**Next Generation Sequencing (NGS): RNA-Seq** Nov 2018 - Jan 2019

*Differential gene expression (DGE) analysis & Gene set enrichment analysis (GSEA) of RNA-Seq data*

*Demo:* <https://github.com/LuchaoQi/NGS>

* Created tools (**Shell script, R, Python**) that can be used to perform one-stop analysis from downloading the raw Sequence Read Archive (**SRA**) gene data to investigating the differentially expressed gene matrix
* Performed gene set enrichment analysis (**GSEA**) of profiles obtained from Gene Expression Omnibus (**GEO**)
* Identified significant (p-value < 0.05) co-occurring or mutually exclusive mutated driver genes across different cancer types using **Fisher’s exact test, Chi-Square test and Permutation test**
* Identified 50 over-represented genes that may have associations with disease phenotypes

**Investigation of Yelp user funnels, Key Performance Indicators (KPIs)** Nov 2018 - Jan 2019

*Performance analysis of Yelp users & restaurant using SQL*

*Demo:* <https://github.com/LuchaoQi/Yelp_Data_Set_SQL>

* Wrote **web crawler** to scrape and parse unstructured data from Yelp using **Xpaths**, **BeautifulSoup** in Python
* Created a database using **MySQL workbench** and imported ~10 GB data file into the database
* Visualized geographic distribution of restaurants with average ratings using **Tableau**
* Performed metrics analysis (**bracket retention, DAU/MAU**) using SQL to measure customer engagement and making suggestions for ways to improve upon KPIs via **A/B testing**

## **PUBLICATIONS**

1. **Qi L**, Zhang Q, Tan Y, et al. Non-contact High-frequency Ultrasound Microbeam Stimulation: A Novel Finding and Potential Causes of Cell Responses. *IEEE Trans Biomed Eng* 2019.
2. **Qi L**, Zhang Q, Lam KH, et al. Calcium fluorescence response of human breast cancer cells by 50-MHz ultrasound microbeam stimulation. Presented at 2017 IEEE International Ultrasonics Symposium (IUS), 6-9 Sept. 2017 2017.

你如果听了我之前关于data简历上的项目讲解的话，可以发现我一直在强调pipeline的问题，那么在做一个data相关的项目的时候，1. 先要告诉你的读者(可能是hr/recruiter/interviewer)，你为什么要做这个项目，也就是motivation，use case或者要达到的目的和项目功能。2. 说明data的获取来源，比如爬虫获得还是怎样，当然了，从raw data开始是工业界的人最希望看到的。如果实在data相对不是很complex，那么也可以先写一下如何进行的预处理(pre-processing)，包括怎么做的extraction，cleaning等等。3. 预处理之后可以想想怎样存取数据或者进行realtime的data互动，比如存到数据库中？那么这个过程包含数据库的设计。4. 开始使用数学模型来完成你的目的，比如用一些ml的常用算法，然后基于常用模型的优化，调参，比对各个模型的优势等等；5. 测试，是如何测试的最后结果？测试的结果怎么样？为什么你做的东西有价值？这个环节能量化就量化。能量化的结果最有说服力。6. 迭代整个pipeline，再不停优化。要说明优化办法。

数据科学简历修改指导：https://youtu.be/XuYUVlvzZPo