanc M (2003). *Logic regression - methods and software*. Nonlinear Estimation and Classification, Lecture Notes in Statistics 171: 333-44.
- **Ruczinski I**, Kooperberg C, LeBlanc M (2003). *Logic regression*. Journal of Computational and Graphical Statistics 12(3): 475-511.
- Bonneau R, Tsai J, **Ruczinski I**, Baker D (2002). *Contact order and ab initio protein structure prediction*. Protein Science 11(8): 1937-44.
- **Ruczinski I**, Kooperberg C, Bonneau R, Baker D (2002). *Distributions of β sheets in proteins with application to structure prediction*. Proteins: Structure, Function and Genetics 48: 85-97.
- Larson S, **Ruczinski I**, Davidson AR, Baker D, Plaxco KW (2002). *Residues participating in the folding nucleus do not exhibit preferential evolutionary conservation*. Journal of Molecular Biology 316: 225-33.
- Kooperberg C, **Ruczinski I**, LeBlanc M, Hsu L (2001). *Sequence analysis using logic regression*. Genetic Epidemiology 21(S1): 626-31.
- Bonneau R, Tsai J, **Ruczinski I**, Chivian D, Rohl C, Strauss C, Baker D (2001). *Rosetta in CASP4: progress in ab initio protein structure prediction*. Proteins: Structure, Function and Genetics 45(S5): 119-26.
- Bonneau R, Tsai J, **Ruczinski I**, Baker D (2001). *Functional inferences from blind ab initio protein structure predictions*. Journal of Structural Biology 134(2-3): 186-90.

- Plaxco KW, Simons KT, **Ruczinski I**, Baker D (2000). *Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics*. *Biochemistry* 39(37): 11177-83.
- Plaxco KW, Larson S, **Ruczinski I**, Riddle DS, Thayer EC, Buchwitz B, Davidson AR, Baker D (2000). *Evolutionary conservation in protein folding kinetics*. *Journal of Molecular Biology* 298: 303-12.
- Simons KT, Bonneau R, **Ruczinski I**, Baker D (1999). *Ab initio protein structure prediction of CASP3 targets using ROSETTA*. *Proteins* 37(S3): 171-6.
- Riddle DS, Grantcharova VP, Santiago J, Alm E, **Ruczinski I**, Baker D (1999). *Experiment and theory highlight role of native state topology in SH3 folding*. *Nature Structural Biology* 6: 1016-24.
- Simons KT, **Ruczinski I**, Kooperberg C, Fox B, Bystroff C, Baker D (1999). *Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins*. *Proteins* 34(1): 82-95.

Book Chapters and Extended Conference Proceedings

- Scharpf RB, **Ruczinski I** (2008). *R classes and methods for SNP array data*. *Bioinformatics in Clinical OMICs Research*, Humana Press (to appear).
- **Ruczinski I**, Kooperberg C (2008). *Logic regression*. *Encyclopedia of Medical Decision Making* (to appear).
- **Ruczinski I** (2007). *Hidden Markov models for the assessment of chromosomal alterations using high-throughput SNP arrays*. INSERM Workshop #179 Proceedings, LaLonde les Maures, France.

Thesis

- PhD in Statistics, University of Washington (2000). *Logic Regression and Statistical Issues Related to the Protein Folding Problem*. Supervisors: Charles Kooperberg, Michael LeBlanc, David Baker.
- Diplom/MA in Mathematics, minor Computer Science, University of Würzburg (1995). *The Degree of Inclusion and Other Process Capability Indices*. Supervisor: Elart von Collani.
- MA in Mathematics, State University of New York at Albany (1994). *Births in the USA: A Time Series Analysis*. Supervisor: Igor Zurbenko.

Software

- Logic regression: Logic regression is a regression methodology primarily for the detection of gene-gene and gene-environment interactions in SNP association studies. The Logic Regression methodology and software was developed in collaboration with Charles Kooperberg and Michael LeBlanc at the Fred Hutchinson Cancer Research Center. Logic Regression is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form.
- Φ estimation: Methods and software for the analysis of kinetic data derived from denaturing experiments were developed in collaboration with Kevin Plaxco at the University of California Santa Barbara, and Tobin Sosnick at the University of Chicago. The software was implemented as a web tool, using Perl, HTML, and the statistical software environment R.

