

## Ruifeng Hu

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College

Email: huruifeng.cn@hotmail.com

Website : huruifeng.github.io

### RESEARCH FOCUSES

My interests include but not limited to developing novel bioinformatics algorithms and tools, whole genome analysis, data mining, biostatistics, and web application development, etc.

Nearly five years' computational biology and four years' computer science and technology experiences made me have the ambitions to dig deeper into life science or biology science. I want to apply my skills to the scientific researches where they are needed, to use computational approaches to discover the mechanism of gene regulation and analyze the variation that affect the gene expression, to have a better understand how our genome organized and worked, to cure diseases and make a better world.

### EXPERIENCES

#### ■ Teaching Assistant

Institute of Medicinal Plant Development, PUMC&CAMS

2016-2017

Served as a Teaching Assistant for the course of Python Programming, Applied Bioinformatics

#### ■ PhD Candidate

Institute of Medicinal Plant Development, PUMC&CAMS

2012.9-2017.7

A 5-year PhD training program in Medical Science, Pharmacognosy, with a concentration in Bioinformatics

### EDUCATION

#### ■ Chinese Academy of Medical Sciences & Peking Union Medical College(PUMC&CAMS) *Beijing, China*

Institute of Medicinal Plant Development(IMPLAD)

*Sep, 2012 - Jul, 2017*

PhD in Medical Science, Pharmacognosy, with a concentration in Bioinformatics

**Overall GPA: 3.58/4.50**

**Weighted Average: 87.20/100**

#### ■ Nanjing Forestry University(NJFU)

*Nanjing, China*

College of Computer Science and Technology

*Sep, 2008 - Jul, 2012*

Bachelor of Engineering in Computer Science and Technology

**Overall GPA: 3.87/4.50**

**Weighted Average: 89.92/100**

### RESEARCH PPOJECTS

#### ■ An Evidence-Based platform for Network Pharmacology analysis(TarNet)

*Mar, 2014 – Mar, 2016*

*PhD Research Project, Institute of Medicinal Plant Development (IMPLAD), Beijing*

- The development of new multicomponent drugs for complex diseases is imperative, and the establishment of a suitable solution for drug group-target protein network analysis is a key scientific problem that must be addressed.
- Information on 894 medicinal plants from Chinese, Japanese, European, and American Pharmacopoeia were gathered and enriched.
- Text mining technologies were used to collect information on plant-compound inclusion relationships and compound-protein interactions from SciFinder Scholar and PubMed abstracts.
- 295,154 publications, 12,187 natural products and 10,763 potential bio-targets were identified. All this data

can be used to build the network for visualization. TarNet integrates 216,587 experimentally confirmed PPI data entries on 16,452 human proteins.

- Python was used as our main data processing programming language. We used the NetworkX (<http://networkx.github.io/>) python package for server computations and analyses. Standard HTML4.01/CSS, Javascript and the Cytoscape-Web library were used to browse and interact with networks.

## ■ **Mathematical modeling of estimating the growth rate of *A. flavus*, and predicting the amount of aflatoxin** *Apr, 2012 - Oct, 2012*

*Research Project, Institute of Medicinal Plant Development (IMPLAD), Beijing*

- The growth of *A. flavus* and the production of aflatoxin under different conditions were widely obtained through a comprehensive literature review.
- Three mathematical models were established to predict the *A. flavus* colony growth rate, lag phase duration and aflatoxin content, as functions of temperature and water activity.
- A web platform was utilized to show the results and display warning information using PHP

## ■ **lncRNA Targets Prediction based on nucleic acid thermodynamics** *Feb, 2015 - Apr, 2015*

*Research Project, Institute of Medicinal Plant Development (IMPLAD), Beijing*

- The nearest-neighbor (NN) model was used to calculate binding-free energy.
- The main principle of NN model for nucleic acid assumes that identity and orientation of neighbor base pairs determine stability of a given base pair.
- Allow use not only for human but also for other animals or plants; processing all lncRNAs in high throughput without RNA size limitation.
- Web-based, user-friendly interface, and colored result displays.

## ■ **Algorithms designing for new Chinese herbal medicine formulae design** *May, 2015 - Sep, 2015*

*Research Project, Institute of Medicinal Plant Development (IMPLAD), Beijing*

- 28 Chinese medicinal herbs of Traditional Chinese Medicines(TCM) for the treatment of diabetes were chosen which have the highest appearance rate in literatures.
- Construction of human protein-protein interaction(PPI) network and diabetes disease protein interaction network has been done
- Obtained information on the interaction of compounds–targets in our own database of TarNet
- Algorithms was designing to intervene the PPI-networks, and test the robustness of networks by using different herb combinations.

