

Curriculum Vitae

Lei Sun

(Last updated August 21, 2009)

Division of Biostatistics	6th Floor, Health Sciences Building
Dalla Lana School of Public Health	155 College Street
Faculty of Medicine	Toronto, ON M5T 3M7, Canada
Department of Statistics	P: (416) 978-7519 F: (416) 978-8299
University of Toronto	sun@utstat.toronto.edu

Degrees

2001/08	Ph.D.	Statistics, University of Chicago, USA
1995/07	B.S.	Mathematics, Fudan University, Shanghai, China
1995/07	Diploma	Economics, Fudan University, Shanghai, China

Academic Positions

2007/07-present	Associate Professor	Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto
2005/07-present	Adjunct Professor	Department of Statistics, University of Toronto
2001/09-2007/07	Research Scientist	Program in Genetics and Genomic Biology, Hospital for Sick Children Research Institute, Toronto
2004/01-2004/04	Adjunct Professor	Department of Statistics and Actuarial Science, University of Waterloo (Graduate Teaching)
2001/09-2007/06	Assistant Professor	Department of Public Health Sciences, Faculty of Medicine, University of Toronto
1998-2001	Lecturer	Department of Statistics, University of Chicago
1996-2001	Research Assistant	Department of Statistics, University of Chicago
1996-2001	Teaching Assistant	Department of Statistics, University of Chicago

Interruptions/Delays

Maternity Leave, 30 weeks from April to November 2004

Maternity Leave, 30 weeks from January to August 2007

Research Grants

2008-2013	NSERC	\$14,000 per annum
2007-2012	CIHR (Co-PIs: Shelley Bull, Radu Craiu)	\$108,606 per annum
2006-2009	NIH (PI: Andrew Paterson, 5 co-investigators)	US \$4,172,781 in total
2003-2006	CIHR (Co-PI: Shelley Bull)	\$87,000 per annum
2002-2007	NSERC	\$14,500 per annum
2001-2002	UT Connaught Start-up Grant	\$10,000

Associate Editorship

2007/07-present Statistical Applications in Genetics and Molecular Biology

Awards

- 2006 Dean's Award, Faculty of Medicine's 5% Merit Pool for academic achievement in research, teaching and service, University of Toronto
- 2004 Dean's Award, Faculty of Medicine's 5% Merit Pool for academic achievement in research, teaching and service, University of Toronto
- 1996-2001 University of Chicago Fellowship
- 1996-1997 Paul Meier Fellowship, Department of Statistics, University of Chicago

Research Contributions (Trainees)

Working Manuscripts

1. Dorfman R, Sandford A, Taylor, Lin F, **Sun L**, Par PD, Berthiaume, Corey M, Durie P, Zielenski J. Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis. (Under Review at Respiratory Research)
2. Paterson AD, Waggott D, Boright AP, Hosseini SM, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary PA, Lachin JM, Canty AJ, **Sun L**, Bull SB, and the DCCT/EDIC Research Group. A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both HbA1c and glucose. (Under revision at Diabetes)
3. Mireia L, **Sun L**, Stafford JE, Bull SB. Using evidence for population stratification bias in combined individual- and family-level genetic association analyses of quantitative traits. (Under review at Genetic Epidemiology)

4. Xu L, Craiu RV, **Sun L**. Bayesian methods to overcome the winner's curse in genetic studies. (Under review at Biostatistics)
5. **Sun L**, Dimitromanolakis A, Faye L, Paterson AD, Waggott D, the DCCT/EDIC Research Group, Bull SB. BRsquared: a practical solution to the winner's curse in genome-wide scans.
6. Faye L, **Sun L**, Dimitromanolakis A, Bull SB. Reducing selection bias: efficiency and robustness of parametric & non-parametric approaches.
7. **Sun L**, Dimitromanolakis A, Paterson AD. PREST-plus: efficient detection of pedigree errors and cryptic relatedness with applications to Framingham and HapMap3 data.
8. Yoo YJ, **Sun L**, Bull SB. Gene-based genetic association analysis by multiple regression tests.

