

Ruowang Li

D207 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104

Email: ruowang@upenn.edu

Google Scholar: <https://scholar.google.com/citations?user=KN62X5gAAAAJ&hl>

CURRENT POSITION

University of Pennsylvania, Philadelphia, PA

Postdoctoral Researcher, Advisors: Jason H. Moore (Primary), Yong Chen

Department of Biostatistics, Epidemiology and Informatics, October 2016-Present

EDUCATION

Pennsylvania State University, University Park, PA

Ph.D., Bioinformatics and Genomics, 2016

Dissertation: "Using Large-scale Genomics Data to Understand the Genetic Basis of Complex Traits"

Committee: Drs. Marylyn Ritchie (Advisor), Ross Hardison, Shaun Mahony, Le Bao

MS, Applied Statistics, 2016

Worcester Polytechnic Institute, Worcester, MA

BS, *High Distinction*, Biology and Biotechnology, 2011

Minor, Computer Science, 2011

RESEARCH INTERESTS

Electronic Health Records, BioBank, Genomics, Polygenic Risk Score, Complex Diseases, Risk Prediction, Data Integration, Biomedical Informatics, Statistical Genetics, Machine Learning

GRANTS AND FELLOWSHIPS

- K99/R00: Pathway to Independence Award, National Library of Medicine, NIH (Impact score of 31 for the 1st submission, resubmitted)
- National Science Foundation Graduate Fellowship (GRFP), NSF, 2013-2016

OTHER AWARDS

- Mack Technology Fellows, Wharton School, University of Pennsylvania, 2019
- Best Informatics Research, Department of Biostatistics, Epidemiology and Informatics Research Day, University of Pennsylvania, 2018
- Best presentation, 2nd place, Biomedical Postdoc Symposium, University of Pennsylvania, 2017
- Best poster award, Translational Bioinformatics Conference, 2017
- Graham Endowed Fellowship, Pennsylvania State University, 2011-2013
- Huck Institute of Life Sciences Fellowship, Pennsylvania State University, 2011-2012
- Excellence in Graduate Recruitment Award, Pennsylvania State University, 2011-2012
- University Graduate Fellowship, Pennsylvania State University, 2011
- Provost's Major Qualifying Project Best Thesis Award, Worcester Polytechnic Institute, 2011

- Chemistry and Biochemistry Fellowship, Worcester Polytechnic Institute, 2007-2011

RESEARCH AND TEACHING EXPERIENCES

Teaching assistant, BMB 498D Human Genomics & Biomedical Informatics, Pennsylvania State University, 2012

Co-instructor, R workshop, Pennsylvania State University, 2014

Research Assistant, Laboratory of Dr. Zhiping Weng, University of Massachusetts Medical School, 2009-2011

- Bioinformatics analysis of gene expression and methylation data in mice

Research Assistant, Laboratory of Dr. William Wu, Worcester Polytechnic Institute, 2008-2009

PROFESSIONAL AND OUTREACH EXPERIENCES

- Organizing committee, Department of Biostatistics, Epidemiology and Informatics Research Day, University of Pennsylvania, 2018-2019
- Co-organizer, Bioinformatics and Genomics Retreat, Pennsylvania State University, 2015
- Peer Mentor, For first-year graduate student, Pennsylvania State University, 2014-2015
- Vice President; Treasurer, GENOMIX club, Pennsylvania State University, 2011-2014
- Co-leader, Ebola outbreak in West Africa presentation. United Nation Association of Centre County. State College, PA, 2014
- Volunteer, Dumpling program for adopted children from China, Worcester Polytechnic Institute, 2008-2011
- Team member, Developed ASSITment teaching methods and problem sets in geometry for 8th grade students, Worcester Polytechnic Institute, 2009-2010
- Reviewer of
EJHG, Genetics, PLOS ONE, BioData Mining, Evolutionary Bioinformatics, Proceedings of the Pacific Symposium on Biocomputing, Frontiers in Genetics, GECCO, ACM-BCB

BOOK

Co-editor, Big Data Analytics in Bioinformatics and Healthcare. ISBN 1466666110, 2014

PUBLICATIONS

Published manuscripts

1. Jiayi Tong, Rui Duan, **Ruowang Li**, Martijn J. Scheuemie, Jason H. Moore and Yong Chen. (accepted) Learning from heterogeneous health systems without sharing patient level data. Proceedings of the Pacific Symposium on Biocomputing.
2. **Ruowang Li**, Rui Duan, Rachel L Kember, Regeneron Genetics Center, Daniel J Rader, Scott M Damrauer, Jason H. Moore, Yong Chen. (2019) A regression framework to uncover pleiotropy in large-scale electronic health record data. Journal of the American Medical Informatics Association. doi: 10.1093/jamia/ocz084

