

Contact Information

Davide Risso
University of California, Berkeley
344 Li Ka Shing Center
MC # 3370
Berkeley, CA 94720-3370
E-mail: risso.davide@gmail.com
Web-site: <http://daviderisso.wordpress.com>
Github: <https://github.com/drisso>

Research and Teaching Interests

Statistical methodology modeling of high-dimensional data, parametric and non-parametric regression, clustering, factor analysis and dimensionality reduction, multiple hypothesis testing, combining data from multiple sources (meta-analysis).

Genomics data analysis design and analysis of high-throughput gene expression experiments, analysis of microarray and next-generation sequencing data (RNA-seq), non-coding RNA annotation and expression quantitation.

Software development visualization and statistical analysis of high-throughput genomic data, reproducible research.

Education

2009 – 2012 Ph.D. in Statistics

Department of Statistical Sciences, University of Padova, Italy.
Thesis title: “Simultaneous Inference for RNA-Seq data”.
Supervisor: Prof. Monica Chiogna.
Co-supervisors: Prof. Sandrine Dudoit, Dr. Chiara Romualdi.

2005 – 2008 MSc (*Laurea Magistrale*) in “Statistics and Informatics”

Faculty of Statistical Sciences, University of Padova, Italy.
Final grade: 110 (out of 110) with honors.
Thesis (in italian): “Analysis of gene expression data: a comparative study on the impact of normalization on statistical inference”, Supervisor: Dr. Chiara Romualdi.

2002 – 2005 BSc (*Laurea*) in “Statistics, Mathematics and Data Management”

Faculty of Mathematical, Physical and Natural Sciences, University of Genova, Italy.

Final grade: 110 (out of 110) with honors.

Thesis (in italian): “Statistical modeling of fever episodes in neutropenic pediatric patients with cancer”, Supervisor: Dr. Vincenzo Fontana.

Research Experience

April 2012 – present Postdoctoral Scholar

Department of Statistics and Division of Biostatistics, School of Public Health, University of California, Berkeley, USA.

Jointly supervised by Prof. Terry Speed and Prof. Sandrine Dudoit.

Currently funded through the National Institutes of Health BRAIN Initiative grant *Classification of Cortical Neurons by Single Cell Transcriptomics* (U01 MH105979, PI: John Ngai).

Fall 2010, Summer 2011 Visiting Student Researcher

Graduate Group in Biostatistics, University of California, Berkeley, USA.

Project: “Genome annotation and transcriptome analysis in *Saccharomyces cerevisiae* using mRNA-Seq”.

Supervisor: Prof. Sandrine Dudoit.

May – Dec. 2008 Graduate Fellow

Department of Chemical Process Engineering, University of Padova, Italy.

Project: “Analysis and integration of gene expression data for studying muscle plasticity” supervised by Dr. Silvio Bicciato, in collaboration with Dr. Chiara Romualdi.

2010 – present Reviewer

for international journals, including *Annals of Applied Statistics*, *BMC Bioinformatics*, *BMC Genomics*, *Computational Statistics and Data Analysis*, *RNA*, *PLoS ONE*.

Teaching Experience

Fall Semester 2014 Lecturer

Division of Biostatistics, School of Public Health, University of California, Berkeley, USA.

Instructor of the graduate-level class *Computational Statistics with Applications in Biology and Medicine*, cross-listed between the Division of Biostatistics and the Department of Statistics, University of California, Berkeley.

Course syllabus at: <http://www.stat.berkeley.edu/biostat/degrees/courses.htm>

Spring 2012 – Spring 2014 Co-Instructor

CGRL workshop on RNA-Seq data analysis (offered every semester). Computational Genomics Resource Laboratory. University of California, Berkeley.

More info: <http://cgrlucb.wikispaces.com/RNASeqSpring2014>

Summer 2013 Instructor

Specialist course on Statistical Analysis of RNA-Seq data. Doctoral Program in Bioinformatics. Biogen. Ariano Irpino, Italy.

Summer 2012 Co-Instructor

Specialist course on Statistical Inference for RNA-Seq data. Doctoral School of Statistics. University of Padova, Italy.