Refereed Journal Publications

1. Paterson AD, Lopes-Virella MF, Waggott D, Boright AP, Hosseini SM, Carter RE, Shen E, Mirea L, Bharaj B, **Sun L**, Bull SB, and the DCCT/EDIC Research Group. Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-Selectin. (accepted by *Arteriosclerosis, Thrombosis, and Vascular Biology*).
2. Dorfman R, Li W, **Sun L**, Lin F, Wang Y, Sandford S, Par PD, McKay K, Kayserova H, Macek M, Bal J, Sands D, Tiddens H, Castro S, Sontag M, Accurso FJ, Blackman S, Cutting GR, Tsui LC, Corey M, Durie P, Zielenski J, Strug L. Modifier gene study of Meconium Ileus in Cystic Fibrosis: statistical considerations and gene mapping results. (accepted by *Human Genetics*)
3. Yoo YJ, Bull SB, Paterson AD, Waggott D, DCCT/EDIC Research Group, **Sun L** (2009). Were genome-wide linkage studies a waste of time? Exploiting candidate regions within genome-wide association studies. (in press at *Genetic Epidemiology*).
4. Yoo YJ, Pinnaduwege D, Waggott D, Bull SB, **Sun L** (2009). Association analyses of genome-wide SNP data of NARAC and FHS utilizing genome-wide linkage results. (Accepted by *BMC Proceedings*).
5. Asimit J, Yoo YJ, Waggott D, **Sun L**, Bull SB (2009). Region-based analysis in GWA of FHS blood lipid phenotypes. (Accepted by *BMC Proceedings*).
6. Craiu RV, **Sun L** (2008). Choosing the lesser evil: trade-off between false discovery rate and non-discovery rate. *Statistica Sinica* 18:861-879.

7. Lee SSF, **Sun L**, Kustra R, Bull SB (2008). EM-random forest and new measures of variable importance for multi-Locus quantitative trait linkage analysis. *Bioinformatics* 24:1603-1610.
8. Dorfman R, Sandford A, Taylor C, Huang B, Frangolias D, Wang Y, Sang R, Pereira L, **Sun L**, Berthiaume Y, Tsui LC, Pare PD, Durie P, Corey M, Zielenski J (2008). Complex two-gene modulation of lung disease severity in children with cystic fibrosis. *Journal of Clinical Investigation* 118:1040-1049.
9. Al-Kateb H, Boright AP, Xie X, Mirea L, Sutradhar R, Mowjoodi A, Bharaj B, Liu M, Buckska JM, Arends VL, Steffes MW, Cleary PA, Sun W, Lachin JM, Thorner PS, Ho M, McKnight AJ, Maxwell PA, Savage DA, Kidd KK, Kidd JR, Speed WC, Orchard TJ, Miller RG, **Sun L**, Bull SB, Paterson AD and the DCCT/EDIC research group (2008). Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy. *Diabetes* 57:218-228.
10. Al-Kateb H, Mirea L, Xie X, **Sun L**, Liu M, Chen H, Bull SB, Boright AP, Paterson AD, The DCCT/EDIC Research Group (2007). Multiple variants in Vascular Endothelial Growth Factor (VEGF) are risk factors for time to severe retinopathy in type 1 diabetes: The DCCT/EDIC genetics studies. *Diabetes* 56:2161-2168.
11. Huang B, Rangreg J, Paterson AD, **Sun L** (2007). The multiplicity problem in linkage analysis of gene expression data - the power of differentiating *cis*- and *trans*-acting regulators. *BMC Proceedings* 1:S142.
12. Greenwood C, Rangrej J, **Sun L** (2007). Optimal selection of markers for validation from genome-wide association studies. *Genetic Epidemiology* 31:396-407.
13. Wu LY, **Sun L**, Bull SB (2006). Locus-specific heritability estimation via the bootstrap in linkage scans for quantitative trait loci. *Human Heredity* 62:84-96.
14. **Sun L**, Craiu RV, Paterson AD and Bull SB (2006). Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genetic Epidemiology* 30:519-530.
15. Wu LY, Lee SSF, Shi HS, **Sun L**, Bull SB (2005). Resampling methods to reduce the selection bias in genetic effect estimation in genome-wide scans. *Genetic Analysis Workshop 14: Microsatellite and single-nucleotide polymorphism. BMC Genetics* 6:S24.
16. Biernacka J, **Sun L**, Bull SB (2005). Tests for the presence of two linked disease susceptibility genes. *Genetic Epidemiology* 29:389-401.
17. **Sun L**, Bull SB (2005). Reduction of selection bias in genome-wide genetic studies by resampling. *Genetic Epidemiology* 28:352-367.

