**Haotian Teng**

E-mail: [havens.teng@gmail.com](mailto:havens.teng@gmail.com) haotiant@andrew.cmu.edu

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

**EDUCATION**

**Carnegie Mellon University (CMU)**,Pittsburgh, PA, United States

*Doctor of Philosophy, Computational Biology, School of Computer Science (SCS)*

Advisor: Ziv Bar-Joseph, Machine Learning and Computational Biology department, SCS, CMU. 2019 - 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 Natural Language Processing 11-711

**University of Queensland**, Brisbane, Queensland, Australia

Master of Bioinformatics 2016 - 2018

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

Advisor: Prof. Geoffrey Goodhill, Queensland Brain Institute, University of Queensland

**Peking University**, Beijing, China

Bachelor of Science, Physics 2011 - 2015

**RESEARCH EXPERIENCE**

**Ph.D. Research Projects, Carnegie Mellon University**

**Target-based drug generation using graph neural networks with probabilistic diffusion model.** Nov 2022 – Present

* Adapted an E(3) equivalent graph neural network with probabilistic diffusion model to generate small molecules with high binding affinity to given target protein pocket and high druggability.

**Detecting N6-methyladenosine (m6A) modification in RNA using long read sequencing**  Jan 2021 – Jan 2023

* Introduced Xron, an interpretable hybrid deep learning model combining non-homogeneous Hidden Markov Model (NHMM) and Convolutional-Recurrent Neural Network (CRNN) that achieve accurate detection of m6A modification in RNA. Obtained a 22% and 7% relative improvement on Yeast and human HEK293T cell lines compared to state-of-the-art.
* Obtained informative chemical parameters of nanopore from the highly interpretable hybrid model. Constructed pore models based on the parameters and completed accurate signal segmentation in an unsupervised manner.

**Clustering the single-cell spatial transcriptomics data with novel probabilistic graphical model.** Aug2019 - Feb 2022

* Developed a robust probabilistic graphical model [FICT](https://academic.oup.com/bioinformatics/article/38/4/997/6384569) that clusters the cell for spatial transcriptomics data. By considering the spatial distribution in addition to the gene expression, the model can give cell clusters variated in spatial domain.
* Discovered 5 novel genes related to neuron differentiation by DGE analysis on clusters identified by FICT.
* Constructed regularized denoise auto-encoder that achieve dimensional reduction faster than UMAP and applicable to any dimensions. It disentangles complicated clusters failed by t-SNE which does not consider nonlinear transformations.

**Neural architecture search using deep reinforcement learning** July2022 - Sept 2022

* Identified 3 novel convolutional neural layer structures using neural architecture search, achieved better performance than usual residual connection layer on several common tasks in openAI gym library with 20% fewer parameters.

**Masters Research Project, Institute for Molecular Bioscience, University of Queensland**

**Using Machine learning algorithm in Nanopore Basecalling** 2017-02 - 2018-07

**Advisor: Prof. Lachlan Coin**

* Built a deep learning-based basecaller **Chiron** using Tensorflow, for Oxford Nanopore sequencer basecalling
* Developed a preprocessing tool **Nanopre** to identify the polyA region in the Nanopore RNA sequencing platform.
* Prepared training dataset of DNA and RNA Nanopore basecalling reads, using Nanoraw and Graphmap to label the data.
* Implemented a pipeline in Google Cloud and Google Compute engine for end-to-end genome analysis.

**Internship and Winter Scholar, Queensland Brain Institute, University of Queensland**

**The development of spontaneous neural activity in the zebrafish** 2016-03 - 2017-02

**Advisor: Prof. Geoffrey Goodhill,**

* Built a pipeline for laboratory automation and data analysis in Zebrafish neuron experiment with Arduino, LabVIEW, and MATLAB.
* Constructed PHANTOM toolbox for projecting visual stimulation with conformal transformation, used for zebrafish tectum research.
* Developed algorithms for functional connectivity reconstruction using regularization method under scale-free assumption, correct the false positive correlation due to common input, transition connection, and latent common input.

