

Department of Computer Science  
Faculty of Engineering, The University of Hong Kong  
Rm307, Haking Wong Bldg  
HKU, Pokfulam Rd., Hong Kong

+852-564-33-527  
[jhsu@connect.hku.hk](mailto:jhsu@connect.hku.hk)  
<https://sujunhao.github.io>

## Education

- |  |                      |
|--|----------------------|
| <b>Ph.D. in Computer Science</b><br>The University of Hong Kong, Hong Kong<br>Supervisor: Ruibang Luo                  | Sept 2019 – Present  |
| <b>M. Eng. in Computer Science and Technology</b><br>Harbin Institute of Technology, Harbin<br>Supervisor: Yadong Wang | Sept 2017 – Jun 2019 |
| <b>Exchange Student in Computer Science and Engineering</b><br>National Chiao Tung University, Taiwan                  | Sept 2015 – Jan 2016 |
| <b>B. Eng. in Bioinformatics</b><br>Harbin Institute of Technology, Harbin   | Sept 2013 – Jun 2017 |

## Work Experience

- |   |                         |
|---|-------------------------|
| <b>Teaching Assistant</b> , The University of Hong Kong<br>▪ COMP2501, Introduction to Data Science and Engineering, Undergraduate Course | Spring 2023             |
| <b>Teaching Assistant</b> , The University of Hong Kong<br>▪ CCST9047, The Age of Big Data, Undergraduate Course, Common Core             | Spring 2020, 2021, 2022 |
| <b>Teaching Assistant</b> , Harbin Institute of Technology<br>▪ Bioinformatics, Undergraduate Course                                      | Fall 2018               |
| <b>Bioinformatics Intern</b> , WeGene (ShenZhen)<br>▪ Topic: SNP applications development and data mining                                 | Summer 2016, 2017       |

## Peer-reviewed Publications

### Corresponding (\*), first or co-first author (†):

1. Z. Zheng<sup>†</sup>, S. Li<sup>†</sup>, J. Su<sup>†</sup>, A.W. Leung, T. Lam, R. Luo\*. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science*, 2022.
  - Built the state-of-the-art germline small variants caller for the Nanopore long-read data.
  - Over 110k downloads. Open source at: <https://github.com/HKU-BAL/Clair3>.
2. J. Su, Z. Zheng, S. Ahmed, T. Lam, R. Luo\*. Clair3-Trio: high-performance Nanopore long-read variant calling in family trios with Trio-to-Trio deep neural networks. *Briefings in Bioinformatics*, 2022.
  - Improved family trio variant calling accuracy of over +10% F1-Score at the low coverage Nanopore long-read data.
  - Over 2.2k downloads. Open source at: <https://github.com/HKU-BAL/Clair3-Trio>.

3. **J. Su**, Y. Wu, H. Ting, T. Lam, R. Luo\*. RENET2: high-performance full-text gene–disease relation extraction with iterative training data expansion. *NAR Genomics and Bioinformatics*, 2021.
  - Extended relation extraction to full-text articles scope and mined all gene-disease relations from all open-access full-text articles (1.8M) with a large GPUs cluster.
  - Over 1.2k downloads. Open source at: <https://github.com/sujunhao/RENET2>.
4. **J. Su**, W. Lui, Y. Lee, Z. Zheng, ..., R. Luo\*. Evaluation of *Mycobacterium tuberculosis* enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Scientific Reports*, 2023.
  - Benchmarked adaptive sequencing performance at metagenomic samples and built a Nextflow pipeline for *Mycobacterium tuberculosis* data analysis.
  - Open source at: <https://github.com/HKU-BAL/ONT-TB-NF>.

**Co-author:**

1. Y. Liu, T. Jiang, **J. Su**, B. Liu\*, T. Zang\*, Y. Wang\*. SKSV: ultrafast structural variation detection from circular consensus sequencing reads. *Bioinformatics*, 2021.

**Presentations**

---

**Talk:**

1. December 2022 – Presenter at HKU Bioinfo-seminars, The University of Hong Kong, Hong Kong
2. June 2023 – Presenter at TRS Symposium, The University of Hong Kong, Hong Kong

**Poster:**

1. April 2023 - ClusterV: Accurate Detection of HIV Quasispecies and Drug Resistance Mutations using ONT Sequencing Data. RECOMB-SEQ, Istanbul, Turkey

**Professional Activities**

---

- Reviewer for ISMB/ECCB, BMC Bioinformatics, Patterns, NeurIPS, ACM BCB.

**References**

---

Ruibang Luo

*PhD Supervisor*

Associate Professor of Computer Science, The University of Hong Kong

[rbluo@cs.hku.hk](mailto:rbluo@cs.hku.hk)

Tak-Wah Lam

*PhD Co-supervisor*

Professor of Computer Science, The University of Hong Kong

[twlam@cs.hku.hk](mailto:twlam@cs.hku.hk)

Yadong Wang

*Master Supervisor*

Professor of Computer Science and Technology, Harbin Institute of Technology

[ydwang@hit.edu.cn](mailto:ydwang@hit.edu.cn)