

CONTACT INFORMATION	Department of Computer Science Princeton University Princeton, NJ USA	website: willtownes.github.io email: ftownes@princeton.edu
SUMMARY	I am a postdoctoral researcher focused on biomedical data science. I work with Barbara Engelhardt in the Princeton Computer Science Department. My interests include unsupervised machine learning, genomics, and infectious disease. My goal is to produce statistical methods and open source software implementations to help scientists achieve their research goals more effectively.	
EDUCATION	2019 Ph.D. Biostatistics , Harvard University, Cambridge, MA USA <i>Doctoral Committee: Rafael Irizarry (Dana-Farber Cancer Institute, Data Science)</i> <i>Jeff Miller (Harvard University, Biostatistics)</i> <i>Martin Aryee (Massachusetts General Hospital, Pathology)</i>	
	2016 A.M. Biostatistics , Harvard University, Cambridge, MA USA	
	2013 M.S. Mathematics & Statistics , Georgetown University, Washington, DC USA	
	2007 B.S. Biology <i>magna cum laude</i> , Washington & Lee University, Lexington, VA USA	
PROFESSIONAL EXPERIENCE	2019 - present Postdoctoral Researcher , Princeton Computer Science Department	
	2011 - 2014 Software Test Engineer , SRA International, Arlington, VA Developed automated testing of web services and ETL processes for the Energy Star Qualified Product Exchange. Trained other testers on the use of APIs and version control.	
	2008 - 2010 Software Analyst , Perrin Quarles Associates, Charlottesville, VA Supported development of the Environmental Protection Agency's National Emissions Inventory System.	
	2007 - 2008 Fulbright Scholar , University of the Philippines Conducted independent forest ecology research while living with an indigenous <i>Aeta</i> community in Bataan Province.	
RESEARCH	Preprints	
	[1] Townes FW , Irizarry RA. Quantile normalization of single-cell RNA-seq read counts without unique molecular identifiers. <i>bioRxiv</i> (2019). https://doi.org/10.1101/817031 .	
	[2] Hecker J, Townes FW , Kachroo P, Lasky-Su J, Ziniti J, Cho MH, Weiss ST, Laird NM, Lange C. An exact, unifying framework for region-based association testing in family-based designs, including higher criticism approaches, SKATs, multivariate and burden tests. <i>bioRxiv</i> (2019). https://doi.org/10.1101/815290 .	
	Peer-reviewed Journal Articles	
	[1] Townes FW , Hicks SC, Aryee MJ, Irizarry RA. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. <i>Genome Biology</i> (2019). https://doi.org/10.1186/s13059-019-1861-6 .	
	[2] Marsh DM, Townes FW , Cotter K, Farroni K, McCreary K, Petry R, Tilghman J. Thermal Preference and Species Range in Mountaintop Salamanders and Their Widespread Competitors. <i>Journal of Herpetology</i> (2019). https://doi.org/10.1670/18-110 .	
	[3] Hicks SC, Townes FW , Teng M, Irizarry RA. Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments. <i>Biostatistics</i> (2018). http://dx.doi.org/10.1101/025528 .	

- [4] Hecker, J, Xu X, **Townes FW**, Fier HL, Corcoran C, Laird N, Lange C. Family-Based Tests for Associating Haplotypes With General Phenotype Data: Application to Asthma Genetics. *Genetic Epidemiology* (2017). <https://doi.org/10.1002/gepi.22094>.
- [5] Valeri L, Patterson-Lomba O, Gurmu Y, Abloh A, Bobb J, **Townes FW**, Harling G. Predicting Subnational Ebola Virus Disease Epidemic Dynamics from Sociodemographic Indicators. *PLoS ONE* (2016). <https://doi.org/10.1371/journal.pone.0163544>.
- [6] **Townes W**. Seed dispersal of the genus *Leea* in forest patches of Bataan, Philippines. *Ecotropica* (2010).

Technical Reports

- [1] **Townes FW**. Review of Probability Distributions for Modeling Count Data. *arXiv* (2020). <https://arxiv.org/abs/2001.04343>.
- [2] **Townes FW**. Generalized Principal Component Analysis. *arXiv* (2019). <https://arxiv.org/abs/1907.02647>.

