CURRICULUM VITAE

Zihuai He

(Updated 3/8/2019)

Quantitative Science Unit
Stanford University School of Medicine
1070 Arastradero Rd, Room 305
Palo Alto, CA 94304

Email: zihuai@stanford.edu
Web: www.zihuai-he.com

Education and Training

Ph.D.	Biostatistics	University of Michigan	2012 - 2016
B.S.	Mathematics and Physics	Tsinghua University	2006 - 2010

Experience

2018 - present	Assistant Professor Department of Neurology & Neurological Sciences Department of Medicine (Biomedical Informatics Research) Stanford University
2016 - 2018	Postdoctoral Research Scientist, Supervisor: Dr. Iuliana Ionita-Laza Department of Biostatistics, Columbia University
2013 - 2016	Research Assistant, Advisors: Dr. Bhramar Mukherjee and Dr. Min Zhang Department of Biostatistics, University of Michigan
2012 - 2013	Graduate Student Instructor Department of Biostatistics, University of Michigan
2011 - 2012	Graduate Student Instructor Department of Statistics and Probability, Michigan State University
2009 - 2010	Research Assistant Tsinghua National Laboratory for Information Science and Technology

Research Interests

Statistical Methodology: statistical genetics and integrative analysis of omics data, high dimensional data analysis, correlated (longitudinal, familial) data analysis, machine learning.

Application Areas: neurological disorders, next generation sequencing data analysis, high-throughput epigenomic functional annotations, gene-environment interactions.

Awards and Honors

2015	Rackham Pre-doctoral Fellowship Award	University of Michigan
2013 - 2015	Rackham Conference Travel Grant	University of Michigan
2013	Best Performance on the Qualifying Exam	University of Michigan

Publications

Peer Reviewed Publications

1. **He, Z.**, Liu, L., Wang, K. and Ionita-Laza, I. (2018). A semi-supervised approach for predicting cell type specific functional consequences of non-coding variation using MPRAs. ***Nature Communications***, 9(1), 5199.
2. Li, M.*, **He, Z.***, Tong, X., Witte, J.S. and Lu, Q. (2018). Detecting Rare Mutations with Heterogeneous Effects Using a Family-Based Genetic Random Field Method. ***Genetics***, genetics-301266.
   **** Joint first author***
3. Backenroth, D., **He, Z.**, Kiryluk, K., Boeva, V., Pethukova, L., Khurana, E., Christiano, A., Buxbaum, J., Ionita-Laza, I. (2018). FUN-LDA: A latent Dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation. ***The American Journal of Human Genetics***, 102(5), 920-942.
4. **He, Z.**, Xu, B., Lee, S., Ionita-Laza, I. (2017). Unified sequence-based association tests allowing for multiple functional annotations, and meta-analysis of noncoding variation in Metabochip data. ***The American Journal of Human Genetics***, 101(3), 340-352.
5. **He, Z.**, Zhang, M., Lee, S., Smith, J.A., Kardia, S.L.R., Diez Roux, A.V. and Mukherjee, B. (2017). Set-based tests for gene-environment interaction in longitudinal studies. ***Journal of the American Statistical Association***, 112(519), 966-978.
6. **He, Z.**, Lee, S., Zhang, M., Smith, J.A., Guo, X., Palmas, W., Kardia, S.L.R., Ionita-Laza, I., and Mukherjee, B. (2017). Rare-variant association tests in longitudinal studies, with an application to the Multi-Ethnic Study of Atherosclerosis (MESA). ***Genetic Epidemiology***, 41(8), 801-810.
7. Zhao, W., Ware, E.B., **He, Z.**, Kardia, S.L.R., Faul, J.D., Smith, J.A. (2017). Social and psychosocial factors modify the effect of genetic variants on body mass index: a gene environment interaction analysis in a longitudinal setting. ***International Journal of Environmental Research and Public Health***, 14(10), 1153.
8. Li, M., Li, J., **He, Z.**, Lu, Q., Witte, J.S., Macleod, S.L., Hobbs, C.A., Cleves, M.A., and the National Birth Defect Prevention Study (2016). Testing allele transmission for a SNP-set with a family-based generalized genetic random field method. ***Genetic Epidemiology***, 40(4), 341-351.
9. Wen, Y., **He, Z.**, Li, M., and Lu, Q. (2016). Risk prediction modeling of sequencing data using a forward random field method. ***Scientific Reports***, 6.
10. Mukherjee, B., Chen, Y., Ko, Y., **He, Z.**, Lee, S., Zhang, M., and Park, S.K. (2016). Statistical strategies for modeling gene-environment interactions in longitudinal cohort studies. ***Statistical Approaches to Gene-Environment Interactions for Complex Phenotypes***, Cambridge, MA: MIT Press, 2016.
11. **He, Z.**, Zhang, M., Lee, S., Smith, J.A., Guo, X., Palmas, W., Kardia, S.L.R., Diez Roux, A.V., and Mukherjee, B. (2015). Set-based tests for genetic association in longitudinal studies. ***Biometrics***, 71(3), 606-615.
12. Li, M.*, **He, Z.***, Schaid D.J., Cleves M.A., Nick T.G., and Lu Q. (2015). A powerful non-parametric statistical framework for family-based association analyses. ***Genetics***, 200 (1), 69-78.
    **** Joint first author***
13. **He, Z.**, Payne, E.K., Mukherjee, B., Lee, S., Smith, J.A., Ware, E.B., Sánchez, B.N., Seeman, T.E., Kardia, S.L.R., and Diez Roux, A.V. (2015). Association between stress response genes and features of diurnal cortisol curves in the Multi-Ethnic Study of Atherosclerosis. ***PLOS ONE***, e0126637.

