Seunggeun (Shawn) Lee, Ph.D.

Graduate School of Data Science
Seoul National University
Bldg 942, Research Park, 1 Gwanak-ro
Seoul, Republic of Korea

Email: lee7801@snu.ac.kr
Web: www.leelabsg.org

Research Interests

Statistical Genetics; Genomics; Health Data Science; Precision Health; Statistics; Machine Learning/AI; Biobank; EHR

Employment

Professor (with tenure), Graduate School of Data Science, Seoul National University 2023 - present

Associate Professor (with tenure), Graduate School of Data Science, Seoul National University 2020 - 2023

Associate Professor (with tenure), Department of Biostatistics, University of Michigan 2018 - 2020

Assistant Professor, Department of Biostatistics, University of Michigan 2013 - 2018

John G Searle Assistant Professor, Department of Biostatistics, University of Michigan 2017 - 2018

Education and Training

Research Fellow, Harvard School of Public Health 2010 - 2013

- Mentor: Dr. Xihong Lin

Ph.D., Biostatistics, University of North Carolina at Chapel Hill 2010

- Dissertation: "Principal Component Analysis for High Dimensional Data, Application for Genomewide-association studies"
- Advisors: Dr. Fei Zou and Dr. Fred A. Wright

B.S., Biology and Statistics, Seoul National University, Korea 2005

Experience

Research Assistant, University of North Carolina, Chapel Hill, NC 2006 - 2010

Software Engineer, ECO Inc., Korea 2000 - 2004

Honor and Awards

- James E. Grizzle Distinguished Alumni Award, UNC Biostatistics. 2017
- NIH Pathway to Independence Award (K99/R00). 2012
- Delta Omega Honorary Society. 2011
- IMS Laha Travel Award. 2010
- Special Commendation for Highest Score in Qualifying Exam, Biostatistics. 2007

- Greenberg Fellowship. 2005
- Merit Assistantship. 2005-2006
- Worak institution Scholarship. 1999,2000
- Merit based Tuition Scholarship. 1996,1997

Publications

Peer reviewed journal publications

1. Wenjian Bi, Wei Zhou, Peipei Zhang, Yaoyao Sun, Weihua Yue, **Seunggeun Lee** (2023) Scalable mixed model methods for set-based association studies on large-scale categorical data analysis and its application to exome-sequencing data in UK Biobank ***American Journal of Human Genetics***, 110 (5), 762-773

2. Jangho Kim*, Junhyeong Lee*, Kisung Nam, **Seunggeun Lee** (2023) Genome-wide study on 72,298 Korean individuals in Korean biobank data for 76 traits identifies hundreds of novel loci, ***Scientific Reports***, 13 (1), 1526

3. Adrian I. Campos, Shinichi Namba, ..., **Seunggeun Lee** ..., Loic Yengo (2023) Boosting the power of genome-wide association studies within and across ancestries by using polygenic scores, ***Nature Genetics***, In press

4. Wei Zhou, Wenjian Bi, Zhangchen Zhao, Kushal K Dey, Karthik A Jagadeesh, Konrad J Karczewski, Mark J Daly, Benjamin M Neale, **Seunggeun Lee** (2022) SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests ***Nature Genetics***, 54 (10), 1466-1469

5. Kisung Nam*, Jangho Kim*, **Seunggeun Lee** (2022) Genome-wide study on 72,298 Korean individuals in Korean biobank data for 76 traits identifies hundreds of novel loci, ***Cell Genomics***, 2 (10), 100189

6. Zhangchen Zhao*, Lars G Fritsche, Jennifer A Smith, Bhramar Mukherjee, **Seunggeun Lee** (2022) The Construction of Multi-ethnic Polygenic Risk Score using Transfer Learning ***American Journal of Human Genetics***, 109 (11), 1998-2008

7. Yongwen Zhuang*, Brooke N Wolford, Kisung Nam, Wenjian Bi, Wei Zhou, Cristen J Willer, Bhramar Mukherjee, **Seunggeun Lee** (2022) Incorporating family disease history and controlling case–control imbalance for population-based genetic association studies, ***Bioinformatics***, 38, 4337-4343

8. Wei Zhou, et al (2022) Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease, ***Cell Genomics***, 2 (10), 100192

9. Rounak Dey, Wei Zhou, et al (2022) Efficient and accurate frailty model approach for genome-wide survival association analysis in large-scale biobanks ***Nature Communications***, 13 (1), 1-13

10. Masahiro Kanai, et al (2022) Meta-analysis fine-mapping is often miscalibrated at single-variant resolution, ***Cell Genomics***, 2 (10), 100210

