




Kuan-Hao Chao

PhD Candidate, Computer Science
Johns Hopkins University, USA

kh.chao@cs.jhu.edu
https://khchao.com

 0000-0003-0099-0692
 kuan-hao-chao

 Kuanhao-Chao
 @KuanHaoChao

Research Interests

Genomics, Self-supervised Learning, Graph Neural Network, Genome Assembly, Gene Annotation, Transcriptomics

Education

- **Ph.D. Candidate in Computer Science**, Johns Hopkins University, Baltimore, Maryland, USA. Advisors: Steven Salzberg Ph.D., Mihaela Pertea Ph.D. (08/2021 - Expected: 06/2025)
- **M.S. in Computer Science**, Johns Hopkins University, Baltimore, Maryland, USA. (08/2021 - 05/2023)
- **B.S. in Electrical Engineering**, National Taiwan University, Taipei, Taiwan. (09/2016 - 01/2021)
- **Exchange Student**, The Australian National University, Canberra, Australia. (07/2019 - 07/2020)

Professional Experience

- **Genomics Machine Learning Intern**, Calico Life Sciences LLC, South San Francisco, California. Advisor: David Kelley Ph.D. (Expected 05/2024 - 08/2024)
- **Post-Baccalaureate**, Institute of Information Science, Academia Sinica, Taipei, Taiwan. Advisor: Huai-Kuang Tsai Ph.D. (07/2020 - 02/2021)
- **Research Assistant**, Division of Ecology and Evolution, The Australian National University, Canberra, Australia. Advisor: Robert Lanfear Ph.D. (08/2019 - 06/2020)
- **Student Researcher**, Centers of Genomic and Precision Medicine, National Taiwan University, Taipei, Taiwan. Advisors: Eric Y. Chuang Ph.D., Tzu-Pin Lu Ph.D. (08/2018 - 07/2019)
- **Software Engineer Intern**, Institute for Information Industry, Taipei, Taiwan. Advisor: Ching-Hao Mao Ph.D. (12/2017 - 01/2018)

Peer-Reviewed Journal Papers

* Corresponding author, † Equal contribution

- [J1] **Kuan-Hao Chao**^{*}, Jakob M. Heinz, Celine Hoh, Alan Mao, Mihaela Pertea, Steven L. Salzberg^{*}. 2024. Combining DNA and protein alignments to improve genome annotation with LiftOn. *bioRxiv [Genome Research under revision]*, [Manuscript]
- [J2] **Kuan-Hao Chao**^{*}, Alan Mao, Steven L Salzberg, Mihaela Pertea^{*}. 2023. Splam: a deep-learning-based splice site predictor that improves spliced alignments. *bioRxiv [Genome Biology under revision]*, [Manuscript]
- [J3] **Kuan-Hao Chao**^{†*}, Pei-Wei Chen[†], Sanjit A. Seshia, Ben Langmead^{*}. 2023. WGT: Tools and algorithms for recognizing, visualizing and generating Wheeler graphs. *iScience*, [Manuscript]
- [J4] Yu-Hsin Chen[†], **Kuan-Hao Chao**[†], Jin Yung Wong, Chien-Fu Liu, Jun-Yi Leu^{*}, Huai-Kuang Tsai^{*}. 2023. A feature extraction free approach for protein interactome inference from co-elution data. *Briefings in Bioinformatics*, [Manuscript]
- [J5] **Kuan-Hao Chao**^{*}, Aleksey V Zimin, Mihaela Pertea, Steven L Salzberg^{*}. 2023. The first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual. *G3: Genes, Genomes, Genetics*, [Manuscript]
- [J6] Ales Varabyou^{*}, Markus J Sommer, Beril Erdogdu, Ida Shinder, Iliia Minkin, **Kuan-Hao Chao**, Sukhwan Park, Jakob Heinz, Christopher Pockrandt, Alaina Shumate, Natalia Rincon, Daniela Puiu, Martin Steinegger, Steven L Salzberg^{*}, Mihaela Pertea^{*}. 2023. CHESS 3: an improved, comprehensive catalog of human genes and transcripts based on large-scale expression data, phylogenetic analysis, and protein structure. *Genome Biology*, [Manuscript]
- [J7] Ida Shinder^{*}, Richard Hu, Hyun Joo Ji, **Kuan-Hao Chao**, Mihaela Pertea^{*}. 2023. EASTR: Correcting systematic alignment errors in multi-exon genes. *Nature Communications*, [Manuscript]

