

## Curriculum Vitae

Yaron Orenstein

Personal web-page: <https://www.ee.bgu.ac.il/~yaronore/> Lab web-page: <https://www.ee.bgu.ac.il/~cb/>

Research Interests: bioinformatics; computational biology; deep learning in genomics; string algorithms

## Employment

- 2017-present **Ben-Gurion University of the Negev**  
Senior Lecturer at the Department of Electrical and Computer Engineering
- 2014-2017 **Massachusetts Institute of Technology**  
Post-doctoral researcher in Computer Science. Advisor: Prof. Bonnie Berger  
Research: Computational modeling of protein-RNA interactions
- 2016 **University of California Berkeley**  
Research fellow, Simons Institute for the Theory of Computing

## Education

- 2010-2014 **Tel-Aviv University**  
PhD in Computer Science. GPA: 96.7  
Thesis: Discovering Motifs Using High-Throughput in Vitro Data. Advisor: Prof. Ron Shamir
- 2007-2010 **Tel-Aviv University**  
MSc in Electrical Engineering. Summa Cum Laude. GPA: 95.87  
Thesis: Property Testing in Directed Graphs. Advisor: Prof. Dana Ron
- 2004-2008 **Tel-Aviv University**  
BSc in Electrical Engineering and Computer Science. Summa Cum Laude. GPA: 96.37

## Awards and Fellowships

- 2018 Excellent teacher in Electrical and Computer Engineering
- 2017 Top 10 papers reading list in Regulatory and Systems Genomics
- 2017 RECOMB travel fellowship
- 2017 ILANIT/FISEB conference travel fellowship
- 2016 ISMB best paper award
- 2013, 2016 ISMB travel fellowship
- 2016 The Simons Institute for the Theory of Computing research fellowship
- 2014 Deutsch Prize for excellence in PhD research
- 2014 Best poster award, Edmond J. Safra Bioinformatics Retreat
- 2012-2014 In each year, enlisted among the top 100 teaching assistants at TAU
- 2010-2014 In each year, Safra fellowship for Ph.D. students in bioinformatics at TAU
- 2013 I-CORE: Israeli Center of Research Excellence fellowship
- 2012 Dan David fellowship for excellent research in the field of Genomics
- 2012 Excellent teaching assistant in Computer Science
- 2011 Best performer, DREAM 6 AML Patient Classification Challenge
- 2010 Best performer, DREAM 5 DNA-Motif Recognition Challenge
- 2009 Academic Excellence Award commemorating fallen Israeli soldiers, Engineering (M.Sc.)
- 2009 Wolf Foundation award for excellent M.Sc. research
- 2004-2008 The Adi Lautman Program for Outstanding Students, fellowship
- 2004-2008 Dean's list of excellent students in Engineering in all four years of B.Sc. studies
- 2004-2008 Dean's list of excellent students in Computer Science in each eligible year of B.Sc. studies
- 2007 Academic Excellence Award commemorating fallen Israeli soldiers, Computer Science (B.Sc.)
- 2007 Cisco Award for excellent students in Computer Science

## Grants

2022-2026	860,000\$ National Institute of Health (NIH) (co-PIs: E. Radisky, N. Papo, T. Caulfield)
2022-2023	425,000\$ Israel Innovation Authority (co-PI: R. Amit)
2021-2025	337,000\$ Israel Science Foundation (ISF)
2021-2022	75,000\$ United States-Israel Binational Science Foundation (BSF) (co-PI: B. Berger)
2021	12,000\$ BGU-NJIT seed grant (co-PI: Z. Wei)
2021	9,350\$ Israeli Council for Higher Education via Data Science Research Center, BGU
2021	13,700\$ Israel Cancer Association
2020-2023	290,000\$ Israel Innovation Authority (CRISPRIL consortium)
2020-2021	20,000\$ MIT-Israel BGU Seed Fund (co-PI: B. Berger)
2020	14,670\$ Israeli Council for Higher Education via Data Science Research Center, BGU
2019	6,000\$ Engineering at BGU research support
2019	20,000\$ Neuroscience at BGU collaboration grant (co-PI: R. Birnbaum)
2019	20,000\$ Health-Engineering at BGU collaboration grant (co-PI: D. Levy)
2019	4,000\$ Life Sciences at BGU multidisciplinary grant (co-PI: R. Birnbaum)
2018	20,000\$ from Augmanity Nano Ltd. for consulting on computational analysis of SELEX data
2018	1,000\$ and computational resources from Intel Ltd. for teaching deep learning in comp. biology

