

Curriculum Vitae

Sandrine Dudoit

Sunday 28th July, 2019

Professor and Chair, Department of Statistics
Professor, Division of Epidemiology and Biostatistics, School of Public Health
University of California, Berkeley
367 Evans Hall, #3860
Berkeley, CA 94720-3860
Website: <http://www.stat.berkeley.edu/~sandrine> E-mail: sandrine@stat.berkeley.edu Fax: (510) 642-7892

AFFILIATIONS

Berkeley Institute for Data Science (BIDS; <http://bids.berkeley.edu>), Senior Fellow, 2015 – Present.

Berkeley Stem Cell Center (<http://stemcellcenter.berkeley.edu>), 2009 – Present.

Bioconductor Project (<http://www.bioconductor.org>), Founding Core Developer, 2001 – Present.

California Institute for Quantitative Biosciences (QB3; <http://qb3.berkeley.edu>), 2006 – Present.

Center for Bioinformatics and Molecular Biostatistics (CBMB; <http://www.biostat.ucsf.edu/cbmb>), UC San Francisco, 2003 – Present.

Center for Computational Biology (CCB; <http://ccb.berkeley.edu>), UC Berkeley, Core Faculty, 2003 – Present.

Graduate Group in Biostatistics (<http://www.stat.berkeley.edu/biostat>), UC Berkeley, 2001 – Present.

Graduate Group in Computational and Data Science and Engineering (CDSE; <http://citris-uc.org/initiatives/decse>), UC Berkeley, Core Faculty, 2008 – Present.

Graduate Group in Computational and Genomic Biology (<http://ccb.berkeley.edu/academics/designated-emphasis>), UC Berkeley, Core Faculty, 2003 – Present.

EDUCATION

PhD in Statistics, University of California, Berkeley, May 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, February 1994.

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

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

Advisor: 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 honors), Lycée Molière, Paris, France, 1988.

EMPLOYMENT

Sabbaticals, Fall 2014, Spring 2016.

Professor, July 2010 – Present.

Division of Biostatistics, School of Public Health, and Department of Statistics, University of California, Berkeley.

Associate Professor, November 2006 – June 2010.

Division of Biostatistics, School of Public Health, and Department of Statistics, University of California, Berkeley.

Associate Professor, July 2005 – October 2006.

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 Researcher, July 2000 – June 2001.

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

Statistical methods and software for the analysis of gene expression data from DNA microarray experiments.

Postdoctoral Researcher, September 1999 – June 2000.

Mathematical Sciences Research Institute, Berkeley.

Statistical methods and software for the analysis of gene expression data from DNA microarray experiments.

Visiting Postdoctoral Fellow, June 1999 – August 1999.

Department of Statistics, University of California, Berkeley.

Linkage analysis of complex human traits; analysis of gene expression data from DNA microarray experiments.

Graduate Student Researcher, January 1997 – May 1999.

Department of Statistics, University of California, Berkeley.

Statistical methods for 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.

English to French translation of various SSC publications.

Translator, Canadian Journal of Statistics, 1993 – 1994.

English to French translation of abstracts.

AWARDS AND FELLOWSHIPS

- Elected Member, International Statistical Institute, 2014.
- Thomson Reuters Highly Cited Researcher (<http://highlycited.com/>), 2014.
- Award, Committee on Teaching Excellence, School of Public Health, UC Berkeley, Fall 2013, Spring 2014, Spring 2015, Fall 2015, Fall 2016, Spring 2017 (on sabbatical Fall 2014 and Spring 2016).
- Fellow, American Statistical Association, 2010.
- 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.
- Evelyn Fix Memorial Medal for PhD, Department of Statistics, UC 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.
- Senate Medal for MSc, Carleton University, 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.
- Chacellor’s Medal for BSc, Carleton University, 1992.
- 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.

GRANTS

Ongoing

- Funding source (Project number): National Institutes of Health, BRAIN Initiative Cell Census Program (1U19MH114830-01)
Project title: *A Comprehensive Whole-Brain Atlas of Cell Types in the Mouse*
Principal investigator: John Ngai, Department of Molecular and Cell Biology, UC Berkeley
Project period: 09/20/2017–06/30/2022
- Funding source (Project number): Environmental Protection Agency (RD83615901)
Project title: *Center for Integrative Research on Childhood Leukemia and the Environment*
Principal investigator: Catherine Metayer, Division of Epidemiology, School of Public Health, UC Berkeley
Project period: 09/01/2015–08/31/2019
- Funding source (Project number): National Institutes of Health/National Institute of Environmental Health Sciences (P50ES018172)
Project title: *Center for Integrative Research on Childhood Leukemia and the Environment*
Principal investigator: Catherine Metayer, Division of Epidemiology, School of Public Health, UC Berkeley
Project period: 09/30/2015–06/30/2019

Completed

- Funding source (Project number): National Institutes of Health (1 R33 CA191159-01A1)
Project title: *Using Adductomics to Characterize Exposures to Carcinogens*
Principal investigator: Stephen Rappaport, Division of Environmental Health Sciences, School of Public Health, UC Berkeley
Project period: 09/01/2015–08/31/2018
- Funding source (Project number): National Institutes of Health/BRAIN Initiative (1U01MH105979-01)
Project title: *Classification of Cortical Neurons by Single Cell Transcriptomics*
Principal investigator: John Ngai, Department of Molecular and Cell Biology, UC Berkeley
Project period: 04/01/2015–03/31/2017
- Funding source (Project number): National Institutes of Health/NIGMS (U01 GM061390)
Project title: *Pharmacogenomics of Membrane Transporters*
Principal investigator: Kathleen M. Giacomini, Department of Biopharmaceutical Sciences, UCSF
Project period: 04/01/2011–03/31/2015
- Funding source: France Berkeley Fund
Project title: *Statistical Methods for the Joint Analysis of High-Throughput Transcription and Genotype Data*
Principal investigator: Sandrine Dudoit, Division of Biostatistics, UC Berkeley; Jean-Philippe Vert, Mines ParisTech and Institut Curie
Project period: 11/15/2011–12/15/2014
- Funding source (Project number): National Institutes of Health/NIAID (5 R01 AI077737-03)
Project title: *Analysis of the C. albicans Transcriptome*

Principal investigator: Gavin J. Sherlock, Department of Genetics, Stanford University
Project period: 07/01/2009–06/30/2013

- Funding source (Project number): National Institutes of Health/NIEHS (P42 ES 04705-18)
Project title: *Toxic Substances in the Environment*, Subproject Core D, Biostatistics and Computing
Principal investigator: Martyn T. Smith, Division of Environmental Health Sciences, UC Berkeley
Project period: 04/01/2008–03/31/2009
- Funding source (Project number): National Institutes of Health/NHGRI (U01 HG004271, Intra-University Transaction # 6823304)
Project title: *Comprehensive Characterization of the Drosophila Transcriptome*
Principal investigator: Steven E. Brenner, Department of Plant and Microbial Biology, UC Berkeley
Project period: 05/07/2007–03/31/2011
- Funding source (Project number): Children with Leukaemia Foundation (Reference # 2005/028)
Project title: *Individual Genetic Susceptibility and Environmental Exposures in the Aetiology of Childhood Leukaemia*
Principal investigator: Patricia A. Buffler, Division of Epidemiology, UC Berkeley
Project period: 12/01/2005–11/30/2008
- Funding source (Project number): National Institutes of Health/NIGMS (R01 GM071397)
Project title: *Data Adaptive Estimation in Genomics and Epidemiology*
Principal investigator: Mark J. van der Laan, Division of Biostatistics, UC Berkeley
Project period: 08/01/2004–07/31/2008
- Funding source: France Berkeley Fund
Project title: *OBELINKS: Combining Machine Learning and Biostatistics to Discover Significant Obesity-Related Genetic Polymorphisms*
Principal investigator: Sandrine Dudoit, Division of Biostatistics, UC Berkeley; Jean-Daniel Zucker, Universite Paris 13, France
Project period: 10/01/2003–09/30/2004
- Funding source (Project number): National Institutes of Health/NCHGR (R33 HG002708)
Project title: *A Statistical Computing Framework for Genomic Data*
Principal investigator: Robert C. Gentleman, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center
Project period: 09/30/2003–06/30/2007
- Funding source (Project number): National Science Foundation (41964-7055)
Project title: *CRCNS: Modeling Pathfinding and Target Recognition in the Olfactory System*
Principal investigator: David M. Lin, Department of Biomedical Sciences, Cornell University
Project period: 10/01/2002–09/30/2006
- Funding source (Project number): National Institutes of Health/NLM (R01 LM07609)
Project title: *Statistical Design and Analysis of Microarray Experiments*

Principal investigator: Sandrine Dudoit, Division of Biostatistics, UC Berkeley
Project period: 08/01/2002–07/31/2006

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

PUBLICATIONS

Co-senior-authored publications are marked with a dagger (†).

Books

- [1] R. C. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and **S. Dudoit**, editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health. Springer, New York, 2005.
- [2] **S. Dudoit** and M. J. van der Laan. *Multiple Testing Procedures with Applications to Genomics*. Springer Series in Statistics. Springer, New York, 2008.
- [3] **S. Dudoit**, editor. *Selected Works of Terry Speed*. Springer Selected Works in Probability and Statistics Series. Springer, New York, 2012.

Refereed Journal Publications

- [1] **S. Dudoit** and T. P. Speed. A score test for linkage using identity by descent data from sibships. *Annals of Statistics*, 27(3):943–986, 1999.
- [2] M. J. Callow, **S. Dudoit**, E. L. Gong, T. P. Speed, and E. M. Rubin. Microarray expression profiling identifies genes with altered expression in HDL deficient mice. *Genome Research*, 10(12):2022–2029, 2000.

-
- [3] **S. Dudoit** and T. P. Speed. A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data from sib-pairs. *Biostatistics*, 1(1):1–26, 2000.
- [4] D. R. Goldstein, **S. Dudoit**, and T. P. Speed. Power of a score test for quantitative trait linkage analysis of relative pairs. *Genetic Epidemiology*, 19(Suppl. 1):S85–S91, 2000.
- [5] D. R. Goldstein, **S. Dudoit**, and T. P. Speed. Power and robustness of a score test for linkage analysis of quantitative traits using identity by descent data on sib pairs. *Genetic Epidemiology*, 20(4):415–431, 2001.
- [6] 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. Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc. Natl. Acad. Sci.*, 99(2):972–977, 2002.
- [7] H. Y. Chang, J.-T. Chi, **S. Dudoit**, C. Bondre, M. van de Rijn, D. Botstein D, and P. O. Brown. Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl. Acad. Sci.*, 99(20):12877–12882, 2002.
- [8] 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 *et al.* Gene expression patterns in human liver cancers. *Molecular Biology of the Cell*, 13(6):1929–1939, 2002.
- [9] **S. Dudoit** and J. Fridlyand. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, 3(7):research0036.1–research0036.21, 2002.
- [10] **S. Dudoit**, J. Fridlyand, and T. P. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.
- [11] **S. Dudoit**, Y. H. Yang, M. J. Callow, and T. P. Speed. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, 12(1):111–139, 2002.
- [12] Y. H. Yang, M. J. Buckley, **S. Dudoit**, and T. P. Speed. Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, 11(1):108–136, 2002.
- [13] Y. H. Yang, **S. Dudoit**, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30(4):e15, 2002.
- [14] **S. Dudoit** and J. Fridlyand. Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, 19(9):1090–1099, 2003.
- [15] **S. Dudoit**, R. C. Gentleman, and J. Quackenbush. Open source software for the analysis of microarray data. *BioTechniques*, 34(3):45–51, 2003. Supplement: Microarrays in Cancer: Research and Applications.
- [16] **S. Dudoit**, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.
-

- [17] **S. Dudoit**, M. J. van der Laan, S. Keleş, A. M. Molinaro, S. E. Sinisi, and S. L. Teng. Loss-based estimation with cross-validation: Applications to microarray data analysis. *SIGKDD Explorations*, 5(2):56–68, 2003. Special Issue: Microarray Data Mining.
- [18] Y. Ge, **S. Dudoit**, and T. P. Speed. Resampling-based multiple testing for microarray data analysis. *TEST*, 12(1):1–77, 2003.
- [19] S. Keleş, M. J. van der Laan, **S. Dudoit**, B. Xing, and M. B. Eisen. Supervised detection of regulatory motifs in DNA sequences. *Statistical Applications in Genetics and Molecular Biology*, 2(1):Article 5, 2003.
- [20] **S. Dudoit**, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 13, 2004.
- [21] 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 *et al.* Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80, 2004. (Highly accessed).
- [22] S. Keleş, M. J. van der Laan, and **S. Dudoit**. Asymptotically optimal model selection method with right censored outcomes. *Bernoulli*, 10(6):1011–1037, 2004.
- [23] A. M. Molinaro, **S. Dudoit**, and M. J. van der Laan. Tree-based multivariate regression and density estimation with right-censored data. *Journal of Multivariate Analysis*, 90(1):154–177, 2004. Special Issue on Multivariate Methods in Genomic Data Analysis.
- [24] M. J. van der Laan, **S. Dudoit**, and S. Keleş. Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 4, 2004.
- [25] M. J. van der Laan, **S. Dudoit**, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 14, 2004.
- [26] M. J. van der Laan, **S. Dudoit**, and K. S. Pollard. 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*, 3(1):Article 15, 2004.
- [27] A. Barrier, P.-Y. Boëlle, A. Lemoine, C. Tse, D. Brault, F. Chiappini, F. Lacaine, S. Houry, M. Huguier, A. Flahault, and **S. Dudoit**. Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. *Diseases of the Colon and Rectum*, 48(12):2238–2248, 2005.
- [28] 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**. Colon cancer prognosis prediction by gene expression profiling. *Oncogene*, 24(40):6155–6164, 2005.

- [29] 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. 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*, 19(12):1617–1626, 2005.
- [30] **S. Dudoit** and M. J. van der Laan. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, 2(2):131–154, 2005.
- [31] K. S. Pollard, M. D. Birkner, M. J. van der Laan, and **S. Dudoit**. Test statistics null distributions in multiple testing: Simulation studies and applications to genomics. *Journal de la Société Française de Statistique*, 146(1–2):77–116, 2005. Numéro spécial Statistique et biopuces.
- [32] 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. Exploration of global gene expression in human liver steatosis by high-density oligonucleotide microarray. *Laboratory Investigation*, 86(2):154–165, 2006.
- [33] T. Hothorn, P. Bühlmann, **S. Dudoit**, A. M. Molinaro, and M. J. van der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006.
- [34] S. Keleş, M. J. van der Laan, **S. Dudoit**, and S. E. Cawley. Multiple testing methods for ChIP-Chip high density oligonucleotide array data. *Journal of Computational Biology*, 13(3):579–613, 2006.
- [35] D. Rubin, **S. Dudoit**, and M. J. van der Laan. A method to increase the power of multiple testing procedures through sample splitting. *Statistical Applications in Genetics and Molecular Biology*, 5(1):Article 19, 2006.
- [36] Y. Wang, L. P. Zhao, and **S. Dudoit**. A fine-scale linkage-disequilibrium measure based on length of haplotype sharing. *American Journal of Human Genetics*, 78(4):615–628, 2006.
- [37] 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**. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of Clinical Oncology*, 24(29):4685–4691, 2006.
- [38] M. J. van der Laan, **S. Dudoit**, and A. W. van der Vaart. The cross-validated adaptive epsilon-net estimator. *Statistics & Decisions*, 24(3):373–395, 2006.
- [39] A. W. van der Vaart, **S. Dudoit**, and M. J. van der Laan. Oracle inequalities for multi-fold cross validation. *Statistics & Decisions*, 24(3):351–371, 2006.
- [40] A. Barrier, P.-Y. Boëlle, A. Lemoine, A. Flahault, **S. Dudoit**, and M. Huguier. Gene expression profiling in colon cancer. *Bull. Acad. Natl. Med.*, 191(6):1091–1101, discussion 1102–1103, 2007.
- [41] A. Barrier, F. Roser, P.-Y. Boëlle, B. Franc, C. Tse, D. Brault, F. Lacaine, S. Houry, P. Callard, C. Penna, B. Debuire, A. Flahault, **S. Dudoit**, and A. Lemoine. Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene*, 26(18):2642–2648, 2007.

