

Yue LIU

Address: 302 W 38th Street #203, Austin, Tx, USA
Email: yliu_625@163.com Phone: 737-610-2945

SUMMARY

Bioinformatics scientist with 7-year experience specializing in oncology, microbiome and evolution data analysis.

EDUCATION

University of Texas at Austin

Austin, USA

Doctor of Philosophy in Biology

Aug. 2019 - Now

Shanghai Jiao Tong University (SJTU)

Shanghai, China

Master of Science in Biology specializing in Bioinformatics

Sep. 2011 - Mar. 2014

- **Overall GPA:** 2.69/3.3; **Ranking:** TOP 1 (Total 17)
- **Awards:** 2014 Outstanding Graduate (TOP 1), SJTU; 2013-2014 Arawana Scholarship (TOP 1%), Wilmar International;
- 2013 Anthony J Leggett Award (TOP 1), Education Development Foundation, SJTU; 2011-2012 Merit Student (Top 1), SJTU

Shanghai Ocean University (SHOU)

Shanghai, China

Bachelor of Science in Biotechnology

Sep. 2007 - Jul. 2011

- *Minor: Economics* in Fudan University
- **Overall GPA:** 3.56/4.00; **Major GPA:** 3.80/4.00; **Ranking:** TOP 10 (Total 486)
- **Awards:** 2009-2010 People's Scholarship First Term First Prize (TOP 2%), SHOU; 2009-2010 People's Scholarship Second Term Second Prize (TOP 10%), SHOU; 2009-2010 Zhu Yuanding Scholarship (TOP 0.5%), SHOU; 2008-2009 People's Scholarship First Term First Prize (TOP 2%), SHOU; 2008-2009 People's Scholarship Second Term First Prize (TOP 2%), SHOU; 2008 Merit Student (Top 5%); 2007-2008 People's Scholarship First Term Second Prize (TOP 10%), SHOU; 2007-2008 People's Scholarship Second Term First Prize (TOP 2%), SHOU

PUBLICATION

Liu Y, Liu BY, Hao P, Li X, Li YX, Wang JF. π - π stacking mediated drug-drug interactions in human CYP2E1. *Proteins*. 2013 Jun; 81(6): 945-54

WORK EXPERIENCE

Illumina Inc., China

Shanghai, China

Bioinformatics Scientist

Jun. 2017 - present

Project: *China Association for Medical Devices Industry 2018 BRCA Mutations Database* (Jun. 2018 - present)

- Assess the quality of BRCA1/2 gene sequencing data from 25 participating Chinese diagnostic companies
- Construct BRCA1/2 gene germline and somatic mutations' database, which will be used for future clinical guideline for OlympiAD

Project: *Illumina Immune Oncology Collaboration with Zhong Shan Hospital* (Jun. 2017 - present)

- Analyze the differences of immune markers between primary and the corresponding metastatic tumors with the goal of establishing the baseline of liver cancer immune biomarkers in China
- Calculate TMB, MSI, HLA and Neoantigen through Illumina in-house algorithm, analyze the consistency between HLA, TMB, and Neoantigen via tumor - normal WES method, such as germline mutations extraction and other synonymous mutations filtering, and define FPKM cutoff for HLA biomarkers through 900 liver samples
- Improve the accuracy for Neoantigen prediction through FPKM score, mutation rate, and protein similarity selection

Project: *Chinese Traditional Medicine Genetics Database with the Institute of Medicinal Plant Development* (Jun. 2017 - present)

- Assemble the chloroplast genome from shotgun sequencing data (CLC, SOAP, Velvet, SPAdes, etc.) for 1000 commonly used Chinese traditional medicinal plants
- Identify medicinal plants to sub-species through chloroplasts de novo assembling results and contrast the results with higher level similar plants to make phylogenetic trees
- Conclude molecular identification method can improve identification accuracy from traditional morphology method

that always distinguishes a medicine from shape or smell

- Compare the assembling results from NGS data with PacBio's results and use the NGS data to revise the PacBio's results

Project: Somatic Variant Calling Pipeline Development (May 2018 - present)

- Develop somatic variant calling pipeline from BWA alignment, mark duplicates to PISCES variants calling and add background correction step, which significantly decrease the background noise from FFPE damage
- Help with developing UMI classification method to call extra-low frequency variants
- Test the accuracy of the pipeline with various FFPE and cell line datasets
- Solve left-alignment problem for indels and complex variants calling, assist in wrapping up the pipeline through docker and push it into China BaseSpace cloud

Project: Nextera Flex Kit Performance Assessment (Jan. 2018 - Jun. 2018)

- Evaluated the performance of Flex with low input (10ng) FFPE DNA in WGS as a part of the NIH NGS quality control standardization project
- Utilized in-house WGS pipeline (BWA and PISCES) to generate variants' gvcf results
- Assessed the performance of Flex kits in genome sequencing of bacteria (*M.tuberculosis*, *M.Kansasii*, *M.Gordonae* and *M. terrae*) that causes common pulmonary diseases for Centers for Disease Control (CDC) in China

Thermo Fisher (Life Tech), China

Shanghai, China

Field Bioinformatics Scientist

Sep. 2015 - May 2017

Project: Meta16s Bacteria Identification Pipeline Development (Sep. 2015 - Dec. 2015)

- Independently developed the Meta16s Identification Pipeline which can detect bacteria types from samples
- Improved meta16s identification accuracy from 97% to 98% in species level by correcting the Greengenes database with NCBI data
- Integrated the python pipeline into Ion Torrent server with UI
- Supported more than 20 Chinese academic customers with this pipeline

