

# RUI YIN

Assistant Professor, University of Florida, Gainesville, FL 32603

Mobile: (+1)6172302038 E-mail: ruiyin@ufl.edu

## EMPLOYMENT

---

**Assistant Professor, USA** Dec.2022 - present  
Department of Health Outcome and Biomedical Informatics, University of Florida

**Research Postdoctoral Fellow, USA** Aug.2020 - Dec.2022  
Department of Biomedical Informatics, Harvard Medical School  
Advisor: Paul Avillach

## EDUCATION

---

**Ph.D. Nanyang Technological University, Singapore** Jun.2020  
Major in Biomedical Informatics, School of Computer Science and Engineering  
Advisor: Kwoh Chee Keong, Associate Professor

**M.S. Central South University, China** Dec.2015  
Major in Automation, School of Automation  
Advisor: Yang Chunhua, Professor, IEEE Fellow

**B.S. Shandong University, China** Jun.2013  
Major in Automation, School of Control Science and Engineering

**B.S. South China University of Technology, China (exchange)** Jul.2011  
Majoring in Automation, School of Automation Science and Engineering

## RESEARCH INTERESTS

---

- Computational modeling in RNA viruses (influenza, coronaviruses, zika, etc.)
- Machine learning modeling to facilitate the diagnosis of rare/undiagnosed diseases
- Genotype-phenotype association study for biomarker identification and novel discovery
- Combination of clinical informatics and genomics to support clinical decision system
- Interpretable and trustworthy AI to advance health equity

## RESEARCH EXPERIENCES

---

- **Postdoctoral Fellow**, Harvard Medical School Aug.2020 - Dec.2022
- Project: AI-assisted framework to accelerate the diagnosis of rare and undiagnosed diseases based on patient-centered information (on-going)
  - a. Designed data-centric methods to integrate genotypic and phenotypic data from patients
  - b. Explored the causal genes and variants for the rare and undiagnosed diseases
  - c. Developed machine learning models for predictive tasks (e.g., variant pathogenicity) to facilitate the diagnosis in undiagnosed disease network
  - d. Investigated mechanism of model interpretability that can be understood in human terms for better clinical-decision making
- Project: Artificial intelligence/machine learning consortium to advance health equity and researcher diversity (on-going)

- a. Concierge service for consortium members on preparing data (EHR, etc.) for research
- b. Regulatory/Concierge service for consortium members on regulation and compliance
- c. Evaluation of data sharing options, tools, hardware, and cloud usage
- **Graduate student**, Nanyang Technological University, Singapore Jan.2016 - Jun.2020
  - Project: Computational virulence model with functional information for influenza viruses
    - a. Built novel computational methods to identify the potential virulent mutation sites and detect the probability of influenza reassortment
    - b. Proposed a new 2D CNN-based model for influenza antigenicity prediction
    - c. Developed an integrative framework to predict influenza virulence incorporating mutation and reassortment information with extensive experiments
    - d. Provided strategies for the inference of viral lethality through meta-analysis of virulence and antigenicity
- **Undergraduate student**, Shandong University, China Sep.2010 - Dec.2012
  - Project: The design of automatic tracking intelligent car system
    - a. Took charge of all the hardware design and test including system circuit board, main circuit board, driver circuit board, signal detection circuit board and speed measuring circuit board
    - b. Completed the program of signal acquisition and control of steering engine
    - c. simulate and fit curve for gaining optimal running parameters
    - d. Debugged and operated the intelligent car on the real racing track repetitively

## TEACHING AND MENTORING

---

- **Teaching assistant** Jan.2016 - Jan.2020  
 School of Computer Science and Engineering, Nanyang Technological University, Singapore  
**Course:** Data Mining, Computational Biology, Computer Architecture, Introduction to Data Science and Artificial Intelligence
  - Graded assignments and tests for over 300 undergraduate students
  - Supervised course projects for over 80 undergraduate students
  - Led weekly lab sessions for a subset of 40 students
  - Interacted with students and answered questions during weekly office hours
  - Completed a graduate level course in effective teaching strategies with excellence
- **Mentoring undergraduates and graduates:** Sep.2016 - present
  - **1. Xianghe Zhu**, University of Oxford, UK, 2021  
 Project: B-cell epitope prediction of potential novel viruses through machine learning techniques
  - **2. Zihan Luo**, Huazhong University of Science and Technology, China, 2020  
 Project: Virulence prediction of influenza A viruses with prior mutation and reassortment knowledge using all 8 segments
  - **3. Jan Dabrowski**, Swansea University, UK, 2019  
 Project: Interpretable modelling of chronic kidney disease progression with recurrent neural networks

