Bowen Tan

+852 9517 9352 | Shenzhen/Hong Kong, China | bowentan78@gmail.com

EDUCATION

City University of Hong Kong Sep. 2018 — Present
Doctor of Philosophy, Computer Science *Hong Kong, China*

- Cumulative GPA: 4.0/4.3

City University of Hong Kong Sep. 2012 — Jul. 2016
Bachelor of Science, Computing Mathematics *Hong Kong, China*

- Cumulative GPA: 4.02/4.3 | Dean's List, CityU Mainland Student Scholarship Scheme, The Department of Mathematics Outstanding Performance Scholarships

SKILLS

- **Web development**: TypeScript/JavaScript (NestJS), Python (Django), GraphQL
- **Database management**: MySQL, PostgreSQL, MongoDB and Redis
- **CI/CD & DevOps**: GitHub Actions, GitLab CI/CD, containerization with Docker, and cloud resources management with Openshift.
- **Software development/Data analysis**: C/C++, Python(NumPy, Pandas, Scikit-learn, Matplotlib), R(ggplot2, dplyr, tidyverse), Julia, Java.
- **Automation scripting**: Shell/Bash, Python, Lua.
- **Language**: Chinese, English, Japanese

WORK EXPERIENCES

Project Associate May 2024 — Present
The Hong Kong Polytechnic University *Hong Kong, China*

- **Backend Development**: Engineered the core architecture of a new admissions system using NestJS and MongoDB. Developed comprehensive RESTful API development for seamless application processing, robust authentication and authorization, as well as dynamic PDF generation and merging from templates, ensuring prompt responses and enhanced user satisfaction.
- The innovative system consolidates the entire application process, including student applications, supervisor nominations, and admission works for clerical staffs, into a single online platform. It drastically reduces most of the time and effort spent on communications among students, various departments, colleges, and supporting units.
- **CI/CD & Deployment Automation**: Streamlined deployment by automating CI/CD pipelines via GitLab CI/CD and deploying containerized applications on Openshift, significantly reducing manual effort.
- **Version Release Management**: Designed and implemented a robust version release workflow integrating Git and GitLab CI/CD. Automated changelog generation and managed release processes across multiple branches and repositories.
- **Data Migration**: Developed resilient data migration scripts to ensure a seamless transition from legacy systems to the new platform.

Research Assistant Nov. 2022 — Apr. 2024
City University of Hong Kong *Hong Kong, China*

- **Algorithm Development**: Designed and implemented innovative clustering algorithms for topologically associated domains (TADs) using dynamic programming and graph theory, enhancing the efficiency of existing TAD-calling tools to guarantee that the optimal detections of TADs were obtained so that the found TADs were meaningful in genomics.
- **Computing Resource Management**: Configured and maintained a SLURM cluster with three servers with totally 352 CPU cores and 3T memory for more than 30 students and more than five platforms in the lab, providing extensive reliability and robustness when using the resources and optimizing high-performance computing resources to facilitate intensive computational tasks by diverse users and platforms for development and deployment.

Research Assistant Aug. 2016 — Aug. 2018
City University of Hong Kong *Hong Kong, China*

- **Bioinformatics Tool Development**: Designed and developed tools for identifying potential short tandem repeat (STR) loci for individual identification and paternity testing, providing 99.9999% evidence in data.
- **Performance Optimization**: Refactored and enhanced existing bioinformatics tools such as tools for fastq cleaning and preprocessing, primarily using C/C++ and multithreading, resulting in at least 2x computational performance and reliability.
- **Lab Resource Administration**: Managed the laboratory's computing environment, including the installation and upkeep of computing clusters, thereby simplifying access to essential software and analytical tools.

Junior software developer (internship) Jul. 2015 — Aug. 2015
BGI Shenzhen *Shenzhen, China*

- **Web Development**: Collaborated on the development of the official website for BGI Forensics, enhancing the organization's digital presence.
- **Data Analysis Tools**: Developed and integrated robust analysis tools for forensic and genetic data. Streamlined the data processing pipeline, thereby improving the overall workflow efficiency.

