

# SHANE CHU

555 Melville Ave, Apt 3N , St. Louis, MO 63130  
+1 917-640-7301 ♦ skchu@wustl.edu

## EDUCATION

---

- Ph.D., Computer Science** Aug 2018 - Present  
Washington University in St. Louis  
Advisor: Gary D. Stormo  
Thesis: Sparse representations for regulatory genomics
- Bachelor of Arts in Mathematics** Aug 2014 - May 2018  
University of Kansas

## WORK EXPERIENCE

---

- Predoctoral researcher** Jan 2019- Present  
Washington University in St. Louis
- Designed new models based on sparse representations for biological sequences.
  - Designed and implemented training algorithms for large-scale problems in regulatory genomics.
  - Designed and implemented fast inference techniques for
    - finding DNA sequence motifs that may exhibit patterns such as variable spacings, multimeric bindings, multiple binding domains, and alternative structural conformations.
    - finding the composition of the regulatory elements in functional DNA sequences.
  - Revealed new statistically significant motifs that show alternative binding modes in the JASPAR database.
- Graduate research assistant** Jan 2019 - Present  
Washington University in St. Louis
- Created several bioinformatics packages in Julia (available from the Julia registry).
  - Developed NGS pipelines using best practices.
- Research assistant** Nov 2017 - May 2018  
Center for Remote Sensing of Ice Sheets (CReSIS)
- Hyperparameter optimization for models on large-scale ice bottom detection.
  - Integrated parameter tuning subroutines to the CReSIS codebase to automate model selections.
- Research Intern** May 2017 - Aug 2017  
ZOLOZ
- Constructed parallel data processing pipelines for data collections.
  - Integrated ZOLOZ's software development kit into web and mobile technologies.

## TEACHING EXPERIENCE

---

- Assistant instructor** Jan 2021 - Jun 2022  
Washington University in St. Louis
- Held weekly recitations and office hours for mathematical optimization (ESE 415).
  - Prepared lecture notes and office hours for Bayesian methods in machine learning (CSE 515).

## PUBLICATIONS

---

- In preparation:**
- S Chu, GD Stormo, "Interpretable Functional DNA Classifications and Regressions With Sparse Representations"
- Preprints:**

- *S Chu, GD Stormo, “Deep Unfolded Convolutional Dictionary Learning for motif discovery”, bioRxiv 2022.11.06.515322; doi: <https://doi.org/10.1101/2022.11.06.515322>*

#### Published:

- *JD Paden, V Berger, M Al-Ibadi, S Chu, M Xu, D Crandall, G Fox, “Subglacial bed topography using machine learning and geostatistical analysis applied to 2D and 3D radar sounding”, Advancing Earth and Space Science, Lawrence, Kansas, KS, USA 2018.*
- *V Berger, M Xu, S Chu, D Crandall, J Paden, G Fox, “Automated Tracking of 2D and 3D Ice Radar Imagery Using Viterbi and TRW-S,” IEEE International Geoscience and Remote Sensing Symposium, Lawrence, Kansas, KS, USA 2018. doi: 10.1109/IGARSS.2018.8519411.*

#### CONFERENCE PRESENTATIONS

---

- *Poster, Machine Learning in Computational and Systems Biology, International Society for Computational Biology (ISMB) July 2022*

#### SOFTWARE ENGINEERING SKILLS

---

- *Languages: Python, Julia, C/C++ , Matlab, R, JavaScript,  $\text{\LaTeX}$ .*
- *High-performance computing: CUDA.jl, Slurm*
- *Pipeline Development: Snakemake*
- *Environments: Linux (Ubuntu)*
- *Mathematical Tools: FFTW*
- *Deep Learning: Flux.jl, PyTorch*
- *Visualizations: Makie.jl, Luxor.jl, D3.js, HTML*

#### OPEN SOURCE CONTRIBUTIONS

---

**Github:** <https://github.com/kchu25>

#### Packages:

- ***UnfoldCDL.jl** Motif discovery with deep unfolded convolutional dictionary learning.*
- ***SeqShuffle.jl** Shuffle strings such that it preserves the k-mer frequency in each string.*
- ***MotifPvalue.jl** Implemented an approximation algorithm that estimates the score thresholds of position weight matrices.*
- ***SimDNA.jl** Create DNA strings that simulate motifs with variable spacing, multimeric bindings, and multiple binding domains.*

#### KEY COURSES

---

- *Washington University in St. Louis: Large-scale optimization for data science, machine learning, Bayesian methods for machine learning, Bayesian inference, advanced algorithms.*
- *University of Kansas: Operating systems, parallel computing, numerical analysis, mathematical analysis, genetics, microbiology.*

#### LANGUAGES

---

*English and Mandarin.*