

# Curriculum Vitae

Yaron Orenstein

Ben-Gurion University of the Negev

## Education and Academic Experience

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

## Awards and Fellowships

- 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.

## 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.

## Journal Publications [\* equal contribution]

- 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.
- 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.
- 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.
- 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.**
- 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.
- Y. Orenstein** and R. Shamir (2016) Modeling protein-DNA binding via high throughput in vitro technologies. *Briefings in Functional Genomics*, elw030.
- 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.**
- 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.
- 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.
- A. Sloutskin, Y. Danino, **Y. Orenstein**, Y. Zehavi, T. Doniger, R. Shamir and T. Juven-Gershon (2015) ElementNT: A computational tool for detecting core promoter elements. *Transcription*, 6(3):41-50.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Y. Orenstein** and D. Ron (2011) Testing Eulerianity and connectivity in directed sparse graphs. *Theoretical Computer Science*, 412(45):6390-6408.

## Conference Proceedings

**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)*.

**Y. Orenstein**, D. Pellow, G. Marcais, 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)*.

**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)*.

## Teaching

- 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.

## Refereed Conference talks

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

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.

## **Review service**

RECOMB, ISMB, Bioinformatics, Brief. in Functional Genomics, Nucleic Acid Research, PLoS Comp. Bio.

**Interests** Bioinformatics, computational biology, graph algorithms, applied machine learning.

## **Conference poster presentations**

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.