RESEARCH EXPERIENCES

CeiTEA: Adaptive Hierarchical Clustering of Single Cells with Topological Entropy

- Developed CeiTEA, a novel adaptive hierarchical clustering algorithm using topological entropy to accurately capture complex cellular heterogeneity from single-cell RNA sequencing data.
- Leveraged a unique multi-nary partition tree with integrated bottom-up and top-down strategies to construct precise hierarchical structures, outperforming traditional clustering methods (e.g., Louvain, Leiden, K-means, SEAT) on both simulated and real-world datasets.
- Validated algorithm performance through superior clustering metrics and biological analyses (gene ontology, cell-cell interactions, pseudo-time), demonstrating its utility in mapping tumor heterogeneity and tissue specification.

CNAHap: a germline haplotyping method using tumor allele-specific copy number alteration

- Developed CNAHap, a user-friendly tool for allele-specific copy number aberration estimation and germline variant phasing in tumor genomes.
- Achieved superior performance by generating longer phased blocks and higher accuracy in allele-specific calling compared to benchmark tools, validated on in silico datasets.
- Demonstrated clinical relevance in a hepatocellular carcinoma case study, producing extensive phased blocks (N50: 25M, N90: 7M) and identifying recurrent gene amplifications.

OmicsSankey: Crossing Reduction of Sankey Diagram on Omics Data

- Designed and implemented OmicsSankey, a novel optimization formulation for Sankey diagram layouts in bioinformatics that uses eigen decomposition as a closed-form solution, which addressed scalability and connectivity challenges in large biological networks by implementing a teleportation mechanism to stabilize eigenvector solutions.
- Validated on synthetic and real datasets to demonstrate enhanced diagram interpretability compared to state-of-the-art layout optimizers, improving visual insights in tools like Cell Layers, BioSankey, and Sequence Flow.
- Visualized the resultant diagrams using Plotly with Python.

Optimality of detections for topologically associated domains

- Summarized the objective functions of existing TAD-calling tools and proposed optimizations for those tools by using dynamic programming.
- Implemented optimized algorithms via dynamic programming with memorization tables.

Search for More Effective Microsatellite Markers for Forensics With Next-Generation Sequencing

- Proposed and developed a novel algorithm that identifies customized, high-discriminative STR markers from next-generation sequencing data for forensic applications.
- Enhanced forensic accuracy by reducing DNA profile frequencies and random matching probabilities compared to traditional CODIS 13 markers, with validation on datasets from 320 Chinese individuals.

LocalHap: a haplotype phasing method using local assembly on conjugate graph

- Implemented a novel haplotype phasing algorithm based on conjugate graph local assembly by C/C++, which utilized a conjugate graph to represent the relationship between genomic segments from structural variations occurring in the chromosomes, to reconstruct the segment sequences of individual haplotypes.
- Applied the new tool to various datasets, including simulated and real gut metagenomic data.

SpecHap: a diploid phasing algorithm based on spectral graph theory

- Designed and improved the methodology for phasing diploid genomes using spectral graph theory.

PUBLICATIONS

Bowen Tan, Shiying Li, Mengbo Wang, Shuai Cheng Li, "CeiTEA: Adaptive Hierarchical Clustering of Single Cells with Topological Entropy", *Advanced Science* (2025): 2503539. *Published (IF = 14.3, 2024)*

Bowen Tan, Lingxi Chen, Xuedong Wang, Wenlong Jia, Yanfei Wang, Hechen Li, Shuo Yang, Guangze Pan, Shuai Cheng Li, "CNAHap: a germline haplotyping method using tumor allele-specific copy number alteration". *Accepted by ISBRA 2025*

Shiying Li, **Bowen Tan**, Si Ouyang, Zhao Ling, Miaozhe Huo, Tongfei Shen, Jingwan Wang, Xikang Feng, "OmicsSankey: Crossing Reduction of Sankey Diagram on Omics Data". *Submitted to ICIBM 2025*

