

Chen-Hao Chen, MD, PhD

88 Linden Street, Boston, MA 02134, United States | +1-650-867-7639 (Mobile) | chen-hao_chen@dfci.harvard.edu

EDUCATION

Harvard University

Doctor of Philosophy in Biological and Biomedical Science (Concentration: Computational Biology)

- **Thesis advisor:** Professor Xiaole Shirley Liu (Biostatistics and Computational Biology)

Cambridge, Massachusetts

Aug 12 – May 17

National Taiwan University

Doctor of Medicine

- Graduated with a cumulative GPA of 3.95/4.00 and 280-semester credits

Taipei, Taiwan

Sep 03 – Jun 09

SELECTED HONORS AND CERTIFICATES

Educational Commission for Foreign Medical Graduates (ECFMG) Certificate

U.S.A.

Issued by Educational Commission for Foreign Medical Graduates

Aug 17

- Qualified physicians graduating from foreign medical school who can enter programs of graduate medical education in the States
- Passed United States Medical Licensing Examination Medical Science (Step 1 & Step 2 CK) and Clinical Skills (Step 2 CS)

Doctor Certificate

Certified by the Department of Health (Taiwan)

- Qualified to practice medicine in Taiwan

Taipei, Taiwan

Aug 09

34th International Physics Olympiad

Gold Medal

- Awarded to the top 6% of participants
- Attended by more than 300 participants from more than 60 countries

Taipei, Taiwan

Aug 03

4th Asian Physics Olympiad

Gold Medal

Most Creative Solution in Theory

- Awarded to the participant who developed the most creative solution for theory problems.

Bangkok, Thailand

Apr 03

SELECTED WORK AND TEACHING EXPERIENCES

Harvard University

Teaching fellow

- Harvard STAT316: Deep Learning and Regulatory Genomics
- Lectured latent variables and derivation of latent representation using deep learning: word embedding in natural language processing and auto-encoder in image compression.
- Designed assignments to implement convolutional neural networks and transformer structures to predict genomic activities from DNA sequence.

Cambridge, Massachusetts

Jan 19 – May 19

Harvard Innovation and Commercialization

Winner Team

- Developed and pitched the business idea to venture capitalists

Boston, Massachusetts

Sep 17 – Nov 17

Department of Pathology, Brigham and Women Hospital

Observership

- Participated in anatomical pathology and molecular pathology diagnosis
- Used lasso regression to deconvolute mutation signatures from exome sequencing data to infer the origin of carcinogens.

Boston, Massachusetts

Jul 17 – Aug 17

National Taiwan University Hospital (NTUH)

Medical Intern (General Clinical Practice)

- Received rigorous clinical training at the best medical center in Taiwan with rotations at local hospitals and rural clinics
- Selected for Accelerating Program: students can finish seven-year medical education within six years

Taipei, Taiwan

Jun 07 – May 09

SELECTED RESEARCH EXPERIENCES

Harvard University

Ph.D. study and Research associate

Professor Xiaole Shirley Liu's Group (Department of Data Sciences/Dana Farber Cancer Institute)

- Constructed a data-driven mathematical framework of gene regulation ([publication 1](#))
 - **Aim:** construction of a succinct framework in interpreting gene regulation and characterization of common properties between gene regulators.
 - Mathematical model of transcriptional regulation combining multiple high-throughput genomic data. Applied maximum likelihood and non-linear distance correlation for model selection and inferring regulatory ranges of gene regulators
 - Identification and characterization of **two** types of gene regulators with long-range and short-range regulatory distances and other biological properties, including distinct binding sites preferences and auto-regulatory properties.
- Imputed regulator binding sites from single-cell ATAC-seq data ([publication 4](#))

