

ELIZABETH PURDOM

Associate Professor
Department of Statistics
University of California at Berkeley

367 Evans Hall #3860
Berkeley, CA 94720-3860

Tel: (510) 642-6154

Fax: (510) 642-7892

www.stat.berkeley.edu/~epurdom

E-mail: epurdom@stat.berkeley.edu

RESEARCH INTERESTS

My research interests lie in developing statistical methods for high-dimensional data arising in the field of biology and genetics. I focus on questions of robust estimation and hypothesis testing for high-throughput biological experiments, in particular sequencing technologies. My application areas include the areas of gene expression analysis, single cell sequencing, cancer biology, microbiome, methylation (bisulfite) sequencing, normalization of sequencing data, and integration of multiple ‘omics data experiments. I am also interested in integration of heterogeneous sources of data, where the data can be multiple experimental platforms or, more generally, arbitrary forms of preexisting biological knowledge such as networks or trees. Statistically, I am interested in questions of high-dimensional inference and multivariate analysis – problems that arise naturally in trying to create a unified understanding of this type of data.

AFFILIATIONS

Department of Statistics, UC Berkeley

Graduate Group in Biostatistics, UC Berkeley

Center for Computational Biology, UC Berkeley

POST-GRADUATE EMPLOYMENT

Jul 2016- Associate Professor
Present *Department of Statistics, UC Berkeley, CA*

Jul 2009- Assistant Professor
Jul 2017 *Department of Statistics, UC Berkeley, CA*

Mar 2007- NSF Post-Doctoral Fellow in Biological Informatics
Jul 2009 *Division of Biostatistics, UC Berkeley, CA*

Sept 2006- Post-Doctoral Scholar
Feb 2007 *Mathematical Sciences Research Institute, Berkeley, CA*

EDUCATION

- 2001-2006 Stanford University Stanford, CA
Ph.D., Statistics
- Advisor: Susan Holmes
 - Thesis Title: Multivariate Kernel Methods in the Analysis of Graphical Structures
- 1996-2000 Yale University New Haven, CT
B.S. Applied Mathematics and History (Double Major)
- Summa Cum Laude
 - Distinction in the major of Applied Mathematics, concentration in Statistics
 - Distinction in the major of History

AWARDS AND HONORS

- 09/2013 Hellman Family Faculty Fund Award
- 09/2008 Stellar Abstract Award, Conference on Emerging Quantitative Issues in Parallel Sequencing, Harvard School of Public Health
- 2007-2009 NSF Post-Doctoral Fellowship in Biological Informatics
- 06/2006 Departmental Teaching Award (Stanford University)
- 2001-2004 Gabilan Stanford Graduate Fellowship in Science and Engineering (Stanford University)
- 06/2003 Departmental Teaching Award (Stanford University)
- 05/2000 Wrexham Prize for the best senior essay in the field of Humanities (Yale University)
John Addison Porter Prize in American History (Department of History, Yale University)
- 11/1999 Phi Beta Kappa (Phi Beta Kappa Society, Yale Chapter)

GRANT ACTIVITIES

Active Awards

- 9/20/17-
6/30/22 A Comprehensive Whole-Brain Atlas of Cell Types in the Mouse.
Lead PI: John Ngai (UC Berkeley) U19 MH114830 (NIH)
Role in project: Co-Investigator Total Direct Costs: \$908,808
- 8/15/15-
9/15/21 Epigenetic Control of Drought Response in Sorghum (EPICON)
Lead PI: Peggy Lemaux (UC Berkeley) DE-SC0014081(DOE)
Role in project: Co-PI Total Direct Costs: \$4,690,531

Completed Awards

9/26/14- 6/30/17	Classification of Cortical Neurons by Single Cell Transcriptomics Lead PI: John Ngai (UC Berkeley) Role in project: Co-Investigator	U01 MH105979 (NIH) Total Direct Costs: \$2,778,372
09/16- 06/17	UC Berkeley Presidential Chair Fellows -- Curriculum Enrichment Grant Lead PIs: Elizabeth Purdom and Adityanand Guntuboyina Award: \$20,000	
09/13- 09/14	Hellman Family Faculty Fund Award Lead PI: Elizabeth Purdom Award: \$24,842	
9/15/10- 8/31/14	SCREMS: Building a Statistical Computing Environment to Support Scientific Research Lead PI: Deborah Nolan (UC Berkeley) Role in project: Co-PI	DMS-1026441(NSF) Total Direct Costs: \$101,213
09/28/2009- 07/31/2014	TCGA Data Analysis Center At Berkeley Lead PI: Paul Spellman Role in project: Collaborator 2009-2010	U24CA143799 Total Direct Costs: \$724,564

Grant Review Panel and Other Activities with Granting Agencies

11/2016	Participant in NSF Review Panel
10/25/2015	Participant in NSF Genomes to Phenomes (G2P) Working Group

PUBLICATIONS

Refereed Journal Publications

1. Gong B, Zhou Y, and **Purdom E** (under revision with *Genome Biology*) "Cobolt: Joint analysis of multimodal single-cell sequencing data"
2. Roux de Bezieux H, Street K, Fischer S, Van den Berge K, Chance R, Risso D, Gillis J, Ngai J, **Purdom E**, and Dudoit S (under revision) "Improving replicability in single-cell RNA-Seq cell type discovery with Dune"
3. BRAIN Initiative Cell Census Network (BICCN) (to appear) "A multimodal cell census and atlas of the mammalian primary motor cortex" *Nature*
4. Yao Z, Liu H, Xie F, Fischer S, et al (to appear) "An integrated transcriptomic and epigenomic atlas of mouse primary motor cortex cell types" *Nature*
5. Hicks SC, Liu R, Ni Y, **Purdom E**, Risso D (2021) "mbkmeans: fast clustering for single cell data using mini-batch k-means" *PLOS Computational Biology* 17(1): e1008625.
6. Varoquaux N and **Purdom E** (2020) "A pipeline to analyse time-course gene expression data" *F1000Research* 9:1447

