Zichen Wang, PhD

333 Ludlow Street
Stamford, CT 06902
(917) 826-8235
wangzc921@gmail.com

ABOUT

Pricipal Scientist in biomedical informatics, with intensive experiences in bioinformatics, systems pharmacology, transcriptomics, web development and Machine Learning (ML). I am passionate about many research areas in ML including deep learning (DL), natural language processing (NLP) and reinforcement learning (RL). I have been actively learning and applying the latest algorithms and techniques from those areas into my research.

EXPERIENCE

Sema4, Stamfort, CT — *Principal Scientist*

SEP 2019 - PRESENT

Conduct research and development using longitudinal EMR data to understand diseases with Deep Learning.

Icahn School of Medicine at Mount Sinai, New York, NY — *Assistant Professor*

JUN 2017 - SEP 2019

Conducted research in Bioinformatics and Computational Systems Biology using high-dimensional molecular and clinical big data.

Icahn School of Medicine at Mount Sinai, New York, NY — *Postdoctoral Fellow*

JAN 2017 - JUN 2017

Developed a DL model leveraging clinical variables from the EMR to study physiological aging.

EDUCATION

Icahn School of Medicine at Mount Sinai, New York, NY — *PhD in Biomedical Science (bioinformatics and genomics)*

AUG 2012 - DEC 2016

Developed computational and mathematical methods to study the complexity of regulatory networks in mammalian systems.

China Agricultural University, Beijing, China — *BS in Biochemistry and Molecular Biology*

SEP 2008 - JUN 2012

SELECTED PUBLICATIONS

SKILLS

ML/Data Science: Univariate and Multivariate Statistics, Ensemble Models (Random forest, Gradient boosting), Network Analysis, Deep Neural Network, NLP, RL

Biomedical Informatics: Electronic Medical Records, Next-gen Sequencing Data, Transcriptomics, Cheminformatics

Software Development:

Web frameworks: Flask, Shiny

Databases: MongoDB, MySQL, PostgreSQL

DevOps: Docker, Apache Mesos

UI: Bootstrap, React.js, Backbone.js, d3.js, THREE.js

LANGUAGES

Programming: Python, R, Javascript, SQL, Bash, php, MATLAB, HTML, CSS

Natural: English, Mandarin

SOCIAL MEDIA

GitHub: wangz10

LinkedIn: zichenwang

Twitter: @ZichenWangPhD

Medium: @wangzc921

Kaggle: zichen

StackOverflow: 2340065/zichen-wang

Google scholar: bwLMCp4AAAAJ

Personal website: wangz10.github.io

<u>Wang Z</u>, Lachmann A, Keenan AB, Ma'ayan A: **L1000FWD: Fireworks visualization of drug-induced transcriptomic signatures.** *Bioinformatics* 2018, 34(12): 2150-2152

Developed a large-scale interactive visualization engine (l1000fwd.net) for over 17,000 drug-induced transcriptomic signatures to facilitate drug discovery.

<u>Wang Z</u>, Li L, Glicksberg BS, Israel A, Dudley JT, Ma'ayan A: **Predicting Age by Mining Electronic Medical Records with Deep Learning Characterizes Differences between Chronological and Physiological Age.** *Journal of Biomedical Informatics* 2017, 76: 59-68

Employed deep neural network to predict physiological age for over 300,000 patients using the Electronic Medical Record (EMR) data and performed large-scale statistical association studies to reveal clinical and genetic factors contribute to disparities in aging.

<u>Wang Z</u>, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW et al: **Extraction and analysis of signatures from the gene expression omnibus by the crowd.** *Nature Communications* 2016 7, 12846.

Built a crowdsourcing microtask platform to enable over 70 citizen scientists to collaboratively work on curation of gene expression signatures from public repository and achieved one of the largest high-quality collection of signatures themed by genetic perturbation, drug treatment and diseases: http://amp.pharm.mssm.edu/CREEDS/.

<u>Wang Z</u>, Ma'ayan A: **An open RNA-Seq data analysis pipeline tutorial with an example of reprocessing data from a recent Zika virus study.** *F1000Research* 2016, 5(1574).

Developed one of the first transparent and reproducible computational pipeline implemented in Dockerized Jupyter notebook with equipped diverse interactive visualization widgets to analyze RNA-seq data from scratch: https://github.com/MaayanLab/Zika-RNAseq-Pipeline.

<u>Wang Z</u>, Clark NR, Ma'ayan A: **Drug-induced adverse events prediction with the LINCS L1000 data.** *Bioinformatics* 2016, 32(15):2338-2345.

Developed a novel multi-label classification framework to predict adverse drug reactions using features from drug-induced gene expression data and chemical structures: http://maayanlab.net/SEP-L1000/.

<u>Wang Z</u>, Clark NR, Ma'ayan A: **Dynamics of the discovery process of protein-protein interactions from low content studies.** *BMC Systems Biology*, 2015, 9(1):26.

Introduced a mathematical network discovery model to investigate the existence and consequences of research focus biases in biological networks from the literature.