ERAN SEGAL

Curriculum Vitæ– October 2003

Computer Science Department
Gates Building, Room 126
Stanford, CA 94305-9010
Phone: (650) 725 8784

eran@cs.stanford.edu
http://cs.stanford.edu/∼eran/
Home phone: (415) 282 4517
Cell phone: (415) 806 8178

EDUCATION

Ph.D. in Computer Science, Stanford University. **1999–2004 (expected)**
Advisor: Professor Daphne Koller.
Co-advisors: Professor Nir Friedman (Hebrew University) and Professor Stuart Kim (Stanford).
Reading committee (additional): Professor David Haussler (University of California Santa Cruz).

Ph.D. Minor in Genetics, Stanford University. **2000–2004 (expected)**

B.Sc. in Computer Science (*summa cum laude, 96/100*), Tel-Aviv University. **1995–1998**

RESEARCH INTERESTS

Computational Molecular Biology. Principled approaches based on probabilistic models to analyzing and integrating heterogeneous genomic data for the purpose of scientific discoveries.

Learning Probabilistic Graphical Models. Foundations and algorithms for representation, inference and learning in complex graphical models, with emphasis on exploiting problem-specific structure for learning in large-scale real-world domains.

AWARDS

Best Paper Award. Intelligent Systems for Molecular Biology (ISMB) 2003: "Genome-wide Discovery of Transcriptional Modules from Gene Expression and Sequence Data", E. Segal, R. Yelensky, and D. Koller.

Best Student Paper Award. Intelligent Systems for Molecular Biology (ISMB) 2003: "Discovering Molecular Pathways from Protein Interaction and Gene Expression Data", E. Segal, H. Wang, and D. Koller.

Runner-up for Best Student Paper Award. Uncertainty in Artificial Intelligence (UAI) 2003: "Learning Module Networks", E. Segal, D. Pe'er, A. Regev, D. Koller, and N. Friedman.

Runner-up for Best Presentation Award. Biomedical Computation at Stanford (BCATS) 2002: "Module Networks: Reconstruction of Molecular Modules and their Regulation from Gene Expression", E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, N. Friedman, and D. Koller.

Stanford Graduate Fellowship (SGF), 1999–2003.

Dean of the Faculty Award of Excellence. Faculty of exact sciences, Tel-Aviv University. Three separate awards: 1995–1996, 1996–1997, 1997–1998.

Dean of the Faculty Scholarship for B.Sc. Students.. Faculty of exact sciences, Tel-Aviv University. Three separate scholarships: 1995–1996, 1996–1997, 1997–1998.

PROFESSIONAL EXPERIENCE

Research Assistant, **Stanford University**. **1999–present**

- Research in computational molecular biology and probabilistic graphical models. Advisor: Professor Daphne Koller.

Research Assistant, **Tel-Aviv University**. **1998–1999**

- Research in database systems. Advisor: Professor Yossi Matias.
- Research in robotics. Advisor: Professor Yehezkel Yeshurun.

PROFESSIONAL ACTIVITIES

Journal reviewing: *Science*, *Bioinformatics*, *AIME: Journal of Artificial Intelligence in Medicine*, *EURASIP Journal on Applied Signal Processing*.

Program committee and conference reviewing: International Conference on Machine Learning (ICML): 2003, Neural Information Processing Systems (NIPS): 2001, 2002, Pacific Symposium on Biocomputing (PSB): 2003, 2004, Research in Computational Molecular Biology (RECOMB): 2002, 2003, Uncertainty in Artificial Intelligence (UAI): 2001, 2002, 2003.

EDITORIAL POSITIONS

Co-chair: Session on joint learning from multiple types of genomic data, Pacific Symposium on Biocomputing (PSB) 2004. Proposed the session to the PSB organizing committee, designed its scope, selected reviewers, and managed the reviewing process.

