

Shoujun Gu, Ph.D

Current Visa Status: Adjust Status Pending

Rockville, Maryland

shoujun.gu@gmail.com | 210-315-2598

<https://shoujungu.github.io/introduction/> | <https://www.linkedin.com/in/shoujungu/>

Technical Skills

Programing Language:

Python: Numpy, Pandas, Matplotlib, SQLAlchemy, Flask, Scikit-learn, Keras
R: DESeq, EdgeR, maSigPro

Interactive Data Visualization & Webpage:

Javascript: D3js, Plotly, Leaflet
HTML: Bootstrap

Database:

MySQL, MongoDB

Bioinformatics & Next Generation Sequencing (NGS):

Data mining from various biology databases: NCBI, Ensembl, TCGA, cBioportal
Hands-on experience in NGS data analysis workflow

Molecular & Cellular Biology:

Extensive experience in most molecular & cellular experiment, as well as animal models

Selected Projects

- **The Cancer Genome Atlas (TCGA) data mining and analysis project**

The manuscript of this project entitled 'A Pan-Cancer Atlas of Genomic, Epigenomic and Transcriptomic Alterations in the TGF- β Pathway' is under reviewing.

- **Built a machine learning model to predict the significance of a research project based on its abstract**

Details available at: https://github.com/shoujungu/Impact_Factor_Pred

Sample app available at: <https://if-pred.herokuapp.com>

- **Interactive visualization of US healthcare providers data**

Details available at: https://github.com/shoujungu/US_Healthcare_Providers

Sample app available at: <http://hcproviders.herokuapp.com>

- **Data mining and analysis on all healthcare publications (Year: 2017) in PubMed database**

Details available at: https://github.com/shoujungu/2017_Pubmed_Analysis

Experience

The George Washington University, Washington, DC

Oct 2015 –

Postdoctoral Researcher

- Studying immune cell profiling in GI cancers by TCGA data mining and using various deconvolution tools;
- Studying cancer genomic, epigenomic and transcriptomic alterations in the TGF- β pathway by TCGA data mining and analysis (manuscript under review);
- Discovered an aberrant regulation pathway between TGF- β signaling and cancer biomarker CEA in colon adenomas by using both whole genomic sequencing and whole transcriptome sequencing methods ¹;
- Discovered that Vitamin D promotes liver tumor progression in TGF- β deficient environment by using both in vivo mouse model and bioinformatics tools ²;
- Published a review article entitled “Alcohol, stem cells and cancer” ³;

National Institutes of Health, Bethesda, MD

Oct 2012 – Sept 2015

Visiting Fellow

- Discovered that Wnt5a-induced Vangl2 phosphorylation is a key step for planar cell polarity (PCP) signaling initiation and establishment by showing phosphorylated Vangl2 exhibits altered interaction with other PCP proteins and forms self-aggregation;
- Generated myc-Prickle1 knock-in mice by using CRISPR/Cas9 genome editing technique;
- Discovered that Casein Kinase 1 ϵ/δ conditional knockout mice exhibit significantly shortened long bones and delayed limb development.

Education

The George Washington University, Washington, DC

Oct 2017 – Apr 2018

Data Analytics Boot Camp

- An intensive 24-week long boot camp dedicated to Data Mining and Analytics Skills on a variety of real-world problems.

UT Health Science Center at San Antonio

Aug 2005 – May 2012

Ph.D. Department of Biochemistry

- Oral presentation in American Society for Bone and Mineral Research 2008 Annual Meeting;
- Discovered that the Twist1 inhibits Sox9 transactivation through protein-protein interaction between Twist1 C-terminal domain and Sox9 DNA binding domain, which results in negative regulation of chondrogenesis ⁴.

Publications

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4839765/>
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4960540/>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5724803/>
4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3375531/>