Kuan-Hao Chao

🆔 0000-0003-0099-0692 | ⬛ Kuanhao-Chao | 💼 kuan-hao-chao | 🐦 @KuanHaoChao

PhD Candidate, Computer Science
Johns Hopkins University, Baltimore, Maryland, USA

kuanhao.chao@gmail.com
https://khchao.com/

Research Interests

Genomics, Machine learning, Self-supervised learning, Genome assembly, Transcriptomics, RNA-Seq, Gene annotation, Annotation lift-over, Sequence alignment, Pan-genomics

Education

- **Ph.D. Candidate in Computer Science**, 08/2021 - Now
  Johns Hopkins University, Baltimore, Maryland, USA
  Advisor: Steven Salzberg Ph.D., Mihaela Pertea Ph.D.

- **M.S. in Computer Science**, 08/2021 - 05/2023
  Johns Hopkins University, Baltimore, Maryland, USA
  Advisor: Steven Salzberg Ph.D., Mihaela Pertea Ph.D.

- **B.S. in Electrical Engineering**, 09/2016 - 01/2021
  National Taiwan University, Taipei, Taiwan

- **Exchange student**, 07/2019 - 07/2020
  The Australian National University, Canberra, Australia

Professional Experience

- **Calico Life Sciences LLC, South San Francisco, CA**
  Kelley Lab, Advisor: David R. Kelley Ph.D.
  Genomics Machine Learning Intern : 05/2024 - 08/2024 (full-time)

- **Institute of Information Science, Academia Sinica, Taipei, Taiwan**
  HK Tsai Lab, Advisor: Huai-Kuang Tsai Ph.D.
  Research Assistant : 07/2020 - 01/2021 (part-time), 01/2021 - 02/2021 (full-time)

- **Division of Ecology and Evolution, Research School of Biology, ANU, Canberra, Australia**
  Lanfear Lab, Advisor: Robert Lanfear Ph.D.
  Student Researcher : 08/2019 - 06/2020 (part-time)

- **Centers of Genomic and Precision Medicine, NTU, Taipei, Taiwan**
  Bioinformatics Core Lab, Advisor: Eric Y. Chuang Ph.D., Tzu-Pin Lu Ph.D.
  Student Researcher : 08/2018-07/2019 (part-time)

- **Institute for Information Industry, Taipei, Taiwan**
  Cybersecurity Technology Institute,   Advisor: Ching-Hao Mao Ph.D.
  Software Engineer Intern : 12/2017-01/2018 (full-time)

Peer-Reviewed Journal Papers

* Corresponding author, † Equal contribution

[J7] **Kuan-Hao Chao**†*, Pei-Wei Chen†, Sanjit A. Seshia, Ben Langmead*. 2023. WGT: Tools and algorithms for recognizing, visualizing and generating Wheeler graphs. *iScience*, https://doi.org/10.1016/j.isci.2023.107402

[J6] 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*, https://doi.org/10.1093/bib/bbad229

[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*, https://doi.org/10.1093/g3journal/jkac321

[J4] Ales Varabyou*, Markus J Sommer, Beril Erdogdu, Ida Shinder, Ilia 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*, https://doi.org/10.1186/s13059-023-03088-4

[J3] Ida Shinder*, Richard Hu, Hyun Joo Ji, **Kuan-Hao Chao**, Mihaela Pertea*. 2023. EASTR: Correcting systematic alignment errors in multi-exon genes. *Nature Communications*, https://doi.org/10.1101/2023.05.10.540179

[J2] **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*, https://doi.org/10.1093/gbe/evab028

[J1] **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*, https://doi.org/10.1109/TCBB.2019.2956708

Preprints

* Corresponding author

[P2] **Kuan-Hao Chao***, Jakob M. Heinz, Celine Hoh, Alan Mao, Alaina Shumate, Mihaela Pertea, Steven L Salzberg*. 2024. Combining DNA and protein alignments to improve genome annotation with LiftOn. *bioRxiv*, https://doi.org/10.1101/2024.05.16.593026

[P1] **Kuan-Hao Chao**[*], Alan Mao, Steven L Salzberg, Mihaela Pertea[*]. 2023. Splam: a deep-learning-based splice site predictor that improves spliced alignments. *bioRxiv*, https://doi.org/10.1101/2023.07.27.550754

Awards and Honors

[A5] 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/, *2024*.

[A4] Best Poster Award. Bioconductor Conference (Bioc2021), *Virtual*, 2021.

[A3] College Student Research Fellowship. Taiwan Ministry of Science and Technology, *Taipei, Taiwan*, 2019

[A2] Outstanding Research Prize (1st prize), NTU Centers of Genomics and Precision Medicine Research Contest, *Taipei, Taiwan*, 2018

[A1] Elite Prize (1st prize), 2017 HackNTU (500+ people Hackathon), *Taipei, Taiwan*, 2017

Invited Talks and Lectures

[T9] "Combining DNA and protein alignments to improve genome annotation with LiftOn" 20-min presentation at Intelligent Systems for Molecular Biology (ISMB), General Computational Biology COSI. *Montréal, Canada*, 2024. Slides: TBD

[T8] "Predicting splice sites in DNA sequences with sequence models." Joint Biostats-Genomics Lab Meeting. *Baltimore, MD, USA*, 2024. Video: https://t.ly/4Gl99; slides: https://t.ly/7VFpO

[T7] "Combining DNA and protein alignments to improve genome annotation with LiftOn" 10-min presentation at Research on Computational Molecular Biology on Biological Sequence Analysis (RECOMB-Seq). *Cambridge, MA*, 2024. Slides: https://t.ly/L92CD

[T6] "WGT: Tools and Algorithms for Recognizing, Visualizing and Generating Wheeler Graphs." Proceeding paper presentation at Research on Computational Molecular Biology on Biological Sequence Analysis (RECOMB-Seq). *Istanbul, Türkiye*, 2023. Video: https://youtu.be/TkX9S024Dk8; slides: https://t.ly/5Q7Ef

[T5] "Enhancing Splice Junction Prediction using Deep Residual Convolutional Neural Networks." Joint Biostats-Genomics Lab Meeting. *Baltimore, MD, USA*, 2023.

