Contact Department of Biostatistics and Computational Biology

Information Dana-Farber Cancer Institute twitter: @stephaniehicks

> 450 Brookline Avenue email: shicks@jimmy.harvard.edu Boston, MA 02215 USA website: stephaniehicks.com

RESEARCH Interests

My research interests are focused on developing statistical methods, tools and software for the analysis of genomics data. Specifically, my research addresses statistical challenges in epigenomics, functional genomics and single-cell genomics such as the pre-processing, normalization, analysis of raw highthroughput data (microarray and next-generation sequencing) leading to an improved quantification and understanding of biological variability. This work resulted in a K99/R00 grant from the National Human Genome Research Institute (NHGRI).

Current APPOINTMENT

## Postdoctoral Research Fellow, Boston MA

2013 - Present

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute (DFCI) Department of Biostatistics, Harvard T.H. Chan School of Public Health (HSPH)

Mentor: Rafael Irizarry, Ph.D. (Biostatistics - DFCI, HSPH)

EDUCATION

2013 PhD., Statistics, Rice University, Houston, TX USA

Doctoral Advisors: Marek Kimmel, Ph.D. (Statistics - Rice) and

Sharon Plon, M.D., Ph.D. (Baylor College of Medicine, TX Children's Hospital)

Dissertation: Probabilistic Models for Genetic and Genomic Data with Missing Information

M.A., Statistics, Rice University, Houston, TX USA 2012

2007 B.S., Mathematics, Louisiana State University, Baton Rouge, LA USA magna cum laude, Phi Beta Kappa

Professional EXPERIENCE

01/2017 -Postdoctoral Research Fellow, DFCI and HSPH, Boston, MA

> Mentors: Rafael Irizarry, Ph.D. (Biostatistics - DFCI, HSPH) Franziska Michor, Ph.D. (Comp Bio - DFCI, HSPH, Broad) Bradley Bernstein, M.D., Ph.D (Pathology - HMS, Mass Gen Hospital, Broad)

08/2013 -Postdoctoral Research Fellow, DFCI and HSPH, Boston, MA

12/2016 Mentor: Rafael Irizarry

06/2013Research Staff, Rice University, Houston, TX

Graduate Student Researcher, Rice University, Houston, TX 2007 - 2013

Advisors: Marek Kimmel and Sharon Plon

2006 Undergraduate Researcher (REU), Statistics, University of Wisconsin, Madison, WI

Mentors: Murray Clayton and Jo Handelsman

2005 Undergraduate Researcher (REU), Statistics, Rice University, Houston, TX

Mentor: Javier Rojo

Undergraduate Researcher, Psychology, LSU, Baton Rouge, LA 2005 - 2007

Mentor: Robert Mathews

2005 - 2007 Undergraduate Tutor, Roadmap 2 Redesign Program, LSU, Baton Rouge, LA

Tutored college algebra, trigonometry and pre-calculus students.

FELLOWSHIPS	2017 - 2021	NIH/NHGRI K99/R00 Pathway to Independence Award - Statistical
		Methods for the Normalization and Quantification of Single-Cell
		RNA-Sequencing Data, Dana-Farber Cancer Institute / HSPH
	2009 - 2011	NIH/NCI T32 Predoctoral Training Program in Biostatistics, Rice University
	2005 - 2007	LA-STEM Research Scholars, LSU
	2004 - 2005	HMMI Professors Program, LSU
Honors and	2017	Travel award for the Ascona Workshop 2017 on
Awards		Statistical Challenges in Single-Cell Biology in Ascona, Switzerland
	2016	Travel award for the Women in Statistics and Data Science Conference 2016 in NC, USA
	2015	Stellar Abstract Award for the 2015 PQG Conference in Boston, MA, USA
	2015	Travel award for the Genome Informatics Meeting 2015 at CSHL
	2014	Travel award for the Women in Statistics Conference 2014 in NC, USA
	2011	Travel and tuition award for the 16th Annual Summer Institute in Statistical Genetics
		at University of Washington (Population Genetics, Coalescent Theory modules)
	2007	Inducted into Phi Beta Kappa
	2007	LSU Austin Chapter Scholarship Award
	2003 - 2007	TOPS Tuition Award, LSU

