Charlotte A. Darby, PhD

cdarby.github.io | charlotte.ay.darby@gmail.com | LinkedIn | @ChaDarby

I'm a computer scientist working in the exciting space of computational biology in discovery, preclinical, and clinical stages of Oncology drug development. I work with omics data at all scales, from 1 million single cells to the first patient in a clinical trial and contribute to multiple cross-functional project teams. I have experience in algorithm development in computational biology and bring a strong academic background in computer science (Johns Hopkins PhD CS, 2020; Carnegie Mellon BS/MS CompBio, 2016). I continuously strive for high-quality technical presentation and communication skills.

Experience

Takeda Pharmaceuticals — Cambridge, MA   3/2021 - present
Research Scientist II, Computational Biology
Process, analyze, and interpret omics data (RNA-seq, scRNA-seq, Nanostring ...)
to inform translational strategy for numerous discovery, preclinical, and clinical
projects in Oncology, particularly Immuno-Oncology. Work with cross-functional
project teams to analyze public and internal datasets.
Build computational biology infrastructure (workflows, packages, pipelines).

New York Genome Center — New York, NY (Remote)
Senior Computational Biologist, Satija Lab   6/2020 - 11/2020
Worked on the Seurat single-cell analysis R package, maintaining and improving
existing code and adding new features and functionality for the v4 release in
October 2020. Developed and deployed the Azimuth Shiny application on AWS.
This app demonstrates the new reference-mapping algorithm in Seurat v4.

                                                                                 6/2019 - 8/2019

10x Genomics — Pleasanton, CA
Computational Biology Intern
Developed scHLAcount, a bioinformatics workflow for allele-specific HLA gene
expression from single-cell RNA-seq.
Designed "dual index" sample indexes for new multiplexing kits.

IBM Research Almaden — San Jose, CA   6/2018 - 8/2018
Graduate Research Assistant in Applied & Industrial Genomics
Project: Machine learning for classifying bacterial genomes

Cold Spring Harbor Laboratory — Cold Spring Harbor, NY   6/2015 - 8/2015
Undergraduate Research Program (URP) Researcher in Doreen Ware's lab
Project: Transcription factor networks in maize and *Arabidopsis*

Education

Ph.D. Computer Science — Johns Hopkins University   9/2016 - 6/2020

Thesis:	Computational methods addressing genetic variation in next-generation sequencing data (May 2020)
Major Projects:	Vargas: fast heuristic-free read alignment
	Samovar: somatic variant calling with linked reads
Other Projects:	Algorithms for polyploid haplotype assembly
	Species phylogeny and comparative genomics of *Trichomonadida*
	LRSim: simulation software for linked reads
	Comparing haplotype assemblies from linked and long reads

Co-Advisors: Michael Schatz, Ben Langmead

M.S. Computational Biology — Carnegie Mellon University   9/2015 - 5/2016

GPA 3.55; Research Honors
Thesis: New nomenclature for horizontal gene transfer
Advisor: Dannie Durand

B.S. Computational Biology — Carnegie Mellon University   9/2012 - 9/2015

GPA 3.70; University Honors

Technical Skills

Programming. *Advanced:* R, Unix. *Proficient:* Python, Rust, C++.

CS. cloud computing (AWS), R Shiny, RStudio, functional programming, algorithms, high-performance computing environments, Git

CompBio. bioinformatics pipelines, bioinformatics algorithm development, phylogenetic analysis/evolutionary biology, genome assembly, structural variant detection, small variant calling, short read alignment, haplotype assembly (phasing)

'Omics. transcriptomics, single-cell 'omics, functional genomics, long read and third-generation sequencing technologies, linked reads

Pharma. matrix environment, multidisclipinary project teams, preclinical drug development, immuno-oncology

Teaching

Johns Hopkins University — Baltimore, MD   2017 — 2019

- Teaching Academy [Certificate of Completion](#)
- Instructor, "Algorithms for hard problems in computer science" (seminar for first-year engineering students, class size 10, 2 semesters)
- Teaching Assistant, "Applied Comparative Genomics" (graduate elective course, class size 20, 1 semester)