- SNPchip: The R package SNPchip contains S4 classes and methods useful storing, visualizing, and analyzing high density SNP data. SNPchip was developed in collaboration with Robert Scharpf at the Johns Hopkins Bloomberg School of Public Health, and Jonathan Pevsner and Jason Ting in the Kennedy Krieger Institute. The R package SNPchip is freely available from the Bioconductor webpage as free software in source code form under the terms of the GNU General Public License of the Free Software Foundation.
- VanillaICE: The R package ICE contains the software for fitting hidden Markov models on genomic array data to infer chromosomal alterations, including deletions, amplifications, and regions with loss of heterozygosity. In addition, measures of uncertainty for the genotype and copy number estimates can be incorporated, which can be crucial for the detection of micro-deletions and micro-amplifications. ICE was developed in collaboration with Robert Scharpf and Giovanni Parmigiani at the Johns Hopkins Bloomberg School of Public Health. The R package is available from the Bioconductor webpage as free software in source code form under the terms of the GNU General Public License of the Free Software Foundation.
- Rosetta ab initio: Rosetta is a software suite relevant for the prediction and design of protein structures, protein folding mechanisms, and protein-protein interactions. Rosetta ab initio is a tool for de novo prediction of protein structures, and was developed in collaboration with members in the laboratory of David Baker at the University of Washington. The license for the Rosetta code is available through the Baker laboratory.

CURRICULUM VITAE

Ingo Ruczinski

Part II

TEACHING

Advisees

- Long, Ezhou. Master of Science, Biostatistics (2005).
- Scharpf, Robert. Doctor of Philosophy, Biostatistics (2007).
- Li, Qing. Doctor of Philosophy, Biostatistics (2008, expected).
- Liu, Yun. Master of Health Sciences, Biostatistics (2008, expected).

Academic Advisees

- Wang, Jian. Master of Science, Biostatistics (2001-2002).
- Shum, Kenny. Doctor of Philosophy, Biostatistics (2002-2004).
- Ho, Yen-Yi. Doctor of Philosophy, Biostatistics (2003-2005).
- Chang, Howard. Doctor of Philosophy, Biostatistics (2004-2006).
- Murakami, Peter. Doctor of Philosophy, Biostatistics (2006-2007).
- Paithankar, Sameer. Master of Science, Bioinformatics (2007-present).

Rotation Students

- Aripirala, Srinivas. Doctor of Philosophy, IMMBI (2007).

Post-Doctoral Fellows

- Scharpf, Robert (2007-present).

Advisory Committees

- Tsai, Yuhjung (2008-present).

Visitors

- Cristian Pattaro, EURAC, Bolzano, Italy (2008).
- Florentina Bunea, Florida State University, Tallahassee, FL (2008).

Thesis Committees / Reader

Tarr, Deirdre Ellen. Doctor of Philosophy, Molecular Microbiology and Immunology (2004). Liu, Youngmei. Doctor of Philosophy, Epidemiology (2004). Sefcovic, Natasha. Doctor of Philosophy, Biology NIH/Johns Hopkins (2006). Wang, Wenyi. Doctor of Philosophy, Biostatistics (2007). Nicodemus, Kristin. Doctor of Philosophy, Epidemiology (2007). Gamston, Courtney. Master of Science, Molecular Microbiology and Immunology (2007). Lam, Tram Kim. Doctor of Philosophy, Epidemiology (2007). Wang, Ming Hsi. Doctor of Philosophy, Epidemiology (2008). Cheng, Yu-Ching. Doctor of Philosophy, Epidemiology (2008). Murakami, Peter. Master of Science, Biostatistics (2008). Carvalho, Benilton. Doctor of Philosophy, Biostatistics (2008). Venkatesan, Meera. Doctor of Philosophy, Molecular Microbiology and Immunology (2008). Roberson, Eli. Doctor of Philosophy, Human Genetics, JHMI

(2008, expected). McCall, Matthew. Doctor of Philosophy, Biostatistics (2009, expected). Chu, Audrey. Doctor of Philosophy, Epidemiology, (2009, expected). Bowers, Katherine. Doctor of Philosophy, Epidemiology (2009, expected). Boca, Simina. Doctor of Philosophy, Biostatistics (2010, expected).