[T4] "Differential Gene Expression Analysis of High-throughput RNA-seq Data in R." EN.580.458 / 658 Teaching assistance course lecture. *Baltimore, MD, USA*, 2023.

[T3] "RNASeqR: an R Package for Automated Two-group RNA-Seq Analysis Workflow." Highlight talk at International Conference on Intelligent Biology and Medicine (ICIBM). *Columbus, OH, USA*, 2019.

[T2] "Analyzing RNA-Seq Data in R." NTU Centers of Genomics and Precision Medicine (NTUCGM). *Taipei, Taiwan*, 2018.

[T1] "Taipei City Smart Transportation." Joint presentation with Taipei City Government at World Congress on Information Technology (WCIT). *Taipei, Taiwan*, 2017. https://youtu.be/2ts_NDdfEgQ

Poster Presentations

[PP7] "Combining DNA and protein alignments to improve genome annotation with LiftOn" 10-min presentation at Research on Computational Molecular Biology (RECOMB). *Cambridge, MA*, 2024. https://storage.googleapis.com/storage.khchao.com/poster/RECOMB_LiftOn.pdf

[PP6] "Splam: a deep-learning-based splice site predictor that improves spliced alignments." Genome Informatics (GI 2023). *Cold Spring Harbor, New York, USA.*, 2023. https://storage.googleapis.com/storage.khchao.com/poster/GI_splam.pdf

[PP5] "Splam: a deep-learning-based splice site predictor that improves spliced alignments." Intelligent Systems for Molecular Biology / European Conference on Computational Biology (ISMB/ECCB 2023). *Lyon, France*, 2023. https://storage.googleapis.com/storage.khchao.com/poster/ISMB_splam.pdf

[PP4] "The first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual." Intelligent Systems for Molecular Biology / European Conference on Computational Biology (ISMB/ECCB 2023). *Lyon, France*, 2023. https://storage.googleapis.com/storage.khchao.com/poster/ISMB_Han1.pdf

[PP3] "WGT: Tools and algorithms for recognizing, visualizing and generating Wheeler graphs." Research in Computational Molecular Biology Conference (RECOMB). *Istanbul, Türkiye*, 2023. https://storage.googleapis.com/storage.khchao.com/poster/RECOMB_WGT.pdf

[PP2] "The first gapless, reference-quality, fully annotated genome from a Southern Han Chinese individual." Cold Spring Harbor Laboratory Meeting on Biological Data Science (BDS). *Cold Spring Harbor, New York, USA.*, 2022. https://storage.googleapis.com/storage.khchao.com/poster/BDS_Han1.pdf

[PP1] "sangeranalyseR: simple and interactive processing of sanger sequencing data in R." Bioconductor Conference 2021 (Bioc 2021). *Virtual*, 2021. https://storage.googleapis.com/storage.khchao.com/poster/Bioc_sangeranalyseR.pdf

Contributed Open Source Software

[S7] LiftOn (2024): combining DNA and protein alignments to improve genome annotation with LiftOn (primary author). https://ccb.jhu.edu/lifton/.

[S6] SpliceAI-toolkit (2024): a versatile framework for splicing prediction across multiple species (primary author). https://ccb.jhu.edu/spliceai-toolkit/.

[S5] 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

[S4] 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

[S3] 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

[S2] sangeranalyseR (2021): simple and interactive processing of Sanger sequencing data in R (primary author; joint with Robert Lanfear). https://github.com/roblanf/sangeranalyseR

[S1] RNASeqR (2019): an R package for automated two-group RNA-Seq analysis workflow (primary author). https://github.com/Kuanhao-Chao/RNASeqR

Teaching

- **Department of Computer Science, Johns Hopkins University**
  Teaching Assistant
  EN.580.458 / 658 Computing the Transcriptome, Spring 2023

- **Department of Electrical Engineering, National Taiwan University**
  Teaching Assistant
  CSX 4001 Data Science Programming, Spring 2019
  EE 1006 Cornerstone EECS Design and Implementation, Fall 2018

Mentorship

- **Undergraduate Research Mentor, Department of Computer Science, Johns Hopkins University**

  - *Mentee:* Alan Mao, 05/2023 - Now
  - Guided Alan Mao in benchmarking Splam, a state-of-the-art deep learning tool designed to accurately predict splice sites.

Professional Service

Reviewer

- 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

Leadership

Johns Hopkins University

- Computer Science Ph.D. Student Council, Orientation Chair: 2023

- Graduate Representative Organization, Computer Science Ph.D. Student Liason: 2023

National Taiwan University

- Department of Electrical Engineering Student Organization, Public Relations Chair: 2018

(Last updated: May 21, 2024)