7. Brann DH, Tsukahara T, Weinreb C, Lipovsek M, Van den Berge K, Gong B, Chance R, Macaulay IC, Chou H-J, Fletcher RB, Das D, Street K, Roux de Bezieux H, Choi YG, Risso D, Dudoit S, **Purdom E**, Mill J, Hachem RA, Matsunami H, Logan DW, Goldstein BJ, Grubb MS, Ngai J, Datta SR (2020) "Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia" *Science Advances* 6: eabc5801.
8. Gao C, Montoya L, Xu L, Madera M, Hollingsworth J, **Purdom E**, Singan V, Vogel J, Hutmacher RB, Dahlberg JA, Coleman-Derr D, Lemaux PG, and Taylor JW (2020) "Fungal community assembly in drought-stressed sorghum shows stochasticity, selection, and universal ecological dynamics" *Nature Communications* 11:34.
9. Varoquaux N, Cole B, Gao C, Pierroz G, Baker CR, Patel D, Madera M, Jeffers T, Hollingsworth J, Sievert J, Yoshinaga Y, Owiti JA, Singan VR, DeGraaf S, Xu L, Blow MJ, Harrison MJ, Visel A, Jansson C, Niyogi KK, Hutmacher R, Coleman-Derr D, O'Malley RC, Taylor JW, Dahlberg J, Vogel JP, Lemaux PG, and **Purdom E** (2019) "Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses" *PNAS* 116(52): 27124-27132.
10. Gong B and **Purdom E** (2019) "MethCP: Differentially Methylated Region Detection with Change Point Models" In: Cowen L. (eds) *Research in Computational Molecular Biology. RECOMB 2019. Lecture Notes in Computer Science*, vol 11467.
11. El Karoui N and **Purdom E** (2019) "The bootstrap, covariance matrices and PCA in moderate and high-dimensions" *Proceedings of Machine Learning Research: The 22nd International Conference on Artificial Intelligence and Statistics*. 89: 2115-2124.
12. Harirchian P, Lee J, Hilz S, Sedgewick AJ, Perez White BE, Kesling MJ, Mully T, Golovato J, Gray M, Mauro TM, **Purdom E**, Kim EA, Sbitany H, Bhutani T, Vaske CJ, Benz SC, Cho RJ, and Cheng JB (2019) "A20 and ABIN1 Suppression of a Keratinocyte Inflammatory Program with a Shared Single-Cell Expression Signature in Diverse Human Rashes," *Journal of Investigative Dermatology*, 139(6), 1264-1273.
13. Cole MB, Risso D, Wagner A, DeTomaso D, Ngai J, **Purdom E**, Dudoit S, Yosef N (2019) "Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq" *Cell Systems* 8(4): 315-328.
14. Cho RJ, Alexandrov LB, den Breems NY, Atanasova VS, Farshchian M, **Purdom E**, Nguyen TN, Coarfa C, Rajapakshe K, Prisco M, Sahu J, Tassone P, Greenawalt EJ, Collisson EA, Wu W, Yao H, Su X, Guttmann-Gruber C, Hofbauer PN, Hashmi R, Fuentes I, Benz SC, Golovato J, Ehli EA, Davis CM, Davies GE, Covington KR, Murrell DF, Salas-Alanis JC, Palisson F, Bruckner AL, Robinson W, Has C, Bruckner-Tuderman L, Titeux M, Jonkman MF, Rashidghamat E, Lwin SM, Mellerio JE, McGrath JA, Bauer JW, Hovnanian A, Tsai KY, South AP (2018) "APOBEC mutation drives early-onset squamous cell carcinomas in recessive dystrophic epidermolysis bullosa" *Science Translational Medicine* 10 (455): eaas9668

15. Cheng JB, Sedgewick AJ, Finnegan AI, Harirchian P, Lee J, Kwon S, Fassett MS, Golovato J, Gray M, Ghadially R, Liao W, White BEP, Mauro TM, Mully T, Kim EA, Sbitany H, Neuhaus IM, Grekin RC, Yu SS, Gray JW, **Purdome E**, Paus R, Vaske CJ, Benz SC, Song JS, Cho RJ (2018). "Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution" *Cell Reports*, 25(4), 871–883.
16. Gao C, Montoya L, Xu L, Madera M, Hollingsworth J, **Purdome E**, Hutmacher RB, Dahlberg JA, Coleman-Derr D, Lemaux PG, and Taylor JW (2018) "Strong succession in arbuscular mycorrhizal fungal communities" *The ISME Journal*
17. Risso D, Purvis L, Fletcher R, Das D, Ngai J, Dudoit S, **Purdome E** (2018) "clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets" *PLoS Comput Biol*. 2018 Sep 4;14(9):e1006378
18. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, **Purdome E**, and Dudoit S (2018) "Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics" *BMC Genomics* 19(1): 477.
19. El Karoui N and **Purdome E** (2018) "Can we trust the bootstrap in high-dimension?" *Journal of Machine Learning Research* 19:1-66.
20. Perraudon F, Risso D, Street K, **Purdome E**, and Dudoit S (2017) "Bioconductor workflow for single-cell RNA sequencing: Normalization, dimensionality reduction, clustering, and lineage inference" *F1000Research* 6:1158.
21. Naylor D, DeGraaf S, **Purdome E**, and Coleman-Derr D (2017) "Drought and host selection influence bacterial community dynamics in the grass root microbiome" *ISME Journal* 11: 2691–2704.
22. Gadye L, Das D, Sanchez MA, Street K, Baudhuin A, Wagner A, Cole MB, Choi YG, Yosef N, **Purdome E**, Dudoit S, Risso D, Ngai J, and Fletcher RB (2017) "Injury Activates Transient Olfactory Stem Cell States with Diverse Lineage Capacities" *Cell Stem Cell* 21(6): 775-790
23. Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, Cole MB, Flores Q, Choi YG, Yosef N, **Purdome E**, Dudoit S, Risso D and Ngai J (2017) "Deconstructing Olfactory Stem Cell Trajectories at Single Cell Resolution" *Cell Stem Cell* 20(6):817-830.
24. Johnson M and **Purdome E** (2017) "Clustering of mRNA-Seq data for detection of alternative splicing patterns" *Biostatistics* 8(2): 295-307.
25. Ruddy S, Johnson M, and **Purdome E** (2016) "Shrinkage of dispersion parameters in the binomial family, with application to differential exon skipping," *Annals of Applied Statistics* 10: 690-725.
26. Sanborn Z, Chung J, **Purdome E**, Wang NJ, Kakavand H, Wilmott J, Butler T, Thompson JF, Mann GJ, Haydu LE, Saw RP, Busam KJ, Lo R, Collisson EA, Hur JS, Spellman P, Cleaver JE, Gray JW, Huh N, Murali R, Scolyer RA, Bastian BC, and Cho RJ (2015) "Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination" *PNAS* 112(35): 10995-11000.