- [42] **S. Dudoit**, H. N. Gilbert, and M. J. van der Laan. Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: Focus on the false discovery rate and simulation study. *Biometrical Journal*, 50(5):716–744, 2008.
- [43] B. Durbin, **S. Dudoit**, and M. J. van der Laan. A deletion/substitution/addition algorithm for classification neural networks, with applications to biomedical data. *Journal of Statistical Planning and Inference*, 138(2):464–488, 2008. Special Issue: Statistical Design and Analysis in the Health Sciences III.
- [44] A. Lee, K. D. Hansen, J. Bullard, **S. Dudoit**, and G. Sherlock. Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species. *PLoS Genetics*, 4(12):e1000299, 2008.
- [45] D. Shilane, J. Martikainen, **S. Dudoit**, and S. Ovaska. A general framework for statistical performance comparison of evolutionary computation algorithms. *Information Sciences*, 178(14):2870–2879, 2008.
- [46] L. Chae, S. Sudat, **S. Dudoit**, T. Zhu, and S. Luan. Diverse transcriptional programs associated with environmental stress and hormones in the *Arabidopsis* receptor-like kinase gene family. *Molecular Plant*, 2(1):84–107, 2009.
- [47] S. Durinck, J. Bullard, P. T. Spellman, and **S. Dudoit**. GenomeGraphs: integrated genomic data visualization with R. *BMC Bioinformatics*, 10:Article 2, 2009. (Highly accessed).
- [48] K. D. Hansen, L. F. Lareau, M. Blanchette, R. E. Green, Q. Meng, J. Rehwinkel, F. L. Gallusser, E. Izaurralde, D. C. Rio, **S. Dudoit**, and S. E. Brenner. Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila*. *PLoS Genetics*, 5(6):e1000525, 2009.
- [49] E. A. Osborne, **S. Dudoit**, and J. Rine. The establishment of gene silencing at single-cell resolution. *Nature Genetics*, 41(7):800–806, 2009.
- [50] J. H. Bullard, Y. Mostovoy, **S. Dudoit**, and R. B. Brem. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci.*, 107(11):5058–5063, 2010.
- [51] J. H. Bullard, E. A. Purdom, K. D. Hansen, and **S. Dudoit**. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, 11:Article 94, 2010. (Highly accessed).
- [52] K. D. Hansen, S. E. Brenner, and **S. Dudoit**. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, 38(12):e131, 2010.
- [53] **The modENCODE Consortium**. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*, 330(6012):1787–1797, 2010.
- [54] A. N. Brooks, L. Yang, M. O. Duff, K. D. Hansen, J. W. Park, **S. Dudoit**, S. E. Brenner, and B. R. Graveley. Conservation of an RNA regulatory map between *Drosophila* and mammals. *Genome Research*, 21(2):193–202, 2011.

- [55] B. R. Graveley, A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature*, 471(7339):473–479, 2011.
- [56] D. Risso, K. Schwartz, G. Sherlock, and **S. Dudoit**. GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, 12:Article 480, 2011. (Highly accessed).
- [57] L. Jacob, P. Neuvial, and **S. Dudoit**. More power via graph-structured tests for differential expression of gene networks. *Annals of Applied Statistics*, 6(2):561–600, 2012.
- [58] T. Steijger, J. F. Abril, P. G. Engström, F. Kokocinski, **The RGASP Consortium**, T. J. Hubbard, R. Guigó, J. Harrow, and P. Bertone. Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, 10(12):1177–1184, 2013.
- [59] A. Cleynen, **S. Dudoit**, and S. Robin. Comparing segmentation methods for genome annotation based on RNA-Seq data. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):101–118, 2014.
- [60] T. Ferreira, S. R. Wilson, Y. G. Choi, D. Risso, **S. Dudoit**, T. P. Speed, and J. Ngai. Silencing of odorant receptor genes by G protein $\beta\gamma$ signaling ensures the expression of one odorant receptor per olfactory sensory neuron. *Neuron*, 81(4):847–859, 2014.
- [61] D. Risso, J. Ngai, T. P. Speed, and **S. Dudoit**. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, 32(9):896–902, 2014.
- [62] L. Petrick, W. Edmands, C. Schiffman, H. Grigoryan, K. Perttula, Y. Yano, **S. Dudoit**, T. Whitehead, C. Metayer, and S. M. Rappaport. An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies. *Metabolomics*, 13(3):27, March 2017. DOI 10.1007/s11306-016-1153-z.
- [63] C. A. Vallejos, D. Risso, A. Scialdone, **S. Dudoit**, and J. C. Marioni. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nature Methods*, 14(6):1–7, 2017. DOI 10.1038/NMETH.4292.
- [64] R. B. Fletcher, D. Das, L. Gadye, K. N. Street, A. Baudhuin, A. Wagner, M. B. Cole, Q. Flores, Y. G. Choi, N. Yosef, E. Purdom, **S. Dudoit**, D. Risso, and J. Ngai. Deconstructing olfactory stem cell trajectories at single-cell resolution. *Cell Stem Cell*, 20(6):817–830, 2017. DOI 10.1016/j.stem.2017.04.003.
- [65] F. Perradeau, D. Risso, K. Street, E. Purdom, and **S. Dudoit**. Bioconductor workflow for single-cell RNA sequencing: Normalization, dimensionality reduction, clustering, and lineage inference. *F1000Research*, 6:1158, July 2017. DOI 10.12688/f1000research.12122.1.
- [66] L. Gadye, D. Das, M. A. Sanchez, K. Street, A. Baudhuin, A. Wagner, M. B. Cole, Y. G. Choi, N. Yosef, E. Purdom, **S. Dudoit**, D. Risso, J. Ngai, and R. B. Fletcher. Injury activates transient olfactory stem cell states with diverse lineage capacities. *Cell Stem Cell*, 21(6):775–790, 2017. DOI 10.1016/j.stem.2017.10.014.
- [67] D. Risso, F. Perradeau, S. Gribkova, **S. Dudoit**, and J.-P. Vert. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nature Communications*, 9(1):284, 2018. DOI 10.1038/s41467-017-02554-5.

- [68] † K. Van den Berge, F. Perraudeau, C. Sonesson, M. I. Love, D. Risso, J.-P. Vert, M. D. Robinson, **S. Dudoit**, and L. Clement. Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biology*, 19(1):24, 2018. ISSN 1474-760X. DOI 10.1186/s13059-018-1406-4. URL <https://doi.org/10.1186/s13059-018-1406-4>.
- [69] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and **S. Dudoit**. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1):477, 2018. ISSN 1471-2164. DOI 10.1186/s12864-018-4772-0. URL <https://doi.org/10.1186/s12864-018-4772-0>
- [70] D. Risso, L. Purvis, R. Fletcher, D. Das, J. Ngai, **S. Dudoit**, and E. Purdom. clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *PLoS Computational Biology*, 14(9):e1006378, 2018. DOI 10.1371/journal.pcbi.1006378.
- [71] K. Perttula, C. Schiffman, W. M. B. Edmands, L. Petrick, H. Grigoryan, X. Cai, M. J. Gunter, A. Naccarati, S. Polidoro, **S. Dudoit**, P. Vineis, and S. M. Rappaport. Untargeted lipidomic features associated with colorectal cancer in a prospective cohort. *BMC Cancer*, 18(1):996, 2018. DOI 10.1186/s12885-018-4894-4.
- [72] † M. B. Cole, D. Risso, A. Wagner, D. DeTomaso, J. Ngai, E. Purdom, **S. Dudoit**, and N. Yosef. Performance assessment and selection of normalization procedures for single-cell RNA-seq. *Cell Systems*, 8(4):315–328, 2019. DOI 10.1016/j.cels.2019.03.010.
- [73] Y. Yano, H. Grigoryan, C. Schiffman, W. Edmands, L. Petrick, K. Hall, T. Whitehead, C. Metayer, **S. Dudoit**, and S. Rappaport. Untargeted adductomics of Cys34 modifications to human serum albumin in newborn dried blood spots. *Analytical and Bioanalytical Chemistry*, 411(11):2351–2362, 2019. DOI 10.1007/s00216-019-01675-8.
- [74] L. M. Petrick, C. Schiffman, W. M. B. Edmands, Y. Yano, K. Perttula, T. Whitehead, C. Metayer, C. E. Wheelock, M. Arora, H. Grigoryan, H. Carlsson, **S. Dudoit**, and S. M. Rappaport. Metabolomics of neonatal blood spots reveal distinct phenotypes of pediatric acute lymphoblastic leukemia and potential effects of early-life nutrition. *Cancer Letters*, 452:71–78, 2019. DOI 10.1016/j.canlet.2019.03.007.
- [75] C. Schiffman, L. Petrick, K. Perttula, Y. Yano, H. Carlsson, T. Whitehead, C. Metayer, J. Hayes, S. Rappaport, and **S. Dudoit**. Filtering procedures for untargeted LC-MS metabolomics data. *BMC Bioinformatics*, 2019. Accepted.

Refereed Book Chapters and Conference Proceedings

- [1] **S. Dudoit** and T. P. Speed. Triangle constraints for sib-pair identity by descent probabilities under a general multilocus model for disease susceptibility. In M. E. Halloran and S. Geisser, editors, *Statistics in Genetics*, Volume 112 of *The IMA Volumes in Mathematics and its Applications*, pages 181–221. Springer, New York, 1999.

- [2] Y. H. Yang, **S. Dudoit**, P. Luu, and T. P. Speed. Normalization for cDNA microarray data. In M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty, editors, *Microarrays: Optical Technologies and Informatics*, Volume 4266 of *Proceedings of SPIE*, pages 141–152. SPIE—The International Society for Optical Engineering, Bellingham, WA, 2001.
- [3] **S. Dudoit** and D. R. Goldstein. Extensions to a score test for genetic linkage with identity by descent data. In D. R. Goldstein, editor, *Statistics and Science: A Festschrift for Terry Speed*, Volume 40 of *Lecture Notes – Monograph Series*, pages 307–319. Institute of Mathematical Statistics, 2003.
- [4] **S. Dudoit** and M. J. van der Laan. Unified cross-validation methodology for estimator selection and applications to genomics. In *Bulletin of the International Statistical Institute, 54th Session Proceedings*, Volume LX, pages 412–415. 2003.
- [5] **S. Dudoit** and Y. H. Yang. 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, editors, *The Analysis of Gene Expression Data: Methods and Software*, Statistics for Biology and Health, Chapter 3, pages 73–101. Springer, 2003.
- [6] **S. Dudoit** and J. Fridlyand. Classification in microarray experiments. In T. P. Speed, editor, *Statistical Analysis of Gene Expression Microarray Data*, Chapter 3, pages 93–158. Chapman & Hall/CRC, 2003.
- [7] **S. Dudoit** and J. Fridlyand. Introduction to classification in microarray experiments. In D. P. Berrar, W. Dubitzky, and M. Granzow, editors, *A Practical Approach to Microarray Data Analysis*, Chapter 7, pages 132–149. Springer, 2003.
- [8] R. Gentleman, B. Ding, **S. Dudoit**, and J. Ibrahim. Distance measures in DNA microarray data analysis. In R. C. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and **S. Dudoit**, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Statistics for Biology and Health, Chapter 12, pages 189–208. Springer, New York, 2005.
- [9] K. S. Pollard, **S. Dudoit**, and M. J. van der Laan. Multiple testing procedures: The `multtest` package and applications to genomics. In R. C. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and **S. Dudoit**, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Statistics for Biology and Health, Chapter 15, pages 249–271. Springer, New York, 2005.
- [10] **S. Dudoit**, S. Keleş, and M. J. van der Laan. Multiple tests of association with biological annotation metadata. In D. Nolan and T. P. Speed, editors, *Probability and Statistics: Essays in Honor of David A. Freedman*, Volume 2 of *Institute of Mathematical Statistics COLLECTIONS*, pages 153–218. Institute of Mathematical Statistics, 2008.
- [11] A. Cleynen, **S. Dudoit**, E. Lebarbier, and S. Robin. Two segmentation methods for genome annotation. In *Proceedings of the 10th International Workshop on Computational Systems Biology*, pages 56–60. 2013.
- [12] D. Risso, J. Ngai, T. P. Speed, and **S. Dudoit**. The role of spike-in standards in the normalization of RNA-seq. In S. Datta and D. Nettleton, editors, *Statistical Analysis of Next*

Generation Sequencing Data, *Frontiers in Probability and the Statistical Sciences*, Chapter 9, pages 169–190. Springer International Publishing, 2014.

Manuscripts Under Revision or Submitted

- [1] J. L. Hayes, W. M. B. Edmands, Y. Yano, H. Grigoryan, C. Schiffman, **S. Dudoit**, and S. Rappaport. adductomicsR: A package for detection and quantification of protein adducts from mass spectra of tryptic digests. *Bioinformatics*, 2019. Submitted.
- [2] † K. Van den Berge, H. Roux de Bézieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, **S. Dudoit**, and L. Clement. Trajectory-based differential expression analysis. *Nature Communications*, 2019. Submitted.
- [3] H. Grigoryan, C. Schiffman, M. J. Gunter, A. Naccarati, S. Polidoro, **S. Dudoit**, P. Vineis, and S. M. Rappaport. Cys34 adductomics provides evidence linking colorectal cancer with the gut microbiota and redox biology. *Cancer Research*, 2019. Submitted.
- [4] Y. Yano, H. Grigoryan, J. L. Hayes, W. M. B. Edmands, T. P. Whitehead L. Petrick, C. Metayer, **S. Dudoit**, and S. M. Rappaport. Untargeted adductomics of newborn dried blood spots identifies modifications to human serum albumin associated with childhood leukemia. *Cancer Epidemiology, Biomarkers & Prevention*, 2019. Submitted.

Technical Reports and Manuscripts in Preparation

- [1] **S. Dudoit** and J. Fridlyand. Applications of resampling methods to estimate the number of clusters and to improve the accuracy of a clustering method. Technical Report 600, Department of Statistics, University of California, Berkeley, Berkeley, CA 94720-3860, 2001.
- [2] **S. Dudoit**, Y. H. Yang, and B. Bolstad. Using R for the analysis of DNA microarray data. *R News*, 2(1):24–32, March 2002.
- [3] **S. Dudoit**, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. Technical Report 110, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2002.
- [4] S. Keleş, M. J. van der Laan, and **S. Dudoit**. Asymptotically optimal model selection method with right censored outcomes. Technical Report 124, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
- [5] M. J. van der Laan, **S. Dudoit**, and S. Keleş. Asymptotic optimality of likelihood based cross-validation. Technical Report 125, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
- [6] **S. Dudoit** and M. J. van der Laan. Asymptotics of cross-validated risk estimation in model selection and performance assessment. Technical Report 126, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.