11. Vivek Sriram, Manu Shivakumar, Sang-Hyuk Jung, Yonghyun Nam, Lisa Bang, Anurag Verma, **Seunggeun Lee**, Eun Kyung Choe, Dokyoon Kim (2022) NETMAGE: A human disease phenotype map generator for the network-based visualization of phenome-wide association study results ***GigaScience***, 11

12. Jiacong Du, Lauren J Beesley, **Seunggeun Lee**, Xiang Zhou, Walter Dempsey, Bhramar Mukherjee (2022) Optimal diagnostic test allocation strategy during the COVID-19 pandemic and beyond ***Statistics in medicine***, 41, 310-327

13. Jingchunzi Shi*, Michael Boehnke, **Seunggeun Lee** (2021) Trans-ethnic meta-analysis of rare variants in sequencing association studies, ***Biostatistics***, 22, 706-722

14. Zhangchen Zhao*, Stephen Salerno, Xu Shi, **Seunggeun Lee**, Bhramar Mukherjee, Lars G Fritsche (2021) Understanding the Patterns of Serological Testing for COVID-19 Pre-and Post-Vaccination Rollout in Michigan, ***Journal of clinical medicine***, 10, 4341

15. Lars G Fritsche, Ying Ma, Daiwei Zhang, Maxwell Salvatore, **Seunggeun Lee**, Xiang Zhou, Bhramar Mukherjee (2021) On cross-ancestry cancer polygenic risk scores, ***PLoS genetics***, 17, e1009670

16. Seungjin Ryu, Jeehae Han, Trina M Norden-Krichmar, Quanwei Zhang, **Seunggeun Lee**, Zhengdong Zhang, Gil Atzmon, Laura J Niedernhofer, Paul D Robbins, Nir Barzilai, Nicholas J Schork, Yousin Suh (2021) Genetic signature of human longevity in PKC and NF-kB signaling, ***Aging cell***, 20, e13362

17. Wenjian Bi*, **Seunggeun Lee** (2021) Scalable and Robust Regression Methods for Phenome-Wide Association Analysis on Large-Scale Biobank Data, ***Frontiers in Genetics***, 12, 682638

18. Wenjian Bi*, Wei Zhou, Rounak Dey, Bhramar Mukherjee, Joshua N Sampson, **Seunggeun Lee** (2021) Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes, ***American Journal of Human Genetics***, 108, 5-6

19. Yatong Li, **Seunggeun Lee** (2021) Novel score test to increase power in association test by integrating external controls, ***Genetic Epidemiology***, 45, 293-304

20. Diptavo Dutta*, Peter VandeHaar, Lars G Fritsche, Sebastian Zöllner, Michael Boehnke, Laura J Scott, **Seunggeun Lee** (2021) A powerful subset-based method identifies gene set associations and improves interpretation in UK Biobank, ***American Journal of Human Genetics***, 108, 669-681

21. Bi., W.*, Fritsche, L.G., Mukherjee, B., Kim, S., **Lee, S.** (2020) A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank, ***American Journal of Human Genetics***, 207, 222-233

22. Zhou, W.*#, Zhao, Z.*#, Nielsen, J.B, Fritsche, L.G., LeFaive, J., Gagliano Taliun, S.A., Bi, W., Gabrielsen, M.E., Daly, M.J., Neale, B.M., Hveem, K., Abecasis, G.R., Willer, C.J., **Lee, S.** (2020) Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts, ***Nature Genetics***,52, 634–639

[#] *Equal contribution*

23. Zhao, Z.*, Bi, W.*, Zhou, W., VandeHaar, P., Fritsche, L.G., **Lee, S.** (2020) UK-Biobank Whole Exome Sequence Binary Phenome Analysis with Robust Region-based Rare Variant Test, ***American Journal of Human Genetics***, 106, 3-12

24. Shi, J.*, Boehnke, M., **Lee, S.** (2020) Trans-ethnic meta-analysis of rare variants in sequencing association studies, ***Biostatistics***, in press, doi:https://doi.org/10.1093/biostatistics/kxz061

25. Zhang, D.*, Dey, R., **Lee, S.** (2020), Fast and robust ancestry prediction using principal component analysis, ***Bioinformatics***, in press, preprint: https://doi.org/10.1101/713172

26. Bi, W.*, Zhao, Z.*, Dey, R., Fritsche, L.G., Mukherjee, B., **Lee, S.** (2019) A Novel Method for Genome-Wide Scale Phenome-Wide GxE Analysis and its Application to UK Biobank, ***American Journal of Human Genetics***, 105, 1182-1192

