Yue Lyu

Email: yuelyu0521@gmail.com   Tel.: (530)761-3312   Add.: 518 Andes Prairie Lane, Rosharon, 77583

TECHNICIAL SKILLS AND QULIFICATION

- R(tidyverse, ggplot2, shiny); SAS(Base, Macro, Statistics, SQL); MySQL/SQL; Python(numpy, matplotlib, sklearn, pytorch); HTML/CSS; SPSS
- Self-learner; Good team-player; Statistical knowledge and analytic skills; Strong communication skills

EDUCATION

University of California Davis, Graduate Group in Biostatistics   *Biostatistics, MS*   09/2018 - 06/2020
Hunan University, Department of Finance and Statistics   *Statistics, BS*   09/2014 - 07/2018

- Survival Analysis, Longitudinal Analysis, Clinical Trial, Statistical Methods (GLM, ANOVA), Time Series, Epidemiology
- Machine Learning in Genomics, Statistical Programming, Computational Statistics, Mathematical Statistics, Data & Web Technologies, Multivariate Data Analysis, Nonparametric Statistic

WORK EXPERIENCES

Biostatistics Specialist, **Bayer**   02/2021 - 07/2021

- Applied statistical tools to support research and development of new products for crop protection and efficiency in **R**
- Identified incorporate algorithms to automate the analysis of bioassay data and field trial results

Research Assistant, **UC Davis Health**   08/2020 - 03/2021

- Conducted analysis of time-to-event, longitudinal, clinical data from UC Davis Health Center by **R&SAS**
- Participated in a longitudinal study for evaluating Effect of Probiotics on Nausea and Vomiting Among Pregnant Women

Bioinformatic Engineer, **Metanotitia Inc.**   09/2020 - 12/2020

- Analyzed mass spectrometry data integrated with clinical data sets to identify significant targeted metabolites, and predict pre-cancer stage using single-sample data
- Generated health examination reports automatically by using **R shiny** to maintain a health management APP based on more than 3000 electronic health records **(EHR)**
- Enhanced, troubleshooted and maintained R code base for metabolomics data processing pipeline

PROFESSIONAL PROJECTS

Prediction of Disease-MiRNA Association with Improved Random Walk

- Implemented random walk with restart to predict potential MiRNA-Disease association of 495 miRNA and 364 diseases based on known MiRNA-Disease association
- Constructed Integrated similarity for miRNAs and diseases matrix given miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity
- Selected high scored miRNA for experimental validation by comparing the results of different assigned initial probability

Coronavirus-19 Information Real-time Tracker Web Dashboard

- Designed an interface dashboard to track COVID-19 data based on **R Shiny** allow users find the real-time information, detailed statistic, daily reported data, breaking news etc.
- Dynamic map by **Leaflet**; Latest News by calling **APIs**; Updated data by **Web Scraping**; Country comparison by **plotly**
- Code and data used to generate this Shiny application are available on **Github**

Identification of T Cell Cycle Stage (naive T cells into T helper 2 cells) based on Single-cell RNA-seq

- Implemented the identification of cell stage (G1, S, G2&M) of 288 T cells with 38693 genes detected
- Pro-processing workflow based on common QC metrics (low-quality detection), detection of highly variable features and data normalization and scaling followed by Seurat
- Classified cell stage based on compressed information representation of the original count input given by hidden layer of an Autoencoder model

High TERT (telomerase reverse transcriptase) expression is associated with poor prognosis of non-small cell lung cancer

- Missing values handled by multiple imputation based on multinomial regression under missing at random assumption
- Kaplan-Meier curves and log-rank test were used to compare overall survival between TERT IHC score groups
- Cox proportional hazards implemented to model the levels of TERT expression on survival time, with adjustment for clinical characteristics (gender, age, race, smoking, histology, differentiation etc.)