- **4. Nyi Nyi Thwin**, Nanyang Technological University, Singapore, 2019  
Project: Web-based design and implementation of multi-type genomic data mining and retrieval
- **5. Ong Wenqi**, Nanyang Technological University, Singapore, 2018  
Project: Predicting the antigenicity and reassortment probability of influenza viruses
- **6. Chan Wei An Darren**, Nanyang Technological University, Singapore, 2018  
Project: The detection of influenza reassortment based on host tropism of genome segments
- **7. Gayle Natalie Ang**, Nanyang Technological University, Singapore, 2017  
Project: Applying machine learning techniques to predict influenza virulence based on its potential origin sequence
- **8. Viet Hung Tran**, Hanoi University of Science and Technology, Hanoi, Vietnam, 2017  
Project: Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model
- **9. Chew Li Ting Chery**, Nanyang Technological University, Singapore, 2016  
Project: The detection of virulence factors of influenza
- **10. Tan Jun Qiu**, Nanyang Technological University, Singapore, 2016  
Project: Inference of homology by BLAST visualization of influenza genome set
- **11. Ho Wei Shen**, Nanyang Technological University, Singapore, 2016  
Project: Web application development for flu analysis that incorporates co-mutation detection, sequence alignment and lookup

## RESEARCH GRANTS

---

- Project title: The discovery of neutralizing antibodies for potential novel coronavirus through machine learning approaches  
Funding: MOE AcRF Tier 1  
Duration: Sep.2020 Aug.2022  
Role: Collaborator (PI: Kwoh Chee Keong, Co-PI: Hsu Li Yang)
- Project title: Host-pathogen protein-protein interaction approaches for predicting virulence  
Funding: MOE AcRF Tier 2  
Duration: Dec.2019 Nov.2022  
Role: Collaborator (PI: Kwoh Chee Keong, Co-PI: Tan Yee Joo)

## INDUSTRIAL EXPERIENCE

---

- Data Science Intern, Holmusk, Singapore Oct.2019 - Jan.2020
  - Designed database to store heterogeneous patient data
  - Developed approaches to provide the severity scores for patients with different diseases
- Software Engineer Intern, Continental, China Dec.2012 - Mar.2013
  - Implemented the mechanical assembly and electrical assembly
  - Integrated and deployed the information of production lines including data files, programs, etc.
  - Added the Counting System and assembled a laser code testing machine to each production line

## HONORS & AWARDS

---

- **Research Scholarship**, Nanyang Technological University, Singapore 2016 - 2020

- **Best oral presentation**, The 6th International Conference on Bioinformatics and Biomedical Science, Singapore, Jun.2017
- **The Excellence in Teaching Assistantship Award**, Nanyang Technological University Mar.2017
- **Research Scholarship**, Central South University, China 2013 - 2015
- **Second Prize**, National Graduate Students Mathematical Modeling Competition, China Sep.2013
- **Science and Technology Innovation Scholarship**, Shandong University, China Apr.2013
- **Research and Innovation Scholarship**, Shandong University, China Dec.2012
- **Grand Prize (No.2, 0.01%)**, National Undergraduate Intelligent Car Competition, China Aug.2012

## PUBLICATIONS

---

- **Under review or in revision**

- **Yin R**, Wack M, Kohane IS, Avillach P, et al. Identification of genotype-phenotype associations in Phelan-McDermid syndrome using family-sourced data from an international registry. *American Journal of Human Genetics*, 2022. (under review)
- **Yin R**, Gutierre A, Kohane IS, Avillach P, et al. VarPPUD: predicting pathogenicity of gene-variant combinations in the undiagnosed disease patients. *Genetics in Medicine*, 2022. (in submission)
- Gutierre A\*, **Yin R**, Serret-Larmande A, et al. Creating a patient-centric information common on undiagnosed patients at a national scale. *Journal of the American Medical Informatics Association*, 2021. (in submission)
- **Yin, R.**, Luo, Z., Zhuang, P., Kwoh, C. K., Lin, Z. (2022). ViPal: A Framework for Virulence Prediction of Influenza Viruses with Prior Viral Knowledge Using Genomic Sequences. *bioRxiv*.