27. Dutta, D.*, Gagliano Taliun, S.A., Weinstock, J.S., Zawistowski, M., Sidore, C. , Fritsche, L.G., Cucca, F., Schlessinger, D., Abecasis, G.R., Brummett, C.M., **Lee, S.** (2019) Meta-MultiSKAT: Multiple phenotype meta-analysis for region-based association test, ***Genetic Epidemiology***, in press

28. Dey, R.*, Nielsen, J.B., Fritsche, L.G., Zhou, W., Zhu, H., Willer, C.J., Lee, S. (2019) Robust meta-analysis of biobank-based genome-wide association studies with unbalanced binary phenotypes, ***Genetic Epidemiology***, doi:https://doi.org/10.1002/gepi.22197

29. Dey, R.*, and Lee, S. (2019) Asymptotic properties of principal component analysis and shrinkage-bias adjustment under the generalized spiked population model, ***Journal of Multivariate Analysis***, doi:https://doi.org/10.1016/j.jmva.2019.02.007

30. Chen, H, Huffman, J.E., [... including **Lee, S.** , ...], Lin, X (2019) Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies, ***American Journal of Human Genetics***, 104, 260-274

31. Graham, S.E., Nielsen, J.B., [... including **Lee, S.** , ...], Willer, C.J. (2019) Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis, ***Nature Communication***, in press

32. Zhou, W.*, Nielsen, J.B., Fritsche, L.G., Dey, R., Elvestad, M.B., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., Bastarache, L.A., Wei, W-Q, Denny, J.C., Lin, M., Hveem, K., Kang, H.M., Abecasis, G.R., Willer, C.J.[#], **Lee, S.**[#] (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies, ***Nature Genetics***, 50, 1335-1341
[#] *Equal contribution*

33. Dutta, D.*, Scott, L., Boehnke, M., **Lee, S.** (2018) Multi-SKAT: General framework to test multiple phenotype associations of rare variants, ***Genetic Epidemiology***, 43(1), 4-23

34. Yu, Y., Xia, Lu, Lee, S., Zhou, X., Stringham, H.M., Boehnke, M., Mukherjee, B. (2018) Subset-Based Analysis using Gene-Environment Interactions for Discovery of Genetic As-

sociations across Multiple Studies or Phenotypes, ***Human Heredity***, in press

35. Nielsen, J.B., Thorolfsdottir, R.B., [... including **Lee, S.** , ...], Willer, C.J. (2018) Biobank-driven genomic discovery yields new insight into atrial fibrillation biology, ***Nature Genetics***, 50, 1234-1239

36. Dey, R.*, Schmidt, E.M., Abecasis, G.R., **Lee, S.** (2017) A fast and accurate algorithm to test for binary phenotypes and its application to PheWAS, ***American Journal of Human Genetics***, 101, 37-49

37. **Lee, S.**, Sun, W., Wright, F.A., Zou, F. (2017) An improved and explicit surrogate variable analysis procedure by coefficient adjustment, ***Biometrika***, 104, 303-316

38. **Lee, S.**, Kim, S., Fuchsberger, C. (2017) Improving power for rare variant tests by integrating external controls, ***Genetic Epidemiology***, 41, 610-619

39. He, Z.*, Zhang, M., **Lee, S.**, Smith, J.A., Kardia, S., Diez Roux, A.V., Mukherjee, B. (2017) Set-based tests for gene-environment interaction in longitudinal studies, ***Journal of the American Statistical Association***, 101, 340-352

40. He, Z.*, **Lee, S.**, Zhang, M., Smith, J.A., Guo, X., Palmas, W., Kardia, S., Iuliana, I., Mukherjee, B. (2017) Rare-variant association tests in longitudinal studies, with an application to the Multi-Ethnic Study of Atherosclerosis (MESA), ***Genetic Epidemiology***, 41, 801-810

41. He, Z., Xu, B., **Lee, S.**, Ionita-Laza, I. (2017) Unified sequence-based association tests allowing for multiple functional annotation scores, and applications to meta-analysis of noncoding variation in Metabochip data, ***American Journal of Human Genetics***, 101, 340-352

42. Schmidt, E.M., Fritsche, L.G., **Lee, S.**, VandeHaar, P., Brummett, C.M., Kheterpal, S., Abecasis, G.R. (2017) The Michigan Genomics Initiative: A Model Framework for Genetic Discovery Using Patient Electronic Health Records, ***Genetic Epidemiology***, 41, 676-677