TALKS

Invited Talks

- [1] Don't Normalize: the GLM-PCA approach to normalization. *Normjam, New York Genome Center*. November 2019.
- [2] Biostatistics and Data Science as a Career Path. *Biology Department, Washington & Lee University*. October 2019.

Contributed Talks

- [1] Dimension Reduction for Single Cell RNA-Seq based on a Multinomial Model. *Joint Statistical Meetings* (Denver, CO). August 2019.
- [2] Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. *Bioconductor Annual Meeting* (Boston, MA). July 2017.

Seminars

- [1] Feature Selection and Dimension Reduction for Single Cell RNA-Seq. *Greene Lab, Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania*. February 2019.
- [2] Unsupervised Learning for Single Cell Gene Expression. *Harvard Biostatistics Cancer Working Group*. September 2018.
- [3] Single Cell Housekeeping Genes and Normalization. *Dana-Farber Cancer Institute cBio Seminar*. April 2018.
- [4] Informative Missing Data in Single Cell RNA-Seq. *Dana-Farber Cancer Institute Genomics Seminar*. October 2016.
- [5] Family Based Association Tests for Rare Variants. *Brigham & Women's Hospital Channing Network Medicine Seminar*. September 2015.

Posters

- [1] Townes FW, Shukla C. Gene Expression Autoencoders. *Harvard Biomedical Informatics 707 (Deep Learning in Healthcare)*. April 2018.
- [2] Townes FW, Marquez-Luna C. Mixture of Experts Analysis of Infectious Disease Outbreak Characteristics. *Harvard Computer Science 281 (Advanced Machine Learning)*. December 2015.
- [3] Townes FW, Karaayvaz M, Gillespie S, Bernstein B, Ellisen L, Aryee M. Single Cell RNA-Seq Technical and Biological Confounders. *Program in Quantitative Genomics Conference, Harvard Medical School*. November 2015.

FUNDING	2018 - 2019	NIH T32 Training Grant: Cancer
	2017	Chan-Zuckerberg Foundation Travel Grant Human Cell Atlas Jamboree. European Bioinformatics Institute, Hinxton, UK
	2016 - 2018	NIH T32 Training Grant: Big Data to Knowledge
	2016	NSF-CBMS Travel Award: Topology, Geometry, and Statistics
	2014 - 2016	NIH T32 Training Grant: Clinical Epidemiology of Lung Diseases
TEACHING EXPERIENCE	2018 Fall	Teaching Assistant , Applied Regression (Harvard - BST 210) Grade homeworks, hold office hours, and teach a weekly lab section.
	2018 Summer	Co-Instructor , StatStart (Harvard) Statistics program for under-represented high school students. Gave two interactive lectures on graphing data and regression.
	2017 Fall	Teaching Assistant , Applied Bayesian Analysis (Harvard - BST 228) Grade homeworks, provide solutions, and hold office hours.
	2017 Summer	Lead Instructor , Introduction to Data Science (PARSE Ltd.) Nonprofit, week-long statistics program for high school students. Developed course material and gave interactive lectures using R.
	2017 Spring	Teaching Assistant , Applied Longitudinal Analysis (Harvard - BST 226) Grade homeworks and hold office hours.
	2016 Fall	Teaching Assistant , Intro to Statistical Methods (Harvard - BST 201) Grade homeworks, provide solutions, hold office hours, and teach a weekly lab section.
	2016 Summer	Project Mentor , Pipelines into Biostatistics (Harvard) Biostatistics program for undergraduates. Guided three students on a pharmacogenomics data analysis.
	2016 Spring	Teaching Assistant , Rates and Proportions (Harvard - BST 210) Grade homeworks, provide solutions, hold office hours, and teach a weekly lab section.
SOFTWARE PACKAGES	[1] glmpca : dimension reduction for non-normally distributed data. R package: https://cran.r-project.org/package=glmpca , python package: https://pypi.org/project/glmpca/	
TECHNICAL SKILLS	Programming : R, Python, SQL, Git, L ^A T _E X, Stata, Julia, Matlab, Shell Operating Systems : Mac OS X, Linux, Windows Machine Learning Frameworks : Caret, Sklearn, Keras, Tensorflow, Stan, JAGS	