Evan M. Cofer

Computational Biology & Computer Science

Lewis-Sigler Institute for Integrative Genomics
Princeton University
Princeton, NJ 08544
✉ *ecofer {at} princeton {dot} edu*
🖳 *www.evancofer.com*

▬▬▬ Education

September 2017 – Present	**Ph.D. in Quantitative and Computational Biology**, *Princeton University*, Princeton, NJ. Certificate in Statistics and Machine Learning
2017 – 2019	**M.A. in Quantitative and Computational Biology**, *Princeton University*, Princeton, NJ.
2012 – 2017	**Bachelor of Science in Computer Science**, *Trinity University*, San Antonio, TX. *Computer Science GPA: 3.93/4.0*

▬▬▬ Awards, Honors, and Scholarships

April 2019 **Graduate Research Fellowship Program Award**, *National Science Foundation (NSF)*.

"The program supports outstanding graduate students who are pursuing research-based master's and doctoral degrees in science, technology, engineering, and mathematics (STEM) or in STEM education. The GRFP provides three years of support for the graduate education of individuals who have demonstrated their potential for significant research achievements in STEM or STEM education."

March 2017 **Graduate Research Fellowship Program: Honorable Mention**, *National Science Foundation (NSF)*.

"The program supports outstanding graduate students who are pursuing research-based master's and doctoral degrees in science, technology, engineering, and mathematics (STEM) or in STEM education. The NSF accords Honorable Mention to meritorious applicants who do not receive Fellowship awards. This is considered a significant national academic achievement."

February 2017 **Trinity Computer Science Department Outstanding Senior Research Award**, *Trinity University*.

"Not awarded every year, this honor is given to recognize that senior who has a particularly distinguished record of Computer Science research."

April 2016 **Mach Research Fellowship**, *Trinity University*.

Yearly, each academic department recommends a rising senior for the Mach, which recognizes outstanding undergraduate research achievement. From these nominees, the Faculty Research Committee selects five recipients. I was nominated by the computer science department, and subsequently chosen as a recipient.

April 2016 & December 2015 **No-Boundary Thinking Contest: First Place**, *Trinity University*.

"Entrants should propose a computational solution to an interdisciplinary problem. In addition to an essay, submissions should include a presentation at the department colloquium." I competed twice, winning both times. My first submission considered the use of authorship/topic networks for improving gene name disambiguation. For the second entry, I described a means of constructing lightweight ontologies to aid in the exploration of large datasets of leaked documents.

December 2015 – May 2017	**Dean's List**, *Trinity University*. Achieved a grade point average of 3.65 or higher while taking 15 or more hours.
August 2012 – May 2017	**Trustee's Scholarship**, *Trinity University*. Scholarship for academic merit.

Research Experience

March 2018 – Present	**Troyanskaya Lab**, *Princeton University*. Researched and developed deep learning methods for analyzing genomics and histopathology data. Initiated collaborations with other labs from Princeton and beyond. Recruited and mentored both graduate and undergraduate students. Interviewed staff candidates at all levels. http://function.princeton.edu/
May 2017 – September 2017	**Kostic Lab**, *Joslin Diabetes Center and Harvard Medical School*. Conducted research in metagenomics and type 1 diabetes. Developed scalable and cost-efficient tools for bioinformatics in the cloud. Interviewed computational research staff candidates. http://www.kosticlab.org/
January 2016 – June 2017	**Hibbs Lab**, *Trinity University*. Quantified the rate at which somatic mutations accumulate in healthy, aging mammals. Developed a deep learning model to distinguish somatic mutations from noise in high volume DNA sequencing data (>13 trillion paired-end reads). http://www.cs.trinity.edu/ mhibbs/HibbsHome/Research.html
April 2015 – July 2015	**Jiang Lab**, *Trinity University*. Modelled the popular vote of "battleground" states in the 2012 United States presidential election. Predicted outcomes both nationally and state-by-state. http://www.cs.trinity.edu/ xjiang/
May 2013 – July 2013	**Glawe Lab**, *Trinity University*. Analyzed the use and accuracy of nutating disc mechanical flow meters for recycling water condensation from air conditioning systems. In addition to research, managed all equipment and developed graphical user interfaces for lab instruments. https://new.trinity.edu/faculty/diana-glawe