**Internship, Feedback in AIY neurons in Thermotaxis behavior of C.elegans** 2015-07 - 2015-12

**Advisor: Prof.Aravinthan D.T. Samuel, Center for Brain Science, Harvard University, Boston**

* Studied thermotaxis in C.elegans with tracking and multi-neuron fluorescent marked.
* Cross & keep the worm, experiment using a spinning disk confocal microscope and the afterward data acquisition & processing with the combination of ImageJ (Miji) and Matlab
* Proved the derivation dependence between AFD neuron and temperature, designed and conducted the experiment to measure the parameters of the AFD-temperature relationship with temperature signal input under different shapes.

**Research Assistant, Locomotion and PH sensoring mechanism in C.elegans & fast reaction tracking System development**

2012-07 - 2015-06

Advisosr: Dr. Louis Tao, Center for Bioinformatics, Peking University, Beijing

* Marked GCaMP6 into the C.elegans ASH, AWC, ASE neurons to testify and determine the neuron responsible for PH sensoring.
* Developed a neuro-muscle model of C.elegans motor system and proved the theoretical prediction of gait adaptation in C.elegans.
* Recorded and analyzed long-term locomotion parameters of C.elegans by using a tracking and photographing system.
* Developed a visualization tool with openGL to describe and simplify the neuron network in C.elegans, and enabled the tool to search the whole neural pathway through any two given neurons.
* Built a tracking system as one of the contributors, which could achieve high-precision (accuracy below 1 micron) tracking and photographing and simultaneous data collection & processing
* Modified and developed a “snake” model based algorithm for robust and precise C.elegans center line extraction.

**Professional EXPERIENCE**

**Algorithm Engineer Winter Intern** 2019-01 - 2019-02

**Alibaba, Hangzhou, China**

* Intelligence cache prediction using deep learning model based on biometric information.

**Bioinformatics Engineer**

**Novogene Europe, Beijing, China** 2018-09 - 2019-01

* Optimized the human resequencing and laboratory automation pipeline.
* Designed and developed the long-read sequencing platform.

**Senior Research Technician**

**Institute for Molecular Bioscience, University of Queensland, Australia** 2017-06 - 2018-07

* Worked on Oxford Nanopore Technologies Long-read Nanopore direct RNA sequencing data processing, improve the sequencing accuracy and efficiency, improve the succeeded sequencing reads ratio by 15X compared to the original pipeline for long poly-A tail reads.

**Intern** 2014-07 - 2014-10

**Biodynamic Optical Imaging Center, PKU, Beijing, China**

* Micro-fluid chip preparation and fabrication.
* Developed a Computational Fluid Dynamics (CFD) module for the microfluid chips fluid field calculation in Fluent, which could draw the flow field from the CAD design sketch.

**PUBLICATIONS**

* **Teng, H.**, Yuan, Y. and Bar-Joseph, Z., 2021. Clustering Spatial Transcriptomics Data.*Bioinformatics.*
* 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), giaa002.*
* **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), giy037.*
* 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*

**HONORS AND AWARDS**

* The 1st Prize at 27th Chinese Physics Olympiad, Zhejiang Province (rank 1/1232 in theory part) 2011
* The Silver Medal at 27th Chinese Physics Olympiad, Finals 2011
* The 1st Prize at 29th Parts of the National College Students Physics Competition 2012

**SKILLS**

* Programming: Python, C, C++, Matlab, R, Linux, LaTeX,
* Packages&Platforms: Tensorflow, MXNet, Caffe, CUDA, cuDNN, OpenGL, BWA, SAMtools, Velvet, DIAMOND, BLAST+, Minimap2, H5py, Psychtoolbox, LabVIEW, Arduino.
* General Machine Learning: Linear Regression, Logistic Regression, SVM, Decision Tree, Random Forest, Probabilistic Graphical Models, RBM, GMM, KNN, K-Means, PCA, t-SNE, UMAP, Bagging, AdaBoost, Gradient Boosting.
* Deep Learning: CNN, RNN, LSTM, GMM, Diffusion model, GAN, VAE, BERT, Transformer, deep Q-learning
* Software: PyMOL, Fluent(ANSYS), Origin, AutoCAD, Primer Premier, DNA Man, Microsoft Office,
* Wet-lab experiment skill: Molecular cloning, Microinjection
* Language: Chinese(Mother Language), English(Fluent), Spanish(basic), German (Pizza-orderable)

TOEFL: Cumulative 103 (R 29, L 29, S 23, W 22); GRE: V 150, Q 169, AW 3.0

* Proficient in Piano playing, learned since 6 years old. Skillful in saxophone.