

# Andrey A. Shabalin, Ph.D.

Assistant Professor  
Center for Biomarker Research and Precision Medicine  
Virginia Commonwealth University

1112 East Clay Street  
McGuire Hall, Room 219  
Richmond, VA 23298

**E-mail:** [ashabalin@vcu.edu](mailto:ashabalin@vcu.edu)

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

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

**GitHub:** [github.com/andreyshabalin](https://github.com/andreyshabalin)

**Google Scholar:** [scholar.google.com/citations?hl=en&user=wXj\\_PpsAAAAJ](https://scholar.google.com/citations?hl=en&user=wXj_PpsAAAAJ)

**NCBI Bibliography:** [ncbi.nlm.nih.gov/sites/myncbi/110\\_9pCfUhMkh/bibliography/51285584/public](https://ncbi.nlm.nih.gov/sites/myncbi/110_9pCfUhMkh/bibliography/51285584/public)

## 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:

2014 – now, Assistant Professor, VCU  
2012 – 2014, Postdoctoral Research Associate, VCU, Professor Edwin J.C.G. van den Oord  
2010 – 2012, Postdoctoral Research Associate, UNC, Professor Fred A. Wright  
2004 – 2005, Economist, Centre for Economic and Financial Research, Moscow, Russia  
2004 – 2005, Research Assistant, New Economic School, Moscow, Russia

## Publications:

Enrichment methods provide a feasible approach to comprehensive and adequately powered investigations of the brain methylome (2017)  
R.F. Chan, **A.A. Shabalin**, L.Y. Xie, D.E. Adkins, M. Zhao, G. Turecki, S.L. Clark, K.A. Aberg, E.J.C.G. van den Oord  
*Nucleic Acids Research*, gkx143  
DOI: 10.1093/nar/gkx143  
<https://academic.oup.com/nar/article/3052767>

Deep sequencing of 71 candidate genes to characterize variation associated with alcohol dependence (2017)  
S.L. Clark, D.E. Adkins, G. Kumar, K.A. Aberg, S. Nerella, L. Xie, A.L. Collins, J.J. Crowley, C.R. Quackenbush, C.E. Hilliard, **A.A. Shabalin**, S.I. Vrieze, R.E. Peterson, W.E. Copeland,

J.L. Silberg, M. McGue, H. Maes, W.G. Iacono, P.F. Sullivan, E.J. Costello, E.J.C.G. van den Oord

*Alcoholism: Clinical and Experimental Research*, 2017 Feb 14

DOI: 10.1111/acer.13352, PMID: 28196272

<http://onlinelibrary.wiley.com/doi/10.1111/acer.13352/full>

Correcting for cell-type effects in DNA methylation studies: reference-based method outperforms latent variable approaches in empirical studies (2017)

M.W. Hattab, **A.A. Shabalín**, S.L. Clark, M. Zhao, G. Kumar, R.F. Chan, L.Y. Xie, R. Jansen, L.K.M. Han, P.K.E. Magnusson, G. van Grootheest, C.M. Hultman, B.W.J.H. Penninx, K.A. Aberg, E.J.C.G. van den Oord

*Genome Biology*, 2017 18:24

DOI: 10.1186/s13059-017-1148-8, PMID: 28137292, PMCID: PMC5282865

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1148-8>

High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction (2015)

J.L. McClay\*, **A.A. Shabalín**\*(co-first authors), ..., K.A. Aberg, and E.J.C.G. van den Oord

*Genome Biology*, 16:291

DOI: 10.1186/s13059-015-0842-7, PMID: 26699738, PMCID: PMC4699364

<http://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0842-7>

Candidate gene methylation studies are at high risk of erroneous conclusions (2015)

**A.A. Shabalín**, K.A. Aberg, and E.J.C.G. van den Oord

*Epigenomics*, 7(1), 13-5

DOI: 10.2217/epi.14.70, PMID: 25687462

<http://www.futuremedicine.com/doi/abs/10.2217/epi.14.70>

An integrated map of structural variation in 2,504 human genomes (2015)