Teaching Experience

January 2021 – May 2021	**Teaching Assistant**, *Princeton University*. Worked as a grader, tutor, and teaching assistant for ISC-234: An Integrated, Quantitative Introduction to the Natural Sciences II.
2018 – Present	**Student Mentor and Supervisor**, *Princeton University*. Directly mentored and supervised several undergraduate and graduate student researchers.
January 2017 – May 2017	**Teaching Assistant**, *Trinity University*. Worked as a grader, tutor, and teaching assistant for CSCI-1323: Discrete Structures.

Additional Academic Experience

Referee.

Science Translational Medicine, PLOS: Computational Biology, Scientific Reports, ICCV: Workshop on Computer Vision for Automated Medical Diagnosis, ICML: Workshop on Interpretable Machine Learning in Healthcare (IMLH), NeurIPS: Workshop on Machine Learning for Health (ML4H), Pacific Symposium on Biocomputing (PSB), Machine Learning for Healthcare (MLHC)

Relevant Coursework

Completed at Princeton University.
Molecular Basis of Cancer, Computer Systems and Machine Learning, Graduate Research Seminar in Statistics and Machine Learning, Foundations of Applied Statistics and Data Science with Applications in Biology, Fundamentals of Machine Learning, Introduction to Genomics and Computational Molecular Biology, Method and Logic in Quantitative Biology

Completed at Trinity University.
Computer Science Thesis I & II, Artificial Intelligence, Compilers, Software Engineering, Principles of Functional Languages, Theoretical Computer Science, Database Systems, UNIX Power Tools, Web Application Development, Advanced Algorithms, Principles of Computer Design, Computer Science Thesis Reading, Data Abstraction, Computational Game Theory and Multi-agent Systems, Discrete Data Structures, Principles of Computer Science I & II, Probability & Statistics for Scientists & Engineers, Differential Equations & Linear Algebra, Calculus I, II, & III, Directed Studies in Computer Science, Competitive Programming, Electronic Circuits & Lab, Physics I & II, Thermodynamics I, Mass and Energy Balance

Proficiencies

Languages and Technologies.
Python, PyTorch, R, TensorFlow (Keras), C++, bash, Scala (Akka, Play, Scala.js), git, Linux, C, SQL, JavaScript, Haskell, HPC (Torque, Slurm), MongoDB, LATEX, CSS/HTML, MIPS Assembly

Academic and Professional Affiliations

February 2017 – Present **Member, Upsilon Pi Epsilon**.
International honor society for computer scientists.
http://upe.acm.org/

August 2015 – Present **Member, Association for Computing Machinery (ACM)**.
Professional and scholarly society for computer scientists.
https://www.acm.org

Publications

Cofer EM, Raimundo J, Tadych A, Yamazaki Y, Wong AK, Theesfeld CL, Levine MS, and Troyanskaya OG. Modeling transcriptional regulation of model species with deep learning. *Genome Research*, 31(6), April 2021. doi:10.1101/gr.266171.120.

Le Goallec A*, Tierney BT*, Luber JM, **Cofer EM**, Kostic AD, and Patel CJ. A systematic machine learning and data type comparison yields metagenomic predictors of infant age, sex, breastfeeding, antibiotic usage, country of origin, and delivery type. *PLOS Computational Biology*, 16(5), May 2020. doi:10.1371/journal.pcbi.1007895. (*Contributed equally)

Chen KM*, **Cofer EM***, Zhou J, and Troyanskaya OG. Selene: a PyTorch-based deep learning library for sequence data. *Nature Methods*, 16(4):315–318, March 2019. doi:10.1038/s41592-019-0360-8. (*Contributed equally)

Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow P, Zietz M, Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Lanchantin J Israeli J, Woloszynek S, Carpenter AE, Shrikumar A, Xu J, **Cofer EM**, Lavender CA, Turaga SC, Alexandari AM, Zhiyong L, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Boca SM, Swamidass J, Huang A, Gitter A[†], and Greene CS[†]. Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society Interface*, 16(141), April 2018. doi:10.1098/rsif.2017.0387. (Random ordering where not marked by [†])

Luber JM*, Tierney BT*, **Cofer EM**, Patel CJ, and Kostic AD. Aether: Leveraging linear programming for optimal cloud computing in genomics. *Bioinformatics*, December 2017. doi:10.1093/bioinformatics/btx787. (*Contributed equally)

Preprints

Zhang Z*, **Cofer EM***, and Troyanskaya OG (Contributed equally). AMBIENT: Accelerated convolutional neural network architecture search for regulatory genomics. *bioRxiv*, January 2021. doi:10.1101/2021.02.25.432960. (*Contributed equally)

Cofer EM, Raimundo J, Tadych A, Yamazaki Y, Wong AK, Theesfeld CL, Levine MS, and Troyanskaya OG. DeepArk: modeling *cis*-regulatory codes of model species with deep learning. *bioRxiv*, April 2020. doi:10.1101/2020.04.23.058040.

Chen KM*, **Cofer EM***, Zhou J, and Troyanskaya OG. Selene: a PyTorch-based deep learning library for sequence-level data. *bioRxiv*, October 2018. doi:10.1101/438291. (*Contributed equally)

Luber JM*, Tierney BT*, **Cofer EM**, Patel CJ, and Kostic AD. Aether: Leveraging linear programming for optimal cloud computing in genomics. *bioRxiv*, July 2017. doi:10.1101/162883 (*Contributed equally)

Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow P, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek S, Carpenter AE, Shrikumar A, Xu J, **Cofer EM**, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Gitter A[†], and Greene CS[†]. Opportunities and obstacles for deep learning in biology and medicine. *bioRxiv*, May 2017. doi:10.1101/142760. (Random ordering where not marked by [†])

Conference Presentations, Posters, and Talks

Cofer EM*, Zhang Z*, and Troyanskaya OG. AMBIENT: Accelerated convolutional neural network architecture search for regulatory genomics [platform talk]. *Machine Learning in Computational Biology (MLCB)*, November 2020. Virtual conference due to COVID-19 pandemic. (*Contributed equally)

Preuss C, Chen X, Chen KM, Theesfeld CL, **Cofer EM**, Uyar A, Cary GA, Pandey RS, Garceau D, Kotredes KP, Logsdon B, Mangravite LM, Howell GR, Sasner M, Troyanskaya OG, and Carter GW. Translating genetic risk variants in disease-associated enhancers into novel mouse models of Alzheimer's disease [poster]. *Alzheimer's Association International Conference*, July 2020. Virtual conference due to COVID-19 pandemic.

Cofer EM. DeepArk: Sequence-based models of *cis*-regulatory logic for model organisms [platform talk]. *The Allied Genetics Conference (TAGC) of the Genetics Society of America (GSA)*, April 2020. Virtual conference due to COVID-19 pandemic.

Cofer EM. Modeling regulatory codes with deep learning [talk]. *The Center for Statistics and Machine Learning Graduate Research Seminar*, March 2020. Princeton University. Princeton, NJ.

Cofer EM. Modeling the *cis*-regulatory codes of model organisms with deep learning [talk]. *Graduate Research Seminar for Quantitative and Computational Biology*, February 2020. Princeton University. Princeton, NJ.

Cofer EM and Troyanskaya OG. Deepark: Sequence-based regulatory models for model organisms [poster]. *Lewis-Sigler Institute for Integrative Genomics Retreat*, September 2019. Princeton University. Princeton, NJ.

Chen KM* and **Cofer EM*** and Zhou J and Troyanskaya OG. Selene: a pytorch-based deep learning library for sequence data [poster]. *International Conference on Research in Computational Molecular Biology*, May 2019. The George Washington University. Washington, DC. (*Contributed equally & co-presented)

Cofer EM and Troyanskaya OG. Predicting alternative splicing with deep neural networks [poster]. *NHGRI Research Training and Career Development Annual Meeting*, April 2019. Washington University in St. Louis. St. Louis, MO.