-
- [7] **S. Dudoit**. IBD configuration transition matrices and linkage score tests for unilineal relative pairs. Technical Report 128, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [8] M. J. van der Laan and **S. Dudoit**. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples. Technical Report 130, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [9] S. Keleş, M. J. van der Laan, **S. Dudoit**, B. Xing, and M. B. Eisen. Supervised detection of regulatory motifs in DNA sequences. Technical Report 131, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [10] A. M. Molinaro, **S. Dudoit**, and M. J. van der Laan. Tree-based multivariate regression and density estimation with right-censored data. Technical Report 135, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [11] **S. Dudoit**, M. J. van der Laan, S. Keleş, A. M. Molinaro, S. E. Sinisi, and S. L. Teng. Loss-based estimation with cross-validation: Applications to microarray data analysis and motif finding. Technical Report 137, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [12] **S. Dudoit**, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. Technical Report 138, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [13] M. J. van der Laan, **S. Dudoit**, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. Technical Report 139, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2003.
 - [14] M. J. van der Laan, **S. Dudoit**, and K. S. Pollard. Multiple testing. Part III. Procedures for control of the generalized family-wise error rate and proportion of false positives. Technical Report 141, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
 - [15] M. J. van der Laan, **S. Dudoit**, and A. W. van der Vaart. The cross-validated adaptive epsilon-net estimator. Technical Report 142, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
 - [16] S. Keleş, M. J. van der Laan, **S. Dudoit**, and S. E. Cawley. Multiple testing methods for ChIP-Chip high density oligonucleotide array data. Technical Report 147, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
 - [17] Y. Wang and **S. Dudoit**. Quantification and visualization of LD patterns and identification of haplotype blocks. Technical Report 150, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
 - [18] K. S. Pollard, **S. Dudoit**, and M. J. van der Laan. Multiple testing procedures: R `multtest` package and applications to genomics. Technical Report 164, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
-

-
- [19] **S. Dudoit**, M. J. van der Laan, and M. D. Birkner. Multiple testing procedures for controlling tail probability error rates. Technical Report 166, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2004.
- [20] M. D. Birkner, K. S. Pollard, M. J. van der Laan, and **S. Dudoit**. Multiple testing procedures and applications to genomics. Technical Report 168, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [21] B. Durbin, **S. Dudoit**, and M. J. van der Laan. Optimization of the architecture of neural networks using a deletion/substitution/addition algorithm. Technical Report 170, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [22] D. Rubin, **S. Dudoit**, and M. J. van der Laan. A method to increase the power of multiple testing procedures through sample splitting. Technical Report 171, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2006.
- [23] T. Hothorn, P. Bühlmann, **S. Dudoit**, A. M. Molinaro, and M. J. van der Laan. Survival ensembles. Technical Report 174, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [24] A. Barrier and **S. Dudoit**. Colon cancer prognosis prediction by gene expression profiling. Technical Report 178, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [25] A. Barrier and **S. Dudoit**. Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. Technical Report 179, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [26] K. S. Pollard, M. D. Birkner, M. J. van der Laan, and **S. Dudoit**. Test statistics null distributions in multiple testing: Simulation studies and applications to genomics. Technical Report 184, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [27] Y. Wang, L. P. Zhao, and **S. Dudoit**. A fine-scale linkage disequilibrium measure based on length of haplotype sharing. Technical Report 192, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2005.
- [28] **S. Dudoit**, S. Keleş, and M. J. van der Laan. Multiple tests of association with biological annotation metadata. Technical Report 202, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2006.
- [29] D. Shilane, J. Martikainen, **S. Dudoit**, and S. Ovaska. A general framework for statistical performance comparison of evolutionary computation algorithms. Technical Report 204, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2006.
- [30] D. Shilane, R. H. Liang, and **S. Dudoit**. Loss-based estimation with evolutionary algorithms and cross-validation. Technical Report 227, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2007.

- [31] **S. Dudoit**, H. N. Gilbert, and M. J. van der Laan. Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: Focus on the false discovery rate and simulation study. Technical Report 228, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2007.
- [32] H. N. Gilbert, M. J. van der Laan, and **S. Dudoit**. Joint multiple testing procedures for graphical model selection with applications to biological networks. Technical Report 245, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2009.
- [33] J. H. Bullard, E. A. Purdom, K. D. Hansen, and **S. Dudoit**. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. Technical Report 247, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2009.
- [34] H. N. Gilbert, K. S. Pollard, M. J. van der Laan, and **S. Dudoit**. Resampling-based multiple hypothesis testing with applications to genomics: New developments in the R/Bioconductor package multtest. Technical Report 249, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2009.
- [35] L. Jacob, P. Neuvial, and **S. Dudoit**. Gains in power from structured two-sample tests of means on graphs. Technical Report 271, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2010.
- [36] D. Risso, K. Schwartz, G. Sherlock, and **S. Dudoit**. GC-content normalization for RNA-Seq data. Technical Report 291, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7358, 2011.
- [37] D. Risso, F. Perraudeau, S. Gribkova, **S. Dudoit**, and J.-P. Vert. ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data. *bioRxiv*, April 2017. DOI 10.1101/125112. URL <http://biorxiv.org/content/early/2017/04/06/125112>.
- [38] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and **S. Dudoit**. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *bioRxiv*, April 2017. DOI 10.1101/128843. URL <http://biorxiv.org/content/early/2017/04/19/128843>.
- [39] F. Perraudeau, **S. Dudoit**, and J. H. Bullard. Accurate determination of bacterial abundances in human metagenomes using full-length 16S sequencing reads. *bioRxiv*, 2017. DOI 10.1101/228619. URL <https://www.biorxiv.org/content/early/2017/12/04/228619>.
- [40] † M. B. Cole, D. Risso, A. Wagner, D. DeTomaso, J. Ngai, E. Purdom, **S. Dudoit**, and N. Yosef. Performance assessment and selection of normalization procedures for single-cell RNA-Seq. *bioRxiv*, 2017. DOI 10.1101/235382. URL <https://www.biorxiv.org/content/early/2017/12/16/235382>.
- [41] † K. Van den Berge, F. Perraudeau, C. Sonesson, M. I. Love, D. Risso, J.-P. Vert, M. D. Robinson, **S. Dudoit**, and L. Clement. Observation weights to unlock bulk RNA-seq tools for zero inflation and single-cell applications. *bioRxiv*, 2018. DOI 10.1101/250126. URL <https://www.biorxiv.org/content/early/2018/01/18/250126>.

- [42] C. Schiffman, L. Petrick, K. Perttula, Y. Yano, H. Carlsson, T. Whitehead, C. Metayer, J. Hayes, W. M. B. Edmands, S. Rappaport, and **S. Dudoit**. Data-adaptive pipeline for filtering and normalizing metabolomics data. *bioRxiv*, 2018. DOI 10.1101/387365. URL <https://www.biorxiv.org/content/early/2018/08/08/387365>.
- [43] J. L. Hayes, W. M. B. Edmands, Y. Yano, H. Grigoryan, C. Schiffman, **S. Dudoit**, and S. Rappaport. adductomicsR: A package for detection and quantification of protein adducts from mass spectra of tryptic digests. *bioRxiv*, 2018. DOI 10.1101/463331. URL <https://www.biorxiv.org/content/early/2018/11/06/463331>.
- [44] † K. Van den Berge, H. Roux de Bézieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, **S. Dudoit**, and L. Clement. Trajectory-based differential expression analysis for single-cell sequencing data. *bioRxiv*, 2019. DOI 10.1101/623397. URL <https://www.biorxiv.org/content/early/2019/05/02/623397>.
- [45] J. A. Mayfield, A. V. R. Madduri, J. Rine, M. Davies, J. Posey, M. Willby, M. Murray, D. B. Moody, and **S. Dudoit**. Metabolomics with limms revealed biomarkers of antibiotic tolerance during tuberculosis. In Preparation.
- [46] A. N. Brooks, K. D. Hansen, A. Hundal, **S. Dudoit**, M. Meyerson, and S. E. Brenner. JuncBASE: a junction-based analysis of splicing events for RNA-Seq data. In Preparation.

Special Issue Edition

- [1] **S. Dudoit**, R. C. Gentleman, and M. J. van der Laan. Multivariate methods in genomic data analysis. *Journal of Multivariate Analysis*, 90(1), 2004.
- [2] A. Labbe and **S. Dudoit**. Special issue on computational statistical methods for genomics and systems biology. *Statistical Applications in Genetics and Molecular Biology*, 11(2), 2012.

PhD Dissertation

- [1] **S. Dudoit**. *Linkage Analysis of Complex Human Traits Using Identity by Descent Data*. PhD thesis, Department of Statistics, University of California, Berkeley, Berkeley, CA 94720-3860, 1999.

SOFTWARE PACKAGES

N.B. Packages released through the Bioconductor Project are refereed. A * indicates a package developed by group members.

- [1] affxparser*: Affymetrix file parsing SDK. H. Bengtsson, J. Bullard, R. Gentleman, K. D. Hansen, J. Hester, and M. Morgan. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/affxparser.html>).

-
- [2] **biomaRt***: Interface to BioMart databases (e.g. Ensembl, COSMIC, Wormbase and Gramene). S. Durinck and W. Huber. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/biomaRt.html>).
 - [3] **clusterExperiment***: Compare clusterings for single-cell sequencing. E. Purdom, D. Risso, and M. Johnson. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/clusterExperiment.html>).
 - [4] **cosmo**: Supervised detection of conserved motifs in DNA sequences (CONstrained Search for MOTifs). O. Bembom, F. Gallusser, and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/2.4/bioc/html/cosmo.html>).
 - [5] **cosmoGUI**: GUI for constructing constraint sets used by the **cosmo** package. F. Gallusser, O. Bembom, and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/2.4/bioc/html/cosmoGUI.html>).
 - [6] **DEGraph**: Two-sample tests on a graph. L. Jacob, P. Neuvial, and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/DEGraph.html>).
 - [7] **EDASeq**: Exploratory data analysis and normalization for RNA-Seq. D. Risso and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/EDASeq.html>).
 - [8] **GenomeGraphs***: Plotting genomic information from Ensembl. S. Durinck and J. Bullard. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/GenomeGraphs.html>).
 - [9] **Genominator***: Analyze, manage and store genomic data. J. Bullard and K. D. Hansen. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/Genominator.html>).
 - [10] **marray**: Exploratory analysis for two-color spotted microarray data (classes and methods, data input, diagnostic plots, normalization procedures, quality assessment). Y. H. Yang, with contributions from A. Paquet and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/marray.html>).
 - [11] **multtest**: Resampling-based multiple hypothesis testing. K. S. Pollard, H. N. Gilbert, Y. Ge, S. Taylor, and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/multtest.html>).
 - [12] **RUVSeq**: Remove unwanted variation from RNA-Seq data. D. Risso and **S. Dudoit**. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/RUVSeq.html>).
 - [13] **scone***: Single Cell Overview of Normalized Expression data. M. Cole and D. Risso. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/scone.html>).
 - [14] **slingshot***: Tools for ordering single-cell sequencing. K. Street. GitHub (<https://github.com/kstreet13/slingshot>).
-

- [15] **sma**: Statistical microarray analysis. **S. Dudoit**, Y. H. Yang, and B. M. Bolstad, with contributions from N. Thorne, I. Lönnstedt, and J. Mar. CRAN (<http://cran.r-project.org>).
- [16] **zinbwave***: Zero-Inflated Negative Binomial-based Wanted Variation Extraction (ZINB-WaVE). D. Risso, S. Gribkova, and J.-P. Vert. Bioconductor Project (<http://www.bioconductor.org/packages/release/bioc/html/zinbwave>).

PRESENTATIONS

Invited Presentations

- [1] Single Cell Analysis Symposium, Molecular Medicine Tri-Conference 2019, San Francisco, CA, March 14, 2019.
Speaker and Session Chair.
Statistical Methods and Software for the Study of Stem Cell Differentiation Using Single-Cell Transcriptome Sequencing.
- [2] Berkeley Microbiome Initiative Workshop: Microbiome Data Science, Lawrence Berkeley National Laboratory, Berkeley, CA, November 14, 2018.
- [3] Computational Medicine Workshop, McGill University, Montreal, PQ, Canada, June 11, 2018.
Computational Medicine: A Statistical Perspective.
- [4] Biostatistics and Bioinformatics Seminar/Institute for Computational Health Sciences Seminar, University of California, San Francisco, April 11, 2018.
Statistical Methods and Software for Investigating Stem Cell Differentiation Using Single-Cell Transcriptome Sequencing.
- [5] Department of Statistics and Actuarial Science, Simon Fraser University, Vancouver, BC, Canada, March 23, 2018.
Statistical Methods and Software for Investigating Stem Cell Differentiation Using Single-Cell Transcriptome Sequencing.
- [6] Berkeley Distinguished Lectures in Data Science, Berkeley Institute for Data Science (BIDS), University of California, Berkeley, November 28, 2017.
Statistical Methods and Software for Investigating Stem Cell Differentiation Using Single-Cell Transcriptome Sequencing.
- [7] Department of Biostatistics, University of Michigan, Ann Arbor, April 13, 2017.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.
- [8] Department of Statistics, University of Washington, Seattle, November 04, 2016.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.
- [9] Institute for Human Genetics, University of California, San Francisco, September 30, 2016.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.

-
- [10] Keynote Presenter, Intelligent Systems for Molecular Biology (ISMB), Orlando, FL, July 11, 2016.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.
- [11] Invited Speaker, Junior Principal Investigator Workshop (WK01), Intelligent Systems for Molecular Biology (ISMB), Orlando, FL, July 10, 2016.
How to outline an individualised career development plan to become a successful PI.
- [12] Invited Speaker, Computational and Systems Immunology Seminar Series, Stanford University, June 28, 2016.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.
- [13] Invited Speaker, BioC 2016: Where Software and Biology Connect, Stanford University, June 25, 2016.
Using Single-Cell Transcriptome Sequencing to Infer Olfactory Stem Cell Fate Trajectories.
- [14] Distinguished Speaker, Department of Biostatistics, University of Texas M. D. Anderson Cancer Center, April 13, 2016.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [15] Plenary Speaker, 12th German Probability and Statistics Days, Bochum, Germany, March 01–04, 2016.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [16] Data Science Lecture Series, Berkeley Institute for Data Science (BIDS), University of California, Berkeley, December 04, 2015.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [17] Statistical Analysis of Massive Genomic Data, Institut de Biologie Génétique et Bioinformatique (IBGBI), Génomole, Evry, France, November 19–21, 2015.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [18] Center for Bioinformatics, Biostatistics and Integrative Biology (C3BI), Institut Pasteur, Paris, France, November 18, 2015.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [19] Festival of Genomics, San Mateo, CA, November 03–05, 2015.
Invited Panelist, *Panel Discussion: Unleashing the Potential of Next Generation Sequencing Data for Therapeutic Development.*
- [20] Algorithmics, Bioinformatics and Statistics for NGS (ABS4NGS) Data Analysis Workshop, Paris, France, June 22–24, 2015.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [21] Keynote Speaker, Complex Trait Community 14th Annual Meeting, Portland, OR, June 08–11, 2015.
Identification of Novel Cell Types in the Brain Using Single-Cell Transcriptome Sequencing.
- [22] Centre de recherche Biologie du développement, Institut Curie, Paris, France, November 19, 2014.
Normalization, Differential Expression, and Controls in RNA-Seq.
-

