

Alexander Statnikov

E-mail: alexander.statnikov@vanderbilt.edu

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

Office: 1 (615) 936-3388

Fax: 1 (615) 936-1427

Cell: 1 (615) 545-3685

Work address:

Eskind Biomedical Library, Room 413,
2209 Garland Avenue,
Vanderbilt University Medical Center,
Nashville, TN 37232-8340, USA

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), <i>Vanderbilt University</i> , Nashville, Tennessee.	December 2008
Master of Science , Biomedical Informatics (GPA 3.9), <i>Vanderbilt University</i> , Nashville, Tennessee.	August 2005
Master of Science , Applied Mathematics (GPA 3.9), <i>Case Western Reserve University</i> , Cleveland, Ohio.	August 2002
Bachelor of Science , Mathematics (GPA 3.6), <i>Case Western Reserve University</i> , Cleveland, Ohio.	August 2001
Applied Mathematics (GPA 3.9), <i>Moscow State Institute of Electronics and Mathematics</i> , Moscow, Russia.	September 1996 - July 1999

Work Experience

Senior Scientific Programmer

Discovery Systems Laboratory, Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee: May 2002 - Present

- Programming complex machine learning and statistical algorithms and optimizing existing implementations;
- Conducting large-scale feature selection, causal discovery, and classification experiments;
- Participating in theoretical design of new algorithms;
- Writing research papers and technical reports;
- Developing software applications and libraries of algorithms for feature selection, causal discovery and classification in biomedical domains;
- Organizing and developing computational infrastructure of the laboratory.

Scientific Programmer

Zyxbio, LLC, Cleveland, Ohio: February 2002 - May 2002

- Developed computer models for prediction of drug absorption using neural networks;
- Consulted major pharmaceutical firms on integration and software design issues of the company's products.

Research Assistant and Programmer

Department of Mathematics, Case Western Reserve University, Cleveland, Ohio: May 2000 - May 2002

- Developed a software system for image reconstruction;

- Programmed numerical mathematics algorithms;
- Participated in NASA Glenn Research Center Wind Tunnel project;
- Created graphics user interface applications for X-Windows.

Research Assistant and Programmer

Mechanical Engineering Research Institute, Russian Academy of Sciences, Moscow, Russia: February 1997 - August 1999

- Developed interface programs for multicriteria optimization;
- Co-authored a comprehensive software system MOVI (Multicriteria Optimization and Vector Identification) for multicriteria analysis.

Programmer and System Administrator

NPO Stroyservice Ltd., Moscow, Russia: December 1998 - August 1999

- Developed accounting and inventory software system;
- Created and maintained the company's website;
- Participated in business negotiations in Taiwan (May 1999).

Consulting

Prediction Sciences, La Jolla, California: 2004

- Developed computational diagnostic models of cerebrovascular disorders based on molecular and clinical data.

Publications

Journal Papers:

- **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.
- 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*, 2008.
- 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*, 2008.
- **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.
- **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.
- 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.
- 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.
- Levy S, **Statnikov A**, Aliferis CF. Biomarker Selection from High-Dimensionality Data. *Pharmaceutical Discovery*. 2005 Microarray Supplement, 2005 Sep; 37-44.

- **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.
- **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.
- 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.

Papers in Conference Proceedings:

- Guyon I, Aliferis CF, Cooper GF, Elisseeff A, Pellet JP, Spirtes P, **Statnikov A**. Design and Analysis of the Causation and Prediction Challenge. Accepted to the *Journal of Machine Learning Research Proceeding of the IEEE World Congress on Computational Intelligence (WCCI)*, 2008.
- **Statnikov A**, Aliferis CF. Are Random Forests Better than Support Vector Machines for Microarray-Based Cancer Classification? *AMIA Annual Symposium*, 2007.
- **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.
- Tsamardinos I, **Statnikov A**, Brown LE, Aliferis CF. Generating Realistic Large Bayesian Networks by Tiling. *19th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2006.
- 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.
- **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. *Medinfo*, 2004.
- Aliferis CF, Tsamardinos I, **Statnikov A**. HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection. *AMIA Annual Symposium*, 2003.
- Tsamardinos I, Aliferis CF, **Statnikov A**. Time and Sample Efficient Discovery of Markov Blankets and Direct Causal Relations. *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.
- 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. *16th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2003.
- Tsamardinos I, Aliferis CF, **Statnikov A**. Algorithms for Large Scale Markov Blanket Discovery. *16th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 2003.
- 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.
- 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.
- 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.

Posters and Abstracts in Conference Proceedings:

- 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.
- 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.
- **Statnikov A**, Tsamardinos I, Aliferis CF. Using GEMS for Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data. *13th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*, 2005.

Software Demonstrations in Conference Proceedings:

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

Papers in Preparation or under Review:

- **Statnikov A**, Aliferis CF. Algorithms for Identification of Multiple Markov Boundaries. In preparation, 2008.
- **Statnikov A**, Aliferis CF. Analysis and Computational Dissection of Molecular Signature Multiplicity. In preparation, 2008.
- Fananapazir N, **Statnikov A**, Aliferis CF. The FAST-AIMS Clinical Mass Spectrometry Analysis System. Submitted, 2008.
- Mani S, Aliferis CF, **Statnikov A**. Bayesian Algorithms for Causal Data Mining. Submitted, 2008.
- 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. Submitted, 2008.

Technical Reports:

- **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.

- **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.
- 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.
- **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.
- 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:

- **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.
- **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.
- **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:

- **Statnikov A**, Aliferis CF. A Method for Determining All Markov Boundaries and Its Application for Discovering Multiple Optimally Predictive and Non-Redundant Molecular Signatures. *United States Provisional Patent Application*, 2008.
- **Statnikov A**, Aliferis CF, Tsamardinos I, Fananapazir N. Method and System for Automated Supervised Data Analysis. *United States Patent Application #20070122347*, 2006.

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

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

Conferences:

- 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009): reviewer of KDD cup proposals
- Pacific Symposium on Biocomputing (PSB) 2009
- 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
- 3rd Indian International Conference on Artificial Intelligence (IICAI 2007)
- American Medical Informatics Association (AMIA) 2006 Symposium
- 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006): Workshop on Data Mining for Biomedical Applications (BioDM 2006)
- 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005)

Books:

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

Service

- Co-Organizer of the “*Causality Challenge #2: Potluck*”: August 2008 - Present
- 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 - Present
- 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

- Tutorial “*Support Vector Machines without Tears*” at the *AMIA Annual Symposium*, 2008
- 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

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