

Qiwen Zheng

+8618823412178 | roy.qiwen@foxmail.com



EDUCATION

| | |
|--|---------------------|
| Nanyang Technological University | Jan 2021 - Jan 2029 |
| PhD in Bioinformatics + AI | |
| University of Cambridge | Sep 2020 - Jul 2021 |
| Master of Engineering in Electrical and Information Sciences | |
| University of Cambridge | Sep 2017 - Jul 2020 |
| Bachelor of Arts in Electrical and Information Sciences | |

WORK EXPERIENCE

| | |
|--|---------------------|
| BGI Genomics | Aug 2021 - Dec 2024 |
| Senior Software Development Engineer | Shenzhen, China |
| <ul style="list-style-type: none">High-performance computing (HPC) software development based on gene sequencing dataProduct development of comprehensive analysis solutions for gene sequencing data | |

PROJECT EXPERIENCE

Bayesian-based genetic variant calling software development

- Due to the continuous increase in the amount of sequencing data, the analysis speed of the open-source components is brutal to meet the production and delivery requirements. Therefore, it is decided to reconstruct the most time-consuming and widely used variant calling part in the process.
- As a core development member, I had a deep insight into the core algorithm principles of variant detection, re-designed the whole software's task management and memory management, reconstructed the software in C/C++, and applied various algorithm optimization methods to complete the component reconstruction.
- The analysis time has been reduced from 19 hours to 0.5 hours, the accuracy of the results is almost consistent with the original version, and more than 100,000 samples have been analyzed since the launch.

Variant calling using machine learning

- Despite the domination of Bayesian methods, accurately calling genetic variants present in an individual genome from billions of short, errorful sequence reads remains challenging. To further improve the accuracy, machine learning techniques are involved in the variant calling process.
- We first developed an interpretable model based on XGBoost, since interpretability is crucial for clinical applications. Compared to traditional Bayesian-based methods, both the recall and precision are improved by 0.1%.
- Currently, we are fine-tuning an open-source CNN-based model to fit long-read data produced by our machines.

INTERSHIP EXPERIENCE

| | |
|--------------------------|---------------------|
| Hikvision | Jun 2020 - Aug 2020 |
| Audio Algorithm Engineer | |
| TWI | Jul 2019 - Aug 2019 |
| Electronic Engineer | |

ADDITIONAL INFORMATION

- Programming:
 - Languages: Proficient in C/C++ (familiar with STL and Boost libraries), Python and Shell
 - Development tools: Proficient in using integrated development environments (e.g., Visual Studio, CLion), version control tools (Git), and debugging tools (GDB)
 - Performance optimization: Knowledge of C++ program performance optimization techniques, including memory optimization, CPU optimization, and use of tools (e.g., Intel VTune).
- Academic papers :
 - Wang X, Wang T, Xie Z, et al. GPMeta: a GPU-accelerated method for ultrarapid pathogen identification from metagenomic sequences. Brief Bioinform. 2023;24(2):bbad092.
 - Wang T, Zhang Y, Wang H, et al. Fast and accurate DNaseq Variant Calling workflow composed of LUSH toolkit. bioRxiv. 2023.03.01.530618
- Two invention patents; one is authorized, and the other one is accepted for potential authorization.
- Languages: English (fluent, IELTS 8.0)
- Soft skills: teamwork, effective communication, strong work ethic, problem-solving, and emotional intelligence