18. Biernacka J, **Sun L**, Bull SB (2005). Simultaneous localization of two linked disease susceptibility genes. *Genetic Epidemiology* 28:33-47.
19. Paterson A, **Sun L**, Liu XQ (2003). Transmission ratio distortion in families from the Framingham Heart Study. *Genetic Analysis Workshop 13: Analysis of longitudinal family data for complex diseases and related risk factors*. *BMC Genetics* 4:S48.
20. Strug L, **Sun L**, Corey M (2003). The Genetics of Cross-Sectional and Longitudinal BMI. *Genetic Analysis Workshop 13: Analysis of longitudinal family data for complex diseases and related risk factors*. *BMC Genetics* 4:S14.
21. **Sun L**, Wilder K, McPeck MS (2002). Enhanced pedigree error detection. *Human Heredity* 54:99-110.
22. **Sun L**, Cox NJ, McPeck MS (2002). A statistical method for identification of polymorphisms that explain a linkage result. *American Journal of Human Genetics* 70:399-411.
23. **Sun L**, Abney M, McPeck MS (2001). Detection of misspecified relationships in inbred and outbred pedigrees. *Genetic Analysis Workshop 12: Analysis of complex genetic traits: Applications to asthma and simulated data*. *Genetic Epidemiology* 21:S36-S41.
24. McPeck MS, **Sun L** (2000). Statistical tests for detection of misspecified relationships by use of genome-screen data. *American Journal of Human Genetics* 66:1076-1094.

Refereed Conference Abstracts

1. Dimitromanolakis A, Paterson AD, **Sun L** (2009). Accurate IBD inference identifies cryptic relatedness in 9 HapMap populations. Abstract presented at *the annual meeting of the American Society of Human Genetics*.
2. **Sun L**, Bull SB, Dimitromanolakis A, Faye L, Waggott D, Paterson AD, the DCCT/EDIC Research Group (2009). BR-squared: a practical solution to the winner's curse in genome-wide scans. Abstract presented at *the annual meeting of the American Society of Human Genetics* (with oral presentation).
3. Faye L, Bull SB, **Sun L** (2009). Effect estimates in a 2-stage design: from genome-wide association to sequencing. Abstract presented at *the annual meeting of the American Society of Human Genetics*.
4. Bull SB, Yoo YJ, **Sun L** (2009). Regression multi-marker tests for gene-based genetic association analysis. Abstract presented at *the annual meeting of the American Society of Human Genetics*.

5. Li W, **Sun L**, Taylor C, Dorfman R, Cojocaru A, Zielenski J, Durie P, Corey M, Strug LJ (2009). Cystic Fibrosis: an inclusive disease demonstrated by analysis of population stratification. Abstract presented at *the annual meeting of the American Society of Human Genetics*.
6. Mirea L, **Sun L**, Stafford JE, Infante-Rivard C, Bull SB (2009). Incorporating evidence for population stratification bias in combined analyses of case-control and case-trio data. IGES abstract.
7. Faye L, **Sun L**, Dimitromanolakis A, Bull SB (2009). A comprehensive look at the likelihood and bootstrap approaches to overcome the winners curse in GWAS. IGES abstract.
8. Dimitromanolakis A, Paterson AD, **Sun L** (2009). Accurate IBD inference identifies cryptic relatedness in 9 HapMap populations. IGES abstract.
9. Paterson AD, Waggott D, Boright A, Hosseini M, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary P, Lachin J, Canty A, Sun L, Bull SB, the DCCT/EDIC Research Group (2009). A genome-wide association study identifies variation near SORCS1 as a major locus for glycemic control in type 1 diabetes, as measured by both HbA1c and glucose. IGES abstract.
10. Yoo YJ, **Sun L**, Bull SB (2009). Gene-based genetic association analysis by multiple regression tests. The annual Joint Statistical Meetings (with oral presentation).
11. Cojocaru A, Li W, **Sun L**, Corey M, Dorfman R, Taylor C, Zielenski J, Durie P, Keenan K, L.J. Strug L (2009). Association analysis for Meconium Ileus incorporating phenotype misclassification. *The Fourth Canadian Genetic Epidemiology and Statistical Genetics Workshop*.
12. Dimitromanolakis A, Paterson AD, **Sun L** (2009). PREST-plus: An improved software tool for the detection of cryptic relations and pedigree errors using high-throughput genotype data. *The Fourth Canadian Genetic Epidemiology and Statistical Genetics Workshop*.
13. Faye L, Bull SB, **Sun L** (2009). Two-stage study designs combining GWA and sequencing: accuracy of genetic effect estimates. *The Fourth Canadian Genetic Epidemiology and Statistical Genetics Workshop* (with oral presentation).
14. Li W, **Sun L**, Taylor C, Dorfman R, Cojocaru A, Zielenski J, Durie P, Corey M, L.J. Strug L (2009). Cystic Fibrosis: an inclusive disease demonstrated by population stratification analysis. *The Fourth Canadian Genetic Epidemiology and Statistical Genetics Workshop*.
15. Yoo YJ, **Sun L**, Bull SB (2009). Regression-based multi-marker tests for gene-based analysis of genetic association data. *The Fourth Canadian Genetic Epidemiology and Statistical Genetics Workshop*.