Project: FluA Identification Pipeline Development (Jan. 2017 - May 2017)

- Independently built Python based metagenomics FluA identification analysis pipeline which can rapidly identify avian influenza types, including removing human genome contaminations, constructing phylogenetic trees and annotating pathogenic loci with CDC database
- Obtained FluA whole sequences through two methods, one is de novo assembling and the other is replacing the reference genome with SNPs and indels detected from the variant calling step
- More than half of the CDCs across China adopt this pipeline

Project: PGMDX CFDA Registration (2017 Best Project in Thermo Fisher) (Jan. 2016 - Dec. 2016)

- Supported Ion Torrent PGM IVD CFDA registration, in charge of testing the performance of PGMDX through 400 real world clinical samples with TMAP alignment and variant calling by TVC and summarizing the performance of the results
- Provided analysis guidelines for Pathology Quality Control Center of China that focuses on somatic variants detection and exon deletion
- Collaborated with AstraZeneca on how to interpret deleterious mutations with HGVS

National Engineering Center for Biochip, China

Shanghai, China

Bioinformatics Engineer

Apr. 2014 - Sep. 2015

Project: Chicken GWAS Analysis for South China Agricultural University (Sep. 2014 - Jan. 2015)

- Analyzed GWAS and QTL of 400 chickens, used genABEL in R to calculate the P-value of the correlation between quantitative and quality characters of chicken in microarray data, removed the outlier samples with PCA and to mine key factors of high-quality broiler chicken, and drew Manhattan plots with R to show the results

Project: Health Scanning Array Development (Jan. 2015 - Sep. 2015)

- Collected chronic disease related biomarkers from PubMed, design Direct to Customer (DTC) microarray for disease risk scanning and Y-STR analysis in Illumina and Affymetrix websites, and communicated with these two companies to improve the probes coverage of the personal design; Developed disease risk score algorithm based on the product

Project: Genetic Disease Analysis (Sep. 2014 - Sep. 2015)

- Analyzed genetic disease array data and mined useful information from linkage analysis and haplotype analysis with PHASE calculated D value, R square value and haplotype frequency
- Designed specific NGS Ampliseq panels for genetic disease identification after getting the potential pathogenic

regions

Project: *WES Research on Adolescent Depression in East China* (Sep. 2014 - Sep. 2015)

- Conducted WES data analysis from BWA alignment to GATK SNVs and indels identification, CNV/SV detection, etc.

LABORATORY EXPERIENCE

Key Laboratory of Systems Biomedicine, SJTU

Shanghai, China

Group Project

Sep. 2011 - Mar. 2014

Project: *The Interaction Mechanism between Cytochrome P450 Enzymes and Drugs* (Sep. 2011 - May 2013)

- Calculated CYP2E1 protein interaction with drugs, employed GPU-accelerated molecular dynamics simulations to study the multiple-binding mode of human CYP2E1 and published a paper in *Proteins*

Project: *Molecular Markers of Gastric Cancer* (Sep. 2011 - Jan. 2014)

- Analyzed RNA-seq and microarray data as well as identified gastric cancer related RNA expression biomarkers and new transcripts with cufflinks at different stages
- Analyzed the relationship between biomarkers with clinical information (age, clinical stage, survival time, etc.), researched gastric cancer related pathway and used SPSS to draw survival curves
- Constructed Weighted Correlation Network Analysis (WGCNA) in R to identify mutual relationship between genes and selected key markers in the network

Project: *Metabolomics Data Analysis* (Sep. 2011 - Jan. 2012)

- Analyzed urine metabolomics data from UPLC and GS-MS with PCA, PLS-DA and cluster analysis in SIMCA and identified potential biomarkers with diabetes

Shanghai Ocean University

Shanghai, China

Undergraduate Graduation Project

Aug. 2010 - Jun. 2011

Project: *Cloning and Sequence Analysis of Grass Carp Bf/C2A Clones* (Sep. 2011 - Jan. 2012)

- Performed sequence alignment of immune related gene clones of Bf/C2A in Grass Carp, which is easy to infect with *A.hydrophila* in the early development stage and analyzed the relationship between different kinds of clones of Bf/C2A and *A.hydrophila*
- constructed phylogenetic trees of Bf/C2A clones
- Identified the RNA expression level in different tissues and different development stages

SOFTWARE PATENTS

- Blood Disease Variants Identification Software
- FluA Identification and Annotation Software
- HBV classification and anti-drug annotation software

SKILLS AND INTERESTS

- **Languages:** Native Chinese; Fluent English; Japanese Beginner
- **Computer Skills:** Python; R; HTML; CSS; JavaScript; Linux
- **Bioinformatics Skills:** Statistics analysis in SPSS and SIMCA; Proficient in bioinformatics tools (BWA, GATK, Pieces, Plink, STAR, bowtie, Cufflinks etc.) and bioinformatics databases (NCBI, Ensembl, TCGA, 1000K, HGMD, Oncomine etc.)
- **Interests:** Piano (Grade 8); Outdoor Camping

EXTRACURRICULAR ACTIVITIES

Formula Development Intern, Dumex, China

Nov. 2012 - Jan. 2013

Commissary in Charge of General Affairs for SBSC, SJTU

Sep. 2011 - Mar. 2014

Volunteer Teacher for Primary School, Teach for China

Aug. 2010 - Aug. 2010

Assistant Minister in Learning Department for the college of Fisheries and Life, SHOU

Aug. 2008 - Aug. 2009