Bowen Tan, Zicheng Zhao, Zhe Zhang, Shengbin Li and Shuai Cheng Li, "Search for More Effective Microsatellite Markers for Forensics With Next-Generation Sequencing", *IEEE Transactions on NanoBioscience*, vol. 16, no. 5, pp. 375-381, July 2017, doi: 10.1109/TNB.2017.2712795. *Published (IF = 3.7, 2024)*

Xinyao Miao, Yuesheng Shen, ..., **Bowen Tan**, ..., "A novel forensic panel of 186-plex SNPs and 123-plex STR loci based on massively parallel sequencing", *International Journal of Legal Medicine, 135*, pp.709-718. *Published (IF = 2.2, 2024)*

AWARDS & ACHIEVEMENTS

Patent: Method for Next Generation Sequencing based Genetic Testing, Department of Computer Science, City University of Hong Kong May 2024

- Awarded a patent for developing innovative algorithms that enable accurate individual identification and paternity testing using next-generation sequencing data.

谭博文

+86 135 3827 8377 | 中国深圳/香港 | bowentan78@gmail.com

教育

香港城市大学 2018 年 9 月 — 至今

计算机科学哲学博士 中国香港

- 累计绩点：4.0/4.3

香港城市大学 2012 年 9 月 — 2016 年 7 月

计算数学理学学士 中国香港

- 累计绩点：4.03/4.3 | 院长名、单香港城市大学内地学生奖学金计划、数学系杰出表现奖学金

技能

- **网页开发**：TypeScript/JavaScript（NestJS），Python（Django），GraphQL
- **数据库管理**：MySQL，PostgreSQL，MongoDB 和 Redis
- **CI/CD 与 DevOps**：GitHub Actions，GitLab CI/CD，使用 Docker 进行容器化，以及使用 Openshift 管理云资源.
- **软件开发/数据分析**：C/C++，Python（NumPy，Pandas，Scikit-learn，Matplotlib），R（ggplot2，dplyr，tidyverse），Julia.
- **自动化脚本编程**：Shell/Bash，Python，Lua.
- **语言**：中文，英文，日文

工作经历

项目助理 2024 年 5 月 — 至今

香港理工大学 中国香港

- **后端开发**：负责设计和构建了新版招生系统的后端核心架构，系统基于 NestJS 框架并利用 MongoDB 实现数据持久化。采用模块化设计与微服务架构，实现了功能解耦和高扩展性。在这一过程中，构建了一整套包含全面 RESTful API 的解决方案，涵盖了申请处理、用户身份验证、权限管理以及基于模板的 PDF 自动生成与合并功能。系统经过严格的单元及集成测试，确保在高并发场景下依然稳定、高效。
- 这一创新系统将整个申请流程整合到一个单一的在线平台上，涵盖了学生申请、导师提名以及文职人员的录取工作，确保给予用户的及时响应并提高用户满意度，并大幅减少了学生与各个部门、学院以及辅助单位之间在沟通上所消耗的大量时间和精力。
- **CI/CD 与部署自动化**：主导了自动化流水线的设计与实施，通过 GitLab CI/CD 构建自动化测试和部署流程。利用 Docker 容器化技术和 Openshift 平台，实现了一键式从代码提交到上线的自动化流程，大大减少了手动操作和人为错误，并且通过智能监控和自动回滚机制提高了系统的鲁棒性。
- **版本发布管理**：设计并实现了一套严谨的版本发布管理流程，整合 Git 与 GitLab CI/CD，实现自动更新日志生成、跨分支版本同步以及错误回滚机制。该流程确保了跨团队合作时代码的无缝集成，降低了版本更新中的冲突风险，提高了发布效率和系统稳定性。
- **数据迁移**：负责旧系统到新平台的数据迁移工作，开发了一系列高效且稳固的数据迁移脚本。这些脚本支持大规模数据的批量处理和实时同步，通过数据验证和错误日志监控确保了数据的完整性与一致性，同时设计了完善的备份与恢复方案，以应对突发数据异常。

