



KUAN-HAO CHAO

+1 667-351-6019 | kh.chao@cs.jhu.edu | <https://khchao.com> |  Kuanhao-Chao |  kuan-hao-chao

RESEARCH INTERESTS AND SKILLS

- **Research Interests:** Computational Genomics, Machine Learning, Genomic Variant, Genome Assembly, Transcriptome Assembly, Transcriptomics (RNA-Seq), Genome Annotation, Sequence Indexing, Lift-over
- **Skills:** C/C++, Python, R, Java, Bash | \LaTeX , Git, make, cmake, Snakemake, Docker, Web, Unity, Android

EDUCATION

Johns Hopkins University (JHU)

Ph.D. in Computer Science, Advisor: Steven Salzberg Ph.D. and Mihaela Pertea Ph.D.

Baltimore, MD

08/2021 – Present

The Australian National University (ANU)

Exchange Student in Engineering & Computer Science

Canberra, Australia

07/2019 - 07/2020

National Taiwan University (NTU)





Bachelor of Science in Electrical Engineering

Taipei, Taiwan

09/2016 - 01/2021

PUBLICATIONS

* *CO-FIRST AUTHOR*, † *CORRESPONDING AUTHOR*

1. **Kuan-Hao Chao**^{†*}, Pei-Wei Chen*, S.A. Seshia, and B. Langmead[†], “WGT: Tools and algorithms for recognizing, visualizing and generating Wheeler graphs” in *bioRxiv*, 2022 
2. **Kuan-Hao Chao**[†], A.V. Zimin, M. Pertea, and S.L. Salzberg[†], “The first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual” in *bioRxiv (G3: Genes, Genomes, Genetics)* in press, 2022 
3. **Kuan-Hao Chao**[†], K. Barton, S. Palmer, and R. Lanfear[†], “sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R” in *Genome biology and evolution*, 2020 
4. **Kuan-Hao Chao**, Y.-W. Hsiao, Y.-F. Lee, C.-Y. Lee, L.-C. Lai, M.-H. Tsai, T.-P. Lu, E.Y. Chuang[†], “RNASeqR: an R package for automated two-group RNA-Seq analysis workflow” in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019 

RESEARCH EXPERIENCE

Salzberg & Pertea Lab, Center for Computational Biology, JHU

Advisors: Steven Salzberg Ph.D. & Mihaela Pertea Ph.D.

Baltimore, MD

08/2021 – Present

Software: multiStringTie  C++ |  , Assembly: Han1  , Publication[2]


- Building a multi-sample assembler upon StringTie, the state-of-the-art transcriptome assembler, to construct a more precise gene catalog by taking advantage of the complementarity and redundancy of multiple samples.
- Improving Liftoff by aligning transcriptome in the exon-level to improve the cross-species mapping accuracy.
- Assembled Han1, the first gapless Southern Han Chinese individual genome, using HG00621 PacBio HiFi reads and Oxford Nanopore reads sequenced by Human Pangenome Reference Consortium (HPRC) guided by T2T-CHM13.
- Conducted the first-ever gene content comparison between two fully annotated individuals and investigated homozygous mutations on Han1 compared to T2T-CHM13.

Langmead Lab, Department of Computer Science, JHU

Advisors: Ben Langmead Ph.D. & Sanjit A. Seshia Ph.D. (University of California, Berkeley)

Baltimore, MD

11/2021 – Present

Software: WGT (Wheeler-Graph-Toolkit)  C++, Publication[1]

- Developing algorithms to improve the graph-based sequence indexing and applying it to the pan-genome graph.
- Initiated the collaboration and proposed a new algorithm combining a renaming heuristic with a Satisfiability Modulo Theory (SMT) on solving the Wheeler Graph recognition problem.
- Improved the state-of-the-art Gibney’s and Thankachan’s exponential Wheeler graph recognition algorithm on all Wheeler Graphs and made this NP-complete problem solvable in practice.

HK Tsai Lab, Institute of Information Science, Academia Sinica

Advisor: *Huai-Kuang Tsai Ph.D.*

Software: *SPIFFED*    

Taipei, Taiwan

08/2020 - 02/2021

- Developed a new software that predicts protein-protein interactions using protein co-elution profiles by applying a convolutional neural network.

