

David Pellow

Curriculum Vitae

Department of Computer Science,
University of Toronto
Toronto, M5S 2E4, ON Canada
✉ dpellow@utoronto.ca
🌐 dpellow.github.io

Employment

- 2024-present **Eric and Wendy Schmidt AI in Science Postdoctoral Fellow**, *Department of Computer Science*, University of Toronto, Toronto, Ontario, Canada
- 2023-2024 **Postdoctoral Fellow**, *Vector Institute for Artificial Intelligence*, Toronto, Ontario, Canada
- 2023-present **Researcher**, *University Health Network*, Toronto, Ontario, Canada
- 2022-2023 **Postdoctoral Fellow**, *Weizmann Institute of Science*, Rehovot, Israel

Education

- 2022 **Ph.D. in Computer Science**, *Department of Computer Science*, Tel Aviv University, Tel Aviv, Israel
- 2015 **MS**, *Language Technologies Institute*, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
- 2013 **BASc in Engineering Science (ECE)**, *Faculty of Applied Science & Engineering*, University of Toronto, Toronto, Ontario, Canada

Leadership & Service

- 2025 **Organizer**: *Foundation Models for Science Workshop*. A funded community initiative of Schmidt Sciences, hosted at University of Toronto.
- 2024 **Organizer**: *Red teaming LLMs for healthcare workshop* (at MLHC)
- Reviewing** ACM BCB, ISMB, Oxford Bioinformatics, Nucleic Acids Research, BMC Bioinformatics, Journal of Computational Biology, Nature Aging, Scientific Reports

Awards & Funding

- 2025 **Schmidt Sciences Community Initiative**, *\$100K event funding to organise Foundation Models for Science Workshop*, Co-organisers: Ashley Dale, Biprateep Dey, Ishrath Mohamed-Irshadeen
- 2024-2026 **Schmidt AI in Sciences Postdoctoral Fellowship**, *University of Toronto*
- 2023-2024 **Vector Institute for AI Postdoctoral Fellowship**
- 2022-2023 **Koshland Postdoctoral Prize**, *Weizmann Institute of Science*
- 2016-2019 **Edmond J Safra Bioinformatics Fellowship**, *Tel Aviv University*

Teaching

2019, 2020 **Teaching Assistant**, *Introduction to computer science for math majors*, Tel Aviv University

Supervision & Mentorship

- 2025-present **Kevin Chen**, *Foundation models for the human gut microbiome*, Summer research internship, Vector Institute for AI
- 2025-present **Kevin Deng**, *Foundation models for the human gut microbiome*, CSC494 undergraduate research course, University of Toronto
- 2025 **Jaehyuk (Jay) Choi**, *Improving prediction with LLM generated privileged information*, CSC494 undergraduate research course, University of Toronto
- 2025 **Ben Cheng**, *Predicting next-day Tacrolimus blood levels*, CSC494 undergraduate research course, University of Toronto
- 2021-2022 **Abhinav Dutta**, *Parameterized syncmer schemes improve long-read mapping*, Guided research, Tel Aviv University

Publications

* indicates co-first authors

Refereed articles

13. **Pellow, D.**, Geva, G.A., Godneva, A., Reisner, Y., Talmor-Barkan, Y., Segal, E. (2025). Analysis of biomarkers in the Human Phenotype Project using disease models from UK Biobank. *Accepted, in revision*
12. Zorea, A., Moraïs, S., **Pellow, D.**, Gershoni-Yahalom, O., Probst, M., Nadler, S., Shamir, R., Rosental, B., Elia, N., Mizrahi, I. (2025). ProFiT-SPEci-FISH: A Novel Approach for Linking Plasmids to Hosts in Complex Microbial Communities at the Single-Cell Level. *Accepted, in press*
11. Zorea, A., **Pellow, D.**, Levin, L., Pilosof, S., Friedman, J., Shamir, R., Mizrahi, I. (2024). Plasmids in the human gut reveal neutral dispersal and recombination that is overpowered by inflammatory diseases. *Nature Communications*, 15(1), 3147.
10. **Pellow, D.**, Pu, L., Ekim, B., Kotlar, L., Berger, B., Shamir, R., Orenstein, Y. (2023). Efficient minimizer orders for large values of k using minimum decycling sets. *Genome Research*, 33(7) 1154-1161.
9. Dutta, A.*, **Pellow, D.***, Shamir, R. (2022). Parameterized syncmer schemes improve long-read mapping. *PLoS Computational Biology*, 18(10), e1010638.
8. Flomin, D., **Pellow, D.**, Shamir, R. (2022). Dataset-adaptive minimizer order reduces memory usage in k-mer counting. *Journal of Computational Biology*, 29(8), 825-839.
7. **Pellow, D.**, Zorea, A., Probst, M., Furman, O., Segal, A., Mizrahi, I., Shamir, R. (2021). SCAPP: An algorithm for improved plasmid assembly in metagenomes. *BMC Microbiome*, 9(1).

