Seunggeun (Shawn) Lee, Ph.D.

Department of Biostatistics
University of Michigan
1415 Washington Heights
Ann Arbor, MI 48109

Phone: 734-764-7884
Email: leeshawn@umich.edu
Web: www-personal.umich.edu/∼leeshawn/

Research Interests

High dimensional data analysis, statistical genetics and genomics, next generation sequencing data analysis, Random Matrix Theory, and population genetics and Coalescent Theory

Postdoctoral Training

Research Fellow, Harvard School of Public Health, MA 2010 - 2013
Mentor: Xihong Lin, PhD

Education

Ph.D., Biostatistics, University of North Carolina, Chapel Hill, NC May, 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 March, 2005

Experience

Research Assistant, University of North Carolina, Chapel Hill, NC 2006 - 2010
Supervisor: Fei Zou, PhD

Software Engineer, ECO Inc., Korea 2000 - 2004

Honor and Awards

- 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

Papers and Publications

Peer reviewed publications (Methodological)

1. **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.
2. 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.
   ∗ *Joint first author*
3. 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.
4. 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.
5. 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.
6. 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.
7. **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.
8. **Lee, S.**, Wu, M. and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. ***Biostatistics***, 13, 762-775.
9. 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.
   ∗ *Joint first author*
10. **Lee, S.**, Wright, F.A. and Zou, F. (2011). Control of population stratification by correlation-selected principal components. ***Biometrics***, 67, 967-974.
11. **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.
    [An earlier version won 2010 IMS Laha Travel Award]
12. 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.
13. 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.
14. **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.

Peer reviewed publications (Collaborative)

15. **Collaborative Cross Consortium** (2011). The Genome Architecture of the Collaborative Cross Mouse Genetic Reference Population, ***Genetics***, 190, 389-401.
16. 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.
17. 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., Par, P., Rommens, J. Sanfrod, A., Stonebraker, J.R., Sun, W., Taylor, C., Vanscoy, L.L., Zou, F., Blangero, J., Zielenski, J., ONeal, 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.
18. 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.
19. 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.
20. 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.

Submitted Papers

21. **Lee, S.**, Zou, F. and Wright, F.A. (2011). Convergence of sample eigenvalues, eigenvectors and PC scores for ultra-high dimensional data, *Biometrika*, minor revision.

Working in Progress

22. **Lee, S.**, Zou, F. and Wright, F.A (2011). A direct approach to surrogate variable analysis for high dimensional data. *in preparation.*

Grant

Principal Investigator, NIH Pathway to Independence Award (K99/R00). "Statistical methods for rare variant effects in sequencing association studies"

Conference 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. "Rare Variants Analysis in Sequencing Studies"', ICSA, Boston, MA, June 2012.
5. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", PQG Short Course, HSPH, Boston, MA, December 2011.
6. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", PQG Short Course, HSPH, Boston, MA, December 2011.
7. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", PMAGE Seminar Series, HSPH, Boston, MA, September 2011.
8. "Convergence and prediction of principal component scores in high dimensional and ultra high dimensional settings", JSM, Miami, FL, August 2011.
9. "Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test", New England Statistics Symposium, Storrs, CT, April 2011.
10. "Principal Component Analysis in High Dimensional Data: Application to Genome-wide Association Studies", High dimensional data seminar series. HSPH, MA, October 2010.
11. "Convergence and prediction of principal component scores in high dimensional settings", ENAR, New Orleans, LA, March 2010.
12. "Principal component score prediction for related and unrelated GWA samples", ASHG meeting, Honolulu, HI, October 2009.
13. "Control of Population Stratification by Correlation-selected principal components", ASHG meeting, Philadelphia, PA, November 2008.
14. "Control of population stratification using correlated SNPs by shrinkage Principal components", Gain Analysis Workshop III, Philadelphia, PA, November 2008.
15. "Preliminary result for association study", Carolina Center for Exploratory Genetic Analysis meeting, Chapel Hill, NC, November 2006.

Teaching Experience

Teaching Assistant, UNC-CH

- *BIOS 663:* Intermediate Linear Model, spring 2009
  - prepared and taught weekly recitation sessions
- *BIOS 761:* Advanced Probability and Statistical Inference II, spring 2008
- *BIOS 663:* Intermediate Linear Model, spring 2008
  - prepared and taught weekly recitation sessions
- *BIOS 600:* Principal of Statistical Inference, fall 2005

Teaching Volunteer, South Korea 1996 - 1999

- Taught Mathematics and English to young students from low income families

Software Developed

MetaSKAT

- R-package for meta-analysis burden test, SKAT and SKAT-O
- Web: http://www.hsph.harvard.edu/skat/metaskat

SKAT

- R-package for SNP-set (Sequence) Kernel Association Test (SKAT)
- Web: http://www.hsph.harvard.edu/skat/

EigenCorr

- R-package for selecting principal components for adjusting population stratification.
- Web: http://www-personal.umich.edu/∼leeshawn/EigenCorr/

Computing Skills

- Computer Language: Proficient with C, C^{++} and Unix Shell Script; some knowledge of Perl, PHP, Fortran and SQL
- Statistical Applications: R, SAS, Matlab
- Other Applications: Genome-wide data analysis tools (Plink, Eigensoft), LaTeX, Microsoft office software, common Linux tools.

Academic Service

- Peer Review: *American Journal of Human Genetics*, *Biostatistics*, *Genetics*, *PLOS One*, *Genetic Epidemiology*, *BMC bioinformatics*, *European Journal of Human Genetics*, and *Statistics in Medicine*
- Organizer of *ICSA* 2012 Invited Session
- Co-organizer of *PQG short courses and tutorials*, HSPH 2010-2011

Professional Affiliations

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