NIII /NIICDI 1/00 /D00 D-+1----- + - I-- 1-- - - - 1-- - - - - 1-- - - - 1

# RESEARCH Pre-prints

GRANTS AND

- [1] **Hicks SC**, Irizarry RA. (2016). A Guide to Teaching Data Science. *arXiv* and under review at *The American Statistician*. https://arxiv.org/abs/1612.07140.
- [2] Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. (2016). Smooth Quantile Normalization. bioRxiv and under review at Biostatistics. doi: http://dx.doi.org/ 10.1101/085175.
- [3] **Hicks SC**, Teng M, Irizarry RA. (2015). On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data. *bioRxiv*. doi: http://dx.doi.org/10.1101/025528.

## Peer-reviewed Journal Articles

- [1] J Saliba, R Zabriskie, R Ghosh, BC Powell, S Hicks, M Kimmel, Q Meng, DI Ritter, DA Wheeler, RA Gibbs, FTF Tsai, SE Plon. (2016). Pharmacogenetic characterization of naturally occurring germline NT5C1A variants to chemotherapeutic nucleoside analogs. *Pharmacogenetics and Genomics* 26: 271-279. PMID: 26906009. PMCID: PMC4853247.
- [2] Hicks SC<sup>1</sup>, Irizarry RA. (2015). quantro: a data-driven approach to guide the choice of an appropriate normalization method. Genome Biol 16:117. PMID: 26040460. PMCID: PMC4495646.
- [3] Osman AA, Neskey DM, Katsonis P, Patel AA, Ward AM, Hsu TK, Hicks SC, McDonald TO, Ow TJ, Alves MO, Pickering CR, Skinner HD, Zhao M, Sturgis EM, Kies MS, El-Naggar A, Perrone F, Licitra L, Bossi P, Kimmel M, Frederick MJ, Lichtarge O, Myers JN. (2015). Evolutionary Action Score of TP53 Coding Variants Is Predictive of Platinum Response in Head and Neck Cancer Patients. Cancer Res 75: 1-11. PMID: 25691460
- [4] Neskey DM, Osman AA, Ow TJ, Katsonis P, McDonald T, **Hicks SC**, Hsu TK, Pickering CR, Ward A, Patel A, Yordy JS, Skinner HD, Giri U, Sano D, Story MD, Beadle BM, El-Naggar AK, Kies MS, William WN, Caulin C, Frederick M, Kimmel M, Myers JN, Lichtarge O. (2015). Evolutionary Action score of TP53 (EAp53) identifies high risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Res* **75**: 1527-1536. PMID: 25634208.
- [5] Berger RL, Li LT, Hicks SC, Liang MK. (2014). Suture versus preperitoneal polypropylene mesh for elective umbilical hernia repairs. J Surg Res 192: 426-431. PMID: 24980854

