

# ROBERT B. SCHARPF <sup>1</sup>

Johns Hopkins University

550 N. Broadway, Suite 1103 · Baltimore, MD 21205-2179

Phone: 410-614-7837 · Email: [rscharpf@jhsph.edu](mailto:rscharpf@jhsph.edu)

Website: <http://www.biostat.jhsph.edu/people/postdocs/scharpf.shtml>

---

## EDUCATION

- |           |                                                                                                                                                                                                                                                   |                  |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| 2010-     | JOHNS HOPKINS UNIVERSITY<br>Assistant Professor, Department of Oncology                                                                                                                                                                           | Baltimore, MD    |
| 2007-2010 | JOHNS HOPKINS BLOOMBERG SCHOOL OF PUBLIC HEALTH<br>Postdoctoral Fellow, Biostatistics<br>Mentors: Ingo Ruczinski, PhD (primary) · Josef Coresh, MD, PhD (co-mentor)                                                                               | Baltimore, MD    |
| 2002-2007 | JOHNS HOPKINS BLOOMBERG SCHOOL OF PUBLIC HEALTH<br>PhD, Biostatistics<br>Advisors: Giovanni Parmigiani, PhD (primary) · Ingo Ruczinski, PhD (co-advisor)<br>Dissertation: <i>Combining high-throughput genomic data: methods and applications</i> | Baltimore, MD    |
| 1995-1998 | JOHNS HOPKINS UNIVERSITY KRIEGER SCHOOL OF ARTS AND SCIENCES<br>MS, Biotechnology                                                                                                                                                                 | Baltimore, MD    |
| 1990-1994 | JAMES MADISON UNIVERSITY<br>BS <i>Magna Cum Laude</i> , Biology                                                                                                                                                                                   | Harrisonburg, VA |

## HONORS AND AWARDS

- 2009 NIH K99/R00 Pathways to Independence Award
- 2008 Young Investigator Travel Award, Markov Chain Monte Carlo in Theory and Practice ("MCM-Ski") Conference, the 2nd International Joint Meeting of the Institute of Mathematical Statistics and the International Society for Bayesian Analysis, Bormio, Italy
- 2007 Margaret Merrell Award for Excellence in Research, Johns Hopkins Bloomberg School of Public Health
- 2007 Helen Abbey Award for Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health
- 2006 Travel Scholarship sponsored by the American Statistical Association, Bioconductor Conference, Seattle, WA
- 2006 Nominated for Teaching Assistant Recognition Award, Johns Hopkins Bloomberg School of Public Health
- 2001 Research Citation Finalist, Anesthesiology and Critical Care Medicine Conference, San Francisco, CA
- 1994 Finalist for Excellence in Biology Award, James Madison University

---

<sup>1</sup>Updated September 9, 2010

**PROFESSIONAL EXPERIENCE**

2007-2008	<i>Visiting Assistant Professor</i> Department of Health Sciences, American University of Armenia	Yerevan, Armenia
2006-2007	<i>Consultant</i> for Department of Health Policy and Management PhD candidates Department of Biostatistics, Johns Hopkins School of Public Health	Baltimore, MD
2002-2007	<i>Predoctoral Research</i> Sidney Kimmel Cancer Center, Johns Hopkins School of Medicine	Baltimore, MD
1998-2002	<i>Laboratory Research Assistant</i> Department of Anesthesiology, Johns Hopkins School of Medicine	Baltimore, MD
1995-1998	<i>Informatics Research Assistant</i> Division of Biomedical Sciences, Johns Hopkins School of Medicine	Baltimore, MD

**TEACHING EXPERIENCE**

## PRIMARY INSTRUCTOR

*Department of Health Sciences, American University of Armenia*

- 2007 (Fall) Inferential Biostatistics I
- 2008 (Fall) Modeling and Sampling

## CO-INSTRUCTOR

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

- 2006 (Summer) The Analysis of Gene Expression Data

## TEACHING ASSISTANT

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

- 2005-2007 Lead Teaching Assistant for Statistical Methods in Public Health I and II
- 2005 Generalized Linear Models III, IV
- 2004 Introduction to Probability I, II
- 2003-2004 Statistical Methods in Public Health I, II, III, IV