-
- [23] Statistics for Systems Biology Group, INRA, Evry, France, November 18, 2014.
Normalization, Differential Expression, and Controls in RNA-Seq.
- [24] Toronto Bioinformatics Users Group (TorBUG), Toronto, ON, Canada, September 24, 2014.
Computationally Reproducible Research and Normalization and Differential Expression in RNA-Seq.
- [25] 2014 Mid-Atlantic Genetic Epidemiology and Statistics (MAGES) Conference, University of Pennsylvania, Philadelphia, PA, May 30, 2014.
Normalization and Differential Expression in RNA-Seq.
- [26] Big Data in Biomedicine Conference, Stanford University, Stanford, CA, May 23, 2014.
Statistics and Big Data.
- [27] Department of Biostatistics Seminar, University of Washington, Seattle, WA, May 15, 2014.
Normalization and Differential Expression in RNA-Seq.
- [28] Computational Biology Semimar, University of Maryland, College Park, MD, April 23, 2014.
Normalization and Differential Expression in RNA-Seq.
- [29] Keynote Address, *Dealing with the Data Deluge*, Center for Genome Research and Biocomputing (CGRB) Spring Conference, Oregon State University, Corvallis, OR, April 07, 2014.
Normalization and Differential Expression in RNA-Seq.
- [30] Future of the Statistical Sciences Workshop, Royal Statistical Society Offices, London, England, November 11-12, 2013.
Invited Participant and Panelist for Breakout Session 2b *Statistics and Science.*
- [31] Penn Bioinformatics Forum, Penn Center for Bioinformatics, The University of Pennsylvania, Philadelphia, PA, October 09, 2013.
Normalization of RNA-Seq Data: Are the ERCC Spike-In Controls Reliable?
- [32] Joint Statistical Meetings, Montreal, Canada, August 03–08, 2013.
Session #574, Invited Papers WNAR, *Statistical Challenges in Cancer Genomics with Next-Generation Sequencing and Microarrays*
Normalization of RNA-Seq Data: Are the ERCC Spike-In Controls Reliable?
- [33] Department of Human Genetics, University of Chicago, Chicago, IL, May 29, 2013.
Normalization and Differential Expression in RNA-Seq.
- [34] Biostatistics Seminar, Stanford University, Stanford, CA, April 18, 2013.
Normalization and Differential Expression in RNA-Seq.
- [35] Biostatistics and Biomathematics Program Seminar, Fred Hutchinson Cancer Research Center, Seattle, WA, April 10, 2013.
Normalization and Differential Expression in RNA-Seq.
- [36] Cancer Epidemiology Seminar, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, February 25, 2013.
Normalization and Differential Expression in RNA-Seq.
-

-
- [37] Cancer Bioinformatics Workshop, Cancéropôle, Île-de-France, October 4–5, 2012.
Normalization and Differential Expression for RNA-Seq.
- [38] Joint Statistical Meetings, San Diego, CA, July 28–August 02, 2012.
Section on Statistical Computing, Topic Contributed Papers, Session 224–*Statistical and Computational Challenges in Metagenomic Analysis of Next-Generation Sequencing Data.*
Normalization and Differential Expression in RNA-Seq.
- [39] Department of Statistics, University of California, Riverside, February 28, 2012.
Normalization and Differential Expression in RNA-Seq.
- [40] San Francisco Bay Area Chapter of the American Statistical Association, University of California, Berkeley, November 17, 2011.
Normalization and Differential Expression in RNA-Seq.
- [41] Department of Statistics, University of California, Davis, November 10, 2011.
Normalization and Differential Expression in RNA-Seq.
- [42] Molecular and Computational Biology Seminar, University of Southern California, October 27, 2011.
Normalization and Differential Expression in RNA-Seq.
- [43] Workshop II: Transcriptomics and Epigenomics, Mathematical and Computational Approaches in High-Throughput Genomics, Institute for Pure and Applied Mathematics, Los Angeles, CA, October 26, 2011.
Normalization and Differential Expression in RNA-Seq.
- [44] Statistical Methods for Very Large Datasets Conference 2011, Baltimore, MD, June 01–03, 2011.
Analysis of an RNA-Seq Experiment in Saccharomyces.
- [45] Centre for Applied Mathematics in Bioscience and Medicine (CAMBAM), McGill University, Montreal, PQ, Canada, April 18, 2011.
Statistical methods and software for high-throughput gene expression experiments using mRNA-Seq.
- [46] Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada, April 15, 2011.
Statistical methods and software for high-throughput gene expression experiments using mRNA-Seq.
- [47] Keynote Speaker, Cambridge Healthtech Institute’s X-Gen Congress and Expo, San Diego, CA, March 14–18, 2011.
Statistical methods and software for high-throughput gene expression experiments using mRNA-Seq.
- [48] CSB 2010–9th Annual International Conference on Computational Systems Bioinformatics, Stanford University, CA, August 16–18, 2010.
Statistical approaches for meeting challenges in the analysis and interpretation of mRNA-Seq.
-

-
- [49] Cambridge Healthtech Institute's Inaugural Beyond Sequencing: Strategies for Success, San Francisco, CA, June 22–23, 2010.
Statistical challenges in mRNA-Seq (Keynote Speaker).
- [50] Statistical and Machine Learning Methods in Computational Biology, Lipari International Summer School on Bioinformatics and Computational Biology, Lipari, Italy, June 12–19, 2010.
Statistical challenges in mRNA-Seq.
- [51] Statistical Modelling: Challenges in Health, Lisbon, Portugal, May 09–12, 2010.
Statistical inference in mRNA-Seq.
- [52] Illumina User Symposium, Napa, CA, April 7–9, 2010.
Statistical approaches for meeting challenges in the analysis and interpretation of mRNA-Seq.
- [53] Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA, March 18, 2010.
Statistical approaches for meeting challenges in the analysis and interpretation of mRNA-Seq.
- [54] Advances in Bioinformatics and Genomics Symposium, Stanford, CA, February 19, 2010.
Statistical challenges in mRNA-Seq (Keynote Speaker).
- [55] Biology and Mathematics in the Bay Area (BaMBA) V, Santa Cruz, CA, November 14, 2009.
Statistical inference in mRNA-Seq (Keynote Speaker).
- [56] Northern California Sequencing User Meeting 2009, Berkeley, CA, August 07, 2009.
Statistical inference in mRNA-Seq.
- [57] Designated Emphasis in Computational and Genomic Biology Student Research Seminar, University of California, Berkeley, April 27, 2009.
Statistical inference in mRNA-Seq.
- [58] Mathematical and Computational Biology Seminar (MATH 290), University of California, Berkeley, April 23, 2009.
Statistical inference in mRNA-Seq.
- [59] Mathematical Genomics Workshop, Mathematical Sciences Research Institute, Berkeley, CA, April 13–15, 2009.
Statistical inference in mRNA-Seq.
- [60] Division of Biostatistics, University of California, San Francisco, March 18, 2009.
Statistical inference in mRNA-Seq.
- [61] International Biometric Society, WNAR Meeting, Davis, CA, June 24, 2008.
Statistical and computational challenges in gene expression studies based on next-generation sequencing.
- [62] CAMDA 2007–The 7th International Conference for the Critical Assessment of Microarray Data Analysis, Centro de Investigación Príncipe Felipe, Valencia, Spain, December 13, 2007.
Statistical challenges in microarray data analysis: Past, present, and future (Keynote Speaker).
-

-
- [63] Frontiers in Genomics Seminar, UNAM Center for Genomic Sciences, Cuernavaca, Mexico, October 15, 2007.
Statistical and computational challenges in genomics.
- [64] MCP 2007—The 5th International Conferences on Multiple Comparison Procedures, Vienna, Austria, July 09–11, 2007.
Multiple testing procedures with applications to genomics.
- [65] QB3—Berkeley Science Lunch, University of California, Berkeley, June 20, 2007.
Statistical and computational challenges in genomics.
- [66] 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.
- [67] Department of Statistics, University of California, Davis, December 01, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- [68] Department of Human Genetics, University of California, Los Angeles, November 21, 2005.
Multiple testing procedures and applications to biomedical and genomic data analysis.
- [69] *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.
- [70] Joint Statistical Meetings, Minneapolis, MN, August 07–11, 2005.
Organizer and Chair, Session 76—*Recent advances in multiple testing.*
- [71] 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.
- [72] 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.*
- [73] Department of Biostatistics, Johns Hopkins University, Baltimore, MD, November 03, 2004.
Multiple testing procedures and applications to genomics.
- [74] Cytokinetics, Inc., South San Francisco, CA, October 27, 2004.
Introduction to R and Bioconductor software.
- [75] *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.
- [76] Statistics in Functional Genomics, Centro Stefano Franscini, Ascona, Switzerland, June 27–July 02, 2004.
Multiple testing procedures and applications to genomics.
-

- [77] The Second Erich L. Lehmann Symposium, Rice University, Houston, TX, May 19–22, 2004.
Multiple testing procedures: Applications to genomics.
- [78] Department of Statistics Colloquium, University of California, Berkeley, March 30, 2004.
Multiple testing procedures: Applications to genomics.
- [79] Department of Mathematics and Statistics, Utah State University, Logan, UT, March 16, 2004.
Multiple testing procedures: Applications to genomics.
- [80] Statistics and Genomics Seminar, University of California, Berkeley, February 19, 2004.
Multiple testing procedures: Applications to genomics.
- [81] Genetics of Complex Disease Workshop, Mathematical Sciences Research Institute, Berkeley, CA, February 11, 2004.
Multiple testing procedures and applications to genomic data analysis.
- [82] Department of Statistics, University of California, Riverside, November 25, 2003.
Multiple hypothesis testing in microarray data analysis.
- [83] ESF Exploratory Workshop, Genomic Approaches to Microarray Data Analysis, Madrid, Spain, October 30, 2003.
Resampling-based multiple testing: Applications to microarray data analysis.
- [84] 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.
- [85] NAS Committee on Emerging Issues and Data on Environmental Contaminants, Seattle, WA, September 16, 2003.
Statistical issues in -omic data analysis.
- [86] 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.*
- [87] 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.*
- [88] 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.
- [89] Statistics and Genomics Seminar, University of California, Berkeley, April 10, 2003.
Unified cross-validation methodology for estimator selection and applications to genomics.

-
- [90] 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.
- [91] 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.
- [92] 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.
- [93] Division of Biostatistics, University of California, San Francisco, January 15, 2003.
The Bioconductor Project: Open-source statistical software for the analysis of microarray data.
- [94] 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.
- [95] Joint Statistical Meetings, New York, NY, August 11–15, 2002.
Classification in DNA microarray experiments.
- [96] 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.
- [97] 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.
- [98] Midwest Biopharmaceutical Statistics Workshop, Muncie, IN, May 20–22, 2002.
Statistical methods for the design and analysis of DNA microarray experiments.
- [99] Bioinformatics and Computational Biology Workshop, Fundación BBVA, Madrid, Spain, April 25–26, 2002.
Statistical methods and software for the analysis of DNA microarray experiments.
- [100] Program in Mathematics and Molecular Biology Meeting VII, Santa Fe, NM, January 05–10, 2002.
Statistics Tutorial.
- [101] Cytokinetics, Inc., South San Francisco, CA, November 27, 2001.
Statistical methods for the design and analysis of DNA microarray experiments.

- [102] 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.
- [103] International Society for Clinical Biostatistics Annual Conference, Stockholm, Sweden, August 2001.
Identifying differentially expressed genes in microarray experiments.
- [104] SSC/WNAR/IMS Meeting, Burnaby, BC, Canada, June 2001.
A score test for the linkage analysis of complex human traits.
- [105] 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.
- [106] Astra Zeneca, Göteborg, Sweden, April 2001.
Identifying differentially expressed genes in microarray experiments.
- [107] Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, April 2001.
Identifying differentially expressed genes in microarray experiments.
- [108] International Biometric Society, ENAR Meeting, Charlotte, NC, March 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
- [109] Department of Statistics, Purdue University, West Lafayette, IN, March 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
- [110] Department of Statistics, University of California, Davis, March 2001.
A score test for the genetic mapping of complex human traits.
- [111] 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.
- [112] Department of Statistics, Stanford University, Stanford, CA, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
- [113] The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, February 2001.
Statistical methods for identifying differentially expressed genes in microarray experiments.
- [114] Department Statistics, University of Toronto, Toronto, ON, Canada, February 2001.
A score test for the genetic mapping of complex human traits.
- [115] Department of Mathematics and Statistics, McGill University, Montreal, PQ, Canada, February 2001.
A score test for the genetic mapping of complex human traits.

-
- [116] 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.
- [117] Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada, February 2001.
A score test for the genetic mapping of complex human traits.
- [118] Department of Biostatistics, Johns Hopkins University, Baltimore, MD, January 2001.
A score test for the genetic mapping of complex human traits.
- [119] Division of Biostatistics, University of California, Berkeley, January 2001.
A score test for the genetic mapping of complex human traits.
- [120] 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.
- [121] 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.
- [122] 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.
- [123] Department of Statistics, University of California, Los Angeles, May 2000.
A linkage score test based on identity by descent data.
- [124] Department of Statistical Science, Southern Methodist University, Dallas, TX, April 2000.
Comparison of discrimination methods for the classification of tumors using gene expression data.
- [125] 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.
- [126] Microarray Algorithms and Statistical Analysis: Methods and Standards, Tahoe City, CA, November 1999.
Statistical methods for the classification of cancer using gene expression data.
- [127] 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.
- [128] Neyman Seminar, Department of Statistics, University of California, Berkeley, May 1999.
A linkage score test based on identity by descent data.
- [129] International Biometric Society, ENAR Meeting, Atlanta, GA, March 1999.
A linkage score test based on identity by descent data.
-

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

Contributed Presentations

- [1] 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.*
- [2] Nature Genetics Microarray Meeting, Scottsdale, AZ, September 1999.
Poster: *Statistical methods for the characterization of tumor classes using cDNA microarray data.*
- [3] Program in Mathematics and Molecular Biology Meeting VI, Santa Fe, NM, January 1999.
Poster: *Sib-pair linkage score test for qualitative and quantitative traits.*
- [4] 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.*
- [5] Program in Mathematics and Molecular Biology Retreat, Warrenton, VA, May 1997.
Oral presentation: *Mapping disease genes using identity by descent data.*
- [6] Program in Mathematics and Molecular Biology Meeting V, Santa Fe, NM, January 1997.
Poster: *Searching for a human obesity gene.*

Abstracts

- [1] O. Bengtsson, T. P. Speed, and **S. Dudoit**. A complete set of constraints and parameterisation of two-locus affected sib-pair IBD probabilities. *American Journal of Human Genetics*, 69(4, Suppl. 1):524–524, 2001.

Organizing and Program Committee Membership

- [1] *useR!– International R User Conference 2016*, Member of Organizing Committee. Stanford University, CA, June 27–30, 2016.
- [2] IMS Program Chair, International Biometric Society WNAR/IMS Meeting, University of California, Los Angeles, Los Angeles, CA, June 16–19, 2013.
- [3] *High-Throughput Genomic Assays*, Invited Session, Chair and Organizer. World Congress in Probability and Statistics, Istanbul, Turkey, July 09–14, 2012.
- [4] *Workshop II: Transcriptomics and Epigenomics*, Mathematical and Computational Approaches in High-Throughput Genomics, Member of Organizing Committee. Institute for Pure and Applied Mathematics, Los Angeles, CA, October 25–28, 2011.