27. Zheng CL, Wang NJ, Chung J, Moslehi H, Sanborn JZ, Hur JS, Collisson EA, Vemula SS, Naujokas A, Chiotti KE, Cheng JB, Fassihi H, Blumberg AJ, Bailey CV, Fudem GM, Mihm FG, Cunningham BB, Neuhaus IM, Liao W, Oh DH, Cleaver JE, LeBoit PE, Costello JF, Lehmann AR, Gray JW, Spellman PT, Arron ST, Huh N, **Purdom E***, and Cho RJ (2014) "Transcription Restores DNA Repair to Heterochromatin, Determining Regional Mutation Rates in Cancer Genomes" *Cell Reports* 9(4): 1228-1234.
**Joint corresponding author*
28. **Purdom E***, Ho C, Grasso CS, Quist MJ, Cho RJ, and Spellman P. (2013) "Methods and challenges in timing chromosomal abnormalities within cancer samples," *Bioinformatics*. 29(24): 3113-20.
**Both first author and corresponding author*
29. **Purdom E**, Restall C, Busuttill RA, Schluter H, Boussioutas A, Thompson EW, Anderson RL, Speed TP, and Haviv I. (2013) "Determining epithelial contribution to in vivo mesenchymal tumour expression signature using species-specific microarray profiling analysis of xenografts." *Genetics Research*. 95(1):14-29.
30. Couvillion M, Bounova G, **Purdom E**, Speed T, and Collins K. (2013) "A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus" *Molecular Cell*. 48:509-20.
31. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, Gibb WJ, Wang NJ, Ziyad S, Tong F, Bayani N, Hu Z, Billig JI, Dueregger A, Lewis S, Jakkula L, Korkola JE, Durinck S, Pepin F, Guan Y, **Purdom E**, Neuvial P, Bengtsson H, Wood KW, Smith PG, Vassilev LT, Hennessy BT, Greshock J, Bachman KE, Hardwicke MA, Park JW, Marton LJ, Wolf DM, Collisson EA, Neve RM, Mills GB, Speed TP, Feiler HS, Wooster RF, Haussler D, Stuart JM, Gray JW, and Spellman PT (2012) "Subtype and pathway specific responses to anticancer compounds in breast cancer," *PNAS* 109(8): 2724-9.
32. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, Pons J, Chan SW, Lam ET, Chu C, Park K, Hong SW, Hur JS, Huh N, Neuhaus IM, Yu SS, Grekin RC, Mauro TM, Cleaver JE, Kwok PY, LeBoit PE, Getz G, Cibulskis K, Aster JC, Huang H, **Purdom E**, Li J, Bolund L, Arron ST, Gray JW, Spellman PT, and Cho RJ (2011). "Temporal Dissection of Tumorigenesis in Primary Cancers." *Cancer Discovery*. 1:137-143.
33. The Cancer Genome Atlas Research Network (2011) "Integrated genomic analyses of ovarian carcinoma," *Nature*. 474(7353):609-15.
34. Wang XV, Verhaak RGW, **Purdom E**, Spellman PT, and Speed TP (2011). "Unifying gene expression measures from multiple platforms using factor analysis" *PloS one*, 6(3): e17691.
35. **Purdom E** (2011) "Analysis Of A Data Matrix And A Graph: Metagenomic Data And The Phylogenetic Tree", *Ann. Appl. Stat.* 5(4) (2011), 2326-2358.