- [J8] **Kuan-Hao Chao***, Kirston Barton, Sarah Palmer, Robert Lanfear*. 2021. sangeranalyseR: Simple and Interactive Processing of Sanger Sequencing Data in R. *Genome biology and Evolution*, [Manuscript]
- [J9] **Kuan-Hao Chao**, Yi-Wen Hsiao, Yi-Fang Lee, Chien-Yueh Lee, Liang-Chuan Lai, Mong-Hsun Tsai, Tzu-Pin Lu, Eric Y. Chuang*. 2019. RNASeqR: an R package for automated two-group RNA-Seq analysis workflow. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, [Manuscript]

Awards and Honors

- [A1] Mark O. Robbins Prize, Advanced Research Computing at Hopkins (ARCH), *Baltimore, MD*, 2024
- [A2] Research Highlighted by JHU Whiting School of Engineering and CS Department, URL: <https://engineering.jhu.edu/news/the-human-genome-is-biased-but-rearranging-it-can-help/>, *Baltimore, MD*, 2024
- [A3] Best Poster Award. Bioconductor Conference (Bioc2021), *Virtual, USA*, 2021
- [A4] College Student Research Fellowship. Taiwan Ministry of Science and Technology, *Taipei, Taiwan*, 2019

Invited Talks & Presentation (Selected)

- [T1] “Combining DNA and protein alignments to improve genome annotation with LiftOn”, 20-min talk, International Conference on Intelligent Systems for Molecular Biology (ISMB), *Montreal, Canada*, 2024, Expected: July
- [T2] “Predicting splice sites in DNA sequences with sequence models.”, JHU Joint Biostats-Genomics Lab Meeting Talk, *Baltimore, MD*, 2024, [Video] [Slides]
- [T3] “WGT: Tools and Algorithms for Recognizing, Visualizing and Generating Wheeler Graphs.” Proceeding paper talk, Research on Computational Molecular Biology on Biological Sequence Analysis (RECOMB-Seq). *Istanbul, Türkiye*, 2023. [Video] [Slides] [Poster]
- [T4] “sangeranalyseR: simple and interactive processing of sanger sequencing data in R.” Bioconductor Conference 2021 (Bioc 2021). *Virtual*, 2021. [Poster] [Best Poster Award]

Contributed Open Source Software (Selected)

- [A1] OpenSpliceAI (2024): a versatile framework for splicing prediction across multiple species (primary author). <https://ccb.jhu.edu/spliceai-toolkit/>.
- [A2] LiftOn (2024): combining DNA and protein alignments to improve genome annotation with LiftOn (primary author). <https://ccb.jhu.edu/lifto/>.
- [A3] Splam (2023): a deep-learning-based splice site predictor that improves spliced alignments (primary author; joint with Alan Mao). <https://github.com/Kuanhao-Chao/splam>, <https://github.com/Kuanhao-Chao/LiftOn>
- [A4] Wheeler Graph Toolkit [WGT] (2023): tools and algorithms for recognizing, visualizing and generating Wheeler graphs (primary author; joint with Pei-Wei Chen). https://github.com/Kuanhao-Chao/Wheeler_Graph_Toolkit
- [A5] Han1 genome and annotation (2022): the first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual (primary author). <https://github.com/JHUCCB/ChineseHanSouthGenome>
- [A6] sangeranalyseR (2021): simple and interactive processing of Sanger sequencing data in R (primary author; joint with Robert Lanfear). <https://github.com/roblanf/sangeranalyseR>
- [A7] RNASeqR (2019): an R package for automated two-group RNA-Seq analysis workflow (primary author). <https://github.com/Kuanhao-Chao/RNASeqR>

Professional Service

- **Reviewer:** Genome Research 2024 | BMC Bioinformatics: 2024 | G3: Genes, Genomes, Genetics: 2024 | Chromatographia: 2023
- **Sub-reviewer:** Genome Research: 2024 | International Society for Computational Biology (ISCB): 2024 | Nature Machine Intelligence: 2023 | G3: Genes, Genomes, Genetics: 2022