- **Published or accepted**

- Biao Ye\*, **Yin R.**\*, Changyu Yin, et al. (2023). CLCAP: Contrastive Learning Improves Antigenicity Prediction for Influenza A Virus Using Convolutional Neural Networks. *The Asia Pacific Bioinformatics Conference*
- Yongfan Ming, Wenkang Wang **Yin R.**, et al. (2023). A review of the enzyme design in catalytic stability by artificial intelligence. *Briefings in Bioinformatics*
- Lin Z, Feng L, Guo X, **Yin R**, Kwoh CK, Xu X. (2023)COMET: Convolutional Dimension Interaction for Deep Matrix Factorization[J]. *ACM Transactions on Intelligent Systems and Technology*
- Li M, Zhao B, **Yin R.**, et al. (2022). GraphLncLoc: long non-coding RNA subcellular localization prediction using graph convolutional networks based on sequence to graph transformation. *Briefings in Bioinformatics*
- **Yin R.**, Zhu X., Zeng M., Wu P., Li M., Kwoh, C. K. (2022). A framework for predicting variable-length epitopes of human-adapted viruses using machine learning methods. *Briefings in Bioinformatics*
- **Yin R.**, Luo Z., Kwoh C. K. (2021). Exploring the Lethality of Human-Adapted Coronavirus Through Alignment-Free Machine Learning Approaches Using Genomic Sequences. *Current Genomics*, 22(8), 583-595.
- Zhang W., Wu P., **Yin R.**, Sun M., Zhang, R., et al. Mendelian Randomization Analysis Suggests No Associations of Herpes Simplex Virus Infections with Multiple Sclerosis. *Frontiers in Neuroscience*
- Wu PF, Zhang X, Zhou P, **Yin, R.**, et al. Growth differentiation factor 15 is associated with Alzheimer's disease risk[J]. *Frontiers in Genetics*, 2021.