36. Lapuk A, Marr H, Jakkula L, Pedro H, Bhattacharya S, **Purdom E**, Hu Z, Simpson K, Pachter L, Durinck S, Wang N, Parvin B, Fontenay G, Speed T, Garbe J, Stampfer M, Bayandorian H, Dorton S, Clark TA, Schweitzer A, Wyrobek A, Feiler H, Spellman P, Conboy J, and Gray JW. (2010) "Exon-level microarray analyses identify alternative splicing programs in breast cancer." *Mol Cancer Res.* 2010 Jul;8(7):961-74.
37. Bullard J*, **Purdom E***, Hansen K, and Dudoit S (2010). "Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments", *BMC Bioinformatics*, 11:94.
*Both authors contributed equally to this work
38. Verhaak R, Hoadley K, **Purdom E**, Wang V, Qi Y, Wilkerson M, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou C, Hayes DN and The Cancer Genome Atlas Research Network (2010) "Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1", *Cancer Cell* 17(1): 98-110.
39. The Cancer Genome Atlas Research Network (2008) "Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways", *Nature* 455(7216): 1061-1068.
40. **Purdom E**, Simpson K, Robinson M, Conboy J, Lapuk A and Speed T (2008) "FIRMA: A Method for Detection of Alternative Splicing from Exon Array Data", *Bioinformatics* 24(15): 1707-1714.
41. Bik E, Eckburg P, Gill S, Nelson K, **Purdom E**, Francois F, Perez-Perez G, Blaser M, and Relman D. (2006) "Molecular Analysis of the Bacterial Microbiota in the Human Stomach", *PNAS* 103(1): 732-737.
42. **Purdom E** and Holmes S (2005) "Error Distribution for Gene Expression Data", *Statistical Applications in Genetics and Molecular Biology* 4(1): Article 16.
<http://www.bepress.com/sagmb/vol4/iss1/art16>
43. Eckburg P, Bik E, Bernstein C, **Purdom E**, Dethlefsen L, Sargent M, Gill S, Nelson K, and Relman D. (2005) "Diversity of the Human Intestinal Microbial Flora", *Science* 308: 1635-1638.
44. Xu T, Xhu C.-T, **Purdom E**, Dang D, Ilsley D, Guo Y, Weber J, Holmes S.P. and Lee P.P (2004) "Microarray Analysis Reveals Differences in Gene Expression of Circulating CD8+ T Cells in Melanoma Patients and Healthy Donors", *Cancer Research* 64: 3661-3667.

Books / Book Chapters

45. **Purdom E** (2020) *Statistical Methods for Data Science*
Online book available at: <https://epurdom.github.io/Stat131A/book/index.html>

46. **Purdom E** and Mukherjee S (2011) “Transcriptomic Technologies and Statistical Data Analysis” in *Handbook of Statistical Systems Biology*. (eds M. P. H. Stumpf, D. J. Balding and M. Girolami), Wiley & Sons, Ltd, Chichester, UK, 135-162.

Technical Reports

47. **Purdom E** (2015) "Can We Trust the Bootstrap in High Dimensions?" Technical Report #824, Department of Statistics, UC Berkeley
48. **Purdom E** (2012) "Annotation-Free Estimates of Gene-Expression from mRNA-Seq," Technical Report #825, Department of Statistics, UC Berkeley
49. **Purdom E** (2008) “Analyzing Data with Graphs: Metagenomic Data and the Phylogenetic Tree”, Technical Report #766, Department of Statistics, UC Berkeley.

Software

50. Cobolt: Joint analysis of multimodal single-cell sequencing data (python package available on Github <https://github.com/boyinggong/cobolt>)
51. moanin: An R Package for Time Course RNASeq Data Analysis (2020) (Bioconductor)
52. mbkmeans: Mini-batch K-means Clustering for Single-Cell RNA-seq (2019) (Bioconductor).
53. methCP: An R package for Differentially Methylated Region Detection (2018) (available on github: boyinggong.github.io/methcp)
54. clusterExperiment: a package for clustering of single-cell sequencing data and other high-throughput experiments with large numbers of samples (2015) (Bioconductor).
55. cancerTiming: Estimation of temporal ordering of cancer abnormalities (2012) (CRAN, <http://cran.r-project.org>).
56. DoubleExpSeq: Differential Exon Usage Test for RNA-Seq data via Empirical Bayes Shrinkage of the Dispersion Parameter (2014) (CRAN, <http://cran.r-project.org>).
57. FIRMA: implemented in the R package *aroma.affymetrix*, part of the *aroma* project (<http://www.aroma-project.org>).

PRESENTATIONS

Invited Presentations (Conferences and Workshops)

1. *Single Cell mRNA-Seq: Subtype discovery and lineage estimation*
ISU BCB Symposium
Iowa State, May 1, 2019
2. *Single Cell mRNA-Seq: Subtype discovery and lineage estimation* (Keynote)
Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM2018)
Marseille, France, July 6, 2018
3. *Lineage Estimation with Single-Cell mRNA-Seq Data*
Statistical Methods for post-genomic data (SMPG2018)
Montpellier, France, January 12, 2018

4. *Statistical Challenges in Analyzing High Dimensional Experiments in Molecular Biology: The example of estimating lineages in single-cell sequencing data*
Special afternoon of Data Science in ENS
Paris, France, December 8, 2017
5. *Lineage Estimation with Single-Cell mRNA-Seq Data*
Statistical and Computational Challenges in Large Scale Molecular Biology
Bamff International Research Station, Mar 28, 2017
6. *Lineage Estimation with Single-Cell mRNA-Seq Data*
DASHU 2017: Data Science and Computational Precision Health
Feb 22, 2017
7. *Expression Analysis of Tumors Based on Patterns in Alternative Splicing*
Workshop on Computational Cancer Biology
Simons Institute, Berkeley, CA, Feb 1-5, 2016
8. *Differential Gene Expression Analysis*
Asia Pacific Bioinformatics Conference
South San Francisco, CA, January 10, 2016
9. *Comparing Temporal Orderings of Chromosomal Abnormalities Using Mutation Data*
Colloque de Bioinformatique Moléculaire
Institut Pasteur, Paris, France, October 7-8, 2015
10. *Practices for Reproducible Research in Bioinformatics*
Panel Discussion, Opening Workshop, SAMSI Program on Beyond Bioinformatics
Research Triangle Park, NC, September 8-12, 2014
11. *Clustering mRNA-Seq Samples for Alternative Splicing Detection*
Invited Session, Joint Statistical Meetings (JSM)
Boston, August 2-7, 2014
12. *Comparing Temporal Orderings of Chromosomal Abnormalities Using Mutation Data*
Information Theory and Applications Workshop
San Diego, February 9-14, 2014
13. *Timing Chromosomal Abnormalities in Tumors using Mutation Data*
Invited Session, Joint Statistical Meetings (JSM)
Montreal, August 7, 2013
14. *Measuring Subclonality in Tumors Using NGS*
Invited Session, Annual Meeting of the Western North American Region of The
International Biometric Society (WNAR)
Los Angeles, June 16-19, 2013
15. *Timing Chromosomal Abnormalities in Tumors using Mutation Data*
Biostatistics Symposium
San Francisco Bay Area Chapter, ASA
Stanford, June 1, 2013
16. *Timing Chromosomal Abnormalities in Tumors using Mutation Data*
Invited Session, Annual Meeting of the Western North American Region of The
International Biometric Society (WNAR)
Fort Collins, June 17-20, 2012

