

SANDRINE DUDOIT

Associate Professor, Step III

Division of Biostatistics
School of Public Health
University of California, Berkeley
140 Earl Warren Hall, #7360
Berkeley, CA 94720-7360

Tel: (510) 643-1108
Fax: (510) 643-5163
E-mail: [sandrine@stat.berkeley.edu](mailto:sandrined@stat.berkeley.edu)
Website: www.stat.berkeley.edu/~sandrine

AFFILIATIONS

Graduate Group in Biostatistics, UC Berkeley (stat-www.berkeley.edu/biostat).

Graduate Group in Computational and Genomic Biology, UC Berkeley (computationalbiology.berkeley.edu).

California Institute for Quantitative Biomedical Research (QB3; www.qb3.org).

Center for Integrative Genomics, UC Berkeley (cigbrowser.berkeley.edu).

Center for Bioinformatics and Molecular Biostatistics, UC San Francisco (CBMB; www.biostat.ucsf.edu/cbmb).

Core Developer, Bioconductor Project (www.bioconductor.org).

RESEARCH INTERESTS

Loss-based estimation with cross-validation: classification, parametric and non-parametric density estimation, prediction, variable selection.

Multiple hypothesis testing: resampling-based multiple testing procedures for controlling Type I error rates defined as tail probabilities and expected values for arbitrary functions of the numbers of Type I errors and rejected hypotheses (e.g., false discovery rate).

Statistical computing (Bioconductor Project; www.bioconductor.org).

Applications to biomedical and genomic research: design and analysis of high-throughput gene expression experiments (e.g., cDNA microarrays, alternative splicing microarrays, ChIP-Chip experiments); nucleotide and protein sequence analysis (e.g., identification of regulatory motifs); genetic mapping of complex human traits; biological annotation metadata analysis (e.g., Gene Ontology (GO) annotation).

EDUCATION

PhD in Statistics, University of California, Berkeley, 1999.

Evelyn Fix Memorial Medal: “Awarded to PhD student showing the greatest promise in statistical research, with preference for applications to biology and problems of health”.

Dissertation title: *Linkage Analysis of Complex Human Traits Using Identity by Descent Data*.
Dissertation Committee: Terence P. Speed, Chair; David A. Freedman; Glenys J. Thomson.

MSc in Mathematics, Carleton University, Ottawa, ON, Canada, 1994.

Senate Medal: “For outstanding academic achievement at the Master’s level”.

BSc Highest Honours in Mathematics, Carleton University, Ottawa, ON, Canada, 1992.

Chancellor’s Medal: “Carleton University’s second-highest honour, awarded annually to a graduating student of outstanding academic achievement”.

Honours project: *Markov Chain Analysis of a Simplified Class of Genetic Algorithms*.

Adviser: Donald A. Dawson.

Diplôme du Baccalauréat de l’Enseignement du Second Degré, Mathématiques et Sciences Physiques, Mention Très Bien (French high-school diploma, with highest honours), Lycée Molière, Paris, France, 1988.

AWARDS AND FELLOWSHIPS

- Insightful Innovation Award for Bioconductor Project, 2003.
- Fellow, Functional Genomics Program, Institute for Pure and Applied Mathematics (IPAM), UCLA, Fall 2000.
- Postdoctoral Fellowship, Program in Mathematics and Molecular Biology (PMMB) – Burroughs Wellcome Fund, 2000.
- Postdoctoral Fellowship, Mathematical Sciences Research Institute (MSRI), Berkeley, 1999.
- Outstanding Graduate Student Instructor Award, UC Berkeley, 1998.
- Graduate Fellowship, Program in Mathematics and Molecular Biology (PMMB) – Burroughs Wellcome Fund, 1997 – 1998.
- Natural Sciences and Engineering Research Council (NSERC) of Canada – PGS B Award for Tenure Abroad (PhD), 1994 – 1996.
- Gertrude Cox Scholarship, 1994.
- Michel and Line Loève Fellowship, UC Berkeley, 1994.
- Dora Garibaldi Scholarship, UC Berkeley, 1994.
- Natural Sciences and Engineering Research Council (NSERC) of Canada – PGS A Award (MSc), 1992 – 1993.
- Carleton University Award for Graduate Studies, 1992 – 1993.
- David and Rachel Epstein Foundation Scholarship, Carleton University, 1993.
- NSERC Undergraduate Student Research Award, Summers 1991, 1992.
- Canada Scholarship, Science and Technology, 1988 – 1992.

- Dean's Honour List, Carleton University, 1989 – 1992.
- Richard J. Semple Memorial Award in Mathematics, Carleton University, 1991.
- Henry Campbell Scholarship, Carleton University, 1990, 1991.
- L. N. Wadlin Scholarship in Mathematics, Carleton University, 1990.
- Paul R. Beesack Memorial Scholarship in Mathematics, Carleton University, 1989.
- Ian H. Griffith Memorial Scholarship, Carleton University, 1989.
- Scholarship from the Crédit Industriel et Commercial for Baccalauréat, France, 1988.

EMPLOYMENT

Associate Professor, July 2005 – Present.

Division of Biostatistics, School of Public Health, University of California, Berkeley.

Assistant Professor, July 2001 – June 2005.

Division of Biostatistics, School of Public Health, University of California, Berkeley.

Postdoctoral Scientist, July 2000 – June 2001.

Laboratory of Professor Patrick O. Brown, Department of Biochemistry, Stanford University.

Research on the design and analysis of gene expression experiments using DNA microarrays.

Postdoctoral Scientist, September 1999 – June 2000.

Mathematical Sciences Research Institute, Berkeley.

Research on the design and analysis of gene expression experiments using DNA microarrays.

Visiting Postdoctoral Fellow, May 1999 – August 1999.

Department of Statistics, University of California, Berkeley.

Research on the linkage analysis of complex human traits and the analysis of gene expression data from DNA microarray experiments.

Graduate Student Researcher, January 1997 – May 1999.

Department of Statistics, University of California, Berkeley.

Research on statistical issues involved in the linkage analysis of complex human traits using identity by descent data.

Statistical Consulting and Lab Rotation, Fall 1997.

Genetics and Bioinformatics Group, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia.

Analysis of genetic data from experimental crosses; affected sib-pair analysis for mapping human diabetes genes; practical lab training: DNA preparation, PCR, gel electrophoresis.

Statistical Consultant, Fall 1996.

Statistical Consulting Service, Department of Statistics, University of California, Berkeley.

Provided statistical consulting to researchers from various groups on the UC Berkeley campus, including the Berkeley Drosophila Genome Project and the Departments of Integrative Biology and Economics.

Statistical Analyst, Summers 1990, 1994.

General Social Survey, Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, ON, Canada.

Translator, Statistical Society of Canada (SSC), 1994.

Translation from English to French of various SSC publications.

Translator, Canadian Journal of Statistics, 1993 – 1994.

Translation of abstracts from English to French.

GRANTS

- *Individual Genetic Susceptibility and Environmental Exposures in the Aetiology of Childhood Leukaemia*
CHILDREN with LEUKAEMIA (Reference #: 2005/028)
PI: Patricia A. Buffler, Division of Epidemiology, UC Berkeley
07/01/2005 – 06/30/2008.
- *Data Adaptive Estimation in Genomics and Epidemiology*
National Institutes of Health (R01 GM071397)
PI: Mark J. van der Laan, Division of Biostatistics, UC Berkeley
08/01/2004 – 07/31/2007.
- *OBELINKS: Combining Machine Learning and Biostatistics to Discover Significant Obesity-Related Genetic Polymorphisms*
France Berkeley Fund
PI: Sandrine Dudoit, Division of Biostatistics, UC Berkeley
10/15/2003 – 10/14/2004.
- *A Statistical Computing Framework for Genomic Data*
National Institutes of Health (R33 HG002708)
PI: Robert C. Gentleman, Program in Computational Biology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center
09/30/2003 – 06/30/2007.
- *CRCNS: Modeling Pathfinding and Target Recognition in the Olfactory System*
National Science Foundation (41964-7055)
PI: David M. Lin, Department of Biomedical Sciences, Cornell University
10/01/2002 – 09/30/2006.
- *Statistical Design and Analysis of Microarray Experiments*
National Institutes of Health (R01 LM07609)
PI: Sandrine Dudoit, Division of Biostatistics, UC Berkeley
08/01/2002 – 07/31/2006.

