NAN XIAO

January 2018

Seven Bridges Genomics, Inc.
1 Main St, 5th Floor, Suite 500
Cambridge, MA 02142, USA

Phone: (312) 450-9898
Email: me@nanx.me
Homepage: nanx.me

EMPLOYMENT

Seven Bridges Genomics, Inc. Cambridge, MA, USA.

Genomic Data Scientist, 2016–present.

EDUCATION

Ph.D. Student. Human Genetics, University of Chicago. Chicago, IL, USA, 2015–2016.

Advisor: Prof. Matthew Stephens.

Ph.D. Candidate. Statistics. Central South University. Changsha, China, 2012–present.

Advisor: Prof. Qing-Song Xu

B.S. Statistics. Central South University. Changsha, China, 2008–2012.

RESEARCH EXPERIENCE

Matthew Stephens Lab. Department of Human Genetics, University of Chicago. Summer 2016.

Graduate student rotation.

Yoav Gilad Lab. Department of Human Genetics, University of Chicago. Spring 2016.

Graduate student rotation.

Computational Biology and Drug Design Group. Xiangya School of Medicine, Central South University. 2013 – 2015.

Graduate student research.

PEER-REVIEWED PUBLICATIONS

1. L.-L. Wang, Y.-W. Lin, X.-F. Wang, **Nan Xiao**, Y.-D. Xu, H.-D. Li, Q.-S. Xu (2017). A Selective Review and Comparison for Interval Variable Selection in Spectroscopic Modeling. *Chemometrics and Intelligent Laboratory Systems*.
2. 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* 168: 62–71.
3. L. Shen, D.-S. Cao, Q.-S. Xu, X. Huang, **Nan Xiao**, Y.-Z. Liang (2016). A Novel Local Manifold-Ranking Based k-NN for Modeling the Regression Between Bioactivity and Molecular Descriptors. *Chemometrics and Intelligent Laboratory Systems* 151: 71–77.
4. **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* 85(18): 3755–3765.
5. **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* 31(11): 1857–1859.
6. 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. *Joint first authors. *Bioinformatics* 31(2): 279–281.

7. 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* 4(9): 498–506.

8. J.-B. Wang, D.-S. Cao, M.-F. Zhu, Y.-H. Yun, **Nan Xiao**, Y.-Z. Liang (2015). *In Silico* Evaluation of logD7.4 and Comparison with Other Prediction Methods. *Journal of Chemometrics* 29(7): 389–398.

PREPRINTS

9. **Nan Xiao**, Q.-S. Xu, and M.-Z. Li (2016). hdnom: Building Nomograms for Penalized Cox Models with High-Dimensional Survival Data. *bioRxiv*.

BOOK TRANSLATIONS

1. 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.

2. 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.

3. 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.

4. 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.

R PACKAGES AUTHORED

msaenet

Multi-step adaptive elastic-net algorithm for high-dimensional feature selection.

Integrated by Max Kuhn's *caret* package for streamlined machine learning modeling.

liftr

Containerize R Markdown documents with Docker.

DockerCon 2017 talk invited by Docker, Inc.

ggsci

Scientific journal and sci-fi themed color palettes for ggplot2.

Downloaded 20k+/month. Top 2% of 11,000+ R packages on CRAN.

enpls

Ensemble partial least squares algorithm for feature screening and outlier detection.

OHPL

Ordered homogeneity pursuit lasso algorithm for group feature selection.

hdnom

Benchmarking and visualization toolkit for high-dimensional survival modeling.

protr

Efficient protein sequence feature extraction for machine learning modeling.

Rcpi

Integrative molecular feature extraction for computational drug discovery.

RECA

Relevant component analysis algorithm for supervised distance metric learning.

grex

Gene ID mapping for Genotype-Tissue Expression (GTEx) data.

sevenbridges-r

Seven Bridges API client, CWL schema, metadata schema, and SDK helper in R.

R PACKAGES CONTRIBUTED

mxnet-r

Contributor to the R binding for Amazon-backed deep learning framework MXNet.

ashr

Empirical Bayes approach for large-scale hypothesis testing and FDR estimation.

dml

Distance metric learning toolkit for dimensionality reduction in computer vision.

WEB APPLICATIONS AUTHORED

DockFlow

Bioconductor workflow containerization and orchestration using Docker and liftr.

hdnom.io

Shiny app for benchmarking and visualizing high-dimensional survival models.

Selected as RStudio Shiny user showcase app.

ImgSVD

Shiny app for image compression via singular value decomposition.

Joint work with *Yihui Xie*, *Yixuan Qiu*, and *Tong He*.

TargetNet

Shiny app for drug target identification by learning from binding affinities data.

ProtrWeb

Shiny app for efficient protein sequence feature extraction.

Signify

Shiny app for making your (>0.05) *p*-values sound significant.

TALKS

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

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

2016 hdnom.io: High-Dimensional Survival Modeling with Shiny. Invited talk. RStudio Shiny Developer Conference (Stanford University).

2015 Introduction to Reproducible Research in Bioinformatics. Invited talk. CRI Annual Bioinformatics Workshop, Center for Research Informatics (University of Chicago).

2015 liftr & sbgr kickstart. Invited workshop. Joint with Dan Tenenbaum and Tengfei Yin. BioC 2015 (Fred Hutchinson Cancer Research Center, Seattle, WA).

2014 Integrated Pipeline for Systems Pharmacology in R/Bioconductor. Invited talk. The 7th China R Conference (Renmin University of China).

2013 Web Scraping with R. Invited talk. The 6th China R Conference (Renmin University of China).

POSTERS

2017 DockFlow: Bioconductor Workflow Containerization and Orchestration with liftr. **Nan Xiao**, Tengfei Yin, and Miaozhu Li. BioC 2017 (Dana-Farber Cancer Institute, Boston, MA).

2017 The Deep Connection between Drugs and Side Effects. ISCB Art in Science Competition. ISMB/ECCB 2017 (Prague, Czech Republic).

2015 liftr: Reproducible Bioinformatics and Statistical Data Analysis with Docker, Rabix, and knitr. **Nan Xiao**, Tengfei Yin, and Miaozhu Li. BioC 2015, (Fred Hutchinson Cancer Research Center, Seattle, WA).

JOURNAL REFEREE

Journal of Statistical Computation and Simulation

Chemometrics and Intelligent Laboratory Systems

Genetic Epidemiology

The R Journal

Bioinformatics

PROFESSIONAL ACTIVITIES

Session Chair. The 10th China R Conference (2017)

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

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

Member, New England Statistical Society, 2018–Present.

Member, Institute of Electrical and Electronics Engineers, 2017–Present.

Member, American Statistical Association, 2016–Present.

Member, American Society of Clinical Oncology, 2015–Present.

Member, International Chinese Statistical Association, 2015–Present.

HONORS & AWARDS

National scholarship for graduate students, 2013 (5%). Highest award for graduate students in China.

Outstanding graduate student award, 2013 (5%). Central South University, Changsha, China.

First prize, Hunan Contest District in China Undergraduate Mathematical Contest in Modeling, 2011 (5%).

Meritorious award, Central China Undergraduate Mathematical Contest in Modeling, 2010 (0.1%).

MISCELLANEOUS

Technology Stack: R, Haskell, Shiny, Docker.

Research Interests: Statistical Machine Learning, High-Dimensional Statistics, Reproducible Research.