Co-chair: Second Biomedical Computation at Stanford Symposium (BCATS) 2002. Co-organized the symposium and managed the reviewing process.

STUDENTS ADVISED

Alexis Battle. Undergraduate.	**2002–2003**
Jonathan Efrat. Undergraduate.	**2003–present**
Todd Fojo. Undergraduate.	**2003–present**
Amit Kaushal. Undergraduate.	**2001–2002**
William Lu. Undergraduate.	**2003–present**
Richard May. Undergraduate.	**2002–2003**
Tuan Pham. Undergraduate.	**2001**
Romain Thibaux. Masters.	**2002–2003**
Mark Tong. Undergraduate.	**2003–present**
Daniel Wright. Undergraduate.	**2001–2002**
Roman Yelensky. Masters.	**2002–2003**

TEACHING EXPERIENCE

Teaching Assistant, Stanford University:

- **CS 221 – Artificial Intelligence: Principles and Techniques** (Fall 2000). Lectured, held regular office hours, assisted in designing assignments, exams and curriculum.
- **CS 228 – Probabilistic Models in Artificial Intelligence** (Winter 2001). Held regular office hours, assisted in designing assignments, exams and curriculum.
- **CS 161 – Design and Analysis of Algorithms** (Fall 2003). Lectured, held regular office hours, assisted in designing assignments and exams.

RESEARCH GRANT EXPERIENCE

Grant Assistant. Assisted Professor Koller in defining our group's research objectives for National Science Foundation (NSF) grant, July 2003.

Grant Assistant. Assisted Professor Koller on administration of National Science Foundation (NSF) grant ACI-0082554 under the Information Technology Research program. Wrote progress and final reports.

SOFTWARE DEVELOPED AND AVAILABLE ONLINE

GeneXPress, a tool for visualization and statistical analysis of genomic data, including gene expression, sequence data, and protein-protein interactions. Publically available for academic use at http://GeneXPress.stanford.edu. Currently, 383 scientists from 41 countries are using GeneXPress.

GeneSets, a web-based application for finding functional information about sets of genes, including enriched functional annotations, cis-regulatory motifs, and relevant microarrays. Publically available at http://GeneSets.stanford.edu

Regulatory Network Predictions from the *Nature Genetics* paper "Module Networks: Identifying Regulatory Modules and their Condition Specific Regulators from Gene Expression Data" are published as part of the primary database of yeast genetics, Saccharomyces Genome Database (SGD; http://www.yeastgenome.org).

INDUSTRY EXPERIENCE

Co-founder and Technical Leader, **SwapStation.com**. **1999**

- Lead the development of two functional sites with over 300,000 active users.
- Lead the development of a marketplace platform for business-to-business (B2B) exchanges, including development of a novel search and match engine.

Technical Leader, **Zapa Digital Arts Co.** **1997–1999**

- Devloped a small rich media Java player for the viewing of multimedia over the Internet.
- Developed a 3D character animation tool, and lead the integration of the tool with a Microsoft technology at Microsoft headquarters in Redmond, WA.

Technical Leader, **Magic Software Co.** **1996–1997**

- Participated in the development of a large scale application generator (MAGIC) with over one million users.
- Developed a generic database engine capable of communicating with various databases including: Oracle, DB2, MSSQL, Sybase, and Btrieve.
- Developed a client-server architecture for remote execution of applications.

Programmer, **Hi-Draw Software Co.** **1995–1996**

- Developed a generic drawing engine for handling complex structures and computations.
- Developed an AutoCAD plug-in application tailored for US engineers and detailers.

Sergeant, **Israeli Defense Forces** **1992–1995**

- Israeli intelligence forces. Devloped a data mining system for efficiently finding recurring statistical patterns in large databases (1993–1995); paratroopers unit (1992–1993).

