LINYING ZHANG

622 West 168th Street PH20, New York, NY 10032
linying.zhang@columbia.edu
HomePage: linyingzhang.com ⋄ GitHub: https://github.com/zhangly811

EDUCATION

Columbia University in the City of New York *August 2018 - Present*
Ph.D. in Biomedical Informatics GPA: 3.96/4.0
Advisors: George Hripcsak and David Blei

Harvard. T.H.Chan School of Public Health *August 2016 - May 2018*
M.S. in Computational Biology and Quantitative Genetics (CBQG) GPA: 3.79/4.0
Advisor: Giovanni Parmigiani
CBQG Program Student Committee Co-chair.
Thesis: *Interactions between multiple myeloma cells and bone marrow stromal cells impact epigenetic profiles of multiple myeloma.*

Boston University *August 2011 - May 2014*
B.A. with Honors (Summa Cum Laude) in Biochemistry and Molecular Biology GPA: 3.74/4.0
Advisor: Ulla Hansen
Awards: Dean's List (6 semesters); Senior Book Award.
Thesis: *Establishing hepatocellular carcinoma cell lines with inducible expression of degradable LSF to investigate LSF regulation in cell cycle.*

RESEARCH INTERESTS

I am interested in developing machine learning models for causal inference from observational data. I develop models that handle high-dimensional data and bias due to unobserved confounding. I am interested in method development as well as application to high-stakes settings where experiments are expensive or even infeasible, such as healthcare.

PUBLICATIONS & PREPRINTS

L. Zhang, Y. Wang, M.J. Schuemie, D.M. Blei, G. Hripcsak. Adjusting for unobserved confounding using large-scale propensity scores. *arXiv preprint*, 2021.

L. Zhang, L.R. Richter, G. Hripcsak. Assessing the impact of race on glomerular filtration rate (GFR) prediction. *medRxiv preprint*, 2021.

L. Zhang, Y. Wang, A. Ostropolets, J. J. Mulgrave, D. M. Blei, and G. Hripcsak. The medical deconfounder: Assessing treatment effects with electronic health records. *Machine Learning for Healthcare Conference*, Ann Arbor, Michigan, 2019.

W. Song, M.J. Kang, **L. Zhang**, W. Jung, J. Song, D.W. Bates, P.C. Dykes. Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *Journal of the American Medical Informatics Association*, 28 (4), 759-765, 2021.

A. Ostropolets, **L. Zhang**, and G. Hripcsak. A scoping review of clinical decision support tools that generate new knowledge to support decision making in real time. *Journal of the American Medical Informatics Association*, 27 (12), 1968-1976, 2020.

A. Ostropolets, R. Chen, **L. Zhang**, and G. Hripcsak. Characterizing physicians information needs related to a gap in knowledge unmet by current evidence. *JAMIA Open*, 3 (2), 281-289, 2020.

O. Gottesman, F. Johansson, J. Meier, J. Dent, D. Lee, S. Srinivasan, **L. Zhang**, Y. Ding, D. Wihl, X. Peng, J. Yao, I. Lage, C. Mosch, L.H. Lehman, M. Komorowski, A. Faisal, L. Celi, D. Sontag, F. Doshi-Velez. Evaluating reinforcement learning algorithms in observational health settings. *arXiv preprint*, 2018.

CONFERENCE PRESENTATIONS & POSTERS

L. Zhang, L.R. Richter, G. Hripcsak. Assessing the impact of race on glomerular filtration rate (GFR) prediction. *Observational Health Data Sciences and Informatics Symposium.* 2021. [Presentation]

L. Zhang, Y. Wang, A. Ostropolets, R. Chen, D.M. Blei, and G. Hripcsak. The Multi-Outcome Medical Deconfounder: Assessing Treatment Effects on Multiple Renal Measures. *American Medical Informatics Association Symposium.* 2020. [Poster]

L. Zhang, Y. Wang, A. Ostropolets, R. Chen, D.M. Blei, and G. Hripcsak. The Multi-Outcome Medical Deconfounder: Assessing Treatment Effects on Multiple Renal Measures. *Observational Health Data Sciences and Informatics Symposium.* 2020. [Poster]

L. Zhang, Y. Wang, A. Ostropolets, J.J. Mulgrave, D.M. Blei, and G. Hripcsak. The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records. *Women in Machine Learning Workshop.* Vancouver, Canada. 2019. [Poster]

L. Zhang, Y. Wang, A. Ostropolets, J.J. Mulgrave, D.M. Blei, and G. Hripcsak. The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records. *Women in Machine Learning Workshop.* Vancouver, Canada. 2019. [Poster]

L. Zhang, Y. Wang, A. Ostropolets, J.J. Mulgrave, D.M. Blei, and G. Hripcsak. The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records. *Machine Learning for Health Workshop.* Vancouver, Canada. 2019. [Poster]