Preliminary Oral Participation

Ziegler, Kathryn. Doctor of Philosophy, Biostatistics (2003). Kittleson, Michelle. Doctor of Philosophy, Clinical Investigation (2004). Scharpf, Robert. Doctor of Philosophy, Biostatistics (2004). Lam, Tram Kim. Doctor of Philosophy, Epidemiology (2005). Naj, Adam. Doctor of Philosophy, Epidemiology (2005). Wang, Ming Hsi. Doctor of Philosophy, Epidemiology (2005). Cheng, Yu-Ching. Doctor of Philosophy, Epidemiology (2006). Li, Qing. Doctor of Philosophy, Biostatistics (2006). Nicodemus, Kristin. Doctor of Philosophy, Epidemiology (2006). Bowers*, Katherine. Doctor of Philosophy, Epidemiology (2007). McCall, Matthew. Doctor of Philosophy, Biostatistics (2007). Chu*, Audrey. Doctor of Philosophy, Epidemiology, (2008). * committee chair

Final Oral Participation

Liu, Youngmei. Doctor of Philosophy, Epidemiology (2004). Tarr, Deirdre Ellen. Doctor of Philosophy, Molecular Microbiology and Immunology (2004). Peila, Rita. Doctor of Philosophy, Epidemiology (2005). Sefcovic, Natasha. Doctor of Philosophy, Biology, NIH/Johns Hopkins (2006). Scharpf, Robert. Doctor of Philosophy, Biostatistics (2007). Nicodemus, Kristin. Doctor of Philosophy, Epidemiology (2007). Lam, Tram Kim. Doctor of Philosophy, Epidemiology (2007). Wang, Ming Hsi. Doctor of Philosophy, Epidemiology (2008). Cheng*, Yu-Ching. Doctor of Philosophy, Epidemiology (2008). Venkatesan*, Meera. Doctor of Philosophy, Molecular Microbiology and Immunology (2008). * committee chair

Classroom Instruction - Principal Instructor

- 140.615 Statistics for Laboratory Scientists I (2007,2008).
- 140.616 Statistics for Laboratory Scientists II (2002,2007,2008).
- 140.751 Advanced Methods in Biostatistics I (2002-2005).
- 140.752 Advanced Methods in Biostatistics II (2002-2005).

Classroom Instruction - Co-Instructor

- 140.668 Statistical Topics in Genetics and Genomics (2004).
- 140.776 Statistical Computing (2003,2004).
- 260.655 Protein Bioinformatics (2004-2006,2008).

Classroom Instruction - Invited Guest Lecturer

- 140.636 Perl for Bioinformatics (2003,2004).
- 140.638 Analysis of Biological Sequences (2005-2007).
- 140.649 Statistical Learning: Algorithmic and Nonparametric Approaches (2007).
- 140.754 Advanced Methods in Biostatistics IV (2002,2004-2007).
- 140.778 Advanced Statistical Computing (2002).
- 260.602 Introduction to Bioinformatics (2006,2007).
- 260.655 Protein Bioinformatics (2007).
- 340.611 Methodologic Issues in Cancer Epidemiology (2005).

- 340.631 Methods for Association Analysis in Genetic Epidemiology (2007,2008).
- 550.865 Public Health Perspectives on Research (2005).

