

Nan Xiao

PhD Student
Department of Human Genetics
The University of Chicago

nanx@uchicago.edu
<http://r2s.name>

Education

- 2015 – PhD student (Human Genetics), The University of Chicago, Chicago, IL.
2012 – 2015 Master student (Statistics), Central South University, Changsha, China.
2008 – 2012 Bachelor of Science (Statistics), Central South University, Changsha, China.

Publications

Journal articles

- 2015 **Nan Xiao** and Q.-S. Xu (2015). “Multi-step adaptive elastic-net: reducing false positives in high-dimensional variable selection”. *Journal of Statistical Computation and Simulation*. doi:10.1080/00949655.2015.1016944
- 2015 **Nan Xiao**, D.-S. Cao, M.-F. Zhu, and Q.-S. Xu (2015). “protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequence”. *Bioinformatics*. doi:10.1093/bioinformatics/btv042
- 2014 D.-S. Cao*, **Nan Xiao***, Q.-S. Xu and A. F. Chen (2014). “Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds, and their interactions”, *Bioinformatics*. *Joint first authors. [Package downloaded over 6,000 times.] doi:10.1093/bioinformatics/btu624
- 2015 D.-S. Cao, **Nan Xiao**, Y.-J. Li, W.-B. Zeng, Y.-Z. Liang, Q.-S. Xu, A. F. Chen (2015). “Integrating multiple evidence sources to predict adverse drug reactions based on systems pharmacology model”. *CPT: Pharmacometrics & Systems Pharmacology*. Accepted.
- 2015 J.-B. Wang, D.-S. Cao, M.-F. Zhu, Y.-H. Yun, **Nan Xiao**, Y.-Z. Liang (2015). “In silico evaluation of logD_{7.4} and comparison with other prediction methods”. *Journal of Chemometrics*. doi:10.1002/cem.2718

Book translations

- 2015 Max Kuhn and Kjell Johnson (2015). *Applied Predictive Modeling*. (En-Chi Ma, Yi-Xuan Qiu, **Nan Xiao**, Hui Lin, & Vivian Zhang, Trans.). China Machine Press (Original work published in 2013). In press.
- 2014 Winston Chang (2014). *R Graphics Cookbook*. (**Nan Xiao**, Yi-Shuo Deng, Tai-Yun Wei, & Yi-Xuan Qiu, Trans.). Posts and Telecom Press (Original work published in 2013).
- 2013 Hadley Wickham (2013). *ggplot2: Elegant Graphics for Data Analysis*. (Tai-Yun Wei, Yi-Xuan Qiu, **Nan Xiao**, Tao Gao, & Wei-Cheng Zhu, Trans.). Xi'an Jiaotong University Press (Original work published in 2010).

- 2013 Robert Kabacoff (2013). *R in Action: Data Analysis and Graphics with R*. (Tao Gao, **Nan Xiao**, & Gang Chen, Trans.). Posts and Telecom Press (Original work published in 2011). [Sold 27,500 copies.]

Conference presentations & posters

- 2015 **Nan Xiao** and Tengfei Yin. **liftr: Reproducible Bioinformatics and Statistical Data Analysis with Docker, Rabix, and knitr** (Poster).
BioC 2015, Fred Hutchinson Cancer Research Center, Seattle, WA. July 2015.
- 2015 **liftr & sbgr kickstart** (Workshop with Dan Tenenbaum and Tengfei Yin).
BioC 2015, Fred Hutchinson Cancer Research Center, Seattle, WA. July 2015.
Workshop details and course material available from:
<http://www.bioconductor.org/help/course-materials/2015/BioC2015/>.
- 2014 **Integrated Pipeline for Systems Pharmacology in R/Bioconductor**
The 7th China R Conference, Renmin University of China, Beijing, China. May 2014.
- 2013 **Web Scraping with R**
The 6th China R Conference, Renmin University of China, Beijing, China. May 2013.
- 2009 **Security Issues and Vulnerabilities of the R Environment**
The 2nd China R Conference, Renmin University of China, Beijing & East China Normal University, Shanghai. December 2009.

