

# Alexander Statnikov

**E-mail:** [Alexander.Statnikov@med.nyu.edu](mailto:Alexander.Statnikov@med.nyu.edu)

**Website:** <http://www.statnikov.org>

**Office:** 1 (212) 263-3641

**Fax:** 1 (212) 263-5995

**Cell:** 1 (615) 545-3685

**Work address:**

333 E38th Street, 6th Floor,

Center for Health Informatics and Bioinformatics,

New York University Langone Medical Center,

New York, NY 10016, USA

---

## Current Academic Affiliation and Positions

- **Assistant Professor**, *Center for Health Informatics and Bioinformatics, Department of Medicine, Division of Clinical Pharmacology*, July 2009 - Present
- **Director**, *Computational Causal Discovery Laboratory*,
- **Benchmarking Director**, *Best Practices Integrative Informatics Consultation Service (BPIC)*

New York University Langone Medical Center, New York, New York.

## Research Interests

- Computational causal discovery in high-dimensional biomedical data;
- Translational bioinformatics and development, optimization, and validation of molecular signatures;
- Machine learning applications in biomedicine.

## Education

**Ph.D.**, Biomedical Informatics (GPA 3.9), *Vanderbilt University*, Nashville, Tennessee, December 2008

**Master of Science**, Biomedical Informatics (GPA 3.9), *Vanderbilt University*, Nashville, Tennessee, August 2005

**Master of Science**, Applied Mathematics (GPA 3.9), *Case Western Reserve University*, Cleveland, Ohio, August 2002

**Bachelor of Science**, Mathematics (GPA 3.6), *Case Western Reserve University*, Cleveland, Ohio, August 2001

Applied Mathematics (GPA 3.9), *Moscow State Institute of Electronics and Mathematics*, Moscow, Russia, September 1996 - July 1999

## Professional Work Experience

**Senior Scientific Programmer**, *Discovery Systems Laboratory, Department of Biomedical Informatics, Vanderbilt University Medical Center*, Nashville, Tennessee, May 2002 - May 2009

**Scientific Programmer**, *Zyxbio, LLC*, Cleveland, Ohio, February 2002 - May 2002

**Research Assistant and Programmer**, *Department of Mathematics, Case Western Reserve University*, Cleveland, Ohio.

May 2000 -  
May 2002

**Research Assistant and Programmer**, *Mechanical Engineering Research Institute, Russian Academy of Sciences*, Moscow, Russia.

February 1997 -  
August 1999

**Programmer and System Administrator**, *NPO Stroyservice Ltd.*, Moscow, Russia.

December 1998 -  
August 1999

## Consulting

**Prediction Sciences, LLC** (La Jolla, CA): Proteomic-based stroke diagnosis biomarker discovery and diagnostics signatures development.

**Sanctis, LLC** (Leesburg, VA): Molecular signatures development.

**Discovery Holdings, LLC** (Nashville, TN): Machine learning algorithms development.

## Publications

### Peer-Reviewed Articles\* :

1. Aliferis CF, **Statnikov A**, Tsamardinos I, Schildcrout JS, Shepherd BE, Harrell FE. Factors Influencing the Statistical Power of Complex Data Analysis Protocols for Molecular Signature Development from Microarray Data. *PLoS ONE*, 2009; 4(3): e4922.
2. Fananapazir N, **Statnikov A**, Aliferis CF. The FAST-AIMS Clinical Mass Spectrometry Analysis System. *Advances in Bioinformatics*, vol. 2009, Article ID 598241, 2009.
3. Aliferis CF, **Statnikov A**, Tsamardinos I, Mani S, Koutsoukos X. Local Causal and Markov Blanket Induction Algorithms for Causal Discovery and Feature Selection for Classification. Part I: Algorithms and Empirical Evaluation. Accepted to the *Journal of Machine Learning Research*, 2009.
4. Aliferis CF, **Statnikov A**, Tsamardinos I, Mani S, Koutsoukos X. Local Causal and Markov Blanket Induction Algorithms for Causal Discovery and Feature Selection for Classification. Part II: Analysis and Extensions. Accepted to the *Journal of Machine Learning Research*, 2009.
5. **Statnikov A**, Aliferis CF. TIED: An Artificially Simulated Dataset with Multiple Markov Boundaries. Accepted to the *Journal of Machine Learning Research Workshop and Conference Proceedings, Workshop on Causality (NIPS 2008)*, 2009.
6. Mani S, Aliferis CF, **Statnikov A**. Bayesian Algorithms for Causal Data Mining. Accepted to the *Journal of Machine Learning Research Workshop and Conference Proceedings, Workshop on Causality (NIPS 2008)*, 2009.
7. **Statnikov A**, Wang L, Aliferis CF. A Comprehensive Comparison of Random Forests and Support Vector Machines for Microarray-Based Cancer Classification. *BMC Bioinformatics*, 2008; 9:319.
8. **Statnikov A**, Li C, Aliferis CF. A Statistical Reappraisal of the Findings of an Esophageal Cancer Genome-Wide Association Study. *Cancer Research*, 2008; 68: 3074-3075.