Lanfear Lab, Division of Ecology and Evolution, ANU

Advisor: *Robert Lanfear Ph.D.*

Software: *sangeranalyseR*  (83 ★)  | 5000+  |  | *Bioconductor 3.16*, *Publication*[3]





Canberra, Australia

07/2019 - 07/2020

- Developed an open-source software to democratize sanger sequencing data analysis in *R* by providing features including reads trimming, base calling, chromatogram plotting, multiple sequence alignments, and phylogenetic analyses.

Bioinformatics Core Lab, Centers of Genomic and Precision Medicine, NTU

Advisors: *Eric Y. Chuang Ph.D.* & *Tzu-Pin Lu Ph.D.*

Software: *RNASeqR*  (11 ★)  | 4500+  |  | *Bioconductor 3.16*, *Publication*[4]

Taipei, Taiwan

01/2018 - 07/2019

- Conducted differential gene expression analysis and pathway enrichment analysis using RNA-Seq data.
- Developed an open-source R package to provide a robust pipeline to automate the standard two-group RNA-Seq analysis in R with several aligner and quantifier options for users to choose from.

INTERNSHIP

Summer Research Intern at Institute of Information Science, Academia Sinica

Supervisor: *Huai-Kuang Tsai Ph.D.*

Taipei, Taiwan

07/2018 - 08/2018

- Investigated tandem repeat lncRNA and double-stranded DNA binding sites by comparing real experiment data to software predictions.

Software Engineer at Institute for Information Industry



Supervisor: *Ching-Hao Mao Ph.D.*

Taipei, Taiwan


12/2017 - 01/2018

- Processed and visualized PC login and logoff history to help the team build a system to find insiders in the company.

PRESENTATIONS

1. **Kuan-Hao Chao**, Y.-W. Hsiao, Y.-F. Lee, C.-Y. Lee, L.-C. Lai, M.-H. Tsai, T.-P. Lu, E.Y. Chuang[†], “RNASeqR: an R package for automated two-group RNA-Seq analysis workflow” at *International Conference on Intelligent Biology and Medicine (ICIBM 2019)*, Columbus, OH, Jun 2019 
2. **Kuan-Hao Chao**^{*}, Y.-S. Chuang^{*}, C.-D. Lin^{*}, “Smart Bus Bell: An Android App to make taking bus easier in Taipei” at *World Congress on Information Technology 2017 (WCIT 2017)*, Taipei, Taiwan, Sep 2017 





























POSTERS

1. **Kuan-Hao Chao**[†], A.V. Zimin, M. Pertea, and S.L. Salzberg[†], “The first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual” in *Biological Data Science 2022 (BDS 2022)*, Cold Spring Harbor, NY, Nov 2022
2. **Kuan-Hao Chao**[†], K. Barton, S. Palmer, and R. Lanfear[†], “sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R” in *Bioconductor Conference 2021 (Bioc 2021)*, Virtual, Aug 2021 

AWARDS



1. **Best Poster Award**, *Bioconductor Conference 2021 (Bioc2021)* 08/2021
2. **College Student Research Fellowship**, *Taiwan Ministry of Science and Technology* 01/2019
3. **Outstanding Prize (1st prize)**, *NTU Centers of Genomics and Precision Medicine Research Contest* 09/2018
4. **Elite Prize (1st prize)**, *2017 HackNTU*, 500+ people Hackathons in Taiwan (TALKS:2 | PROJECTS9) 08/2017

