

CONTACT INFORMATION	Department of Statistics and Data Science Carnegie Mellon University Pittsburgh, PA USA	website: <a href="https://willtownes.github.io">willtownes.github.io</a> email: <a href="mailto:ftownes@andrew.cmu.edu">ftownes@andrew.cmu.edu</a>
SUMMARY	I am an applied statistician working mostly in the biomedical and public health domains. As part of the Delphi infectious disease tracking and forecasting team, I focus primarily on wastewater data. I am also working on projects in the areas of genomics and spatiotemporal statistics. My side interests include wearable devices, ecology, and economics.	
PROFESSIONAL EXPERIENCE	2022 - present	<b>Assistant Professor</b> , Department of Statistics and Data Science Carnegie Mellon University, Pittsburgh, PA
	2019 - 2022	<b>Postdoctoral Researcher</b> , Department of Computer Science Princeton University, Princeton, NJ (advisor: Barbara Engelhardt)
	2011 - 2014	<b>Software Test Engineer</b> , SRA International, Arlington, VA
	2008 - 2010	<b>Software Analyst</b> , Perrin Quarles Associates, Charlottesville, VA
	2007 - 2008	<b>Fulbright Scholar</b> , University of the Philippines Conducted independent forest ecology research while living with an indigenous <i>Aeta</i> community in Bataan Province.
EDUCATION	2019	<b>Ph.D. Biostatistics</b> , Harvard University, Cambridge, MA USA <i>Dissertation 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>
	2013	<b>M.S. Mathematics &amp; Statistics</b> , Georgetown University, Washington, DC USA
	2007	<b>B.S. Biology</b> , Washington & Lee University, Lexington, VA USA
RESEARCH	<b>In Review</b>	
	<ol style="list-style-type: none"> <li>[1] Rumack A, Rosenfeld R, <b>Townes FW</b>. Correcting for heterogeneity in real-time epidemiological indicators (2023). Preprint: doi:<a href="https://doi.org/10.48550/arXiv.2309.16546">10.48550/arXiv.2309.16546</a>.</li> <li>[2] Gewirtz ADH, Townes FW, Engelhardt BE. Expression QTLs in single-cell sequencing data (2022). Preprint: doi:<a href="https://doi.org/10.1101/2022.08.14.503915">10.1101/2022.08.14.503915</a>.</li> </ol>	
	<b>Peer-reviewed Journal Articles</b>	
	<ol style="list-style-type: none"> <li>[1] Jones A, Townes FW, Li D, Engelhardt BE. Alignment of spatial genomics data using deep Gaussian processes. <i>Nature Methods</i> (2023). doi:<a href="https://doi.org/10.1038/s41592-023-01972-2">10.1038/s41592-023-01972-2</a></li> <li>[2] Verpeut JL, Bergeler S, Kislin M, Townes FW, Klibaite U, Dhanerawala ZM, Hoag A, Janarthanan S, Jung C, Lee J, Pisano TJ, Seagraves KM, Shaevitz JW, Wang SSH. Cerebellar contributions to a brainwide network for flexible behavior in mice. <i>Communications Biology</i> (2023). doi:<a href="https://doi.org/10.1038/s42003-023-04920-0">10.1038/s42003-023-04920-0</a></li> <li>[3] <b>Townes FW</b>, Engelhardt BE. Nonnegative spatial factorization applied to spatial genomics. <i>Nature Methods</i> (2023). doi:<a href="https://doi.org/10.1038/s41592-022-01687-w">10.1038/s41592-022-01687-w</a></li> <li>[4] Gewirtz ADH, Townes FW, and Engelhardt BE. Telescoping bimodal latent Dirichlet allocation to identify expression QTLs across tissues. <i>Life Science Alliance</i> (2022). doi:<a href="https://doi.org/10.26508/lsa.202101297">10.26508/lsa.202101297</a></li> <li>[5] Jones A, Townes FW, Li D, Engelhardt BE. Contrastive latent variable modeling with application to case-control sequencing experiments. <i>Annals of Applied Statistics</i> (2022). doi:<a href="https://doi.org/10.1214/21-AOAS1534">10.1214/21-AOAS1534</a></li> </ol>	

- [6] 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
- [7] **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
- [8] **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
- [9] **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.
- [10] 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.
- [11] Hicks SC, Townes FW, Teng M, Irizarry RA. Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments. *Biostatistics* (2018). doi:10.1093/biostatistics/kxx053.
- [12] 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.
- [13] 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.
- [14] **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] Community Wastewater-based Infectious Disease Surveillance - **Data Analysis and Integration** (panelist). National Academies of Sciences, Engineering, and Medicine. October 2023.
- [2] Biomedical Informatics Department, University of Colorado Anschutz Medical Center. October 2022.
- [3] Advanced Biomedical Computation Seminar, Computational Pathology Department, Brigham and Women's Hospital. March 2022.
- [4] Biomedical Data Science Department, Stanford University. February 2022.
- [5] Biostatistics Department, University of Michigan. January 2022.
- [6] Population and Public Health Sciences Department, University of Southern California. January 2022.
- [7] Emerging Concepts in Microbial Informatics, Molecular Medicine and Biotechnology (panelist). Trinity University of Asia. November 2021.