17. Invited speaker, unable to attend due to birth of child
Transcriptomics and Epigenomics
Los Angeles, October 25-28, 2011
18. *Statistical Methods for normalization and DE in mRNA-Seq experiments*
Illumina Expert Panel on Sequencing
Toronto, July 22, 2010
19. *Detecting alternative splicing with RNAseq: Some Thoughts on the Pipeline*
Joint Statistical Meetings
Vancouver, July 31 – August 5, 2010
20. *Connecting Microarray and Sequencing Experiments*
Invited Session, Annual Meeting of the Western North American Region of The
International Biometric Society (WNAR)
Seattle, June 20-23, 2010
21. *Exon Array Data for Tumor Profiling*
Approaches for Analysis of mRNA-Seq on Tumor Samples
Genomic Aspects of Glioblastoma Multiforme
CEIT, San Sebastian, Spain. June 14-16, 2010
22. *Data Analysis with Graphs*
Massive Modern Data Sets
Stanford University, June 2008
23. *Use of the Graph Laplacian to Analyze Network Data*
Invited Session, Joint Statistical Meetings
Denver, August 5, 2008
24. *Estimating Transcription Using Next Generation Sequencing*
Statistical and Computational Challenges in Next-Generation Sequencing
MSRI, October 9, 2008
25. *Introduction to Exon Arrays*
Detection of Alternative Splicing Using Exon Arrays
Mathematical Systems Biology of Cancer II
MSRI, November, 2007

Invited Presentations (Seminars)

26. *Topics in the Analysis of High-Throughput Molecular Data and the Bootstrap in High Dimensions* (Nov 11, 2019)
Department of Statistics
U. of Chicago
27. *Analyzing Epigenetic Time Course Data for Drought Response* (Mar 26, 2018)
MIA Séminaire
Agro Paris Tech, Paris, France
28. *Analysis of a High-Throughput Experiment:
the example of single-cell sequencing* (Feb 13, 2018)
Data Science Colloquium
Ecole Normale Supérieure, Paris, France

29. *Can We Trust the Bootstrap in (Moderately) High Dimensions?* (Feb 5, 2018)
SMILE Workshop
Paris, France
30. *Lineage Estimation with Single-Cell mRNA-Seq Data* (Nov 30, 2017)
Bioinformatics/Genetics Seminar
Institut Cochin, Paris, France
31. *Robust Strategies for the analysis of Single-Cell mRNA-Seq Data* (October 20, 2017)
Laplace Data Science Reading Group
Ecole Normale Supérieure, Paris France
32. *Robust Strategies for Analysis of Single-Cell mRNA-Seq Data* (Dec 1, 2016)
Statistics and Genomics Seminar
UC, Berkeley
33. *Robust Strategies for Analysis of Single-Cell mRNA-Seq Data* (Nov 10, 2016)
Biostatistics Seminar
Stanford University
34. *Robust Strategies for Analysis of Single-Cell mRNA-Seq Data* (Nov 2, 2016)
Visiting Lectures Series
Department of Bioinformatics and Computational Biology
MD Anderson Cancer Center
35. *Robust strategies for analysis of single cell mRNA-Seq data* (June 2, 2016)
Science & Technology Seminar
Joint Genome Institute
36. *Comparing Temporal Orderings of Chromosomal Abnormalities Using Mutation Data*
(November 14, 2014)
Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute
37. *Shrinkage of Dispersion Estimates for Analysis of Exon Usage in mRNA-Seq Experiments* (May 13, 2014)
Department of Statistics
UC, Davis
38. *Timing Chromosomal Abnormalities using Mutation Data* (October 15, 2013)
Department of Statistics
UC, Riverside
39. *Timing Chromosomal Abnormalities using Mutation Data* (May 1, 2013)
Department of Epidemiology and Biostatistics
UC, San Francisco
40. *Timing Chromosomal Abnormalities using Mutation Data* (July 26, 2012)
Laboratoire Statistique et Génome, CNRS
41. *Timing Chromosomal Abnormalities using Mutation Data* (June 29, 2012)
Institute Curie, Paris, France
42. *Simplifying mRNA-Seq* (October 28, 2010)
Biostatistics Workshop
Stanford

