BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Hicks, Stephanie C.

eRA COMMONS USER NAME (credential, e.g., agency login): schicks

POSITION TITLE: Assistant Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Louisiana State University, Baton Rouge, LA	BS	05/2007	Mathematics
Rice University, Houston, TX	MA	12/2011	Statistics
Rice University, Houston, TX	PhD	05/2013	Statistics
Dana-Farber Cancer Institute, Boston, MA Harvard T.H. Chan School of Public, Boston, MA	Postdoctoral	01/2018	Biostatistics and Computational Biology

A. Personal Statement

I am an applied statistician interested in developing statistical methodology, tools and software for biomedical data analysis**,** which often contains noisy or missing data and systematic biases. Specifically, my research addresses statistical challenges in epigenomics, functional genomics and single-cell genomics such as the pre-processing, normalization, analysis of raw, noisy high-throughput data (microarray and next-generation sequencing) leading to an improved quantification and understanding of biological variability.

I am an expert in developing fast, memory-efficient statistical methods and software for the analysis of single-cell RNA-Sequencing (scRNA-seq) data. I have a large amount of experience examining publicly available scRNA-seq data and identifying systematic biases that can explain a substantial amount of observed cell-to-cell expression variability (1). More recently, I developed a dimensionality reduction method, Varying-censoring Aware Matrix Factorization (VAMF), which permits the identification of low-dimensional representations of cells in the presence of cell-specific censoring. This allows for the correction for batch effects if they are mediated through a varying censoring mechanism in either confounded or unconfounded study designs, which is not possible using standard batch correction methods (2). This work resulted in a K99/R00 grant from the National Human Genome Research Institute (NHGRI). In addition, I have developed normalization methods (3) and data-driven methods to guide the choice in normalization methods for bulk genomics datasets (4).

Four publications that highlight experience and qualifications for this project:
1.   Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2017. Epub 2017/11/10. doi: 10.1093/biostatistics/kxx053. PubMed PMID: 29121214; PMCID: PMC29121214.
2.   Townes FW, Hicks SC, Aryee MJ, Irizarry RA. Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. bioRxiv. [Preprint]. In press 2017.
3.   Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. Smooth quantile normalization. Biostatistics. 2017. doi: 10.1093/biostatistics/kxx028. PubMed PMID: 29036413; PMCID: PMC5862355.
4.   Hicks SC, Irizarry RA. quantro: a data-driven approach to guide the choice of an appropriate normalization method. Genome Biol. 2015;16:117. doi: 10.1186/s13059-015-0679-0. PubMed PMID: 26040460; PMCID: 4495646.

B. Positions and Honors

Positions and Employment
2007 – 2013   Graduate Student Researcher
              Department of Statistics, Rice University, Houston, TX
2013 – 2018   Postdoctoral Research Fellow
              Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute
              Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
2018 –   Assistant Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public
              Health, Baltimore, MD

Honors
2003 – 2007   TOPS Tuition Award, Louisiana State University
2004 – 2005   HMMI Professors Program, Louisiana State University
2005 – 2007   LA-STEM Research Scholars Program, Louisiana State University
2007   LSU-Austin Chapter Scholarship Award
2007   Phi Beta Kappa, Louisiana State University
2009 – 2011   NIH/NCI T32 Predoctoral Training Program in Biostatistics, Rice University
2011   Travel and tuition award for the 16th Annual Summer Institute in Statistical Genetics
                  at University of Washington
2014   Travel award for the Women in Statistics Conference 2014
2015   Travel award for the Genome Informatics Meeting 2015
2015   Stellar Abstract Award for the Program in Quantitative Genomics Conference 2015
2016   Travel award for the Women in Statistics and Data Science Conference 2016
2016 – 2020   NIH/NHGRI K99/R00 Pathway to Independence Award
2017   Travel award for the Ascona Workshop 2017 on Statistical Challenges in Single-Cell Biology
2017   Travel award from the Caucus for Women in Statistics to attend the Joint Statistical Meetings

Professional Societies and Public Advisory Committees
2005-Present Member, American Statistical Association
2007-Present Member, American Mathematical Society
2015-Present Board Member, Cards Against Humanity and SMBC Science Ambassador Scholarship
2017-2019   Member, American Statistical Association Committee on Women in Statistics

C. Contributions to Science

Google Scholar profile: https://scholar.google.com/citations?user=4T4qnL4AAAAJ
Full list of publications here: http://www.ncbi.nlm.nih.gov/pubmed/?term=Hicks+SC+or+Hicks+S

1. Statistical methods for high-throughput data
Normalization is an essential step for the analysis of genomics high-throughput data. Quantile normalization is one of the most widely used multi-sample normalization tools for applications including genotyping arrays, RNA-Sequencing (RNA-Seq), DNA methylation, ChIP-Sequencing (ChIP-Seq) and brain imaging. However, quantile normalization relies on assumptions about the data-generation process that are not appropriate in some contexts. I developed a data-driven method to test these assumptions and guide the choice of an appropriate normalization method (1). Our freely available software (2) has been downloaded over 3700 times (distinct IPs) from Bioconductor since 2014 and has helped researchers test the assumptions of global normalization methods in the analysis of their own data**.** To address the scenario when the assumptions of quantile normalization are not appropriate, I have developed a generalization of quantile normalization, referred to as smooth quantile normalization, which allows for global differences between biological groups (3). More recently, I collaborated with researchers from the University of Maryland to correct for compositional biases found in sparse metagenomic sequencing data (4).

