

# Andrey A. Shabalin, Ph.D.

Postdoctoral Research Associate,  
Department of Biostatistics, University of North Carolina at Chapel Hill

2308 McGavran-Greenberg, 135 Dauer Drive,  
Chapel Hill, NC 27599, USA

**E-mail:** [shabalin@email.unc.edu](mailto:shabalin@email.unc.edu)

**Phone:** (919) 923-8325

**Website:** [shabal.in](http://shabal.in)

## Education:

2005 – 2010, Ph.D. in Statistics, Department of Statistics and Operations Research,  
University of North Carolina at Chapel Hill. Professor Andrew B. Nobel, adviser.

2002 – 2004, M.A. in Economics, New Economic School, Moscow, Russia.

1997 – 2002, M.S. in Mathematics (with honors), Moscow State University, Moscow, Russia.

## Positions:

Postdoctoral Research Associate, UNC, Professor Fred A. Wright, 2010 – present.

Graduate Research Assistant, UNC, Professor Ivan Rusyn, 2008.

Graduate Research Assistant, UNC, Professor Andrew B. Nobel, 2006 – 2008, 2010.

Graduate Research Assistant, UNC, Professor Charles M. Perou, 2007.

Economist, Centre for Economic and Financial Research, Moscow, Russia, 2004 – 2005.

Research Assistant, New Economic School, Moscow, Russia, 2004 – 2005.

## Honors:

Hoeffding award, Department of Statistics and Operations Research, Fall 2006  
(awarded to the student with the best first year performance).

## Grant activity:

Participated in writing of the following successful NSF and NIH grant applications

- NSF grant number DMS-0907177: “Significance Based Procedures for Mining and Prediction of Large Data Sets”. PI: Andrew B. Nobel
- NIH grant number 1R01MH090936-01 “Facilitating GTEx, Disease, and GxE Analyses via Fast Expression (e)QTL Mapping”. PIs: Andrey B. Nobel, Ivan Rusyn, Fred A. Wright.

## Publications:

seeQTL: A searchable database for human eQTLs.

K. Xia, **A.A. Shabalín**, *et al.*

*Bioinformatics*, 2011, in print.

Basal-like Breast Cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival.

V.J. Weigman, H. H. Chao, **A.A. Shabalín**, *et al.*

*Breast Cancer Research and Treatment*, 2 Nov 2011 pp. 1-16.

Detection of Low Rank Signals in Noise and Fast Correlation Mining with Applications to Large Biological Data.

**A.A. Shabalín**

Ph.D. Dissertation. UNC-CH, Department of Statistics and Operations Research, 2010.

Sex-specific Gene Expression in BXD Mouse Liver.

D. Gatti, N. Zhao, E. Chesler, B. Bradford, **A.A. Shabalín**, R. Yordanova, L. Lu, and I. Rusyn.

*Physiological Genomics*, August 2010 vol. 42 no. 3 pp. 456-468.

Finding large average submatrices in high dimensional data.

**A.A. Shabalín**, V.J. Weigman, C.M. Perou, and A.B. Nobel.

*Annals of Applied Statistics*, 3(3), 985-1012, 2009.

FastMap: Fast eQTL mapping in homozygous populations.

D.M. Gatti\*, **A.A. Shabalín**\* (co-first authors), T.C. Lam, F.A. Wright, I. Rusyn, and A.B. Nobel.

*Bioinformatics*, 25(4):482, 2009.

The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency.

C. Lickwar, B. Rao, **A.A. Shabalín**, A.B. Nobel, B.D. Strahl, and J.D. Lieb.

*PLoS ONE*, 4(3), e4886, 2009.

Merging two gene-expression studies via cross-platform normalization.

**A.A. Shabalín**, H. Tjelmeland, C. Fan, C.M. Perou, and A.B. Nobel.

*Bioinformatics*, 24(9):1154, 2008.

## Submitted papers:

Matrix eQTL: Ultra fast eQTL analysis via large matrix operations.

**A.A. Shabalín**.

*Arxiv preprint* at <http://arxiv.org/abs/1105.5764>.

Computational Tools for Discovery and Interpretation of Expression Quantitative Trait Loci (eQTL).

F.A. Wright, **A.A. Shabalín**, and I. Rusyn.

Reconstruction of a Low-rank Matrix in the Presence of Gaussian Noise.

**A.A. Shabalín** and A.B. Nobel.

*Arxiv preprint* at <http://arxiv.org/abs/1007.4148>. Submitted.

### **Working papers:**

FastMap 2.0: Fast Association Mapping in Heterozygous Populations

D.M. Gatti, **A.A. Shabalín**, M. Sypa, T.C. Lam, F.A. Wright, A.B. Nobel, and I. Rusyn.

### **Programming languages:**

Matlab (multithreaded and GPU programming, vectorization, see LAS, Matrix eQTL),  
R (creating packages for CRAN submission, vectorization, see Matrix eQTL),  
Java, C# (multithreaded programming, complex data structures, see LAS, FastMap),  
Gauss, SPSS, Stata.

### **Software developed:**

Matrix eQTL. Ultra fast eQTL analysis via large matrix operations.

R and Matlab, [http://www.bios.unc.edu/research/genomic\\_software/Matrix\\_eQTL/](http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/)

LAS Biclustering. Finding large average submatrices in high-dimensional data.

C# (with GUI) and Matlab, <https://genome.unc.edu/las/>

FastMap. Fast gene expression quantitative trait loci mapping tool.

Java (with GUI), <http://comptox.unc.edu/resources.html>

XPN. Cross-platform normalization method for combining gene expression data from different studies.

Matlab, <https://genome.unc.edu/xpn/>

SWITCHdna. SupWald identification of dna-copy changes.

R, <https://genome.unc.edu/pubsup/SWITCHdna/>

### **Teaching Experience:**

Introduction to Statistics, Instructor, UNC-CH, Spring - Fall 2009.

Responsibilities: conduct lectures, design homework and exams, grade exams, supervise instructional assistant and grader.

Econometrics and Time-series analysis, Instructor, State University of Humanities, Moscow, Russia, Fall 2003 – Spring 2004.

Responsibilities: complete charge of the course – choose textbook, conduct lectures and sessions, design and grade homework, quizzes, and exams.

Applied Time Series Econometrics, Econometrics IV, and Continuous Time Finance,  
Instructor Assistant, New Economic School, Moscow, Russia, Fall 2004 – Spring 2005.  
Responsibilities: conduct sessions, grade homework and exams.

Elements of Statistics, Instructor Assistant, International College of Economics and Finance,  
Moscow, Russia, Fall 2003 – Spring 2004.  
Responsibilities: conduct sessions, grade homework and exams.

**Reviewed for:**

Annals of Applied Statistics,  
Bioinformatics,  
BMC Bioinformatics,  
IEEE/ACM Transactions on Computational Biology and Bioinformatics,  
IEEE Transactions on Information Theory,  
Quantile (Russia)