## Students

2020-Present	Reut Moshe, MSc (joint with Prof. Niv Papo)
2020-Present	Maor Turner, MSc
2020-Present	Ori Feldman, MSc
2019-Present	Ofir Yaish, PhD
2019-Present	Shai Elkayam, MSc
2019-Present	Sofia Aizenshtein, MSc
2019-2021	Maor Asif, MSc
2018-2021	Yael Ben-Ari, MSc (joint with Prof. Ron Shamir)
2018-2021	Mira Barshai, MSc
2018-2020	Eitamar Tripto, MSc (joint with Prof. Ilana Nisky)
2018-2020	Yifat Dar, MSc (joint with Dr. Kobi Cohen)
2018-2020	Yahel Salomon (joint with Dr. Ramon Birnbaum)
2018-2021	Shaked Naor, MSc (joint with Prof. Michal Ziv-Ukelson)
2018-2019	Ilia Goland, MSc
2014-2018	Maor Dan, MSc (joint with Prof. Ron Shamir)

## Visitors

2021	Stephan Sinn, post-doc, KIT, Germany
2021	Evelyn G. Peters, BSc, MIT, USA
2020	Alice Aubert, BSc, Polytechnique, France
2019	Baris Ekim, BSc, MIT, USA
2019	Tyler Shimko, PhD, Stanford, USA

## Grant Reviews

Israel Science Foundation (ISF)  
United States-Israel Binational Science Foundation (BSF)  
French National Research Agency – ANR  
Biotechnology and Biological Science Research Council (BBSRC)

## Program Committees

2021	Machine Learning in Computational Biology (MLCB2021)
2021	International Conference on Research in Computational Molecular Biology (RECOMB2021)
2021	Intelligent Systems for Molecular Biology and European Conference on Computational Biology (ISMB/ECCB2021)
2020	Intelligent Systems for Molecular Biology (ISMB2020)
2019	Intelligent Systems for Molecular Biology and European Conference on Computational Biology (ISMB/ECCB2019)

## Conference Reviews

2017-Present	International Conference on Research in Computational Molecular Biology (RECOMB)
2017-Present	Intelligent Systems for Molecular Biology (ISMB)

## Paper Reviews

Bioinformatics, Briefings in Functional Genomics, Briefings in Bioinformatics, Nucleic Acids Research, PLoS Computational Biology, Frontiers in Genetics, Nature Biotechnology, Genome Biology

## Associate Editor

2019-Present Frontiers in Genetics

## Conference Organization

2022	RECOMB - poster track chair
------	-----------------------------

## Teaching

2018-	Lecturer, course on Deep Learning in Computational Biology
2017-	Lecturer, course on Computer Architecture for Computer Engineering students
2011-2014	TA, courses on Computational Genomics for Computer Science students
2010-2014	TA, courses on Data Structures for Computer Science students
2010, 2013	TA, workshop in bioinformatics for Computer Science students
2007-2010	TA, courses on Data Structures and Algorithms for Electrical Engineering students

## Conference Proceedings

1. Y. Ben-Ari, D. Flomin, L. Pu, **Y. Orenstein** and R. Shamir (2021) Improving the efficiency of de Bruijn graph construction using compact universal hitting sets. *The 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*
2. J. Karin, H. Michel and **Y. Orenstein** (2021) MultiRBP: multi-task neural network for protein-RNA binding prediction. *The 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*
3. B. Ekim, B. Berger and **Y. Orenstein** (2020) A randomized parallel algorithm for efficiently finding near-optimal universal hitting sets. *Proceedings of the 24<sup>th</sup> Annual International Conference on Research in Computational Molecular Biology (RECOMB)*
4. M. Barshai and **Y. Orenstein** (2019) Predicting G-quadruplexes from DNA sequences using multi-kernel convolutional neural networks. *The 10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*
5. **Y. Orenstein** (2018) Reverse de Bruijn: utilizing reverse peptide synthesis to cover all amino acid k-mers. *Proceedings of the 22<sup>nd</sup> Annual International Conference on Research in Computational Molecular Biology (RECOMB)*
6. **Y. Orenstein**, R. Kim, P. Fordyce and B. Berger (2017) Joker de Bruijn: sequence libraries to cover all k-mers using joker characters. *Proceedings of the 21<sup>st</sup> Annual International Conference on Research in Computational Molecular Biology (RECOMB)*
7. **Y. Orenstein**, D. Pellow, G. Marçais, R. Shamir and C. Kingsford (2016) Compact universal k-mer hitting sets. *Proceedings of the 16<sup>th</sup> Annual Workshop on Algorithms in Bioinformatics (WABI)*

