Erin Wilson

Website
erinhwilson
erinhwilson@gmail.com
LinkedIn

Computer Science PhD student with 3+ additional years working in the biotech industry ▪ Approachable, enthusiastic scientist who can connect ideas and skillsets across machine learning and biology ▪ Graduate research focus: using deep learning methods to discover DNA regulatory elements in methane-eating bacteria ▪ **Seeking a data scientist position where I can be part of sustainable solutions for environmental & climate challenges and continue growing as a creative ML researcher.**

Key words: *data science, synthetic biology, machine learning, genomics, RNA-seq, metabolic engineering*
Research Tools: *Python, PyTorch, Altair, Git, AWS, HPC, BLAST, BWA, Samtools*

Education

PhD Candidate, Computer Science, University of Washington, Seattle, WA **2017 - Present**
Advanced Data Science option; Computational Molecular Biology certificate
Advisors: Dr. Mary Lidstrom, Dr. David Beck

- **Thesis:** (in progress) *Towards sustainable biomolecule production: Computational approaches to accelerate genetic tool development for engineering metabolism in microorganisms*
  - Building deep learning models to predict RNA-seq expression directly from DNA sequences (upstream promoter regions) in the methanotroph *M. buryatense*.
  - Developed computational framework to identify strong, constitutive promoters in non-model organisms from bulk RNA-seq data (link)
- **M.S. in Computer Science (2019)**. Thesis: *Using microorganisms to solve macro problems: untangling the genetic circuitry of methane-eating bacteria.*
- **Visiting PhD student:** Center for Biosustainability at the Technical University of Denmark (2022)
  - Using Independent Component Analysis (ICA) to identify independently regulated gene modules (iModulons) from a compendium of bulk RNA-seq data in *M. buryatense*

B.A. Computational Biology, Carleton College, Northfield, MN **2010 - 2014**
Advisors: Dr. David Liben-Nowell, Dr. Jennifer Wolff

- **Thesis:** *Developing a benchmark standard for predicting gene targets of chemical perturbants.*
- **Graduation Honors:** *Magna Cum Laude*; awarded Distinction on senior thesis.
- **Studies Abroad:** Coastal Marine Ecology program with Dr. Annie Bosacker in Queensland, Australia

Work & Research Experience

Intern, Data Science, Zymergen **Summer 2018**
Mentor: Trent Hauck

- Prototyped deep learning models for predicting the presence of DNA regulatory features in non-model microbe genomes (using: Keras/Tensorflow, CNN, sklearn).

Associate Scientist, Scientific Computing, Amyris **July 2014 - July 2017**
Mentors: Dr. Amoolya Singh, Dr. Darren Platt

- **Whole Genome Sequencing:** maintained automated sequence analysis pipeline; provided technical and troubleshooting support for internal Automated Strain Engineering (ASE) service's DNA QC workflow and for individual scientists interpretting mutation and coverage data.
- **Genotype Specification Language (GSL):** implemented software features in DNA compiler code base; trained biologists to use GSL syntax for designing DNA constructs; co-authored journal article and textbook chapter introducing open-source version of the GSL compiler; collaborated with Autodesk engineers to add GSL extension into open-source Genetic Constructor interface (blog post)
- **Genotype Generator:** developed software pipeline to interface between Design and Build phases of DARPA-funded project to produce 450 molecules ("Automated Scientist"); implemented code and database schema to translate high level designs for metabolic pathways into instructions for building strains to carry out pathway design.
- **Outreach and Communication:** helped facilitate communication of software needs between biologists and software engineers; held weekly office hours for 1x1 sequencing and GSL support; gave technical presentations to Amyris R&D (∼60 people), Strain Engineering (∼30 people), and Automation & Computing groups (∼30 people); engaged with attendees at local Bay Area science nights about Amyris' sustainable technology.

Intern, Scientific Computing, Amyris **December 2013**
Mentor: Dr. Amoolya Singh
- Implemented data visualization tool to overlay experimental data on yeast metabolic pathway

Research Assistant, Computational Biology, University of Minnesota **Summer 2013**
Principal Investigator: Dr. Chad Myers
- Analyzed genetic and chemical-genetic interaction data to predict gene targets for chemical perturbants; designed and coded target prediction pipeline; used pipeline to create benchmark for prediction accuracy using chemicals with known targets

Research Assistant, Evolutionary Computing, Carleton College **Summer 2012**
Principal Investigator: Dr. Sherri Goings
- Executed experiments with populations of mutating digital organisms to examine the effects of limited CPU resources on the populations' ability to evolve complex Boolean logic functions

Research Assistant, Genetics, University of California, San Francisco **Summer 2011**
Principal Investigator: Dr. Nadav Ahituv
- Performed chromatin immunoprecipitation sequencing experiments on mouse limb tissue to find enhancer candidates involved in limb patterning and development