43. Liu, G., Mukherjee, B., **Lee, S.**, Lee, A.W, Wu, A.H., Bandera, E.V., Jensen, A., Rossing, M.A, Moysich, K.B., Chang-Claude, J., Doherty, J., Gentry-Maharaj, A., Kiemeney, L., Modugno, F., Massuger, L., Goode, E.L., Fridley, B., Terry, K.L., Cramer, D.W., Anton-Culver, H., Ziogas, A., Tyrer, J.P., Schildkraut, J.M., Kjaer, S.K., Webb, P.M., Ness, R.B., Pike, M.C., Menon, U., Berchuck, A., Pharoah, P.D., Risch, H., Pearce, C.L, the Ovarian Cancer Association Consortium (2017) Robust Tests for Additive Gene-Environment Interaction in Case-Control Studies Using Gene-Environment Independence, ***American Journal of Epidemiology***, 187, 366-377

44. Gauderman, W.Z., Mukherjee, B., Aschard, H., Hsu, L., Lewinger, J.P., Patel, C.J., Witte, J.S., Amos, C., Tai, C., Conti, D., Torgerson, D.G., **Lee, S.**, Chatterjee, N. (2017) Update on the State of the Science for Analytical Methods for Gene-Environment Interactions (GxE), ***American Journal of Epidemiology***, 186, 762-770

45. Kim, D., Basile, A., Bang, L., **Lee, S.**, Ritchie, M., Saykin, A., Nho, K. (2017) Knowledge-

driven binning approach for rare variant association analysis: Application to neuroimaging biomarkers in Alzheimer's disease, ***BMC Medical Informatics and Decision Making***, DOI:10.1186/s12911-017-0454-0

46. X Wang, Z Zhang, N Morris, T Cai, **S Lee**, C Wang, TW Yu, CA Walsh, X Lin. (2016) Rare variant association test in family based sequencing studies, ***Briefings in Bioinformatics***, bbw083.

47. Shi, J.* and **Lee, S.** (2016) A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis, ***Biometrics***, 72(3):945-54

48. **Lee, S.**, Fuchsberger, C., Kim, S., Scott, L. (2016) An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies, ***Biostatistics***, 17, 1-15.

49. Lin, X, **Lee, S.**, Wu, M.C., Wang, C., Chen, H., Li, Z., Lin, X. (2016) Test for rare variants by environment interactions in sequencing association studies, ***Biometrics***, 72(1):156-64.

50. Mensah-Ablorh, A., Lindstrom, S., Haiman, C.A., Henderson, B.E., Marchand, L.L, **Lee, S.**, Stram, D.O., Eliassen, H., Price, A., Kraft, P. (2016) Meta-analysis of rare variant association tests in multi-ethnic populations, ***Genetic Epidemiology***, 40, 57-65.

51. Ware, E.B., Smith, J.A., Mukherjee, B., **Lee, S.**, Kardia, S., Diez Roux, A.V. (2016) Applying novel methods for assessing individual- and neighborhood-level social and psychosocial environment interactions with genetic factors in the prediction of depressive symptoms in the Multi-Ethnic Study of Atherosclerosis, ***Behavior Genetics***, 46, 89-99.

52. Ma, C.*, Boehnke, M., **Lee, S.** and the GoT2D investigators (2015) Evaluating the calibration and power of three gene-based association tests for the X chromosome, ***Genetic Epidemiology***, 39, 499-508.

53. Urrutia, E., **Lee, S.**, Maity, A., Zhao, N., Shen, J., Li, Y., Wu, M.C. (2015) Rare variant testing across methods and thresholds using the multi-kernel sequence kernel association test (MK-SKAT), ***Statistics and Its Interfaces***, 8, 495-505.

54. He, Z.*, Zhang, M., **Lee, S.**, Smith, JA, Guo, X., Palmas, W., Kardia, S., Rouz, A, Mukherjee, B. (2015) Set-based tests for genetic association in longitudinal studies, ***Biometrics***, 71, 606-15

55. He, Z.*, Payne, EK, Mukherjee, B., **Lee, S.**, Smith, JA, Ware, EB, Sanchez, BN, Seeman, TE, Kardia, S, Roux, A. (2015) Association between stress response genes and features of diurnal cortisol curves in the Multi-Ethnic Study of Atherosclerosis: a new multi-phenotype approach for gene-based association tests, ***Plos One***, 10(5):e0126637

56. **Lee, S.** Abecasis, G., Boehnke, M., Lin, X. (2014) Rare-Variant Association Analysis: Study Designs and Statistical Tests, ***American Journal of Human Genetics***, 95, 5-23. [cite $>$ 200]

57. **Lee, S.**, Zou, F. and Wright, F.A. (2014).Convergence of sample eigenvalues, eigenvectors

and PC scores for ultra-high dimensional data, ***Biometrika***, 101, 484-490.