8. **Y. Orenstein** and B. Berger (2015) Efficient design of compact unstructured RNA libraries covering all k-mers. *Proceedings of the 15<sup>th</sup> Annual Workshop on Algorithms in Bioinformatics (WABI)*

## Book Chapters

9. **Y. Orenstein** (2021) Improved analysis of high-throughput sequencing data using small universal k-mer hitting sets. *Deep Sequencing Data Analysis*

## Journal Publications [\* equal contribution]

10. O. Yaish and **Y. Orenstein** (2021) Computational modeling of mRNA degradation dynamics using deep neural networks. *Bioinformatics*, btab800
11. E. Tripto and **Y. Orenstein** (2021) A comparative analysis of RNA-binding proteins binding models learned from RNAcompete, RNA Bind-n-Seq and eCLIP data. *Briefings in Bioinformatics*, bbab149
12. M. Barshai\*, A. Aubert\* and **Y. Orenstein** (2021) G4detector: Convolutional neural network to predict DNA G-quadruplexes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*
13. N. Katz, E. Tripto, N. Granik, S. Goldberg, O. Atar, Z. Yakhini, **Y. Orenstein** and R. Amit (2021) Overcoming the design, build, test bottleneck for synthesis of nonrepetitive protein-RNA cassettes. *Nature Communications* 12(1):1:15. **Covered in ynet:** <https://www.ynet.co.il/health/article/B1W006115FO>
14. M. Asif and **Y. Orenstein** (2020) DeepSELEX: Inferring DNA-binding preferences from HT-SELEX data using multi-class CNNs. *Bioinformatics*, 36(Supplement\_2):i634-i642
15. M. Barshai, E. Tripto and **Y. Orenstein** (2020) Identifying regulatory elements via deep learning. *Annual Reviews of Biomedical Data Science*, 3:315-338
16. **Y. Orenstein** (2020) Reverse de Bruijn: Utilizing Reverse Peptide Synthesis to Cover All Amino Acid k-mers. *Journal of Computational Biology*, 27(3), 376-385
17. O. Adato, **Y. Orenstein**, J. Kopolovic, T. Juven-Gershon and R. Unger (2020) Quantitative Analysis of Differential Expression of HOX Genes in Multiple Cancers. *Cancers*, 12(6):1572
18. T.C. Shimko, P.M. Fordyce and **Y. Orenstein** (2020) DeCoDe: degenerate codon design for complete protein-coding DNA libraries. *Bioinformatics*, 35(11):3357-3364
19. S. Ray, D. Tillo, R.E. Boer, N. Assad, M. Barshai, G. Wu, **Y. Orenstein**, D. Yang, J.S. Schneekloth Jr and C. Vinson (2020) Custom DNA Microarrays Reveal Diverse Binding Preferences of Proteins and Small Molecules to Thousands of G-Quadruplexes. *ACS Chemical Biology*, 15(4),925-935
20. **Y. Orenstein**, Y.W. Yu and B. Berger (2018) Joker de Bruijn: covering k-mers using joker characters. *Journal of Computational Biology*, 25(11):1171-1178
21. 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
22. D. Le, T. Shimko, A. Aditham, A. Keys, **Y. Orenstein** and P. Fordyce (2018) Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proceedings of the National Academy of Sciences*, 201715888
23. **Y. Orenstein**, U. Ohler and B. Berger (2018) Finding RNA structure in the unstructured RBPome. *BMC Genomics*, 19(1):154
24. **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
25. 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**
26. G. Marcais, D. Pellow, D. Bork, **Y. Orenstein**, R. Shamir and C. Kingsford (2017) Improving the performance of minimizers and winnowing schemes. *Bioinformatics*, 33(14):i110-i117
27. **Y. Orenstein**, R. Puccinelli, R. Kim, P. Fordyce and B. Berger (2017) Optimized sequence library design for efficient in vitro interaction mapping. *Cell Systems*, 5(3):230-236
28. **Y. Orenstein** and R. Shamir (2016) Modeling protein-DNA binding via high throughput in vitro technologies. *Briefings in Functional Genomics*, elw030