Primary Instructor: Sandrine Dudoit

Fall 2011 Teaching Assistant

Applied Statistics, Master degree in Healthcare Biology. Faculty of Mathematical, Physical and Natural Sciences, University of Padova, Italy.

Primary Instructor: Prof. Guido Masarotto.

Duties included preparing and grading computer labs.

2006 – 2008 Junior Tutor

at the Faculty of Statistical Sciences, University of Padova.

Awards and Scholarships

April 2013 Visiting scientist fellowship, Department of Statistical Sciences, University of Padova.

January 2012 Research award, Telethon Foundation, Italy.

January 2009 Three-year doctoral scholarship, Department of Statistical Sciences, University of Padova.

Language Skills

Italian: Native Speaker

English: Fluent

Computer Skills

Operating Systems: Unix/Linux, MAC OS X, MS Windows

Programming languages: R, python, shell scripting, SQL

Professional Memberships

American Statistical Association (ASA): Member (since 2012).

International Society for Computational Biology (ISCB): Member (since 2015).

Publications

Refereed Journal Publications

1. **D. Risso**, J. Ngai, T.P. Speed and S. Dudoit (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, **32**, 896-902.
2. T. Ferreira, S.R. Wilson, J. Choi, **D. Risso**, S. Dudoit, T.P. Speed and J. Ngai (2014). Silencing of odorant receptor gene expression by G protein $\beta \gamma$ signaling ensures the expression of one odorant receptor per olfactory sensory neuron. *Neuron*, **81**, 847-859.
3. **D. Risso**, K. Schwartz, G. Sherlock, S. Dudoit (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, **12**, 480. *Highly accessed*.

4. P. Martini, **D. Risso**, G. Sales, C. Romualdi, G. Lanfranchi, S. Cagnin (2011). Statistical Test of Expression Pattern (STEPath): a new strategy to integrate gene expression data with genomic information in individual and meta-analysis studies. *BMC Bioinformatics*, **12**, 92. *Highly accessed*.
5. R. De Bin, **D. Risso** (2011). A novel approach to the clustering of microarray data via nonparametric density estimation. *BMC Bioinformatics*, **12**, 49. *Highly accessed*.
6. **D. Risso**, M.S. Massa, M. Chiogna, C. Romualdi (2009). A modified LOESS normalization applied to microRNA arrays: a comparative evaluation. *Bioinformatics*, **25**(20), 2685-2691.
7. A. Bisognin, A. Coppe, F. Ferrari, **D. Risso**, C. Romualdi, S. Biciato, S. Bortoluzzi (2009). A-MADMAN: annotation-based microarray data meta-analysis tool. *BMC Bioinformatics*, **10**, 201. *Highly accessed*.
8. M. Chiogna, M.S. Massa, **D. Risso**, C. Romualdi (2009). A comparison on effects of normalisations in the detection of differentially expressed genes. *BMC Bioinformatics*, **10**, 61. *Highly accessed*.

Refereed Book Chapters

1. **D. Risso**, J. Ngai, T.P. Speed and S. Dudoit (2014). The role of spike-in standards in the normalization of RNA-seq. In D. Nettleton and S. Datta (Editors) *Statistical Analysis of Next Generation Sequence Data*. Springer, New York. http://link.springer.com/chapter/10.1007/978-3-319-07212-8_9.
2. **D. Risso**, G. Sales, C. Romualdi, M. Chiogna (2013). A hierarchical Bayesian model for RNA-Seq data. In Grigoletto, Lisi, Petrone (Editors) *Complex Models and Computational Methods in Statistics*. Springer-Verlag, Italy. http://link.springer.com/chapter/10.1007/978-88-470-2871-5_17.

Preprints and Technical Reports

- **D. Risso**, K. Schwartz, G. Sherlock, S. Dudoit (2011). GC-Content Normalization for RNA-Seq Data. *Technical Report 291. Division of Biostatistics, University of California, Berkeley*. <http://www.bepress.com/ucbbiostat/paper291/>
- R. De Bin, **D. Risso** (2010). Clustering via nonparametric density estimation: an application to microarray data. *WP 3-2010. Department of Statistical Sciences, University of Padova*.
- L. Peixoto, **D. Risso**, S. G. Poplawski, M. E. Wimmer, T. P. Speed, M. A. Wood, T. Abel (2015). Power gain: how normalization affects reproducibility and biological insight of RNA-seq studies in Neuroscience. *Submitted*.