研究助理 2022 年 11 月 — 2024 年 4 月

香港城市大学 中国香港

- **算法开发**：在基因组数据处理的前沿研究中，针对现有拓扑关联域（TAD）调用工具的局限性，设计并改善了现有的聚类算法。结合了动态规划与图论的核心原理，针对 TAD 结构的特殊性进行优化，保证了最终得到的 TAD 是最优及更具意义的区域。
- **计算资源管理**：负责配置、监控并维护了一个由三台服务器组成的 SLURM 集群，总共拥有 352 个 CPU 核心和 3T 内存，为实验室中超过 30 名学生以及 5 个以上的平台提供服务。通过优化资源调度策略和合理分配计算节点，该集群在资源使用过程中提供了极高的可靠性和稳健性，显著提升了任务并行处理能力，提高了 90% 的资源利用率，满足了不同用户和平台在开发与部署过程中对密集计算任务的需求。

研究助理 2016 年 8 月 — 2018 年 8 月

香港城市大学 中国香港

- **生物信息学工具开发**：深入研究人类基因序列中的结构特征，设计并开发了一套专用于个体识别和亲子鉴定的短串联重复（STR）位点检测工具。该工具采用先进的算法与数据处理技术，能够从大规模测序数据中快速、精准地提取 STR 区域信息，使得数据证据的可信度达到 99.9999%，为后续生物信息研究打下基础。
- **性能优化**：针对原有工具在处理海量基因数据时存在的性能瓶颈问题（例如用于 fastq 数据的清洗和预处理），采用 C/C++对软件进行了全面重构和优化。通过改进内存管理、引入并行计算技术和优化算法逻辑，不仅显著提升了至少 2 倍的运行效率，同时稳定性和可靠性也得到了大幅改善。多次单元测试和性能测试确保了优化后的工具在各种环境下的卓越表现。
- **实验室资源管理**：负责管理实验室计算环境，涵盖高性能计算集群的安装、配置和维护。制定了一整套标准化运维流程，确保科研人员能够方便快捷地访问所需软件和分析工具，有效减少因环境配置问题导致的科研延误，并显著提升了整体数据处理效率。

初级软件开发工程师（实习） 2015 年 7 月 — 2015 年 8 月

BGI 深圳 中国深圳

- **网页开发**：在实习期间参与 BGI 法医部门官方网站的开发，与设计师和开发团队紧密协作，共同打造了响应式、用户友好的网页界面。采用 HTML5、CSS3 和 JavaScript 等现代前端技术，不仅提升了页面加载速度和交互体验，还使网站在各类终端设备上都能保持优良呈现，进一步强化了机构的数字品牌形象。
- **数据分析工具**：针对法医与基因数据分析的实际需求，设计并集成了一套高效的数据处理工具。该工具支持多种数据格式的导入及转换，涵盖数据清洗、统计分析和数据可视化功能。通过优化数据处理流程和算法效率，实现了快速响应及高准确度的数据分析，为后续的数据报告和决策支持提供了有力保障。

研究经历

CeiTEA：基于拓扑熵的单细胞自适应层次聚类

- 设计并实现了 CeiTEA，一种基于拓扑熵的新型自适应层次聚类算法，能够准确捕捉单细胞 RNA 测序数据中复杂的细胞异质性。
- 利用独特的多叉分割树，结合自下而上与自上而下的策略构建精确的层次结构，在模拟数据和真实数据集上均超越了传统聚类方法（如 Louvain、Leiden、K-means、SEAT）。
- 通过聚类指标和生物学分析（基因本体论、细胞间相互作用、伪时间分析）验证了算法性能，证明其在肿瘤异质性与组织分型映射中的应用价值。