**PUBLICATIONS**

## PEER-REVIEWED ARTICLES

1. Leek et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, to appear.
2. **Scharpf RB**, Iacobuzio-Donahue CA, Cope, L, Ruczinski I, Garrett-Mayer E, Lakkur S, Campagna D, and Parmigiani G (2010). Cross-platform comparison of two pancreatic cancer phenotypes. *Cancer Informatics*, to appear.
3. **Scharpf RB**, Ruczinski I, Carvalho B, Doan B, Chakravarti A, and Irizarry RA (2010). A multilevel model to address batch effects in copy number estimation using SNP arrays. *Biostatistics* July 12 (Epub).
4. **Scharpf RB** and Ruczinski I (2010). R classes and methods for SNP array data. *Methods Mol Biol*, 593:67-69. PMID: 19957145.
5. Carvalho B, Irizarry RA, **Scharpf RB**, and Carey VJ (2009). Processing and analyzing Affymetrix SNP chips with Bioconductor. *Statistics In Biosciences*, 1(2), 160-180 .
6. Caffo B, Liu D, **Scharpf RB**, and Parmigiani G (2009). Likelihood estimation of conjugacy relationships in linear models with applications to high-throughput genomics. *International Journal of Biostatistics*, 5(1), Article 18. DOI:10.2202/1557-4679.1129.

7. **Scharpf RB**, Tjelmeland H, Parmigiani G, and Nobel AB (2009). A Bayesian model for cross-study differential gene expression, *Journal of the American Statistical Association*, 140(488), p1295–1310 .
8. **Scharpf RB**, Tjelmeland H, Parmigiani G, and Nobel AB (2009). Rejoinder, *Journal of the American Statistical Association*, 140(488), p1318-1323 .

\* These authors contributed equally.

† These authors jointly directed.

9. Kottgen A\*, Glazer NL\*, Dehghan A\*, Hwang SJ\*, . . . , **Scharpf RB**, . . . , Witteman JC†, Coresh J†, Shlipak MG†, and Fox CS† (2009). Multiple loci associated with indices of renal function and chronic kidney disease. *Nature Genetics*, v41, p712-717. PMID: 19430482.
- \* These authors contributed equally.
10. Levy D\*, Ehret G\*, Rice K \*, Verwoert GC \*, Launer LJ, . . . , **Scharpf RB**, . . . , Gudnason V\*, Larson MG\*, Chakravarti A\*, Psaty BM\*, and van Duijn CM\* (2009). Genome-wide association study of blood pressure and hypertension. *Nature Genetics*, v41, p677-687. PMID: 19430479
11. **Scharpf RB**, Parmigiani G, Pevsner J, and Ruczinski I (2008). Hidden Markov models for the assessment of chromosomal alterations using high-throughput SNP arrays. *Annals of Applied Statistics*, 2(2) 687-713 [pdf].
12. Zahurak ML, Parmigiani G, Yu W, **Scharpf RB**, Berman D, Schaeffer E, Shabbeer S, and Cope L (2007). Pre-processing Agilent microarray data. *BMC Bioinformatics*, 8:142 [pdf].
13. **Scharpf RB**, Ting JC, Pevson J, and Ruczinski I (2007). SNPchip: R classes and methods for SNP array data. *Bioinformatics*, 23(5):627-628 [pdf].
14. **Scharpf RB**, Iacobuzio-Donahue CA, Sneddon JB, and Parmigiani G (2007). When should one subtract background fluorescence in two color microarrays? *Biostatistics*; 8(4):695-707 [pdf].
15. Morrell CN, Matsushita K, Chiles K, **Scharpf RB**, Yamakuchi M, Mason RJ, Bergmeier W, Mankowski JL, Baldwin WM 3rd, Faraday N, Lowenstein CJ (2004). Regulation of platelet granule exocytosis by S-nitrosylation. *Proceedings of the National Academy of Sciences*, 102:3782-7 [pdf].
16. Faraday N, Martinez EA, **Scharpf RB**, Kasch-Semenza L, Dorman T, Pronovost PJ, Perler B, Thompson RE, Gerstenblith G, Bray PF, Fleisher LA (2004). Platelet gene polymorphisms and perioperative cardiac risk assessment in vascular surgical patients. *Anesthesiology*, 101(6):1291-7 [PubMed].
17. **Scharpf R**, Garrett ES, Hu J, Parmigiani G. Statistical modeling and visualization of molecular profiles in cancer. *Biotechniques*, 34:S22–S29, 2003 [pdf].
18. Faraday N, Gullar E, Sera VA, Bolton ED, **Scharpf RB**, Cartarius AM, Emery K, Concord J, Kickler TS. Utility of whole blood hemostatology using the clot signature analyzer for assessment of hemostasis in cardiac surgery. *Anesthesiology*, 96(5):1115-22, 2002 [PubMed].
19. Faraday N, **Scharpf RB**, Dodd-o JM, Martinez EA, Rosenfeld BA, Dorman T. Leukocytes can enhance platelet-mediated aggregation and thromboxane release via interaction of P-selectin glycoprotein ligand 1 with P-selectin. *Anesthesiology*, 94(1):145-51, 2001 [PubMed].
20. Anagnostopoulos AV, **Scharpf RB**. It's a knockout! *Trends in Genetics*. 14(2):82, Feb 1998 [PubMed].
21. Anagnostopoulos AV, **Scharpf RB**. It's a knockout! *Trends in Genetics*. 13(12):499-500, Dec 1997 [PubMed].