29. D. Chen\*, **Y. Orenstein\***, R. Golodnitsky, M. Pellach, D. Avrahami, C. Wachtel, A. Ovadia-Shochat, H. Shir-Shapira, A. Kedmi, T. Juven-Gershon, R. Shamir and D. Gerber (2016) SELMAP - SELEX Affinity Landscape Mapping of transcription factor binding sites using integrated microfluidics. *Scientific Reports*, 6
30. **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**
31. **Y. Orenstein** and B. Berger (2015) Efficient design of compact unstructured RNA libraries covering all k-mers. *Journal of Computational Biology*, 23(2):67-79
32. Y. Glick\*, **Y. Orenstein\***, D. Chen, D. Avrahami, T. Zor, R. Shamir and D. Gerber (2015) Integrated microfluidic approach for quantitative high-throughput measurements of transcription factor binding affinities. *Nucleic Acid Research*, 44(6):e51
33. A. Sloutskin, Y. Danino, **Y. Orenstein**, Y. Zehavi, T. Doniger, R. Shamir and T. Juven-Gershon (2015) ElemeNT: A computational tool for detecting core promoter elements. *Transcription*, 6(3):41-50
34. A. Kedmi, Y. Zehavi, Y. Glick, **Y. Orenstein**, D. Ideses, C. Wachtel, T. Doniger, H. Waldman Ben-Asher, N. Muster, J. Thompson, S. Anderson, D. Avrahami, J.R. Yates, R. Shamir, D. Gerber and T. Juven-Gershon (2014) Drosophila TRF2 is a preferential core promoter regulator. *Genes and Development*, 28(19):2163-2174
35. **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
36. **Y. Orenstein** and R. Shamir (2013) Design of shortest double-stranded DNA sequences covering all k-mers with applications to protein-binding microarrays and synthetic enhances. *Bioinformatics*, 29(13):i71-i79
37. **Y. Orenstein**, E. Mick and R. Shamir (2013) RAP: accurate and fast motif finding based on protein binding microarray data. *Journal of Computational Biology*, 20(5):375-382
38. MT. Weirauch, *et al.* (**Y. Orenstein** as part of DREAM5 consortium) (2013) Evaluation of methods for modeling transcription factor sequence specificity. *Nature Biotechnology*, 31(2):126-134
39. **Y. Orenstein**, C. Linhart and R. Shamir (2012) Assessment of algorithms for inferring positional weight matrix motifs of transcription factor binding sites using protein binding microarray data. *PLoS One*, 7(9):e46145
40. **Y. Orenstein** and D. Ron (2011) Testing Eulerianity and connectivity in directed sparse graphs. *Theoretical Computer Science*, 412(45):6390-6408

## Refereed Conference talks

DeCoDe: degenerate codon design for complete protein-coding DNA libraries. ACM-BCB, Online, August 2020

Algorithms for universal, unbiased and compact DNA and amino acid libraries. The 2<sup>nd</sup> Israeli Synthetic Biology Meeting, Rehovot, Israel, March 2019

Reverse de Bruijn: utilizing reverse peptide synthesis to cover all amino acid k-mers. RECOMB, Paris, France, April 2018

Computational modeling of protein-RNA interactions. ACM-BCB, Boston, August 2017

Joker de Bruijn: sequence libraries to cover all k-mers using joker characters. RECOMB, Hong Kong, China, May 2017

Structure-based prediction of protein-RNA interactions on a transcriptome-wide scale. ILANIT, Eilat, Israel, February 2017

Structural analysis of the compendium of RNA-binding proteins. RECOMB-RSG, Phoenix, November 2016.

DNA shape readout specificities of different TF families. ISMB, Orlando, July 2016

RCK: accurate and efficient inference of sequence and structure-based protein-RNA binding models from RNAcompete data. ISMB, Orlando, July 2016

Efficient design of compact unstructured RNA libraries covering all k-mers. WABI, Atlanta, September 2015

The use of HT-SELEX to infer TF binding models: comparison to PBM and an improved algorithm. ISMB, Boston, July 2014

Design of shortest double-stranded DNA sequences covering all k-mers with applications to protein binding microarrays and synthetic enhancers. ISMB, Berlin, Germany, July 2013

RAP: accurate and fast motif finding based on protein binding microarray data. RECOMB Regulatory Genomics, San Francisco, November 2012

Identifying TFBS motifs and predicting probe binding intensities from protein binding microarray data. RECOMB DREAM, New York, November 2010

### **Non-refereed conference talks**

Interpretability of deep neural networks in genomics. 2<sup>nd</sup> Summer School on Machine Learning in Bioinformatics, Virtual, August 2021

Convolutional neural networks for protein-DNA/RNA binding. 1<sup>st</sup> Summer School on Machine Learning in Bioinformatics, Virtual, August 2020

Deep learning for protein-RNA interactions. Rules of Protein-DNA Recognition: Computational and Experimental Advances, Oxaca, Mexico, June 2018

Utilizing de Bruijn graphs in universal sequence design for discovery of regulatory elements. Regulatory Genomics and Epigenomics workshop, Simons Institute, UC Berkeley, March 2016