- *Statistical and Computational Methods for the Classification of Tumors Using Gene Expression Data from DNA Microarray Experiments*
Regents Junior Faculty Fellowship, UC Berkeley
PI: Sandrine Dudoit, Division of Biostatistics, UC Berkeley
Summer 2002.
- *Development of Statistical Methods for the Design and Analysis of DNA Microarray Experiments*
Research Enabling Grant, Committee on Research, UC Berkeley
PI: Sandrine Dudoit, Division of Biostatistics, UC Berkeley
2001 – 2002.
- *Statistical Methods for Gene Mapping*
National Institutes of Health (R01 GM59506)
PI: Terence P. Speed, Department of Statistics, UC Berkeley
1998 – 2001.

PUBLICATIONS

Books

1. **S. Dudoit** and M. J. van der Laan (In preparation). *Multiple Testing Procedures with Applications to Genomics*, Springer Series in Statistics.

Refereed Journal Publications

2. **S. Dudoit** and T. P. Speed (1999). Triangle constraints for sib-pair identity by descent probabilities under a general multilocus model for disease susceptibility. In M. E. Halloran and S. Geisser (eds), *Statistics in Genetics*, Vol. 112 of *IMA Volumes in Mathematics and its Applications*, Springer-Verlag, New York, p. 181–221.
3. **S. Dudoit** and T. P. Speed (1999). A score test for linkage using identity by descent data from sibships. *Annals of Statistics*, Vol. 27, No. 3, p. 943–986.
4. **S. Dudoit** and T. P. Speed (2000). A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data on sib-pairs. *Biostatistics*, Vol. 1, No.1, p. 1–26.
5. D. R. Goldstein, **S. Dudoit**, and T. P. Speed (2000). Power of a score test for quantitative trait linkage analysis of relative pairs. *Genetic Epidemiology*, Vol. 19, Suppl. 1, p. S85–S91.
6. M. J. Callow, **S. Dudoit**, E. L. Gong, T. P. Speed, and E. M. Rubin (2000). Microarray expression profiling identifies genes with altered expression in HDL deficient mice. *Genome Research*, Vol. 10, No. 12, p. 2022–2029.
7. D. R. Goldstein, **S. Dudoit**, and T. P. Speed (2001). Power and robustness of a score test for linkage analysis of quantitative traits using identity by descent data on sib pairs. *Genetic Epidemiology*, Vol. 20, No. 4, p. 415–431.

8. **S. Dudoit**, Y. H. Yang, T. P. Speed, and M. J. Callow (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, Vol. 12, No. 1, p. 111–139.
9. J. C. Boldrick, A. A. Alizadeh, M. Diehn, **S. Dudoit**, C. L. Liu, C. E. Belcher, D. Botstein, L. M. Staudt, P. O. Brown, and D. A. Relman (2002). Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc. Natl. Acad. Sci.*, Vol. 99, No. 2, p. 972–977.
10. **S. Dudoit**, Y. H. Yang, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, Vol. 30, No. 4, e15.
11. **S. Dudoit**, J. Fridlyand, and T. P. Speed (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, Vol. 97, No. 457, p. 77–87.
12. Y. H. Yang, M. J. Buckley, **S. Dudoit**, and T. P. Speed (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, Vol. 11, No. 1, p. 108–136.
13. X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K.-M. Lai, J. Ji, **S. Dudoit**, I. O. L. Ng, M. van de Rijn, D. Botstein, and P. O. Brown (2002). Gene expression patterns in human liver cancers. *Molecular Biology of the Cell*, Vol. 13, No. 6, p. 1929–1939.
14. **S. Dudoit** and J. Fridlyand (2002). A prediction-based resampling method to estimate the number of clusters in a dataset. *Genome Biology*, Vol. 3, No. 7, p. 0036.1–0036.21.
15. H. Y. Chang, J. T. Chi, **S. Dudoit**, C. Bondre, M. van de Rijn, D. Botstein, D and P. O. Brown (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl. Acad. Sci.*, Vol. 99, No. 20, p. 12877–12882.
16. **S. Dudoit** and D. R. Goldstein (2003). Extensions to a score test for genetic linkage with identity by descent data. In D. R. Goldstein (ed), *Science and Statistics: A Festschrift for Terry Speed*, Vol. 40 of *Institute of Mathematical Statistics, Lecture Notes – Monograph Series*, p. 307–319.
17. **S. Dudoit**, R. C. Gentleman, and J. Quackenbush (2003). Open source tools for microarray analysis. *Biotechniques Supplements, Microarrays and Cancer: Research and Applications*, p. 45–51.
18. **S. Dudoit** and J. Fridlyand (2003). Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, Vol. 19, No. 9, p. 1090–1099.
19. **S. Dudoit**, J. P. Shaffer, and J. C. Boldrick (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, Vol. 18, No. 1, p. 71–103.
20. Y. Ge, **S. Dudoit**, and T. P. Speed (2003). Resampling-based multiple testing for microarray data analysis. *TEST*, Vol. 12, No. 1, p. 1–44 (with discussion p. 44–77).

21. S. Keleş, M. J. van der Laan, **S. Dudoit**, B. Xing, and M. B. Eisen (2003). Supervised detection of regulatory motifs in DNA sequences. *Statistical Applications in Genetics and Molecular Biology*, Vol. 2, No. 1, Article 5.
22. **S. Dudoit**, M. J. van der Laan, S. Keleş, A. M. Molinaro, S. E. Sinisi, and S. L. Teng (2003). Loss-based estimation with cross-validation: Applications to microarray data analysis. In G. Piatetsky-Shapiro and P. Tamayo (eds), *Microarray Data Mining*, Special Issue of *SIGKDD Explorations*, Vol. 5, No. 2, p. 56–68.
23. M. J. van der Laan, **S. Dudoit**, and S. Keleş (2004). Asymptotic optimality of likelihood based cross-validation. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 4.
24. A. M. Molinaro, **S. Dudoit**, and M. J. van der Laan (2004). Tree-based multivariate regression and density estimation with right-censored data. In **S. Dudoit**, R. C. Gentleman, and M. J. van der Laan (eds), *Multivariate Methods in Genomic Data Analysis*, Special Issue of *Journal of Multivariate Analysis*, Vol. 90, No. 1, p. 154–177.
25. **S. Dudoit**, M. J. van der Laan, and K. S. Pollard (2004). Multiple testing. Part I. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 13.
26. M. J. van der Laan, **S. Dudoit**, and K. S. Pollard (2004). Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 14.
27. M. J. van der Laan, **S. Dudoit**, and K. S. Pollard (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 15.
28. S. Keleş, M. J. van der Laan, and **S. Dudoit** (2004). Asymptotically optimal model selection method with right censored outcomes. *Bernoulli*, Vol. 10, No. 6, p. 1011–1037.
29. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, **S. Dudoit**, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. K. Smyth, L. Tierney, Y. H. Yang, and J. Zhang (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, Vol. 5, No. 10, Article R80.
30. **S. Dudoit** and M. J. van der Laan (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, Vol. 2, No. 2, p. 131–154.
31. A. Barrier, A. Lemoine, P.-Y. Boëlle, C. Tse, D. Brault, F. Chiappini, J. Breittschneider, F. Lacaine, S. Houry, M. Huguier, M. J. van der Laan, T. P. Speed, B. Debuire, A. Flahault, and **S. Dudoit** (2005). Colon cancer prognosis prediction by gene expression profiling. *Oncogene*, Vol. 24, No. 40, p. 6155–6164.

