BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Bar-Joseph, Ziv

eRA COMMONS USER NAME (credential, e.g., agency login): ZBARJO

POSITION TITLE: FORE Systems Professor of Computational Biology and Machine Learning

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Hebrew University, Jerusalem, Israel	B. Sc.	07/1997	Computer Science, Math
Hebrew University, Jerusalem, Israel	M. Sc.	08/1999	Computer Science
Massachusetts Institute of Technology	Ph. D.	06/2003	Computer Science

A. Personal Statement

My work focuses on the analysis, processing and modeling of large scale high throughput biological data and biological networks using machine learning techniques. We have specifically been focused on the analysis of bulk and single cell gene expression data and integrating that data with other types of biological data. We have developed methods for the analysis of RNA-Seq data, for integrating static interaction data with time series gene expression data and methods to obtain and improve condition and species specific data so that it can be used to model dynamic regulation. Our methods are implemented as open code software and are used by hundreds of groups word-wide to analyze large genomic datasets. All of our work is performed in close collaboration with experimental groups and several of the predictions made by our computational models have been experimentally validated.

a. Afek, N. Alon, O, Barad, E. Hornstein, N. Barkai, and Ziv Bar-Joseph. A biological solution to a fundamental distributed computing problem. *Science*, 331(6014):183-5, 2011.
b. G.E. Zinman, S. Naiman, Y. Kanfi, H. Cohen, and Z. Bar-Joseph. ExpressionBlast: Mining large, unstructured, expression databases. *Nature Methods*,10:925–926, 2013.
c. Kleyman*, E. Sefer*, T. Nicola, C. Espinoza, D. Chhabra, J.S. Hagood, N. Kaminski, N. Ambalavanan, and Z. Bar-Joseph. Selecting the most appropriate time points to profile in high-throughput studies. *eLife* pii: e18541, 2017
d. J. Ding, B. Aronow, N. Kaminski, J. Kitzmiller, J. Whitsett and Z. Bar-Joseph. Reconstructing differentiation networks and their regulation from time series single cell expression data. *Genome Res.* 28: 383-395, 2018

B. Positions and Honors

Positions and Employment

1999-2003 Research Assistant (Teaching Assistant, 2001), EECS, MIT, Cambridge, MA
2003 Postdoctoral Associate, Rick Young and David Gifford Lab, Whitehead Institute, MIT, Cambridge, MA
2003-2009 Assistant professor, Machine Learning Dept. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
2009-2015 Associate professor, Lane Center for Computational Biology, Machine Learning Dept., School of Computer Science, Carnegie Mellon University, Pittsburgh, PA

2015- Professor, Department of Computational Biology, Machine Learning Dept., School of Computer Science, Carnegie Mellon University, Pittsburgh, PA

2017- FORE Systems Professor of Computational Biology and Machine Learning, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA

Other Experience and Professional Memberships

2001- Member, International Society for Computational Biology

2009-2012 Editorial Board, Bioinformatics

2009, 2010 PC chair, Recomb Regulatory Genomics and Systems Biology

2013- 2017 Associate Editor, Bioinformatics

2013-2015 PC chair: Workshop on Biological Distributed Algorithms (BDA)

2014-2017 Stirring Committee Member: National Institutes of Allergy and Infectious Diseases (NIAID) Systems Biology Program

2014 - Standing member: NIH Modeling and Analysis of Biological Systems (MABS) study section

2015- Co-Director, Big Data for Better Health (BD4BH) in Pennsylvania

2017- Director, joint CMU-Pitt PhD Program in Computational Biology (CPCB)

Honors

1998- Best student paper award, Twenty-Second ACM Symposium on Principles of Distributed Computing (PODC)

2001-2003 Program in Mathematics and Molecular Biology (PMMB), national fellow

2003 Dimacs-Celera Genomics graduate student award in computational molecular biology

2005 National Science Foundation CAREER award

2012 Overton Prize, International Society for Computational Biology (ISCB)

2016 Best paper award, 20th Annual International Conference on Research in Computational Molecular Biology (RECOMB)

2017 FORE Systems Chair, School of Computer Science, Carnegie Mellon University

C. Contribution to Science

1. My early work focused on developing one of the first methods for integrating the (then very new) protein-DNA interaction data (primarily from ChIP-CHIP experiments) with gene expression data. As part of this work we have developed a number of methods and applied them to study the cell cycle regulation in human and yeast. This work identified a number of novel yeast regulators and have also led to the first published list of cycling genes in primary human cells. Some of the papers related to that work are:

a. Z. Bar-Joseph, G. K. Gerber, T. I. Lee, ..., T.S. Jaakkola, R.A. Young and D. K. Gifford. Computational Discovery of Gene Modules and Regulatory Networks. *Nature Biotechnology*. 21(10), pp 1337-1442, 2003.