## BOOK CHAPTERS

1. **Scharpf RB** and Garrett-Mayer E. Models for Probability of Under and Overexpression: The POE Scale. In *Bayesian Inference for Gene Expression and Proteomics* (Kim-Anh Do, Peter Mueller, and Marina Vannucci), Cambridge University Press, 2006

## PRESENTATIONS

### INVITED TALKS

- *Copy number estimation and batch effects in SNP arrays*, BIRS Statistical Genomics In Biomedical Research, Banff, Alberta. July 20, 2010.
- *Batch effects in SNP arrays: solutions for genotyping and copy number estimation*, Vassar University, November 17, 2009.
- *Batch effects in SNP arrays: solutions for genotyping and copy number estimation*, Computational Biology and Bioinformatics Seminar, M.D. Anderson Cancer Center, Houston, TX, Sept 30, 2009.
- *Discussant*, Biostatistics Grand Rounds, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, February 4, 2009.
- *Genome-scale estimation of copy number and inference*, Second Annual Young Investigator Symposium on Genomics and Bioinformatics, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, September 5, 2008.
- *Estimation problems in high throughput SNP platforms*, Genome-Wide Association Studies Symposium, Graduate Institute in Epidemiology and Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (June 16, 2008).
- *A Bayesian model for cross-study differential gene expression*. MCMSki II, Bormio, Italy (January 11, 2008).
- *A Bayesian model for cross-study differential gene expression*. 39th Conference on Interface of Statistics and Computing Science, Philadelphia, PA (May 24, 2007).
- *Combining estimates of copy number and genotype in high-throughput SNP chips*. Eastern North American Region (ENAR) International Biometric Society Conference, Atlanta, GA (March 12, 2007).
- *Visualizing and analyzing high density SNP data with SNPscan*. ENAR Conference, Tampa, FL (March 28, 2006).

### OTHER TALKS

- *Batch effects in SNP arrays: solutions for genotyping and copy number estimation*, Computational Biology and Genomics Seminar, Johns Hopkins University, Sept 21, 2009.
- *A hidden Markov model for trios using the Illumina genotyping platform*. The Expressionists Seminar Series, Department of Biostatistics, Johns Hopkins University, Baltimore, MD (August 27, 2008).
- *Estimating copy number from high-throughput Affymetrix SNP platforms*. The Expressionists Seminar Series, Department of Biostatistics, Johns Hopkins University, Baltimore, MD (April 30, 2008).
- *Assessing copy number alterations in high throughput SNP platforms*, Welch Center for Prevention Epidemiology and Research, Johns Hopkins University, Baltimore, MD (March 21, 2008).
- *Assessing copy number alterations using high throughput SNP platforms*. American Heart Association, 48th Cardiovascular Disease Epidemiology and Prevention Conference, Colorado Springs, CO (March 12, 2008).
- *Genotype calling: uncertainty and copy number*. Atherosclerosis Risk in Communities (ARIC) Study Workshop, Houston, TX (June, 2007).
- *Combining high-throughput genomic data: methods and applications*. Thesis defense, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (February 07, 2007).
- *Visualizing and analyzing high density SNP Data with SNPscan*. Biostatistics Departmental Retreat, Baltimore, MD (May, 2006).
- *Leukocytes can enhance platelet-mediated aggregation and thromboxane release via interaction of P-selectin glycoprotein ligand 1 with P-selectin..* Anesthesiology and Critical Care Medicine Conference, San Francisco, CA (February, 2001)

## POSTERS

- *A multilevel model to address batch effects for copy number estimation in SNP arrays.* Statistical Advances in Genome Scale Data Analysis, Ascona, Switzerland (May, 2009).
- *Hidden Markov models for characterizing chromosomal alterations with high-throughput SNP arrays.* 3<sup>rd</sup> Annual Postdoctoral Fellows' Research Poster Competition, Johns Hopkins Bloomberg School of Public Health (June, 2007).
- *Cross-linking of Fcγ RIIa and Fcγ RIIb on human polymorphonuclear neutrophils inhibits platelet aggregation.* Anesthesiology/Critical Care Medicine Poster Presentation Day (October, 2001).

## SOFTWARE

PEER-REVIEWED R PACKAGES AVAILABLE AT [BIOCONDUCTOR](#):

- SNPchip**     **Scharpf RB** and Ruczinski I. Classes and methods for visualization of high-throughput SNP chip data.
- VanillaICE**     **Scharpf RB** and Ruczinski I. Hidden Markov Models for identifying chromosomal aberrations using high-throughput SNP chip data.
- XDE**     **Scharpf RB**, Tjelmeland H, Nobel AB, and Parmigiani G. A Bayesian multilevel model for cross-study differential gene expression.
- crlmm**     Carvalho B and **Scharpf RB**, Genotyping and copy number estimation for high-throughput SNP arrays.
- oligoClasses**     Carvalho B and **Scharpf RB**, Infrastructure for analysis of high-throughput genotyping platforms.