3. **Ruowang Li**, Yong Chen, Jason H. Moore. (2019) Integration of genetic and clinical information to improve imputation of data missing from electronic health records. *Journal of the American Medical Informatics Association*. doi: 10.1093/jamia/ocz041
4. Shefali S Verma, Anastasia Lucas, Xinyuan Zhang, Yogasudha Veturi, Scott Dudek, Binglan Li, **Ruowang Li**, Ryan Urbanowicz, Jason H Moore, Dokyoon Kim, Marylyn D Ritchie. (2018) Collective feature selection to identify crucial epistatic variants. *BioData Mining*. doi: 0.1186/s13040-018-0168-6
5. **Ruowang Li**, Dokyoon Kim, Scott Dudek, Heather Wheeler, Eileen Dolan, Marylyn Ritchie. (2018) Integration of genetic and functional genomics data to uncover chemotherapeutic induced cytotoxicity. *The Pharmacogenomics Journal*. doi: 10.1038/s41397-018-0024-6
6. **Ruowang Li**, Dokyoon Kim, Marylyn D Ritchie. (2017) Methods to analyze big data in pharmacogenomics research. *Pharmacogenomics*. doi: 10.2217/pgs-2016-0152
7. **Ruowang Li**, Scott Dudek, Dokyoon Kim, Molly A Hall, Yuki Bradford, Peggy Peissig, Murray Brilliant, James Linneman, Catherine A McCarty, Le Bao, Marylyn D Ritchie. (2016). Identification of genetic interaction networks via an evolutionary algorithm evolved bayesian network. *BioData Mining*. doi: 10.1186/s13040-016-0094-4
8. Dokyoon Kim, **Ruowang Li**, Anastasia Lucas, Shefali S. Verma, Scott M. Dudek, Marylyn D. Ritchie. (2016). Using knowledge-driven genomic interactions for multi-omics data analysis: meta-dimensional models for predicting clinical outcomes in ovarian carcinoma. *Journal of the American Medical Informatics Association*. doi:10.1093/jamia/ocw165
9. Dokyoon Kim, Anastasia Lucas, Joseph Glessner, Shefali S. Verma, Yuki Bradford, **Ruowang Li**, Alex T. Frase, Hakon Kakonarson, Catherine A Mccarty, Peggy Peissig, Murray Brilliant, Marylyn D. Ritchie. (2016). Biofilter as a functional annotation pipeline for common and rare copy number burden. *Proceedings of the Pacific Symposium on Biocomputing*.
10. Dokyoon Kim, **Ruowang Li**, Scott M. Dudek, Marylyn D. Ritchie (2015). Predicting censored survival data based on the interactions between meta-dimensional omics data in breast cancer. *Journal of Biomedical Informatics*. doi:10.1016/j.jbi.2015.05.019
11. Marylyn D Ritchie, Emily R Holzinger, **Ruowang Li**, Sarah A, Pendergrass, Dokyoon Kim. (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nature Reviews Genetics*, 16(2), 85–97. doi: 10.1038/nrg3868
12. Dokyoon Kim, **Ruowang Li**, Scott M. Dudek, Alex A. Frase, Sarah A. Pendergrass, Marylyn D. Ritchie. (2014). Knowledge-driven genomic interactions: an application in ovarian cancer. *BioData mining*, 7, 20. doi:10.1186/1756-0381-7-20
13. Dokyoon Kim, **Ruowang Li**, Scott M. Dudek., John R. Wallace, Marylyn D. Ritchie (2014). Binning somatic mutations based on biological knowledge for predicting survival: an application in renal cell carcinoma. *Proceedings of the Pacific Symposium on Biocomputing*.
14. **Ruowang Li**, Dokyoon Kim, Scott M. Dudek, Marylyn D. Ritchie. (2014) An Integrated Analysis of Genome-Wide DNA Methylation and Genetic Variants Underlying Etoposide-Induced

Cytotoxicity in European and African Populations. Applications of Evolutionary Computation (Vol. 8602). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-45523-4