Publications

- L. He, J. D. Groom, **E. H. Wilson,** J. Fernandez, M. C. Konopka, D. A. C. Beck, M. E. Lidstrom. "A methanotrophic bacterium to enable direct methane capture for climate mitigation." [*Under Review*]
- **E. H. Wilson,** M. E. Lidstrom, and D. A. C. Beck. (2021) "A multi-task learning approach to enhance sustainable biomolecule production in engineered microorganisms." Tackling Climate Change with Machine Learning, workshop at *ICML*. (recording)
- **E. H. Wilson,** J. D. Groom, M. C. Sarfatis, S. M. Ford, M. E. Lidstrom, and D. A. C. Beck. (2021) "A Computational Framework for Identifying Promoter Sequences in Nonmodel Organisms Using RNA-seq Data Sets." *ACS Synthetic Biology.*
- **E. H. Wilson,** C. Macklin, and D. Platt. (2018) Engineering genomes with Genotype Specification Language. In *Methods in Molecular Biology, Synthetic Biology*. J.C. Braman, ed. Springer Publishing Company, New York, NY. In Press.
- **E. H. Wilson,** S. Sagawa, J. Weis, M. Shubert, M. Bissell, B. Hawthorne, C. Reeves, J. Dean, and D. Platt. (2016) Genotype Specification Language. *ACS Synthetic Biology.*
- A. H. Singh, B. B. Kaufmann-Malaga, J. A. Lerman, D. P. Dougherty, Y. Zhang, A. L. Kilbo, **E. H. Wilson**, C. Y. Ng, O. Erbilgin, K. A. Curran, C. D. Reeves, J. E. Hung, S. Mantovani, Z. A. King, M. J. Ayson, J. R. Denery, C. Lu, P. Norton, C. Tran, D. M. Platt, J. R. Cherry, S. S. Chandran, A. L. Meadows. (2023) "An Automated Scientist to Design and Optimize Microbial Strains for the Industrial Production of Small Molecules" bioRxiv
- S. W. Simpkins, J. Nelson, R. Deshpande, S.C. Li, J. S. Piotrowski, **E. H. Wilson**, A. A. Gebre, R. Okamoto, M. Yoshimura, M. Costanzo, Y. Yashiroda, Y. Ohya, H. Osada, M. Yoshida, C. Boone, C. L. Myers. (2018) "Predicting bioprocess targets of chemical compounds through integration of chemical-genetic and genetic interactions." *PLoS Computational Biology.*

Presentations & Posters

- **E. H. Wilson.**, M. E. Lidstrom, D. A. C. Beck. "Methane, Microbes, and Machine Learning: Engineering biology to combat climate change." Poster at Industry Affiliates Research Symposium, University of Washington, November 2022.
- **E. H. Wilson.** "Using microorganisms to mitigate macro-problems." Talk at Virtual Women's Research Day, University of Washington, April 2020. (recording)
- **E. H. Wilson,** M. E. Lidstrom, D. A. C. Beck. "Using microorganisms to solve macro problems: untangling the genetic circuitry of methane-eating bacteria." Invited talk, MIDAS Data Science Symposium, University of Michigan, November 2020.
- **E. H. Wilson.** "Can deep learning help us program biology?" Presentation at Industry Affiliates Research Symposium, University of Washington, November 2018.
- **E. H. Wilson,** D. Platt. "Genotype Specification Language: *Programming in DNA!*" Poster at Synthetic Biology, Engineering, Evolution & Design (SEED) conference in Chicago, July 2016.

Distinctions

- **Scan Design Foundation Fellowship** **2022**
  - Support for research and cultural exchange between Danish and American students
- **NSF Graduate Research Fellow** **2019**
  - Three years of research funding from the National Science Foundation
- **Marilyn Fries Fellowship** for graduate students in Computer Science & Engineering **2017**
  - Awarded first year graduate research funding
- **Clare Boothe Luce Scholarship** for Women in Physics and Computer Science **2012**
  - Received funding for summer research in Evolutionary Computing

Leadership & Outreach

Scientific Communication & Tutorials

- *"Modeling DNA Sequences with PyTorch."* (2022) Article in Towards Data Science, a Medium publication.
- *"The Light Side of Genetic Engineering."* (2019) Article in OneZero, a Medium publication.
- *"Genetic Constructor and GSL - The Best of Both worlds."* (2016) Blog post with Autodesk Bionano Research.

Volunteering & Outreach

- Research mentor for undergraduate student (2020-2023)
- Pre-application Review Service mentor (2021-2022)
  - Provided early application feedback to support prospective PhD students from diverse backgrounds
- Peer mentor for groups of incoming PhD students (2018-2022)
  - Mentorship Program Organizer (2019-2020)
- New Grad Orientation Organizer (2018)
- *"Programming Organisms with DNA Puzzles!"*
  - Developed interactive activity to teach elementary/middle schoolers about metabolic engineering at UW's "Engineering Discovery Days" and "Introduce a Girl to CoRDS (Coding, Robotics, and Data Science)"

Interests & Activities

- **Recreational data visualization**
  - "Mistborn: The Final Eyebrow." (2021) Article in Towards Data Science.
    - An analysis of social dynamics in the fantasy novel *Mistborn: The Final Empire* by Brandon Sanderson as conveyed through characters' eyebrow raising behavior.
- **Wildlife rescue hospital volunteer**, PAWS (Seattle, WA) and Wildcare (San Rafael, CA)
  - Treat and care for injured songbirds in hospital; Co-led youth nature hikes with Education Department
- **MeadoWatch Field Data collector**, UW Biology, citizen science project
  - Collect wildflower blooming data in Mount Rainier National Park
- **Carleton Varsity Athletics (Div III)**
  - Women's Soccer (4 seasons), Women's Tennis (1 season)
  - Named to MIAC (Minnesota) Academic All-Conference *(2011, 2012, 2013)*