

# Nan Xiao

Ph.D. Student  
Department of Human Genetics  
University of Chicago

[nanx@uchicago.edu](mailto:nanx@uchicago.edu)  
[nanx.me](http://nanx.me)

## Education

2015 –	Ph.D. student (Human Genetics), University of Chicago.	Chicago, IL, U.S.
2012 – 2015	Master student (Statistics), Central South University.	Changsha, China
2008 – 2012	Bachelor of Science (Statistics), Central South University.	Changsha, China

## Work experience

2015	<b>Intern.</b> Seven Bridges Genomics, Inc. Advisor: Tengfei Yin.	Cambridge, MA, U.S.
	<ul style="list-style-type: none"><li>• Developed R/Bioconductor package <i>sbgr</i>: R client for SBG Platform / NCI Cancer Genomics Cloud Platform API; the first shipped official language binding for SBG API.</li><li>• Developed R package <i>liftr</i>: dockerizing R Markdown documents with support for Rabix. First enabled OS-level reproducibility for bioinformatics &amp; statistical data analysis.</li><li>• Presented the packages at the BioC 2015 conference (workshop session and poster session), Fred Hutchinson Cancer Research Center, Seattle, WA.</li></ul>	

## Publications

### Journal articles

2015	<b>Nan Xiao</b> and Q.-S. Xu (2015). “Multi-step adaptive elastic-net: reducing false positives in high-dimensional variable selection”. <i>Journal of Statistical Computation and Simulation</i> . doi:10.1080/00949655.2015.1016944	
2015	<b>Nan Xiao</b> , 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”. <i>Bioinformatics</i> . doi:10.1093/bioinformatics/btv042	
2014	D.-S. Cao*, <b>Nan Xiao</b> *, Q.-S. Xu and A. F. Chen (2014). “Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds, and their interactions”, <i>Bioinformatics</i> . *Joint first authors. [Package downloaded over 6,000 times.] doi:10.1093/bioinformatics/btu624	
2015	D.-S. Cao, <b>Nan Xiao</b> , Y.-J. Li, W.-B. Zeng, Y.-Z. Liang, A.-P. Lu, Q.-S. Xu, A. F. Chen (2015). “Integrating multiple evidence sources to predict adverse drug reactions based on systems pharmacology model”. <i>CPT: Pharmacometrics &amp; Systems Pharmacology</i> . doi:10.1002/psp4.12002	
2015	J.-B. Wang, D.-S. Cao, M.-F. Zhu, Y.-H. Yun, <b>Nan Xiao</b> , Y.-Z. Liang (2015). “In silico evaluation of logD <sub>7.4</sub> and comparison with other prediction methods”. <i>Journal of Chemometrics</i> . 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.
- 2014 **Integrated Pipeline for Systems Pharmacology in R/Bioconductor** The 7<sup>th</sup> China R Conference, Renmin University of China, Beijing, China. May 2014.
- 2013 **Web Scraping with R** The 6<sup>th</sup> China R Conference, Renmin University of China, Beijing, China. May 2013.
- 2009 **Security Issues and Vulnerabilities of the R Environment** The 2<sup>nd</sup> China R Conference, Renmin University of China, Beijing & East China Normal University, Shanghai. December 2009.

## Software and web applications

### R/Bioconductor packages

#### **hdnom**

- 2015 – R package building nomograms for high-dimensional Cox models, with support for model validation and calibration. Project website: <http://hdnom.org>. Availbale from CRAN: <https://cran.r-project.org/web/packages/hdnom/>.

#### **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 Bioconductor: <http://bioconductor.org/packages/sbgr/> and  
GitHub: <https://github.com/road2stat/sbgr>.
- 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/Rcpi>.
- 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

- hdnom.io**  
2015 Web application for building penalized Cox models for high-dimensional data with sur-  
vival outcomes. The web app streamlined the process of nomogram building, model val-  
idation, model calibration, and reproducible report generation.  
Available at: <http://hdnom.io>.  
Included by Shiny user showcase: <https://rstudio.com/products/shiny/shiny-user-showcase/>
- 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>.
- ProtrWeb**  
2013 – 2014 Web application for computing 20 types of protein sequence-derived structural and physic-  
ochemical features in bioinformatics. [Over 2,000 users from 40 countries.]  
Available at <http://protrweb.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 *Miaozhu 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 1<sup>st</sup> 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 (2<sup>nd</sup> in 1,000+) of Central China Undergraduate Mathematical Contest in Modeling (CCUMCM). With *Tao Gao* and *Chen-Xi Guo*.

## Professional services

- 2015 – **Reviewer.** *Journal of Statistical Computation and Simulation*
- 2014 **Contributing Editor.** *ICSA (International Chinese Statistical Association) Bulletin*
- 2014 **Organizing Committee Member.** *The 7<sup>th</sup> China R Conference (Beijing)*. <http://china-r.org>
- 2013 **Co-Organizer.** *The 6<sup>th</sup> China R Conference (Beijing)*
- 2012 **Volunteer.** *The 5<sup>th</sup> China R Conference (Beijing)*
- 2010 – **Editor.** *Capital of Statistics*, <http://cos.name>. Largest online community on statistics in China.
- 2009 **Volunteer.** *The 2<sup>nd</sup> China R Conference (Beijing, Shanghai)*

## Research interests

- Methodology
- High-dimensional linear models and variable selection
  - Supervised distance metric learning and dimensionality reduction
- Applications
- High-dimensional data integration and network inference
  - Systems genetics and translational bioinformatics

## Professional skills

- Research
- High-dimensional linear models
  - Statistical machine learning methods
  - High-performance statistical computing
  - Statistical graphics and biological data visualization
- Programming
- R/Bioconductor package development (6 years)
  - Complex web application development with R/Shiny (3 years)
  - Good knowledge of Python, PostgreSQL, T<sub>E</sub>X and shell programming (4 years)
  - Experience of Linux server setup, management and application deployment (4 years)

Last revision: September, 2015