---

\* Following the practice in biomedical informatics and computer science, this section includes full-length peer-reviewed articles published in journals and in major, rigorously peer-reviewed biomedical informatics and computer science conference proceedings (the acceptance rates of most computer science conferences in the present vitae range from less than 25% to less than 5%).

9. Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P, **Statnikov A**. Design and Analysis of the Causation and Prediction Challenge. *Journal of Machine Learning Research Workshop and Conference Proceedings, Volume 3: Causation and Prediction Challenge (WCCI 2008)*: 1-33, 2008.
10. **Statnikov A**, Li C, Aliferis CF. Effects of Environment, Genetics and Data Analysis Pitfalls in an Esophageal Cancer Genome-Wide Association Study. *PLoS ONE*, 2007; 2(9): e958.
11. **Statnikov A**, Aliferis CF. Are Random Forests Better than Support Vector Machines for Microarray-Based Cancer Classification? *AMIA Annual Symposium*, 2007.
12. Aliferis CF, **Statnikov A**, Tsamardinos I. Challenges in the Analysis of Mass-Throughput Data: A Technical Commentary from the Perspective of Statistical Machine Learning. *Cancer Informatics*. 2006; 2: 133-162.
13. Aphinyanaphongs Y, **Statnikov A**, Aliferis CF. A Comparison of Citation Metrics to Machine Learning Filters for the Identification of High Quality MEDLINE Documents. *Journal of the American Medical Informatics Association*. 2006 Jul-Aug; 13: 446-455.
14. **Statnikov A**, Hardin D, Aliferis CF. Using SVM Weight-Based Methods to Identify Causally Relevant and Non-Causally Relevant Variables. *Neural Information Processing Systems (NIPS) 2006 Workshop on Causality and Feature Selection*, 2006.
15. Tsamardinos I, **Statnikov A**, Brown LE, Aliferis CF. Generating Realistic Large Bayesian Networks by Tiling. *19<sup>th</sup> International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2006.
16. Levy S, **Statnikov A**, Aliferis CF. Biomarker Selection from High-Dimensionality Data. *Pharmaceutical Discovery*. 2005 Microarray Supplement, 2005 Sep; 37-44.
17. **Statnikov A**, Tsamardinos I, Dosbayev Y, Aliferis CF. GEMS: A System for Automated Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data. *International Journal of Medical Informatics*. 2005 Aug; 74(7-8): 493-501.
18. **Statnikov A**, Aliferis CF, Tsamardinos I, Hardin D, Levy S. A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis. *Bioinformatics*. 2005 Mar; 21(5): 631-43.
19. Aphinyanaphongs Y, Tsamardinos I, **Statnikov A**, Hardin D, Aliferis CF. Text Categorization Models for High-Quality Article Retrieval in Internal Medicine. *Journal of the American Medical Informatics Association*. 2005 Mar-Apr; 12(2): 207-16.
20. Duda S, Aliferis CF, Miller R, **Statnikov A**, Johnson K. Extracting Drug-Drug Interaction Articles from MEDLINE to Improve the Content of Drug Databases. *AMIA Annual Symposium*, 2005.
21. **Statnikov A**, Aliferis CF, Tsamardinos I. Methods for Multi-Category Cancer Diagnosis from Gene Expression Data: A Comprehensive Evaluation to Inform Decision Support System Development. *Studies in Health Technology and Informatics*. 2004; 107(Pt 2): 813-817.
22. Aliferis CF, Tsamardinos I, **Statnikov A**. HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection. *AMIA Annual Symposium*, 2003.
23. Tsamardinos I, Aliferis CF, **Statnikov A**. Time and Sample Efficient Discovery of Markov Blankets and Direct Causal Relations. *9<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.
24. Aliferis CF, Tsamardinos I, Massion P, **Statnikov A**, Fananapazir N, Hardin D. Machine Learning Models for Classification of Lung Cancer and Selection of Genomic Markers Using Array Gene Expression Data. *16<sup>th</sup> International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2003.
25. Tsamardinos I, Aliferis CF, **Statnikov A**. Algorithms for Large Scale Markov Blanket Discovery. *16<sup>th</sup> International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2003.