PUBLICATIONS

PAPERS IN PEER-REVIEWED JOURNALS

[1] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D.Koller, N. Friedman. "Module Networks: Identifying Regulatory Modules and their Condition Specific Regulators from Gene Expression Data". *Nature Genetics* 34(2): 166-76, June 2003.

[2] J. Stuart*, E. Segal*, D. Koller, S. Kim (* – co-first authors). "A Gene Co-expression Network for Global Discovery of Genetic Modules". *Science*, August 2003.

[3] E. Segal, R. Yelensky, D. Koller. "Genome-wide Discovery of Transcriptional Modules from DNA Sequence and Gene Expression". *Bioinformatics* 19(Suppl 1): 1273–82, July 2003 (via ISMB).

[4] E. Segal, H. Wang, D. Koller. "Discovering Molecular Pathways from Protein Interaction and Gene Expression Data". *Bioinformatics* 19(Suppl 1): 1264–72, July 2003 (via ISMB).

[5] E. Segal, B. Taskar, A. Gasch, N. Friedman, D. Koller. "Rich Probabilistic Models for Gene Expression". *Bioinformatics* 17(Suppl 1): S243–52, July 2001 (via ISMB).

PAPERS IN PEER-REVIEWED CONFERENCES

[6] E. Segal, D. Pe'er, A. Regev, N. Friedman, D. Koller. "Learning Module Networks". *Nineteeth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2003.

[7] E. Segal, A. Battle, D. Koller. "Decomposing Gene Expression into Cellular Processes". *Eighth Pacific Symposium on Biocomputing (PSB)*, 2003.

[8] S. Rosset, E. Segal. "Boosting Density Estimation". *Sixteenth International Conference on Neural Information Processing Systems (NIPS)*, 2002.

[9] E. Segal, Y. Barash, I. Simon, N. Friedman, D. Koller. "From Promoter Sequence to Expression: A Probabilistic Framework". *Sixth International Conference on Research in Computational Molecular Biology (RECOMB)*, 2002.

[10] E. Segal D. Koller. "Probabilistic Hierarchical Clustering for Biological Data". *Sixth International Conference on Research in Computational Molecular Biology (RECOMB)*, 2002.

[11] E. Segal, D. Koller, D. Ormoneit. "Probabilistic Abstraction Hierarchies". *Fifteenth International Conference on Neural Information Processing Systems (NIPS)*, 2001.

[12] B. Taskar, E. Segal, D. Koller. "Probabilistic Classification and Clustering in Relational Data". *Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2001.

[13] U. Lerner, E. Segal, D. Koller. "Exact Inference in Networks with Discrete Children of Continuous Parents". *Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2001.

[14] Y. Matias, E. Segal, J. Vitter. "Efficient Bundle Sorting". *In Eleventh Annual SIAM/ACM Symposium on Discrete Algorithms (SODA)*, 2000.

PATENTS

[15] Y. Matias, E. Segal. Method of sorting data sets with a predetermined number of distinct keys, U.S. Patent 6,519,593.

[16] E. Segal, within a group. Method for rendering an animation object, including a plurality of sub-objects, on a display of a client computer under control of a remote sever, U.S. Patent 6,351,267.

SELECT UNREVIEWED PAPERS

[17] E. Segal S. Kim. "The Modular Era of Functional Genomics". *Genome Biology* 4(5):317, April 2003.

[18] L. Getoor, E. Segal, B. Taskar, D. Koller. "Probabilistic Models of Text and Link Structure for Hypertext Classification". *In IJCAI Workshop on Text Learning*, 2001.

[19] E. Segal, Y. Matias. "Approximate Iceberg Queries". *TR*, Tel-Aviv University, 1999.

PAPERS IN PREPARATION

[20] E. Segal, D. Pe'er, A. Regev, D. Koller, N. Friedman. "Learning Module Networks". Submitted to *Journal of Machine Learning Research* (August 2003).

[21] E. Segal, R. Sharan. "A Discriminative Model for Identifying Spatial Cis-Regulatory Modules". Submitted to *RECOMB* (September 2003).