P.H. Sudmant, ..., **A.A. Shabalín** (co-author 62 of 83)

*Nature*, 526, 75–81

DOI: 10.1038/nature15394, PMID: 26432246, PMCID: PMC4617611

<http://www.nature.com/nature/journal/v526/n7571/abs/nature15394.html>

The Genotype-Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans (2015)

K.G. Ardlie, ..., **A.A. Shabalín** (co-author 22 or 139)

*Science* 348 (6235), 648-660

DOI: 10.1126/science.1262110, PMID: 25954001, PMCID: PMC4547484

<http://www.ncbi.nlm.nih.gov/pubmed/25954001>

A Whole Methylome CpG-SNP Association Study of Psychosis in Blood and Brain Tissue (2015)

E.J.C.G. van den Oord, S.L. Clark, L.Y. Xie, **A.A. Shabalín**, M.G. Dozmorov, G. Kumar, Swedish Schizophrenia Consortium, V.I. Vladimirov, P.K.E. Magnusson, and K.A. Aberg  
*Schizophrenia Bulletin*, 182

DOI: 10.1093/schbul/sbv182, PMID: 26656881, PMCID: PMC4903046  
<http://schizophreniabulletin.oxfordjournals.org/content/early/2016/02/05/schbul.sbv182>

Deep Sequencing of Three Loci Implicated in Large-Scale Genome-Wide Association Study Smoking Meta-Analyses (2015)

S.L. Clark, J.L. McClay, D.E. Adkins, K.A. Aberg, G. Kumar, S. Nerella, L. Xie, A.L. Collins, J.J. Crowley, C.R. Quakenbush, C.E. Hillard, G. Gao, **A.A. Shabalin**, R.E. Peterson, W.E. Copeland, J.L. Silberg, H. Maes, P.F. Sullivan, E.J. Costello, and E.J.C.G. van den Oord  
*Nicotine & Tobacco Research*, 166

DOI: 10.1093/ntr/ntv166  
<http://ntr.oxfordjournals.org/content/early/2015/08/16/ntr.ntv166.short>

Combined whole methylome and genomewide association study implicates CNTN4 in alcohol use (2015)

S.L. Clark, K.A. Aberg, S. Nerella, G. Kumar, J.L. McClay, W. Chen, L.Y. Xie, A. Harada, **A.A. Shabalin**, G. Gao, S.E. Bergen, C.M. Hultman, P.K.E. Magnusson, P.F. Sullivan, and E.J.C.G. van den Oord

*Alcoholism: Clinical and Experimental Research*, 39 (8), 1396-1405

DOI: 10.1111/acer.12790, PMID: 26146898, PMCID: PMC4515164  
<http://onlinelibrary.wiley.com/doi/10.1111/acer.12790/full>

Refinement of schizophrenia GWAS loci using methylome-wide association data (2015)

G. Kumar, S.L. Clark, J.L. McClay, **A.A. Shabalin**, D.E. Adkins, L. Xie, R. Chan, S. Nerella, Y. Kim, P.F. Sullivan, C.M. Hultman, P.K.E. Magnusson, K.A. Aberg, and E.J.C.G. van den Oord  
*Human Genetics*, 134 (1), 77-87

DOI: 10.1007/s00439-014-1494-5, PMID: 25284466, PMCID: PMC4282961  
<http://link.springer.com/article/10.1007%2Fs00439-014-1494-5>

Quantitative trait locus mapping methods for diversity outbred mice (2014)

D.M. Gatti, K.L. Svenson, **A.A. Shabalin**, L.Y. Wu, W. Valdar, P. Simecek, N. Goodwin, R. Cheng, D. Pomp, A. Palmer, E.J. Chesler, K.W. Broman, G.A. Churchill  
*G3: Genes Genomes Genetics*, 2014 vol. 4 no. 9, 1623-1633