## ■ **PacBio long reads error correction by short reads alignment** *Oct, 2015 - May, 2016*

*Research Project, Institute of Medicinal Plant Development (IMPLAD), Beijing*

- The raw data from SMRT sequencing are of relatively low quality, with a random error rate of approximately 15 %.
- To eliminate mistakes consisting of an uncertain number of a particular base in polymer segments, an HC transformation strategy was applied to increase the sensitivity of SR-LR alignment. The poor-quality sequences are excluded. HC reads with lengths less than 40 bp and reads with too many 'N's.
- Bowtie2 is embedded as the default alignment program.
- Sample calculations are provided illustrating the precision and efficiency of this method regarding error correction and isoform detection.

## **PROFESSIONAL STRENGTHS AND SKILLS**

- Proficiency in Python, C++, C, JS, PHP, SQL, and Proficiency in Web development
- Statistical analysis and data mining using R and MATLAB
- Rich experience in bioinformatics

- Rich experience in text mining using python, Familiar with machine learning method.
- Good problem-solving/analytical skills
- Familiar with Molecular biology techniques and experiments

### **AWARDS & SCHOLARSHIP**

- National Scholarship for Graduate Students, from Chinese Ministry of Education and Ministry of Finance (CAMS&PUMC) 2016.10
- First Grade Scholarship for Academic Excellence, two times (CAMS&PUMC) 2014 - 2015
- Outstanding Graduate (NJFU) 2012.6
- Pacemaker to Merit Student of Nanjing Forestry University (In three consecutive years) 2009 - 2011
- National Scholarship for undergraduate Students, from Chinese Ministry of Education and Ministry of Finance(NJFU) 2009.10

### **EXTRACURRICULAR ROLES**

- **Department Head** Graduate Union of Peking Union Medical College
- **Class Monitor** Class 0808012, College of Computer Science and Technology, NJFU

### **PUBLICATIONS**

- [6] **Hu R**, Sun X. Design of new traditional chinese medicine herbal formulae for treatment of type 2 diabetes mellitus based on network pharmacology[J]. Chinese Journal of Natural Medicines, 2017,15(6):0436-0441.
  - [5] **Hu R**, Sun G, Sun X. LSCplus: a fast solution for improving long read accuracy by short read alignment[J]. BMC Bioinformatics, 2016, 17(1):451.
  - [4] **Hu R**, Ren G, Sun G, et al. TarNet: An Evidence-Based Database for Natural Medicine Research[J]. PloS one, 2016, 11(6): e0157222
  - [3] **Hu R**, Sun X. lncRNATargets: A platform for lncRNA target prediction based on nucleic acid thermodynamics. [J]. Journal of Bioinformatics & Computational Biology, 2016, 14(04):1650016.
  - [2] **Hu R**, Xing X, Sun G, et al. [The prospect of using bioinformatics technology in the field of biological medicine in the era of big data] [J]. Yao xue xue bao= Acta pharmaceutica Sinica, 2014, 49(11): 1512-1519
  - [1] **Hu R**, Zeng X, Gao W, et al. HRAS: a webserver for early warning of human health risk brought by aflatoxin[J]. Molecular biology reports, 2013, 40(2): 1181-1187
- 
- Yang D, Shao J, Hu R, et al. Angiotensin II promotes the anticoagulant effects of rivaroxaban via angiotensin type 2 receptor signaling in mice[J]. Scientific Reports, 2017,7(1): 369
  - Zhang L, Jing J, Zhang L, **Hu R**, Li G, Huo X, et al. Quantitative analysis of differential protein expression in cervical carcinoma cells after zeolite treatment by stable isotope labeling with amino acids in cell culture[J]. Journal of Proteomics, 2015, 126:279-287.
  - Ai Q, Sun G, Luo Y, Dong X, **Hu R**, Meng X, et al. Ginsenoside Rb1 prevents hypoxia-reoxygenation-induced apoptosis in H9c2 cardiomyocytes via an estrogen receptor-dependent crosstalk among the Akt, JNK, and ERK 1/2 pathways using a label-free quantitative proteomics analysis[J]. Rsc Advances, 2015, 5(33):26346-26363..