RESEARCH GRANT PARTICIPATION

Ongoing Research Support

- *Preprocessing and Analysis Tools for Contemporary Microarray Applications (NIH)*.
Dates: September 2007 - November 2012. Principal Investigator: Rafael Irizarry, PhD. Objective: Develop rigorous statistical procedures for five microarray applications with a rapidly growing user base: high-throughput genotyping, whole-genome chromosomal abnormality detection, discovering promoter binding sites with ChIPchip technology, high-throughput detection of methylated sites, and alternative splicing. Responsibility: Co-Principal Investigator (10%).
- *Institute for Clinical and Translational Research (NIH/NCRR)*.
Dates: September 2007 - May 2012. Principal Investigator: Daniel Ford, MD. Objective: This CTSA grant will support clinical and translational research throughout Johns Hopkins. It includes support for education and training of new translational investigators, facilities in which clinical research can take place and infrastructure support of patient recruitment, bioinformatics, biostatistics and translational core centers. Responsibility: Co-investigator (15%).
- *Novel Statistical Methods for Gene-Environment Interactions in Complex Diseases (NHLBI)*.
Dates: September 2007 - July 2010. Principal Investigator: Kung-Yee Liang, PhD. Objective: The goal of this proposal is to develop and evaluate new statistical methods to prioritize genes through proper ranking in genome-wide association studies that address GxE interactions. Responsibility: Co-Investigator (20%).
- *Hierarchical Models in Health Services Research (NIDDK)*.
Dates: September 2007 - July 2010. Principal Investigator: Tom Louis, PhD. Objective: The proposed research will produce innovative statistical methodology, provide guidance on trade-offs in addressing multiple goals, communicate interesting and informative case studies, educate and acculturate a pre-doctoral student. Responsibility: Co-Investigator (10%).
- *Genome-Wide Association Studies of Asthma In Populations Of African Descent (NIH)*.
Dates: October 2006 - September 2009. Principal Investigator: Kathleen Barnes, PhD. Objective: Performing a genome-wide association study for asthma and associated quantitative traits in a family-based African Caribbean population and an independent population of African American asthmatics and non-asthmatics, and validating the findings in large panels of English families selected for asthma, and evaluating the generalizability of these findings in the larger US white population. Responsibility: Co-Investigator (10%).
- *DNA Repair, Skin Cancer and Overall Cancer Risk (NIH)*.
Dates: July 2005 - June 2009. Principal Investigator: Anthony Alberg, PhD. Objective: To investigate the associations between SNPs in DNA repair genes and skin cancer. Responsibility: Co-Investigator (20%).
- *Genotypic Determinants of Aspirin Response in High Risk Families (NIH)*.

Dates: January 2008 - May 2009. Principal Investigator: Lewis Becker, MD. Objective: To determine the genotypic determinants of aspirin response in high risk families. Responsibility: Co-Investigator (5%).

Completed

- *Genetic Influences on Age-Related Decline In Strength (NIH)*.
Dates: February 2005 - December 2007. Principal Investigator: Brock Beamer, MD. Objective: Investigate the associations between genotype, environmental variables, and age-related decline in strength. Responsibility: Co-Investigator (5%).
- *Novel Approaches to Studying the In Situ Bioremediation Potential of Complex Mixtures (NIEHS)*.
Dates: September 2006 - December 2007. Principal Investigator: Rolf Halden, PhD. Objective: Evaluation of In Situ Microcosm Array Technology for the study of the Biodegradation Potential of Complex Chemical Mixtures in Contaminated Environments. Responsibility: Co-Investigator (3%).
- *Johns Hopkins Center in Urban Environmental Health (NIEHS)*
Dates: April 2003 - September 2007. Principal Investigator: Scott Zeger, PhD. Objective: Consult with environmental health scientists including molecular toxicologists, physiologists, environmental engineers, and epidemiologists. Responsibility: Co-Investigator (10%).
- *Adaptive Function Estimation for Genomic Data (NIH)*.
Dates: September 2003 - August 2007. Principal Investigator: Ingo Ruczinski, PhD. Objective: Enhance the existing software packages, address some standing methodological problems, and implement the respective solutions. Responsibility: Principal Investigator (15%).
- *Gene-Gene Interaction Involved In Nasopharyngeal Carcinoma (NCI)*.
Dates: September 2005 - August 2006. Principal Investigator: Yin Yao Shugart, PhD. Objective: Investigate associations between single nucleotide polymorphisms and nasopharyngeal carcinoma in an Asian population. Responsibility: Co-Investigator (5%).
- *Analysis of SNP Data using Logic Regression (MDHMM)*.
Dates: July 2002 - June 2004. Principal Investigator: Ingo Ruczinski, PhD. Objective: Develop software and methodology for logic regression, and analyze cancer related data. Responsibility: Principal Investigator (40%).
- *Gene-Environment Interactions: The Odyssey Cohort (NCI)*.
Dates: September 1999 - June 2004. Principal Investigator: Kathy Helzlsouer, MD. Objective: Investigate associations between single nucleotide polymorphisms, environmental variables, and disease incidence. Responsibility: Co-Investigator (20%).
- *Modeling of Mass Spectrometry MALDI Data (DARPA)*.
Dates: July 2003 - December 2003. Principal Investigator: Fernando Pineda, PhD. Objective: BioTOF Systems Modeling and Data Analysis. Responsibility: Co-Investigator (5%).
- *Using All-Atom Potentials to Improve Ab Initio Protein Structure Prediction (JHU Faculty Innovation)*.
Dates: July 2002 - June 2003. Principal Investigator: Ingo Ruczinski, PhD. Objective: Develop filters for ROSETTA to improve ab initio protein structure prediction, and develop tools for the analysis of other protein related data. Responsibility: Principal Investigator (20%).