[22] A. Battle, E. Segal, D. Koller. "Probabilistic Discovery of Overlapping Cellular Processes and Their Regulation Using Gene Expression Data". Submitted to *RECOMB* (September 2003).

[23] E. Segal, D. Koller. "Association Webs: A Foundation for Networks of Genomic Interactions". *In preparation*.

[24] E. Segal, N. Friedman, A. Regev, D. Koller. "The Cancer Module Map: Combinatorial Organization of Cancer Revealed by the Unification of Genomic Data". *In preparation*.

[25] E. Segal, R. Yelensky, N. Friedman, D. Koller. "Identifying Modules of Cis-Regulatory Motifs in Human". *In preparation*.

SEMINARS AND INVITED TALKS

COMPUTER SCIENCE AND BIOINFORMATICS

[1] "Global Discovery of Conserved Genetic Modules using Gene Expression". *In Fifth Biopathways Consortium meeting*, Brisbane, Australia, June 2003.

[2] "Module Networks: Reconstruction of Molecular Modules and their Regulation from Gene Expression", *In Third Biomedical Computation at Stanford*, Stanford, California, October 2002.

[3] "Finding Regulatory Modules from Gene Expression Profiles and Sequence Data". *In Fourth Biopathways Consortium meeting*, Edmonton, Canada, July 2002.

[4] "From Promoter Sequence to Expression: A Probabilistic Framework", *In the Learning Workshop at Snowbird*, Snowbird, Utah, April 2002.

[5] "From Promoter Sequence to Expression: A Probabilistic Framework", *In Bioinformatics Workshop at NIPS*, Vancouver, British Columbia, Canada, December 2001.

[6] "Rich Probabilistic Models for Gene Expression", *In University of Colorado Health Sciences Center Seminar*, Denver, Colorado, October 2001.

[7] "Two Sided Clustering for Yeast Gene Expression Using Probabilistic Relational Models", *In First Biomedical Computation at Stanford*, Stanford, California, October 2000.

BIOLOGY

[8] "Module Networks: Identifying Regulatory Modules and their Condition Specific Regulators from Gene Expression Data". In *Systems Biology: Genomic Approaches to Transcriptional Regulation* Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, March 2003.

[9] "Module Networks: Identifying Regulatory Modules and their Condition Specific Regulators from Gene Expression Data." In *Functional Genomics: Global Analysis of Complex Biological Systems* Keystone conference, Santa Fe, New Mexico, February 2003.

[10] "Learning module networks from expression profiles." In *2002 Yeast Genetics and Molecular Biology Meeting*, Madison, Wisconsin, July 2002.

REFERENCES

Professor Douglas Brutlag, research collaborator
Department of Biochemistry
Beckman Center, B40, Mail Code 5307
Stanford, CA 94305-5307
brutlag@stanford.edu, (650) 723 6593

Professor Nir Friedman, co-advisor
School of Computer Science and Engineering
Hebrew University
Ross Building, Room 203
Givat Ram, Jerusalem 91904, Israel
nir@cs.huji.ac.il, +972 2 658 4720

Professor David Haussler, thesis committee member
Center for Biomolecular Science and Engineering
321 Baskin Engineering Building
University of California Santa Cruz, CA, 95064
haussler@cse.ucsc.edu (831) 459 2105

Professor Stuart Kim, co-advisor
Department of Developmental Biology
Beckman Center, B300
Stanford, CA 94305-5329
kim@cmgm.stanford.edu, (650) 725 7671

Professor Daphne Koller, principal advisor
Computer Science Department
Gates Building 142
Stanford, CA 94305-9010
koller@cs.stanford.edu, (650) 723 6598

Professor Matthew Scott, research collaborator
Department of Developmental Biology
Clark Center W252
Stanford, CA 94305-5439
scott@cmgm.stanford.edu, (650) 725 7656