16. Paterson AD, Waggott D, Boright AP, Hosseini SM, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary PA, Lachin JM, Canty AJ, **Sun L**, Bull SB, and the DCCT/EDIC Research Group (2009). A genome-wide association study identifies variation near SORCS1 as a major locus for glycemic control in type 1 diabetes, as measured by both HbA1c and glucose. *The American Diabetes Association meeting* (with oral presentation).
17. Boright AP, Shen E, Waggott D, Hosseini SM, Cleary PA, Lachin JM, Canty AJ, **Sun L**, Bull SB, Paterson AD and the DCCT/EDIC Research Group (2009). Genome Wide Association Analysis identifies genetic risk factors for Coronary Calcium in subjects with Type 1 Diabetes in the DCCT/EDIC Genetics Study. *The American Diabetes Association meeting*.
18. **Sun L**, Chiang T (2008). PREST-plus: detection of pedigree error and cryptic relatedness among individuals that allows for high-throughput genotype data and adjusts for large-scale multiple hypothesis testing. Abstract 2295 presented at *the annual meeting of the American Society of Human Genetics*.
19. Faye L, Bull SB, **Sun L** (2008). Bootstrap- vs. Likelihood-based methods to reduce selection bias: variance, confidence interval estimation and computational feasibility in large-scale genetic mapping studies. Abstract 2396 presented at *the annual meeting of the American Society of Human Genetics*.
20. Mirea L, **Sun L**, Stafford JE, Bull SB (2008). Combined Individual- and Family-for Population Stratification Bias. Abstract 2439 presented at *the annual meeting of the American Society of Human Genetics*.
21. Li W, **Sun L**, Corey M, Dorfman R, Zielenski J, Durie P, Strug LJ (2008). Complexities in Linkage Analysis for Modifier Genes. Abstract 2505 presented at *the annual meeting of the American Society of Human Genetics*.
22. Strug L, Li W, **Sun L**, Huang B, Corey M, Dorfman R, Blackman S, Cutting G, Zielenski J, Durie P (2008). Modifier gene loci for meconium ileus map to chromosomes 12 and 4. *2008 North American Cystic Fibrosis Conference* (with oral presentation).
23. **Sun L**, Yoo YJ, Bull SB, Paterson AD, Waggott D (2008) Preserving candidate regions in the era of genome-wide association via stratified false-discovery rate control improves power. *Genetic Epidemiology* 32:675-675 (with oral presentation at the International Genetic Epidemiology Society annual meeting 2008).
24. Paterson AD, Boright AP, Waggott D, Zuo Y, Shen E, Mirea L, Zhu L, Huang O, Yoo YJ, Hosseini M, Cleary PA, Lachin JM, **Sun L**, Bull SB (2008). Genome-wide association study of time to long-term diabetic complications. *Genetic Epidemiology* 32:710-711.