PROJECTS

- I. Multi-sample transcriptome assembly and quantification: multiStringTie** **Baltimore, MD**
Advisors: Steven Salzberg Ph.D. & Mihaela Pertea Ph.D. |  C++ |  *09/2021 - Present*
- Building a multi-sample assembler upon StringTie, the state-of-the-art transcriptome assembler, to construct a more precise gene catalog by taking advantage of the complementarity and redundancy of multiple samples.
 - Inferring transcripts from “universal splicing graphs”, multi-samplewise refined directed splicing graphs representing each gene locus, to achieve better sensitivity-precision tradeoff.
- II. Human genome assembly and annotation: Han1** **Baltimore, MD**
Advisors: Steven Salzberg Ph.D. & Mihaela Pertea Ph.D. |  |  *09/2021 - Present*
- Assembled Han1, the first gapless Southern Han Chinese individual genome, using HG00621 PacBio HiFi reads and Oxford Nanopore reads sequenced by Human Pangenome Reference Consortium (HPRC) guided by T2T-CHM13.
 - Conducted the first-ever gene content comparison between two fully annotated individuals and investigated homozygous mutations on Han1 compared to T2T-CHM13.
- III. Genome annotation refinement and liftover: liftoff2** **Baltimore, MD**
Advisors: Steven Salzberg Ph.D. & Mihaela Pertea Ph.D. |  C++ |  *09/2022 - Present*
- Improving Liftoff by aligning transcriptome in the exon-level to improve the cross-species mapping accuracy.
 - Lifting over T2T-CHM13 annotations to Han1 and created the first gene catalog for a Southern Han Chinese individual.
 - Lifting over CHES annotation from GRCh38 to T2T-CHM13 and compared their gene content difference.
- IV. Wheeler graph fast graph-based sequence indexing algorithm: WGT** **Baltimore, MD**
Advisor: Ben Langmead Ph.D. |   *10/2021 - Present*
- Proposed a new algorithm combining a renaming heuristic with a Satisfiability Modulo Theory (SMT) to improve the Gibney’s and Thankachan’s state-of-art exponential algorithm on solving the Wheeler Graph recognition problem.
 - My Wheeler graph recognition algorithm is several folds faster than G & T’s algorithm on all Wheeler Graphs with different levels of hardness and made this NP complete problem solvable in practice.
 - Developed Wheele-Graph-Toolkit, an first open-source software, to recognize, generate, and visualize Wheeler graphs.
- V. Interactive Sanger sequencing data analysis in R: sangeranalyser** **Canberra, Australia**
Advisor: Robert Lanfear Ph.D. |  (83 ★)  | 5000+  |  | *Bioconductor* *09/2019 - 07/2020*
- Developed an open-source software to democratize sanger sequencing data analysis in *R* by providing features including reads trimming, base calling, chromatogram plotting, mutiple sequence alignments, and phylogenetic analyses.
- VI. Automated two-group RNA-Seq analysis in R: RNASeqR** **Taipei, Taiwan**
Advisors: Eric Y. Chuang Ph.D. & Tzu Pin Lu Ph.D. |  (11 ★)  | 4500+  |  | *Bioconductor* *09/2019 - 07/2020*
- Developed an open-source R package to provide a robust pipeline to automate the standard two-group RNA-Seq analysis in R with several aligner and quantifier options for users to choose from.
- VII. Elution profile-based protein complexes inference algorithm: SPIFFED** **Taipei, Taiwan**
Advisor: Huai-Kuang Tsai Ph.D. |   |  *09/2019 - 07/2020*
- Developing a new software, SPIFFED, by applying a convolutional neural network to improve current state-of-the-art tools, EPIC and PrInCE, on predicting protein-protein interactions (PPIs) using protein co-elution profiles.
- VIII. Unity Games development: Biobaby, Flappy penguin, Survival shooter** **Taipei, Taiwan**
Self-project |  (17 ★) C# |  *Biobaby* |  *Flappy penguin* |  *Survival shooter* *09/2019 - 07/2020*
- Developed three Unity games and deployed them in WebGL on my website.
- IX. Android App development: Smart Bus Bell** **Taipei, Taiwan**
Hackathon project |    |  |  *07/2017 - 09/2017*
- Proposed a novel bus bell idea and built it into two Android apps, user-end and driver-end apps, using Taipei City open data to provide real-time bus information in Taipei.
 - Won the 1st prize in 2017 HackNTU and invited for a talk at *World Congress on Information Technology 2017 (WCIT 2017)*.

TEACHING ASSISTANT

EE-1006: EECS Design and Implementation Cornerstone Course, NTU



Instructor: Shi-Chung Chang Ph.D. |  | 

Taipei, Taiwan

02/2019 - 06/2019

- Lead TA and developer of a 9-people team. Designed lecture materials, developed final project self-tracking car program, and maintained flask server for the final project contest.

CSX-4001: Data Science R Programming, NTU

Instructor: Yun-Cheng Tsai Ph.D. |  | 

Taipei, Taiwan

09/2018 - 01/2019

- Designed and taught workshops explaining basics of Git, R Shiny application, R package development, and Bioconductor R package submission and designed programming assignments.

REVIEWER

G3: Genes, Genomes, Genetics