ACADEMIC SERVICE

Department of Biostatistics

- Member, Committee for Biostatistics Information Technology (2001-present).
- Member, Second Year Oral Examination Committee (2003-2005).
- Chair, Honors and Awards Committee (2003-present).
- Seminar Coordinator (2004-2005).
- Faculty Senate Representative (2004-2006).
- Member, Intellectual Environment Committee (2005-present).

Bloomberg School of Public Health

- Member, Steering Committee, Malaria Research Institute (2002-present).
- Member, Curriculum Committee MHS in Bioinformatics (2003-present).
- Member, Bioinformatics and Biostatistics Core, Center in Urban Environmental Health (2003-present).
- Affiliate, Cancer Epidemiology Prevention and Control Training Fellowship Program (2005-present).
- Co-Director, MHS in Bioinformatics Program (2007-present).
- Committee on Information Technology (2007-present).
- Co-Chair, CTSA Innovative Working Group on Genome-wide Association Studies (2008-present).

PRESENTATIONS

Scientific Meetings

- *On Missing Data and Genotyping Errors in Association Studies*. GEI Genome-wide Association: Analyze This! Workshop, National Institutes of Health, Bethesda, MD (2008).
- *On Missing Data and Genotyping Errors in Association Studies*. BIRS Workshop on Emerging Statistical Challenges in Genome and Translational Research, Banff, Canada (2008).
- *Novel Statistical Methods for Gene-Environment Interactions in Complex Diseases*. National Institutes of Health, Bethesda, MD (2008).
- *On Genotyping Errors and Missing Data in Genome-Wide Association Studies*. GENEVA Investigator Meeting, National Institutes of Health, Bethesda, MD (2008).
- *Assessment of Chromosomal Alterations using Copy Number and Genotype Estimates*. INSERM Workshop, LaLonde les Maures, France (2007).
- *An Integrated Approach for the Assessment of Chromosomal Abnormalities*. BIRS Workshop Statistical Methods for High-Throughput Genetic Data, Banff, Canada (2007).
- *An Integrated Approach for the Assessment of Chromosomal Abnormalities*. Workshop on Statistics for Biomolecular Data Integration and Modeling, Ascona, Switzerland (2007).
- *Detecting Genetic Interactions in Disease*. Dementia Consortium Retreat, JHM Alzheimer's Disease Research Center, Baltimore, MD (2007).
- *An Integrated Approach for the Assessment of Chromosomal Abnormalities*. Chapter Meeting of the International Society of Genetic Epidemiology, Harvard University, Boston, MA (2007).

- *On the Precision of Experimentally Determined Protein Folding Rates and Φ Values*. Structural Bioinformatics & Computational Biophysics Meeting, Fortaleza, Brazil (2006).
- *Computational and Statistical Tools Relevant for the Exploration of the Protein Folding Process*. BIRS Workshop on Computational and Statistical Genomics, Banff, Canada (2006).
- *Visualizing and Analyzing High Density SNP Data with SNPscan*. International Biometric Society Spring Meeting, Tampa, FL (2006).
- *On Missing Data and Interactions in SNP Association Studies*. Workshop on Statistical Methods in Molecular Epidemiology, Ruhr University, Bochum, Germany (2005).
- *Logic Regression and its Applications in SNP Association Studies*. 26th Annual Conference of the International Society for Clinical Biostatistics, Szeged, Hungary (2005).
- *Logic Regression in SNP Association Studies*. BIRS Workshop on Statistical Science for Genome Biology, Banff, Canada (2004).
- *Protein Folding and Structure Prediction*. North American New Researchers Conference, York University, Toronto, Canada (2004).
- *Improvements for Logic Regression*. Meeting of the International Federation of Classification Societies IFCS, Chicago, IL (2004).
- *Statistical and Computational Issues in Protein Folding and Structure Prediction*. Interface 2004: Computational Biology and Bioinformatics, Baltimore, MD (2004).
- *Protein Structure Prediction using Rosetta*. IPAM Workshop on Structural Proteomics at UCLA, Los Angeles, CA (2004).
- *Finding Interactions and Assessing Variable Importance in SNP Association Studies*. DIMACS Working Group on Data Mining and Epidemiology, Rutgers University, New Brunswick, NJ (2004).
- *Finding Simple Classification Rules in Risk Analysis*. Joint Stat. Meetings, San Francisco, CA (2003).
- *Exploring Interactions in Genomic Data*. International Biometric Society Meeting, Tampa, FL (2003).
- *Statistical and Computational Issues in Ab Initio Protein Structure Prediction*. Joint Statistical Meetings, New York, NY (2002).
- *Classification using Boolean Functions*. 8th Conference of the International Federation of Classification Societies IFCS, Cracow, Poland (2002).
- *Logic Regression*. Annual Meeting, Classification Society of North America, Madison, WI (2002).
- *Computational and Statistical Issues in Ab Initio Protein Structure Prediction*. American Statistical Association Chapter Meeting, Albany, NY (2002).
- *Logic Regression*. International Biometric Society ENAR Spring Meeting, Arlington, VA (2002).
- *A New Regression Methodology using Boolean Logic*. Joint Statistical Meetings, Atlanta, GA (2001).
- *Logic Regression*. MSRI Workshop on Nonlinear Estimation and Classification, Berkeley, CA (2001).