14. Li, M.*, **He, Z.***, Zhang, M., Zhan, X., Wei, C., Elston, R.C., and Lu, Q. (2014). A generalized genetic random field method for the genetic association analysis of sequencing data. ***Genetic Epidemiology***, 38 (3), 242-253.
    **** Joint first author***
15. **He, Z.****, Zhang, M.**, Zhan, X., and Lu, Q. (2014). Modeling and testing for joint association using a genetic random field model. ***Biometrics***, 70 (3), 471-479.
    ***** Joint corresponding author***
16. Wei C., Li, M., **He, Z.**, Vsevolozhskaya O., Schaid, D.J., and Lu, Q. (2014). A weighted U-statistic for genetic association analyses of sequencing data. ***Genetic Epidemiology***, 38 (8), 699-708.

Preprints

17. **He, Z.**, Xu, B., Buxbaum, J. and Ionita-Laza, I.. GenoScan: a genome-wide scan statistic framework for whole-genome sequence data analysis with applications to data from the Simons Simplex Collection.
18. Zhang, M., Wang, S., **He, Z.** and Mukherjee, B.. Interaction analysis under misspecification of main effects: some common mistakes and simple solutions.

Presentations

Invited presentations

1. Novarits Pharmaceuticals, Division of Advanced Exploratory Analytics, East Hanover, NJ. "A semi-supervised approach for predicting organism level and cell/tissue specific functional consequences of noncoding variants", September 2017.
2. University of Michigan, Department of Statistics, Ann Arbor, MI. "Set-based tests for gene-environment interaction", November 2015.
3. The 2014 WNAR/IMS conference, Honolulu, HI. "Set-based tests for gene-environment interaction in longitudinal studies", June 2014.
4. University of Michigan, Department of Statistics, Ann Arbor, MI. "Modeling and testing for joint association using a genetic random field model and its extensions", October 2013.
5. Michigan State University, Department of Statistics and Probability, East Lansing, MI. "Modeling and testing for joint association using a genetic random field model and its extensions", October 2013.
6. The Kidney Disease Research (KDR) meeting, Ann Arbor, MI. "Random field modeling of association and correlation", November 2013.

Contributed presentations

7. The American Society of Human Genetics (ASHG) Annual Meeting, San Diego, CA. Poster presentation, "A semi-supervised approach for predicting cell type specific functional consequences of non-coding variation using MPRAs", October 2018.
8. The American Society of Human Genetics (ASHG) Annual Meeting, Orlando, FL. Poster presentation, "A semi-supervised approach for predicting organismal level and cell type/tissue specific functional consequences of noncoding variation", October 2017.
9. The 2017 IGES conference, Cambridge, United Kingdom. Selected platform presentation, "Unified sequence-based association tests allowing for multiple functional annotations, and meta-analysis of noncoding variation in Metabochip data", September 2017.

10. The 2015 ENAR/IMS conference, Miami, FL. “Set-based tests for genetic association in longitudinal studies”, March 2015.
11. The Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS), Ann Arbor, MI. “Set-based tests for genetic association in longitudinal studies”, March 2015.
12. The 2014 ENAR/IMS conference, Baltimore, MD. “Random field modeling of genetic association in longitudinal studies”, March 2014.
13. The Center for Statistical Genetics (CSG) meeting, Ann Arbor, MI. “Testing gene-environment interaction in longitudinal studies”, October 2014.
14. The Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS), Ann Arbor, MI. Poster presentation, “Modeling and testing for joint association using a genetic random field model”, March 2013.
15. The American Society of Human Genetics (ASHG) Annual Meeting, Boston, MA. Poster presentation, “Joint association analysis of family-based sequencing data using a family-genetic random field model”, October 2013.
16. The Center for Statistical Genetics (CSG) meeting, Ann Arbor, MI. “Modeling and testing for joint association using a genetic random field model and its extensions”, November 2013.

Software Developed

GenoNet: web interface for predicting cell type specific functional consequences of non-coding variation across 127 tissues/cell types.

- http://www.funlda.com/genonet

FSTpackage: R-package for unified sequence-based association tests allowing for multiple functional annotation scores

- Individual gene / gene-set / genome-wide scan for coding and noncoding variants using summary statistics (available for meta-analysis) and functional genomics data

LGEWIS: R-package for longitudinal gene-environment-wide interaction studies

- Single variant/set-based score tests for genetic association/gene-environment interaction in studies with time varying outcomes and environmental exposures

LGRF: R-package for the Longitudinal Genetic Random Field model (LGRF)

- Test the association between a longitudinally measured quantitative outcome and a set of genetic variants in a gene/region

Academic Service

- Peer Review: *American Journal of Human Genetics*, *Journal of the American Statistical Association*, *Statistics in Medicine*, *Genetic Epidemiology*, and *PLOS Computational Biology*
- Guest Associate Editor: *PLOS genetics*

Other Professional Activities

- Bachelor’s Thesis on Statistical Analysis of Microarray Data. Tsinghua University, Beijing, China, 2010.
  Thesis Supervisor: Shao Li, M.D.
- National Graduate Students Summer School in Applied Mathematics and Statistics, Tsinghua University, Beijing, China, 2009.