58. Mukherjee, B., Chen, Y-H, Ko, Y-A, He, Z., **Lee, S.**, Zhang, M., Park, SK. (2014) Statistical strategies for modeling gene-environment interactions in longitudinal cohort studies, ***Statistical Approaches to Gene-Environment Interactions for Complex Phenotypes***, Cambridge, MA: MIT Press, in press

59. Wang, X., **Lee, S.**, Zhu, X., Redline, S., Lin, X. (2013). GEE-based SNP Set Association Test for Continuous and Discrete Traits in Family Based Association Studies, ***Genetic Epidemiology***, 37:778-786.

60. **Lee, S.**, Teslovich, T., Boehnke, M., Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies, ***American Journal of Human Genetics***, 93, 42-53. [cite > 50]

61. Ionita-Laza, I.[#], **Lee, S.**[#], Makarov, V., Buxbaum, J. Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants, ***American Journal of Human Genetics***, 92, 841-853. [cite > 100]
[#] *Joint first author*

62. Lin, X., **Lee, S.**, Christiani, D. and Lin, X. (2013). Test for interactions between a Gene/SNP-set and Environment/Treatment in generalized linear models, ***Biostatistics***, doi: 10.1093/biostatistics/kxt006.

63. Wu, M.C., Maity, A., **Lee, S.**, Simmons, E.M., Molldrem, J.J. and Armistead, P.M. (2013) Kernel machine SNP-set testing under multiple candidate kernels. ***Genetic Epidemiology***, 37, 267-275.

64. **Lee, S.**, Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J. Nickerson, D.A., NHLBI GO Exome Sequencing ProjectESP Lung Project Team, Christiani, D.C., Wurfel, M.M. and Lin, X. (2012). Optimal unified approach for rare variant association testing with application to small sample case-control whole-exome sequencing studies. ***American Journal of Human Genetics***, 91, 224-237. [cite > 200]

65. **Lee, S.**, Wu, M. and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. ***Biostatistics***, 13, 762-775. [cite > 200]

66. Ionita-Laza, I., **Lee, S.**, Makarov, V., Buxbaum, J. Lin, X. (2012). Family-based association tests for sequence data, and comparisons with population-based association tests. ***European Journal of Human Genetics***, doi: 10.1038/ejhg.2012.308.

67. Barnett, I., **Lee, S.** and Lin, X. (2012). Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. ***Genetic Epidemiology***, 37, 142-151.

68. Wu, M.[#], **Lee, S.**[#], Cai, T., Li, Y., Boehnke, M., Lin, X. (2011). Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test (SKAT). ***American Journal of Human Genetics***, 89, 82-93. [cite > 1000]
[#] *Joint first author*

69. **Lee, S.**, Wright, F.A. and Zou, F. (2011). Control of population stratification by correlation-selected principal components. ***Biometrics***, 67, 967-974.

70. **Collaborative Cross Consortium** (2011). The Genome Architecture of the Collaborative Cross Mouse Genetic Reference Population, ***Genetics***, 190, 389-401.

71. Sun, W., **Lee, S.**, Zhabotynsky, V., Zou, F., Wright, F.A., Crowley, J.J., Yun, Z. , Buus, R., Miller, D., Wang, J., McMillan, L., de Villena, F. and Sullivan, P.F. (2011). Transcriptome atlases of mouse brain reveals differential expression across brain regions and genetic backgrouds. ***G3: Genes, Genomes, Genetics***, 2, 203-211.

72. Wright, F.A., Strug, L.J., Doshi, V.K., Commander, C.W., Blackman, S.M., Sun, L., Berthiaume, Y., Cutler, D., Cojocaru, A., Collaco, J.M, Corey, M., Dorfman, R., Goddard, K., Green, D., Kent Jr, J.W., Ethan, Lange, **Lee, S.**, Li, W., Luo, J., Mayhew, G., Naughton, K., Pace, R., Paré, P., Rommens, J. Sanfrod, A., Stonebraker, J.R., Sun, W., Taylor, C., Vanscoy, L.L., Zou, F., Blangero, J., Zielenski, J., O'Neal, W.K., Drumm, M.L., Durie1, P.R., Knowles, M.R., Cutting, G.R. (2011). Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. ***Nature Genetics***, 43, 539-546.