26. Frey L, Fisher D, Tsamardinos I, Aliferis CF, **Statnikov A**. Identifying Markov Blankets with Decision Tree Induction. *Third IEEE International Conference on Data Mining (ICDM)*, 2003.
27. Aliferis CF, Tsamardinos I, **Statnikov A**, Brown LE. Causal Explorer: A Probabilistic Network Learning Toolkit for Biomedical Discovery. *International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS)*, 2003.
28. Aliferis CF, Tsamardinos I, Massion P, **Statnikov A**, Hardin D. Why Classification Models Using Array Gene Expression Data Perform So Well: A Preliminary Investigation of Explanatory Factors. *International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS)*, 2003.

#### Books:

29. Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P, **Statnikov A** (editors). Challenges in Causality. Volume 1: Causation and Prediction Challenge. (In press) Brookline, Massachusetts: *Microtome Publishing*, 2009.

#### Book Chapters:

30. Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P, **Statnikov A**. Causality Workbench. In *Causality in the Sciences*. Edited by Illari PM, Russo F and Williamson J. (In press) *Oxford University Press*, 2010.
31. **Statnikov A**, Tsamardinos I, Brown LE, Aliferis CF. Causal Explorer: A Matlab Library of Algorithms for Causal Discovery and Variable Selection for Classification. In *Challenges in Causality. Volume 1: Causation and Prediction Challenge*. Edited by Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P and Statnikov A. (In press) Brookline, Massachusetts: *Microtome Publishing*, 2009.

#### Peer-Reviewed Abstracts and Software Presentations:

32. Zimmerman LJ, Coleman JA, Hardin DP, **Statnikov A**, Aliferis CF, Liebler DC. Development of Metrics for Assessment of Plasma Quality. *57th ASMS Conference on Mass Spectrometry and Allied Topics*, 2009.
33. Kokkotou E, Lois A, Triggs C, Conboy L, McDougall L, **Statnikov A**, Aliferis CF, Pothoulakis C, Kaptchuk T, Lembo A. Serum Biomarker Analysis of Placebo Responses in Patients with Irritable Bowel Syndrome. *Neurogastroenterology and Motility Joint International Meeting*, 2008.
34. Aliferis CF, **Statnikov A**, Tsamardinos I, Kokkotou E, Massion PP. Application and Comparative Evaluation of Causal and Non-Causal Feature Selection Algorithms for Biomarker Discovery in High-Throughput Biomedical Datasets. *Neural Information Processing Systems (NIPS) 2006 Workshop on Causality and Feature Selection*, 2006.
35. Aliferis CF, **Statnikov A**, Massion PP. Pathway Induction and High-Fidelity Simulation for Molecular Signature and Biomarker Discovery in Lung Cancer Using Microarray Gene Expression Data. *APS Conference: Physiological Genomics and Proteomics of Lung Disease*, 2006.
36. **Statnikov A**, Tsamardinos I, Aliferis CF. Using GEMS for Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data. *13<sup>th</sup> Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 2005.

37. **Statnikov A**, Tsamardinos I, Aliferis CF. Using the GEMS System for Supervised Analysis of Cancer Microarray Gene Expression Data. *AMIA Annual Symposium*, 2005.
38. **Statnikov A**, Tsamardinos I, Aliferis CF. Using the GEMS System for Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data. *12<sup>th</sup> National Conference on Artificial Intelligence (AAAI)*, 2005.
39. **Statnikov A**, Tsamardinos I, Aliferis CF. Using GEMS for Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data. *13<sup>th</sup> Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 2005.