DOI: 10.1534/g3.114.013748, PMID: 25237114, PMCID: PMC4169154  
<http://www.g3journal.org/content/4/9/1623>

The Genotype-Tissue Expression (GTEx) project (2013)

J. Lonsdale, J. Thomas, ..., **A.A. Shabalin** (co-author 98 of 127)  
*Nature Genetics*, 45 (6), 580-585

DOI: 10.1038/ng.2653, PMID: 23715323, PMCID: PMC4010069  
<http://www.nature.com/ng/journal/v45/n6/abs/ng.2653.html>

Heritability and genomics of gene expression in peripheral blood (2014)

F.A. Wright, ..., **A.A. Shabalin** (co-author 25 of 39)  
*Nature Genetics*, 46 (5), 430-437

DOI: 10.1038/ng.2951, PMID: 24728292, PMCID: PMC4012342  
<http://www.nature.com/ng/journal/v46/n5/full/ng.2951.html>

Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies (2013)

A. Ramasamy, D. Trabzuni, J.R. Gibbs, A. Dillman, D.G. Hernandez, S. Arepalli, R. Walker, C. Smith, G.P. Illori, **A.A. Shabalín**, Y. Li, A.B. Singleton, M.R. Cookson, J. Hardy, M. Ryten, M.E. Weale

*Nucleic Acids Research*, 41 (7), e88-e88

DOI: 10.1093/nar/gkt069, PMID: 23435227, PMCID: PMC3627570

<http://nar.oxfordjournals.org/content/41/7/e88>

Reconstruction of a low-rank matrix in the presence of Gaussian noise (2013)

**A.A. Shabalín** and A.B. Nobel

*Journal of Multivariate Analysis*, 118, 67-76

DOI: 10.1016/j.jmva.2013.03.005

<http://www.sciencedirect.com/science/article/pii/S0047259X13000328>

Matrix eQTL: Ultra fast eQTL analysis via large matrix operations (2012)

**A.A. Shabalín**

*Bioinformatics*, 28 (10): 1353-1358

DOI: 10.1093/bioinformatics/bts163, PMID: 22492648, PMCID: PMC3348564

<http://bioinformatics.oxfordjournals.org/content/28/10/1353>

seeQTL: A searchable database for human eQTLs (2012)

K. Xia, **A.A. Shabalín**, ..., and F.A. Wright

*Bioinformatics*, 28 (3): 451-452

DOI: 10.1093/bioinformatics/btr678, PMID: 22171328, PMCID: PMC3268245

<http://bioinformatics.oxfordjournals.org/content/28/3/451>

Computational tools for discovery and interpretation of Expression Quantitative Trait Loci (eQTL) (2012)

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

*Pharmacogenomics*, 13 (3), 343-352

DOI: 10.2217/pgs.11.185, PMID: 22048815, PMCID: PMC3295835

<http://www.futuremedicine.com/doi/abs/10.2217/pgs.11.185>

Basal-like Breast Cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival (2012)

V.J. Weigman, H.H. Chao, **A.A. Shabalín**, ..., and C.M. Perou

*Breast Cancer Research and Treatment*, 133 (3), 865-880

DOI: 10.1007/s10549-011-1846-y, PMCID: PMC3387500

<http://link.springer.com/article/10.1007/s10549-011-1846-y>

Sex-specific Gene Expression in BXD Mouse Liver (2010)

D. Gatti, N. Zhao, E. Chesler, B. Bradford, **A.A. Shabalín**, R. Yordanova, L. Lu, and I. Rusyn  
*Physiological Genomics*, 42 (3), 456-468

DOI: 10.1152/physiolgenomics.00110.2009, PMID: 20551147, PMCID: PMC2929887

<http://physiolgenomics.physiology.org/content/42/3/456>

Finding large average submatrices in high dimensional data (2009)

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

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

DOI: 10.1214/09-AOAS239

<http://projecteuclid.org/euclid.aoas/1254773275>

<http://www.jstor.org/stable/30242874>

FastMap: Fast eQTL mapping in homozygous populations (2009)