32. A. Barrier, N. Olaya, F. Chiappini, F. Roser, O. Scatton, C. Artus, B. Franc, **S. Dudoit**, A. Flahault, B. Debuire, D. Azoulay, and A. Lemoine (2005). Ischemic preconditioning modulates the expression of several genes, leading to the overproduction of IL-1Ra, iNOS, and Bcl-2 in a human model of liver ischemia-reperfusion. *The FASEB Journal*, Vol. 19, No. 12, p. 1617–1626.
33. A. Barrier, P.-Y. Boëlle, A. Lemoine, C. Tse, D. Brault, F. Chiappini, F. Lacaine, S. Houry, M. Huguier, A. Flahault, and **S. Dudoit** (2005). Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. *Diseases of the Colon and Rectum*, Vol. 48, No. 12, p. 2238–2248.
34. K. S. Pollard, M. D. Birkner, M. J. van der Laan, and **S. Dudoit** (2005). Test statistics null distributions in multiple testing: Simulation studies and applications to genomics. Numéro double spécial *Statistique et Biopuces, Journal de la Société Française de Statistique*, Vol. 146, No. 1–2, p. 77–115.
35. F. Chiappini, A. Barrier, R. Saffroy, M.-C. Domart, N. Dagues, D. Azoulay, M. Sebah, B. Franc, S. Chevalier, B. Debuire, **S. Dudoit**, and A. Lemoine (2006). Exploration of global gene expression in human liver steatosis by high-density oligonucleotide microarray. *Laboratory Investigation*, Vol. 86, No. 2, p. 154–165.
36. T. Hothorn, P. Bühlmann, **S. Dudoit**, A. M. Molinaro, and M. J. van der Laan (2006). Survival ensembles, Vol. 7, No. 3, p. 355–373. *Biostatistics*.
37. Y. Wang, L. P. Zhao, and **S. Dudoit** (2006). A fine-scale linkage-disequilibrium measure based on length of haplotype sharing. *American Journal of Human Genetics*, Vol. 78, No. 4, p. 615–628.
38. S. Keleş, M. J. van der Laan, **S. Dudoit**, and S. E. Cawley (2006). Multiple testing methods for ChIP-Chip high density oligonucleotide array data. *Journal of Computational Biology*, Vol. 13, No. 3, p. 579–613.
39. D. Rubin, M. J. van der Laan, and **S. Dudoit** (2006). A method to increase the power of multiple testing procedures through sample splitting. *Statistical Applications in Genetics and Molecular Biology*, Vol. 5, No. 1, Article 19.
40. A. Barrier, P.-Y. Boëlle, F. Roser, J. Gregg, C. Tse, D. Brault, F. Lacaine, S. Houry, M. Huguier, B. Franc, A. Flahault, A. Lemoine, and **S. Dudoit** (2006). Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of Clinical Oncology*, Vol. 24, No. 29, p. 4685–4691.
41. B. Durbin, **S. Dudoit**, and Mark J. van der Laan (2006). A Deletion/Substitution/Addition algorithm for classification neural networks, with applications to biomedical data. In S. Gupta and R. Mukerjee (eds), *Statistical Design of Medical Experiments III*, Special Issue of *Journal of Statistical Planning and Inference*. (Accepted).
42. M. J. van der Laan, **S. Dudoit**, and A. W. van der Vaart (2006). The cross-validated adaptive epsilon-net estimator. *Statistics & Decisions*. (Accepted).

43. A. W. van der Vaart, M. J. van der Laan, and **S. Dudoit** (2006). Oracle inequalities for k -fold cross-validation. *Statistics & Decisions*. (Accepted).

Book Chapters and Conference Proceedings

44. **S. Dudoit**, Y. H. Yang, P. Luu, and T. P. Speed (2001). Normalization for cDNA microarray data. In M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty (eds), *Microarrays: Optical Technologies and Informatics*, Vol. 4266 of *Proceedings of SPIE*, p. 141–152.
45. **S. Dudoit** and J. Fridlyand (2003). Introduction to classification in microarray experiments. In D. P. Berrar, W. Dubitzky, and M. Granzow (eds), *A Practical Approach to Microarray Data Analysis*, Kluwer, Chapter 7, p. 132–149.
46. **S. Dudoit** and J. Fridlyand (2003). Classification in microarray experiments. In T. P. Speed (ed), *Statistical Analysis of Gene Expression Microarray Data*, Chapman & Hall/CRC, Chapter 3, p. 93–158.
47. **S. Dudoit** and Y. H. Yang (2003). Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In G. Parmigiani, E. S. Garrett, R. A. Irizarry, and S. L. Zeger (eds), *The Analysis of Gene Expression Data: Methods and Software*, Springer, New York, p. 73–101.
48. **S. Dudoit** and M. J. van der Laan (2003). Unified cross-validation methodology for estimator selection and applications to genomics. *Bulletin of the International Statistical Institute, 54th Session Proceedings*, Vol. LX, Book 2, p. 412–415.
49. R. Gentleman, B. Ding, **S. Dudoit**, and J. Ibrahim (2005). Distance measures in DNA microarray data analysis. In R. C. Gentleman, V. J. Carey, W. Huber, R. Irizarry, and **S. Dudoit** (eds), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer-Verlag, New York, Chapter 12, p. 189–208.
50. K. S. Pollard, **S. Dudoit**, and M. J. van der Laan (2005). Multiple testing procedures: the `multtest` package and applications to genomics. In R. C. Gentleman, V. J. Carey, W. Huber, R. Irizarry, and **S. Dudoit** (eds), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer-Verlag, New York, Chapter 15, p. 249–271.

Technical Reports and Submitted Manuscripts

51. **S. Dudoit**, Y. H. Yang, and B. Bolstad (2002). Using R for the analysis of DNA microarray data. *R News*, Vol. 2, No. 1, p. 24–32.
52. **S. Dudoit** (2003). IBD configuration transition matrices and linkage score tests for unilineal relative pair. Technical Report #128, Division of Biostatistics, UC Berkeley.
53. M. J. van der Laan and **S. Dudoit** (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive ϵ -net estimator: Finite sample oracle inequalities and examples. Technical Report #130, Division of Biostatistics, UC Berkeley.
54. M. J. van der Laan, **S. Dudoit**, and A. W. van der Vaart (2004). The cross-validated adaptive ϵ -net estimator. Technical Report #142, Division of Biostatistics, UC Berkeley.

55. Y. Wang and **S. Dudoit** (2004). Quantification and visualization of LD patterns and identification of haplotype blocks. Technical Report #150, Division of Biostatistics, UC Berkeley.
56. **S. Dudoit**, M. J. van der Laan, and M. D. Birkner (2004). Multiple testing procedures for controlling tail probability error rates. Technical Report #166, Division of Biostatistics, UC Berkeley.
57. M. D. Birkner, K. S. Pollard, M. J. van der Laan, and **S. Dudoit** (2005). Multiple testing procedures and applications to genomics. Technical Report #168, Division of Biostatistics, UC Berkeley.
58. B. Durbin, **S. Dudoit**, and M. J. van der Laan (2005). Optimization of the architecture of neural networks using a Deletion/Substitution/Addition algorithm. Technical Report #170, Division of Biostatistics, UC Berkeley.
59. D. Rubin, **S. Dudoit**, and Mark J. van der Laan (2005). A method to increase the power of multiple testing procedures through sample splitting. Technical Report #171, Division of Biostatistics, UC Berkeley.
60. A. Barrier, M. J. van der Laan, and **S. Dudoit** (2005). Colon cancer prognosis prediction by gene expression profiling. Technical Report #178, Division of Biostatistics, UC Berkeley.
61. A. Barrier, M. J. van der Laan, and **S. Dudoit** (2005). Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. Technical Report #179, Division of Biostatistics, UC Berkeley.
62. **S. Dudoit**, S. Keleş, and M. J. van der Laan (2006). Multiple tests of association with biological annotation metadata. Technical Report #202, Division of Biostatistics, UC Berkeley.
63. D. Shilane, J. Martikainen, **S. Dudoit**, and S. Ovaska (2006). A general framework for statistical performance comparison of evolutionary computation algorithms. Technical Report #204, Division of Biostatistics, UC Berkeley.