Introduction to regulatory genomics and epigenomics: motif finding. Algorithmic Challenges in Genomics Boot Camp, Simons Institute, UC Berkeley, January 2016

Computational inference of binding site models from high-throughput SELEX data. Rules of Protein-DNA Recognition: Computational and Experimental Advances, Oxaca, Mexico, June 2015

Inferring binding site motifs from high-throughput in vitro data. Edmond J. Safra Bioinformatics Retreat, Kinneret, Israel, May 2014

Design of shortest double-stranded DNA sequence covering all k-mers. Workshop on Interdisciplinary Applications of Graph Theory, Combinatorics and Algorithms, Haifa, Israel, May 2013

Inferring binding site motifs from protein binding microarray data. Students Conference in Genetics, Genomics and Evolution, Beer-Sheva, Israel, September 2012

### **Conference poster presentations**

A deep learning approach for predicting protein-RNA binding. Seventh Broad – Israel Science Foundation (ISF) Symposium, Jerusalem, July 2019

A deep learning approach for predicting protein-RNA binding. RECOMB, Paris, April 2018

A deep learning approach for predicting protein-RNA binding. Frontiers in Genetics XI, Tel-Aviv, February 2018

Sequence libraries to cover all k-mers using degenerate characters. Systems Biology: Global Regulation of Gene Expression, Cold Spring Harbor, March 2017

Structure-based prediction of RNA-protein interactions on a transcriptome-wide scale. Systems Biology: Global Regulation of Gene Expression, Cold Spring Harbor, March 2017

Transcription factor family-specific DNA shape readout revealed by quantitative specificity models. Systems Biology: Global Regulation of Gene Expression, Cold Spring Harbor, March 2017

DNA shape readout specificities of different TF families. RECOMB-RSG, Phoenix, November 2016

Sequence libraries to cover all k-mers using degenerate characters. RECOMB-RSG, Phoenix, November 2016

Compact universal k-mer hitting sets. High-Throughput Sequencing Algorithms and Application Special Interest Group, Orlando, July 2016

Inferring binding site motifs from HT-SELEX data. Safra Retreat, Kinneret, Israel, May 2014

A comparative analysis of TF binding models learned from PBM, HT-SELEX and ChIP data. RECOMB/ISCB Regulatory and Systems Genomics, Toronto, Canada, November 2013

Inferring binding site motifs from high-throughput in vitro data. EMS Autumn School on Computational Aspects of Gene Regulation, Bedlewo, Poland, October 2013

Inferring binding site motifs from high-throughput technologies. Regulatory Genomics Special Interest Group, Berlin, Germany, July 2013

Inferring binding site motifs from high-throughput in vitro data. Israel Bioinformatics Symposium, Beer-Sheva, Israel, June 2013

Inferring binding site motifs from high-throughput technologies. DNA@60 conference, Haifa, Israel, May 2013

Inferring binding site motifs from high-throughput technologies. Safra Retreat, Kfar Giladi, Israel, May 2013

Accurate and user-friendly tools for discovering transcription factor microRNA binding sites. Frontiers in Genetics, Haifa, Israel, January 2013

RAP: accurate prediction of cis-regulatory motifs from protein binding microarrays. Safra Retreat, Ginosar, Israel, May 2012

RAP: accurate prediction of cis-regulatory motifs from protein binding microarrays. Systems Biology: Global Regulation of Gene Expression, Cold Spring Harbor, March 2012

Accurate and user-friendly tools for discovering transcription factor and microRNA binding sites. Frontiers in Genetics, Rehovot, Israel, February 2011

Amadeus-PBM: fast and accurate motif discovery from protein binding microarray data. RECOMB Regulatory Genomics, Barcelona, Spain, October 2011

Regression based classification of AML flow cytometry profiles. RECOMB DREAM, Barcelona, Spain, October 2011

Discovering transcription factor binding site motifs from protein binding microarrays. Safra Retreat, Kibutz Hagoshrim, Israel, May 2011

Accurate and user-friendly tools for discovering transcription factor and microRNA binding sites. ILANIT/FISEB conference, Eilat, Israel, February 2011

Identifying TFBS motifs and predicting probe binding intensities from protein binding microarray data. RECOMB DREAM, New York, November 2010

## **Professional Experience**

2017	Bioinformatics consultant, Augmanity Nano, Israel
2007-2010	Student position in IBM Haifa Research Labs, Israel
2006-2007	Student position in wide-band firm-ware group, Intel Labs, Israel
1999-2004	Worked in a technological unit of the Intelligence Corps, Israeli Defense Force