

# Yunqiu Yao

100 Haven Avenue, New York, NY 10032  
(646)-683-6625 | yy2827@cumc.columbia.edu

## EDUCATION

---

### Columbia University

New York, NY

*Master of Science, Biostatistics/Theory and Method Track*

May 2019

- Relevant Coursework: Data Science (R Programming), Probability, Inference, Linear Regression, Generalized Linear Regression, Survival Analysis, Statistical Learning, etc.
- Teaching Assistant for *Applied Regression Models*: tutoring students on SAS

### Shanghai Jiao Tong University

Shanghai, China

*Bachelor of Science, Food Science and Engineering*

June 2017

- Georgia Institute of Technology (Summer Program)
- Academic Progress Scholarship, Academic Excellence Scholarship

## RELEVANT EXPERIENCE

---

### Data Analyst Intern

Shanghai, China

*Edenred - Accentv' China*

Dec. 2016 – Feb. 2017

- Wrote SAS and SQL scripts to produce weekly sales report for THE NORTH FACE and SEPHORA
- Implemented cluster analysis in SAS to analyze 1.86 million purchase records from Haagen-Daz
- Presented key findings to supervisor and contributed to recommendations for future marketing strategies for different groups with different purchase preference and habits
- Collaborated with 3 teammates on the construction of a recommendation engine for L'OCCITANE, including the data cleaning, cluster analysis and association analysis in SAS Enterprise Miner

### Empirical Bayes (EB) Method for Haplotype-based GWAS

Shanghai, China

*Research Assistant, Shanghai Jiao Tong University*

Jan. 2016 – Oct. 2016

- Utilized R to collaborate on the construction of a linear mixed model (LMM) for GWAS with EB
- Spearheaded the initiative to apply the constructed EB-LMM model to haplotype-based GWAS, and applied to the genome of 1092 subjects
- Identified 17 significant haplotype blocks and inferred 3 genes associated with the trait of interest
- Drafted report based on the study findings for presentation to PI overseeing the research

### Proteome-wide Association Study for Colorectal Cancer

Shanghai, China

*Research Assistant, Shanghai Jiao Tong University*

Jan. 2016 – Oct. 2016

- Standardized and converted proteomic data from 95 cases and 32 controls into two integrated matrices
- Introduced the Protein Interaction Network (PPI) paired the mutant proteins with the neighbors
- Applied GLM in R to identify the key mutant proteins significantly affecting neighbors' expression values
- Performed BH and found 58 proteins out of 1189 candidates statistically significant for colorectal cancer

### Logistic Regression for Metastases in Renal Cell Carcinoma

Shanghai, China

*Project team member, Shanghai Jiao Tong University*

Nov. 2015 – Dec. 2015

- Designed a logistic analysis of case-control study on metastasis status of renal cell carcinoma
- Conducted the study for 58 patients with binary metastasis value and other 5 clinical features
- Performed logistic regression and identified 3 key determinants of metastases, and applied the model to other samples for the prediction of metastasis based on clinical features

## SKILLS

---

- R, SQL, SAS, Python, Perl, MS Office (Word, Excel, PowerPoint)