Editor

64. **S. Dudoit**, R. C. Gentleman, and M. J. van der Laan (eds) (2004). *Multivariate Methods in Genomic Data Analysis*, Special Issue of *Journal of Multivariate Analysis*, Vol. 90, No. 1.
65. R. C. Gentleman, V. J. Carey, W. Huber, R. Irizarry, and **S. Dudoit** (eds) (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer-Verlag, New York.

Manuscripts in Preparation

66. M. D. Birkner, M. Courtine, **S. Dudoit**, M. J. van der Laan, K. Clément, and J-D. Zucker (2006). Statistical methods for detecting genotype/phenotype associations in the ObeLinks Project.

Software Packages

1. **S. Dudoit**, Y. H. Yang, and B. M. Bolstad, with contributions from N. Thorne, I. Lönnstedt, and J. Mar. `sma`: R package for the analysis of microarray data (cran.r-project.org).
2. **S. Dudoit** and Y. H. Yang. `marrayClasses`, `marrayInput`, `marrayNorm`, and `marrayPlots`: Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data (www.bioconductor.org).
3. K. S. Pollard, Y. Ge, and **S. Dudoit**. `multtest`: Bioconductor R package for multiple hypothesis testing (www.bioconductor.org).

PRESENTATIONS

Invited Presentations

- Spring 2006 School of Public Health Research Symposium, University of California, Berkeley, February 14, 2006.
Multiple Testing Procedures and Applications to Biomedical and Genomic Data Analysis.
- Department of Statistics, University of California, Davis, December 01, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- Department of Human Genetics, University of California, Los Angeles, November 21, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- *Emerging Issues in Molecular Epidemiology* Short Course, The 9th International Conference on Environmental Mutagens, San Francisco, CA, September 03, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- Joint Statistical Meetings, Minneapolis, MN, August 07 – 11, 2005.
Organizer and Chair, Session 76 – *Recent advances in multiple testing.*
- Workshop in Biostatistics, Department of Health Research and Policy, Stanford University, Stanford, CA, April 07, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- Gordon Research Conference on Quantitative Genetics and Genomics, Ventura, CA, February 20 – 25, 2005.
Discussion Leader for session on *False Discovery Rate (FDR) for Genetic and Genomic Data.*
- Department of Biostatistics, Johns Hopkins University, Baltimore, MD, November 03, 2004.
Multiple testing procedures and applications to genomics.
- Cytokinetics, Inc., South San Francisco, CA, October 27, 2004.
Introduction to R and Bioconductor software.
- *Advances in Computational Biology and Bioinformatics* Workshop, 26th Annual International Conference IEEE Engineering in Medicine and Biology Society (EMBS), San Francisco, CA, September 01, 2004.
Multiple testing procedures and applications to genomics.

-
- Statistics in Functional Genomics, Centro Stefano Franscini, Ascona, Switzerland, June 27 – July 02, 2004.
Multiple testing procedures and applications to genomics.
 - The Second Erich L. Lehmann Symposium, Rice University, Houston, TX, May 19 – 22, 2004.
Multiple testing procedures: Applications to genomics.
 - Department of Statistics Colloquium, University of California, Berkeley, March 30, 2004.
Multiple testing procedures: Applications to genomics.
 - Department of Mathematics and Statistics, Utah State University, Logan, UT, March 16, 2004.
Multiple testing procedures: Applications to genomics.
 - Statistics and Genomics Seminar, University of California, Berkeley, February 19, 2004.
Multiple testing procedures: Applications to genomics.
 - Genetics of Complex Disease Workshop, Mathematical Sciences Research Institute, Berkeley, CA, February 11, 2004.
Multiple testing procedures and applications to genomic data analysis.
 - Department of Statistics, University of California, Riverside, November 25, 2003.
Multiple hypothesis testing in microarray data analysis.
 - ESF Exploratory Workshop, Genomic Approaches to Microarray Data Analysis, Madrid, Spain, October 30, 2003.
Resampling-based multiple testing: Applications to microarray data analysis.
 - IMA Workshop 1: Statistical Methods for Gene Expression: Microarrays and Proteomics, Minneapolis, MN, September 30, 2003.
Data-adaptive loss-based estimation with cross-validation: Applications to microarray and DNA sequence analysis.
 - NAS Committee on Emerging Issues and Data on Environmental Contaminants, Seattle, WA, September 16, 2003.
Statistical issues in – omic data analysis.
 - International Statistical Institute, 54th Session, Berlin, Germany, August 13 – 20, 2003.
 - Speaker, Invited Paper Meeting 76 – *Gene expression data: Unified cross-validation methodology for estimator selection and applications to genomics.*
 - Discussant, Invited Paper Meeting 17 – *Statistical methods for high dimensional data.*
 - Joint Statistical Meetings, San Francisco, CA, August 03 – 07, 2003.
 - Invited poster presenter, Session 64: *The Bioconductor Project: Open-source statistical software for the analysis of genomic data.*
 - Organizer and Chair, Session 351 – *Statistics and Genomics.*
 - Workshop in Biostatistics, Department of Health Research and Policy, Stanford University, Stanford, CA, May 29, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of genomic data.

- Statistics and Genomics Seminar, University of California, Berkeley, April 10, 2003.
Unified cross-validation methodology for estimator selection and applications to genomics.
- Genomics and Computational Biology Seminar, University of California, Berkeley, February 19, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of microarray data.
- American Association for the Advancement of Science Annual Meeting, Microarrays and Functional Genomics (organized by the Microarray Gene Expression Data (MGED) Society and Science magazine), Denver, CO, February 15, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of microarray data.
- Department of Computer Engineering, University of California, Santa Cruz, January 29, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of microarray data.
- Department of Biostatistic, University of California, San Francisco, January 15, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of microarray data.
- Department of Biostatistics and Department of Statistics Joint Seminar, University of Michigan, Ann Arbor, MI, November 21, 2002.
Statistical methods and software for the analysis of DNA microarray experiments.
- Joint Statistical Meetings, New York, NY, August 11 – 15, 2002.
Classification in DNA microarray experiments.
- Applications of Bioinformatics in Cancer Detection Workshop, National Cancer Institute, National Institutes of Health, Bethesda, MD, August 06 – 07, 2002.
Statistical methods and software for the analysis of DNA microarray experiments.
- European Society of Mathematical and Theoretical Biology, 5th tri-annual conference, Mathematical Modelling & Computing in Biology and Medicine, Milano, Italy, July 02 – 06, 2002.
Statistical methods for the design and analysis of DNA microarray experiments.
- Midwest Biopharmaceutical Statistics Workshop, Muncie, IN, May 20 – 22, 2002.
Statistical methods for the design and analysis of DNA microarray experiments.
- Bioinformatics and Computational Biology Workshop, Fundación BBVA, Madrid, Spain, April 25 – 26, 2002.
Statistical methods and software for the analysis of DNA microarray experiments.
- Program in Mathematics and Molecular Biology Meeting VII, Santa Fe, NM, January 05 – 10, 2002.
Statistics Tutorial.