43. *Normalizing mRNA-Seq Data* (March 3, 2010)
Mathematical and Computational Biology Seminar
UC Berkeley
44. *Simplifying mRNA-Seq* (February 24, 2010)
Statistics and Genomics Seminar
UC, Berkeley
45. *Estimating Transcription Using Next Generation Sequencing* (November 19, 2009)
Stanford Genome Technology Center
46. *Statistical Problems in Estimating Alternative Splicing* (March 19, 2009)
Division of Biostatistics
New York University
47. *Statistical Problems in Estimating Alternative Splicing* (March 18, 2009)
Department of Biostatistics
Columbia University
48. *Statistical Problems in Estimating Alternative Splicing* (March 9, 2009)
Department of Bioinformatics
Emory University
49. *Statistical Problems in Estimating Alternative Splicing* (February 26, 2009)
The Broad Institute
50. *Statistical Problems in Estimating Alternative Splicing* (February 10, 2009)
Program in Biostatistics and Biomathematics
Fred Hutchinson Cancer Research Center
51. *Statistical Problems in Estimating Alternative Splicing* (February 6, 2009)
Department of Statistics
Columbia University
52. *Statistical Problems in Estimating Alternative Splicing* (February 2, 2009)
Department of Bioinformatics and Computational Biology
MD Anderson
53. *Estimating Splicing Events Using Sequencing Data* (December 5, 2008)
Department of Statistics
Stanford, University
54. *Estimating Transcription Using Next Generation Sequencing* (September 22, 2008)
Correlated and High-Dimensional Data Seminar
Biostatistics, Harvard Medical School
55. *Detection of Alternative Splicing Using Exon Arrays* (March 6, 2008)
Statistics and Genomics Seminar
UC, Berkeley
56. *Data Analysis Using Graphical Structures* (January 8, 2007)
Applied Mathematics Seminar
Institut National de la Recherche Agronomique, Jouy-en-Josas, France

57. *A Look at Data Analysis with Graphs: By Way of Spectral Decompositions of Graphs* (March 7, 2007)
Neyman Seminar, Department of Statistics
UC Berkeley
58. *Data Analysis Using Graphical Structures* (May 23, 2006)
Department of Statistics
Stanford University
59. *Data Analysis Using Graphical Structures* (February 15, 2006)
Division of Biostatistics
UC Berkeley.
60. *Analysis of Ecological Data: Use of Phylogenetic Trees with Diversity Measurements* (December 8, 2005)
Statistics and Genomics Seminar
UC Berkeley

Contributed Presentations

61. *Clustering Tumors Based on Patterns in Alternative Splicing*
Beyond the Genome 2014
Boston, October 8-10, 2014
62. *Estimating Transcription Using Next Generation Sequencing*
Emerging Quantitative Issues in Parallel Sequencing
Harvard Medical School, September 25, 2008

Posters Presented

63. Purdom E, Ho C, Grasso CS, Quist MJ, Cho RJ, and Spellman P (November 25, 2013)
“Timing Chromosomal Abnormalities in Tumors Using Mutation Data” Impact of Large-Scale Genomic Data on Statistical and Quantitative Genetics, U. Washington, Seattle, WA
64. Purdom E., Ho C., Huang H, Durinck S., Spellman PT, and Cho, RJ. (June 20-25, 2011)
“Timing Chromosomal Abnormalities using Mutation Data” Changing Landscape of the Cancer Genome, Boston, MA
65. Purdom E., Bullard J., Hansen K., and Dudoit S. (October 6, 2009) “Improving Differential Expression Detection with mRNA-Seq” 12th International MGED Meeting, Phoenix, Arizona.
66. Purdom E. and Holmes S. (June 2005) “Multivariate Analysis of Data Collected on Species” Mathematics of Evolution and Phylogeny Conference, Paris, France.
67. Dethlefsen L, Purdom E, Eckburg PB, Bik EM, Bernstein CN, Sargent M, Gill SR, Nelson KE, and Relman DA (April 2005) “Comparisons of 16S rDNA clone libraries using multiple methods to investigate within- and between-subject variability of the colonic bacterial communities in 3 human individuals” ASM Beneficial Microbes Conference, Lake Tahoe, Nevada.

68. Purdom, E. and Holmes, S (August 2004) “Asymmetric Laplace Distribution for Gene Expression Data” BIRS Statistical Science for Genome Biology Workshop, Banff, Canada.

CONFERENCE ORGANIZING

1. “Statistical Problems in Cancer Genomics” (Session Organizer)
Joint Statistical Meetings (JSM)
Boston, August 2-7, 2014
2. “Cancer Genomics” (Session Organizer)
Annual Meeting of the Western North American Region of The International Biometric Society (WNAR)
Los Angeles, June 16-19, 2013
3. RNA-Seq Data Analysis Workshop (Organizer)
Computational Genomics Resource Laboratory, UC Berkeley
September 26, 2011 and Spring 2013
4. Berkeley Sequencing Group (Organizer)
Weekly, Fall 2010 – Fall 2011
5. “Statistical Methods for High-throughput Sequencing” (Session Organizer)
Annual Meeting of the Western North American Region of The International Biometric Society (WNAR)
Seattle, Washington, June 20-23, 2010
6. “Mathematical Systems of Biology of Cancer II” (Conference Organizer)
Mathematical Sciences Research Institute, Berkeley, CA, November 2007

EDITORIAL AND ORGANIZATIONAL ACTIVITIES

Associate Editor

- *Statistical Applications in Genetics and Molecular Biology* (www.bepress.com/sagmb)
Fall 2009 – Fall 2011

Referee

- Annals of Applied Statistics
- Bioinformatics
- BMC Genomics
- Biometrics
- Biostatistics
- Computational Statistics and Data Analysis
- Genetics
- Genome Biology
- Genome Medicine
- Journal of the American Statistical Association
- Journal of the Royal Statistical Society
- Journal of Statistical Computation and Simulation