<sup>&</sup>lt;sup>1</sup> 'Highly Accessed' on BioMed Central

- [6] Brahmbhatt R, Carter SA, Hicks SC, Berger DH, Liang MK. (2014). Identifying Risk Factors for Surgical Site Complications after Laparoscopic Ventral Hernia Repair: Evaluation of the Ventral Hernia Working Group Grading System. Surg Infect (Larchmt) 15: 187-193. PMID: 24773169
- [7] Carter SA, Hicks SC, Brahmbhatt R, Liang MK. (2014). Recurrence and Pseudorecurrence after Laparoscopic Ventral Hernia Repair: Predictors and Patient-focused Outcomes. Am Surg 80: 138-148. PMID: 24480213
- [8] Hicks SC. (2014). When Women in Statistics Come to Know Their Power. Chance 27.
- [9] Li LT, Hicks SC, Davila JA, Kao LS, Berger RL, Arita NA, Liang MK. (2014). Circular Closure is Associated with the Lowest Rate of Surgical Site Infection Following Stoma Reversal: A Systematic Review and Multiple Treatment Meta-analysis. Colorectal Dis 16: 406-416. PMID: 24422861
- [10] Li LT, Brahmbhatt R, Hicks SC, Davila JA, Berger DH, Liang MK. (2014). Prevalence of Surgical Site Infection at the Stoma Site following Four Skin Closure Techniques: A Retrospective Cohort Study. Dig Surg 31: 73-78. PMID: 24776653
- [11] Li LT, Jafrani RJ, Becker NS, Berger RL, Hicks SC, Davila JA, Liang MK. (2014). Outcomes of acute versus elective primary ventral hernia repair. J Trauma Acute Care Surg 76: 523-528. PMID: 24458061
- [12] Liang MK, Li LT, Nguyen MT, Berger RL, Hicks SC, Kao LS. (2014). Abdominal reoperation and mesh explantation following open ventral hernia repair with mesh. Am J Surg 204: 670-676. PMID: 25241955.
- [13] Liang MK, Berger RL, Nguyen MT, Hicks SC, Li LT, Leong M. (2014). Outcomes with Porcine Acellular Dermal Matrix versus Synthetic Mesh and Suture in Complicated Open Ventral Hernia Repair. Surg Infect 15: 506-512. PMID: 25215466.
- [14] Nguyen MT, Berger RL, Hicks SC, Davila JA, Li LT, Kao LS, Liang MK. (2014). Comparison of Outcomes of Synthetic Mesh vs Suture Repair of Elective Primary Ventral Herniorrhaphy: A Systematic Review and Meta-analysis. JAMA Surg 31: 73-78. PMID: 24554114
- [15] Nguyen MT, Phatak UR, Li LT, Hicks SC, Moffett JM, Arita NA, Berger RL, Kao LS, Liang MK. (2014). Review of stoma site and midline incisional hernias after stoma reversal. J Surg Res 190: 504-509. PMID: 24560428
- [16] Berger RL, Li LT, Hicks SC, Davila JA, Kao LS, Liang MK. (2013). Development and validation of a risk-stratification score for surgical site occurrence and surgical site infection after open ventral hernia repair. J Am Coll Surg. 217: 974-982. PMID: 24051068
- [17] Clapp ML, Hicks SC, Awad SS, Liang MK. (2013). Trans-cutaneous Closure of Central Defects (TCCD) in Laparoscopic Ventral Hernia Repairs (LVHR). World J Surg. 37: 42-51. PMID: 23052806.
- [18] Liang MK, Clapp M, Li LT, Berger RL, Hicks SC, Awad S. (2013). Patient Satisfaction, Chronic Pain, and Functional Status following Laparoscopic Ventral Hernia Repair. World J Surg. 37: 530-537. PMID: 23212794.
- [19] Liang MK, Li LT, Avellaneda A, Moffett JM, Hicks SC, Awad SS. (2013). Outcomes and Predictors of Incisional Surgical Site Infection in Stoma Reversal. JAMA Surg 148: 183-189. PMID: 23426597.
- [20] Liang MK, Berger RL, Li LT, Davila JA, Hicks SC, Kao LS. (2013). Outcomes of laparoscopic vs open repair of primary ventral hernias. JAMA Surg 148: 1043-1048. PMID: 24005537
- [21] Neskey DM, Klein JD, Hicks S, Garden AS, Bell DM, El-Naggar Ak, Kies MS, Weber RS, Kupferman ME. (2013). Prognostic Factors Associated With Decreased Survival in Patients With Acinic Cell Carcinoma. JAMA Otolaryngol Head Neck Surg 139: 1195-1202. PMID: 24076756

- [22] Nguyen MT, Li LT, Hicks SC, Davila JA, Suliburk JW, Leong M, Kao LS, Berger DH, Liang MK. (2013). Readmission following open ventral hernia repair: incidence, indications, and predictors. Am J Surg 206: 942-948. PMID: 24296099
- [23] Subramanian A, Clapp ML, Hicks SC, Awad SS, Liang MK. (2013). Laparoscopic ventral hernia repair: Primary versus secondary hernias. J Surg Res 181: e1-5. PMID: 22795342.
- [24] Hicks S, Plon SE, Kimmel M. (2013). Statistical Analysis of Missense Mutation Classifiers. Hum Mut 34: 405-406. PMID: 23086893.
- [25] Cheung HC<sup>2</sup>, San Lucas FA, Hicks S, Chang K, Bertuch AA, Ribes-Zamora A. (2012). An S/T-Q cluster domain census unveils new putative targets under Tel1/Mec1 control. BMC Genomics 13: 664. PMID: 23176708.
- [26] Hicks S<sup>3</sup>, Wheeler DE, Plon SE, Kimmel M. (2011). Prediction of Missense Mutation Functionality Depends on both the Algorithm and Sequence Alignment Employed. *Hum Mut.* 32: 661-668. PMID: 21480434.