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

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

DOI: 10.1093/bioinformatics/btn648, PMID: PMC2642639

<http://bioinformatics.oxfordjournals.org/content/25/4/482>

The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency (2009)

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

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

DOI: 10.1371/journal.pone.0004886, PMID: 19295910, PMCID: PMC2654109

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004886>

Merging two gene-expression studies via cross-platform normalization (2008)

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

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

DOI: 10.1093/bioinformatics/btn083

<http://bioinformatics.oxfordjournals.org/content/24/9/1154>

Detection of Low Rank Signals in Noise and Fast Correlation Mining with Applications to Large Biological Data (2010)

**A.A. Shabalin**

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

<https://cdr.lib.unc.edu/record/uuid:6a46e597-9e72-44f1-bd73-f643501bc0db>

#### Submitted and working papers:

Estimation of Interpretable eQTL Effect Sizes Using a Log of Linear Model

J. Palowitch, **A.A. Shabalin**, Y. Zhou, A.B. Nobel, and F.A. Wright

<https://arxiv.org/abs/1605.08799>

Local genetic effects on gene expression across 44 human tissues

F. Aguet, ..., **A.A. Shabalin** (co-author 24 of 50)

<http://biorxiv.org/content/early/2016/09/09/074450>

A non-parametric statistical framework for integrating heterogeneous prior information in large-scale multiple testing

**A.A. Shabalin**, J. Bukzar, J.L. McClay, K.A. Aberg, and E.J.C.G. van den Oord

An empirical bayes approach for Multiple Tissue eQTL Analysis  
G. Li, **A.A. Shabalin**, I. Rusyn, F.A. Wright, and A.B. Nobel  
<http://arxiv.org/abs/1311.2948>

FastMap 2.0: Fast Association Mapping in Heterozygous Populations  
D.M. Gatti, **A.A. Shabalin**, M. Sypa, T.C. Lam, F.A. Wright, A.B. Nobel, and I. Rusyn  
<http://comptox.unc.edu/fastmap.php>

### Programming languages:

R (creating packages for CRAN and Bioconductor, see Matrix eQTL, RaMWAS)  
Matlab (multithreaded and GPU programming, vectorization, see LAS, XPN, 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 at CRAN, [http://www.bios.unc.edu/research/genomic\\_software/Matrix\\_eQTL/](http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/)

RaMWAS. Fast Methylome-Wide Association Study Pipeline for Enrichment Platforms

R at Bioconductor, <https://bioconductor.org/packages/ramwas/>

Filematrix. File-Backed Matrix Class with Convenient Read and Write Access

R at CRAN, <https://github.com/andreyshabalin/filematrix>

fastCircularPermutations. Very fast circular permutation analysis on binary outcomes

R at github, <https://github.com/andreyshabalin/fastCircularPermutations>

ACMEeqtl. Estimation of Interpretable eQTL Effect Sizes Using a Log of Linear Model

R at GitHub, <https://github.com/andreyshabalin/ACMEeqtl>

CorrMeta: Fast Association Analysis for eQTL with Related Samples

R at [http://www.bios.unc.edu/research/genomic\\_software/CorrMeta/](http://www.bios.unc.edu/research/genomic_software/CorrMeta/)

DOQTL. Genotyping and QTL Mapping in Diversity Outbred Mice

R at Bioconductor, <https://bioconductor.org/packages/release/bioc/html/DOQTL.html>

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/fastmap.php>

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

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.

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, etc.

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  
BMC Genomics  
GigaScience  
Human Molecular Genetics  
IEEE/ACM Transactions on Computational  
Biology and Bioinformatics  
IEEE Transactions on Information Theory  
Journal of Computational and Graphical  
Statistics

Journal of Multivariate Analysis  
Nucleic Acids Research  
Pattern Recognition  
PLOS Genetics  
PLOS One  
Quantile (Moscow, Russia)  
Statistics and Computing  
Transactions on Computational Biology and  
Bioinformatics