



## Innovations in computational biology: RECOMB 2024 Special Issue

Jian Ma and Bonnie Berger

*Genome Res.* 2024 34: vii-viii

Access the most recent version at doi:[10.1101/gr.280036.124](https://doi.org/10.1101/gr.280036.124)

---

**References** This article cites 20 articles, 20 of which can be accessed free at:  
<http://genome.cshlp.org/content/34/9/vii.full.html#ref-list-1>

**Creative Commons License** This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

# Innovations in computational biology: RECOMB 2024 Special Issue

Jian Ma<sup>1</sup> and Bonnie Berger<sup>2,3</sup>

<sup>1</sup>Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; <sup>2</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; <sup>3</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The 28th International Conference on Research in Computational Molecular Biology (RECOMB 2024) took place from April 29 to May 2, 2024, in Cambridge, Massachusetts, USA. This year, the conference received 352 full paper submissions, with 57 ultimately accepted after a rigorous peer-review process involving at least three reviewers per paper. Authors had the option to submit full-length papers or concise versions to the Conference proceedings while pursuing journal publication elsewhere. A subset of these papers was invited for further review at *Genome Research*, ultimately leading to the publication of this Special Issue.

RECOMB is the leading international conference on algorithmic computational biology, bridging computational, mathematical, statistical, and biological sciences. It is held annually in the Spring and is paired with its typically adjoining satellite conferences. It provides a scientific forum for the cutting-edge theoretical advances in computational biology and their applications in molecular biology and medicine, emphasizing advancements in computational biology methodologies, ranging from algorithmic innovations to developments in AI and machine learning.

*Genome Research* has been a leading journal for publishing high-impact studies on genome structure, function, and their roles in evolution, population variation, and disease. The partnership between RECOMB and *Genome Research* is crucial in bringing novel computational methodologies to a broader audience, fostering their application to significant genomic challenges. This collaboration exemplifies how new computational methods presented at RECOMB can drive discoveries in genomics and molecular biology as a whole.

In this Special Issue of *Genome Research*, we introduce a diverse collection of 20 papers from RECOMB 2024. These include algorithmic innovations in genomic variation analysis, privacy-preserving algorithms, DNA structural properties, cancer genomics, transcriptomic studies, gene regulatory networks, biomolecular representation learning, and metagenomic data analysis, all of which reflect the field's rapid evolution. This year, two cross-cutting algorithmic themes have emerged in the field: (1) Clever sketching and summarization techniques, along with advanced data structures, to handle the massive data sets generated by modern genomic studies, including single-cell, metagenomic, and collaborative large-scale databases; we are seeing papers rise to this challenge. (2) More sophisticated probabilistic models, many inspired by recent advances in statistics and AI/ML, including large language models.

The first five papers in this Special Issue focus on new algorithms for analyzing genomic variation. Chandra et al. (2024)

introduce new algorithms for haplotype-aware sequence-to-pangenome graph alignment. Sens et al. (2024) present a novel framework to integrate multiple genetic risk factors to improve disease risk predictions. Jeong et al. (2024) develop SUM-RHE to estimate heritability using summary statistics, and Fu et al. (2024) introduce QuadKAST to detect epistasis using a scalable algorithm for large data sets. Burch et al. (2024) use matrix sketching to accelerate linear mixed model computations for genome-wide association studies.

The importance of genomic privacy algorithms is highlighted in the next two papers. Hong et al. (2024) introduce SF-Relate, a secure federated algorithm for identifying genetic relatives across large, distributed genomic data sets while ensuring privacy. This paper received the "Best Student/Young Scientist Paper Award." Goldenberg et al. (2024) present a new approach to improve both privacy and efficiency for inferring biological age from DNA methylation data.

DNA structural properties are key to understanding genome function and aberrations. Yang et al. (2024) report SEM to estimate nucleosome positions and subtypes from MNase-seq data. Zhu et al. (2024) and Giurgiu et al. (2024) present algorithms, CoRAL and Decoil, respectively, that leverage long-read sequencing to resolve extrachromosomal DNA structure, with significant implications for cancer research.