1.   Hicks SC, Irizarry RA. quantro: a data-driven approach to guide the choice of an appropriate normalization method. Genome biology. 2015;16:117. doi: 10.1186/s13059-015-0679-0. PubMed PMID: 26040460; PMCID: 4495646.
2.   Hicks S, Irizarry R. quantro: A test for when to use quantile normalization. 2014.
3.   Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. Smooth quantile normalization. Biostatistics. 2017. doi: 10.1093/biostatistics/kxx028. PubMed PMID: 29036413; PMCID: PMC5862355.
4.   Kumar MS, Slud EV, Okrah K, Hicks SC, Hannenhalli S, Corrada Bravo H. Analysis and correction of compositional bias in sparse sequencing count data. bioRxiv. [Preprint]. In press 2017.

2. Statistical methods for single-cell RNA-Sequencing

I have begun working with single-cell RNA-Sequencing (scRNA-seq) data, which has become the most widely used high-throughput method for transcription profiling of individual cells. This technology has created an unprecedented opportunity to investigate important biological questions that can only be answered at the single-cell level. However, this technology also brings new statistical, computational and methodological challenges. In contrast to bulk RNA-seq experiments, the majority of reported expression levels in scRNA-seq data are zeros, which could be either biologically-driven, genes not expressing RNA at the time of measurement, or technically-driven, genes expressing RNA, but not at a sufficient level to be detected by sequencing technology. In addition, systematic errors, including batch effects, have been widely reported as a major challenge in high-throughput technologies, however, surprisingly, these issues have received minimal attention in published studies based on scRNA-seq technology. To investigate this, I examined data from fifteen published scRNA-seq studies and demonstrated that systematic errors can explain a substantial percentage of observed cell-to-cell expression variability, which in turn can lead to false discoveries, for example, when using unsupervised learning methods(1). More recently, I developed a dimensionality reduction method, Varying-censoring Aware Matrix Factorization (VAMF), which permits the identification of low-dimensional representations of cells in the presence of cell-specific censoring. This allows for the correction for batch effects if they are mediated through a varying censoring mechanism in either confounded or unconfounded study designs, which is not possible using standard batch correction methods (2).

1.   Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2017. Epub 2017/11/10. doi: 10.1093/biostatistics/kxx053. PubMed PMID: 29121214; PMCID: PMC29121214.
2.   Townes FW, Hicks SC, Aryee MJ, Irizarry RA. Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. bioRxiv. [Preprint]. In press 2017.

3. Assessing the functional impact of coding mutations

Understanding the functional consequences of coding mutations is important for assessing their disease-association or clinical importance. In particular, the interpretation of missense mutations (point mutations in which a single nucleotide change results in an amino acid change) has remained a difficult task because missense mutations do not necessarily impact protein function. Many computational or *in silico* algorithms have been developed to predict the impact of missense mutations on protein function, but they often lead to conflicting results leaving the researcher without guidance in how to prioritize the mutations identified for further evaluation in biological assays. As a graduate student, I investigated how functional predictions vary between using different *in silico* methods and between using different protein sequence alignment (1). This work was relevant in characterizing germline mutations (2). I highlighted the difficulty in accurately predicting the functionality of missense mutations and showed how *in silico* methods often have a high degree of disagreement, which is particularly relevant to the users of these methods. My extensive background and experience of working with these *in silico* methods led me to critically assess many other *in silico* methods. In particular, we evaluated and validated the *in silico* method evolutionary action (EAp53) as a tool to successfully stratify patients with head and neck tumors harboring TP53 mutations as high risk or low risk, which may be useful in clinical prognosis of tumors with TP53 mutations (3). Similarly, we evaluated and demonstrated the successful use of EAp53 to predict the impact of TP53 mutations on response to chemotherapy (4).

1.   Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Human mutation. 2011;32(6):661-8. doi: 10.1002/humu.21490. PubMed PMID: 21480434; PMCID: 4154965.