-
- [5] *CAMDA 2011–The 11th International Conference for the Critical Assessment of Massive Data Analysis*, Member of Scientific Committee.
Vienna, Austria, July 15–16, 2011.
<http://camda.bioinfo.cipf.es/camda2011>
- [6] *Statistical Methods for High-Throughput Sequencing Gene Expression Studies*, Invited Session, Organizer and Chair.
International Biometric Society, WNAR/IMS Meeting, San Luis Obispo, CA, June 19–22, 2011.
- [7] *Computational Statistical Methods for Genomics and Systems Biology*, Member of Scientific Committee.
Centre de recherches mathématiques, Université de Montréal, Montréal, PQ, Canada, April 18–22, 2011.
- [8] *Biological Knowledge Discovery and Data Mining (BIOKDD'10)/Expert Systems Applications (DEXA 2010)*, Member of Program Committee.
Bilbao, Spain, August 30–September 3, 2010.
<http://www.dexa.org>
- [9] *DIMACS Workshop on Next Generation Sequencing: Making the most of what you read*, Member of Organizing Committee.
DIMACS Center, CoRE Building, Rutgers University, August 25, 2010.
<http://dimacs.rutgers.edu/Workshops/Sequencing>
- [10] *CSB 2010–9th Annual International Conference on Computational Systems Bioinformatics*, Member of Program Committee.
Stanford University, CA, August 16–18, 2010.
<http://www.csb2010.org>
- [11] *Statistical Genomics in Biomedical Research*, Member of Organizing Committee.
Banff International Research Station, Banff, Canada, July 18–23, 2010.
Workshop funded on a competitive basis; Co-organizers: Jennifer Bryan, Jane Fridlyand, Darlene Goldstein, Sündüz Keleş, and Katherine S. Pollard.
<http://www.birs.ca/events/2010/5-day-workshops/10w5076>
- [12] *RGASP - The RNAseq Genome Annotation Assessment Project*, Member of Steering Committee.
Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom, November 10–11, 2009.
- [13] *CAMDA 2009–The 9th International Conference for the Critical Assessment of Massive Data Analysis*, Member of Scientific Committee.
Northwestern University, Chicago, IL, October 5–6, 2009.
<http://camda2009.bioinformatics.northwestern.edu>
- [14] *Next-Generation High-Throughput Sequencing*, Joint WNAR/IMS Invited Session 2, Chair and Organizer.
International Biometric Society, WNAR/IMS Meeting, Portland, OR, June 16, 2009.
-

-
- [15] *CAMDA 2008–The 8th International Conference for the Critical Assessment of Microarray Data Analysis*, Member of Scientific Committee.
Boku University, Vienna, Austria, December 4–6, 2008.
<http://camda.bioinfo.cipf.es/camda08>
- [16] *Statistical and Computational Challenges in Next-Generation Sequencing*, Member of Organizing Committee.
Mathematical Sciences Research Institute, Berkeley, CA, October 10, 2008.
<http://www.stat.berkeley.edu/~seqmtg>
- [17] *Emerging Statistical Challenges in Genome and Translational Research*, Member of Organizing Committee.
Banff International Research Station, Banff, Canada, June 01–06, 2008.
Workshop funded on a competitive basis; Co-organizers: Jennifer Bryan, Jane Fridlyand, Darlene Goldstein, Sündüz Keleş, and Katherine S. Pollard.
<http://www.birs.ca/events/2008/5-day-workshops/08w5062>
- [18] *Fourth Annual Retreat of the UC Berkeley Graduate Group in Computational and Genomic Biology*, Committee Chair.
Mathematical Sciences Research Institute, Berkeley, CA, April 26, 2008.
- [19] *CAMDA 2007–The 7th International Conference for the Critical Assessment of Microarray Data Analysis*, Member of Scientific Committee.
Centro de Investigación Príncipe Felipe, Valencia, Spain, December 13–14, 2007.
<http://camda.bioinfo.cipf.es/camda07>
- [20] *MCP 2007–The 5th International Conferences on Multiple Comparison Procedures*, Member of Organizing Committee.
Vienna, Austria, July 09–11, 2007.
<http://mcp-conference.org/2007>
- [21] *CIBB 2007–The Fourth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, Member of Scientific Program Committee.
Portofino, Italy, July 07–10, 2007.
<http://cibb07.dsi.unimi.it>
- [22] *Computational and Statistical Methods for Genome-Wide Studies of Biological Systems*, Invited Session Organizer.
Interface 2007–39th Symposium on the Interface: Computing Science and Statistics; Theme: Systems Biology, Philadelphia, PA, May 23–26, 2007.
<http://www.galaxy.gmu.edu/stats/IFNA.html>
- [23] *RECOMB 2007–Eleventh Annual International Conference on Research in Computational Molecular Biology*, Chair of Organizing Committee.
San Francisco Bay Area, April 21–25, 2007.
<http://www.qb3.org/recomb2007>
- [24] *Computational and Statistical Genomics*, Member of Organizing Committee.
Banff International Research Station, Banff, Canada, July 08–13, 2006.
-

- Workshop funded on a competitive basis; Co-organizers: Jennifer Bryan, Sündüz Keleş, Mark J. van der Laan, and Katherine S. Pollard.
<http://www.birs.ca/events/2006/5-day-workshops/06w5076>
- [25] *useR!—The R User Conference 2006*, Member of Program Committee.
Vienna, Austria, June 15–17, 2006.
<http://www.r-project.org/useR-2006>
- [26] *Second Annual Retreat of the UC Berkeley Graduate Group in Computational and Genomic Biology*, Committee Chair.
Bodega Bay Spa and Lodge, Bodega Bay, CA, April 29–30, 2006.
- [27] *First Annual Retreat of the UC Berkeley Graduate Group in Computational and Genomic Biology*, Committee Chair.
Marconi Conference Center, Marshall, CA, April 01–03, 2005.
- [28] *Statistical Science for Genome Biology*, Member of Organizing Committee.
Banff International Research Station, Banff, Canada, August 14–19, 2004.
Workshop funded on a competitive basis; Co-organizers: Jennifer Bryan and Mark J. van der Laan.
<http://www.birs.ca/events/2004/5-day-workshops/04w5519>
- [29] *Challenges in the Statistical Analysis of Genomic Data*, Session Organizer.
American Association for the Advancement of Science Annual Meeting, Denver, CO, February 14, 2003.
Session funded on a competitive basis; Co-organizer: Juliet P. Shaffer.
- [30] *Mathematics and Computational Biology of Genome Analysis*, Member of Organizing Committee.
Mathematical Sciences Research Institute, Berkeley, CA, June 2000.

Short Course Development and Instruction

- [1] *Modern Statistical Methods for Big Data and Big Questions in Genomics*.
Short Course 3, The Fourteenth Asia Pacific Bioinformatics Conference, San Francisco Bay Area, January 10, 2016.
- [2] *Statistical Inference for RNA-Seq*.
Specialist Course, Department of Statistical Sciences, University of Padua, Italy, June 19–22, 2012.
http://www.stat.unipd.it/uploads/File/dottorato/corsi%202012/Dudoit_June2012.pdf
- [3] *Statistical Methods and Software for mRNA-Seq and ChIP-Seq*.
Centro de Investigación Príncipe Felipe, Valencia, Spain, November 08–10, 2010.
- [4] *Statistical Methods and Software for the Analysis of Microarray Experiments*.
Mathematical Biosciences Institute, Ohio State University, Columbus, OH, September 20–24, 2004.
<http://www.stat.berkeley.edu/~sandrine/Docs/Talks/MBI04/mbi.html>

-
- [5] *Tutorial III: Hands on Analysis.*
7th International Meeting of the Microarray Gene Expression Data (MGED) Society, Toronto, Canada, September 08, 2004.
- [6] *CBMB and QB3 Short Course: Analysis of Gene Expression Microarray Data.*
Genentech Hall Auditorium, Mission Bay, University of California, San Francisco, November 15, 2003.
<http://www.biostat.ucsf.edu/cbmb/courses/course.html>
- [7] *Practical Statistical Analysis of DNA Microarray Data.*
KolleKolle, Denmark, October 26–28, 2003.
<http://www.biostat.ku.dk/~pd/bioC-2003>
- [8] *Bioconductor Tutorial.*
6th International Meeting of the Microarray Gene Expression Data (MGED) Society, Aix-en-Provence, France, September 03–05, 2003.
<http://www.bioconductor.org/workshops/MGED6/mged6.html>
- [9] *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.
- [10] *Statistical Methods and Software for the Analysis of DNA Microarray Experiments.*
International Biometric Society, ENAR Meeting, Tampa, FL, March 30, 2003.
<http://www.bioconductor.org/workshops/ENAR03/enar03.html>
- [11] *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.
<http://www.bioconductor.org/workshops/EMB003/embo03.html>
- [12] *Analyzing DNA Microarray Data Using Bioconductor.*
Fred Hutchinson Cancer Research Center, Seattle, WA, December 04–06, 2002.
<http://www.bioconductor.org/workshops/Seattle02/index.html>
- [13] *Analyzing DNA Microarray Data Using Bioconductor.*
Microarray Core Facility, Johns Hopkins Medical Institutions, Baltimore, MD, October 28–29, 2002.
<http://www.bioconductor.org/workshops/Hopkins02/index.html>
- [14] *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.
<http://www.bioconductor.org/workshops/JAX02/index.html>
- [15] *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.
<http://www.bioconductor.org/workshops/Summer02Course/index.html>
-

- [16] *Statistics and Genomics*.
Department of Biostatistics, Harvard School of Public Health, Boston, MA, January 23–25, 2002.
<http://www.bioconductor.org/workshops/ShortCourse012302/index.html>
- [17] *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.
- [18] *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.

TEACHING

Primary Instructor, Fall 2001 – Present.
University of California, Berkeley.

Spring 2019

COMPSCI C100/C200A/STAT C100/C200C (Data 100/200), Section 001, *Principles and Techniques of Data Science*.

Co-taught with Professor John DeNero.

PB HLTH 298, Section 013, *Group Study*.

Instructor of Record: Sandrine Dudoit.

Course Facilitator: Hector Roux de Bézieux.

STAT 278B, Section 00,3 *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/STAT278B.S19>.

Fall 2018

MCELLBI 293S, Section 001, *Foundations of Biostatistical Practice*.

Instructor of Record: Sandrine Dudoit.

Instructor: Partow Imani.

PB HLTH 298, Section 028, *Group Study*.

Instructor of Record: Sandrine Dudoit.

Course Facilitator: Hector Roux de Bézieux.

STAT 278B, Section 004, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/STAT278B.F18>.

PB HLTH C240C/STAT C245C, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*.

Spring 2018

MCELLBI 293S, Section 001, *Foundations of Biostatistical Practice*.

Instructor of Record: Sandrine Dudoit.

Instructor: Kelly Street.

STAT 278B, Section 003, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/STAT278B.S18>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.

Fall 2017

PB HLTH 295, Section 001, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.F17>.

PB HLTH C240D/STAT C245D, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine II*.

Spring 2017

PB HLTH 298, Section 013, *Special Topics in Applied Biostatistical Practice*.

Instructor of Record: Sandrine Dudoit.

Course Facilitators: Suzanne Dufault, Lucia Petito, Courtney Schiffman, and Kelly Street.

PB HLTH 295, Section 001, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.S17>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.

Fall 2016

PB HLTH 298, Section 008, *Special Topics in Biostatistical Theory*.

Instructor of Record: Sandrine Dudoit.

Course Facilitators: Suzanne Dufault, Courtney Schiffman, and Kelly Street.

PB HLTH 295, Section 003, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.F16>.

PB HLTH C240C/STAT C245C, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine I*.

Spring 2016

PB HLTH 298, Section 073, *Special Topics in Applied Biostatistical Practice*.

Instructor of Record: Professor Sandrine Dudoit.

Course Facilitators: Lucia Petito and Kelly Street.

Fall 2015

PB HLTH 295, Section 003, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.F15>.

PB HLTH C240D/STAT C245D, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine II*.

Spring 2015

PB HLTH 295, Section 004, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.S15>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.

Spring 2014

PB HLTH 295, Section 003, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.S14>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.

Fall 2013

PB HLTH 295, Section 001, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH295.F13>.

PB HLTH C240D/STAT C245D, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine II*.

Spring 2013

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S13>.

PB HLTH C240A/STAT C245A, Section 001, *Introduction to Modern Biostatistical Theory and Practice*.

Co-taught with Dr. Marco Carone.

Fall 2012

PB HLTH 292, Section 008, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F12>.

PB HLTH C240C/STAT C245C, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*.

Spring 2012

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S12>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.
Co-taught with Professor Haiyan Huang.

Fall 2011

PB HLTH 292, Section 008, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F11>.

Spring 2011

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S11>.

Fall 2010

PB HLTH 292, Section 008, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F10>.

PB HLTH C240C/STAT C245C, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*.

Spring 2010

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.
<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S10>.

PB HLTH C240F/STAT C245F, Section 001, *Statistical Genomics II*.
Co-taught with Professor Haiyan Huang.

Fall 2009

PB HLTH 292, Section 008, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F09>.

PB HLTH C240E/STAT C245E, Section 001, *Statistical Genomics I*.

Co-taught with Professor Yun Song.

Spring 2009

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S09>.

PB HLTH C240D/STAT C245D and STAT 246, Section 001, *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.

Co-taught with Professor Haiyan Huang.

Fall 2008

PB HLTH 292, Section 013, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F08>.

PB HLTH C240C/STAT C245C, Section 001, *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine*.

Spring 2008

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S08>.

Fall 2007

PB HLTH 292, Section 013, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F07>.

PB HLTH 240A, Section 001, *Biostatistical Methods: Advanced Categorical Data Analysis*
Team-taught course, with primary instructor Professor Nicholas P. Jewell.

Lecture: *Tests of Association in a 2×2 Table*, September 18, 20, and 25, 2007.

Spring 2007

PB HLTH 296, Section 036, *Biostatistics Laboratory*.

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S07>.

PB HLTH 240D, Section 001, *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.

Fall 2006

PB HLTH 296, Section 033, *Biostatistics Laboratory*.

PB HLTH 292, Section 013, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F06/mainPH292F06.html>.

PB HLTH 240C, Section 001, *Biostatistical Methods: Computational Techniques with Applications in Biology and Medicine*.

Spring 2006

PB HLTH 296, Section 036, *Biostatistics Laboratory*.

PB HLTH 292, Section 020, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.S06/sem.html>.

Fall 2005

PB HLTH 296, Section 033, *Biostatistics Laboratory*.

PB HLTH 292, Section 013, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F05/sem.html>.

Spring 2005

PB HLTH 240D, Section 001, *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.

Fall 2004

PB HLTH 292, Section 013, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH292.F04/sem.html>.

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, Section 001, *Introduction to Statistical Methods in Computational and Genomic Biology*.

Fall 2003

PB HLTH 296, Section 033, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH296.F03/ph296.F03.html>.

Spring 2003

PB HLTH 296, Section 036, *Biostatistics Laboratory*.

PB HLTH 240D, Section 001, *Biostatistical Methods: Applications of Statistics to Genetics and Molecular Biology*.

Fall 2002

PB HLTH 296, Section 033, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH296.F02/ph296.F02.html>.

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 036, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH296.S02/ph296.S02.html>.

Fall 2001

PB HLTH 296, Section 033, *Statistics and Genomics Seminar*.

<http://www.stat.berkeley.edu/~sandrine/Teaching/PH296.F01/ph296.F01.html>.

Invited Guest Lecturer, Fall 2001 – Present.

University of California, Berkeley.

CMPBIO 201, *Classics in Computational Biology*, October 29 and 31, 2013.

Primary instructor: Professor Rasmus Nielsen, Department of Integrative Biology and Department of Statistics.

Lecture: *Caveat Emptor: On the Use, Mis-use, and Re-use of Statistics in Computational Biology*.

PB HLTH 258, *Epidemiology of Neoplastic Diseases*, March 20, 2007.

Primary instructor: Professor Patricia A. Buffler, Division of Epidemiology.

Lecture: *Multiple testing procedures with applications to genomics*.

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 assignments 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 undergraduate Calculus, Algebra, and Statistics courses. Duties included conducting discussion sections, grading, and preparing solutions for tests and assignments.

ADVISING

Co-Advised Bachelor Students

- **Sichao (Kevin) He** (co-advised with Mariel Vazquez, Department of Mathematics, SFSU), Department of Statistics, UC Berkeley (Spring 2013).
Senior thesis title: *Modeling Unknotting Mechanism of Type II Topoisomerases in the Cubic Lattice*.