#### **Invited Talks**

- [1] Batch effects and technical biases in scRNA-Seq data. HSCI Single-Cell Workshop. 2016 Nov 29-30. Harvard Medical School, Boston, MA, USA.
- [2] Towards progress in batch effects and biases in single-cell RNA-seq data. *Single-Cell Genomics Conference*. 2016 Sept 14-16. Wellcome Genome Campus, Hinxton, Cambridge, UK.
- [3] On the widespread and critical impact of systemic bias and batch effects in single-cell RNA-seq data. *Joint Statistical Meetings*. 2016 July 31-Aug 4. Chicago, IL, USA.
- [4] On the widespread and critical impact of systemic bias and batch effects in single-cell RNA-seq data. Boston Single-Cell Network Meeting. 2016 March 15. Boston, MA, USA.
- [5] Introduction to R and R Package Demonstration. Dept of Biostatistics and Computational Biology Training Sessions. Dana-Farber Cancer Institute. 2016 March 9. Boston, MA, USA.
- [6] Why Statistics Matters in the Analysis of Genomics Data. Computational Biology Seminar, Department of Biology, Louisiana State University. 2015 Feb 11. Baton Rouge, LA, USA.
- [7] Normalization of DNA methylation and Gene Expression Data in the Context of Global Variation. Bioinformatics Meeting, Division of Immunology, Harvard Medical School. 2014 Sept 18. Boston, MA, USA.

#### Contributed Talks

- [1] **Hicks SC**. Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments. Ascona Workshop 2017 on Statistical Challenges in Single-Cell Biology. 2017 Apr 30-May 5. Ascona, Switzerland.
- [2] Hicks SC. Transforming the Classroom to Teach Statistics and Data Science with Active Learning. Women in Statistics and Data Science Conference. 2016 Oct 20-22. Charlotte, NC, USA.
- [3] **Hicks SC**<sup>4</sup>, Teng M, Irizarry R. On the widespread and critical impact of batch effects and systematic bias in single-cell RNA-Seq data. PQG Conference: Single-cell Genomics: Technology, Analysis, and Applications. 2015 Nov 5-6. Boston, MA, USA.
- [4] **Hicks SC**, Irizarry R. quantro: When should you use quantile normalization? Flashlight talk at *Bioconductor Conference*. 2014 Jul 30-Aug 1. Boston, MA, USA.

<sup>&</sup>lt;sup>2</sup> 'Highly Accessed' on BioMed Central

<sup>&</sup>lt;sup>3</sup>Cited and Discussed in Nature **482**: 257-262. 09 Feb 2012. PMID: 22318607

<sup>&</sup>lt;sup>4</sup>Selected for a Stellar Abstract Award

- [5] **Hicks S**, Plon SE, Kimmel M. Modeling Discovery Of Functional SNPs From Genome Scale Data. *Joint Statistical Meetings*. 2011 Aug 5. Miami, FL, USA.
- [6] Hicks S, Plon SE, Wheeler DE, Kimmel M. Prediction of Missense Mutation Functionality Depends on both the Algorithm and Sequence Alignment Employed. Human Genome Variation Society's Exploring the Functional Consequences of Genomic Variation Meeting. 2010 Nov 1. Washington, D.C., USA.

## Conference Proceedings

- [1] Saliba J, Trevino LR, Meng Q, Zabriskie R, Powell B, **Hicks S**, Kimmel M, Cheung H, Muzny DM, Reid JG, Wheeler D, Gibbs RA, and Plon SE. Abstract 5113: Functional analysis of genomic variants identified through whole exome sequencing for susceptibility to lymphocytic leukemia. *AACR 103rd Annual Meeting*. 2012 Mar 31-Apr 4. Chicago, IL.
- [2] Saliba J, Zabriskie R, Powell B, **Hicks S**, Kimmel M, Cheung H, Ritter D, Muzny DM, Reid JG, Wheeler DA, Gibbs RA, Plon SE. Abstract A8: Functional analysis of genomic variants identified through whole exome sequencing of pediatric lymphocytic leukemia kindreds. *AACR Special Conference: Pediatric Cancer at the Crossroads: Translating Discovery into Improved Outcomes.* 2013 Nov 3-6. San Diego, CA.

