Nan Xiao

me@nanx.me
https://nanx.me

Work experience

2016 –   Genomic Data Scientist. Seven Bridges Genomics, Inc.   *Cambridge, MA, U.S.*

Academic degrees

2008 – 2012   Bachelor of Science (Statistics). Central South University.   *Changsha, China*

Studies

2015 – 2016   Ph.D. Student (Human Genetics). The University of Chicago.   *Chicago, IL, U.S.*
2012 – 2015   Ph.D. Student (Statistics). Central South University.   *Changsha, China*

Publications

Preprints

2016   **Nan Xiao**, Q.-S. Xu, and M.-Z. Li (2016). "hdnom: Building nomograms for penalized Cox models with high-dimensional survival data". *bioRxiv*. doi: 10.1101/065524.

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.

2015   D.-S. Cao*, **Nan Xiao***, Q.-S. Xu and A. F. Chen (2015). "Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds, and their interactions", *Bioinformatics*. *Joint first authors. doi: 10.1093/bioinformatics/btu624.

2015   D.-S. Cao, **Nan Xiao**, 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". *CPT: Pharmacometrics* & *Systems Pharmacology*. doi: 10.1002/psp4.12002.

2017   Y.-W. Lin, **Nan Xiao**, L.-L. Wang, C.-Q. Li, Q.-S. Xu (2017). "Ordered homogeneity pursuit lasso for group variable selection with applications to spectroscopic data". *Chemometrics and Intelligent Laboratory Systems*. doi: 10.1016/j.chemolab.2017.07.004.

2015   J.-B. Wang, D.-S. Cao, M.-F. Zhu, Y.-H. Yun, **Nan Xiao**, Y.-Z. Liang (2015). "*In silico* evaluation of $\log D_{7.4}$ and comparison with other prediction methods". *Journal of Chemometrics*. doi: 10.1002/cem.2718.

2015   L. Shen, D.-S. Cao, Q.-S. Xu, X. Huang, **Nan Xiao**, Y.-Z. Liang (2015). "A novel local manifold-ranking based k-NN for modeling the regression between bioactivity and molecular descriptors". *Chemometrics and Intelligent Laboratory Systems*. doi: 10.1016/j.chemolab.2015.12.005.

Book translations

2016   Max Kuhn and Kjell Johnson (2016). *Applied Predictive Modeling*. (Hui Lin, Yi-Xuan Qiu, En-Chi Ma, **Nan Xiao**, & Vivian Zhang, Trans.). China Machine Press (Original work published in 2013). ISBN: 978-7-1115-3342-9.

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). ISBN: 978-7-115-34227-0.

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). ISBN: 978-7-5605-4969-9.

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). ISBN: 978-7-115-29990-1.

Software and web applications

R/Bioconductor packages

msaenet

2016 –   R package implementing the multi-step adaptive elastic-net (MSAENet) algorithm for feature selection in high-dimensional regressions. Website: https://msaenet.com. Available from CRAN: https://cran.r-project.org/package=msaenet.

ggsci

2016 –   R package offering a collection of ggplot2 color palettes inspired by scientific journals, data visualization libraries, and science fiction TV shows. Website: https://ggsci.net. Available from CRAN: https://cran.r-project.org/package=ggsci.

protr

2012 –   R package generating various numerical representation schemes of protein sequence for bioinformatics and proteochemometrics research. Website: https://nanx.me/protr/. Available from CRAN: https://cran.r-project.org/package=protr.

Rcpi

2013 – R/Bioconductor package with comprehensive integration of bioinformatics and cheminformatics tools for drug discovery. Website: https://nanx.me/Rcpi/.
Available from Bioconductor: https://bioconductor.org/packages/Rcpi.

liftr

2015 – R package for containerizing R Markdown documents. Website: https://liftr.me.
Available from CRAN: https://cran.r-project.org/package=liftr.

OHPL

2017 – R package implementing the ordered homogeneity pursuit lasso (OHPL) algorithm for group variable selection. Website: https://OHPL.io.
Available from CRAN: https://cran.r-project.org/package=OHPL.

hdnom

2015 – R package building nomograms for high-dimensional survival data, with support for model validation, calibration, and comparison. Website: https://hdnom.org.
Available from CRAN: https://cran.r-project.org/package=hdnom.

sevenbridges-r

2016 – Seven Bridges API client, CWL schema, Meta schema, and SDK helper in R.
Website: https://sbg.github.io/sevenbridges-r/
Available from Bioconductor: https://bioconductor.org/packages/sevenbridges.

sbgr

2015 – R Client for Seven Bridges Genomics API (v1). Website: https://nanx.me/sbgr/.
Available from GitHub: https://github.com/road2stat/sbgr.

enpls

2014 – R package for ensemble partial least squares regression, a unified framework for feature selection, outlier detection, and ensemble learning. Website: https://enpls.org.
Available from CRAN: https://cran.r-project.org/package=enpls.

grex

2016 – R package for mapping Ensembl gene IDs in Genotype-Tissue Expression (GTEx) data to Entrez IDs, HGNC symbols, and UniProt IDs. Website: https://nanx.me/grex/.
Available from CRAN: https://cran.r-project.org/package=grex.

RECA

2012 – R package for relevant component analysis (RCA) in supervised distance metric learning and supervised dimensionality reduction. Website: https://nanx.me/RECA/.
Available from CRAN: https://cran.r-project.org/package=RECA.

Web applications

hdnom.io

2015 – Shiny web application for building penalized Cox models for high-dimensional data with survival outcomes. Streamlined the process of nomogram building, model validation, calibration, and comparison. Available at: http://hdnom.io.

ProtrWeb

2013 –   Shiny web application for computing 20 types of protein sequence-derived structural and physicochemical features in bioinformatics. Available at http://protr.org.

TargetNet

2014 –   Shiny web application for drug target identification with large-scale public binding affinities data. Available at http://targetnet.org.

ImgSVD

2014 –   Shiny web application for image compression via singular value decomposition.
Collaboration with *Yihui Xie*, *Yixuan Qiu*, and *Tong He*.
Available at https://nanx.shinyapps.io/imgsvd/.

Signify

2015   Shiny web application for making your (>0.05) p-values sound significant.
Available at https://nanx.shinyapps.io/signify/.

Academic services

2017 –   **Reviewer.** *Chemometrics and Intelligent Laboratory Systems*

2016 –   **Reviewer.** *The R Journal*

2015 –   **Reviewer.** *Journal of Statistical Computation and Simulation*

2017   **Session Chair on Software Tools.** *The 10th China R Conference (Beijing)*

2014   **Organizing Committee Member.** *The 7th China R Conference (Beijing)*

2013   **Co-Organizer.** *The 6th China R Conference (Beijing)*

Conference & workshop presentations

2017   **Persistent Reproducible Reporting with Docker and R**.
Invited talk. The 10th China R Conference, Tsinghua University, China. May 2017.

2017   **Reproducible Dynamic Report Generation with Docker and R**.
Invited talk. DockerCon 2017, Austin, TX. April 2017.

2016   **hdnom.io: High-Dimensional Survival Modeling with Shiny**.
Invited lightning talk. Shiny Developer Conference, Stanford University. January 2016.

2015   **Introduction to Reproducible Research in Bioinformatics**.
Invited talk. CRI Bioinformatics Workshop, Center for Research Informatics,
The University of Chicago. December 3, 2015.

2015   <u>Nan Xiao</u>, Tengfei Yin, and Miaozhu Li. **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.

2013 **Web Scraping with R**
Invited talk. The 6th China R Conference, Renmin University of China. May 2013.

Honors and awards

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, 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*.

Last revision: July, 2017