Masters 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: Chief Data Scientist, Cdiscount.
- **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; Bioconductor R package *cosmo* for CONstrained Search for MOtifs in DNA (*cosmo* Project, <http://cosmoweb.berkeley.edu/intro.html>).
Current position: Head of Analytics, Google Hardware, Mountain View, CA.
- **Partow Imani**, Graduate Group in Biostatistics, UC Berkeley (Spring 2018).
Thesis title: *Evaluation of Methods for the Detection of Dynamic Temporal Gene Expression*.
Current position: PhD Student, Graduate Group in Biostatistics, UC Berkeley.
- **Sylvia E. Keuter Sudat**, Graduate Group in Biostatistics, UC Berkeley (Spring 2009).
Thesis title: *An Application of Statistical Clustering Methods to the Investigation of Patterns of Gene Expression within the Arabidopsis Receptor-like Kinase Superfamily*.
Current position:

- **Kenneth Jarrod Millman**, Graduate Group in Biostatistics, UC Berkeley (Spring 2015).
Thesis title: *permute — A Python Package for Permutation Tests and Confidence Sets*.
Current position: PhD Student, Graduate Group in Biostatistics, UC Berkeley.
- **Jennifer L. Poirier**, Graduate Group in Biostatistics, UC Berkeley (Spring 2007).
Thesis title: *A Comparative Analysis of Regression Methods*.
Current position:
- **Steven Pollack**, Graduate Group in Biostatistics, UC Berkeley (Spring 2014).
Thesis title: *Boost: a practical generalization of AdaBoost*.
Current position:
- **Hector Roux de Bézieux**, Graduate Group in Biostatistics, UC Berkeley (Spring 2018 – Fall 2018).
Thesis title: *Temporal patterns of gene expression during embryogenesis*.
Current position: PhD Student, Graduate Group in Biostatistics, UC Berkeley.
- **Kelly N. Street**, Graduate Group in Biostatistics, UC Berkeley (Spring 2014).
Thesis title: *Analysis of Control Sequence Reliability in Single-Cell RNA-Seq Data*.
Current position: Postdoctoral Fellow, Dana-Farber Cancer Institute.

Co-Advised Masters Students

- **Benjamin A. Goldstein** (co-advised with Lisa Barcellos), MPH Degree Program in Epidemiology/Biostatistics, School of Public Health, UC Berkeley (Spring 2007).
Thesis title: *Comparative Analysis of FDR-Controlling Procedures to Identify Genetic Associations in Complex Diseases*.
Current position:

PhD Students

- **James H. Bullard**, Graduate Group in Biostatistics, UC Berkeley (Fall 2006 – Fall 2009).
Graduated: December 2009.
Dissertation title: *Statistical Methods and Software for High-throughput Gene Expression Experiments*.
Project: Statistical and computational methods for the design and analysis of high-throughput gene expression experiments: allele-specific expression and polygenic directional evolution of gene expression in *Saccharomyces* using RNA-Seq; genome annotation and transcriptome analysis in *Saccharomyces cerevisiae* using tiling microarrays and RNA-Seq; microarray-based assays (16s small-subunit rRNA microarrays) for the quantitative detection of microorganisms in complex environmental and medical samples; software development for the analysis and visualization of genomic data.
Current position: Co-Founder and CTO, Whole Biome, Inc.
- **Houston N. Gilbert**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006 – Spring 2009).
Graduated: May 2009.
Dissertation title: *Multiple Hypothesis Testing: Methodology, Software Implementation, and Applications to Genomics*.

Project: Multiple hypothesis testing; graphical models; DNA microarray-based genetic mapping and gene expression studies; genome-wide association studies.

Current position: Senior Director of Biostatistics, Bellicum Pharmaceuticals.

- **Kasper D. Hansen**, Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Fall 2009).
Graduated: August 2009.
Dissertation title: *Analyses of High-Throughput Gene Expression Data*.
Project: Statistical and computational methods for the design and analysis of high-throughput gene expression experiments based on DNA microarrays and next-generation sequencing; transcriptome analysis using RNA-Seq; genome annotation in *S. cerevisiae* using tiling microarrays; microarray-based study of gene regulation by alternative splicing in *D. melanogaster*; microarray-based assays (16s small-subunit rRNA microarrays) for the quantitative detection of microorganisms in complex environmental and medical samples; ChIP-Chip experiments for identifying transcription factor binding sites and histone modification.
Current position: Associate Professor, Department of Biostatistics, Johns Hopkins University.
- **Fanny Perraudieu**, Graduate Group in Biostatistics, UC Berkeley (Fall 2014 – Spring 2018).
Graduated: Spring 2018.
Dissertation title: *Statistical and Computational Methods for Single-Cell Transcriptome Sequencing and Metagenomics*.
Current position: Senior Data Scientist, Whole Biome.
- **David A. Shilane**, Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Summer 2008).
Graduated: August 2008.
Dissertation title: *Optimization Studies with Multiple Testing, Loss-Based Estimation, and Confidence Intervals for Negative Binomials of High Dispersion*.
Project: Loss-based estimation with evolutionary algorithms and cross-validation.
Current position: Consultant, End-to-End Analytics.
- **Kelly N. Street**, Graduate Group in Biostatistics, UC Berkeley (Fall 2014 – Fall 2018).
Graduated: Fall 2018.
Dissertation title: *Trajectory Inference and Analysis in Single-Cell Genomics*.
Current position: Postdoctoral Fellow, Dana-Farber Cancer Institute.
- **Yan Wang**, Graduate Group in Biostatistics, UC Berkeley (Fall 2002 – Spring 2006).
Graduated: May 2006.
Dissertation title: *Statistical Methods for Evaluating 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: Genentech.

Co-Advised PhD Students

- **Raúl E. Aguilar Schall** (co-advised with Alan E. Hubbard), Graduate Group in Biostatistics, UC Berkeley (Fall 2005 – Fall 2010).

Graduated: September 2010.

Dissertation title: *Semi-Parametric Graphical Computation Approach Using Loss-Based Estimation to Estimate Exposure Effects: Applications on Infant Developmental Outcomes.*

Project: Statistical methods for the analysis of miRNA microarray experiments; statistical methods for relating pesticide exposure to clinical outcomes (CHAMACOS study).

Current position: Manager Biostatistics, Gilead Sciences.

- **Merrill D. Birkner** (co-advised 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, <http://www.obelinks.org>).
Current position: Genentech.
- **Alice Cleynen** (co-advised with Stéphane Robin), Mathematics, Université Paris-Sud 11, France (January 2011 – November 2013).
PhD defense: November 15, 2013.
Dissertation title: *Statistical Approaches for Segmentation : Application to Genome Annotation.*
Current position: Chargée de recherche CNRS, Institut Montpellierain Alexander Grothendieck.
- **Nicholas Everetts** (co-advised with Iswar Hariharan and Nir Yosef), PhD Program in Computational Biology, UC Berkeley (Summer 2017 – Present).
Project: Study of growth regulation in *Drosophila* imaginal discs using RNA-Seq.
- **Miles E. Lopes** (co-advised with Peter J. Bickel), Department of Statistics, UC Berkeley (Summer 2010 – Summer 2012).
Project: Structured tests in high dimensions; sparsity in compressed sensing; tests for biological annotation metadata; inference of time-varying networks; *Drosophila* developmental gene expression based on RNA-Seq.
Current position: Assistant Professor, Department of Statistics, UC Davis.
- **Luca Pozzi** (co-advised with Nicholas P. Jewell and Alan E. Hubbard), Graduate Group in Biostatistics, UC Berkeley (Spring 2010 – Fall 2010).
Project: Statistical and computational methods for the study of alternative splicing using high-throughput sequencing.
Current position: Data Scientist, Airbnb.
- **Davide Risso** (co-advised with Monica Chiogna and Chiara Romualdi), Department of Statistical Sciences, University of Padua, Italy (August 2010 – February 2012).
Graduated: February 2012.
Dissertation title: *Simultaneous Inference for RNA-Seq.*
Current position: Assistant Professor, Rita Levi Montalcini Fellow, Department of Statistical Sciences, University of Padova, Italy
Adjunct Assistant Professor, Healthcare Policy and Research, Division of Biostatistics and Epidemiology, Department of Healthcare Policy and Research, Weill Cornell Medicine.

- **Courtney Schiffman** (co-advised with Haiyan Huang), Graduate Group in Biostatistics, UC Berkeley (Fall 2016 – Spring 2019).
Dissertation title: *The Role of Exploratory Data Analysis and Pre-processing in Omics Studies*.
Project: Adductomics and metabolomics for the untargeted characterization of exposures (based on nano liquid chromatography high-resolution mass spectrometry, nano-LC-HRMS) and for use in exposome-wide association studies (EWAS) to discover exposures that contribute to disease risks.
Current position: Genentech.
- **Siew Leng Teng** (co-advised with Haiyan Huang), Graduate Group in Biostatistics, UC Berkeley (Fall 2002 – Fall 2004).
Graduated: Fall 2007.
Dissertation title: *Statistical Methods in Integrative Analysis of Gene Expression Data with Applications to Biological Pathways*.
Project: Loss-based variable importance measures; applications to the prediction of biological and clinical outcomes using microarray gene expression measures.
Current position: Genentech.

Rotation PhD Students

- **Tal Ashuach**, PhD Program in Computational Biology, UC Berkeley (January 09, 2017 – March 10, 2017).
Project: *The Effect of Culturing Platform on Stem Cell Biology*.
- **Calvin Chi**, PhD Program in Computational Biology, UC Berkeley (November 16, 2015 – February 05, 2016).
Project: Cluster analysis for single-cell RNA-Seq.

Postdoctoral Researchers

- **Dr. Abha Bais**, Division of Biostatistics, UC Berkeley (Spring 2011 – Spring 2013).
Project: Genome annotation and transcriptome analysis in *Candida* using RNA-Seq.
Current position: Staff Research Scientist, Department of Developmental Biology, University of Pittsburgh.
- **Dr. Gergana Bounova** (co-advised with Terence P. Speed), Department of Statistics, UC Berkeley (Summer 2010 – Fall 2011).
Project: Genome surveillance by small RNAs in *Tetrahymena*; inference of time-varying networks in genomics.
Current position: Computational Scientist, Cancer Genomics, The Netherlands Cancer Institute.
- **Professor Mathias Drton** (co-advised 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: Professor, Department of Statistics, University of Washington, Seattle.

- **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 (DSA) algorithm.
Current position: Principal Statistician, UC Davis.
- **Dr. Steffen Durinck** (co-advised with Paul T. Spellman), Division of Biostatistics, UC Berkeley, and Life Sciences Division, Lawrence Berkeley National Laboratory (Fall 2007 – Spring 2008).
Project: Statistical and computational methods for transcriptome analysis using next-generation sequencing.
Current position: Scientist (Bioinformatics), Genentech.
- **Dr. Josie Hayes** (co-advised with Stephen M. Rappaport), Division of Biostatistics, UC Berkeley (Fall 2017 – Spring 2019).
Project: Adductomics and metabolomics for the untargeted characterization of exposures (based on nano liquid chromatography high-resolution mass spectrometry, nano-LC-HRMS) and for use in exposome-wide association studies (EWAS) to discover exposures that contribute to disease risks.
Current position: Data Scientist, REVOLUTION Medicines.
- **Professor Laurent Jacob** (co-advised with Terence P. Speed), Genentech Innovation Fellow, Center for Computational Biology and Department of Statistics, UC Berkeley (Spring 2010 – Fall 2013).
Project: Statistical and computational methods for transcriptome analysis using high-throughput microarray and sequencing; structured tests in high dimensions.
Current position: Chargé de recherche CNRS, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon I.
- **Professor Sündüz Keleş** (co-advised 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: Professor, Departments of Biostatistics and Statistics, University of Wisconsin, Madison.
- **Professor Katherine S. Pollard**, Division of Biostatistics, UC Berkeley (Summer 2003 – 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: Professor, Division of Biostatistics, UC San Francisco, and Associate Investigator, Gladstone Institutes.
- **Professor Elizabeth A. Purdom**, Division of Biostatistics and Department of Statistics, UC Berkeley (Spring 2007 – Spring 2009).
Project: Statistical and computational methods for transcriptome analysis using next-generation

sequencing.

Current position: Associate Professor, Department of Statistics, UC Berkeley.

- **Professor Davide Risso**, Division of Biostatistics and Department of Statistics, UC Berkeley (Spring 2012 – December 2016).

Project: Statistical methods and software for the analysis of bulk and single-cell RNA-Seq data.

Current position: Assistant Professor, Rita Levi Montalcini Fellow, Department of Statistical Sciences, University of Padova, Italy

Adjunct Assistant Professor, Healthcare Policy and Research, Division of Biostatistics and Epidemiology, Department of Healthcare Policy and Research, Weill Cornell Medicine.

Visiting Students

- **Alice Cleynen**, Visiting PhD Student Researcher, Ecole Normale Supérieure Cachan/AgroParisTech, France (January 2011 – July 2011, February 2012 – July 2012).

Project: Genome annotation and transcriptome analysis in *Saccharomyces* using RNA-Seq.

- **Matthieu Doutreligne**, Visiting Student Researcher, Ecole Polytechnique, France (April 04, 2017 – August 10, 2017).

Project: Identification of differentially expressed genes along cell lineages based on single-cell RNA-Seq data.

- **Nathan Grinsztajn**, Visiting Student Researcher, Ecole Polytechnique, France (April 02, 2018 – August 15, 2018).

Project: Identification of differentially expressed genes along cell lineages based on single-cell RNA-Seq data.

- **Kasper D. Hansen**, Visiting Student Researcher, Department of Biostatistics, Copenhagen University, Denmark (July 2004 – August 2005).

Project: Statistical software design for loss-based estimation with cross-validation.

- **Anne-Claire Haury**, Visiting PhD Student Researcher, Center for Computational Biology, Mines ParisTech/Institut Curie, France (January 2011 – July 2011).

Project: Molecular characterization of cancer using joint analysis of high-throughput transcription and genotype data.

- **Lorenzo Maragoni**, Visiting PhD Student Researcher, Department of Statistical Sciences, University of Padua, Italy (April – June 2013).

Project: Joint analysis of ChIP-Seq and RNA-Seq data.

- **Edouard Pauwels**, Visiting PhD Student Researcher, Center for Computational Biology, Mines ParisTech, France (February 2012 – July 2012).

Project: Molecular characterization of cancer using joint analysis of high-throughput transcription and genotype data.

- **Davide Risso**, Visiting PhD Student Researcher, Department of Statistical Sciences, University of Padua, Italy (August 2010 – August 2011).

Project: Genome annotation and transcriptome analysis in *Saccharomyces* using RNA-Seq.

- **Geneviève Robin**, Visiting Student Researcher, Ecole Polytechnique, France (March 23, 2015 – July 31, 2015).
Project: Dimensionality Reduction in Clustering Single-Cell RNA-Sequencing Data.
- **Koen van den Berge**, Visiting Student Researcher, Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium (April – June 2018).
Project: Identification of differentially expressed genes along cell lineages based on single-cell RNA-Seq data.

