CONTACT
INFORMATION

Department of Statistics and Data Science
Carnegie Mellon University
Pittsburgh, PA USA

website: willtownes.github.io
email: ftownes@andrew.cmu.edu

SUMMARY

I am an applied statistician working primarily with biomedical data. I am involved with infectious disease as part of the Delphi Group and have developed methods for spatial and single-cell genomics. I am also interested in wearable devices, ecology, and economics. I strive to produce open source software implementations to help scientists achieve their research goals more effectively.

PROFESSIONAL
EXPERIENCE

2022 - present	**Assistant Professor**, Department of Statistics and Data Science Carnegie Mellon University, Pittsburgh, PA
2019 - 2022	**Postdoctoral Researcher**, Department of Computer Science Princeton University, Princeton, NJ (advisor: Barbara Engelhardt)
2011 - 2014	**Software Test Engineer**, SRA International, Arlington, VA
2008 - 2010	**Software Analyst**, Perrin Quarles Associates, Charlottesville, VA
2007 - 2008	**Fulbright Scholar**, University of the Philippines Conducted independent forest ecology research while living with an indigenous *Aeta* community in Bataan Province.

EDUCATION

2019	**Ph.D. Biostatistics**, Harvard University, Cambridge, MA USA
	Dissertation Committee: Rafael Irizarry (Dana-Farber Cancer Institute, Data Science)
	Jeff Miller (Harvard University, Biostatistics)
	Martin Aryee (Massachusetts General Hospital, Pathology)
2013	**M.S. Mathematics & Statistics**, Georgetown University, Washington, DC USA
2007	**B.S. Biology**, Washington & Lee University, Lexington, VA USA

RESEARCH

In Review

[1] Jones A, **Townes FW**, Li D, Engelhardt BE. Alignment of spatial genomics and histology data using deep Gaussian processes (2022). Preprint: doi:10.1101/2022.01.10.475692.

[2] **Townes FW**, Engelhardt BE. Nonnegative spatial factorization (2021). Preprint: https://arxiv.org/abs/2110.06122.

[3] Verpeut JL, Bergeler S, Kislin M, **Townes FW**, Klibaite U, Dhanerawala ZM, Hoag A, Jung C, Lee J, Pisano TJ, Seagraves KM, Shaevitz JW, Wang SSH. Cerebellar contributions to a brainwide network for flexible behavior (2021). Preprint: doi:10.1101/2021.12.07.471685.

In Press

[1] Gewirtz ADH, **Townes FW**, and Engelhardt BE. Telescoping bimodal latent Dirichlet allocation to identify expression QTLs across tissues *Life Science Alliance* (2021). Preprint: doi:10.1101/2021.10.27.466156.

Peer-reviewed Journal Articles

[1] Jones A, **Townes FW**, Li D, Engelhardt BE. Contrastive latent variable modeling with application to case-control sequencing experiments. *Annals of Applied Statistics* (2022). doi:10.1214/21-AOAS1534.

[2] Hecker J, **Townes FW**, Kachroo P, Lasky-Su J, Ziniti J, Cho MH, Weiss ST, Laird NM, Lange C. A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests. *Bioinformatics* (2020). doi:10.1093/bioinformatics/btaa1055.

[3] **Townes FW**, Carr K, Miller JW. Identifying longevity associated genes by integrating gene expression and curated annotations. *PLOS Computational Biology* (2020). doi:10.1371/journal.pcbi.1008429.

[4] **Townes FW**, Irizarry RA. Quantile normalization of single-cell RNA-seq read counts without unique molecular identifiers. *Genome Biology* (2020). doi:10.1186/s13059-020-02078-0.

[5] **Townes FW**, Hicks SC, Aryee MJ, Irizarry RA. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biology* (2019). doi:10.1186/s13059-019-1861-6.

[6] 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. *Journal of Herpetology* (2019). doi:10.1670/18-110.

[7] Hicks SC, **Townes FW**, Teng M, Irizarry RA. Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments. *Biostatistics* (2018). doi:10.1101/025528.

[8] 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). doi:10.1002/gepi.22094.