- Nature Biotechnology
- Nature Communications
- Nature Scientific Reports
- Nucleic Acid Research
- Statistical Applications in Genetics and Molecular Biology

DEPARTMENTAL AND CAMPUS LEADERSHIP

Campus Roles

- 2020-21 CDSS Research Advisory Committee
Ad hoc research advisory committee to help inform Associate Provost of the Division of Computing, Data Science and Society (CDSS)
- Spring 2019 University Fellowship Committee, Physical Sciences & Engineering section
Campus-wide committee charged with reviewing and ranking nominations of admitted graduate students for University multi-year fellowships (Berkeley Fellowships & Chancellor's Fellowships).
- Spring 2011 Advisory committee for planning of the Computational Genomics Resource Laboratory (CGRL)

Department of Statistics Roles

- 2018-21 Head Graduate Advisor (HGA)
 2016-17 *Oversight of the PhD Graduate Program, including advising first year students, deciding on students for awards and fellowships, and general oversight of students meeting milestones. During my time as HGA, I developed a new mentoring system for our PhD program, initiated a new doctoral seminar for the first year students, revised our requirements for the first year program, and updated our PhD policies, and updated our funding strategy to be more competitive.*
- 2018-20 (Chair) PhD Admissions Committee
 2016-17 (Chair) *The main duties of this committee are the PhD Admissions and recruitment. We receive roughly 300 applications a year and admit around 20 students, and the faculty committee reviews all these applications. As chair, I would additionally coordinate the review process and oversee the recruitment of the admitted students*
- 2014-16 (Member)
- Summer 2017 Ad hoc committee for the development of new data science major
- Fall 2015 Ad hoc “Next Decade Committee” committee
Committee tasked to prepare for department’s ten year review

- 2013-2014 MA Admissions Committee, Member
The main duties of this committee are the MA Admissions
- 2012-2013 Development & Alumni relations
This committee created the department's Industrial Alliance Program, organizing its inaugural meeting, and developing our alumni connections.
- 2011-2012 Course curriculum committee
This year the committee was tasked with collecting and analyzing our past enrollment and GSI allocations to determine appropriate guidelines for future allocation, especially in the presence of huge growth in our undergraduate major. I also spearheaded our effort to support faculty in introducing new undergraduate course offerings, resulting in two new courses being proposed for the 2012-2013 academic year (one of which was proposed by myself). Both courses have been approved by COCI and are now part of our standard offering.

Center for Computational Biology (CCB) Roles

- 2015-present Executive Committee
This committee makes strategic decisions for CCB and advises the Director of the CCB (faculty member equivalent to the Chair) in the running of the center
- 2019-present Program Chair & Head Graduate Advisor
Oversee both the PhD and Designated Emphasis (PhD minor) graduate programs that are run through CCB. This involved biweekly meetings with the staff and Director of CCB, as well as advising and approving graduate student curriculum affairs
- 2018-present Curriculum Committee
This committee oversees decisions regarding curriculum for the graduate programs offered by CCB
- 2016-2017 (Chair) PhD Admissions Committee
2015-2016 *The main duties of this committee are the PhD Admissions*
- 2016-2017 PhD Program Advisor

ADVISING

Post-Doctoral Students

- Nelle Veroquaux (Summer 2016 – Spring 2019)
Department of Statistics and BIDS Fellow
Project: Integrative analysis of epigenetic ‘omics data in Sorghum (part of the EPICON project)
- Sean Ruddy (Fall 2015 – Fall 2016)
Department of Statistics
Project: Normalization of mRNA-Seq data
- Gergana Bounova (Spring 2011 – Fall 2011)
(co-advised with Terry Speed and Sandrine Dudoit)
Department of Statistics, UC Berkeley
Project: Analysis of data produced from sequenced sRNAs in Tetrahymena (data from collaboration with Kathleen Collins Laboratory)

PhD Students

- Boying Gong (Fall 2017-present)
Department of Biostatistics, UC Berkeley
Thesis Title: Statistical Analysis and Integration of Multi-modal Sequencing Data
- Stephanie DeGraaf (Spring 2016 – Spring 2020)
Department of Statistics, UC Berkeley
Thesis Title: Time-Course Analysis and Clustering of Gene Expression Data Time-Course Analysis and Clustering of Gene Expression Data
- Marla Johnson (2010 – 2017)
Division of Biostatistics, UC Berkeley
Thesis Title: Clustering of mRNA-Seq Data for Detection of Alternative Splicing Patterns
- Sean Ruddy (2010 – 2014)
Department of Statistics, UC Berkeley
Thesis Title: Shrinkage of dispersion parameters in the double exponential family of distributions with applications to genomic sequencing
- Christine Ho (Summer 2010 – Fall 2016)
(co-advised with Haiyan Huang)
Department of Statistics, UC Berkeley
Thesis Title: Statistical modeling and analysis for biomedical applications.

MA Students

- Hao Wang (Fall 2020 – present)
Biostatistics, UC Berkeley
Project: Analysis of patient-level single-cell sequencing data
- Namita Trikannad (Summer 2019 – Summer 2020)
Biostatistics, UC Berkeley
Project: Analysis of gene expression time course data for drought effects
- Boying Gong (2016-2017)
Department of Statistics, UC Berkeley
Project: Analysis of differential methylation in plants using bisulfite sequencing data.

