

HAIDONG YI

600 MARTIN LUTHER KING JR BLVD APT 121, Chapel Hill, NC, 27514

+1 984-215-1411 ✧ haidyi@cs.unc.edu

EDUCATION

University of North Carolina at Chapel Hill

Aug. 2019 - present

Ph.D. Candidate in Computer Science, **Expected Graduate Date: May, 2024**

Teaching Assistant of COMP411 (Compute Organization)

Advisor: [Junier B. Oliva](#)

Nankai University

Sept. 2016 - Jun. 2019

Master student in College of Computer Science

Advisor: [Han Zhang](#) & [Zhenglu Yang](#)

Nankai University

Sept. 2012 - Jun. 2016

B.E. in Intelligent Science and Technology¹,

B.S. in Mathematics and Applied Mathematics² (Co-Major)

Advisor: [Han Zhang](#)¹ & [Changliang Zou](#)²

PUBLICATIONS

Peer reviewed publications

- **Haidong Yi**, Ayush T. Raman, Han Zhang, Genevera I. Allen, Zhandong Liu, "Detecting hidden batch factors through data-adaptive adjustment for biological effects", *Bioinformatics*, 2018, 34(7):1141-1147.
- Huang Le, Han Zhang, Peizhi Wu, Entwistle Sarah, Li Xueqiong, Yohe Tanner, **Haidong Yi**, Yang Zhenglu, Yanbin Yin, "dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation", *Nucleic Acids Research*, 2018, 46(Database issue): D516-D521.
- Liu Wei, Ma Shunjian, Sun Mingwei, **Haidong Yi**, Wang Zenghui, Zengqiang Chen, "Sequential quadratic programming-based fast path planning algorithm subject to no-fly zone constraints", *Engineering Optimization*, 2015, 48(8):1401-1418.

AWARDS

- [Meritorious Winner](#) of Interdisciplinary Contest In Modeling Certificate of Achievement (ICM), 2015.
- The First Prize of Excellent Graduate Student (**College Top 3%**), Nankai University, 2018, RMB 8000.
- Honor in recognition of Excellent Undergraduate Thesis (**College Top 5%**), 2016, RMB 3000.
- The Third Prize as an excellent project in the "100 Projects" of Creative Research for an undergraduate at Nankai University, 2015 (**College Top 1**, Funding: RMB 5000).
- The Second Prize of Excellent Undergraduate Scholarship, Nankai University, 2014, RMB 3000.
- National Encouragement Scholarship, 2013, RMB 5000.

RESEARCH EXPERIENCE

Developing Deep Neural Networks for Density Estimation.

Sept. 2019 - Now

- Develop a new neural network combining the flow model and autoregressive model for density estimation (in progress).

Develop computing tools to identify anti-CRISPR (Acr) and anti-CRISPR associated (Aca) loci in gene sequence dataset

Sept. 2018 - Jun. 2019

Project Leader

- Identify the Acr and Aca with homology-based and guilt-by-association methods;

- Developed a bioinformatics tool using Python and easily installed via Docker;
- Developing a web server using Google Compute Engine (To appear).

Detect Batch Factors in High Dimensional Gene Expression Dataset

Sept. 2016 - Sept. 2017

Project Leader

- Modeled a data-adaptive method to estimate biological effects using convex optimization.
- Detected and classified batch factors using semi-nonnegative matrix factorization method.
- Developed an R package, DASC, to identify and classify batch factors.

Visualization of Convex Bi-clustering

July 2015 - Sept. 2015

Summer Research, advised by Zhandong Liu and Genevera I. Allen

- Wrote an R package to visualize the process of convex bi-clustering.

Design Constraint Conditions of Arc No-fly Zones

Sept. 2013 - Sept. 2014

Project Leader

- Modeled a constraint of arc no-fly zone using line segment approximation.
- Expanded a C++ software to solve the constrained optimization using sequential quadratic programming.

dbCAN-seq: a database of CAZyme sequence and annotation ([website](#))

Apr. 2017 - Jun. 2017

Project Member

- Preprocessed and annotated a part of data using HMMER.
- Optimized the speed of querying in MySQL using the search engine, Sphinx.

SKILLS

Development Languages:	Python, R, C, C++, \LaTeX
Development Framework:	Tensorflow, PyTorch, OpenMP
Platforms	Linux/Mac

SOFTWARES

DASC

[Link to Github Repository](#)

An R package used for identifying batches and classifying samples into different batches in high dimensional gene expression dataset.

Acrfiner

[Link to Github Repository](#)

An python code used for identifying anti-CRISPR (Acr) and Acr-associated (Aca) genomic loci in gene sequence data.