Visiting Scholars

- **Dr. Alain Barrier**, 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. Ana Conesa**, Head, Genomics of Gene Expression Lab, Centro de Investigación Príncipe Felipe, Valencia, Spain (Summer 2012).
- **Dr. Mélanie Courtine**, 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, <http://www.obelinks.org>).
- **Professor Niels Richard Hansen**, Department of Mathematical Sciences, University of Copenhagen, Denmark (October 2010 – June 2011).
- **Professor Hongkai Ji**, Department of Biostatistics, Johns Hopkins University (April 2011 – June 2011).
- **Dr. David O. Nelson**, Lawrence Livermore National Laboratory (January 2003 – July 2006).
Project: Microarray data analysis.
- **Professor Alejandro Sanchez Pla**, Department of Statistics, Faculty of Biology, University of Barcelona, Spain (May 2003 – September 2003).
Project: Microarray data analysis; gene annotation.
- **Professor Jean-Philippe Vert**, Mines ParisTech and Institut Curie, Paris, France (April 2015 – August 2016).

SERVICE ON DISSERTATION/THESES AND EXAMINATION COMMITTEES

Masters Thesis Committee Service

- [1] **Maxwell R Murphy**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2019).
Thesis title: *Simultaneous Estimation of Multiplicity of Infection and Allele Frequencies from Complex Malaria Infection Genetic Data*.

-
- [2] **Fausto Bustos**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2019).
Thesis title: *Case Definitions and Clinical Manifestations of Zika in a Large Pediatric Cohort in Managua, Nicaragua.*
- [3] **Partow Imani**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2018).
Thesis title: *Evaluation of Methods for the Detection of Dynamic Temporal Gene Expression.*
- [4] **Rachael Phillips**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2018).
Thesis title: *Data Adaptive Evaluation of Preprocessing Methods using Ensemble Machine Learning.*
- [5] **Suzanne M. Dufault**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2017).
Thesis title: *LASSO for Public Health Data: An Examination of Prevalent Variable Selection Methods and Demonstration of LASSO in R.*
- [6] **Courtney Schiffman**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2016).
Thesis title : *Single Cell RNA-Seq: A Study on Normalization and Sub-Population Identification Techniques.*
- [7] **Kenneth Jarrod Millman**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2015).
Thesis title: *permute — A Python Package for Permutation Tests and Confidence Sets.*
- [8] **Steven Pollack**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2014).
Thesis title: *Boost: a practical generalization of AdaBoost.*
- [9] **Kelly N. Street**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2014).
Thesis title: *Analysis of Control Sequence Reliability in Single-Cell RNA-Seq Data.*
- [10] **Morgan Campbell Levy**, Department of Statistics, UC Berkeley, Committee Member (Fall 2013).
Thesis title: *Analysis of Environmental Model Sensitivity to Correlated Input Variables.*
- [11] **Erin K. Hartman**, Department of Statistics, UC Berkeley, Committee Member (Spring 2013).
Thesis title: *From SATE to PATT: Combining Experimental with Observational Studies to Estimate Population Treatment Effects.*
- [12] **Xiaorong Shao**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Summer 2012).
Thesit title: *Identifying Genetic Predictors of Disease Severity in Multiple Sclerosis using Targeted Maximum Likelihood Estimation Methodology in a Genome-wide Association Study.*

-
- [13] **Malinka J. Jansson**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2011).
Thesis title: *Searching for Genetic Variants in a Case-Control Study of Acute Kidney Injury*.
- [14] **Inna Gerlovina**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Spring 2010).
Thesis title: *Bioinformatic Approaches to Parallel Deletion Analysis in Yeast: An example Studying Environmental Exposure*.
- [15] **Maria Carolina Nizarala Martinez Caetano**, Department of Statistics, UC Berkeley, Committee Member (Spring 2010).
Thesis title: *Discontinuity Test of Endogeneity*.
- [16] **Ruxi Zhang**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (Fall 2009).
Thesis title: *Evaluation of Chemotherapy effect on Ovarian Cancer Patients Survival Times via Hazard Regression*.
- [17] **Sylvia E. Keuter Sudat**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2009).
Thesis title: *An Application of Statistical Clustering Methods to the Investigation of Patterns of Gene Expression within the Arabidopsis Receptor-like Kinase Superfamily*.
- [18] **Jennifer L. Poirier**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2007).
Thesis title: *A Comparative Analysis of Regression Methods*.
- [19] **Benjamin A. Goldstein**, MPH Degree Program in Epidemiology/Biostatistics, School of Public Health, UC Berkeley, Co-Advisor (Spring 2007).
Thesis title: *Comparative Analysis of FDR-Controlling Procedures to Identify Genetic Associations in Complex Diseases*.

Masters Comprehensive Examination Committee Service

- [1] **Vanessa M. Viggiano**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 09, 2014).
- [2] **Emily Chang**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 13, 2013).
- [3] **Morgan Ann Thompson**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 01, 2012).
- [4] **Carolyn Cotterman**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (August 11, 2011).
- [5] **Elizabeth Secor**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (May 06, 2011).
- [6] **Virginia Chen**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 05, 2009).

- [7] **Karen L. Waitman**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (April 30, 2007).
- [8] **Fabian L. Gallusser**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 3, 2006).
- [9] **James H. Bullard**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 2, 2006).
- [10] **Raúl E. Aguilar Schall**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 13, 2005).
- [11] **Peter Dimitrov**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 4, 2005).
- [12] **Kathryn Vaughn Steiger**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 3, 2004).
- [13] **Srikesh G. Arunajadai**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (April 26, 2004).
- [14] **Jingrong (Michelle) Yang**, Graduate Group in Biostatistics, UC Berkeley, Committee Member (May 7, 2002).

PhD Dissertation Committee Service

- [1] **Tal Ashuach**, Graduate Group in Computational Biology, UC Berkeley.
- [2] **Chenling Antelope Xu**, Graduate Group in Computational Biology, UC Berkeley.
- [3] **Calvin Chi**, Graduate Group in Computational Biology, UC Berkeley.
- [4] **Carlos Fernando Buen Abad Najar**, Graduate Group in Computational Biology, UC Berkeley.
- [5] **Courtney Schiffman**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2019).
Dissertation title: *The Role of Exploratory Data Analysis and Pre-processing in Omics Studies*.
- [6] **Kelly N. Street**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Fall 2018).
Dissertation title: *Trajectory Inference and Analysis in Single-Cell Genomics*.
- [7] **Diya Das**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (Spring 2018).
Dissertation title: *Identification of Olfactory Stem Cell Lineage Trajectories and Cell State Transitions*.

-
- [8] **Fanny Perraudou**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2018).
Dissertation title: *Statistical and Computational Methods for Single-Cell Transcriptome Sequencing and Metagenomics*.
- [9] **Brian Perea**, Department of Chemical and Biomolecular Engineering, UC Berkeley, Outside Member (Spring 2017).
Dissertation title: *An Imaging-Based Human Embryonic Stem Cell Assay for Teratogenic Activity*.
- [10] **Courtney E. French**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (Spring 2016).
Dissertation title: *Prevalence and Significance of Nonsense-Mediated mRNA Decay Coupled with Alternative Splicing in Diverse Eukaryotic Organisms*.
- [11] **Curt Hansen**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Fall 2015).
Dissertation title: *The LITSE Algorithm: Theory and Application*.
- [12] **Matahi Moarii**, École nationale supérieure des mines de Paris, Spécialité “Bio-informatique”, Examineur and Président du Jury (Summer 2015).
Dissertation title: *Learning from Multiple Genomic Information in Cancer for Diagnosis and Prognosis*.
- [13] **Sofia Medina-Ruiz**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (Spring 2015).
Dissertation title: *Transcriptional Regulation of the Vertebrate Neural Crest*.
- [14] **Sheila Adams-Sapper**, Graduate Group in Infectious Diseases and Immunity, UC Berkeley, Outside Member (Spring 2015).
Dissertation title: *Characteristics and Coordinated Mechanisms of Carbapenem Heteroresistance in KPC-Producing Enterobacteriaceae*.
- [15] **Marcus H. Stoiber**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Spring 2015).
Dissertation title: *Biological Networks: Dynamics, Mechanisms and Responses*.
- [16] **Samuel D. Lendle**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Spring 2015).
Dissertation title: *Targeted Minimum Loss Based Estimation: Applications and Extensions in Causal Inference and Big Data*.
- [17] **Ma’ayan Bresler**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Inside Member (Spring 2014).
Dissertation title: *Statistical Methods for Genome Assembly*.
- [18] **Erin K. Hartman**, Department of Political Science, UC Berkeley, Outside Member (Fall 2013).
Dissertation title: *Validating Causal Estimates in Experimental and Observational Designs*.
-

-
- [19] **Edouard Pauwels**, École nationale supérieure des mines de Paris, Examineur and Président du Jury (Fall 2013).
Dissertation title: *Applications of Machine Learning in Computational Biology*.
- [20] **Kathryn L. (Benton) Colborn**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Fall 2013).
Dissertation title: *Statistical Models for Longitudinal Analysis of Single and Mixed Species Infections*.
- [21] **Stephan J. Ritter**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Fall 2013).
Dissertation title: *Software for Prediction and Estimation with Applications to High-Dimensional Genomic and Epidemiologic Data*.
- [22] **Emily A. Whiston**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (Spring 2013).
Dissertation title: *Comparative Genomics and Transcriptomics in the Mammalian Fungal Pathogen *Coccidioides* spp.*
- [23] **Jingyi (Jessica) Li**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Spring 2013).
Dissertation title: *Statistical and Computational Methods for Analyzing High-Throughput Genomic Data*.
- [24] **Dapo Omidiran**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (Spring 2013).
Dissertation title: *Low-Dimensional Models for PCA and Regression*.
- [25] **Kevin S. McLoughlin**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Spring 2013).
Dissertation title: *Modeling and Analysis of Oligonucleotide Microarray Data for Pathogen Detection*.
- [26] **Matthew D. Davis**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (Fall 2012).
Dissertation title: *Computational and Experimental Investigations of the Principles of Eukaryotic Transcriptional Regulation Before, During, and After Open Complex Formation*.
- [27] **Garmay E. Leung**, Department of Bioengineering, UC Berkeley, Outside Member (Fall 2012).
Dissertation title: *Ab Initio Discovery of Regulatory Sequences and Characterization of Their Evolutionary Patterns*.
- [28] **Johann Gagnon-Bartsch**, Department of Statistics, UC Berkeley, Inside Member (Fall 2012).
Dissertation title: *Removing Unwanted Variation from Microarray Data with Negative Controls*.

-
- [29] **Sylvia Keuter Sudat**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Summer 2012).
Dissertation title: *Causal Inference and Prediction in Health Studies: Environmental Exposures and Schistosomiasis, HIV-1 Genotypic Susceptibility Scores and Virologic Suppression, and Risk of Hospital Readmission for Heart Failure Patients.*
- [30] **Angela N. Brooks**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (Spring 2011).
Dissertation title: *RNA Splicing Regulation in Drosophila melanogaster.*
- [31] **Wei-Chun Kao**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (Spring 2011).
Dissertation title: *Algorithms for Next-Generation High-Throughput Sequencing Technologies.*
- [32] **Valerie L. Wong**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (Spring 2011).
Dissertation title: *Investigations into the Mycoheterotrophic Relationship between Rhizopogon salebrosus and Pterospora andromedeae and Development of Bioinformatic Tools Related to Non-Assembled Fungal Genomes.*
- [33] **Oleg S. Mayba**, Department of Statistics, UC Berkeley, Inside Member (Spring 2011).
Dissertation title: *Statistical Aspects of ChIP-Seq Data Analysis.*
- [34] **Kedar K. Patel**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (Fall 2010).
Dissertation title: *Intrinsic and Systematic Variability in Nanometer CMOS Technologies.*
- [35] **Raúl E. Aguilar Schall**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Fall 2010).
Dissertation title: *Semi-Parametric Graphical Computation Approach Using Loss-Based Estimation to Estimate Exposure Effects: Applications on Infant Developmental Outcomes.*
- [36] **Nancy Naichao Wang**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (Fall 2009).
Dissertation title: *Statistical Problems in DNA Microarray Data Analysis.*
- [37] **James H. Bullard**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Fall 2009).
Dissertation title: *Statistical Methods and Software for High-throughput Gene Expression Experiments.*
- [38] **Kasper D. Hansen**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Fall 2009).
Dissertation title: *Analyses of High-Throughput Gene Expression Data.*
- [39] **Margaret A. Taub**, Department of Statistics, UC Berkeley, Inside Member (Fall 2009).
Dissertation title: *Analysis of High-Throughput Biological Data: Some Statistical Problems in RNA-Seq and Mouse Genotyping.*

-
- [40] **Houston N. Gilbert**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Spring 2009).
Dissertation title: *Multiple Hypothesis Testing: Methodology, Software Implementation, and Applications to Genomics*.
- [41] **Thomas J. Sharpton**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (Spring 2009).
Dissertation title: *Investigations of Natural Genomic Variation in the Fungi*.
- [42] **Benjamin N. Blum**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (Fall 2008).
Dissertation title: *Resampling Methods for Protein Structure Prediction*.
- [43] **David A. Shilane**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (Fall 2008).
Dissertation title: *Optimization Studies with Multiple Testing, Loss-Based Estimation, and Confidence Intervals for Negative Binomials of High Dispersion*.
- [44] **Lee Heil Chae**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (Spring 2008).
Dissertation title: *Genomic Analysis of the Receptor-Like Kinase Superfamily in Plants*.
- [45] **Parvez Ahammad**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (Spring 2008).
Dissertation title: *Learning Data Driven Representations from Large Collections of Multidimensional Patterns with Minimal Supervision*.
- [46] **Luz B. Gilbert**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (Spring 2008).
Dissertation title: *Comparative Genomic Hybridizations in the Genus Neurospora and Associated Analysis*.
- [47] **Benjamin P. Berman**, Department of Molecular and Cell Biology, UC Berkeley (Fall 2006).
Dissertation title: *Gene Expression Diversity and Cis-Regulatory Sequence Models in the Transcriptional Network of Drosophila Embryogenesis*.
- [48] **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*.
- [49] **Jing Yi**, Department of Statistics, UC Berkeley (Summer 2006).
Dissertation title: *Absolute and Relative Quantification of Fluorescently Labeled DNA*.
- [50] **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*.
- [51] **Sandra E. Sinisi**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006).
Dissertation title: *Data-Adaptive Prediction with the Deletion/Substitution/Addition Algorithm*.
-

- [52] **Merrill D. Birkner**, Graduate Group in Biostatistics, UC Berkeley (Spring 2006).
Dissertation title: *Statistical Hypothesis Testing and Application to Biological Data*.
- [53] **Boonchai Bobby Boonyaratanakornkit**, Department of Chemical Engineering, UC Berkeley (Fall 2005).
Dissertation title: *Adaptation of the Deep-Sea Thermophile Methanocaldococcus jannaschii to Extreme Temperatures and Pressures*.
- [54] **Chao Chen**, Department of Statistics, UC Berkeley (Fall 2005).
Dissertation title: *Topics in Random Forests*.
- [55] **Yu Chuan Tai**, Graduate Group in Biostatistics, UC Berkeley (Spring 2005).
Dissertation title: *Multivariate Empirical Bayes Models for Replicated Microarray Time Course Data*.
- [56] **Ingileif B. Hallgrimsdottir**, Department of Statistics, UC Berkeley (Spring 2005).
Dissertation title: *Statistical Methods for Gene Mapping in Complex Diseases*.
- [57] **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*.
- [58] **Vivian Wai Ying Ng**, Department of Statistics, UC Berkeley (Fall 2004).
Dissertation title: *Univariate and Bivariate Variable Selection in High Dimensional Data*.
- [59] **François Collin**, Department of Statistics, UC Berkeley (August 2004).
Dissertation title: *Analysis of Oligonucleotide Data with a View to Data Quality Assessment*.
- [60] **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*.
- [61] **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*.
- [62] **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

- [1] **Andrew Sharo**, Graduate Group in Biophysics, UC Berkeley, Outside Member (February 14, 2019).
- [2] **Xiao Li**, Department of Statistics, UC Berkeley, Committee Chair (January 14, 2019).
- [3] **Suzanne Dufault**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 13, 2018).