-
- Cytokinetics, Inc., South San Francisco, CA, November 27, 2001.
Statistical methods for the design and analysis of DNA microarray experiments.
 - 3rd Annual Research Symposium, School of Public Health, University of California, Berkeley, October 11, 2001.
Statistical methods for the design and analysis of DNA microarray experiments.
 - International Society for Clinical Biostatistics Annual Conference, Stockholm, Sweden, August 2001.
Identifying differentially expressed genes in microarray experiments.
 - SSC/WNAR/IMS Meeting, Burnaby, BC, Canada, June 2001.
A score test for the linkage analysis of complex human traits.
 - The Association for Research in Vision and Ophthalmology (ARVO) Annual Meeting, Fort Lauderdale, FL, April 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Astra Zeneca, Göteborg, Sweden, April 2001.
Identifying differentially expressed genes in microarray experiments.
 - Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, April 2001.
Identifying differentially expressed genes in microarray experiments.
 - International Biometric Society, ENAR Meeting, Charlotte, NC, March 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Department of Statistics, Purdue University, West Lafayette, IN, March 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Department of Statistics, University of California, Davis, March 2001.
A score test for the genetic mapping of complex human traits.
 - Bioinformatics Forum, Center for Bioinformatics, Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, PA, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Department of Statistics, Stanford University, Stanford, CA, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Department Statistics, University of Toronto, Toronto, ON, Canada, February 2001.
A score test for the genetic mapping of complex human traits.
 - Department of Mathematics and Statistics, McGill University, Montreal, PQ, Canada, February 2001.
A score test for the genetic mapping of complex human traits.

-
- Genetics Seminar, McGill University Health Center – Research Institute, Montreal Genome Center, Montreal, PQ, Canada, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
 - Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada, February 2001.
A score test for the genetic mapping of complex human traits.
 - Department of Biostatistics, Johns Hopkins University, Baltimore, MD, January 2001.
A score test for the genetic mapping of complex human traits.
 - Division of Biostatistics, University of California, Berkeley, January 2001.
A score test for the genetic mapping of complex human traits.
 - Expression Arrays, Genetic Networks and Diseases, Institute for Pure and Applied Mathematics, University of California, Los Angeles, November 2000.
Applications of resampling methods for the cluster analysis of tumors using microarray data.
 - Department of Statistics, Stanford University, Stanford, CA, July 2000.
A score test for the linkage analysis of complex human traits using identity by descent data.
 - Mathematical and Computational Challenges in Molecular and Cell Biology, Mathematical Sciences Research Institute, Berkeley, CA, June 2000.
Linkage analysis of complex human traits using identity by descent data.
 - Department of Statistics, University of California, Los Angeles, May 2000.
A linkage score test based on identity by descent data.
 - Department of Statistical Science, Southern Methodist University, Dallas, TX, April 2000.
Comparison of discrimination methods for the classification of tumors using gene expression data.
 - Computational Biology Seminar, Division of Biostatistics, University of California, Berkeley, March 2000.
Comparison of discrimination methods for the classification of tumors using gene expression data.
 - Microarray Algorithms and Statistical Analysis: Methods and Standards, Tahoe City, CA, November 1999.
Statistical methods for the classification of cancer using gene expression data.
 - Bay Area Population Genetics and Genetic Epidemiology Seminar, University of California, San Francisco, May 1999.
A linkage score test based on identity by descent data.
 - Neyman Seminar, Department of Statistics, University of California, Berkeley, May 1999.
A linkage score test based on identity by descent data.
 - International Biometric Society, ENAR Meeting, Atlanta, GA, March 1999.
A linkage score test based on identity by descent data.

- Department of Mathematics and Statistics, Simon Fraser University, Vancouver, BC, Canada, February 1998.
Mapping disease genes using identity by descent data.
- Genetics Seminar, Walter and Eliza Hall Institute/Murdoch Institute, Melbourne, Australia, November 1997.
The affected sib-pair method.

Contributed Presentations

- NIH Statistical Genetics Initiative Workshop, Half Moon Bay, CA, October 1999.
Oral presentation: *A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data from sibships.*
- Nature Genetics Microarray Meeting, Scottsdale, AZ, September 1999.
Poster: *Statistical methods for the characterization of tumor classes using cDNA microarray data.*
- Program in Mathematics and Molecular Biology Meeting VI, Santa Fe, NM, January 1999.
Poster: *Sib-pair linkage score test for qualitative and quantitative traits.*
- Doing Science at the Interface, Burroughs Wellcome Fund Workshop, Berkeley, CA, June 1998.
Poster: *A linkage score test for use with identity by descent data on sib-pairs.*
- Program in Mathematics and Molecular Biology Retreat, Warrenton, VA, May 1997.
Oral presentation: *Mapping disease genes using identity by descent data.*
- Program in Mathematics and Molecular Biology Meeting V, Santa Fe, NM, January 1997.
Poster: *Searching for a human obesity gene.*

Organizing and Program Committee Membership

- *MCP 2007 – The 5th International Conferences on Multiple Comparison Procedures.*
Vienna, Austria, July 09 – 11, 2007 (Member of Organizing Committee).
mcp-conference.org/2007.
- *CIBB 2007 – The Fourth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics.*
Portofino, Italy, July 07 – 10, 2007 (Member of Scientific Committee).
- *RECOMB 2007 – Eleventh Annual International Conference on Research in Computational Molecular Biology.*
San Francisco Bay Area, April 21 – 25, 2007 (Chair of Organizing Committee).
www.recomb2007.com.
- *Computational and Statistical Genomics.*
Banff International Research Station, Banff, Canada, July 08 – 13, 2006 (Member of Organizing Committee, with Jennifer Bryan, Sündüz Keleş, Katherine S. Pollard, Mark J. van der Laan; funded on a competitive basis).
www.pims.math.ca/birs/birspages.php?task=displayevent&event_id=06w5076.

- *useR! – The R User Conference 2006.*
Vienna, Austria, June 15 – 17, 2006 (Member of Program Committee).
www.r-project.org/useR-2006.
- *Second Annual Retreat of the UC Berkeley Graduate Group in Computational and Genomic Biology.*
Bodega Bay Spa and Lodge, Bodega Bay, CA, April 29 – 30, 2006 (Committee Chair).
- *First Annual Retreat of the UC Berkeley Graduate Group in Computational and Genomic Biology.*
Marconi Conference Center, Marshall, CA, April 01 – 03, 2005 (Committee Chair).
- *Statistical Science for Genome Biology.*
Banff International Research Station, Banff, Canada, August 14 – 19, 2004 (Member of Organizing Committee, with Jennifer Bryan and Mark J. van der Laan; funded on a competitive basis).
www.stat.berkeley.edu/~sandrine/Public/BIRS04.
- *Challenges in the Statistical Analysis of Genomic Data.*
American Association for the Advancement of Science Annual Meeting, Denver, CO, February 14, 2003 (Session Organizer, with Juliet P. Shaffer; funded on a competitive basis).
- *Mathematics and Computational Biology of Genome Analysis.*
Mathematical Sciences Research Institute, Berkeley, CA, June 2000 (Member of Organizing Committee, with Kevin Atteson, Richard M. Karp, Terence P. Speed, De Witt L. Sumners).

Short Course Development and Instruction

- *Statistical Methods and Software for the Analysis of Microarray Experiments.*
Mathematical Biosciences Institute, Ohio State University, Columbus, OH, September 20 – 24, 2004.
www.stat.berkeley.edu/~sandrine/Docs/Talks/MBI04/mbi.html.
- *Tutorial III: Hands on Analysis.*
7th International Meeting of the Microarray Gene Expression Data (MGED) Society, Toronto, Canada, September 08, 2004.
- *CBMB and QB3 Short Course: Analysis of Gene Expression Microarray Data.*
Genentech Hall Auditorium, Mission Bay, University of California, San Francisco, November 15, 2003.
www.biostat.ucsf.edu/cbmb/courses/course.html.
- *Practical Statistical Analysis of DNA Microarray Data.*
KolleKolle, Denmark, October 26 – 28, 2003.
www.biostat.ku.dk/~pd/bioC-2003.
- *Bioconductor Tutorial.*
6th International Meeting of the Microarray Gene Expression Data (MGED) Society, Aix-en-Provence, France, September 03 – 05, 2003.
www.bioconductor.org/workshops/MGED6/mged6.html.

