
Haoyu Chenghaoyu.cheng@yale.edu

EDUCATION/TRAINING

Postdoctoral Fellow , Dana Farber Cancer Institute, Harvard Medical School, USA Mentor: Heng Li	2019-present
Ph.D., Computer Science , University of Science and Technology of China, China Mentor: Yun Xu	2013-2019
B.E., Computer Science , Hebei University of Technology, China	2009-2013

POSITIONS AND EMPLOYMENT

Assistant Professor , Biomedical Informatics and Data Science, Yale University, USA	2024-present
Postdoctoral Fellow , Dana Farber Cancer Institute, Harvard Medical School, USA	2019-2024

RESEARCH INTERESTS

I am a tenure-track assistant professor specializing in the development of computational methods for genomic data, particularly focusing on haplotype-resolved genome assembly and its applications. My hifiasm algorithm has been widely used and has already become the dominant long-read genome assembler since its publication in 2021. In under three years, hifiasm has garnered over 2,000 citations. In addition, I have contributed extensively on the practical applications of genome assemblies, working closely with large-scale sequencing projects and consortiums such as the Human Pangenome Reference Consortium (HPRC), the Telomere-to-Telomere (T2T) consortium, and the Vertebrate Genomes Project (VGP).

RESEARCH GRANTS

K99/R00 Pathway to Independence Award , National Institutes of Health (NIH) <i>Robust and cost-effective computational methods for haplotype-resolved genome assemblies</i> Role: PI; Funding: ~\$970,000; Project Number: K99HG012798	2023–2028
Essential Open Source Software for Science , The Chan Zuckerberg Initiative (CZI) <i>Improving computational methods for high-throughput sequence data analysis</i> Role: Key Personnel; Funding: ~\$295,000; Project Number: 2021-237653	2021–2023

PUBLICATIONS AND PREPRINTS

Google Scholar profile: <https://scholar.google.com/citations?user=Vff5EiwAAAAJ&hl=en&oi=sra>

Lead author publications and preprints:

- Cheng H**, Asri M, Lucas J, Koren S, Li H. (2024) Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat Methods*, online ahead of print.
- Cheng H**, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. (2022) Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*, 40(9):1332-1335.
- Cheng H**, Concepcion GT, Feng X, Zhang H, Li H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*, 18:170-175.

4. **Cheng H**, Zhang Y, Xu Y. (2018) BitMapper2: A GPU-Accelerated All-Mapper Based on the Sparse q-Gram Index. *IEEE/ACM Trans Comput Biol Bioinform.* 3;16(3):886-97.
5. **Cheng H**, Wu M, Xu Y. (2018) FMtree: a fast locating algorithm of FM-indexes for genomic data. *Bioinformatics*, 34:416-424.
6. **Cheng H**, Xu Y. (2018) BitMapperBS: a fast and accurate read aligner for whole-genome bisulfite sequencing. *bioRxiv preprint*, bioRxiv:442798.
7. **Cheng H**, Jiang H, Yang J, Xu Y, Shang Y. (2015) BitMapper: an efficient all-mapper based on bit-vector computing. *BMC Bioinformatics*, 16:192.

Co-author publications and preprints:

