MAHLER REVSINE

3400 N Charles St, Baltimore, MD 21218

mrevsin1@jh.edu

EDUCATION

Johns Hopkins University, *Baltimore, MD, USA*   Fall 2023 - Present
PhD Student in Computer Science
<u>Selected coursework</u>: Applied Comparative Genomics, Machine Learning,
Sketching and Indexing for Sequences, Advances in Self-Supervised Statistical Models,
Advanced Topics in Genomic Data Analysis, Interpretable Machine Learning Design

University of North Carolina, *Chapel Hill, NC, USA*   Fall 2017 - Spring 2021
B.S. in Computer Science, B.A. in Biology

PROFESSIONAL EXPERIENCE

PhD Student   Fall 2023 - Present
Johns Hopkins University, *Baltimore, MD, USA*
Advisor: Michael Schatz

CRTA Postbaccalaureate Fellow   Spring 2021 - Spring 2023
National Cancer Institute, NIH, *Bethesda, MD, USA*
Advisors: Xin Wei Wang, Lichun Ma

Undergraduate Researcher   Winter 2019 - Spring 2021
University of North Carolina, *Chapel Hill, NC, USA*
Advisor: Jeremy Wang

Software Developer Intern   Summer 2020
Epic Systems Corporation, *Verona, WI, USA*
Genomics Team

Software Engineering Intern, NASA Contractor   Summer 2019
Arctic Slope Regional Corporation, *Greenbelt, MD, USA*

Research Assistant   Summer 2018
University of Maryland, *College Park, MD, USA*
Advisor: Adam Fishbein

AWARDS AND HONORS

National Science Foundation Graduate Research Fellowship Program (NSF GRFP)	2025 - Present
Cancer Research Training Award	2021 - 2023
Phi Beta Kappa Society	2021
Graduated with Highest Distinction, University of North Carolina	2021
Dean's List all semesters, University of North Carolina	2017-2021

CURRENT RESEARCH

Dataset deduplication for genomic Large Language Models: Dataset deduplication, the process of removing repeated samples from a training dataset, has been shown to improve the quality of natural language LLMs. Relatedly, DNA datasets contain substantial redundancy, both in terms of repeat elements within genomes and conserved elements shared between them. We are therefore investigating whether dataset deduplication leads to performance gains when applied to genomic LLMs. We developed an algorithm for genomic deduplication and tested it on training datasets spanning from human to the tree of life. Preliminary results show benefits across a range of downstream tasks including splice site classification, promoter detection, and variant effect prediction.

PUBLICATIONS

See Google Scholar for more information.

Breslow, N., Mishra, A., **Revsine, M.**, Schatz, M. C., Liu, A., & Khashabi, D. (2025). Genomic Next-Token Predictors are In-Context Learners. *arXiv preprint arXiv: 2511.12797.*

Behrens, S., Do, W. L., Wang, L., **Revsine, M.**, Maestri, E., Jacob, A., Chang, C., Forgues, M., Sardoo, A. M., Budhu, A., Argemí, J., Sogbe, M., Sangro, B., Greten, T. F., & Wang, X. W. (2025). Pan-microbial serological repertoire as a biomarker of immunotherapy response in hepatocellular carcinoma. *Journal for Immunotherapy of Cancer, 13*(10), e011742.

Ge, Y., Lu, J., Puiu, D., **Revsine, M.**, & Salzberg, S. L. (2025). Comprehensive analysis of microbial content in whole-genome sequencing samples from The Cancer Genome Atlas project. *Science translational medicine, 17*(814), eads6335.

Myojin, Y., Babaei, S., Trehan, R., Hoffman, C., Kedei, N., Ruf, B., Benmebarek, M., Bauer, K. C., Huang, P., Ma, C., Monge, C., Xie, C., Hrones, D., Duffy, A. G., Armstrong, P., Kocheise, L., Desmond, F., Buchalter, J., Galligan, M., Cantwell, C., Ryan, R., McCann, J., Bourke, M., Nicholas, R. M., McDermott, R., Awosika, J., Cam, M., Krebs, R., Budhu, A., **Revsine, M.**, Figg, W. D., Kleiner, D. E., Redd, B., Wood, B. J., Wang, X. W., Korangy, F., Claassen, M., & Greten, T. F. (2025). Multiomics analysis of immune correlatives in hepatocellular carcinoma patients treated with tremelimumab plus durvalumab. *Gut, 74*(6), 983-995.