- **Yin R**, Luo Z, Zhuang P, et al. VirPreNet: a weighted ensemble convolutional neural network for the virulence prediction of influenza A virus using all 8 segments[J]. *Bioinformatics*, 2021.
- **Yin R**, Thwin N N, Zhuang P, et al. "IAV-CNN: a 2D convolutional neural network model to predict antigenic variants of influenza A virus." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2021).
- Wu PF, Du B, Wang B, **Yin R**, et al. Joint Analysis of Genome-Wide Association Data Reveals No Genetic Correlations Between Low Back Pain and Neurodegenerative Diseases[J], *Frontiers in genetics, Statistical Genetics and Methodology*, 2021
- Lin Z, Feng L, **Yin R**, et al. "GLIMG: Global and local item graphs for top-N recommender systems." *Information Sciences* 580 (2021): 1-14.
- **Yin R**, Luusua E, Dabrowski J, et al. Tempel: Time-series Mutation Prediction of Influenza A Viruses via Attention-based Recurrent Neural Networks[J]. *Bioinformatics*, 2020.
- Zhang Y, Long Y, **Yin R**, et al. DL-CRISPR: A Deep Learning Method for Off-Target Activity Prediction in CRISPR/Cas9 With Data Augmentation[J]. *IEEE Access*, 2020, 8: 76610-76617.
- **Yin R**, Zhang Y, Zhou X, et al. Time series computational prediction of vaccines for influenza A H3N2 with recurrent neural networks[J]. *Journal of Bioinformatics and Computational Biology*, 2020.
- **Yin R**, Zhou X, Rashid S, et al. HopPER: an adaptive model for probability estimation of influenza reassortment through host prediction[J]. *BMC Medical Genomics*, 2020, 13(1): 9.
- Zhou X, **Yin R**, Zheng J, et al. An encoding scheme capturing generic priors and properties of amino acids improves protein classification[J]. *IEEE Access*, 2019, 7: 7348-7356.
- **Yin R**, Chee Keong Kwoh, and Jie Zheng. Whole Genome Sequencing Analysis: Computational Pipelines and Workflows in Bioinformatics, *Encyclopedia of Bioinformatics and Computational Biology*, 176-183, 2019.
- **Yin R**, Tran V H, Zhou X, et al. Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model[J]. *PLoS one*, 2018, 13(12): e0207777.
- Zhou X, **Yin R**, Kwoh C K, et al. A context-free encoding scheme of protein sequences for predicting antigenicity of diverse influenza A viruses[J]. *BMC genomics*, 2018, 19(10): 936.
- **Yin R**, Zhou X, Zheng J, et al. Computational identification of physicochemical signatures for host tropism of influenza A virus[J]. *Journal of bioinformatics and computational biology*, 2018: 1840023-1840023.
- Ding P, **Yin R**, Luo J, et al. Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological and Network Knowledge[J]. *IEEE journal of biomedical and health informatics*, 2018.
- **Yin R**, Tan J, Akhila D, et al. Inference of Sequence Homology by BLAST visualization of Influenza Genome set[C]//*Proceedings of the 9th International Conference on Computational Systems-Biology and Bioinformatics*. ACM, 2018: 5.
- Ivan F X, Zhou X, Deshpande A, **Yin R**, et al. Phylogenetic tree based method for uncovering computational site-pairs in influenza viruses[C]//*Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2017: 21-26.
- **Yin R**, Zhou X, Ivan F X, et al. Identification of Potential Critical Virulent Sites Based on Hemagglutinin of Influenza A Virus in Past Pandemic Strains[C]//*Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science*. ACM, 2017: 30-36. **(Best presentation)**

- **Patent**

- The apparatus and method of the online sedimentation of the dense thick complicated feed liquid matrix of high temperature (CN103645085B) Dec.2013
- Model construction method and device for linear sweep polarography (CN103473475A) Jun.2013

## PRESENTATIONS

---

- Joint 30th International Conference on Genome Informatics and Australian Bioinformatics And Computational Biology Society Annual Conference, Sydney, Australia, Dec. 9-11, 2019  
Title: Time series computational prediction of vaccines for inuenza A H3N2 with recurrent neural networks
- The 9th International Conference on Computational Systems-Biology and Bioinformatics Proceedings, Bangkok, Thailand, Dec. 10 - 13, 2018  
Title: Inference of sequence homology by BLAST visualization of Influenza genome set
- The 29th Genome Informatics , KunMing, Yunnan, China, Dec. 3-5, 2018  
Title: Computational identification of physicochemical signatures for host tropism of influenza A virus
- The 6th International Conference on Bioinformatics and Biomedical Science, Singapore, Jun. 22-24, 2017  
Title: Identification of potential critical virulent sites based on hemagglutinin of influenza A virus in past pandemic strains

## ACADEMIC AND COMMUNITY SERVICE

---

- Program Committee member: The 21th Asia Pacific Bioinformatics Conference (APBC 2023)
- Program Committee member: IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM 2022)
- Guest Editor for Research Topic: Computational Methods to Analyze RNA Data for Human Diseases in the journal of Frontiers In Genetics.
- Member of Harvard Postdoc Association
- Member of NHLBI BioData Catalyst
- Reviewer for Briefings in Bioinformatics, Bioinformatics, Scientific report, Computational and Structural Biotechnology Journal, Expert Systems with Applications, Journal of Biomedical and Health Informatics, Artificial intelligence in medicine, PeerJ, IEEE Access, Journal of Theoretical and Computational Chemistry
- Reviewer for ACM Conference on Bioinformatics, Computational Biology, AAAI Workshop, International Conference on Bioinformatics, The Asia Pacific Bioinformatics Conference, The International Conference on Genome informatics