25. Yoo YJ, **Sun L**, Bull SB (2008). On using prior information to improve power of genome-wide studies: weighted p-value versus stratified false discovery control. *XXIVth International Biometric Conference* (with oral presentation).
26. Li W, Huang B, **Sun L**, Corey M, Rorfman R, Zielenski J, Durie, P, Strug L (2008). Non-parametric linkage results must be interpreted with caution when searching for modifier genes. *The Third Canadian Genetic Epidemiology and Statistical Genetics Workshop*.
27. Faye L, **Sun L**, Bull SB (2008). Reducing selection bias: comparisons of parametric & non-parametric effect estimation in genetic association studies. *The Third Canadian Genetic Epidemiology and Statistical Genetics Workshop* (**Winner of the "Best Trainee Oral Presentation Award"**).
28. Yoo YJ, Bull SB, Paterson AD, **Sun L** (2008). On using prior information to improve power of genome-wide studies: weighted p-value versus stratified false discovery control. *The Third Canadian Genetic Epidemiology and Statistical Genetics Workshop* (with oral presentation).
29. Faye L, **Sun L**, Bull SB (2007). Reducing selection bias: efficiency and robustness of parametric and non-parametric effect estimation. Abstract 2152 presented at *the annual meeting of the American Society of Human Genetics* (<http://www.ashg.org/genetics/ashg07s/index.shtml>).
30. Mirea L, Bull SB, Stafford JE, **Sun L** (2007). A Logistic regression model for combined individual- and family-level association analyses of binary traits. *Genetic Epidemiology* 31:638-639.
31. Al-Kateb H, Boright AP, Xie X, Mirea L, Sutradhar R, **Sun L**, Bull SB, Paterson AD, and the DCCT/EDIC Research Group (2007). Multiple SOD1 SNPs are associated with the development and progression of diabetic nephropathy. *Genetic Epidemiology* 31:615-615.
32. Bull SB, Xie X, Faye L, **Sun L**, Paterson AD (2007). Bias reduction in genome-wide association studies with time-to-event phenotypes. *Genetic Epidemiology* 31:606-606.
33. **Sun L** (2007). On using linkage signals to improve genome-wide association studies - weighted or stratified false discovery control? *Genetic Epidemiology* 31:498-499.
34. Mirea L, Bull SB, Stafford J, **Sun L** (2007). Combined individual- and family-level association analyses of quantitative traits. *Genetic Epidemiology* 31:489-489.
35. Lee SSF, **Sun L**, Bull SB (2007). A random forest approach to identify important interacting markers in quantitative trait linkage analysis. *Genetic Epidemiology* 31:482-482.

36. Bull SB, **Sun L**, Xie X, Wu LY, Paterson AD (2007). Two-stage genome-wide association: power and sample size for replication. *Genetic Epidemiology* 31:464-464.
37. Al-Kateb H, Mirea L, Xie XL, **Sun L**, Sutradhar R, Bharaj B, Wong I, Liu XQ, Bull SB, Boright AB, Paterson AD (2007). Association of multiple variants in superoxide dismutase 1 (SOD1) with diabetic nephropathy: The DCCT/EDIC genetics study. *Diabetes* 56:A91-A91.
38. Kateb H, Mirea L, Xie XL, **Sun L**, Bull SB, Liu M (2007). Multiple variants in vascular endothelial growth factor (VEGF) gene are risk factors for severe retinopathy in type I diabetes: The DCCT/EDIC genetics study. *Diabetes* 56:A25-A25
39. Lee SSF, Bull SB, **Sun L** (2006). Weighted random forest to map genes in multi-locus linkage analysis. *The Second Canadian Genetic Epidemiology and Statistical Genetics Workshop*.
40. Al-Kateb H, Mirea L, Xie X, Mowjoodi A, Poloumienko A, **Sun L**, Bull SB, Boright AP, Paterson AD, and DCCT/EDIC Study Group (2006). Testing of common variations in 212 candidate genes for association with diabetic nephropathy and retinopathy: the DCCT/EDIC Genetics Study. *Diabetes* 55:A176-A176.
41. Lee SSF, Bull SB, **Sun L** (2005). Efficiency comparisons of estimates from Classical and EM Haseman-Elston regressions when IBD sharing is ambiguous. *Genetic Epidemiology* 29:261-261.
42. **Sun L**, Craiu VR (2005). Joint analysis of false discovery rate and non-discovery rate. *American Journal of Human Genetics* Supplement:A439.
43. Wu LY, **Sun L**, Bull SB (2005). Bias-reduced QTL effect size estimation via statistical resampling. *American Journal of Human Genetics* Supplement:A426.
44. Wu LY, **Sun L**, Bull SB (2004). Robustness of resampling methods to reduce selection bias of genetic effect estimates. *Genetic Epidemiology* 27:304.
45. Biernacka J, **Sun L**, Bull SB (2004). A GEE approach for disease gene localization: Using IBD sharing proportions versus mean IBD. *Genetic Epidemiology* 27:262.
46. Biernacka J, **Sun L**, Bull SB (2004). Localization of linked genes for type 1 diabetes: A simultaneous search for two genes. *American Journal of Human Genetics* Supplement:A531.
47. Tevtoushenko I, Markiewicz D, Deng G, Patel M, Dorfman R, Corey M, Tan M, Li F, **Sun L**, Sandford A, Pare P, Durie P, Tsui LC, Zielenski J (2004). Preliminary analysis of human chromosome 6q27 region as a potential modifier locus for pulmonary disease in cystic fibrosis. *American Journal of Human Genetics* Supplement:A469.

