Courtney Shearer

☎ *(+1) 864 538 96 76*
✉ *courtney.a.shearer@gmail.com*
🖱 *www.courtneyshearer.com*

Education

2021 - 2025 **PhD in Systems, Synthetic, and Quantitative Biology**, *Harvard University*, Cambridge MA, USA.

2015 - 2019 **Bachelor of Science in Computer Science, Minor Genetics**, *Clemson University*, Clemson SC, USA, Magna Cum Laude, Honors College.

Experience

October 2025 - Present **Research Engineer**, *Google DeepMind*, London, UK.
- 9 month visit to the Genomics Team within the Science Unit

2021 - 2025 **PhD Candidate**, *Debora Marks*, Cambridge, MA, USA.
- Thesis title: Generative Evolutionary Models for Variant Effect Prediction
- Major Research Interests: Clinical Variant Effect Prediction, Therapeutic Design, DNA/RNA language modeling, Generative Sequence Models, Protein Language Models
- DeboraMarksLab: `https://www.deboramarkslab.com/`

August 2020 - August 2021 **Bioinformatics Engineer**, *CRISPR Therapeutics*, Boston, MA, USA.
- GACT Computational Biology Team Member
- Analyzed high throughput CRISPR screens for a variety of in-vivo experiments
- Tech Lead for incorporating Novaseq and UMI deduplication data handling in high throughput pipeline for Hybrid Capture site confirmation assay.
- Drastically Refactored high throughput pipeline for in vitro CIRCLEseq assay for detecting off target CRISPR/Cas9 cleavage
- Performed adhoc analysis for promoter screens, long range pcr, bulkRNAseq, inversion detection, variant analysis of plasmid libraries

June 2019 - July 2020 **Bioinformatics Engineer**, *Zymergen*, San Francisco, CA, USA.
- Building and Utilizing Genome Assembly/Annotation/Analysis Pipelines at scale
- Dockerized a plethora of bioinformatics tools to be used across the company
- Onboarded novel genomes of interest from the sequencing level to fully annotated genomes
- Wrote custom algorithms for performing genomic lift-overs.
- Co-developed WGS Variant pipeline for analyzing variants across many genomes in parallel
- Interfaced with Software Engineers, Domain Specialists, Product Managers
- Created Genome Analysis team as the inaugural Scrum Master
- Python, Bioconda, Pytest, Pandas, Numpy, React, Jenkins, Docker, AWS, Bash, CWL

2015 - 2019 **Bioinformatics/Machine Learning Researcher**, *Alex Feltus & Melissa Smith*, SC, USA.
- NGS analysis workflow development.
- Processed hundreds of SRA files for many organisms into co-expression networks.
- Developed software to compare multiple gene co-expression networks for evolutionary relatedness using C++/Qt/OpenGl.
- Developed deep learning algorithm to categorize cancer tumors based on RNAseq gene expression levels in numerous datasets, resulted in co first author publication.
- FCTL: `https://cufctl.github.io/#people`

Summer 2017, 2018 **Software Developer**, *Ancestry*, Lehi, UT / San Francisco, CA, USA.

- Full Stack Web Development
- Developed a Video Upload Service using Spring Boot to persist large video files to AWS S3 using SOLID principles and MVC design. Included Throttling and Circuit Breaking.
- Rewrote a Full Stack application from C# to Java. Decoupled the business layer logic from the controller into a service layer. Used ThymeLeaf, Spring Boot, and MySQL.
- Used Mock Frameworks(Mockito) with JUnit tests to increase code coverage from 45 % to 72 % for some critical code modules
- Conducted statistical analysis using a machine-learned XGBoost model.
- Built a monitoring tool to track changes in a live graph of cluster relationships.

Summer 2016 **Bioinformatics REU Student**, *Gavin Conant*, Missouri, USA.

- Determine different metabolic reactions between high and low RFI(Residual Feed Efficiency) groups that are important to feed efficiency for sheep.
- Data from collaborators was analyzed using R and Perl programming techniques.
- A poster presentation held at end of the summer at the MU Summer Research Expo.
- Resulted in Publication.