73. Li, W., Sun, L., Corey, M., Zou, F., **Lee, S.**, Cojocaru, A.L., Taylor, C., Blackman, S.M., Stephenson, A., Sandford, A.J., Dorfman, R., Drumm, M.L., Cutting, G.R., Knowles, M.R., Durie, P., Wright F.A., and Strug L.J. (2011). Understanding the population structure of North American patients with cystic fibrosis. ***Clinical Genetics***, 79, 136-46.

74. **Lee, S.**, Zou, F. and Wright, F.A. (2010). Convergence and prediction of principal component scores in high dimensional settings. ***Annals of Statistics***, 38, 3605-3629. [cite > 50]

75. Zou, F., **Lee, S.** and Wright, F.A. (2010). Control of population stratification using correlated SNPs by shrinkage principal components. ***Human Heredity***, 70, 9-22.

76. Zou, F., Huang, H., **Lee, S.**, and Hoeschele, I. (2010). Nonparametric bayesian variable selection with applications to multiple quantitative trait loci mapping with epistasis and gene-environment interaction, ***Genetics***, 186, 385-394.

77. **Lee, S.**, Sullivan, P.F., Zou, F. and Wright, F.A. (2008). Comment on a simple and improved correction for population stratification. ***American Journal of Human Genetics***, 82, 524-526.

78. Sullivan, P.F., Lin, D., Tzeng, J-Y, E van den Oord, Perkins, D., Stroup, T.S., Wagner, M., **Lee, S.**, Wright, F.A., Zou, F., Liu, W., Downing, A.M., Lieberman, J. and Close S.L. (2008). Genomewide association for schizophrenia in the CATIE study: results of stage 1. ***Molecular Psychiatry***, 13, 570-84.

79. Jeong, J., Choi, M. , Cho, Y., **Lee, S.**, Oh, J., Park, J., Cho, Y., Lee, I., Kim, S., Han, S., Choi, K. and Chung, I. (2008). Chronic gastrointestinal symptoms and quality of life in the Korean population. ***World Journal of Gastroenterology***, 14, 6388-6394.

* doctoral student/research assistant under my supervision

Grants

Current

BP+ 2020H1D3A2A03100666 (NRF of Korea), "*Genomic data science methods and system for precision health*", (PI: Lee), 07/01/20-12/31/23
Role: Principal Investigator

R01-HG-008773-01 (NIH), "*Statistical and computational methods for rare variant association analysis*", (PI: Mukherjee), 05/17/16-04/30/21
Role: Subcontract Principal Investigator.

R01-LM-012535-01 (NIH), "*Integrating Neuroimaging, Multi-omics, and Clinical Data in Complex Disease*", (Subcontract PI: Lee), 05/17/16-04/30/21
Role: Subcontract Principal Investigator.

Past

R01-HL142023-01 (NIH), "*Integrative analysis to uncover biology of blood lipids and coronary heart disease*", (PI: Lee, Willer and Zhou), 05/01/18-04/30/20
Role: Principal Investigator (Contact PI)

R00-HL-113164-03 (NIH), "*Statistical methods for rare variant effects in sequencing association studies*", (PI: Lee), 09/01/13-06/30/17
Role: Principal Investigator.

K99-HL113164 (NIH),"*Statistical methods for rare variant effects in sequencing association studies*", (PI: Lee), 06/2012 - 08/2013
Role: Principal Investigator.

3U01DK082345-08S1 (NIH),"*University of Michigan MAPP research network discovery site*", (PI: Clauw)
Role: Co-investigator

U01 HL137182-01 (NIH), "*Scalable and Translational Analysis Tools on the Cloud for Deep Integrative Omics Data*", (PI: Kang), 04/15/17-03/31/20
Role: Co-investigator

5216271-5500000816/1-U01-DK-105554-01 (NIH), "*Proposal for the AMP T2D-GENES Data Coordination Center and Web Portal*", (PI: Boehnke), 05/01/15-05/31/17
Role: Co-investigator

R01-HL117626 (NIH), "*Studies of Rare Genetic Variation in the Isolated Population of Sardinia*", (PI: Abecasis), 01/15/13-12/31/17
Role: Co-investigator

Paper/Poster Presentations

Invited Presentations

1. "Biobank data and analysis methods", KOBIO registry, 2022

2. "Biobank data and analysis methods", Department of Statistics Seminar, Sookmyung Women's University, 2022

3. "Biobank data and analysis methods", Department of Statistics Seminar, Chung-Ang University, 2022

4. "The Construction of Multi-ethnic Polygenic Risk Score using Transfer Learning", Korean Statistical Society Meeting, 2022

5. "Biobank data and analysis methods", SAIHST, 2022

6. "Biobank data and analysis methods", Yonsei University College of Medicine, 2022

