

# Jiawei Gu

LinkedIn: <https://www.linkedin.com/in/jiawei-gu/> | GitHub: <https://github.com/jiawku>

Cell: (801)-637-4021 | Email: [jiawku17@gmail.com](mailto:jiawku17@gmail.com) | Salt Lake City, UT

---

## Summary

---

I had **four** years experience of **data analysis and engineering** in the realm of biocomputing in academic and industry. I obtained a **MS in computer science** and a **MS in biology** from University of Texas at Dallas. With a Certificate from **Amazon Web Service**, I can boost your data analysis and engineering on the cloud.

## Education

---

**01/2017-05/2018**

**MS, Computer Science**; University of Texas at Dallas(UTD); **3.91/4**

**08/2013-12/2016**

**MS, Bioinformatic**; University of Texas at Dallas(UTD); **3.73/4**

## Proficient Skills

---

### *Programming languages*

Python, Java, JavaScript, C++, R, SAS, SQL, PHP, Bash

### *Frameworks & Tools*

- BigData: Apache Spark, Apache Hadoop
- Web: MEAN stack(MongoDB, Express JS, Angular JS, Node.js), Java Spring
- Deep Learning: TensorFlow, Keras
- Containerization: Docker, Kubernetes
- Cloud Service: **Amazon Web Service**(*certified*)

## Certification

- ***AWS Certified Solutions Architect - Associate***  
Credential ID: 6WQVVZSKE24EQLK9
- *Coursera Deep Learning Specialization*
- *SAS Certified Base & Advanced Programmer for SAS 9*
- *edX Verified Certificate for Big Data Analysis with Apache Spark*

## Experience

---

## Bioinformatics Analyst, Navican Genomics, Inc. Salt Lake City, 08/2018-Present

- **Production & Automation**

- Daily clinical control sequencing quality assurance.
- Implemented and improved the **Tumor Mutation Burden(TMB)** calculation algorithm within the bioinformatics team inspired by
- Built an **automated pipeline** for clinical control assurance in **python**. It connected multiple **cloud service platforms**(Philips IntelliSpace Genomics Medicine, SalesForce, SeraCare IQ, Slack, and Amazon AWS) via **REST API**.
- Created an **Illumina BaseSpace Native APP** for sequencing quality control using **Docker**.

- **Data Engineering**

- Created tools to query and clean data from **cloud service platforms** (Philips IntelliSpace Genomics Medicine, SalesForce, Clarity LIMS) via **REST API** for ETL and R&D.
- Designed a **Postgresql Database** to store bioinformatics data.

- **Data Analysis**

- Implemented and run validation workflows for Illumina TruSight™ Tumor 170 Sequencing Assay.
- Designed **statistical tests** for lab operation quality metrics.
- Help with other ad-hoc data analysis requests.

## Research Assistant, Biology Department, UTD 01/2014-12/2016

I focused on the research of Next-Generation Sequencing analysis, especially on ChIP-seq, RNA-seq, and ChIA-PET Seq result analysis, and using machine learning tools to analyze and predict from data.

- Machine Learning Prediction for RNA-chromosome interaction
  - Collect **genomic and epigenomic data** from online resources by a self coded **web crawler**.
  - Cleaned data and used Random Shuffle to generate more negative data points.
  - Applied **Support Vector Machine** to generate a model
- DNA Sequencing Analysis of Brg1 in Cancer Cell, UTD
  - Built pipelines for data process and ChIP-seq analysis using Linux bash shell
  - Implemented a protein binding motif scan and enrichment analysis program in R
  - Analyzed and Visualized data in **R** and **Python**
  - **Article publication:** [Shi, X., et al. "SMARCA4/Brg1 coordinates genetic and epigenetic networks underlying Shh-type medulloblastoma development." Oncogene \(2016\).](#)

## Side Projects

---

### *Dog Breed Identification, Kaggle.com 05/2018*

- A Kaggle machine learning competition Project
  - Design and built a **convolutional neural network(CNN)** in **keras** to determine the breed of a dog in an image (total 120 breeds)
  - Applied a **integrated model** of Xception and InceptionV3 to extract bottleneck features from image
  - Used three fully connected layers with **drop out** and **batch normalization** to get predicted probabilities
  - Used **Adam** optimization algorithm to train the final model with **cross entropy** as loss function
  - Get the final validation accuracy of 99.76%, can rank 102 out of 1286 teams
-