Cambridge, Massachusetts

Aug 13 – Present

- Aim: since no available experimental method to profile gene regulator genomic binding in single cells, we computationally impute them from single cell chromatin accessibility (i.e., ATAC-seq) data.
- Denoising auto-encoder approach to extract chromatin accessibility features.
- Transformer-like structure combining sequence features (using convolution layer), cell-type embedding, and relative positional embedding to predict chromatin accessibility.
- Constructed a unified genomic “coordinate system” using a compendium of public genomic data (ongoing)
 - Aim: a genomic version of “knowledge graph”: joint mapping of public genomic profiles into low-dimension manifold with annotated cell identities or tissue origins. The resulting annotated genomic “coordinate system” allows for fine tracing of cell states transitions (e.g., from epithelium-like to neural-like) along with development, disease pathogenesis, and drug treatment.
 - Selection of genomic features with rich cell-state specific information.
 - A variational auto-encoder model to project genomic profiles (i.e. binary binding or gene regulation) into a low-dimensional cell-state space, a manifold parameterized by multi-variant Gaussian distribution.
 - Map of all available data comprising cell state transitions and inference of the feasible transition paths along the manifold (e.g., what kind of cell state transitions are permitted).
- Developed statistical model and computational pipelines for genome-wide CRISPR-Cas9 knockout screens ([publication 3](#))
 - Aim: an accurate analytic pipeline of genome-wide knock screens
 - Elastic net regression to predict CRISPR-Cas9 knockout efficiency using sequence features of single guide-RNA.
 - Gaussian mixture model to characterize the on-target and off-target activities of CRISPR-Cas9 knockout, and further characterization of the genomic basis underlying off-target activity
 - Proposal of a new normalization strategy to decrease false-positive rates in assigning genes essential cell survival.
- Investigated non-canonical mechanisms of Enhancer of zeste homolog 2 (EZH2) in advanced prostate cancer ([publication 2](#))
 - Aim: dissect the mechanism of EZH2 in driving prostate cancer tumorigenesis.
 - Discover that EZH2 regulates DNA damage repair in a non-canonical manner
 - Predict and validate synergistically therapeutic effect of EZH2 inhibitor and PARP1 inhibitor in advanced prostate cancer

SELECTED PUBLICATIONS

- **Chen-Hao Chen**, Myles Brown, Clifford A. Meyer, X. Shirley Liu. Determinants of transcription factor regulatory range. *Nature Communications*. In revision. 2019
- Yiji Liao*, **Chen-Hao Chen***, Mei Yang, Tengfei Xiao, Philip W. Kantoff, Henry Long, Steven P. Balk, X. Shirley Liu, Myles Brown, Kexin Xu. Pharmacological EZH2 inhibition enhances cancer cell sensitivity to genotoxic insults through suppressing DNA damage repair. *Genes and Development*. In review. 2019[* Equal contribution]
- **Chen-Hao Chen***, Tengfei Xiao*, Han Xu, Peng Jiang, Clifford A. Meyer, Wei Li, Myles Brown, X. Shirley Liu. Integrative analysis and refined design of CRISPR knockout screens. *Bioinformatics*, bty450. 2018 [* Equal contribution]
- **Chen-Hao Chen***, Jingyu Fan*, Clifford A. Meyer, X. Shirley Liu. S. Imputation of transcription factor binding profiles from single cell ATAC-seq using the self-attention model. In preparation.
- Wei Li, Johannes Koster, Han Xu, **Chen-Hao Chen**, Tengfei Xiao, Jun S. Liu, Myles Brown, X. Shirley Liu. Quality Control, Modeling, and Visualization of CRISPR Screens with MAGeCK-VISPR. *Genome Biology*. 2015.
- Han Xu, Tengfei Xiao, **Chen-Hao Chen**, Wei Li, Clifford A. Meyer Qiu Wu, Di Wu, Le Cong, Feng Zhang, Jun S. Liu, Myles Brown, X. Shirley Liu. Sequence determinants of improved CRISPR sgRNA design. *Genome Research*. 2015; gr.191452.115

SELECTED PRESENTATIONS

“Determinant of transcription factor regulatory ranges.”

Center for Functional Cancer Epigenetics (CFCE) seminar

Oct 25, 19

“Auto-encoder approach to decipher the basic unit of DNase landscape.”

The annual meeting of the Center of Functional Cancer Epigenetics (CFCE)

Aug 13, 19

COURSES

Statistics and machine learning: Statistical inference, Machine learning, Deep learning, Stochastic process

Mathematics: Calculus, Linear Algebra, Differential equations, Multivariate analysis

Medicine and biology: Genetics, Quantitative genomics, Computational biology and bioinformatics, Pathology, Physiology, Immunology, Neurology, Internal medicine, Microbiology, Anatomy, Radiology.

SKILLS, AND OTHERS

Computational languages: Python, R, Bash

Tools: PyTorch, RStudio, git

Github: <https://github.com/hyalin1127>