

SHANE K. CHU

555 Melville Ave, Apt 3N , St. Louis, MO 63130
+1 917-640-7301 ♦ skchu@wustl.edu ♦ kchu25.github.io

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
- Overcome the limitation of traditional motif discovery methods and revealed a substantial presence of cooperative binding patterns and transposable elements in in-vivo datasets such as ChIP-Seq from databases such as JASPAR, FactorBook, and ReMap.
 - Designed new deep learning models with inherent interpretability based on hierarchical sparse representations for biological sequences.
 - Finding DNA sequence to expression map using Bayesian nonparametrics and design an optimization framework for counterfactual inference.
- 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).

SELECTED PUBLICATIONS

Published:

- Chu, S. K., & Stormo, G. D. (2023). Finding motifs using DNA images derived from sparse representations. *Oxford Bioinformatics*, 39(6), btad378.
- Paden, J. D., Berger, V., Al-Ibadi, M., Chu, S. K., Xu, M., Crandall, D., & Fox, G. (2018, December). Subglacial bed topography using machine learning and geostatistical analysis applied to 2D and 3D radar sounding. In *AGU Fall Meeting Abstracts* (Vol. 2018, pp. C43A-07).

- Berger, V., Xu, M., Chu, S. K., Crandall, D., Paden, J., & Fox, G. C. (2018, July). Automated tracking of 2D and 3D ice radar imagery using VITERBI and TRW-S. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 4162-4165). IEEE.

Preprints:

- Chu, S. K., & Stormo, G. (2022). Deep unfolded convolutional dictionary learning for motif discovery. *bioRxiv*, 2022-11.

In preparation:

- S Chu, GD Stormo, "Interpretable Functional DNA Classifications and Regressions With Sparse Representations"

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, \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:

- **MOTIFs.jl** A motif discovery method that is capable of identifying long, gapped, and cooperative binding patterns, in addition to finding the primary binding sites.
- **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.