-
- [4] **Stephanie DeGraaf**, Department of Statistics, UC Berkeley, Committee Chair (November 14, 2018).
 - [5] **Boying Gong**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (October 30, 2018).
 - [6] **Romain Lopez**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (September 11, 2018).
 - [7] **Nicholas Everetts**, Graduate Group in Computational Biology, UC Berkeley, Committee Chair (June 26, 2018).
 - [8] **Alexander Crits-Christopher**, Graduate Group in Microbiology, UC Berkeley, Outside Member (May 21, 2018).
 - [9] **Andrew Sharo**, Graduate Group in Biophysics, UC Berkeley, Outside Member (May 15, 2018).
 - [10] **Tal Ashuach**, Graduate Group in Computational Biology, UC Berkeley, Committee Chair (April 25, 2018).
 - [11] **Chenling Antelope Xu**, Graduate Group in Computational Biology, UC Berkeley, Committee Chair (December 11, 2017).
 - [12] **Courtney Schiffman**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (September 11, 2017).
 - [13] **Carlos Fernando Buen Abad Najar**, Graduate Group in Computational Biology, UC Berkeley, Committee Chair (May 10, 2017).
 - [14] **Calvin Chi**, Graduate Group in Computational Biology, UC Berkeley, Committee Chair (May 03, 2017).
 - [15] **Kelly N. Street**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (September 12, 2016).
 - [16] **Fanny Perraudau**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 05, 2016).
 - [17] **Jeremy R. Coyle**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (August 26, 2015).
 - [18] **Curt Hansen**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 01, 2014).
 - [19] **Lihn Tran**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (October 28, 2014).
 - [20] **Nathan Boley**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (May 12, 2014).

-
- [21] **Samuel D. Lendle**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (April 29, 2014).
 - [22] **Marcus H. Stoiber**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (March 17, 2014).
 - [23] **Ekaterina A. Eliseeva**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (October 22, 2013).
 - [24] **Inna Gerlovina**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 16, 2013).
 - [25] **Stephanie Sapp**, Department of Statistics, UC Berkeley, Committee Chair (April 24, 2013).
 - [26] **Shannon Hateley**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (April 02, 2013).
 - [27] **Curt Hansen**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (November 16, 2012).
 - [28] **Stephan Ritter**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (September 26, 2012).
 - [29] **Christine Ho**, Department of Statistics, UC Berkeley, Inside Member (May 10, 2012).
 - [30] **Miles E. Lopes**, Department of Statistics, UC Berkeley, Inside Member (April 24, 2012).
 - [31] **Kathryn L. (Benton) Colborn**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (February 14, 2012).
 - [32] **Jordan C. Brooks**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (December 15, 2011).
 - [33] **Sylvia Sudat Keuter**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (September 29, 2011).
 - [34] **Jennifer Creasman**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (September 13, 2011).
 - [35] **Stephen Yee**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (May 26, 2011).
 - [36] **Ma'ayan Bresler**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Inside Member (April 28, 2011).
 - [37] **Sofia Medina-Ruiz**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (April 27, 2011).
 - [38] **Dapo Omidiran**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (February 04, 2011).

-
- [39] **Kyungpil Kim**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (August 19, 2010).
 - [40] **Kevin S. McLoughlin**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (February 03, 2010).
 - [41] **Daisy (Yan) Huang**, Department of Statistics, UC Berkeley, Committee Chair (January 20, 2010).
 - [42] **Kedar K. Patel**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (August 28, 2009).
 - [43] **Jingyi (Jessica) Li**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (August 05, 2009).
 - [44] **Garmay E. Leung**, Department of Bioengineering, UC Berkeley, Outside Member (June 10, 2009).
 - [45] **Sherri Rose**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 11, 2009).
 - [46] **Oleg S. Mayba**, Department of Statistics, UC Berkeley, Committee Chair (May 08, 2009).
 - [47] **Wei-Chun Kao**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (May 04, 2009).
 - [48] **James H. Bullard**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (October 08, 2008).
 - [49] **Daniel B. Rubin**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (August 29, 2008).
 - [50] **Ying Xu**, Department of Statistics, UC Berkeley, Inside Member (August 12, 2008).
 - [51] **Catherine A. Tuglus**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 19, 2008).
 - [52] **Benjamin N. Blum**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (May 7, 2008).
 - [53] **Xin Victoria Wang**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (April 30, 2008).
 - [54] **Nancy Naichao Wang**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (April 28, 2008).
 - [55] **Matthew D. Davis**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (April 22, 2008).
 - [56] **Raúl E. Aguilar Schall**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (September 14, 2007).

-
- [57] **Houston N. Gilbert**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (August 27, 2007).
- [58] **David A. Shilane**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 7, 2007).
- [59] **Parvez Ahammad**, Department of Electrical Engineering and Computer Sciences, UC Berkeley, Outside Member (April 20, 2007).
- [60] **Matthew Lin**, Department of Bioengineering, UC Berkeley, Outside Member (March 14, 2007).
- [61] **Margaret A. Taub**, Department of Statistics, UC Berkeley, Inside Member (January 29, 2007).
- [62] **Xia Jiang**, Department of Bioengineering, UC Berkeley, Outside Member (December 8, 2006).
- [63] **Peter Dimitrov**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (August 24, 2006).
- [64] **Kasper D. Hansen**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (April 24, 2006).
- [65] **Sandra E. Sinisi**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 8, 2005).
- [66] **Thomas J. Sharpton**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (October 11, 2005).
- [67] **Merrill D. Birkner**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (May 11, 2005).
- [68] **Srikesh G. Arunajadai**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (April 29, 2005).
- [69] **Anat Caspi**, Department of Bioengineering, UC Berkeley, Outside Member (September 15, 2004).
- [70] **Jing Yi**, Department of Statistics, UC Berkeley, Inside Member (May 7, 2004).
- [71] **Lee Heil Chae**, Department of Plant and Microbial Biology, UC Berkeley, Outside Member (April 30, 2004).
- [72] **Benjamin P. Berman**, Department of Molecular and Cell Biology, UC Berkeley, Outside Member (January 21, 2004).
- [73] **Annette M. Molinaro**, Graduate Group in Biostatistics, UC Berkeley, Committee Chair (December 12, 2003).
- [74] **Yan Wang**, Graduate Group in Biostatistics, UC Berkeley, Inside Member (October 23, 2003).

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

EDITORIAL AND REFEREE SERVICE

Associate Editor

- *Annals of Applied Statistics* (<http://www.imstat.org/aoas>), Summer 2006 – Present.
- *Biology Direct* (<http://www.biology-direct.com>), Spring 2006 – Present.
- *Biometrics* (<http://www.biometrics.tibs.org>), July 01, 2009 – December 01, 2009.
- *BMC Bioinformatics* (<http://www.biomedcentral.com/bmcbioinformatics>), Spring 2005 – Fall 2009.
- *Genomics* (<http://www.journals.elsevier.com/genomics>), Summer 2006 – Summer 2010.
- *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (<http://www.computer.org/tcbb>), Spring 2004 – Present.
- *Journal of Statistical Software* (<http://www.jstatsoft.org>), Fall 2005 – Fall 2006.
- *Statistical Applications in Genetics and Molecular Biology* (<http://www.bepress.com/sagmb>), Spring 2002 – Summer 2012.

Other Editorial Role

- *Bioconductor Project Working Papers*, Administrator (<http://biostats.bepress.com/bioconductor/>), Spring 2004 – Present.
- *BMC Bioinformatics*, Editorial Advisor (<http://www.biomedcentral.com/bmcbioinformatics>), January 2012 – Present.

- *BMC Bioinformatics*, Section Editor for *Transcriptome Analysis* (<http://www.biomedcentral.com/bmcbioinformatics>), Fall 2009 – Fall 2011.

Referee

- *Bioinformatics*.
- *Biology Letters*.
- *Biometrical Journal*.
- *Biometrics*.
- *Biostatistics*.
- *BMC Genomics*.
- *France Berkeley Fund*.
- *Genetic Epidemiology*.
- *Institute of Mathematical Statistics Lecture Notes – Monograph Series*.
- *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).
- *Scandinavian Journal of Statistics*.
- *Springer*.
- *Statistical Methods in Medical Research*.
- *Statistica Sinica*.

External Advisor

- *Computational Systems Biology of Cancer*, U900 INSERM – Institute Curie – PSL Research University/CBIO Mines ParisTech – PSL Research University, Paris, France, Member of Scientific Advisory Board, 2016 – Present.
- *Center for Personal Dynamic Regulomes* (CPDR; <http://cpdr.stanford.edu>), HGRI Center of Excellence in Genomic Science (CEGS), Stanford University School of Medicine, Member of Advisory Board, 2014 – Present.

- *Algorithmics, Bioinformatics and Software solutions for the analysis of Next Generation Sequencing data* (ABS4NGS; <https://sites.google.com/site/abs4ngs/>), France, Member of Scientific Advisory Board, 2010 – Present.
- *NGS Leaders*, Member of Advisory Board, Summer 2010 – Present.
- *Centre for Applied Mathematics in Bioscience and Medicine* (<http://www.mcgill.ca/cambam>), McGill University, Montreal, Canada, Member of Scientific Advisory Board, Fall 2009 – Present.
- *Delphi International Expert Panel*, Evaluation and Prospective Valencian Agency (AVAP), Spain, 2009 – 2010.
- *Gene Expression Pattern Analysis* (GEPAS), Bioinformatics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain, Fall 2005.
- *First Call for Research Proposals in Bioinformatics*, Fundación BBVA (<http://www.fbbva.es>), Spain, Fall 2003.

SERVICE TO PROFESSIONAL SOCIETIES

- Member, Committee on Nominations, Institute of Mathematical Statistics, August 2018 – August 2019.
- Faculty Advisor, California and Nevada Regional Student Group (RSG), International Society for Computational Biology (ISCB), 2016 – Present.
- Member, International Society for Computational Biology (ISCB), 2016 – Present.
- Member, Council Subcommittee on Meeting Co-Sponsorship, Institute of Mathematical Statistics, August 2011 – August 2014.
- Elected Member, Council, Institute of Mathematical Statistics, August 02, 2011 – July 09, 2014.
- Regional Committee Representative, Western North American Region (WNAR) of the International Biometric Society (IBS), 2004 – 2006.
- Member, Western North American Region (WNAR) of the International Biometric Society (IBS), 2004 – Present.
- Member, Institute of Mathematical Statistics (IMS), 1996 – Present.
- Member, American Statistical Association (ASA), 1995 – Present.

UNIVERSITY SERVICE

- Chair, Department of Statistics, UC Berkeley (July 01, 2019 – Present)

- Member, Faculty Working Group, Division of Data Science and Information, UC Berkeley (2019).
- Statistics Chair Delegate, Governance Committee for Degree Programs of the Division of Data Sciences (DDSDP), UC Berkeley (2018 – Present).
- Member, Academic Personnel Committee, Department of Statistics, UC Berkeley (2018 – 2019).
- Member, Interdepartmental Committee on the Formation of the Division of Data Sciences, UC Berkeley (Spring 2018).
- Member, Center for Computational Biology/Electrical Engineering and Computer Sciences Faculty Search Committee, UC Berkeley (2018 – 2019).
- Member, Executive Committee, Berkeley Institute for Data Science (BIDS; <http://bids.berkeley.edu>), UC Berkeley (2017 – Present).
- Member, Masters Program Committee, Department of Statistics, UC Berkeley (2017 – 2018).
- Member, Educational Policy and Curriculum Committee (EPCC), School of Public Health, UC Berkeley (2017 – 2018).
- Member, Curriculum Committee, Center for Computational Biology, UC Berkeley (December 2016 – Present).
- Member, Advisory Board, Berkeley Institute for Data Science (BIDS; <http://bids.berkeley.edu>), UC Berkeley (2016 – Present).
- Member, Academic Personnel Committee, School of Public Health, UC Berkeley (2015 – 2018).
- Member, Development, Alumni Relations, Industry Relations and Media Relations Committee, Department of Statistics, UC Berkeley (2015 – 2016).
- Chair, Neyman Visiting Assistant Professor Search Committee, Department of Statistics, UC Berkeley (2015 – 2016).
- Member, Academic Personnel Committee, Department of Statistics, UC Berkeley (2014 – 2016).
- Chair, Biostatistics Faculty Search Committee, UC Berkeley (2013 – 2014).
- Member, Computational Biology Initiative Faculty Search Committee, UC Berkeley (2012 – 2013).
- Member, Development, Alumni Relations, Industry Relations and Media Relations Committee, Department of Statistics, UC Berkeley (2012 – 2014).
- Acting Head, Division of Biostatistics, UC Berkeley (July 01, 2012 – July 31, 2012).
- Member, Administrative Committee of the Graduate Council, UC Berkeley (2011 – 2018).

-
- Member, Graduate Council, UC Berkeley (2010 – 2018).
 - Member, Advisory Committee, Computational Genomics Research Facility (CGRL), QB3-Berkeley (2010 – 2011).
 - Member, Curriculum Committee, School of Public Health, UC Berkeley (2010 – 2013).
 - Member, Graduate Advising Committee, Graduate Group in Computational and Genomic Biology, UC Berkeley (2008 – Present).
 - Member, Faculty Council, School of Public Health, UC Berkeley (2008 – 2010).
 - Chair and Head Graduate Advisor, Graduate Group in Biostatistics, UC Berkeley (Fall 2007 – Fall 2017).
 - Faculty Advisor for GSI Affairs, Graduate Group in Biostatistics, UC Berkeley (Fall 2007 – Fall 2010).
 - Equity Advisor, Graduate Group in Biostatistics, UC Berkeley (Fall 2007 – Spring 2010).
 - Advisor, Designated Emphasis in Computational and Genomic Biology, Department of Statistics, UC Berkeley (2007 – 2008).
 - Member, Executive Committee, Center for Computational Biology (CCB), UC Berkeley (Fall 2007 – Fall 2009).
 - Member, Research Committee, School of Public Health, UC Berkeley (2007 – 2008).
 - Acting Chair and Head Graduate Advisor, Graduate Group in Biostatistics, UC Berkeley (Spring 2007).
 - Faculty Advisor, International Association for the Exchange of Students for Technical Experience (IAESTE) Local Committee, UC Berkeley (2006 – 2010).
 - Facilitator, National Research Council Assessment of Research Doctoral Programs, UC Berkeley (2006 – 2007).
 - Member, Computational Biology Initiative Faculty Search Committee, UC Berkeley (2006 – 2007).
 - Member, Social Epidemiology Faculty Search Committee, School of Public Health, UC Berkeley (2006 – 2007).
 - Member, QB3 Computing Resources Committee, UC Berkeley (2006 – 2007).
 - Member, Computational Biology Initiative Faculty Search Committee, UC Berkeley (2005 – 2006).
 - Chair, Seminars, Publicity, and Retreat Committee, Graduate Group in Computational and Genomic Biology, UC Berkeley (2003 – 2008).
 - Member, Curriculum Committee, School of Public Health, UC Berkeley (2003 – 2007).

- Member, Executive Committee, Graduate Group in Computational and Genomic Biology, UC Berkeley (2003 – 2005).
- Co-author with Steven E. Brenner, Proposal for a Designated Emphasis in Computational and Genomic Biology, UC Berkeley (2003).

CONSULTING

- AbbVie, May 2015 – Present.
- Whole Biome, May 2015 – Present.

COMPUTING SKILLS

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

Operating Systems: Linux, UNIX, Mac OS X.

LANGUAGE SKILLS

Fluent in English and French.

Intermediate Italian and Spanish.