Publications

- **C. A. Shearer**, F. Teufel, R. Orenbuch, D. Ritter, A. Spinner, E. Xie, J. Frazer, M. Dias, P. Notin, D. S. Marks, LOL-EVE: Predicting promoter variant effects from evolutionary sequences, ICML GenBio Workshop Best Paper, Spotlight & Oral at MLCB 2025. Proceedings of Machine Learning Research MLCB 2025.
- E. Xie*, **C. A. Shearer***, R. Weitzman*, P. Notin, D. S. Marks, Homology Based Prompting for Promoter Sequence Generation, ICML GenBio Workshop (2025).
- R. Arora*, M. Angelo*, C. A. Choe*, **C. A. Shearer***, A. W. Kollasch, F. Qu, R. Weitzman, S. Gurev, E. Xie, D. S. Marks, P. Notin, RNAGym: Benchmarks for RNA Fitness and Structure Prediction. (2024).
- B. J. Livesey, M. Badonyi, M. Dias, J. Frazer, S. Kumar, K. Lindorff-Larsen, D. M. McCandlish, R. Orenbuch, **C. A. Shearer**, L. Muffley, J. Foreman, A. M. Glazer, B. Lehner, D. S. Marks, F. P. Roth, A. F. Rubin, L. M. Starita, J. A. Marsh, Guidelines for releasing a variant effect predictor, arXiv [q-bio.OT] (2024). http://arxiv.org/abs/2404.10807.
- R. Orenbuch, **C. A. Shearer**, A. W. Kollasch, H. D. Spinner, , T. A. Hopf, D. Franceschi, M. Dias, J. Frazer, D. S. Marks, Deep generative modeling of the human proteome reveals over a hundred novel genes involved in rare genetic disorders, bioRxiv (2023). https://doi.org/10.1101/2023.11.27.23299062.
- Chaudhari, Hemangi G, Penterman, Jon, Whitton, Holly J, Spencer, Sarah J, Flanagan, Nicole, Lei Zhang, Maria C, Huang, Elaine, Khedkar, Aditya S, Toomey, J Mike, **Shearer, Courtney A**, Needham, Alexander W, Ho, Tony W, Kulman, John D, Cradick, Thomas J, and Kernytsky, Andrew. Poor Evaluation of Homology-Independent CRISPR-Cas9 Off-Target Assessment Methods. CRISPR Journal. Dec 2020.
- Colin A. Targonski* and **Courtney A. Shearer***, Benjamin T. Shealy, Melissa C. Smith & F. Alex Feltus. Uncovering Biomarker genes with enriched classification potential from Hallmark gene sets. Nature Scientific Reports. July 2019.
- Rocky D Patil, Melinda J Ellison, Sara M Wolff, **Courtney A. Shearer**, Anna M Wright, Rebecca R Cockrum, Kathy J Austin, William R Lamberson, Kristi M Cammack, Gavin C Conant. Poor Feed Efficiency in Sheep is Associated with Several Structural Abnormalities in the Community of Metabolic Network of their Ruminal Microbes. Journal of Animal Science. May 2018.

Talks

- **Courtney A. Shearer** "Learning a language model across promoter evolutionary sequences for variant effect prediction" Spotlight Talk MLCB 2025 September 10th 2025
- **Courtney A. Shearer** "Learning a language model across promoter evolutionary sequences for variant effect prediction" ICML GenBio Spotlight Talk July 18th 2025
- **Courtney A. Shearer** "Learning a language model across promoter evolutionary sequences for variant effect prediction" Cambridge Wellcome Trust AIxBio June 23rd 2025
- **Courtney Shearer** "Protein Language Models for Clinical Variant Effect Prediction" Tutorial for Marinka Zitniks AI in Biomedicine course Spring 2025
- **Courtney A. Shearer** "Can we computationally predict variant effects so we can do fewer experiments???" at Broad Medical Population and Genomics Group Primer Talk November 21st 2024.
- **Courtney A. Shearer** "Learning a language model across promoter evolutionary sequences for variant effect prediction" at Broad Medical Population and Genomics Group June 6th 2024.
- **Courtney A. Shearer** & Sarah Gurev "Methods for Genomics in Medicine" at MIT as Guest Lecture for Machine Learning in for Healthcare. April 25th 2024.
- Anush Chiappino-Pepe, Lisa Dratva, **Courtney A. Shearer**, Huseyin Tas, Jorge A. Marchand, Kamesh Narasimhan, Russel V. Miranda, Alexandra I. Rudolph, and George M. Church. "A workflow for the systems-level analysis, design, and engineering of genomically recoded organisms". 2022 AIChE Annual Meeting.

