

# *Yuliya Karpievitch*

Department of Statistics, Texas A&M University  
3143 TAMU,  
College Station, TX 77843-3143, USA

office phone: 979-845-8882  
email: yuliya@stat.tamu.edu  
<http://stat.tamu.edu/~yuliya>

## **EDUCATION**

**Ph.D.**, Bioinformatics, December 2007

Department of Biostatistics, Bioinformatics and Epidemiology  
Medical University of South Carolina, Charleston, SC 29425

*Thesis Title: Computational Tools for MS-based Proteomics*

*Research Interests:* Machine learning using non-parametric algorithms,  
clustered mass spectrometry and microarray data analysis and normalization,  
biomarker identification

**MS**, Computer Science, 2003

College of Charleston, Charleston, SC 29424

Project Title: Solving Global Optimization Problems Using Interval Arithmetic

**BS**, Computer Information Systems, 2001

College of Charleston, Charleston, SC 29424

## **EXPERIENCE**

Aug 2007 - present

Bioinformatics Post Doctoral Fellow, Department of Statistics, Texas A&M University

Feb 2006 – Aug 2007

Graduate Research assistant, Department of Bioinformatics and Computational Biology, M. D. Anderson Cancer Center, University of Texas

- RF++: Random Forest for clustered mass spectrometry and microarray data classification and biomarker identification
- PrepMS: Graphical mass spectrometry data preprocessing tool

Aug 2003 - Dec 2007

Graduate student, Medical University of South Carolina, Charleston, SC

Jan 2003 - Dec 2005

Adjunct Instructor, Computer Science Department, College of Charleston

- Object Oriented Design with C++
- Java Programming Lab
- Introductory Computer Fluency

Aug 2001 - Dec 2003

Graduate Teaching Assistant (teaching beginning Java Lab),  
Computer Science Department, College of Charleston

Spring 2001

Programming Intern, Information Technology Lab, MUSC,  
22-C Ehrhardt St, Charleston, SC

- Provide restricted access to MySQL database via web using Perl, PHP

Summer 2000

Programming Intern (C++), Automated Trading Desk, Mt. Pleasant

- Develop parser to process incoming text data stream
- Use Object Oriented Design Patterns in software development process

Fall 1999 - Spring 2001

Teaching Assistant (assisting instructor in Java Programming Lab),  
Computer Science Department, College of Charleston

**PEER-REVIEWED  
PUBLICATIONS**

**Karpievitch YV**, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KR, Almeida JS PrepMS: Mass Spectrometry Graphical Preprocessing Tool. *Bioinformatics* 2006.

**Karpievitch YV**, Almeida JS. mGrid: A parallel Matlab library for user code distribution. *BMC Bioinformatics* 2006, 7:139

**PUBLICATIONS IN  
PROGRESS**

**Karpievitch YV**, Callister S, Smith RD, Dabney AR. Normalization of peak intensities from tandem mass spectrometry using Singular Value Decomposition.

**Karpievitch YV**, Huang J, Jaitly N, Manes N, Qian W-J, Metz TO, Smith RD, Dabney AR. Quantitation of protein abundance levels in tandem mass spectrometry with missing peaks.

**Karpievitch YV**, Hill EG, Leclerc AP, Almeida JS. RF++: Generalized Random Forest-based classifier for cluster-correlated data.

**PRESENTATIONS**

**Karpievitch YV**, Hill EG, Leclerc AP, Almeida JS. RF++: Generalized Random Forest-based classifier for cluster-correlated data. NLM annual trainee meeting 2007.

**Karpievitch YV**, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KR, Almeida JS PrepMS: Mass Spectrometry Graphical Preprocessing Tool. NLM annual trainee meeting 2006.

**Karpievitch YK**, Almeida JS. mGrid: A parallel Matlab library for user code distribution. NLM annual trainee meeting 2004.

**POSTER  
PRESENTATIONS**

**Karpievitch YV**, Dabney AR. Normalization of peak intensities from tandem mass spectrometry using Singular Value Decomposition. ISMB 2008.

**Karpievitch YV**, Hill EG, Leclerc AP, Almeida JS. RF++: Generalized Random Forest-based classifier for cluster-correlated data. ISMB 2007.

**Karpievitch YV**, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KR, Almeida JS PrepMS: MS-TOF Data Graphical Preprocessing Tool. HUPO 2006.

**Karpievitch YV**, Hill EG, Millar TP, Smolka AJ, Hoffman BJ, Almeida JS. Modified Random Forest Algorithm for Biomarker Identification in Clustered Mass Spectrometry Data. ISMB 2006.

Millar TP, **Karpievitch YK**, Hill EG, Hammond CE, Hoffman BJ, Smolka AJ. Identification of Human Esophageal Cancer Subtypes by Serum Protein Analysis Using MALDI-TOF Mass Spectroscopy. Digestive Disease Week 2006.

**Karpievitch YK**, Schey KL, Schwacke JH. Improving the Performance of a Protein Identification Algorithm for Intact Protein MALDI/TOF/TOF Spectra Using Parallel Scoring Algorithm. NLM annual trainee meeting 2005.

**PROGRAMMING  
SKILLS**

C++, C, PHP, Matlab, R, Java, Bash shell scripting, Perl  
Developing Matlab stand-alone (compiled) GUI executables  
PHP and Perl scripting for HTTP database front-end, SQL  
Linux system programming including system calls and kernel modules  
LEX and YACC application in implementing a bottom-up parser  
MPI, Distributed System Architectures including Java RMI, SOAP, HTTP

**GRANT SUPPORT**

National Cancer Institute Postdoctoral Bioinformatics Training Grant  
(2007-2009)

National Library of Medicine Training Grant: Toolmakers for Bioinformatics  
(2003-2007) T15-LN07438

ISMB 2007 Travel Fellowship, Vienna, Austria

FASEB(MARC) Travel Award for ISMB 2006, Fortaleza, Brazil

F. Hoffmann-La Roche AG Travel Award for HUPO 2006, Long Beach, CA