L. Zhang, M.K. Samur, R. Szalat, C.B. Epstein, R. Prabhala, M. Fulciniti, N.C. Munshi*, G. Parmigiani*. Interactions between multiple myeloma cells and bone marrow stromal cells impact epigenetic profiles of multiple myeloma. *Program in Quantitative Genomics Conference.* Boston, MA. 2017. [Poster]

L. Zhang, M.K. Samur, R. Szalat, C.B. Epstein, R. Prabhala, M. Fulciniti, N.C. Munshi*, G. Parmigiani*. Interactions between multiple myeloma cells and bone marrow stromal cells impact epigenetic profiles of multiple myeloma. *Dana-Farber/Harvard Cancer Center Celebration of Junior Investigators.* Boston, USA. 2017. [Poster]

R. Chen, M. Schuemie, M. Suchard, A. Ostropolets, **L. Zhang**, P. Ryan, G. Hripcsak. Evaluation of Large-scale Propensity Score Modeling and Covariate Balance on Potential Unmeasured Confounding in Observational Research. *American Medical Informatics Association Symposium.* 2020. [Poster]

A. Ostropolets, **L. Zhang**, J.J. Mulgrave, G. Hripcsak. Investigating female-male differences in risk factors for myocardial infarction using OHDSI tools. *American Medical Informatics Association Symposium.* 2019. [Poster]

W. Song, **L. Zhang**, E. Gill, J.Z. Liu, A. Wright. Personalized treatment for type 2 diabetes using weighted k-nearest neighbors. *American Medical Informatics Association Symposium.* 2019. [Poster]

R. Szalat, M.K. Samur, C.J. Ott, M. Lawlor, C. Epstein, B.J. Abraham, C.Y. Lin, **L. Zhang**, R. Prabhala, N. Farrell, K. Wes, Y.T. Tai, M. Fulciniti, G. Parmigiani, R.A. Young, K.C. Anderson, and N.C. Munshi. Integrative Oncogenomic Analysis Combining Whole Genome, Transcriptome and Epigenome Identifies Altered Chromatin Accessibility Landscape in Multiple Myeloma. *American Society of Hematology Annual Meeting.* 2018. [Oral presentation]

TALKS

Algorithmic fairness in medicine: A case study in glomerular filtration rate (GFR) prediction.
DBMI Research Seminar at Columbia University. 2021.

Assessing the impact of race on glomerular filtration rate (GFR) prediction.
OHDSI Health Equity Working Group. 2021.

Adjusting for unobserved confounding using large-scale propensity scores.
DBMI Research Seminar at Columbia University. 2020.

The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records.
Causal Inference Panel at AMIA Symposium. 2020.

The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records.
SC-TRM Working Group Meeting at Perelman School of Medicine, University of Pennsylvania. 2020.

PROFESSIONAL SERVICE

Reviewer for International Conference on Machine Learning (ICML) 2020.

Reviewer for Women in Machine Learning (WiML) 2019.

Reviewer for Machine Learning for Health (ML4H) 2019 and 2020.

RESEARCH & WORK EXPERIENCE

Dana-Farber Cancer Institute April 2017 - May 2018
Graduate Researcher (Advisor: Giovanni Parmigiani)

- Investigated the impact of interactions between multiple myeloma (MM) cells and bone marrow stromal cells (BMSC).
- Built a pipeline for ChIP-Seq and RNA-Seq analysis in Linux and R.
- Identified epigenetic changes in both MM and BMSC via differential binding analysis and motif analysis.
- Integrated expression data and epigenetic data to study the impact of epigenetic changes on gene regulation.

Alnylam Pharmaceuticals Inc. June 2014 - June 2016
Research Scientist (Supervisor: Stuart Milstein)

- Supported drug candidate selection by performing high-throughput screening and data analysis.
- Supported Clinical Trial Application filling of ALN-AAT by assay development.
- Investigated the possibility of targeting APP with siRNA for treatment of Alzheimers disease.

Boston University March 2013 - May 2014
Undergraduate Researcher (Advisor: Ulla Hansen)

- Identified the role of transcription factor LSF in cell cycle progression in liver cancer cells.
- Initiated the establishment of an inducible cell line by optimizing transfection condition.
- Designed DNA constructs to achieve instant protein degradation in cell cycle by PCR.

TEACHING EXPERIENCE

Computational Methods. *Columbia University.* Spring 2020.

Computer Applications in Health Care Biomedicine. *Columbia University.* Fall 2019.

Principles of Biostatistics I&II. *Harvard T.H.Chan School of Public Health.* Summer 2017.

RELEVANT COURSES

Statistics
Probability (Biostatistics PhD core courses)
Statistical Inference (Biostatistics PhD core courses)
Methods (Biostatistics PhD core courses)
Longitudinal Analysis
Calculus I&II and Linear Algebra

Computer Science
Machine Learning
Data Science
Deep learning
Foundations of Graphical Models
Reinforcement Learning

INTERESTS

Ballroom dancing: 1st place at BADC 2016 at Columbia University; 3rd place at MIT Open 2017.

Piano: won the highest amateur level certificate from Chinese Musicians Association at age 15.

Scuba diving: Certified Open Water diver since 2015.