Gustafson, J. A., Gibson, S. B., Damaraju, N., Zalusky, M. P., Hoekzema, K., Twesigomwe, D., Yang, L., Snead, A. A., Richmond, P. A., Coster, W. D., Olson, N. D., Guarracino, A., Li, Q., Miller, A. L., Goffena, J., Anderson, Z., Storz, S. H., Ward, S. A., Sinha, M., Gonzaga-Jauregui, C., Clarke, W. E., Basile, A. O., Corvelo, A., Reeves, C., Helland, A., Musunuri, R. L., **Revsine, M.**, Patterson, K. E., Paschal, C. R., Zakarian, C., Goodwin, S., Jensen, T. D., Robb, E., The 1000 Genomes ONT Sequencing Consortium, University of Washington Center for Rare Disease Research (UW-CRDR), Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium, McCombie, W. R., Sedlazeck, F. J., Zook, J. M., Montgomery, S. B., Garrison, E., Kolmogorov, M., Schatz, M. C., McLaughlin, R. N., Jr., , Dashnow, H., Zody, M. C., Loose, M., Jain, M., Eichler, E. E., & Miller, D. E. (2024). High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome research, 34*(11), 2061-2073.

Green, B. L., Myojin, Y., Ma, C., Ruf, B., Ma, L., Zhang, Q., Rosato, U., Qi, J., **Revsine, M.**, Wabitsch, S., Bauer, K., Benmebarek, M., McCallen, J., Nur, A., Wang, X., Sehra, V., Gupta, R., Claassen, M., Wang, X. W., Korangy, F., & Greten, T. F. (2024). Immunosuppressive CD29+ Treg accumulation in the liver in mice on checkpoint inhibitor therapy. *Gut, 73*(3), 509-520.

Revsine, M., Wang, L., Forgues, M., Behrens, S., Craig, A. J., Tran, B., Kelly, M., Budhu, A., Monge, C., Xie, C., Hernandez, J. M., Greten, T. F., Ma, L., & Wang., X. W. (2024). Lineage and ecology define liver tumor evolution in response to treatment. *Cell Reports Medicine, 5*(2), 101394.

Wang, L.#, **Revsine, M.**#, Wang, X. W., & Ma, L. (2024). Single-Cell Characterization of the Tumor Ecosystem in Liver Cancer. In *Liver Carcinogenesis: Methods and Protocols* (pp. 153-166). New York, NY: Springer US.

Craig, A. J., Silveira, M. A. D., Ma, L., **Revsine, M.**, Wang, L., Heinrich, S., Rae, Z., Ruchinskas, A., Dadkhah, K., Do, W., Behrens, S., Mehrabadi, F. R., Dominguez, D. A., Forgues, M., Budhu, A., Chaisaingmongkol, J., Hernandez, J. M., Davis, J. L., Tran, B., Marquardt, J. U., Ruchirawat, M., Kelly, M., Greten, T. F., & Wang, X. W. (2023). Genome-wide profiling of transcription factor activity in primary liver cancer using single-cell ATAC sequencing. *Cell reports, 42*(11).

Ruf, B., Bruhns, M., Babaei, S., Kedei, N., Ma, L., **Revsine, M.**, Benmebarek, M., Ma, C., Heinrich, B., Subramanyam, V., Qi, J., Wabitsch, S., Green, B. L., Bauer, K. C., Myojin, Y., Greten, L. T., McCallen, J. D., Huang, P., Trehan, R., Wang, X., Nur, A., Soika, D. Q. M., Pouzolles, M., Evans, C. N., Chari, R., Kleiner, D.