7. "Biobank data and analysis methods", SNU School of Dentistry, 2022

8. "Set-based rare variant association tests for biobank scale data", Genomics at Columbia (Columbia University), 2022

9. "Set-based rare variant association tests for biobank scale sequencing data sets", Korean Statistical Society Meeting, 2021

10. "Public DBs and New Analytic Tools, Potentials for Future Research", Korean Society of Cardiology, 2021

11. "Dimension reduction in high-dimensional data analysis and Two-stage PLS", EcoSTAT, 2021

12. "Biobank data analysis, with consideration on strategies for privacy protection", KOSHIS meeting, 2020

13. "Make Biobank Data in action", KoNECT International Conference, 2020

14. "Scaleable and accurate rare variant test for big biobank data analysis", IBC, 2020

15. "Scalable and accurate association analysis of big biobank data", JSM, July, 2019

16. "Integrating external controls to association test', ENAR , Mar, 2019

17. "Scalable and accurate analysis of big genetics and biomedical data", Seoul National University, Jan, 2019

18. "Scalable and accurate association analysis of Biobank data", PQG Conference, Nov, 2018

19. "Statistical methods for Biobank data analysis", JSM, Aug, 2018

20. "Statistical methods and tools for analyzing 100,000 samples", ENAR, May, 2018

21. "PCA in the high-dimensional genetics data", Department of Biostatistics Seminar, University of North Carolina, April, 2017

22. "Improving power for rare variant tests by integrating external controls", ICSA Conference, December, 2016

23. "Statistical methods for large scale genetics data analysis", Joint Conference of Korean FDA and Biostatistical society, December, 2016

24. "Asymptotic properties of PCA and shrinkage-bias adjustment in high-dimensional data", Department of Statistics Seminar, Michigan State University, September, 2016

25. "Improving power for rare variant tests by integrating external controls", IISA Conference, August, 2016

26. "Improving power for rare variant tests by integrating external controls", JSM, August, 2016

27. "Statistical Methods For Rare Variant Association Analysis", Center for Systems Genomics, Penn State University, May, 2015

28. "Convergence of principal component scores in high dimensional setting", Department of Statistics Seminar, Younsei University, Korea, July, 2014

29. "Rare Variant Analysis in Sequencing-based Association Studies", Department of Biostatistics Seminar, Washington University in St.Louis, December, 2013

30. "Rare Variant Analysis in Sequencing-based Association Studies", Department of Biostatistics Seminar, University of Minnesota, October, 2013

31. "Rare Variant Analysis in Sequencing-based Association Studies", Department of Biostatistics Seminar, Columbia University, October, 2013

32. "Rare Variants Analysis in Sequencing Studies", ICSA, Boston, MA, June 2012.

33. "Rare Variant Association Analysis for Sequencing Data with the Kernel Association Tests", Department of Biostatistics Seminar, University of Michigan, March, 2012

34. "Statistical methods for high dimensional genomic data", Department of Biostatistics Seminar, University of Pittsburgh, March, 2012

35. "Rare Variant Association Analysis for Sequencing Data with the Kernel Association Tests", Department of Biostatistics Seminar, FHCRC, Feburary, 2012

36. "Statistical methods for high dimensional genomic data", Department of Biostatistics Seminar, University of Pittsburgh, March, 2012

37. "Statistical methods for high dimensional genomic data", Department of Biostatistics Seminar, NCSU, January, 2012

38. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", PQG Short Course, HSPH, Boston, MA, December 2011.

39. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", PMAGE Seminar Series, HSPH, Boston, MA, September 2011.

40. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", New England Statistics Symposium, Storrs, CT, April 2011.

41. "Principal Component Analysis in High Dimensional Data: Application to Genome-wide Association Studies", High dimensional data seminar series. HSPH, MA, October 2010.

Other Presentations

1. "General framework for meta-analysis of rare variants in sequencing association studies", ENAR, Orlando, FL, March 2013.

2. "General framework for meta-analysis of rare variants in sequencing association studies", PQC Conference, Boston, MA, November 2012.

3. "General framework for meta-analysis of rare variants in sequencing association studies", ASHG meeting, San Francisco, CA, November 2012.

4. "Convergence and prediction of principal component scores in high dimensional and ultra high dimensional settings", JSM, Miami, FL, August 2011.

5. "Convergence and prediction of principal component scores in high dimensional settings", ENAR, New Orleans, LA, March 2010.

6. "Principal component score prediction for related and unrelated GWA samples", ASHG meeting, Honolulu, HI, October 2009.