Invited Seminars

- *Hidden Markov Models for the Assessment of Chromosomal Alterations using High-throughput SNP Arrays*. Section on Statistical Genetics, Department of Biostatistics, University of Alabama, Birmingham AL (2008).

- *Hidden Markov Models for the Assessment of Chromosomal Alterations using SNP Arrays*. Statistics and Genomics Seminar, University of California, Berkeley CA (2007).
- *On Missing Data and Interactions in SNP Association Studies*. Institute of Genetic Medicine, EURAC, Bolzano, Italy (2007).
- *An Integrated Approach for the Assessment of Chromosomal Abnormalities*. Institut für Statistik, Ludwig-Maximilians Universität München, Munich, Germany (2007).
- *On Missing Data in SNP Association Studies, and What to Do About Them*. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (2007).
- *Logic Regression as a Statistical Tool to Assess Interactions in SNP Association Studies*. Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, MD (2007).
- *On Missing Data and Interactions in SNP Association Studies*. Department of Statistics, University of British Columbia, Canada (2006).
- *On Missing Data and Interactions in SNP Association Studies*. Dept. of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University, Washington, DC (2006).
- *Logic Regression*. Department of Statistics, Florida State University, Tallahassee, FL (2006).
- *Logic Regression*. Department of Statistics and Applied Probability, University of California, Santa Barbara, CA (2006).
- *Uncertainty about Φ Values*. Dept. of Biophysics, Johns Hopkins University, Baltimore, MD (2005).
- *Assessing Interactions in High-Dimensional Genomic Data*. Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT (2005).
- *Missing Data and Gene Interactions in SNP Association Studies*. Centre National de Genotypage, Paris-Evry, France (2005).
- *Protein Folding and Structure Prediction - A Statistician's View*. Seminar für Statistik, Eidgenössische Technische Hochschule Zürich, Switzerland (2005).
- *A Statistician's View on Protein Folding*. Fred Hutchinson Cancer Research Center, Seattle, WA (2005).
- *Protein Folding and Structure Prediction - A Statistician's View*. Department of Statistics, Rutgers University, Piscataway, NJ (2005).
- *Protein Folding and Structure Prediction - A Statistician's View*. Department of Mathematics and Statistics, McGill University, Montreal, Canada (2005).
- *Functional Inferences from Blind Ab Initio Protein Structure Predictions*. Institut für Statistik, Ludwig-Maximilians Universität München, Munich, Germany (2004).
- *Logic Regression*. US Census Bureau, Suitland, MD (2004).
- *Logic Regression and Interactions in High Dimensional Genomic Data*. Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD (2003).
- *Logic Regression and Interactions in High Dimensional Genomic Data*. Memorial Sloan-Kettering Cancer Center, New York, NY (2003).
- *Logic Regression and its Applications to SNP Association Studies*. NCI National Cancer Institute, Rockville, MD (2002).
- *Logic Regression*. General Electric, Schenectady, NY (2002).