b. Z. Bar-Joseph, S. Farkash, D. K. Gifford, I. Simon, R. Rosenfeld. Deconvolving Cell Cycle Expression Data with Complementary Information. *Bioinformatics (Proceedings of ISMB)*, 20 Suppl 1, pp. I23-I30, 2004.

c. I. Simon, Z. Siegfried, J. Ernst and Z. Bar-Joseph. Combined Static and Dynamic Analysis for Determining the Quality of Time-Series Expression Profiles. *Nature Biotechnology*, 23(12), pp 1503-1508, 2005.

d. Z. Bar-Joseph, Z. Siegfried, M. Brandeis, B. Brors, Y. Lu, R. Eils, B.D. Dynlacht and I. Simon. Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proceedings of the National Academy of Science (PNAS)*, 105(3), pp 956-961, 2008.

2. Following our early work on the analysis of cell cycle data we have realized that there was a lack of tools that are specifically focused on the analysis of high throughput time series biological data. Initial methods that analyzed such data treated it in a similar way to static data which meant that several of the features related to such data were not adequately utilized. We developed a set of methods and tools to analyze such data (both for supervised and unsupervised applications) and these have been widely used (some, such as STEM, have been downloaded more than 10000 times over the last decade and have been used in hundreds of publications by other groups to display their clustering results).

a. J. Ernst and Z. Bar-Joseph. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, 7:19, 2006.
widely used software, available at: http://www.sb.cs.cmu.edu/stem/
b. G.E. Zinman, S. Naiman, Y. Kanfi, H. Cohen, and Z. Bar-Joseph. ExpressionBlast: Mining large, unstructured, expression databases. *Nature Methods*,10:925–926, 2013.
widely used webserver at: http://www.expression.cs.cmu.edu/
c. E. Sefer, M. Kleyman and Z. Bar-Joseph. Tradeoffs between Dense and Replicate Sampling Strategies for High-Throughput Time Series Experiments. *Cell Systems*, 3(1):35-42, 2016.
(preliminary version in *Proceedings of the 20th Annual International Conference on Research in Computational Molecular Biology (RECOMB), 2016).*
d. J. Ding, J.S. Hagood, N. Ambalavanan, N. Kaminski, Z. Bar-Joseph. iDREM: Interactive visualization of dynamic regulatory networks*. PLoS Comput Biol.* 14(3):e1006019, 2018
widely used software, available at: http://www.sb.cs.cmu.edu/drem/

3. In addition to developing computational methods for the analysis of high throughput biological data we have initiated a new research direction which focuses on what we term 'bi-directional studies'. These studies focus on a biological process from the information processing point of view. The goal is to identify various biological systems that rely on such processing and then both, try to explain them using a computational model and at the same time use the insights provided by the model to improve the way we solve computational problems. We have applied this approach to several systems in multiple species, some of the outcomes are discussed n the following papers:

a. Afek, N. Alon, O, Barad, E. Hornstein, N. Barkai, and Ziv Bar-Joseph. A biological solution to a fundamental distributed computing problem. *Science*, 331(6014):183-5, 2011.
* Perspective discussing this paper appeared in Science. Selected as a highlight paper by the editors at Science Signaling and Cell.
We are maintaining a **repository for papers** on this subject at: www.algorithmsinnature.org/
b. S. Navlakha, X. He, C. Faloutsos, and Z. Bar-Joseph. Topological Properties of Robust Biological and Computational Networks. J. R. Soc. Interface, 11(96):20140283, 2014.
c. S. Navlakha, A. Barth, and Z. Bar-Joseph. Decreasing-Rate Pruning Optimizes the Construction of Efficient and Robust Distributed Networks. *PLoS Computational Biology*, 11(7):e1004347, 2015
d. S. Navlakha, Z. Bar-Joseph, A.L. Barth. Network Design and the Brain. *Trends Cogn Sci.*, 22(1):64-78, 2018

4. Our major focus over the last couple of years have been on developing and applying methods for reconstructing dynamic regulatory networks. Our methods extend HMMs to enables them to be used for integrating timer series and static data including mRNA and miRNA gene expression data (either from microarray or RNA-Seq experiments), epigenetic data and static protein-DNA interaction data. These methods were applied, by us and several other groups, to model various response on multiple species, to study development and differentiation and to model several other dynamic processes.