### **Selected Poster Presentations**

- [1] **Hicks S**, Teng M, Irizarry R. On the widespread and critical impact of batch effects and systematic bias in single-cell RNA-Seq data. Poster presented at 2015 Genome Informatics. 2015 Oct 28-31. Cold Spring Harbor Laboratories.
- [2] **Hicks S**, Irizarry R. When to use Quantile Normalization? Poster presented at: 2014 Women in Statistics Conference. 2014 May 15-17. Cary, NC.
- [3] **Hicks S**, Irizarry R. Estimating cell composition of whole blood using differentially methylated regions. Poster presented at: 2013 PQG Conference. 2013. Nov 14-15. Boston, MA.
- [4] **Hicks S**, Plon SE, Kimmel M. postMUT: A Statistical Tool for Combining Predictions of Missense Mutation Functionality using Capture-Recapture Methods. Poster presented at: 63nd Annual Meeting of The American Society of Human Genetics. 2013. Oct 22-26. Boston, MA.
- [5] Kimmel M, **Hicks S**, Plon SE. Applications of Hidden Markov Models with Conditional Emission Probabilities to Identify Regions of Identity-By-Descent in Whole-Exome Sequencing Data. Poster presented at: 63nd Annual Meeting of The American Society of Human Genetics. 2013. Oct 22-26. Boston, MA.
- [6] **Hicks S**, Plon SE, Kimmel M. Capture-Recapture Models for Evaluation of Algorithms Estimating Functionality of Missense Mutations. Poster presented at: 62nd Annual Meeting of The American Society of Human Genetics. 2012 Nov 6-12. San Francisco, CA.
- [7] **Hicks S**, Plon SE, Kimmel M. Bernoulli mixture models in application of the evaluation of algorithms estimating functionality of missense mutations. Poster presented at: *Beyond The Genome*. 2012 Sept 27-29. Boston, MA.
- [8] **Hicks S**, Plon SE, Kimmel M. Bernoulli Mixture Models in Application of the Evaulation of Algorithms Estimating Functionality of Missense Mutations. Poster presented at: *International Conference in Stochastic Processes.* 2012. Aug 21-25. Houston, TX.
- [9] **Hicks S**, Plon SE, Kimmel M. Using a Second-order Hidden Markov Model to Identify Regions of Identity-By-Descent in Exome Sequencing Data. Poster presented at: 12th International Congress of Human Genetics/61st Annual Meeting of The American Society of Human Genetics. 2011 Nov 12-15. Montreal, Canada.
- [10] **Hicks S**, Plon SE, Wheeler DE, Kimmel, M. Prediction of Missense Mutation Functionality Depends on both the Algorithm and Sequence Alignment Employed. Poster presented at: 60th Annual Meeting of the American Society of Human Genetics. 2010 Nov 2-6. Washington, D.C.

- [11] **Hicks S**, Plon SE, Wheeler DE, Kimmel, M. Prediction of Missense Mutation Functionality Depends on both the Algorithm and Sequence Alignment Employed. Poster presented at: 4th Annual Meeting of the Genomics of Common Diseases. 2010 Oct 6-9. Houston, TX.
- [12] Cheung HC, San Lucas F, Hicks S, Chang K, Plon SE, Bertuch AA, Ribes-Zamora A. S/T-Q Cluster Domains in Saccharomyces cerevisiae: a bioinformatic census and analysis. Poster presented at: BioScience Research Collaborative Grand Opening. 2010 Apr 1. Houston, TX.
- [13] **Hicks S**, Plon SE, Wheeler DE, Kimmel, M. Predicting Functionality of Missense Mutations with Varying Levels of Evolutionary Depth. Poster presented at: *Computational & Theoretical Biology Symposium*. 2009 Dec 4-6. Houston, TX.
- [14] **Hicks S**, Plon SE, Wheeler DE, Kimmel, M. Predicting Functionality of Missense Mutations with Varying Levels of Evolutionary Depth. Poster presented at: 59th Annual Meeting of the American Society of Human Genetics. 2009 Oct 20-24. Honolulu, HI.

