

# Spencer Nystrom, PhD

✉ [nystromdev@gmail.com](mailto:nystromdev@gmail.com) |  [snystrom](https://github.com/snystrom) |  [NystromSpencer](https://twitter.com/NystromSpencer)

## Education

---

- PhD, Genetics & Molecular Biology** Aug. 2015 – June 2021  
*University of North Carolina* Chapel Hill, NC  
- Thesis: *Epigenetic Control of Developmental Timing by Temporal Transcription Factors*
- B.A. Biology (Cellular & Molecular Emphasis)** Aug. 2011 – May 2015  
*Austin College* Sherman, TX  
- German (Minor)  
- Magna Cum Laude

## Experience

---

- Data Scientist II** May. 2022 – December 2022  
*Locus Biosciences* Hybrid Remote / Research Triangle Park, NC
- Developed modular, distributed genomics analysis pipelines in the cloud using Nextflow and Google Batch
  - Implemented cloud security & data management policies for a newly developed clinical laboratory to meet HIPAA and 21 CFR pt. 11 compliance
  - Developed custom Terraform modules to manage cloud-native container registries, pipeline infrastructure, data warehouses, and company-wide access control policies
  - Designed and implemented CI/CD frameworks, code review best practices, and development standards for internal software development efforts in Python, Julia, R, and Rust
  - Developed end-to-end ETL workflows for data processing & visualization using Google BigQuery, dbt, and Tableau
  - Nominated for and completed 11-week Management & Leadership training program
  - Mentored junior team members on software development, data analysis, and security best-practices
- Data Scientist** June. 2021 – May 2022  
*Locus Biosciences* Hybrid Remote / Research Triangle Park, NC
- Started and led a new Systems Biology team focusing on assay design & data collection efforts to support dynamical systems modeling
  - Drafted regulatory documents for IND filings related to bioinformatic screening algorithms
  - Wrote custom genomics analysis pipelines for novel short-read (Illumina) and long-read (PacBio HiFi & Oxford Nanopore) genomics assays. Worked closely with bench scientists to design sample prep methods and assay conditions to ensure success
  - Built modular long read genome assembly pipeline for bacterial and phage genomes using PacBio or Oxford Nanopore data followed by short-read polishing using Illumina sequencing
  - Developed internal software packages in R, Python, and Julia to simplify development work with internal APIs, databases, and electronic lab notebooks
  - Strengthened reproducibility and stability by hardening software versioning & container usage best-practices in production pipelines
  - Improved internal codebase security by implementing credential management systems in production pipelines
  - Wrote custom software to support integration of laboratory equipment into end-to-end data capture framework
  - Filled *ad-hoc* analysis request tickets & wrote detailed summary reports for R&D team members and external partners

## Graduate Research Student

August 2015 – June 2021

Advisor: Dr. Daniel J. McKay

UNC Chapel Hill

- Integrated multiomic datasets to investigate molecular mechanisms of development, resulting in 2 co-first author publications, and 1 first author software publication
- Designed & built scalable, distributed analysis pipelines for reproducible processing of RNA-, ATAC-, FAIRE-, ChIP-seq, and CUT&RUN data using Snakemake, Python, & R
- Developed open-source software tools for genomics data analysis using R and Rust (see "Software" below)
- Independently designed and carried out wetlab experiments including: *Drosophila* genetics, *in-vivo* enhancer activity assays, genomics bench experiments & associated library preps
- Built automated reports of computational methods & results for non-expert recipients

## Skills

---

**Bioinformatics:** RNA-seq, CUT&RUN, ChIP-seq, ATAC-seq, genome assembly & annotation, variant calling, pangenomics

**Data Engineering:** Terraform, Docker, Google Compute Engine, Google Batch, Google Kubernetes Engine, Google BigQuery, Google Cloud Storage, dbt

**Programming Languages:** R, Python, Rust, Julia, Bash / UNIX commandline, SQL

## Software

---

### CRAN and Bioconductor Packages

*Maintainer*

- cmdfun: Build seamless commandline wrappers in R
- memes: an R interface to the MEME Suite

*Contributor:* universalmotif, plyranges

### Experimental / Works in Progress

- bamf: Manipulate sequencing reads by fragment size
- vplot: Summarize and visualize fragment size distributions at genomic loci

## Awards & Honors

---

Best Graduate Student Poster

UNC Genetics Retreat 2019

Best Talk

Triangle Fly Symposium 2018

Best Graduate Student Poster

Triangle Fly Symposium 2017

Best Undergraduate Student Poster

Texas Genetics Society 2015

Phi Beta Kappa

Austin College 2015

## Teaching

---

### R for Data Analysis

June - July 2018, June 2019

*How to Learn to Code Series*

UNC Chapel Hill

- Designed & taught course for graduate students and post-docs

### Learn to Code Python 3

June 2017 - July 2017

*How to Learn to Code Series*

UNC Chapel Hill

### ATAC-seq & ChIP-seq data analysis workshop

June 2017

*UNC Center for Bioinformatics*

UNC Chapel Hill

- Designed & led 3 day workshop teaching graduate-students, post-docs, and UNC faculty how to process and analyze ATAC-seq data
- Taught introduction to R programming, plotting, and other sequencing analysis tools
- Assisted teaching students to use UNC compute resources

## Publications

---

**Nystrom, Spencer L.** and Daniel J. McKay. Memes: A motif analysis environment in R using tools from the MEME Suite. PLOS Computational Biology, 17(9):e1008991, sep 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008991. URL <https://dx.plos.org/10.1371/journal.pcbi.1008991>.

Robin E. Harris, Michael J. Stinchfield, **Spencer L. Nystrom**, Daniel J. McKay, and Iswar K. Hariharan. Damage-responsive, maturity-silenced enhancers regulate multiple genes that direct regeneration in drosophila. eLife, 9:1–26, jun 2020. ISSN 2050084X. doi: 10.7554/eLife.58305.

**Spencer L. Nystrom\***, Matthew J. Niederhuber\*, and Daniel J. McKay. Expression of E93 provides an instructive cue to control dynamic enhancer activity and chromatin accessibility during development. Development (Cambridge), mar 2020. ISSN 14779129. doi: 10.1242/dev.181909.

Christopher M Uyehara\*, **Spencer L. Nystrom\***, Matthew J Niederhuber, Mary Leatham-Jensen, Yiqin Ma, Laura A Buttitta, and Daniel J McKay. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes and Development, 31(9):862–875, may 2017. ISSN 15495477. doi: 10.1101/gad.298182.117.

\* denotes equal contribution