a. Z. Bar-Joseph, A. Gitter, I. Simon. Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, 13, 552-564, 2012
b. M. H. Schulz, K.V. Pandit, C.L. Lino Cardenas, N. Ambalavanan, N. Kaminski, and Z. Bar-Joseph. Reconstructing dynamic microRNA-regulated interaction networks. *Proceedings of the National Academy of Science (PNAS)*, 110(39):15686-91, 2013.
* Highlighted on the cover of PNAS and discussed in a published commentary in the same issue: U. Ohler, Using machine learning to identify disease-relevant regulatory RNAs, PNAS 110(39) 15516–15517, 2014.
c. A. Gitter, M. Carmi, N. Barkai, Z. Bar-Joseph. Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Research*, 23: 365-376, 2013
* Selected as one of the Top Ten Papers in the field of Regulatory and Systems Genomics for 2013 by the RECOMB/ISCB Conference on Regulatory and Systems Genomics.

d. Kleyman*, E. Sefer*, T. Nicola, C. Espinoza, D. Chhabra, J.S. Hagood, N. Kaminski, N. Ambalavanan, and Z. Bar-Joseph. Selecting the most appropriate time points to profile in high-throughput studies. *eLife* pii: e18541, 2017

5. We have further extended the methods mentioned above to enable them to jointly model dynamic regulatory and signaling networks in bulk *and single cell data*. These models use both regulatory information (expression, epigenetic and protein-DNA interactions) and post-transcriptional and post-translational data (protein interactions, post-transnational modification data) to infer pathways that are activated during development, disease progression and various responses.

a. L. Song, S.C. Huang, A. Wise, R. Castanon, J.R. Nery, H. Chen, M. Watanabe, J. Thomas, Z. Bar-Joseph and J.R. Ecker. A transcription factor hierarchy defines an environmental stress response network. *Science*, 354(6312). pii: aag1550, 2016.
b. C. Lin, S. Jain, H. Kim, Z. Bar-Joseph. Using neural networks for reducing the dimensions of single-cell RNA-Seq data *Nucleic Acids Research* , 45(17):e156, 2017.
c. S. Rashid, D. Kotton, and Z. Bar-Joseph. Tasic: Determining branching models from time series single cell data. *Bioinformatics*, 33(16):2504-2512, 2017
d. J. Ding, B. Aronow, N. Kaminski, J. Kitzmiller, J. Whitsett and Z. Bar-Joseph. Reconstructing differentiation networks and their regulation from time series single cell expression data. *Genome Res.* 28: 383-395, 2018

D. Ongoing Research Support

1R01GM122096 Bar-Joseph (PI) 09/01/2017-08/31/2021
Reconstructing regulatory networks from time series single cell data
Goals – Develop methods for modeling iPSC differentiation to lung cells using scRNA-Seq data

1R01HL128172 Kotton (PI) 07/01/2015-06/30/2019
Epigenenomic and transcriptomic networks in normal and defective lung development
Goals: Study epigenetic modifications in population of cells for a lung development model
Role: Investigator

CURE, PA Department of Health Bar-Joseph, Cooper (PIs) 06/01/2015-05/31/2019
Big Data for Better Health (BD4BH) in Pennsylvania
Goals: Developm machine learning methods for feature construction and classification of genomic and clinical cancer data.

1R01HL127349 Kaminski (PI) 06/01/2015-05/31/2019
Genomic Analysis of Tissue and Cellular Heterogeneity in IPF
Goals: Reconstruct dynamic regulatory networks for IPF patients using gene expression data.
Role: Investigator

NSF DBI- 1356505 Bar-Joseph (PI) 07/14- 07/18
ABI Innovation: BCSP: Understanding the design and usage of distributed biological networks
Goals: Develop machine learning methods to model and understand bacterial communication.

1U01 HL122626-01 Ambalavanan (PI) 04/01/14- 03/30/19.
Alveolar DevMAP
Goals: Model lung development in mice using time series transcription data from population of cells
Role: Investigator

James S. McDonnell Foundation Bar-Joseph (PI) 10/01/13 - 09/31/19
Scholars Award in Studying Complex Systems
Goals: Study bi-directional computational and biological systems to improve our understanding of brain development and the robustness and efficiently of networks.

Completed Research Support within the last 3 years

1. 1 U54 HL127624-01 National Institute of Health (NIH) “Multi-Task Learning of Cancer Response Pathways” to Ma’ayan(PI). Role: Investigator. 2014-2016
2. U01 HL108642-01 (NIH) “Gene networks beyond organ boundaries; heart, lung and pulmonary vascular disease” to Ahmad (PI). Role: Investigator, 2011 - 2015