

Haotian Teng

E-mail: haotiant@cs.cmu.edu

Personal Website: <https://haotianteng.github.io/>

RESEARCH EXPERIENCE

Detecting N6-methyladenosine (m6A) modification in RNA using long read sequencing

Jan 2021 – Present

Advisor: Prof. Ziv Bar-Joseph & Prof. Carl Kingsford, Computational Biology Department, School of Computer Science, Carnegie Mellon University

- Developed a hybrid model that combine non-homogeneous Hidden Markov Model (NHMM) and Convolutional-Recurrent Neural Network (CRNN) to do unsupervised training on ONT sequencing signal to detect m6A modification in RNA.

Clustering and visualize the single-cell spatial transcriptomics data.

Aug 2019 - Feb 2022

Advisor: Prof. Ziv Bar-Joseph

- Developed a Gaussian-Multinomial mixture probabilistic graphical model to clustering the cell from single-molecule Fluorescence in situ hybridization (FISH) data (Project page: <https://github.com/haotianteng/FICT>).
- Developed a denoise auto-encoder for profiling gene expression. (Project page: <https://github.com/haotianteng/GECT>).
- Produce a simulation pipeline based on Metropolis-Hasting to validate the spatial transcriptomics clustering tools and benchmark&visualize the clustering result. (Project page: <https://github.com/haotianteng/FICT-SAMPLE>)

Using Machine learning algorithm in Nanopore long-read sequencing Basecalling

Feb 2017 - Jul 2018

Advisor: Prof. Lachlan Coin, Institute for Molecular Bioscience, University of Queensland

- Built a deep learning basecaller **Chiron** using Tensorflow, for Oxford Nanopore sequencer basecalling (Project page: <https://github.com/haotianteng/Chiron>)
- Developed a segmentation tool **BoostNano** to identify the polyA region in the Nanopore RNA sequencing platform. (Project page: <https://github.com/haotianteng/BoostNano>)
- Prepared training dataset of DNA and RNA Nanopore basecalling reads, perform data labeling and data washing.
- Implemented a pipeline in Google Cloud and Google Compute Engine for end-to-end genome analysis.

PROFESSIONAL EXPERIENCE

Bioinformatics Engineer @ Novogene Europe

Sep 2018 - Feb 2019

- Optimized the human resequencing and laboratory automation pipeline cooperatively in a medium size group.
- Designed and developed the long-read sequencing platform.

PUBLICATIONS

- Teng, H.**, Yuan, Y., & Bar-Joseph, Z. (2021). Clustering Spatial Transcriptomics Data. *Bioinformatics*, 38(4), 997-1004.
- Pitt, M. E., Nguyen, S. H., Duarte, T. P., **Teng, H.**, Blaskovich, M. A., Cooper, M. A., & Coin, L. J. (2020). Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA Nanopore sequencing. *GigaScience*, 9(2), g1aa002.
- Teng, H.**, Cao, M. D., Hall, M. B., Duarte, T., Wang, S., & Coin, L. J. (2018). Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. *GigaScience*, 7(5), g1y037.
- Avitan, L., Pujic, Z., Mölter, J., Van De Poll, M., Sun, B., **Teng, H.**, Amor, R., Scott, E.K. and Goodhill, G.J., 2017. Spontaneous activity in the zebrafish tectum reorganizes over development and is influenced by visual experience. *Current Biology*, 27(16), pp.2407-2419.
- Teng, H.** "A neuron-muscle circuit model of *C.elegans*'s locomotion." *Bachelor of Science Thesis: Peking University*, 2015

EDUCATION

Carnegie Mellon University, Pittsburgh, United States

Ph.D., Computational Biology

- Advisor: Ziv Bar-Joseph, Machine Learning Department and Computational Biology department, School of Computer Science, Carnegie Mellon University.

2020 - present

Core Courses: Computational Genomics 02-710, Machine Learning 10-701, Deep Reinforcement Learning 10-703, Probabilistic Graphical Models 10-708, Convex Optimization 10-725, Advanced NLP 11-711

University of Queensland, Queensland, Australia

M.S., Bioinformatics

2016 - 2019

Peking University, Beijing, China

B.S., Physics

2011 - 2015

SKILLS

- Programming: Python, C++, Bash, MATLAB, R
- Packages&Platforms: Tensorflow, PyTorch, OpenGL, BWA, SAMtools, Velvet.