7. "Control of Population Stratification by Correlation-selected principal components", ASHG meeting, Philadelphia, PA, November 2008.

8. "Control of population stratification using correlated SNPs by shrinkage Principal components", Gain Analysis Workshop III, Philadelphia, PA, November 2008.

9. "Preliminary result for association study", Carolina Center for Exploratory Genetic Analysis meeting, Chapel Hill, NC, November 2006.

Short course

1. "Rare variant association analysis", STOM: Genomics workshop, Seoul, Korea, December 2016.

2. "Analysis of Genetic Association Studies Using Sequencing Data and Related Topics",

STOM: Genomics workshop, Seoul, Korea, July 2014.

3. "Analysis of Genetic Association Studies Using Sequencing Data and Related Topics", ICSA-KISS meeting , Portland, OR, June 2014.

Software Developed

SKAT

- R-package for SNP-set (Sequence) Kernel Association Test (SKAT)
- Downloaded > 170000

MetaSKAT

- R-package for meta-analysis burden test, SKAT and SKAT-O
- Downloaded > 80000

SAIGE

- R-package for logistic mixed effect model with large sample sizes

POLMM

- R-package for GWAS of proportional odds model for categorical phenotypes

iECAT

- R-package for integrating external controls to association tests

SPAtest

- R-package for fast saddle point approximation based score test

dSVA

- R-package for fast saddle point approximation based score test

TransMeta

- R-package for trans-ethnic meta-analysis

EigenCorr

- R-package for selecting principal components for adjusting population stratification.

Academic Service

- Peer Review: *Nature*, *Nature Genetics*, *Nature Methods*, *Nature Communications*, *Cell Genomics*, *American Journal of Human Genetics*, *JASA*, *Biometrics*, *Biostatistics*, *Genetics*, *PLOS Genetics*, *PLOS One*, *Genetic Epidemiology*, *BMC bioinformatics*, *European Journal of Human Genetics*, *Nature Communication*, *Heredity* and *Statistics in Medicine*
- Member, Program Committee, 2016 *ENAR* Spring Meetings
- Organizer, *JSM* 2016 Invited Session
- Organizer, *ICSA* 2012 Invited Session
- Co-organizer, *HSPH PQG short courses and tutorials*, 2011-2012

Teaching Experience

- BIOSTAT 830: Special topic course, Genetic Associations and Interactions 2014
  - This is a special topic course on genetic associations and interactions
- BIOSTAT 651: Applied Statistics II: Extensions for Linear Regression 2015-2017
  - This is a required course for the first year BIOSTAT students.
- BIOSTAT 666: Statistical Models and Numerical Methods in Human Genetics 2018-2019
  - This course covers statistical methods for human genetics data analysis.

Postdoctroal Researcher

- Jack Flanagan, Graduate School of Data Science, Seoul National University
  - Jack is working on methods for WGS

PhD Students, Dissertation Committee Chair

- Kisung Nam, Graduate School of Data Science, Seoul National University
- Na Yeon Kim, Graduate School of Data Science, Seoul National University
- Seokho Jeong, Graduate School of Data Science, Seoul National University
- Jeonghyun Moon, Graduate School of Data Science, Seoul National University
- Junhyeong Lee, Graduate School of Data Science, Seoul National University
- Wonyoung Jang, Graduate School of Data Science, Seoul National University
- Chanhee Lee, Graduate School of Data Science, Seoul National University
- Hohyun Kim, Graduate School of Data Science, Seoul National University
- Changhoon Kang, Graduate School of Data Science, Seoul National University

Former PhD students and postdocs

- Minjung Kho (Postdoc), Visiting Faculty, Seoul National University
- Wenjian Bi (Postdoc), Assistant Professor, Peking University
- Yongwen Zhuang, Graduated 2022, Illumina
- Yatong Li, Graduated 2022, Google
- Zhangchen Zhao, Graduated 2020, Eli Lilly
- Jingchunzi Shi, Graduated 2018, 23 and Me
- Rounak Dey, Graduated 2018, Research fellow at HSPH
- Wei Zhou, Graduated 2018, Research fellow at MGH/Broad
- Diptavo Dutta, Graduated 2019, Research fellow at John's Hopkins University

Dissertation Committee Member

- Clement Ma, Department of Biostatistics, University of Michigan, 2014

- Zihuai He, Department of Biostatistics, University of Michigan, 2016
- Brooke Wolford, Department of Computational Medicine and Bioinformatics, University of Michigan, 2018 -

Professional Affiliations

- American Statistical Association
- Institute of Mathematical Statistics
- ENAR : Eastern North American Region, International Biometric Society
- American Society of Human Genetics