CNAHap：基于肿瘤等位基因特异性拷贝数变化的种系单倍型构建方法

- 设计并实现了 CNAHap，一款用于估计肿瘤基因组中的等位基因特异性拷贝数异常并进行种系变异相位分析的工具。
- 通过生成更长的相位块和实现更高的等位基因检测准确性，相较于基准工具取得了显著性能提升，并在计算机模拟数据集上得到验证。
- 在肝细胞癌病例研究中展示了其临床相关性，产生了大面积相位块（N50：25M，N90：7M）并识别出反复出现的基因扩增事件。

OmicsSankey：基于组学数据的桑基图交叉减少

- 设计并实现了 OmicsSankey，一种桑基图布局优化模型，采用特征分解作为封闭解，并通过引入传送机制解决大规模生物网络中的可扩展性与连通性挑战，从而稳定特征向量解。
- 在合成数据及真实数据集上验证，证明其相较于已有的布局优化工具具有更高的图形可解释性，从而提升了例如 Cell Layers、BioSankey 与 Sequence Flow 等工具的视图效果。
- 使用 Python 中的 Plotly 对结果图进行了可视化。

拓扑关联域检测的最优性

- 总结了现有 TAD 检测工具的目标函数，并提出了利用动态规划优化这些工具的方法。
- 采用带记忆化表的动态规划实现了优化算法。

利用二代测序寻找更有效的法医微卫星标记

- 提出并开发了一种新型算法，从二代测序数据中识别定制化、高区分度的短串联重复序列（STR）标记，用于法医应用。
- 通过降低 DNA 型频率和随机匹配概率，相较于传统 CODIS 13 标记显著提高了法医鉴定准确性，并在 320 名中国个体的数据集上获得验证。

LocalHap：基于共轭图局部组装的单倍型构建方法

- 使用 C/C++实现了一种基于共轭图局部组装的新型单倍型构建算法，该算法利用共轭图描述染色体结构变异中各基因组片段之间的关系，以重构个体单倍型的片段序列。

- 将该工具应用于多种数据集，包括模拟数据和真实肠道宏基因组数据。

SpecHap：基于谱图理论的单倍体构建算法

- 设计并改进了利用谱图理论对单倍体基因组进行相位构建的方法。

文章发表

Bowen Tan, Shiying Li, Mengbo Wang, Shuai Cheng Li, “CeiTEA: Adaptive Hierarchical Clustering of Single Cells with Topological Entropy”, Advanced Science (2025): 2503539. 已出版（IF = 14.3, 2024）

Bowen Tan, Lingxi Chen, Xuedong Wang, Wenlong Jia, Yanfei Wang, Hechen Li, Shuo Yang, Guangze Pan, Shuai Cheng Li, “CNAHap: a germline haplotyping method using tumor allele-specific copy number alteration”. 已被 ISBRA 2025 接收

Shiying Li, **Bowen Tan**, Si Ouyang, Zhao Ling, Miaozhe Huo, Tongfei Shen, Jingwan Wang, Xikang Feng, “OmicsSankey: Crossing Reduction of Sankey Diagram on Omics Data”. 已提交至 ICIBM 2025

Bowen Tan, Zicheng Zhao, Zhe Zhang, Shengbin Li and Shuai Cheng Li, “Search for More Effective Microsatellite Markers for Forensics With Next-Generation Sequencing”, IEEE Transactions on NanoBioscience, vol. 16, no. 5, pp. 375-381, July 2017, doi: 10.1109/TNB.2017.2712795. 已出版（IF = 3.7, 2024）

Xinyao Miao, Yuesheng Shen, …, **Bowen Tan**, …, “A novel forensic panel of 186-plex SNPs and 123-plex STR loci based on massively parallel sequencing”, International Journal of Legal Medicine, 135, pp.709-718. 已出版（IF = 2.2, 2024）

奖项与成就

专利：基于下一代测序的基因检测方法, 香港城市大学计算机科学系 2024 年 5 月

- 因开发出基于下一代测序数据的创新算法，实现精确的个体识别和亲子鉴定而获得专利。