#### Technical Reports:

40. **Statnikov A**, Kasparova E, Aliferis CF. Applying Decision Support Models in the Presence of Incomplete Evidence. Technical Report DSL TR-06-02, *Department of Biomedical Informatics, Vanderbilt University*, 2006.
41. **Statnikov A**, Tsamardinos I, Aliferis CF. New Efficient and Correct Algorithms for Identification of Direct Causal Relationships and Markov Blankets from Data. Technical Report DSL TR-06-01, *Department of Biomedical Informatics, Vanderbilt University*, 2006.
42. Tsamardinos I, Aliferis CF, **Statnikov A**, Brown LE. Scaling-Up Bayesian Network Learning to Thousands of Variables Using Local Learning Technique. Technical Report DSL TR-03-02, *Department of Biomedical Informatics, Vanderbilt University*, 2003.
43. **Statnikov A**, Tsamardinos I, Aliferis CF. An Algorithm for Generation of Large Bayesian Networks. Technical Report DSL TR-03-01, *Department of Biomedical Informatics, Vanderbilt University*, 2003.
44. Aliferis CF, Tsamardinos I, **Statnikov A**. Large-Scale Feature Selection Using Markov Blanket Induction for the Prediction of Protein-Drug Binding. Technical Report DSL TR-02-06, *Department of Biomedical Informatics, Vanderbilt University*, 2002.

#### Theses:

45. **Statnikov A**. Algorithms for Discovery of Multiple Markov Boundaries: Application to the Molecular Signature Multiplicity Problem. Ph.D. Thesis. *Department of Biomedical Informatics, Vanderbilt University*, Advisor: Dr. Constantin F. Aliferis, Committee Members: Dr. Gregory F. Cooper, Dr. Douglas P. Hardin, Dr. Daniel R. Masys, Dr. Ioannis Tsamardinos, December 2008.
46. **Statnikov A**. Automatic Cancer Diagnostic Decision Support System for Gene Expression Domain. Master's Thesis. *Department of Biomedical Informatics, Vanderbilt University*, Advisors: Dr. Constantin F. Aliferis and Dr. Ioannis Tsamardinos, August 2005.
47. **Statnikov A**. Numerical Methods for Image Reconstruction for the Calibration of the NASA-Glenn Icing Research Wind Tunnel: A Computer-Based Approach. Master's Thesis. *Department of Mathematics, Case Western Reserve University*, Advisor: Dr. Steven H. Izen, August 2002.

#### Patents:

48. **Statnikov A**, Aliferis CF. A Computer Implemented Method for Determining All Markov Boundaries and Its Application for Discovering Multiple Maximally Accurate and Non-Redundant Predictive Models. *United States Patent Application*, 2009.
49. **Statnikov A**, Aliferis CF. Methods for Discovery of Markov Boundaries from Datasets with Hidden Variables. *United States Provisional Patent Application*, 2009.

50. **Statnikov A**, Aliferis CF, Tsamardinos I, Fananapazir N. Method and System for Automated Supervised Data Analysis. *United States Patent Application #20070122347*, 2006.

#### Papers in Preparation or under Review:

- **Statnikov A**, Aliferis CF. Analysis and Computational Dissection of Molecular Signature Multiplicity. Submitted, 2009.
- Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P, **Statnikov A**. Development Projects for the Causality Workbench. Submitted, 2009.
- Mani S, Cooper GF, **Statnikov A**. Discovering Causal Relationships in the Presence of Hidden Variables. In Preparation, 2009.
- **Statnikov A**, Gebre-Amlak K, Aliferis CF. On Improving Computational Efficiency of Local Causal Discovery and Markov Boundary Induction Algorithms. In Preparation, 2009.

(The list of publications in the field of *Multicriteria Analysis* is available upon request)

### Grants

#### Active:

3RO1 AR056667201S1 (Cronstein) NIH “The Pharmacology of Dermal Fibrosis” Role: Co-Investigator	1/15/09-12/31/13 \$220,000	1.20 calendar
1 I01 CX000242-01 (Massion) Department of Veterans Affairs “Molecular Biomarkers of Small Cell Lung Cancer Behaviour” Role: Co-Investigator	10/1/09 – 9/30/13 \$188,000	2.4 calendar
1 U54 RR024386-01A2 (Cronstein) NIH/NCRR “NYU-HHC Clinical Translation Science Award” Role: Bioinformatics Service Manager	07/14/09 – 03/31/14 \$3,860,305	3.6 calendar

#### Pending:

ECCS 1001171 (Guyon) NSF “Challenges in Active Learning and Experimental Design” Role: Co-PI	01/01/10 – 12/31/11 \$10,651 (subcontract)	0.6 calendar
---	---	--------------

