

CAMERON DUFAULT

{lastnamefirstname}@cs.toronto.edu [◇ linkedin.com/in/cameron-dufault/](#) [◇ dufaultc.github.io](#) [◇ github.com/dufaultc](#)

EDUCATION

PhD Computer Science, University of Toronto
Supervisor: Dr. Alan Moses

Feb 2025 – Present

MSc Computer Science, University of Toronto
Supervisor: Dr. Alan Moses

Sept 2023 – Jan 2025

MSc Paper: *Unsupervised whole-genome representation learning captures bacterial phenotypes*

Cumulative GPA: 3.93/4.0

BEng Software Engineering Co-op, McMaster University
Cumulative GPA: 3.78/4.0 (3.96/4.0 in 20 most recent courses)

Sept 2017 – Apr 2023

KEY SKILLS

Languages: Python, R, SQL, **ML:** PyTorch, NumPy, Scikit-learn, Pandas, Huggingface Transformers/Datasets, Weights & Biases **Bio modalities:** DNA/protein sequences (BioPython, MMSeqs, NCBI Datasets), scRNA-seq, medical imaging, RNA-seq **Computing:** Unix, Slurm, Git

PUBLICATIONS

- **C. Dufault**, S. Xu, and A. M. Moses, “C3P: Contrastive promoter-protein pretraining yields representations capturing bacterial gene regulation,” *The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS 2026)*, Dec. 2026. Accepted as **Oral** (top ~0.4% of submitted papers).
- A. G. Sangster, **C. Dufault**, H. Qu, D. Le, J. D. Forman-Kay, and A. M. Moses, “Zero-shot segmentation using embeddings from a protein language model identifies functional regions in the human proteome,” *PLOS Computational Biology*, vol. 21, no. 11, e1012929, Nov. 2025.
- **C. Dufault** and A. M. Moses, “Unsupervised Whole-Genome Representation Learning Captures Bacterial Phenotypes,” *ICLR 2025 Workshop on Learning Meaningful Representations of Life*, Singapore, Apr. 2025.
- M. E. Consens, **C. Dufault**, M. Wainberg, D. Forster, M. Karimzadeh, H. Goodarzi, F. J. Theis, A. M. Moses, and B. Wang, “Transformers and genome language models,” *Nature Machine Intelligence*, vol. 7, no. 3, pp. 346–362, Mar. 2025.

RESEARCH EXPERIENCE

Graduate Research Student
Alan Moses Lab, University of Toronto

Sept 2023 – Present

- Researching self-supervised learning techniques for understanding microbial genomes.
- Introduced a novel contrastive approach for self-supervised learning of regulatory DNA, which shows multi-fold improvement over existing models at capturing bacterial regulatory function. Trained small transformer model on ~88 million promoter-protein pairs extracted from ~40,000 genomes. *NeurIPS oral*.
- Introduced first approach for self-supervised learning of whole-genome representations, finding learned representations more predictive of 23/25 bacterial phenotypes than conventional representations. Trained whole-genome encoder through novel reconstruction task on 100,000 unique bacterial genomes. Made possible by the use of an efficient sequence modeling architecture (Mamba) and by tokenizing genomes as sequences of conserved genomic elements (e.g., protein domains). *ICLR workshop paper*.
- Contributor to high-impact review on transformers and genome language models. *Nature Machine Intelligence*.
- Demonstrated that a transformer for symbolic expression can be used to infer Boolean gene regulatory networks from single-cell data. *ISMB poster*.

Undergraduate Research Student

May 2022 – June 2023

Kieran Campbell Lab, Lunenfeld-Tanenbaum Research Institute

- Built a single-cell RNA sequencing data analysis pipeline and applied machine learning techniques to identify gene expression patterns across tumour samples.
- Revealed that tumour cells are stratified by cancer superclasses, recapitulating experimental findings.

Research Assistant / Undergraduate Research Student

May 2020 – Jan 2022

Nathan Magarvey Lab, McMaster University

- Developed an application for automated control of laboratory equipment via real-time mass-spectrometry data analysis (Sept 2021 – Jan 2022).
- Trained transformer models predict enzyme function from sequence (May 2020 – Apr 2021).

POSTERS

- **C. Dufault**, S. Xu, and A. M. Moses, “Learning functional representations of bacterial regulatory sequences with contrastive promoter-protein pretraining,” *XC Cold Spring Harbor Symposium on Quantitative Biology: AI in Biology*, Cold Spring Harbor, New York, 2026.
- **C. Dufault** and A. M. Moses, “Single-Cell Gene Regulatory Network Inference with a Transformer for Symbolic Regression of Boolean Functions,” *International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Montreal, 2024.

SELECTED INDUSTRY EXPERIENCE

Machine Learning Software Engineer Co-op

Jan 2022 – Apr 2022

Borealis AI (RBC’s AI Research Institute)

- Collaborated with ML researchers and engineers on pricing financial products using machine learning. Produced production-grade code in industry AI lab environment.
- Built an automatic model documentation tool, explored impacts of neural network optimization techniques (quantization, compilation).

Data Scientist Co-op

May 2021 – Aug 2021

RBC (Royal Bank of Canada)

- Developed machine learning-based tools for monitoring technical infrastructure.

VOLUNTEERING & LEADERSHIP

Co-founder/Leader

Sept 2025 – Present

Toronto Computational Biology Graduate Student Group

- Co-founded an interdepartmental organization where computational biology graduate students at the university (~40 members) can meet and discuss their work and exciting ideas in our field, with a focus on theory.
- Organized ~monthly meetings where students take turns giving “chalk talks”, describing research ideas to other students with a chalkboard.

Advisor

Feb 2026 – Present

University of Toronto iGEM

- Advising undergraduate synthetic biology competition students on applying deep learning tools to design and characterize proteins.

Director of Projects (2021–2022) / Project Leader (2020–2021)

May 2020 – Apr 2022

McMaster Artificial Intelligence Society

- Organized four student projects for our ~20 technical members, in collaboration with McMaster faculty research.
- Led a project to train a brain tumour segmentation model, applying our model to real-world patient data in collaboration with a neuroradiology researcher.