Michael J. Wilson, PhD

Princeton, NJ · Open to NYC & Philadelphia (onsite/hybrid) · US Permanent Resident
michaeljwilson.github.io · wilson.michael.j.22@gmail.com · +1 (609) 255-9122 · LinkedIn · GitHub

Computational physicist (PhD) and data scientist with 10+ years building high-performance software and production data pipelines for simulation and statistical inference. At genomics startup biomodal, built the raw data processing, detection method and validation framework for allele-specific methylation, a commercial product. Co-led a ∼10-person team on a $70M Department of Energy experiment, designing production algorithms for survey efficiency and a new strategy for a five-year extension.

Technical skills

Engineering:	data analysis & curation, workflows (Nextflow), Docker, cloud computing (GCP), scientific computing (SLURM), Linux, Git, AI coding (Claude)
Frameworks:	Python (NumPy, SciPy, pandas, polars, Dask, scikit-learn), automated testing (pytest), documentation, Python packaging (uv), PyTorch, JAX, Rust, C
Statistics/ML/AI:	probabilistic modeling & Bayesian inference, optimization, simulation, signal processing (FFT), machine learning, model validation
Communication:	technical writing, visualization, literature review, presentation, mentoring

Education

PhD in Physics, University of Edinburgh (UK) 2012 – 2016
Master of Physics – First Class Honours, University of Oxford (UK) 2008 – 2012

Experience

Associate Research Scholar | Princeton University (USA) Jan. 2024 – Present

- Lead developer of `cna-maste`, open-source software that detects cancer copy-number aberrations (CNAs: DNA gains and losses that drive tumors) in high-resolution spatial transcriptomics (maps of gene expression across tissue by 10x Visium HD, Illumina StrataMap).
- Improved sensitivity and recall by ∼20% over existing methods. Validated with simulation and an ovarian cancer sample. First-author manuscript submitted.
- Scaled inference to ∼100K pixels in minutes (>20× current) by exploiting sparsity, graph cuts, JAX and inductive bias.

Computational Scientist | biomodal, Cambridge (UK) Dec. 2022 – Dec. 2023

- Core developer of the flagship allele-specific methylation (ASM: methylation differences between parental gene copies) software, from prototype to commercial product: read pre-processing and QC, statistical ASM calling, analysis and reporting.
- Led analysis of protocol performance, biological validation and experimental design; results presented at leading international genomics conferences (AGBT, Biology of Genomes). Represented biomodal at the Seqera (Nextflow) Summit.

- Built and managed a scalable, reproducible Nextflow pipeline on GCP with Docker to classify colorectal cancer from epigenetic sequencing of a large patient cohort; submitted to *PNAS*.

Postdoctoral Associate | Durham University (UK) Jul. 2021 – Oct. 2022

- Designed and built the core inference pipeline for the DESI Bright Galaxy Survey luminosity function (how many galaxies exist at each brightness in the local Universe), developing optimal statistical estimators on HPC for 10M+ objects.
- Delivered the most precise measurement to date, reaching unprecedentedly faint galaxies with much smaller errors than the previous benchmark (GAMA); *MNRAS* (2026).

Postdoctoral Fellow | Lawrence Berkeley National Lab (USA) Aug. 2017 – Jun. 2021

- Co-led the Bright Galaxy Survey team for the $70M DESI survey, owning science goals, survey design, and software that met strict Department of Energy performance requirements. *Awarded DESI 'Builder' status for critical contributions* to the mission.
- Developed statistical estimators and target selection algorithms that maximized survey efficiency; deployed in production on every observation and foundational to downstream science.
- Led the science case for a next-generation high-redshift survey (simulation, forecasts, instrument comparison); foundation for the five-year DESI-II extension.

PhD Researcher | University of Edinburgh (UK) Sep. 2012 – Aug. 2017

- Designed a calibration method with 1,000× faster runtime and better accuracy for galaxy-survey inference; adopted as standard by major collaborations (*MNRAS*).
- Reduced a ∼100K-galaxy survey (VIPERS) to summary statistics and measured the cosmic growth rate to ∼10% precision, consistent with General Relativity.

Selected publications

Complete list of 15+ publications (≃ 500 citations) available upon request.

[1] C. Ma, M. Balaban, ..., **M. J. Wilson**, et al. (2024). *"Inferring allele-specific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics". Nat. Methods.*

[2] F. Puddu, ..., **M. J. Wilson**, et al. (2025). *"5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer".* Submitted to *PNAS.*

[3] C. Hahn, **M. J. Wilson**, O. Ruiz, et al. (DESI) (2022). *"DESI Bright Galaxy Survey: final target selection, design and validation". ApJ.*

[4] **M. J. Wilson** & M. White (2019). *"Cosmology with dropout selection: straw-man surveys & CMB lensing". Astronomy & Astrophysics.*

[5] **M. J. Wilson**, J. Peacock, A. Taylor, S. de la Torre (2016). *"Rapid modelling of the redshift-space power spectrum multipoles for a masked density field". MNRAS.*