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

### Contributed

- [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] Introduction to cellular and molecular biology. (informal) *CMU Statistics Department*. October 2023.
- [2] Introduction to single cell genomics. (informal) *CMU Machine Learning Department lunch chats*. May 2023.
- [3] Wastewater based epidemiology. *Auton lab*. January 2023.
- [4] Human tumor atlas network steering committee meeting. April 2022.
- [5] Nonnegative spatial factorization. *Hansen and Hicks labs, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*. November 2021.
- [6] Primer: Generalized linear models and latent factor models. *Broad Institute Models, Inference, and Algorithms*. October 2020.
- [7] 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.
- [8] Unsupervised Learning for Single Cell Gene Expression. *Harvard Biostatistics Cancer Working Group*. September 2018.
- [9] Single Cell Housekeeping Genes and Normalization. *Dana-Farber Cancer Institute cBio Seminar*. April 2018.
- [10] Informative Missing Data in Single Cell RNA-Seq. *Dana-Farber Cancer Institute Genomics Seminar*. October 2016.
- [11] 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>

PEER REVIEW    **Journals:** *American Journal of Human Genetics, Bioinformatics, Biostatistics, Epidemics, Genome Biology, Journal of Machine Learning Research, NAR Genomics and Bioinformatics, Nature Biotechnology, Nature Communications, Nature Methods*

DOCTORAL  
THESIS  
COMMITTEES

- Julia Walchessen, “Neural Inference for Complex Spatial Processes Motivated by Challenges in Climate Science”. CMU Statistics (2023-)
- Aaron Rumack, “Bias Correction in Epidemiological Data and Forecasts”. CMU Machine Learning (2023)
- Timothy Barry, “Robust Inference for Single-Cell CRISPR Screens”. CMU Statistics (2023)

OTHER  
SERVICE  
ACTIVITIES

- Seminar Committee (2022-present, chair 2023-present)
- STATGEN 2024- Conference on Statistics in Genomics and Genetics. Local organizing committee member

PROFESSIONAL SOCIETIES    American Statistical Association, International Society for Computational Biology, Sigma Xi

FUNDING

- The Delphi Center for Outbreak Analytics and Disease Modeling in Public Health Response: Innovation and Coordination (CDC). PI: Roni Rosenfeld. Responsibility: Co-PI.
- Digital Public Health Surveillance for the 21st Century (CDC 75D30123C15907). 2023-2024. PI: Roni Rosenfeld. Responsibility: Co-Investigator.
- NIH T32 Training Grant: Cancer. 2018-2019.
- Chan-Zuckerberg Foundation Travel Grant. Human Cell Atlas Jamboree. European Bioinformatics Institute, Hinxton, UK. 2017.
- NIH T32 Training Grant: Big Data to Knowledge. 2016-2018.
- NSF-CBMS Travel Award: Topology, Geometry, and Statistics. 2016.
- NIH T32 Training Grant: Clinical Epidemiology of Lung Diseases. 2014-2016.

TEACHING  
EXPERIENCE

2023 Fall	<b>Lead Instructor</b> , Introduction to Statistical Computing (CMU - 36-350)
2023 Spring	<b>Lead Instructor</b> , Introduction to Statistical Computing (CMU - 36-350)
2021 Fall	<b>Guest lecture</b> , Advanced Computational Genomics (Princeton - COS 597D)
2020 Fall	<b>Guest lecture</b> , Computational Biology of Single Cells (Princeton - COS 597F)
2018 Fall	<b>Teaching Assistant</b> , Applied Regression (Harvard - BST 210) Grade homeworks, hold office hours, and teach a weekly lab section.
2018 Summer	<b>Co-Instructor</b> , StatStart (Harvard) Statistics program for under-represented high school students. Gave two interactive lectures on graphing data and regression.
2017 Fall	<b>Teaching Assistant</b> , Applied Bayesian Analysis (Harvard - BST 228) Grade homeworks, provide solutions, and hold office hours.
2017 Summer	<b>Lead Instructor</b> , 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	<b>Teaching Assistant</b> , Applied Longitudinal Analysis (Harvard - BST 226) Grade homeworks and hold office hours.
2016 Fall	<b>Teaching Assistant</b> , Intro to Statistical Methods (Harvard - BST 201) Grade homeworks, provide solutions, hold office hours, and teach a weekly lab section.
2016 Summer	<b>Project Mentor</b> , Pipelines into Biostatistics (Harvard) Program for under-represented undergraduates. Guided three students on a pharmacogenomics data analysis.
2016 Spring	<b>Teaching Assistant</b> , 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, L<sup>A</sup>T<sub>E</sub>X, Shell  
**Operating Systems**: MacOS, Linux, Windows  
**Machine Learning Frameworks**: Caret, Scikit-learn, Keras, Tensorflow, Stan