6. **Pellow, D.**, Mizrahi, I., Shamir, R. (2020). PlasClass improves plasmid sequence classification. *PLoS Computational Biology*, 16(4), e1007781.
5. Orenstein, Y.*, **Pellow, D.***, Marçais, G., Shamir, R., Kingsford, C. (2017). Designing small universal k-mer hitting sets for improved analysis of high-throughput sequencing, *PLoS Computational Biology*, 13(10).
4. Marçais, G., **Pellow, D.**, Bork, D., Orenstein, Y., Shamir, R., Kingsford, C. (2017). Improving the performance of minimizers and winnowing schemes, *Bioinformatics*, 33(14) 110-117.
3. **Pellow, D.**, Filippova, D., Kingsford, C. (2017). Improving Bloom filter performance on sequence data using k-mer Bloom filters, *Journal of Computational Biology*, 24(6) 547-557.
2. Orenstein, Y., **Pellow, D.**, Marçais, G., Shamir, R., Kingsford, C. (2016). Compact Universal k-mer Hitting Sets. In: Frith, M., Storm Pedersen, C. (eds) *Algorithms in Bioinformatics. WABI 2016. Lecture Notes in Computer Science* vol 9838. Springer.
1. **Pellow, D.**, & Eskenazi, M. (2014). An open corpus of everyday documents for simplification tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* pp. 84-93.

Non-refereed articles

1. Balazadeh, V.*, Cooper, M.*, **Pellow, D.***, Assadi, A., Bell, J., Fackler, J., ... & Krishnan, R. G. (2025). Red Teaming Large Language Models for Healthcare. *arXiv preprint* arXiv:2505.00467.

Refereed published abstracts

1. **Pellow, D.**, & Eskenazi, M. (2014). Tracking human process using crowd collaboration to enrich data. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 2) 52-53.

Book chapters

1. **Forthcoming:** Morais, S., Zorea, A., **Pellow, D.**, Mizrahi, I. (2025). Unraveling Plasmid Ecology and Dispersal with ProFiT-SPEci-FISH. In *Horizontal Gene Transfer: Methods and Protocols*, 2nd ed.

Presentations

Refereed talks

5. Development of a Readily Deployable Prognostication Tool for Predicting Major Adverse Cardiovascular Events Following Liver Transplantation *Temerty Centre for AI Research and Education in Medicine (T-CAIREM) Trainee Rounds*. May 2025.
4. SCAPP: An algorithm for improved plasmid assembly in metagenomes. *29th conference on intelligent systems for microbiology (ISMB)*. July 2021.

3. SCAPP: An algorithm for improved plasmid assembly in metagenomes. *24th RECOMB conference, Recomb-SEQ satellite workshop*. September 2021.
2. Improving Bloom filter performance on sequence data using k-mer Bloom filters. *20th annual research in computational biology (RECOMB) conference*. April 2016.
1. An open corpus of everyday documents for simplification tasks. *14th conference of the European Association of Computational Linguistics, 3rd workshop on predicting and improving text readability for target reader populations*. April 2014.

Refereed posters

6. Modelling long-term cardiovascular risk in post liver-transplant patients. *2025 Ajmera Transplant Centre Annual Research Day*. June 2025.
5. Parameterized syncmer schemes improve long read mapping. *Computational challenges in very large scale 'omics conference, Computational innovation and data-driven biology program at the Simons Institute for the Theory of Computing*. July 2022.
4. Improved complete plasmid reconstruction from metagenomic samples. *Ilanit-Fiseb Federation of Israeli societies of experimental biology conference*. February 2020.
3. Improving assembly of plasmid sequences in metagenomic samples. *Frontiers in genetics symposium at the meeting of the Israel Genetics Society*. January 2019.
2. Compact covering of sequences with universal hitting k-mer sets. *Frontiers in genetics symposium at the meeting of the Israel Genetics Society*. February 2018.
1. Tracking human process using crowd collaboration to enrich data. *2nd AAAI conference on human computation and crowdsourcing*. November 2014.