48. **Sun L**, Craiu RV (2003). Sequential testing methods for pedigree error detection based on genome-screen data. *Genetic Epidemiology* 25:271-272.
49. Zielenski J, Markiewicz D, Yuan X, Patel M, **Sun L**, Liu X, Aznarez I, Tsui L-C, and the CF Modifier Collaborative Group (2003). Enhanced haplotype association for the identification of the cystic fibrosis modifier 1 gene using additional genetic markers in the region containing KCNN4. *Pediatric Pulmonology Supplement* 25:93.
50. **Sun L**, Bull S (2003). Resampling-based statistical methods to improve the estimation of locus-specific effects from genome-wide studies. *American Journal of Human Genetics* Supplement 73:609-609.
51. Biernacka J, **Sun L**, Stafford J, Bull SB (2003). Joint localization of two linked disease genes: Derivation, evaluation, and application of a new method. *American Journal of Human Genetics* Supplement 73:193-193.
52. **Sun L**, Cox NJ, McPeck MS (2000). A statistical method for identification of a functional polymorphism in a gene. *American Journal of Human Genetics* Supplement 67:A315.
53. **Sun L**, McPeck MS (1998). Detection of pedigree relationship errors from genotype data. *American Journal of Human Genetics* Supplement 63:A310.

Others

1. Faye L, Bull SB, **Sun L** (2009). Effect estimates in 2-stage design: from genome-wide association to sequencing. *Statistical Society of Canada* annual meeting.
2. Paterson AD, Boright AP, Waggott D, Zuo Y, Mirea L, Zhu L, Huang OL, Yoo YJ, Hosseini M, **Sun L**, Bull SB, and DCCT/EDIC Research Group (2008). Genome-wide association study of 867,874 SNPs with time to long-term diabetic complications in The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (DCCT/EDIC). Late-breaking abstract for *American Diabetes Association*.
3. Yoo YJ, **Sun L**, Bull SB (2008). On using prior information to improve power of genome-wide studies: weighted p-value versus stratified false discovery control. *Statistical Society of Canada* annual meeting (with oral presentation).
4. Faye L, **Sun L**, Bull SB (2008). Reducing selection bias: efficiency and robustness of parametric & non-parametric approaches. *Statistical Society of Canada* annual meeting (with oral presentation).
5. Bull SB, **Sun L**, Paterson, Xie X, Wu LY (2007). Issues in genome-wide association: power and bias for multi-stage designs. *BIRDS Workshop on Statistical Methods for High-throughput Genetic Data*.