1. Larivière D, Abueg L, Brajuka N, Gallardo-Alba C, Grüning B, Ko BJ, Ostrovsky A, Palmada-Flores M, Pickett BD, Rabbani K, Balacco JR, Chaisson M, **Cheng H**, Collins J, et al. Scalable, accessible, and reproducible reference genome assembly and evaluation in Galaxy. (2024) *Nat Biotechnol*, 42(3):367-370.
2. Zhang Y, Chu J, **Cheng H**, Li H. De novo reconstruction of satellite repeat units from sequence data. (2024) *Genome Research*, 33(11):1994–2001.
3. Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, Buonaiuto S, Chang XH, **Cheng H**, Chu J, et al. A draft human pangenome reference. (2023) *Nature*, 617(7960):312-24.
4. Jarvis ED, Formenti G, Rhie A, Guarracino A, Yang C, Wood J, Tracey A, Thibaud-Nissen F, Vollger MR, Porubsky D, **Cheng H**, Asri M, et al. Semi-automated assembly of high-quality diploid human reference genomes. (2022) *Nature*, 611(7936):519-531.
5. Feng X, **Cheng H**, Portik D, Li H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. (2022) *Nat Methods*, 19(6):671-674.
6. Wagner J, Olson ND, Harris L, McDaniel J, **Cheng H**, Functammasan A, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. (2022) *Nat Biotechnol*, 40(5):672-680.
7. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt SJ, Diekhans M, Logsdon GA, Alonge M, Antonarakis SE, Borchers M, Bouffard GG, Brooks SY, Caldas GV, Chen NC, **Cheng H**, Chin CS, et al. The complete sequence of a human genome. (2022) *Science*, 376(6588):44-53.
8. Zhang H, Song L, Wang X, **Cheng H**, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, Liu XS, Li H. Fast alignment and preprocessing of chromatin profiles with Chromap. (2021) *Nat Commun*. 12(1):6566.
9. Zhang H, Li H, Jain C, **Cheng H**, Au KF, Li H, Aluru S. Real-time mapping of nanopore raw signals. (2021) *Bioinformatics*. 37(Suppl_1):i477-i483.
10. Garg S, Functammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P, Hatas E, Ghurye J, Maguire J, Mahmoud M, **Cheng H**, Heller D, Zook JM, Moemke T, Marschall T, Sedlazeck FJ, Aach J, Chin CS, Church GM, Li H. (2021) Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol*, 39(3):309-312.

PRESENTATIONS

Oral Presentations:

1. *Scalable Telomere-to-Telomere Assembly for Complex Genomes with HiFiasm*. **International Plant and Animal Genome Conference (PAG 31)**, 2024
2. *Scalable telomere-to-telomere assembly with hifiasm*. **Biodiversity Genomics Academy 2023 (BGA23)**, 2023
3. *Scalable telomere-to-telomere assembly for diploid, polyploid and cancer genomes with double graph*. **Human Pangenome Reference Consortium Annual Meeting**, 2023
4. *An integrated algorithm for robust and cost-effective telomere-to-telomere genome assembly*. **Biological Data Science**, 2022
5. *Robust haplotype-resolved assembly of diploid individuals without parental data*. **Genome Informatics**, 2021
6. *Robust haplotype-resolved assembly of diploid individuals without parental data*. **Human Pangenome Reference Consortium Annual Meeting**, 2021
7. *Haplotype-resolved de novo assembly with phased assembly graphs*. **Genome Informatics**, 2020

Poster Presentations:

1. *Scalable telomere-to-telomere assembly for diploid, polyploid and cancer genomes with double graph*. **Genome Informatics**, 2023
2. *Haplotype-resolved de novo assembly with accurate long reads*. **Genome Informatics**, 2019

AWARDS AND HONORS

National Human Genome Research Institute - Pathway to Independence Award (K99/R00) 2023

TEACHING EXPERIENCE

Scalable telomere-to-telomere assembly with hifiasm, Biodiversity Genomics Academy 2023
Role: Invited Speaker for a hands-on session on how hifiasm works and how to use it

Design and Analysis of Algorithms, University of Science and Technology of China 2016–2018
Role: Teaching Assistant

Parallel Computing, University of Science and Technology of China 2017
Role: Teaching Assistant

PROFESSIONAL SERVICE

1. Ad Hoc Reviewer
Nature Methods, Nature Plants, Nature Communications, Nucleic Acids Research, Bioinformatics, PLOS Computational Biology, Genome Biology and Evolution, NAR Genomics and Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics
2. Society Membership
Human Pangenome Reference Consortium, Telomere to Telomere consortium, Vertebrate Genomes Project