- *Logic Regression with Application to SNP Data Analysis*. Wadsworth Center for Labs and Research, Albany, NY (2002).
- *A New Regression Methodology using Boolean Logic*. Department of Biostatistics, Johns Hopkins University, Baltimore, MD. (2001).
- *A New Regression Methodology using Boolean Logic*. Department of Statistics, Columbia University, New York, NY (2001).

Other Meetings and Events

- *On Missing Data and Genotyping Errors in Association Studies*. Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, MD (2008).
- *Novel Statistical Methods for Gene-Environment Interactions in Complex Diseases*. National Institutes of Health, Bethesda MD (2008).
- *Bioinformatics - Some Selected Examples and a Bit of an Overview*. EnviroHealth Connections Summer Institute, Baltimore MD (2007).
- *Statistical Approaches to assess Gene Copy Number and Loss of Heterozygosity*. Expressionists Meeting, Genome Cafe, Johns Hopkins University, Baltimore, MD (2007).
- *Statistics Schmatistics: On the Folded, the Unfolded, and the Transition State*. Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins University, Baltimore, MD (2006).
- *On Missing Genotype Data*. Expressionists Meeting, Johns Hopkins University, Baltimore, MD (2006).
- *Visualizing and Analyzing High Density SNP Data with SNPscan*. Dept. of Biostatistics Retreat, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (2006).
- *SNPscan: A Web Based Tool and R Package to Visualize and Analyze Affymetrix SNP Chip Data*. SNP Working Group, School of Public Health, Johns Hopkins University, Baltimore, MD (2005).
- *Are You Gellin'?* Bayview NHLBI Proteomics Center, Johns Hopkins University, Baltimore, MD (2005).
- *MCMC Methods to Detect Gene-Gene Interactions*. SNP Working Group, School of Public Health, Johns Hopkins University, Baltimore, MD (2005).
- *Are You Gellin'?* Genome Cafe Opening, Bloomberg School of Public Health, Baltimore, MD (2005).
- *An Add-on R Package for Rosetta*. Rosetta Developers Meeting, Leavenworth, WA (2005).
- *The Standard Error of the Lab Scientist*. Wall of Wonder, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (2005).
- *Some SNP Related Stuff I am Working On*. Pulver Lab, Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD (2005).
- *How Do Proteins Fold?* Department of Biostatistics Faculty Meeting, Johns Hopkins University, Baltimore, MD (2004).
- *Protein Structure and Folding*. Expressionists Meeting, Genome Cafe, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (2004).
- *On β Sheets in Proteins with Application to Structure Prediction*. Rosetta Developers Meeting, Leavenworth, WA (2003).
- *A Journey from Gene Expression to Protein Structure*. American Statistical Association Chapter Meeting, Albany, NY (2002).

Posters

- *Some Basic Considerations with Regards to Chevron Curves and Φ -value Estimation.* Gordon Research Conference on Protein Folding Dynamics, Ventura, CA (2008).
- *Inference in Gene Association Studies of Cancer Risks with Partially Missing Family History Data.* The 97th Annual Meeting of the American Association for Cancer Research, Washington, DC (2006).
- *Uncertainty about Φ Values.* The 8th Johns Hopkins Folding Meeting, St.Michaels, MD (2005).
- *Statistical Issues in Protein Folding.* Johns Hopkins Biostatistics Retreat, St.Michaels, MD (2004).
- *Interactions and Variable Importance in Genomic Data.* Workshop on Data Mining, Statistical Learning & Bioinformatics, University of Florida, Gainesville, FL (2004).
- *Distributions of β Sheets in Proteins with Application to Structure Prediction.* Johns Hopkins Biostatistics Retreat, St.Michaels, MD (2003).
- *Exploring Interactions in Genomic Data using Logic Regression Poster.* International Conference on Research in Computational Molecular Biology RECOMB, Berlin, Germany (2003).

ADDITIONAL INFORMATION

Personal Statement of Research Objectives

I develop methods and software for the analysis of data arising from complex biological processes, predominantly concerning DNA (gene-gene and gene-environment interactions in SNP association studies; chromosomal alterations such as aneuploidies and loss of heterozygosity) and proteins (protein folding, folding kinetics, and protein structure).

Keywords

Bioinformatics, computational biology, copy number, folding kinetics, genetics, genomics, high throughput, interactions, loss of heterozygosity, machine learning, protein folding, protein structure, SNP association studies, statistical computing.