[9] Valeri L, Patterson-Lomba O, Gurmu Y, Ablorh A, Bobb J, **Townes FW**, Harling G. Predicting Subnational Ebola Virus Disease Epidemic Dynamics from Sociodemographic Indicators. *PLOS One* (2016). doi:10.1371/journal.pone.0163544.

[10] **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**

[1] Advanced Biomedical Computation Seminar, Computational Pathology Department, Brigham and Women's Hospital. March 2022.

[2] Statistics and Data Science Department, Carnegie Mellon University. March 2022.

[3] Biomedical Data Science Department, Stanford University. February 2022.

[4] Biostatistics Department, University of Michigan. January 2022.

[5] Population and Public Health Sciences Department, University of Southern California. January 2022.

[6] Emerging Concepts in Microbial Informatics, Molecular Medicine and Biotechnology (panelist). Trinity University of Asia. November 2021.

[7] Don't Normalize: the GLM-PCA approach to normalization. *Normjam*, New York Genome Center. November 2019.

[8] Biostatistics and Data Science as a Career Path. Biology Department, Washington & Lee University. October 2019.

Contributed Talks

[1] Latent factorization methods for genomics. *New England Statistical Symposium* (virtual presenter). May 2022.

[2] Nonnegative process factorization for multivariate spatial count data. *Joint Statistical Meetings* (virtual conference). August 2021.

[3] Quantile normalization of single-cell RNA-seq read counts without unique molecular identifiers. *Joint Statistical Meetings* (virtual conference). August 2020.

[4] Dimension reduction for massive single-cell datasets. *Bioconductor Annual Meeting* (virtual conference). July 2020. slides doi:10.7490/f1000research.1118084.1

[5] Dimension Reduction for Single Cell RNA-Seq based on a Multinomial Model. *Joint Statistical Meetings* (Denver, CO). August 2019.

[6] Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. *Bioconductor Annual Meeting* (Boston, MA). July 2017.

Seminars

[1] Human tumor atlas network steering committee meeting. April 2022.

[2] Nonnegative spatial factorization. *Hansen and Hicks labs, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.* November 2021.

[3] Primer: Generalized linear models and latent factor models. *Broad Institute Models, Inference, and Algorithms.* October 2020.

[4] 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.

[5] Unsupervised Learning for Single Cell Gene Expression. *Harvard Biostatistics Cancer Working Group.* September 2018.

[6] Single Cell Housekeeping Genes and Normalization. *Dana-Farber Cancer Institute cBio Seminar.* April 2018.

[7] Informative Missing Data in Single Cell RNA-Seq. *Dana-Farber Cancer Institute Genomics Seminar.* October 2016.

[8] Family Based Association Tests for Rare Variants. *Brigham & Women's Hospital Channing Network Medicine Seminar.* September 2015.

Interviews

[1] The Bioinformatics Chat. March 2020. https://bioinformatics.chat/glm-pca

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.

PEER REVIEW **Journals**: *American Journal of Human Genetics, Bioinformatics, Biostatistics, Genome Biology, NAR Genomics and Bioinformatics, Nature Communications*

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

PROFESSIONAL SOCIETIES American Statistical Association, International Society for Computational Biology

TEACHING EXPERIENCE	2021 Fall	**Guest lecture**, Advanced Computational Genomics (Princeton - COS 597D)
	2020 Fall	**Guest lecture**, Computational Biology of Single Cells (Princeton - COS 597F)
	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)
		Program for under-represented 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/`

[2] **scry**: Small-Count Analysis Methods for High-Dimensional Data. Bioconductor R package: `https://bioconductor.org/packages/release/bioc/html/scry.html`

[3] **quminorm**: Quantile normalization of non-UMI single cell read counts. R package: `https://github.com/willtownes/quminorm`

[4] **spatial-factorization**: Spatially-aware probabilistic factor models. Python package: `https://github.com/willtownes/spatial-factorization-py`

TECHNICAL SKILLS

Programming: R, Python, SQL, Git, LaTeX, Shell
Operating Systems: MacOS, Linux, Windows
Machine Learning Frameworks: Caret, Scikit-learn, Keras, Tensorflow, Stan