Jake Crawford

CONTACT
INFORMATION

Email: jjc2718@gmail.com
Web: http://jjc2718.github.io
GitHub: https://github.com/jjc2718

OVERVIEW

I am a fourth-year computational biology PhD student at Penn, with a strong background in cancer genomics and machine learning. My thesis work has focused on learning from multi-omics datasets in cancer, particularly public databases such as TCGA. Using my background in biology and experience in industry software engineering, my work has benefitted from reproducible and scalable computational analysis pipelines.

EDUCATION

University of Pennsylvania, Philadelphia, PA

Ph.D. Candidate, Genomics and Computational Biology
Fall 2018 - present | Advisor: Casey Greene

Tufts University, Medford, MA

M.S. in Computer Science, awarded Spring 2017
B.S. in Biology, awarded Spring 2012

EMPLOYMENT

Microsoft Research New England, Cambridge, MA   May 2017 - July 2018
Research Assistant, Computational Biology Group

Wayfair, Boston, MA   January 2014 - August 2015
Software Engineer

PUBLICATIONS

J. Crawford, B.C. Christensen, M. Chikina, C.S. Greene. Widespread redundancy in -omics profiles of cancer mutation states. *Genome Biology*, 2022.

M. Tomasoni, S. Gomez, **J. Crawford**, W. Zhang, S. Choobdar, D. Marbach, S. Bergmann. MONET: a toolbox integrating top-performing methods for network modularisation. *Bioinformatics*, 2020.

J. Crawford, C.S. Greene. Incorporating biological structure into machine learning models in biomedicine. *Current Opinion in Biotechnology*, 2020.

S. Choobdar*, M. E. Ahsen*, **J. Crawford***, M. Tomasoni, D. Lamparter, J. Lin, B. Hescott, X. Hu, J. Mercer, T. Natoli, R. Narayan, The DREAM Module Identification Challenge Consortium, A. Subramanian, G. Stolovitzky, Z. Kutalik, K. Lage, D. Slonim, J. Saez-Rodriguez, L. J. Cowen, S. Bergmann, D. Marbach. Assessment of network module identification across complex diseases. *Nature Methods*, 2019.
(* = co-first author)

J. Listgarten, M. Weinstein, B. Kleinstiver, A. A. Sousa, J. K. Joung, **J. Crawford**, K. Gao, M. Elibol, L. Hoang, J. Doench, N. Fusi. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering*, 2018.

PREPRINTS

B. J. Heil, **J. Crawford**, C. S. Greene. The effects of nonlinear signal on expression-based prediction performance. *bioRxiv*, 2022.

K. W. Govek, **J. Crawford**, A. B. Saturnino, K. Zoga, M. P. Hart, P. G. Camara. Multi-modal analysis and integration of single-cell morphological data. *bioRxiv*, 2022.

A. A. Hippen, **J. Crawford**, J. R. Gardner, C. S. Greene. wenda_gpu: fast domain adaptation for genomic data. *bioRxiv*, 2022.

A. J. Lee, D. L. Mould, **J. Crawford**, D. Hu, R. K. Powers, G. Doing, J. C. Costello, D. A. Hogan, C. S. Greene. Generative neural networks separate common and specific transcriptional responses. *bioRxiv*, 2021.

J. Crawford, C.S. Greene. Graph biased feature selection of genes is better than random for many genes. *bioRxiv*, 2020. (no peer-reviewed journal submission currently planned for this work)

MENTORING | *Mentoring experience in Greene Lab:*
- Fall 2021: Mentored Erik Serrano (PhD rotation student at CU Anschutz) on a 3-month rotation project related to multi-omics predictive modeling in cancer
- Summer/Fall 2021: Mentored Matt Ayala (Biomedical Data Science MS student at CU Denver) on a summer data science internship project comparing different cross-validation schemes on TCGA data

TEACHING EXPERIENCE | **Teaching Assistant**, Tufts University
COMP 167: Computational Biology — Spring 2017
COMP 40: Machine Structure and Assembly Language — Fall 2015, Spring 2016

SKILLS AND SOFTWARE | *Programming Languages* (from most to least proficient):
Python, R, C/C++, Bash, JavaScript, PHP, Scheme

Software (in no particular order):
Vim, Git, LaTeX, NumPy, SciPy, scikit-learn, pandas, PyTorch, Jupyter Lab

COURSEWORK | University of Pennsylvania:
- Machine Learning (CIS 520)
- Statistical Inference (STAT 431)
- Bayesian Modeling and Computation (STAT 927)
- Numerical Linear Algebra (AMCS 602)
- Signaling Pathways in Cancer (BMB 585)
- Regulation of the Genome (BIOM 555)

Tufts University:
- Probabilistic Systems Analysis (EE 104)
- Algorithms (COMP 160)

SERVICE AND SOCIETIES | GCB Advising Committee, student representative (2020-2022)
Reviewer for: *Bioinformatics Advances* (2022); *Bioinformatics* (2019-2022); *Briefings in Bioinformatics* (2021-2022); *PLoS Computational Biology* (2022); *Nature Machine Intelligence* (2021); *Computational and Structural Biotechnology* (2020)
Member, International Society for Computational Biology (ISCB)