Software and web applications

R/Bioconductor packages

- liftr**
2015 – R package for dockerizing R Markdown documents with support for Rabix in reproducible research. Project website: <http://liftr.me>.
Availbale from CRAN: <https://cran.r-project.org/web/packages/liftr/>.
- sbgr**
2015 – R Client for Seven Bridges Genomics API.
Availbale from GitHub: <https://github.com/road2stat/sbgr> and Bioconductor:
<http://bioconductor.org/packages/release/bioc/html/sbgr.html>.
- Rcpi**
2013 – 2014 R/Bioconductor package as a comprehensive molecular informatics platform for systems pharmacology and drug discovery. [Downloaded over 6,000 times by February 2015.]
Availbale from Bioconductor: <http://bioconductor.org/packages/release/bioc/html/Rcpi.html>.
- protr**
2012 – 2014 R package generating various numerical representation schemes of protein sequence for bioinformatics and proteochemometrics research.
Available from CRAN: <http://cran.r-project.org/web/packages/protr/>.

enpls

- 2014 R package for ensemble partial least squares regression, a unified framework for feature selection, outlier detection, and ensemble learning.
Available from CRAN: <http://cran.r-project.org/web/packages/enpls/>.

RECA

- 2012 – 2014 R package for relevant component analysis (RCA) in supervised distance metric learning and supervised dimensionality reduction.
Available from CRAN: <http://cran.r-project.org/web/packages/RECA/>.

Web applications

ProtrWeb

- 2013 – 2014 Web application for computing 20 types of protein sequence-derived structural and physicochemical features in bioinformatics. [Over 2,000 users from 40 countries.]
Available at <http://protrweb.scbdd.com>.

TargetNet

- 2014 Web application for predicting the binding of 600+ drug targets for any given molecules. Driven by large-scale machine learning modeling of chemogenomics data.
Available at: <http://targetnet.scbdd.com>.

ImgSVD

- 2014 Web application for image compression via singular value decomposition. Collaborated with Yihui Xie (RStudio Inc.), Yixuan Qiu (Purdue University), and Tong He (Simon Fraser University). Available at <https://yihui.shinyapps.io/imgsvd/>.

Honors and awards

- 2014 Best poster award. *Genes and associated environmental factors affecting skin aging in different ethnic groups*. CAS-MPG Partner Institute for Computational Biology Scientific Exchange Meeting, Shanghai. With Miao Zhu Li.
- 2013 National scholarship for graduate students (5%).
Highest award the government established for graduate students in China.
- 2013 Outstanding graduate student award (5%). Central South University, Changsha, China.
- 2011 1st Prize (top 5%) in Hunan Contest District in China Undergraduate Mathematical Contest in Modeling (CUMCM). With Guan-Chun Liu and Tian-Yu Zhao.
- 2010 Meritorious award (2nd in 1,000+) of Central China Undergraduate Mathematical Contest in Modeling (CCUMCM). With Tao Gao and Chen-Xi Guo.

Internships & professional services

- 2015 **Intern.** Seven Bridges Genomics, Boston, MA. <https://www.sbgenomics.com>
- 2015 – **Reviewer.** *Journal of Statistical Computation and Simulation*
- 2014 **Contributing Editor.** ICSA (*International Chinese Statistical Association*) *Bulletin*

2014 **Organizing Committee Member.** *The 7th China R Conference (Beijing).* <http://china-r.org>
 2013 **Co-Organizer.** *The 6th China R Conference (Beijing)*
 2012 **Volunteer.** *The 5th China R Conference (Beijing)*
 2010 – **Editor.** *Capital of Statistics*, <http://cos.name>. Largest online community on statistics in China.
 2009 **Volunteer.** *The 2nd China R Conference (Beijing, Shanghai)*

Research interests

Methodology Statistical machine learning, theory and methodology:
 • High-dimensional linear models and feature selection
 • Supervised distance metric learning and dimensionality reduction

Applications Statistical machine learning, applications in biomedicine:
 • High-dimensional data integration and network inference
 • Systems genetics and translational bioinformatics

Professional skills

Research
 • Statistical machine learning methods
 • High-dimensional variable selection methods
 • High-performance statistical computing
 • Statistical graphics and data visualization for biological data

Programming
 • R/Bioconductor package development (6 years)
 • Complex web application development with R/Shiny (3 years)
 • Good knowledge of Python, PostgreSQL, T_EX and shell programming (4 years)
 • Experience of Linux server setup, management and application deployment (4 years)

Last revision: August, 2015