E., Telford, W., Dadkhah, K., Ruchinskas, A., Stovroff, M. K., Kang, J., Oza, K., Ruchirawat, M., Kroemer, A., Wang, X. W., Claassen, M., Korangy, F., & Greten, T. F. (2023). Tumor-associated macrophages trigger MAIT cell dysfunction at the HCC invasive margin. *Cell, 186*(17), 3686-3705.

Budhu, A., Pehrsson, E. C., He, A., Goyal, L., Kelley, R. K., Dang, H., Xie, C., Monge, C., Tandon, M., Ma, L., **Revsine, M.**, Kuhlman, L., Zhang, K., Baiev, I., Lamm, R., Patel, K., Kleiner, D. E., Hewitt, S. M., Tran, B., Shetty, J., Wu, X., Zhao, Y., Shen, T., Choudhari, S., Kriga, Y., Ylaya, K., Warner, A. C., Edmonson, E. F., Forgues, M., Greten, T. F., & Wang, X. W. (2023). Tumor biology and immune infiltration define primary liver cancer subsets linked to overall survival after immunotherapy. *Cell Reports Medicine, 4*(6).

Wang, J., Bhakta, N., Miller, V. A., **Revsine, M.**, Litzow, M. R., Paietta, E., Fedoriw, Y., Roberts, K. G., Gu, Z., Mullighan, C. G., Jones, C. D., & Alexander, T. B. (2022). Acute leukemia classification using transcriptional profiles from low-cost nanopore mRNA sequencing. *JCO Precision Oncology, 6*, e2100326.

PRESENTATIONS

Invited Talks

Mahler Revsine, Alex Ostrovsky, Xingyu Chen, Daniel Khashabi, Michael C. Schatz. "Pre-training Dataset Deduplication Improves Genomic LLMs." RECOMB-seq. Santa Fe, NM, USA. (2025)

Poster presentations

Mahler Revsine, Alex Ostrovsky, Xingyu Chen, Daniel Khashabi, Michael C. Schatz. "Pre-training Dataset Deduplication Improves Genomic LLMs." AI in Molecular Biology. Santa Fe, NM, USA. (2025)

Mahler Revsine, Xingyu Chen, Daniel Khashabi, Michael C. Schatz. "Pre-training Dataset Deduplication Improves Genomic LLMs." RECOMB. Yonsei University, Seoul, South Korea. (2025)

Mahler Revsine, Michael C. Schatz. "Transformer transparency: deciphering the decision-making process of genomic large language models." Biological Data Science. Cold Spring Harbor Laboratory, Laurel Hollow, NY, USA. (2024)

Mahler Revsine, Lichun Ma, Xin Wei Wang. "Single-cell Dissection of Liver Timor Evolution in Response to Immunotherapy." NIH Lab of Human Carcinogenesis Seminar. Bethesda, MD, USA. (2022)

Mahler Revsine, Lichun Ma, Xin Wei Wang. "Exploring the Single-Cell Evolution of Liver Tumors during Immunotherapy." NIH Cancer Data Science Lab Seminar. Virtual. (2022)

Mahler Revsine, Lichun Ma, Xin Wei Wang. "Understanding Liver Tumor Evolution after Immunotherapy via Single-Cell Analysis." NIH Postbac Poster Day. Virtual. (2022)

TEACHING AND MENTORSHIP

Teaching Assistant, EN.601.449/649 Applied Comparative Genomics   Fall 2025
Professor: Michael Schatz

First year graduate student mentor   Fall 2025
Mentee: Alex Ostrovsky

First year graduate student mentor   Spring 2025
Mentee: Xingyu Chen

SERVICE

Co-founder and organizer of the Johns Hopkins Deep Learning in Genomics Journal Club   Fall 2024 - Present
Description: A student-run organization for discussing the latest
literature at the intersection of Deep Learning and Genomics.
Visit our website to follow along with our biweekly meetings!

SKILLS

Programming languages: Python, Bash, R, Java, C
Deep Learning Platforms: PyTorch, Composer, HuggingFace, W&B