Transcriptome and gene regulatory networks remain an active area of research. Zahin et al. (2024) develop TERRACE to assemble full-length circular RNAs from RNA-seq data using a splice graph model. Schrod et al. (2024) introduce SpaCeNet to model both intra- and intercellular molecular interactions from spatial transcriptomics data. DIISCO, by Park et al. (2024), infers dynamic cell-cell interactions from single-cell RNA-seq data, and BONOBO, by Saha et al. (2024), estimates sample-specific gene regulatory networks.

New methods for biomolecular representation learning are also featured in this issue. Lal et al. (2024) present regLM, which uses autoregressive language models to design synthetic *cis*-regulatory elements. Li et al. (2024) propose the ANDES framework for improving gene set similarity analysis using a new gene embedding approach. Additionally, Iovino et al. (2024) introduce a novel approach for protein similarity search based on protein domain embeddings, and Zeng et al. (2024) leverage parameter-efficient fine-tuning of large protein language models to enhance signal peptide prediction.

© 2024 Ma and Berger This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Corresponding authors:** [jianma@cs.cmu.edu](mailto:jianma@cs.cmu.edu), [bab@mit.edu](mailto:bab@mit.edu)  
Article and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280036.124>.

Advancements in metagenomic sequencing analysis are presented in the last two papers. Şapcı and Mirarab (2024) develop KRANK to optimize memory usage for *k*-mer selection in large genomic reference libraries, and Azizpour et al. (2024) introduce GraSSRep, which classifies DNA sequences as repetitive or nonrepetitive using graph neural networks and self-supervised learning in metagenomic assembly graphs.

We would like to thank the authors, reviewers, and the *Genome Research* editorial team, especially Executive Editor Dr. Hillary Sussman, for their efforts and support for the RECOMB–*Genome Research* partnership. We hope readers enjoy these excellent RECOMB 2024 papers and look forward to future submissions from the computational biology community to RECOMB in the coming years.

## Competing interest statement

J.M. served as Program Chair for RECOMB 2024 and had access to earlier versions of all papers included in this Special Issue of *Genome Research* prior to publication. B.B. is the Chair of the Steering Committee for the RECOMB series of conferences and was Co-chair of the Organizing Committee for RECOMB 2024.

