

# Yukun Huang

514 West, 110th St. #5B

New York, NY, 10025

(646) 617-8105, [yukun.huang@columbia.edu](mailto:yukun.huang@columbia.edu)

---

## EDUCATION

**Columbia University**  
MS, Computer Science

New York, NY  
Sep 2017- Dec 2018

**The University of Hong Kong**

BE, Computer Science, June 2016

GPA: 3.60/4.00

Hong Kong, CN

2012 – 2016

*Honors:* Final Year Project Competition 2015-16 Champion and Best presentation – “Metagenomics gene assembly”

*Publication:* MegaGTA: A Sensitive and Accurate Metagenomic Gene-Targeted Assembler Using Iterative de Bruijn Graphs [\[pdf\]](#)

*Leadership:* President – Computer Society

## TECHNICAL SKILLS

*Programming/Scripting Languages:* Proficient - C++; Familiar - Ruby, Java, JavaScript, Bash, SQL, Haskell

*Frameworks and tool:* Ruby on Rails, Ember.js, Hadoop, TensorFlow, OpenMP, Git - <https://github.com/HuangYukun>

## PROFESSIONAL EXPERIENCE

**L3 Bioinformatics Limited**

Software Engineer

Hong Kong, CN

Nov 2016 - Jul 2017

- Worked on the Hong Kong Government ITF funded project: Genomic Database for the Chinese Population with Applications to Precision Medicine.
- Managed the sequencing data, performed case-control studies and calculated correlations for clinical usage.
- Designed and implemented front-end GUIs with Ember.js to show a better visualization for doctors, biologists.

**DemystData**

Software Engineer Intern

Hong Kong, CN

Jul 2014 - Jun 2015

- Collected and cleaned data, built predictive models in R to perform credit ratings for approving loans from SMEs and individuals in UK and Indonesia.
- Integrated data sources and trained models into back-end framework Ruby on Rails, managed relational and NoSQL databases such as MySQL, Cassandra, and MongoDB.

## RESEARCH EXPERIENCE

**The University of Hong Kong Bioinformatics Algorithms Research Laboratory**

Hong Kong, CN

**MegaGTA: a sensitive and accurate metagenomics Gene-Targeted Assembler using iterative de Bruijn graphs**

Jun 2015 - Feb 2016

- Co-developed MegaGTA, which is scalable to handle Next-Generation Sequencing DNA datasets that take up over ~500GB of memory footprint. It's at least 7-fold faster than the existing algorithms and being equally space-efficient.
- Distributed the computationally intensive components to run on multiple cores in a CPU using OpenMP.
- Improved the de Bruijn graph data structure to eliminate false positive assembly.

**The University of Hong Kong Bioinformatics Algorithms Research Laboratory**

Hong Kong, CN

**Long Noisy Single Molecule Real Time (SMRT) sequences de novo assembly**

Jul 2016 - Nov 2016

- Utilized the coverage of read-to-read local alignments to differentiate resolvable repeats against unresolvable ones.
- Resolved some repeat regions existing software failed to and thus increased the NG50.

**The University of Hong Kong Bioinformatics Algorithms Research Laboratory**

Hong Kong, CN

**Identify Antibiotic Resistance Genes of Pathogens from Patient's extracted DNA data**

Feb 2016 - Jul 2016

- Collaborated with Hong Kong Sanatorium & Hospital (HKSH) to help doctors detect pathogens precisely.
- Built Hidden Markov Models for antibiotics resistance genes in existing gene databases to guide gene assemblers such as MegaGTA.