### Honors and Awards

- ISMB 2005 Best Poster Award (June 2005)
- Medinfo 2004 Student Paper Competition, Gold Medal (September 2004)
- Vanderbilt University Medical School Dean’s Scholarship (September 2004 - December 2008)
- American Mathematical Society Waldemar J. Trjitzinsky National Award (May 2000)

- Case Western Reserve University Dean's Honors List (2000)

### **Professional and Scientific Societies**

- AAAI (American Association for Artificial Intelligence): 2005 - Present
- AAAS (American Association for the Advancement of Science): 2007 - Present
- ACM (Association for Computing Machinery): 2008 - Present
- AMIA (American Medical Informatics Association): 2003 - Present
- AMS (American Mathematical Society): 2001 - Present
- ASA (American Statistical Association): 2008 - Present
- ISCB (International Society for Computational Biology): 2004 - Present

### **Scientific Reviewer Activities**

#### **Journals:**

- Bioinformatics
- BMC Bioinformatics
- British Journal of Cancer
- Computational Statistics and Data Analysis
- Decision Support Systems
- FEBS Letters
- IEEE Transactions on Knowledge and Data Engineering
- Journal of the American Medical Informatics Association
- Journal of Artificial Intelligence Research
- Journal of Biomedical Informatics
- Journal of Clinical and Laboratory Medicine
- Journal of Machine Learning Research
- Methods of Information in Medicine
- Nucleic Acids Research
- Pattern Recognition Letters
- PLoS Genetics

#### **Conferences:**

- American Medical Informatics Association (AMIA) 2009 Symposium
- 15<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009): Reviewer of KDD cup proposals
- Pacific Symposium on Biocomputing (PSB) 2009
- 22<sup>nd</sup> Annual Conference on Neural Information Processing Systems (NIPS 2008): Causality Workshop
- American Medical Informatics Association (AMIA) 2008 Symposium
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2008)
- IEEE World Congress on Computational Intelligence (WCCI 2008)
- American Medical Informatics Association (AMIA) 2007 Symposium
- 3<sup>rd</sup> Indian International Conference on Artificial Intelligence (IICAI 2007)
- American Medical Informatics Association (AMIA) 2006 Symposium
- 10<sup>th</sup> Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006): Workshop on Data Mining for Biomedical Applications (BioDM 2006)
- 11<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005)

#### **Books:**

- “*Computational Intelligence in Bioinformatics*” (Editor in Chief: Arpad Kelemen)

## Scientific Community Service

- Co-Organizer of the “*Causality Challenge #2: Potluck*”: August 2008 - December 2008
- Co-Organizer of the “*Causality Challenge #1: Causation and Prediction*” (jointly with Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, and Spirtes P): March 2007 - September 2008
- NIPS 2008 (Neural Information Processing Systems) Causality Workshop Committee Member
- AMIA (American Medical Informatics Association) Genomics Workgroup Student Representative: April 2005 - November 2007

## Teaching Experience

### Courses:

- Co-instructor of the “*Introduction of Translational and Clinical Bioinformatics*” section in the Pathology Resident Training, Department of Pathology, New York University School of Medicine, New York, New York: September - November 2009
- Co-instructor of the “*Biomedical Artificial Intelligence*” course, Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee: January - February 2009
- Taught a lecture “*Applications of Support Vector Machines in Clinical Bioinformatics*” in the “*Biomedical Artificial Intelligence*” course, Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee: March 2008
- Instructor of undergraduate Mathematics courses “*Elementary Functions and Analytic Geometry*” and “*Calculus I*”, Department of Mathematics, Case Western Reserve University, Cleveland, Ohio: August 2001 - May 2002
- Teaching assistant, Department of Mathematics, Case Western Reserve University, Cleveland, Ohio: January 2000 - May 2001

### Tutorials:

- Tutorial “*A Gentle Introduction to Support Vector Machines in Biomedicine*” at the *AMIA Annual Symposium*, 2009
- Tutorial “*Support Vector Machines without Tears*” at the *AMIA Annual Symposium*, 2008

## Other Experience

- **Webmaster**, *Department of Mathematics, Case Western Reserve University*, Cleveland, Ohio: January 2000 - May 2002
- **Webmaster**, *Department of Mechanical Engineering, Cleveland State University*, Cleveland, Ohio: December 2000 - May 2001

## Personal

- Born in Moscow, Russia, 1979
- United States citizen
- Fluent in English and Russian