## References

- Azizpour A, Balaji A, Treangen TJ, Segarra S. 2024. Graph-based self-supervised learning for repeat detection in metagenomic assembly. *Genome Res* (this issue) **34**: 1468–1476. doi:10.1101/gr.279136.124
- Burch M, Bose A, Dexter G, Parida L, Drineas P. 2024. Matrix sketching framework for linear mixed models in association studies. *Genome Res* (this issue) **34**: 1304–1311. doi:10.1101/gr.279230.124
- Chandra G, Gibney D, Jain C. 2024. Haplotype-aware sequence alignment to pangenome graphs. *Genome Res* (this issue) **34**: 1265–1275. doi:10.1101/gr.279143.124
- Fu B, Anand P, Anand A, Mefford J, Sankararaman S. 2024. A scalable adaptive quadratic kernel method for interpretable epistasis analysis in complex traits. *Genome Res* (this issue) **34**: 1294–1303. doi:10.1101/gr.279140.124
- Giurgiu M, Wittstruck N, Rodriguez-Fos E, Chamorro González R, Brückner L, Kriemelke-Szymansky A, Helmsauer K, Hartebrod A, Euskirchen P, Koche RP, et al. 2024. Reconstructing extrachromosomal DNA structural heterogeneity from long-read sequencing data using Decoil. *Genome Res* (this issue) **34**: 1355–1364. doi:10.1101/gr.279123.124
- Goldenberg M, Mualem L, Shahar A, Snir S, Akavia A. 2024. Privacy-preserving biological age prediction over federated human methylation data using fully homomorphic encryption. *Genome Res* (this issue) **34**: 1324–1333. doi:10.1101/gr.279071.124
- Hong MM, Froelicher D, Magner R, Popic V, Berger B, Cho H. 2024. Secure discovery of genetic relatives across large-scale and distributed genomic data sets. *Genome Res* (this issue) **34**: 1312–1323. doi:10.1101/gr.279057.124
- Iovino BG, Tang H, Ye Y. 2024. Protein domain embeddings for fast and accurate similarity search. *Genome Res* (this issue) **34**: 1434–1444. doi:10.1101/gr.279127.124
- Jeong M, Pazokitoroudi A, Liu Z, Sankararaman S. 2024. Scalable summary-statistics-based heritability estimation method with individual genotype level accuracy. *Genome Res* (this issue) **34**: 1286–1293. doi:10.1101/gr.279207.124
- Lal A, Garfield D, Biancalani T, Eraslan G. 2024. Designing realistic regulatory DNA with autoregressive language models. *Genome Res* (this issue) **34**: 1411–1420. doi:10.1101/gr.279142.124
- Li L, Dannenfels R, Cruz C, Yao V. 2024. A best-match approach for gene set analyses in embedding spaces. *Genome Res* (this issue) **34**: 1421–1433. doi:10.1101/gr.279141.124
- Park C, Mani S, Beltran-Velez N, Maurer K, Huang T, Li S, Gohil S, Livak KJ, Knowles DA, Wu CJ, et al. 2024. A Bayesian framework for inferring dynamic intercellular interactions from time-series single-cell data. *Genome Res* (this issue) **34**: 1384–1396. doi:10.1101/gr.279126.124
- Saha E, Fanfani V, Mandros P, Ben Guebila M, Fischer J, Shutta KH, DeMeo DL, Lopes-Ramos CM, Quackenbush J. 2024. Bayesian inference of sample-specific coexpression networks. *Genome Res* (this issue) **34**: 1397–1410. doi:10.1101/gr.279117.124
- Şapcı AOB, Mirarab S. 2024. Memory-bound *k*-mer selection for large and evolutionarily diverse reference libraries. *Genome Res* (this issue) **34**: 1455–1467. doi:10.1101/gr.279339.124
- Schrod S, Lück N, Lohmayer R, Solbrig S, Völkl D, Wipfler T, Shutta KH, Ben Guebila M, Schäfer A, Beißbarth T, et al. 2024. Spatial Cellular Networks from omics data with SpaCeNet. *Genome Res* (this issue) **34**: 1371–1383. doi:10.1101/gr.279125.124
- Sens D, Shilova L, Gräf L, Grebenschikova M, Eskofier BM, Casale FP. 2024. Genetics-driven risk predictions leveraging the Mendelian randomization framework. *Genome Res* (this issue) **34**: 1276–1285. doi:10.1101/gr.279252.124
- Yang J, Yen K, Mahony S. 2024. Size-based expectation maximization for characterizing nucleosome positions and subtypes. *Genome Res* (this issue) **34**: 1334–1343. doi:10.1101/gr.279138.124
- Zahin T, Shi Q, Zang XC, Shao M. 2024. Accurate assembly of circular RNAs with TERRACE. *Genome Res* (this issue) **34**: 1365–1370. doi:10.1101/gr.279106.124
- Zeng S, Wang D, Jiang L, Xu D. 2024. Parameter-efficient fine-tuning on large protein language models improves signal peptide prediction. *Genome Res* (this issue) **34**: 1445–1454. doi:10.1101/gr.279132.124
- Zhu K, Jones MG, Luebeck J, Bu X, Yi H, Hung KL, Wong IT, Zhang S, Mischel PS, Chang HY, et al. 2024. CoRAL accurately resolves extrachromosomal DNA genome structures with long-read sequencing. *Genome Res* (this issue) **34**: 1344–1354. doi:10.1101/gr.279131.124