#### Ph.D. Dissertation

[1] **Hicks SC**. Probabilistic Models for Genetic and Genomic Data with Missing Information, Ph.D. Thesis, Rice University, 2013.

Teaching
Experience

- 2016 **Co-instructor**, Introduction to Data Science (HSPH BIO 260) Co-taught lectures with Rafael Irizarry, developed course material
- 2015 **Teaching Assistant**, Statistical Methods for Functional Genomics (CSHL) Contributed course material, gave a lecture, answered questions.
- 2015 Contributor and Discussion Leader
  - (1) Statistics and R for the Life Sciences (Harvard edX PH525.1x)
  - (2) Case study: DNA methylation data analysis (Harvard edX PH525.8x)

Contributed course material, created screencasts, developed assessments, answered questions. Courses taught by Rafael Irizarry with Teaching Assistant Mike Love.

2014 Lead Teaching Fellow, Data Science (Harvard - CS109)

Held weekly lab sections. Developed course material, homework and solutions. Coordinated and directed 25 teaching fellows. All material available on: Github Course taught by Rafael Irizarry and Verena Kaynig-Fittkau.

- 2014 **Discussion Leader**, Data Analysis for Genomics (Harvard edX PH525x) Course taught by Rafael Irizarry with Teaching Assistant Mike Love.
- 2010, 2011 **Teaching Assistant**, Probability in Bioinformatics and Genetics (Rice STAT 423/623) Grade homeworks, provide solutions, hold office hours and gave several guest lectures.
- 2010 **Teaching Assistant**, Applied Stochastic Processes (Rice STAT 552)
  Grade homeworks, provide solutions, hold office hours and gave several guest lectures.
- 2008, 2009 **Lab Instructor**, Introduction to Statistics for the Biosciences (Rice STAT 305) Hold labs 3hr/week to teach students to use R.
- 2007, 2008 **Teaching Assistant**, Probability and Statistics (Rice STAT 310)
  Grade homeworks, provide solutions, hold office hours and gave several guest lectures.

#### SOFTWARE Bioconductor and Github

[1] qsmooth: An R-package that implements a generalization of quantile normalization, referred to as smooth quantile normalization (qsmooth), which is based on the assumption that the

- statistical distribution of each sample should be the same (or have the same distributional shape) within biological groups or conditions [Available on GitHub]
- [2] explainr: An R-package to translate S3 objects into text using standard templates in a simple and convenient way. Developed at ROpenSci Unconference with Hilary Parker, David Robinson and Roger Peng. [Available on GitHub]
- [3] quantro: An R-package that can be used to test for differences between groups of distributions to guide the choice if quantile normalization should be used [Available on Bioconductor]
- [4] quantroSim: A supporting data simulation R-package for the quantro R-package to simulate gene expression and DNA methylation data [Available on GitHub]
- [5] postMUT: A tool implemented in Perl and R to predict the functionality of missense mutations [Available on GitHub]

Community
Coding
EXPERIENCE

Sept 2015 Mozilla Open Science Leadership Summit (website), Mozilla, Toronto, Canada Invited to work with community leaders furthering open practice and open science through creating community events, tools for collaboration and learning resources.

Mar 2015 **ROpenSci Unconference** (website), GitHub, San Francisco, CA Invited to work with over 40 R enthusiasts from industry, academia, non-profits and government on projects supporting open data, open science and data visualization in R.

SERVICE 2017 - 2019 Member of ASA Committee on Women in Statistics
2015 - Present Board Member of Cards Against Humanities Science Ambassador Scholarship

TECHNICAL Programming: R, Python, Git, Perl, LaTeX, SAS, Matlab, Shell scripting Skills Operating Systems: Mac OS X, Unix, Windows

JOURNAL Bioinformatics, Biometrics, BMC Medical Genetics, Genome Biology, Human Mutation, PLoS Computational Biology, PLoS One
REVIEWER

PROFESSIONAL Member, American Statistical Association (ASA)
MEMBERSHIPS Member, American Mathematical Society (AMS)
Member, American Society of Human Genetics (ASHG)