Software

- **RUVSeq**: Remove Unwanted Variation from RNA-Seq Data. *R/Bioconductor package*.
Role: co-author, maintainer.
Release: <http://www.bioconductor.org/packages/release/bioc/html/RUVSeq.html>
Development: <https://github.com/drisso/RUVSeq>
- **EDASeq**: Exploratory Data Analysis and Normalization for RNA-Seq. *R/Bioconductor package*.
Role: co-author, maintainer.
Release: <http://www.bioconductor.org/packages/release/bioc/html/EDASeq.html>
Development: <https://github.com/drisso/EDASeq>

- **BioinfoTree**: an open source framework for building reproducible bioinformatic workflows.
Role: contributing author.
Project website: <http://bioinfotree.org>

Invited Presentations

- May 18–20, 2015** Bioinformatic Summer Course, Division of Bioinformatics, Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA, USA.
Title: *TBD*
- March 28–31, 2015** The Association of Biomolecular Resource Facilities Annual Meeting (ABRF 2015), St. Louis, MO, USA.
Title: *RNA-seq Normalization and Batch Effect Removal*
- May 13, 2013** Department of Statistics, University of Padova, Italy.
Title: *Can we use spike-in sequences to normalize RNA-Seq data?*
- Jan 13, 2012** Symposium: *Computational Sciences and Bioinformatics*, Padova, Italy.
Title: *From data normalization to regulatory networks.*

Contributed Presentations

- January 12–13, 2015** European Bioconductor Developers' Meeting, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.
Title: *Removing unwanted variation from RNA-seq data.*
- Jul 29 – Aug 2, 2012** Joint Statistical Meetings, San Diego, CA. Section on Statistical Computing, Session 224 – *Statistical and Computational Challenges in Metagenomic Analysis of Next-Generation Sequencing Data.*
Title: *Normalization and Differential Expression in RNA-Seq.*
- Sept 19–21, 2011** 7th Conference on Statistical Computation and Complex Systems (SCo2011), Padova, Italy.
Title: *A hierarchical Bayesian model for RNA-Seq data.*
- July 28, 2011** Bioconductor Developers' Meeting (BioC 2011), Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
Title: *EDASeq: Exploratory Data Analysis and Quality Control for RNA-Seq data*

References

Monica Chiogna

Associate Professor
Department of Statistical Sciences
University of Padova
via C. Battisti 241, 35121 Padova, Italy
Phone: +39 049 827 4183
e-mail: monica@stat.unipd.it

Sandrine Dudoit

Professor
Division of Biostatistics and
Department of Statistics
University of California, Berkeley
101 Haviland Hall, # 7358
Berkeley, CA 94720-7358, U.S.A.
Phone: +1 (510) 643-1108
e-mail: sandrine@stat.berkeley.edu

Terry Speed

Professor
Department of Statistics
University of California, Berkeley
and Walter and Eliza Hall Institute
Melbourne, Victoria, Australia
376 Evans Hall, #3860
Berkeley, CA 94720-3860, U.S.A.
Phone: +1 (510) 642-0613
e-mail: terry@stat.berkeley.edu

John Ngai

Coates Family Professor of Neuroscience
Helen Wills Neuroscience Institute
Director, QB3 Functional Genomics Laboratory
University of California, Berkeley
142 Life Sciences Addition # 3200
Berkeley, CA 94720-3200, U.S.A.
Phone: +1 (510) 642-9885
e-mail: jngai@berkeley.edu

Lin He

Assistant Professor
Department of Molecular and Cell Biology
University of California, Berkeley
330F Li Ka Shing Center, # 3370
Berkeley, CA 94720-3370, U.S.A.
Phone: +1 (510) 642-9338
e-mail: lhe@berkeley.edu

Gavin Sherlock

Associate Professor
Department of Genetics
School of Medicine
Stanford University
300 Pasteur Drive,
Stanford, CA 94305-5120, U.S.A.
Phone: +1 (650) 498-6012
e-mail: gsherloc@stanford.edu