# RESEARCHER PROFILE

**Prof. Yaron Orenstein**

**Email:** yaron.orenstein@biu.ac.il

## HORIZON EUROPE TOPIC(S) OF INTEREST:

***HORIZON-HLTH-2023-TOOL-05-03: Integrated, multi-scale computational models of patient patho-physiology ('virtual twins') for personalised disease management***

## CONTRIBUTIONS TOWARD CALL TOPIC

The Orenstein lab develops algorithms to infer predictive models of molecular interactions based on high-throughput biological data. We are mostly experienced with deep neural network applications on genomic datasets of thousands to millions of measurements. We are also experienced in interpreting the trained models to discover new biological hypotheses. Moreover, we develop algorithms to design efficient high-throughput experiments so that the resources are utilized to retrieve as much information as possible.

Our research focuses on computational modeling of protein-DNA and -RNA interactions. Protein-DNA and -RNA interactions are key regulators of gene expression, and as such are involved in almost any process in the cell, including many diseases. Technologies measuring these interactions produce thousands and millions of data points in a single experiment, which can only be analyzed computationally. In our research, we develop efficient computational methods to analyze the data and produce accurate models to predict new interactions and better understand the process at test. To handle the vast amounts of data, I develop efficient methods to process the data and extract relevant information on which I apply learning methods to infer accurate predictive models.

We are very excited to utilize the most advanced machine learning methods to generate more accurate protein-DNA, -RNA and -peptide binding models. The recent advancement in neural networks, termed deep learning, has attracted much attention in the computational biology field. We are applying it successfully to many high-throughput datasets, and plan to take it even further by incorporating several orthogonal sources to improve in vivo binding prediction.

In addition, we develop algorithms to generate compact universal sequence libraries under different biological constraints to improve experimental throughput and enable novel discoveries. We use our experience and knowledge in graph theory to solve combinatorial problems in sequence design, which include, for example, flow and matching algorithms, ILP formulations and greedy heuristics.

## BRIEF PROFILE

In my research, I develop algorithms and bioinformatics methods to solve biological problems arising from high-throughput genomic datasets. In this era of cheap and easy production of big biological datasets, it has become critical to develop methods to extract information from these datasets in order to decipher the biological phenomenon at test and enable prediction to new unobserved molecules and species. I utilize the most advanced machine-learning techniques, i.e., deep neural networks, to learn the molecular principles underlying these biological processes. By learning these principles, I can then predict similar processes over new DNA or RNA molecules and proteins. Just like the revolution induced by deep neural networks in image processing and text analysis, I witness the same revolution taking place in the application of artificial intelligence in genomics, and my research is at the center of this revolution.

My BSc, MSc, and PhD were all conducted in Computer Science and Electrical Engineering at Tel-Aviv University. I continued to post-doctoral training at the Computer Science and Artificial Intelligence Laboratory at Massachusetts Institute of Technology. I also spent a semester as a research fellow at the Simons Institute for Theory of Computing at UC Berkeley.

Bar-Ilan University has a thriving bioinformatics community spanning Life Sciences, Computer Sciences, Mathematics, and Engineering. Thus, it serves as an ideal professional environment for me to develop and lead my research group with excellent students coming from both computational and bioinformatics backgrounds. My shared appointment at the Faculty of Life Sciences and Department of Computer Sciences provides a unique opportunity to build a Computational Biology Center at Bar-Ilan University spanning researchers from various faculties to collaborate and benefit this multi-disciplinary field of research. Moreover, Bar-Ilan University’s researchers and central location in Israel enable it to reach out to community activities and industrial links, with both enhancing the quality and throughput of my research.

## RELEVANT PUBLICATIONS

1. **Y. Orenstein** and R. Shamir (2014) A comparative analysis of transcription factor binding models learned from PBM, HT-SELEX and ChIP data. *Nucleic Acid Research,* 42(8):e63
2. L. Yang*, **Y. Orenstein***, A. Jolma, Y. Yimeng, J. Taipale, R. Shamir^ and R. Rohs^ (2017) Transcription factor family-specific DNA shape readout revealed by quantitative specificity models. *Molecular Systems Biology*, 13(2):910. **Top 10 papers reading list in Regulatory and Systems Genomics**
3. **Y. Orenstein**, Y. Wang and B. Berger (2016) RCK: accurate and efficient inference of sequence and structure-based protein-RNA binding models from RNAcompete data. *Bioinformatics*, 32(12):i351-i359. **Best paper award**
4. I. Ben-Bassat, B. Chor and **Y. Orenstein** (2018) A deep neural network approach for learning intrinsic protein-RNA binding preferences. *Bioinformatics*, 34(17):i638-i646
5. **Y. Orenstein***, D. Pellow*, G. Marcais, R. Shamir and C. Kingsford (2017) Designing small universal k-mer hitting sets for improved analysis of high-throughput sequencing. *PLoS Computational Biology*, 13(10):e1005777

## BAR ILAN UNIVERSITY PROFILE

Established in 1955, Bar Ilan University (BIU) is currently one of Israel's largest universities with a total undergraduate and graduate student enrollment of 19,000. With more than 1,600 senior and junior faculty members, BIU has achieved an international reputation for academic and research excellence, especially, but not limited to the fields of artificial intelligence, renewable energy, bio-medicine, brain sciences, cancer, cyber security, cognitive sciences, environment, quantum technologies, medicine, archaeology, nanotechnology and advanced materials.

Building on our past and current successes in FP6, FP7, H2020 and ERC projects, BIU is committed to strengthening its research and innovation infrastructure and supporting multidisciplinary innovative research initiatives with its 55 research centers and 60 endowed chairs. In addition, both the Bar Ilan Center for Smart Cities and Bar Ilan's Institute of Nanotechnology and Advanced Materials (BINA) are recognized by the EU SMART SPECIALISATION PLATFORM as Digital Innovation Hubs.

The Orenstein lab resides jointly in the Department of Computer Science and Faculty of Life Sciences. It has access to vast computational power via both on-premise dedicated servers, the BIU data science compute cloud, the Computer Science infrastructure, and BIU high-performance cluster. Moreover, it has access to compute cloud of common vendors, such as AWS, GCP, and Azure. There are currently six graduate students, and more are expected to be recruited.