- *Statistical Methods and Software for the Analysis of DNA Microarray Data.*
Making and Using DNA Microarrays Short Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, June 21, 2003.
- *Statistical Methods and Software for the Analysis of DNA Microarray Experiments.*
International Biometric Society, ENAR Meeting, Tampa, FL, March 30, 2003.
www.bioconductor.org/workshops/ENAR03/enar03.html.
- *The Bioconductor Project: Open-source Statistical Software for the Analysis of Microarray Data.*
EMBO Practical Course on Analysis and Informatics of Microarray Data, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, March 18, 2003.
www.bioconductor.org/workshops/EMBO03/embo03.html.
- *Analyzing DNA Microarray Data Using Bioconductor.*
Fred Hutchinson Cancer Research Center, Seattle, WA, December 04 – 06, 2002.
www.bioconductor.org/workshops/Seattle02/index.html.
- *Analyzing DNA Microarray Data Using Bioconductor.*
Microarray Core Facility, Johns Hopkins Medical Institutions, Baltimore, MD, October 28 – 29, 2002.
www.bioconductor.org/workshops/Hopkins02/index.html.
- *Bioconductor: A Statistical Computing Project for the Analysis of Genomic Data.*
Short Course on Mathematical Approaches to the Analysis of Complex Phenotypes, The Jackson Laboratory, Bar Harbor, ME, September 18 – 24, 2002.
www.bioconductor.org/workshops/JAX02/index.html.
- *Statistical Methods and Software for the Analysis of DNA Microarray Experiments.*
8th Applied Statistics Week, Statistical Genetics, Pompeu Fabra University, Barcelona, Spain, June 28 – 29, 2002.
www.bioconductor.org/workshops/Summer02Course/index.html.
- *Statistics and Genomics.*
Department of Biostatistics, Harvard School of Public Health, Boston, MA, January 23 – 25, 2002.
www.bioconductor.org/workshops/ShortCourse012302/index.html.
- *Introduction to the Statistical Analysis of Microarray Data.*
Temple University at Fort Washington, Fort Washington, PA, June 15, 2001.
Co-designed and co-taught this one-day course sponsored by the ASA Philadelphia Chapter. Topics included image analysis, normalization, identification of differentially expressed genes, cluster and discriminant analysis applied to the classification of tumors using gene expression data.
- *Program in Mathematics and Molecular Biology Short Course.*
Berkeley, CA, June 1998, June 2000.
Co-designed and co-taught the computational biology labs. Topics included protein structure prediction and visualization, database searches, sequence alignment, computational gene finding, and phylogeny.

EDITORIAL AND REFEREE SERVICE**Associate Editor**

- *Annals of Applied Statistics* (), Summer 2006 – Present.
- *Biology Direct* (www.biology-direct.com), Spring 2006 – Present.
- *BMC Bioinformatics* (www.biomedcentral.com/bmcbioinformatics), Spring 2005 – Present.
- *Genomics* (www.elsevier.com/wps/find/journaldescription.cws_home/622838/description?navopenmenu=-2), Summer 2006 – Present.
- *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (www.computer.org/tcbb), Spring 2004 – Present.
- *Journal of Statistical Software* (www.jstatsoft.org), Fall 2005 – Fall 2006.
- *Statistical Applications in Genetics and Molecular Biology* (www.bepress.com/sagmb), Spring 2002 – Present.

Administrator

- *Bioconductor Project Working Papers* (www.bepress.com/bioconductor), Spring 2004 – Present.

Referee

- *Bioinformatics*.
- *Biometrics*.
- *Biostatistics*.
- *Genetic Epidemiology*.
- *Journal of the American Statistical Association*.
- *Journal of Computational Biology*.
- *Journal of Statistical Planning and Inference*.
- *Proceedings of the National Academy of Science*.
- *RECOMB Proceedings* (Annual Conference on Research in Computational Molecular Biology)
- *Springer*.
- *Statistical Methods in Medical Research*.
- *Statistica Sinica*.

External Advisor

- *Gene Expression Pattern Analysis (GEPAS)*, Bioinformatics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain; bioinfo.ochoa.fib.es (Fall 2005).
- *First Call for Research Proposals in Bioinformatics*, Fundación BBVA, Madrid, Spain; www.fbbva.es (Fall 2003).

PROFESSIONAL MEMBERSHIPS

- Western North American Region (WNAR) of the International Biometric Society (IBS), Member and Regional Committee Representative.
- American Statistical Association, Member.
- Institute of Mathematical Statistics, Member.

TEACHING

- **Instructor**, Fall 2001 – Present.
Division of Biostatistics, University of California, Berkeley.
 - **Fall 2006:** PB HLTH 296, Section 033 – *Biostatistics Laboratory*.
 - **Fall 2006:** PB HLTH 292, Section 013 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH292.F06/mainPH292F06.html.
 - **Fall 2006:** PB HLTH 240C, Section 001 – *Biostatistical Methods: Computational Techniques with Applications*.
www.stat.berkeley.edu/~sandrine/Teaching/PH240C.F06/PH240CF06.html.
 - **Spring 2006:** PB HLTH 296, Section 036 – *Biostatistics Laboratory*.
 - **Spring 2006:** PB HLTH 292, Section 020 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH292.S06/sem.html.
 - **Fall 2005:** PB HLTH 296, Section 033 – *Biostatistics Laboratory*.
 - **Fall 2005:** PB HLTH 292, Section 013 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH292.F05/sem.html.
 - **Spring 2005:** PB HLTH 240D – *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.
www.stat.berkeley.edu/~sandrine/Teaching/PH240D.S05.
 - **Fall 2004:** PB HLTH 292, Section 013 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH292.F04/sem.html.
 - **Fall 2004:** BIO ENG 131/231 – *Introduction to Computational Molecular and Cellular Biology* (regularized from BIO ENG 190C).
Primary instructor: Professor Teresa Head-Gordon, Department of Bioengineering.
Professor Dudoit participated in the design of this team-taught course on computational biology and is responsible for the instruction of a two-week module on statistical methods for microarray data analysis.

-
- **Spring 2004:** PB HLTH 143 – *Introduction to Statistical Methods in Computational and Genomic Biology*.
www.stat.berkeley.edu/~sandrine/Teaching/PH143.S04.
 - **Fall 2003:** PB HLTH 296, Section 33 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH296.F03/ph296.F03.html.
 - **Spring 2003:** PB HLTH 296, Section 36 – *Biostatistics Laboratory*.
 - **Spring 2003:** PB HLTH 240D – *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.
www.stat.berkeley.edu/~sandrine/Teaching/PH240D.S03.
 - **Fall 2002:** PB HLTH 296, Section 33 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH296.F02/ph296.F02.html.
 - **Fall 2002:** BIO ENG 190C – *Introduction to Computational Molecular and Cellular Biology* (regularized to BIO ENG 131/231).
Primary instructor: Professor Teresa Head-Gordon, Department of Bioengineering.
Professor Dudoit participated in the design of this team-taught course on computational biology and was responsible for the instruction of a three-week module on statistical methods for microarray data analysis.
 - **Spring 2002:** PB HLTH 296, Section 36 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH296.S02/ph296.S02.html.
 - **Fall 2001:** PB HLTH 296, Section 33 – *Statistics and Genomics Seminar*.
www.stat.berkeley.edu/~sandrine/Teaching/PH296.F01/ph296.F01.html.
- **Invited Guest Lecturer**, Fall 2001 – Present.
University of California, Berkeley.
 - PB HLTH 256 – *Molecular and Genetic Epidemiology*, November 12, 2004.
Primary instructors: Professor Nina T. Holland, Division of Environmental Health Sciences; Professor Lisa Barcellos, Division of Epidemiology.
Lecture: *Statistical methods and software for the analysis of DNA microarray experiments*.
 - INTEGBI 296 – *Computational Genomics Tools for Ecologists and Evolutionary Biologists*, March 9, 2004.
Primary instructor: Dr. Beth Slikas, Department of Integrative Biology.
Lecture: *Statistical methods and software for genomic data analysis*.
 - PB HLTH 256 – *Molecular and Genetic Epidemiology*, November 26, 2002.
Primary instructor: Professor Nina T. Holland, Division of Environmental Health Sciences.
Lecture: *Statistical methods and software for the analysis of DNA microarray experiments*.
 - PB HLTH 293, Section 9 – *Epidemiology Doctoral Seminar*, February 25, March 4, and March 11, 2002.
Primary instructor: Professor Ira B. Tager, Division of Epidemiology.
Lecture 1: *Introduction to Genetics*.
Lecture 2: *Genetic mapping – Allelic association studies*.