Cofer EM and Troyanskaya OG. Modeling tissue-specific pre-mRNA splicing with deep neural networks [poster]. *Lewis-Sigler Institute for Integrative Genomics Retreat*, November 2018. Princeton University. Princeton, NJ.

Cofer EM and Troyanskaya OG. Modeling RNA splicing with deep learning [poster]. *The Computational Systems for Integrative Genomics (CSIG) Conference*, July 2018. Broad Institute of MIT and Harvard. Cambridge, MA.

Cofer EM and Hibbs MA. Determining the age-associated somatic mutation rate in diverse mouse tissues [talk]. *The Computational Systems for Integrative Genomics (CSIG) Conference*, June 2017. Princeton University. Princeton, NJ.

Cofer EM. The rate of somatic mutations in aging mice [undergraduate thesis]. *Trinity University Computer Science Undergraduate Honors Thesis*, April 2017. Trinity University. San Antonio, TX.

Cofer EM. Accurately measuring the mammalian somatic mutation rate using deep learning [talk]. *Computer Science Colloquium*, March 2017. Trinity University. San Antonio, TX.

Cofer EM and Hibbs MA. Accurately measuring the mammalian somatic mutation rate using deep learning [platform talk & poster]. *Conference for Science at the Forefront of Basic and Translational Research*, October 2016. University of Texas at San Antonio. San Antonio, TX.

Cofer EM. Using deep learning to classify mutations [talk]. *Computer Science Colloquium*, September 2016. Trinity University. San Antonio, TX.

Cofer EM, Kennedy RM, and Hibbs MA. Accurately measuring the mammalian somatic mutation rate [poster]. *International Conference on Intelligent Systems for Molecular Biology*, July 2016. Orlando, FL.

Cofer EM and Hibbs MA. Using deep learning to classify mutations [talk]. *Trinity University Summer Undergraduate Research Conference*, July 2016. Trinity University. San Antonio, TX.

Cofer EM* and Witecki I*. Ontology construction: a means of improving data-driven journalism [talk]. *Computer Science Colloquium*, April 2016. Trinity University. San Antonio, TX. (*Contributed equally)

Cofer EM. Gene name disambiguation: a novel approach [talk]. *Computer Science Colloquium*, December 2015. Trinity University. San Antonio, TX.

Cofer EM, Kwessi EA, Nguyen HV, Nishikawa KA, and Jiang AX. Modeling the 2012 presidential election's battleground states [talk]. *Trinity University Summer Undergraduate Research Conference*, July 2015. Trinity University. San Antonio, TX.

Cofer EM*, Ybarra T*, and Glawe DD. Positive displacement meter performance [talk]. *Trinity University Summer Undergraduate Research Conference*, July 2013. Trinity University. San Antonio, TX. (*Contributed equally)

Software

2020 **DeepArk**, *https://github.com/FunctionLab/DeepArk*.
Developed a deep learning model capable of predicting epigenetic features from DNA sequences for model organisms such as the fruit fly.

2020 **DeepArk web server**, *https://DeepArk.princeton.edu*.
Constructed a public Django-based web server hosting DeepArk, my deep learning model of epigenetic features. The web server is free, GPU-accelerated, and provides users with a clean interface and clear documentation. GPU-based jobs are dispatched to a queue running on Amazon Web Services.

2018 **Selene**, *https://github.com/FunctionLab/selene*.
A PyTorch-based software library to enable and enhance the development of deep learning models of genomic sequences and other biological sequencing data.

2016 **Reddigest**, *https://github.com/evancofer/reddigest-scalafied*.
Collaborated with two other developers to create Reddigest, a web application based around Reddit. Reddigest filters out links on Reddit that a user has already viewed. Initially built with PHP/javascript/Apache HTTP, but rewritten in Scala/Scala.js/Play to improve maintainability, scalability, and overall quality.

Various **Miscellaneous Code and Projects**, *https://github.com/evancofer*.
Source code available upon request.