## RESEARCH SUPPORT

## ONGOING RESEARCH SUPPORT

- 4/01/2010 - 4/01/2013     *R00 Pathways to Independence Award* (NIH/NHGRI), grant 4R00HG005015  
Principal Investigator: Robert B. Scharpf, PhD
- 4/01/2010 - present     *Bioinformatics Core, Division of Oncology*  
Principal Investigator: Michael Ochs, PhD.
- 7/01/2010 - present     *Institutional Clinical and Translational Science Award*, (NIH/NCRR) grant U54RR023561  
Principal Investigator: Daniel Ford, MD
- 9/03/2010 - 8/31/2011     *Next-Generation Medical Resequencing of Gout Disease Genes in the ARIC Cohort*, (NIH/NHGRI) grant HG005697  
subcontract Principal Investigator : Linda Kao

## PREVIOUS SUPPORT

- Apr. 2009 - Apr. 2010     *K99 Pathways to Independence Award* (NHGRI), grant 1K99HG005015  
Principal Investigator: Robert B. Scharpf, PhD  
Mentors: Ingo Ruczinski, PhD (primary) · Josef Coresh, MD, PhD (co-mentor)
- Feb. 2007 - Feb. 2009     *Cardiovascular Disease and Epidemiology Fellowship*  
NHLBI grant T32 HL007024,  
Director: Josef Coresh, MD, PhD.
- 2004, 2006-2008     *Bioinformatics training grant*  
NSF grant T32GM074906,  
Principal Investigator: Giovanni Parmigiani, PhD
- 2004-2006     *Environmental Biostatistics and Epidemiology Training Grant*  
NIH grant T32ES12871,  
Principal Investigator: Thomas Louis, PhD

## SERVICE

### PEER REVIEW ACTIVITIES

**Scientific Journals:** American Journal of Epidemiology (2) · Bioinformatics (3) · Biometrics (1) · Biostatistics (3) · BMC Bioinformatics (2) · BMC Genomics (1) · Computational Statistics and Data Analysis (2) · Genomics (1) · Journal of the American Statistical Association (2) · Journal of Statistical Software (1) · Statistics in Medicine (1) · Pharmacogenomics (1) · Statistical Applications in Genetics and Molecular Biology (1) · Statistical Methods and Applications (1)

**Statistical Software:** Bioconductor R packages (1).

**Conferences/Symposia:** Pacific Symposium on Biocomputing (1).

### ORGANIZER OF AN INVITED SESSION

March, 2010 ENAR Statistical models and practice that improve reproducibility in genomics research.

### SESSION CHAIR

March, 2010 ENAR Statistical models and practice that improve reproducibility in genomics  
August, 2010 Joint Statistical Meetings Topic: Genomics and Mutation Classification

## MENTORING

### Dissertation Reader for MPH candidates

2009 Lilit Khachatryan Department of Health Sciences, American University of Armenia  
2009 Kushagra Mahansaria Department of Health Sciences, American University of Armenia  
2009 Mikhayil Melikou Department of Health Sciences, American University of Armenia

### Summer research students

2010 Kinjal Basu Topic: copy number estimation co-advised with Aravinda Chakravarti  
2010 Jack Yu Topic: copy number estimation co-advised with Sarah Wheelan and Jonathan Pevsner

### Research scientists

2010 - Lynn Mireless Research area: copy number analysis for ARIC cohort study co-advised with Linda Kao

### Masters of Bioinformatics candidates

2010 - Moiz Bootwalla Johns Hopkins University

## ADDITIONAL INFORMATION

### PERSONAL STATEMENT OF RESEARCH OBJECTIVES

I am interested in developing statistical methods and software for characterizing diseases and complex traits from DNA and RNA measured on a genomic scale. Two technologies that provide high throughput measurements of genomic information are the focus of my current research: single nucleotide polymorphism (SNP) and gene expression arrays. For SNP arrays, I am developing statistical methods to estimate copy number at a single locus, and to infer regions of copy number variants (CNV) spanning multiple loci. In addition, I am developing methods to assess the association of CNV with phenotypes in genome-wide association studies (GWAS). For gene expression data, I am developing methods for the cross-study analysis of expression data to identify concordant and discordant patterns of differential gene expression. The software developed for each of these methods facilitate reproducible research, unify the steps in complex analyses of genomic data, and are freely available under the GNU General Public Licence of the Free Software Foundation.