Lecture 3: *Genetic mapping – Linkage analysis using identity by descent data.*

Professor Dudoit designed and presented a six-hour lecture series on the statistical analysis of genetic data. The lectures included an introduction to genetics and molecular biology, and a discussion of statistical methods for genetic association and linkage studies.

- MATH 191 – *Applied Mathematics Seminar*, November 30, 2001.

Lecture: *Statistical methods for the design and analysis of DNA microarray experiments.* The audience consisted of about 40 mathematics students, ranging from sophomores to seniors. The lecture included an introduction to genome biology and emphasized mathematical aspects in the analysis of microarray data.

- MCELLBI 299 – *Current Research in Computational Biology and Genomics Seminar*, October 31, 2001.

Lecture: *Statistical methods for the design and analysis of DNA microarray experiments.* The audience consisted primarily of graduate students and postdoctoral researchers with biological backgrounds. The lecture emphasized the importance of sound statistical methods for improving the efficiency and reliability of microarray experiments.

- **Graduate Student Instructor**, August 1994 – May 1999.

Department of Statistics, University of California, Berkeley.

Graduate student instructor for PhD level Applied Statistics course and undergraduate Probability and Statistics courses. Duties included conducting discussion sections, preparing and grading computer labs, grading midterms and final exams, and occasionally lecturing.

- **Lecturer**, Spring 1994.

Department of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada.

Responsible for the instruction of the course *Introduction to Differential and Integral Calculus*. Duties included lecturing, preparing problem sets and exams, grading exams, and supervising teaching assistants.

- **Teaching Assistant**, 1990 – 1993.

Department of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada.

Teaching assistant for Calculus, Algebra, and Statistics courses. Duties included conducting discussion sections, grading, and preparing solutions for tests and problem sets.

ADVISING

MA Students

- **Matthieu Cornec**, Department of Statistics, UC Berkeley/École Nationale de la Statistique et de l'Administration Économique, Malakoff, France (Fall 2002 – Spring 2003).

Project: Risk inference using cross-validation.

Current position: Institut National de la Statistique et des Études Économiques (INSEE), Paris, France.

- **Fabian L. Gallusser**, Graduate Group in Biostatistics, UC Berkeley (Summer 2005 – Summer 2006).
Project: Statistical and computational methods for the analysis of alternative splicing microarray experiments; R package `cosmo` for COnstrained Search for MOTifs in DNA.
Current position: ZS Associates, San Mateo, CA.

PhD Students

- **Raul E. Aguilar Schall**, Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Present).
Project: Statistical methods for the analysis of miRNA microarray experiments; statistical methods for relating pesticide exposure to clinical outcomes (CHAMACOS study).
- **Merrill D. Birkner** (jointly with Mark J. van der Laan), Graduate Group in Biostatistics, UC Berkeley (Spring 2004 – Spring 2006).
Graduated: May 2006.
Dissertation title: *Statistical Hypothesis Testing and Application to Biological Data*.
Project: Multiple hypothesis testing; applications to genetic studies of human obesity (ObeLinks Project).
Current position: Biostatistician, BioOncology, Genentech, Inc.
- **Houston N. Gilbert**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006 – Present).
Project: Multiple testing methods and applications to DNA microarray-based genetic mapping and gene expression studies.
- **Kasper D. Hansen**, Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Present).
Project: Statistical and computational methods for the analysis of high-throughput microarray gene expression experiments: alternative splicing microarray experiments; ChIP-Chip experiments for identifying transcription factor binding sites and histone modification; microarray-based assays (16s small-subunit rRNA microarrays) for the quantitative detection of microorganisms in complex environmental and medical samples.
- **David A. Shilane**, Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Present).
Project: Loss-based estimation with evolutionary algorithms and cross-validation.
- **Siew Leng Teng**, Graduate Group in Biostatistics, UC Berkeley (Fall 2002 – Fall 2004).
Project: Loss-based variable importance measures; applications to the prediction of biological and clinical outcomes using microarray gene expression measures.
- **Yan Wang**, Graduate Group in Biostatistics, UC Berkeley (Fall 2002 – Spring 2006).
Graduated: May 2006.
Dissertation title: *Statistical Methods for Evaluation Linkage Disequilibrium and Its Patterns Using Length of Haplotype Sharing*.
Project: Genetic mapping of complex human traits using single nucleotide polymorphisms (SNP); quantification and visualization of genome-wide linkage disequilibrium patterns; identification of haplotype blocks.
Current position: Biostatistician, ICOS.

Postdoctoral Researchers

- **Dr. Mathias Drton** (jointly with Mark J. van der Laan, Lior Pachter, and Bernd Sturmfels), Department of Mathematics, UC Berkeley (Summer 2004 – Spring 2005).
Project: Multiple testing procedures in graphical model selection.
Current position: Assistant Professor, Department of Statistics, University of Chicago.
- **Dr. Blythe P. Durbin**, Division of Biostatistics, UC Berkeley (Fall 2003 – Summer 2005).
Project: Optimization of neural network architecture using loss-based Deletion/Substitution/Addition (D/S/A) algorithm; loss-based variable selection; applications to protein secondary structure prediction.
Current position: Biostatistician, Early Clinical Development, Oncology, Genentech, Inc.
- **Dr. Sündüz Keleş** (jointly with Mark J. van der Laan), Division of Biostatistics, UC Berkeley (Fall 2003 – Spring 2004).
Project: Loss-based estimation with cross-validation; supervised detection of regulatory motifs in DNA sequences; identification of transcription factor binding sites in ChIP-Chip experiments; tests of association between gene expression measures and Gene Ontology annotation; software for automating access to data from HTML forms.
Current position: Assistant Professor, Departments of Biostatistics and Statistics, University of Wisconsin, Madison.
- **Dr. Katherine S. Pollard**, Division of Biostatistics, UC Berkeley (Summer and Fall 2003).
Project: Multiple hypothesis testing: software implementation in the Bioconductor R package `multtest` and applications to the identification of differentially expressed and co-expressed genes in high-throughput gene expression experiments.
Current position: Assistant Professor, Genome Center and Department of Statistics, UC Davis.