Posters

- ASHG 2025 2025 "A Genomic Language Model for Zero-Shot Prediction of Promoter Variant Effects"
- MIT MOML 2024 "LOL-EVE: Predicting Promoter Variant Effects from Evolutionary Sequences"
- Mutational Scanning Sympossium 2024 "Learning a language model across promoter evolutionary sequences for variant effect prediction"
- Broad MPG 2024 & Lightening Talk "Uncovering promoter variant disease associations with machine learning"
- Broad MPG 2023 & Lightening Talk "Fitness prediction in noncoding regions using evolutionary data"

Awards/Scholarships Graduate

- 2023 NSF-GFRP Fellowship Recipient
- 2025 ICML GenBio Best Paper

Volunteer, Outreach, Mentorship

July 2019 - present **Volunteer**, *Research to the People*, CA, USA.

- Assist with a non-profit focused on organizing medical research cases in the bay area in hackathon style events, logo design, database management
- `https://www.researchtothepeople.org/`

July 2020 **Mentor**, *Codelabs*, USA.

- Mentored 3 students through a full stack web development project that had internships cut due to COVID-19 Pandemic
- Used Agile/Software Engineering Best Practices
- Held twice weekly meetings and 1:1 sessions

Oct 2017 - May 2019 **Director and Founder**, *CUhackit*, USA.

- Founder and 2017-2019 Director of CUhackit - Clemson University's first official Hackathon
- Led a team of 30 undergraduates to throw the largest hackathon in SC state history
- Helped the organization become a MLH official hackathon
- Helped raise over 60k in funding
- Managed 6 different core teams
- Held byweekly 1:1 meetings with leaders, general meetings, and leadership meeting
- `https://cuhack.it/`

May 2017 **Volunteer**, *Workshop in Computer Studies for Young Women*, Waterloo, Canada.

This workshop is geared at 9th/10th grade high school girls with CS interest. Students worked on programming with Processing, attended lectures, worked with hardware, and learned about opportunities for the future. I assisted with the girls throughout the entire week, helping them to complete activities and on their code.

Sept 2016 - Sept 2017 **Volunteer**, *Future Engineers*, Clemson, SC.

The Clemson Future Engineers group works in Elementary schools in the Clemson area to provide fun, interactive, and educational after school activities that revolve around topics such as computer science, physics, and engineering.

Awards/Scholarships Undergraduate

Undergraduate

- Outstanding Program of the Year: CUhackit (2019)
- Departmental Honors in Computer Science (2019)
- Velux Academic Excellence Scholarship (2018)
- Full Silver Scholarship to Open Data Science Conference West 2018 (2018)
- Finalist: Student Leader of the Year (2018)
- Clemson Running Club: Service Club of the Year (2018)
- Educational Enrichment Travel Grant Recipient (Summer 2017)
- Global Policy Scholars Fellowship ('16-'19)
- NSF REU: Biocomplexity (NSF-DBI-1358997)
- Finalist Tiger Speak Out Public Speaking Competition (2016)
- SC STEM Enhancement Scholarship ('15-'19)
- SC Palmetto Fellows Scholarship ('15-'19)
- Clemson Scholars Award ('15-'19)

Teaching Experience

- TF for Marinka Zitnik AI in BioMedicine Course @ HMS Spring 2025
- Inaugural TA for Advanced Biomedical Informatics Course @ Clemson.
- Head Tutor for CS for Women in Science and Engineering/FIRSTgen students.
- Lead Lecture series: Code and Code for First year students in Honors College.