15. **Ruowang Li**, Emily R. Holzinger, Scott M. Dudek, Marylyn D. Ritchie. (2014). Evaluation of parameter contribution to neural network size and fitness in ATHENA for genetic analysis. Genetic Programming Theory And Practice. Springer, New York
16. Dokyoon Kim, **Ruowang Li**, Scott M. Dudek, Marylyn D. Ritchie. (2013) ATHENA: Identifying interactions between different levels of genomic data using grammatical evolution neural network. BioData mining, 6(1), 23. doi:10.1186/1756-0381-6-23
17. Ozlem Yildirim, **Ruowang Li**, Jui-Hung Hung, Poshen B. Chen, Xianjun Dong, Ly-Sha Ee, Zhiping Weng, Oliver J. Rando, and Thomas G. Fazio. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell, 147 (2011), pp. 1498–1510. doi: 10.1016/j.cell.2011.11.054
18. Benjamin R. Carone, Lucas Fauquier, Naomi Habib, Jeremy M. Shea, Caroline E. Hart, **Ruowang Li**, Christoph Bock, Chengjian Li, Phillip D. Zamore, Alexander Meissner, Zhiping Weng, Hans A. Hofmann, Nir Friedman, and Oliver J. Rando. (2010). Paternally-induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell 143, 1084–1096. doi: 10.1016/j.cell.2010.12.008
19. Quanqin Dai, Yingnan Wang, Yu Zhang, Xinbi Li, **Ruowang Li**, Bo Zou, JaeTae Seo, Yiding Wang, Manhong Liu, William W. Yu. (2009) Stability Study of PbSe Semiconductor Nanocrystals over Concentration, Size, Atmosphere, and Light Exposure. Langmuir 25 (20), 12320-12324. doi: 10.1021/la9015614

Manuscripts under review

1. **Ruowang Li**, Yong Chen, Marylyn D. Ritchie, Jason H. Moore. Polygenic prediction of disease risks in electronic health record data. (in revision)
2. **Ruowang Li**, Jiayi Tong, Rui Duan, Yong Chen, Jason H. Moore. Evaluation of phenotyping errors on polygenic risk score predictions. (2019) <https://dx.doi.org/10.1101/724534>.

Manuscripts in preparation

1. **Ruowang Li**, Rui Duan, Xinyuan Zhang, Marylyn D. Ritchie, Yong Chen, Jason H. Moore. Sum-Share: Summary statistics via multiple electronic health records for identifying pleiotropy

CONFERENCE PRESENTATIONS

- A computational method to improve missing data imputation in Electronic Health Record. ICSA. Qingdao, China, 2018 (oral)
- A computational method to improve missing data imputation in Electronic Health Record. Biomedical Postdoctoral Research Symposium. Translational Bioinformatics Conference 2017. Long Beach, CA.

- Integration of genetic and functional genomics data to uncover chemotherapeutic induced cytotoxicity. P-STAR Pharmacogenomics Analysis Workshop. State College, PA, 2015 (oral)
- Identification of genetic interaction networks via an evolutionary algorithm evolved Bayesian Network. American Society for Human Genetics 2015. Baltimore, MD.
- Integration of genetic and functional genomics data to uncover chemotherapeutic induced cytotoxicity. Genome of Biology 2015. Cold Spring Harbor, NY.
- Integration of DNA Sequence variation and functional genomics data to infer causal variants underlying chemotherapeutic induced cytotoxicity response. International Genetic Epidemiology Society meeting 2014. Vienna, Austria
- An integrated analysis of genome-wide DNA methylation and genetic variants underlying etoposide-induced cytotoxicity in European and African populations. Granada, Spain, 2014. (oral)
- A genome-wide integrated analysis of chemotherapeutic-induced cytotoxicity in European and African populations. P-STAR Pharmacogenomics Analysis Workshop. Nashville, TN, 2013. (oral)
- Evaluation of parameter contribution to neural network size and fitness in ATHENA for genetic analysis. Genetic Programming Theory And Practice Conference. Ann Arbor, Michigan, 2013 (oral)
- Integration of “-omics” data. Epistasis Discovery in Genetic Epidemiology Conference. Key West, Florida, 2013. (oral)
- Genome-wide analysis identifies heat shock induced gene and chromatin regulatory protein network in *Saccharomyces Cerevisiae*. American Society for Human Genetics 2013. Boston, MA.

SOFTWARE

- The Analysis Tool for Heritable and Environmental Network Associations (ATHENA) <https://ritchielab.org/software/athena-downloads>
Role: Created a new Bayesian Network method (GEBN) to detection genetic interaction networks.

MEMBERSHIPS

- American Society of Human Genetics
- American Statistical Association
- International Genetic Epidemiology Society
- American Medical Informatics Association

REFERENCES

Jason H. Moore, Ph.D.
Edward Rose Professor of Informatics
University of Pennsylvania
D202 Richards, 3700 Hamilton Walk
Philadelphia, PA 19104-6116
215-573-4411
jhmoore@upenn.edu

Yong Chen, Ph.D.
Associate Professor of Biostatistics
University of Pennsylvania
602 Blockley Hall, 423 Guardian Drive
Philadelphia, PA 19104-6116
215-746-8155
ychen123@upenn.edu

Marylyn D. Ritchie, Ph.D.
Professor of Genetics
University of Pennsylvania
A301 Richards Building, 3700 Hamilton Walk
Philadelphia, PA 19104-6116
215-573-2438
marylyn@upenn.edu