- Jonathan Ackerman (Fall 2016)
Department of Statistics, UC Berkeley
Project: Evaluation of Normalization of microbiome data

Undergraduate and Other Advising

- Mentor as part of the Regents' & Chancellor's Mentorship Program (2019-present):
Nevin Richards, Jiwon Shin, Catey Vera, Maddy Wang
- Zhe Chen (Summer 2019)
Visiting Undergraduate from Southern University of Science and Technology, China
Project: Comparison of doublet-finding methods
- Liam Purvis (Summer 2016)
Undergraduate, UC Berkeley
Project: Writing Shiny app for `clusterExperiment` package
- Tae-Kyung (Bryan) Jung (Independent Study, Spring 2010)
Undergraduate, UC Berkeley
Project: “Estimating the Number of Medical Data Breaches”
- Eric Chi (Summer 2009)
Visiting student, LBL
PhD student, Rice University

SERVICE ON DISSERTATION AND EXAMINATION COMMITTEES

PhD Dissertation Committee Service

- Boying Gong, Biostatistics, UC Berkeley (Chair)
- Stephanie DeGraaf, UC Berkeley (Chair)
- Hector Roux de Bezieux, Biostatistics, UC Berkeley
- Rebecca Sarto Basso, Industrial Eng & Operation Research, UC Berkeley
- Magali Jaillard Dancette, University of Lyon
- Amy Ko, Computational Biology, UC Berkeley
- Kelly Street, Biostatistics, UC Berkeley
- Fanny Perraudau, Biostatistics, UC Berkeley
- Marla Johnson, Biostatistics, UC Berkeley (Chair)
- Taryn M. Sumabat, Molecular and Cell Biology, UC Berkeley
- Marina P. Volegova, Molecular and Cell Biology, UC Berkeley
- Christine Ho, Statistics, UC Berkeley (Co-Chair)
- Sean Ruddy, Statistics, UC Berkeley (Chair)

PhD Qualifying Examination Committee Service

- Jinyan Zhao, Civil & Environmental Engineering (May 11, 2021)

- Yun Zhou, Statistics, UC Berkeley (May 8, 2019)
- Stephanie DeGraaf, Statistics, UC Berkeley (November 14, 2018)
- Boying Gong, Biostatistics, UC Berkeley (October 30, 2018)
- Shinya Iguchi, MCB, UC Berkeley (May 10, 2017)
- Jonathan R. Fischer, Statistics, UC Berkeley (December 2, 2016)
- David M. DeTomaso, Computational Biology, UC Berkeley (October 2016)
- Kelly Street, Biostatistics, UC Berkeley (September 12, 2016)
- Fanny Perraudou, Biostatistics, UC Berkeley (May 5, 2016)
- Amy Ko, Computational Biology, UC Berkeley (June 17, 2015)
- Marla Johnson, Biostatistics, UC Berkeley (December 11, 2014)
- Diya Das, Molecular and Cell Biology, UC Berkeley (May 16, 2014)
- Marcus Stoiber, Biostatistics, UC Berkeley (March 17, 2014)
- Angie Zhu, Statistics, UC Berkeley (December 17, 2013)
Committee Chair
- Christine Ho, Statistics, UC Berkeley (May 10, 2012)
- Sean Ruddy, Statistics, UC Berkeley (April 20, 2012)
- Tessa Lorraine Childers-Day, Statistics, UC Berkeley (November 19, 2010)
Committee Chair
- Daisy (Yan) Huang, Statistics, UC Berkeley (January 20, 2010).
- Cagla Meral, Civil and Environmental Engineering, School of Engineering, UC Berkeley (January 15, 2010)

Master's Thesis Committee Service

- Hao Wang, Spring 2021
- Aidan McLoughlin, Spring 2021
- Philippe Boileau, Spring 2020
- Hector Roux de Bezieux, Fall 2018
- Steven Pollack, Biostatistics, Spring 2014
- Nathan P. Boley, Biostatistics, Fall 2013
- Danielle Champney, Statistics, Spring 2010

TEACHING

Spring 2010 Primary Instructor
 - Present *Department of Statistics, UC, Berkeley, CA*

Courses:

Introduction to Statistics at an Advanced Level (STAT 201B)
Fall 2020

Statistical Methods for Data Science (STAT 28 / STAT 131A)
Fall 2019, Spring 2017

Statistical Consulting (STAT 272)
Fall 2018, Fall 2016, Fall 2015, Fall 2013, Fall 2011

Survey Sampling (STAT 152)
Spring 2019, Spring 2014, Spring 2011

Design and Analysis of Experiments (STAT 158)
Spring 2016, Spring 2015, Fall 2013, Spring 2013

Linear Models (STAT 151A)
Spring 2014

Applied Statistics (STAT 215B)
Spring 2021, Spring 2020, Spring 2011, Spring 2010

Fall 2003 Co-Instructor
Department of Statistics, Stanford University, Stanford, CA

Biostatistics course. Prepared and presented a third of the classroom lectures, developed homework and exam questions, in addition to duties as head Teaching Assistant for the course (e.g. grading, leading labs, office hours, coordinating TAs).

04/2001 – Teaching Assistant
05/2006 *Department of Statistics, Stanford University, Stanford, CA*

Taught sections for students (10-20 per section), created and ran R computing labs, led review sessions prior to exams, held office hours for questions about course material, graded assignments.

Courses:

- Introduction to Statistical Methods
- Biostatistics
- Consulting
- Survival Analysis
- Multivariate Statistics
- Linear Models

OUTSIDE CONSULTING

01/2019 -- Independent Statistical Consultant to *Blue Star Genomics*
present

10/12/2010 Seminar to R&D department of *Veracyte, Inc*

04/2003 – Independent Statistical Consultant

05/2006 Provided data analysis for various clients by reviewing and implementing appropriate statistical analyses for the data provided.

04/2003 – Consultant

05/2006 *Statistics Department Consulting Service, Stanford University, Stanford, CA*

Served as a statistical consultant for the drop-in service provided by the Stanford Statistics Department and at times served as the consultant in charge of running the program for the quarter.