2.   Saliba J, Zabriskie R, Ghosh R, Powell BC, Hicks S, Kimmel M, Meng Q, Ritter DI, Wheeler DA, Gibbs RA, Tsai FT, Plon SE. Pharmacogenetic characterization of naturally occurring germline NT5C1A variants to chemotherapeutic nucleoside analogs. Pharmacogenet Genomics. 2016;26(6):271-9. PubMed PMID: 26906009; PMCID: PMC4853247.
3.   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. Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer. Cancer research. 2015;75(7):1527-36. PubMed PMID: 25634208; PMCID: PMC4383697.
4.   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. Evolutionary Action Score of TP53 Coding Variants Is Predictive of Platinum Response in Head and Neck Cancer Patients. Cancer research. 2015;75(7):1205-15. PubMed PMID: 25691460; PMCID: 4615655.

4. Statistical analyses for interdisciplinary biomedical research
As a graduate student, I collaborated with a diverse set of biomedical researchers including surgeons, geneticists, bioinformaticians and academic faculty, at both universities and hospitals who analyze different types of biomedical data. My contribution in these interdisciplinary projects was to perform the statistical analyses including statistical inference and developing statistical models such as Cox proportional hazards models (1), univariate (2) and multivariate meta-analyses (3), propensity matching and logistic regression (4).

1.   Neskey DM, Klein JD, Hicks S, Garden AS, Bell DM, El-Naggar AK, Kies MS, Weber RS, Kupferman ME. Prognostic factors associated with decreased survival in patients with acinic cell carcinoma. JAMA Otolaryngol Head Neck Surg. 2013;139(11):1195-202. doi: 10.1001/jamaoto.2013.4728. PubMed PMID: 24076756; PMCID: PMC5555308.
2.   Li LT, Hicks SC, Davila JA, Kao LS, Berger RL, Arita NA, Liang MK. 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. 2014;16(6):406-16. doi: 10.1111/codi.12556. PubMed PMID: 24422861.
3.   Nguyen MT, Berger RL, Hicks SC, Davila JA, Li LT, Kao LS, Liang MK. Comparison of outcomes of synthetic mesh vs suture repair of elective primary ventral herniorrhaphy: a systematic review and meta-analysis. JAMA Surg. 2014;149(5):415-21. doi: 10.1001/jamasurg.2013.5014. PubMed PMID: 24554114.
4.   Liang MK, Berger RL, Li LT, Davila JA, Hicks SC, Kao LS. Outcomes of laparoscopic vs open repair of primary ventral hernias. JAMA Surg. 2013;148(11):1043-8. doi: 10.1001/jamasurg.2013.3587. PubMed PMID: 24005537.

5. Using amino acid motifs to identify proteins and pathways controlled by Tel1/Mec1 kinases
Tel1/Mec1 kinases play a major role in DNA damage response (DDR) pathways yet the complete set of proteins and pathways controlled by Tel1/Mec1 remains unknown. We developed a method and pipeline to identify all proteins in humans and yeast controlled by Tel1/Mec1 kinases (1). My contribution was to analyze the distribution of S/T-Q motifs (serine (S) or threonine (T) followed by a glutamine (Q)) within verified Tel1/Mec1 targets and to define a novel criteria of at least 3 S/T-Q motifs within a stretch of 50 amino acid residues to identify new Tel1/Mec1 targets. Simply using motifs, we identified a new set of proteins that are involved in not only known DDR pathways, but also several other pathways under Tel1/Mec1 control suggesting new putative targets for these kinases.

1.   Cheung HC, San Lucas FA, Hicks S, Chang K, Bertuch AA, Ribes-Zamora A. An S/T-Q cluster domain census unveils new putative targets under Tel1/Mec1 control. BMC Genomics. 2012;13:664. doi: 10.1186/1471-2164-13-664. PubMed PMID: 23176708; PMCID: PMC3564818.

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

R00HG009007 (Hicks)   12/23/2016 – 02/28/2021
NIH/NHGRI

Statistical Methods for the Normalization and Quantification of Single-Cell RNA-Sequencing Data
To develop novel statistical methods that will 1) remove systematic errors in single-cell RNA-sequencing data by accounting for variability visible only at the single-cell level and 2) quantify biological variability such as transcriptional heterogeneity within and between populations of cells.
Role: Principal Investigator

182891(Davide Risso)   03/01/2018 – 02/28/2019
Cornell University sub, (Prime - CZI) JHSPH: PI Hicks
Fast and efficient implementation of common algorithms for large single-cell data
The purpose of this project is to (A) provide a coherent programmatic interface to the HCA and to (B) enable scalable interactive statistical analysis of single-cell data for Bioconductor users and developers. We accomplish these aims by initiating development of software to (1) access HCA data; (2) efficiently represent and manipulate HCA data within R / Bioconductor; (3) assess and ameliorate data quality; (4) adapt well-established Bioconductor infrastructure for statistical analysis of HCA; and (5) provide facilities for binding and working with HCA data through ontology bindings.
Role: Sub PI

Completed Support

N/A

Overlap

None