Pre Graduate School Talks

- "Intro to APIs" - CUhackit 2020
- "Big Data and Biology Workshop" - Pearl Hacks 2019
- "BioHackathons" Podcast with PraxisAI - Youtube link

Events Moderated

- "How We Cracked the Code", Clemson University - Clemson, SC
- "HDBScan vs. Isolation Forests for anomaly detection", WIDS - Seattle, WA

Hackathons

Judge **CUhackit 2022**, *Clemson, SC*.
- Judged Various Categories Virtually due to COVID19

Judge **CUhackit 2021**, *Clemson, SC*.
- Judged Various Categories Virtually due to COVID19

Mentor/Judge **CUhackit 2020**, *Clemson, SC*.
- Judged Projects for "Best Use of Big Data" and assisted with participants.

Participant **MIT Hacking Medicine**, *San Francisco, CA*.
- Technical Therapies for Restless Leg Syndrome and Business Plan/Pitch for Invention.

Organizer **Research to the People**, *Children's Tumor Foundation/ NF2*, San Francisco, CA.
- Assisted with Project Sync for Researchers.

Director **CUhackit 2019**, Clemson, SC.
- Organized and directed the largest college hackathon in South Carolina history.

Mentor **PearlHacks**, Chapel Hill, NC.
- Held "Big Data in Biology Workshop" and assisted teams.
- `https://github.com/csheare/pearlhacks2019`

Participant **SheHacks Boston**, Boston, MA.
- Project: CarbonFoot, Recycling Application for Tracking Carbon Footprint

Director **CUhackit 2018**, Clemson, SC.
- Inaugural Clemson Only Hackathon with 100 participants.

Participant **Research to the People**, *Papillary Renal Cell Carcinoma*, San Francisco, CA.
- Analysis of tumor type and tumor stage, for three specific kidney tumor types.
- `https://github.com/SVAI/AIzheng`

Director **HelloWorld Hackathon 2018**, San Clemson, SC.
- Directed Clemson's first freshman only hackathon.

Participant **HackNC2017**, *Chapel Hill*, Chapel Hill, NC.
- Cinema Sentiments: `https://devpost.com/software/cinemasentiments`

Participant **HackGT 2017**, *Georgia Tech*, Chapel Hill, NC.
- Recipe Revealer: `https://devpost.com/software/recipe-revealer`

Participant **Mhacks X**, *Michigan State*, Ann Arbor, Michigan.
- Sendator: `https://devpost.com/software/sendator`

Participant **Boilermake 2017**, *Purdue*, West Lafayette, Indiana.
- Chicago Police Tracker: `https://tinyurl.com/y57ybo2u`
- 2nd Place overall

Participant **SFHacks 2017**, *Purdue*, San Francisco, CA.
- Forte Music Analyzer: `https://github.com/IrynaShv/4te`

Professional Memberships

- International Society Computational Biology Membership
- Harvard Therapeutics Graduate Program

Coursework

Computer Science

- Algorithms and Data Structures
- Design and Analysis of Algorithms
- Software Development Foundations
- Foundation of Computer Science
- Computer Systems Organization
- Network Programming
- Operating Systems
- 2-D Game Engine Construction
- Programming Systems

Maths/Statistics

- Calculus of One Variable I/II
- Calculus of Several Variables
- Linear Algebra
- Stats for Science and Engineering
- Applied Data Science
- Theory of Probability @ Clemson
- STAT 110 @ Harvard

Sciences

- Physics with Cal I/II
- General Biology
- General Chemistry I/II
- Molecular & General Genetics
- Survey of Organic Chemistry
- Biochemistry
- Introduction to Genomic Technologies
- Algorithms for DNA Sequencing
- Machine Learning in Genomics @MIT

Political Science/Ethics

- World of Ideas
- Global Policy Process
- Science of Ethics
- Migrant, Exiles, and Refugees

Other Talents/Interests

Basic	Swing Dancer, Violin, Guitar, Whistler
Intermediate	Vocals, Coffee, Backpacker/Camper, Rock Climber
Advanced	Long Distance Running

High Level Interests

Topic	Systems Biology, Protein Engineering, Biology Inspired Design, Synthetic Biology
Methods	Machine Learning, Genetic Algorithms, Deep Learning
Random	Learning Motivation, Creativity, Community Development, Human Psychology and Behavior, Artistic Expression, System Design, Minimalism, Effective Mentorship