Visiting Scholars and Student Researchers

- **Dr. Alain Barrier**, Visiting Scholar, Praticien Hospitalo-Universitaire, Hopital Tenon and Université Pierre et Marie Curie – Paris 6, France (January 2004 – July 2006).
Project: Development and application of statistical and computational methods for cancer prognosis prediction based on microarray gene expression measures.
- **Dr. Mélanie Courtine**, Visiting Scholar, Laboratoire d’Informatique, Université Pierre et Marie Curie – Paris 6, France (Fall 2003).
Project: Development and application of statistical and computational methods for genetic studies of human obesity (ObeLinks Project).
- **Kasper D. Hansen**, Visiting Student Researcher, Copenhagen University, Denmark (July 2004 – Summer 2005).
Project: Statistical software design for loss-based estimation with cross-validation.
- **Professor Alejandro Sanchez Pla**, Visiting Scholar, Department of Statistics, Faculty of Biology, University of Barcelona, Spain (Summer 2003).
Project: Microarray data analysis, gene annotation.

SERVICE ON DISSERTATION AND EXAMINATION COMMITTEES

PhD Dissertation Committee Service

- **Thomas J. Sharpton**, Department of Plant and Microbial Biology, UC Berkeley.
- **Valerie Wong**, Department of Plant and Microbial Biology, UC Berkeley.
- **Luz B. Gilbert**, Department of Plant and Microbial Biology, UC Berkeley.
- **Benjamin P. Berman**, Department of Molecular and Cell Biology, UC Berkeley.
- **Yun Zhou**, Graduate Group in Biostatistics, UC Berkeley (Fall 2006).
Dissertation title: *Statistical Issues in a Case-Control Study of Gene Expression in Post-mortem Human Brains*.
- **Jing Yi**, Department of Statistics, UC Berkeley (Summer 2006).
Dissertation title: *Absolute and Relative Quantification of Fluorescently Labeled DNA*.
- **Yan Wang**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2006).
Dissertation title: *Statistical Methods for Evaluation Linkage Disequilibrium and Its Patterns Using Length of Haplotype Sharing*.
- **Sandra E. Sinisi**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006).
Dissertation title: *Data-Adaptive Prediction with the Deletion/Substitution/Addition Algorithm*.
- **Merrill D. Birkner**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006).
Dissertation title: *Statistical Hypothesis Testing and Application to Biological Data*.
- **Boonchai Bobby Boonyaratankornkit**, Department of Chemical Engineering, UC Berkeley (Fall 2005).
Dissertation title: *Adaptation of the Deep-Sea Thermophile Methanocaldococcus jannaschii to Extreme Temperatures and Pressures*.
- **Chao Chen**, Department of Statistics, UC Berkeley (Fall 2005).
Dissertation title: *Topics in Random Forests*.
- **Yu Chuan Tai**, Graduate Group in Biostatistics, UC Berkeley (Spring 2005).
Dissertation title: *Multivariate Empirical Bayes Models for Replicated Microarray Time Course Data*.
- **Ingileif B. Hallgrimsdottir**, Department of Statistics, UC Berkeley (Spring 2005).
Dissertation title: *Statistical Methods for Gene Mapping in Complex Diseases*.
- **Biao Xing**, Graduate Group in Biostatistics, UC Berkeley (Fall 2004).
Dissertation title: *Statistical Methods for Detecting Structured Cis-Regulatory Motifs and Constructing Transcriptional Regulatory Networks*.
- **Vivian Wai Ying Ng**, Department of Statistics, UC Berkeley (Fall 2004).
Dissertation title: *Univariate and Bivariate Variable Selection in High Dimensional Data*.
- **François Collin**, Department of Statistics, UC Berkeley (August 2004).
Dissertation title: *Analysis of Oligonucleotide Data with a View to Data Quality Assessment*.

- **Annette M. Molinaro**, Graduate Group in Biostatistics, UC Berkeley (Spring 2004).
Dissertation title: *Novel Approaches to Prediction of Survival in Cancer Research: Focus in Genomics*.
- **Benjamin M. Bolstad**, Graduate Group in Biostatistics, UC Berkeley (Spring 2004).
Dissertation title: *Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization*.
- **Olof Bengtsson**, Licentiate of Philosophy, Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden (April 2001).
Thesis title: *Two-Locus Affected Sib-Pair Identity by Descent Probabilities: Constraints, Parameterisation and Estimation*.

PhD Qualifying Examination Committee Service

- **Peter Dimitrov**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (August 24, 2006).
- **Kasper D. Hansen**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (April 24, 2006).
- **Sandra E. Sinisi**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 8, 2005).
- **Thomas J. Sharpton**, Department of Plant and Microbial Biology, UC Berkeley, Committee Member (October 11, 2005).
- **Merrill D. Birkner**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (May 11, 2005).
- **Srikesh G. Arunajadai**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (April 29, 2005).
- **Anat Caspi**, Department of Bioengineering, UC Berkeley, Committee Member (September 15, 2004).
- **Jing Yi**, Department of Statistics, UC Berkeley, Committee Member (May 7, 2004).
- **Lee Chae**, Department of Plant and Microbial Biology, UC Berkeley, Committee Member (April 30, 2004).
- **Benjamin P. Berman**, Department of Molecular and Cell Biology, UC Berkeley, Committee Member (January 21, 2004).
- **Annette M. Molinaro**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 12, 2003).
- **Yan Wang**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (October 23, 2003).

-
- **Yu Chuan Tai**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (October 8, 2003).
 - **Yun Zhou**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (August 1, 2003).
 - **Yue Wang**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 7, 2003).
 - **Benjamin M. Bolstad**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (May 5, 2003).
 - **Chao Chen**, Department of Statistics, UC Berkeley, Committee Member (March 10, 2003).
 - **Vivian Ng**, Department of Statistics, UC Berkeley, Committee Member (February 12, 2003).
 - **Ingileif B. Hallgrimsdottir**, Department of Statistics, UC Berkeley, Committee Member (October 10, 2002).
 - **Biao Xing**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (August 22, 2002).

MA Comprehensive Examination Committee Service

- **Fabian L. Gallusser**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 3, 2006).
- **James H. Bullard**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 2, 2006).
- **Raul E. Aguilar Schall**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 13, 2005).
- **Peter Dimitrov**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 4, 2005).
- **Kathryn Vaughn Steiger**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 3, 2004).
- **Srikesh G. Arunajadai**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (April 26, 2004).
- **Jingrong (Michelle) Yang**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (May 7, 2002).

UNIVERSITY SERVICE

- *International Association for the Exchange of Students for Technical Experience Local Committee* (Faculty Advisor), UC Berkeley (2006 – Present).
- *Social Epidemiology Faculty Search Committee* (Member), School of Public Health, UC Berkeley, 2006 – 2007.

-
- *Computational Biology Initiative Faculty Search Committee* (Member), UC Berkeley, 2006 – 2007.
 - *QB3 Computing Resources Committee* (Member), UC Berkeley, 2006 – Present.
 - *Computational Biology Initiative Faculty Search Committee* (Member), UC Berkeley, 2005 – 2006.
 - *Curriculum Committee* (Member), School of Public Health, UC Berkeley, 2003 – Present.
 - *Executive Committee* (Member), Graduate Group in Computational and Genomic Biology, UC Berkeley, 2003 – 2005.
 - *Seminars, Publicity, and Retreat Committee* (Chair), Graduate Group in Computational and Genomic Biology, UC Berkeley, 2003 – Present.
 - Co-author with Steven E. Brenner of the *Proposal for a Designated Emphasis in Computational and Genomic Biology*, UC Berkeley, 2003.
 - Co-author of the UC Berkeley proposal for a planning grant for a “National Programs of Excellence in Biomedical Computing” center (pre-NPEBC). Title: *Evolutionary Genomics, Cis-regulation and Genetic Variation*; Principal Investigator: Richard M. Karp; Application number: P20 GM068136; Dates: 01/01/2003 – 12/31/05.

COMPUTING EXPERIENCE

Operating systems: UNIX, Linux, Windows.

Languages: C, R/S-Plus, Matlab, Mathematica, L^AT_EX, HTML.

LANGUAGE SKILLS

Fluent in French and English.

Intermediate Italian and Spanish.