Center for Talented Youth — Baltimore, MD   2016 — 2017

- Teaching Assistant, Genetics & Genomics (residential summer academic program for high schoolers, class size 6-18, three 3-week sessions)

Carnegie Mellon University — Pittsburgh, PA

- Teaching Assistant, Introduction to Computational Biology   2015 — 2016 (undergraduate core course, taught recitation sections and held office hours, class size 20/section, 2 semesters)

Conference Posters and Presentations

Vargas: heuristic-free alignment for assessing graph and linear aligners (poster)	Genome Informatics 2019
Samovar: single-sample mosaic SNV calling with linked reads (talk) Best Paper Award	RECOMB-SEQ 2019
A heuristic algorithm for the k-ploid haplotype assembly problem (poster with undergraduate mentee George Botev)	Biological Data Science 2018
Leveraging linked reads for single-sample somatic variant calling (poster)	Genome Informatics 2017
Identification and analysis of somatic variants using linked read sequencing (poster)	Biology of Genomes 2017
Classifying Xenologs (talk)	GL-Bio/CCBC 2016
Classifying Xenologs (poster)	RECOMB-CG 2014

Outreach

Maryland New Directions Assist job-seeking clients in selecting professional clothing sourced from community donations	2018 — 2020
Johns Hopkins SABES Volunteer bi-weekly as a mentor for after-school elementary STEM program	2016 — 2017
Women@SCS (Carnegie Mellon University) Facilitate TechNights workshops for middle school students	2014 — 2016

Networking

CRA-W (Computing Research Association) Grad Cohort for Women Workshop attendee	April 2018

Honors

Academic Achievement Fellowship Carnegie Mellon University (M.S. Computational Biology program)	2015 — 2016
National Merit Scholarship	2012

Publications

- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck III, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., **Darby, C.**, … Smibert, P., Satija, R. (2020). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), P3573-3587.E29. https://doi.org/10.1016/j.cell.2021.04.048

- **Darby, C. A.**, Gaddipati, R., Schatz, M. C., & Langmead, B. (2020). Vargas: heuristic-free alignment for assessing linear and graph read aligners. *Bioinformatics*, 36(12), btaa265. https://doi.org/10.1093/bioinformatics/btaa265

- **Darby, C. A.**, Stubbington, M. J. T., Marks, P. J., Barrio, Á. M., & Fiddes, I. T. (2020). scHLAcount: Allele-specific HLA expression from single-cell gene expression data. *Bioinformatics*, 36(12), btaa264. https://doi.org/10.1093/bioinformatics/btaa264

- **Darby, C. A.**, Fitch, J. R., Brennan, P. J., Kelly, B. J., Bir, N., Magrini, V., … Langmead, B., Schatz, M. C. (2019). Samovar: Single-Sample Mosaic Single-Nucleotide Variant Calling with

Linked Reads. *IScience*, *18*, 1–10. https://doi.org/10.1016/J.ISCI.2019.05.037

- Sedlazeck, F. J., Lee, H., **Darby, C. A.**, & Schatz, M. C. (2018). Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics*, *19*(6), 329–346. https://doi.org/10.1038/s41576-018-0003-4

- Luo, R., Sedlazeck, F. J., **Darby, C. A.**, Kelly, S. M., & Schatz, M. C. (2017). LRSim: A Linked-Reads Simulator Generating Insights for Better Genome Partitioning. *Computational and Structural Biotechnology Journal*, *15*, 478–484. https://doi.org/10.1016/j.csbj.2017.10.002

- **Darby, C. A.***, Stolzer, M.*, Ropp, P. J., Barker, D., & Durand, D. (2017). Xenolog classification. *Bioinformatics*, *27*(5), btw686. https://doi.org/10.1093/bioinformatics/btw686