6. Al-Kateb H, Mirea L, Xie X, **Sun L**, Bull SB, Liu M, Chen HT, Boright AP, Paterson AD, DCCT/EDIC Research Group (2007). Multiple variants in Vascular Endothelial Growth Factor (VEGF) gene are risk factors for severe retinopathy in type 1 diabetes: The DCCT/EDIC genetics study. Invited Contribution, *American Diabetes Association*.
7. Lee SSF, **Sun L**, Bull SB (2006). Weighted random forest to map genes jointly. *CAIMS-MITACS Joint Annual Conference*, (**2nd Prize Winner in Student Poster Competition**).
8. Craiu RV, Bull SB, Paterson AD, **Sun L** (2006). Separating the wheat from the chaff: statistical methods for false discovery control. Invited Contribution, *Statistical Society of Canada annual meeting*.
9. Lee SSF, **Sun L**, Bull SB (2006) Weighted random forest to map genes in multi-locus linkage analysis. *the Public Health Sciences Research Day, University of Toronto*, (2nd Prize Winner in the poster competition).
10. Bull SB, **Sun L**, Wu LY (2005). Reduction of effect estimate bias in genome-wide studies by resampling. Invited Contribution, *Joint Statistical Meetings*.
11. Bull SB, Biernacka J, **Sun L**, Wu LY (2005). Improving the validity and efficiency of statistical genetic analysis. Invited Contribution, *Statistical Society of Canada annual meeting*.
12. Wu LY, Lee SSF, Shi HS, Lewinger JP, **Sun L**, Bull SB (2004). Resampling methods to reduce the selection bias in genetic effect estimation in genome-wide scans. *Genetic Analysis Workshop 14*.
13. Biernacka JM, Bull SB, **Sun L** (2004). Joint analyses of linked disease genes: Location estimation and hypothesis testing methods. Invited Contribution, *the Eastern North American Region of the International Biometric Society annual meeting*.
14. Biernacka JM, Bull SB, **Sun L**, Stafford JE (2003). A novel approach for estimating locations of multiple linked disease genes. *MITACS 4th annual conference* (**1st Prize Winner in the poster competition**).
15. **Sun L** (2001). Two statistical problems in human genetics. Ph.D. thesis (supervisor: Mary Sara McPeck), Department of Statistics, University of Chicago.

Publicly-Released Research Software (<http://utstat.utoronto.ca/sun/>)

1. Dimitromanolakis A, Faye L, Bull SB, **Sun L** (2009). *BR-squared - Bias-Reduced estimates via Bootstrap Resampling* for overcoming the winner's curse in genome-wide association studies of either disease status or quantitative traits.

2. Yoo YJ, **Sun L** (2008). *SFDR - Stratified False Discovery Rate* control to improve power of genome-wide association studies by utilizing genome-wide linkage scan results.
3. Lee SSF, **Sun L**, Kustra R, Bull SB (2007). *EMRF - EM-Random Forest* approach for multi-locus quantitative trait linkage analysis.
4. Wu LY, **Sun L**, Bull SB (2006). *BR-squared - Linkage* for calculation of Bias-Reduced Bootstrap Resampling-based estimates of locus-specific gene-effect size in linkage analyses
5. Wen W, **Sun L** (2004). *STEP C - Statistical Explanation of Positional Cloning* for identification of polymorphisms that explain a linkage result.
6. **Sun L**, Wilder K, McPeck MS (2000). *PREST - Pedigree Relationship Statistical Test* and ALTERTEST for detection of pedigree errors and estimation of relationships using genome-wide marker data.

Academic Activities

Invited Talks - Conferences and Workshops

- Joint Statistical Meetings (JSM), Washington DC, August 2009
- Institute of Mathematical Statistics - China Conference, Weihai, July 2009
- Statistical Society of Canada annual meeting, St. John's, Newfoundland, June 2007
- SAMSI Workshop on Multiplicity and Reproducibility in Scientific Studies, Research Triangle Park, July 2006
- International Workshop on Applied Probability (IWAP), U. of Connecticut, May 2006
- The First Canadian Genetic Epidemiology and Statistical Genetics Workshop, Toronto, March 2006
- Applied Statistics Symposium, International Chinese Statistical Association (ICSA), Chicago, June 2001

Invited Talks - Seminars

- Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, April 2009

- CRM-ISM-GERAD Statistics Colloque (Universite de Montreal, McGill University, Concordia University, and Universite du Quebec a Montreal), March 2009
- Department of Mathematics and Statistics, U. of Guelph, September 2008
- Department of Statistics, U. of British Columbia, April 2008
- Genetics & Genomic Biology, HSC Research Institute, Toronto, October 2007
- Center for Statistical Genetics, U. of Michigan at Ann Arbor, April 2006
- Department of Statistics, U. of Chicago, April 2006
- Department of Mathematics and Statistics, Laval U. February 2006
- Department of Statistics, U. of Toronto, December 2005
- Department of Statistics, Colorado State U. December 2004
- Department of Mathematics and Statistics, York U. September 2003
- Department of Statistics and Actuarial Science, U. of Waterloo, March 2003
- Genetics & Genomic Biology, HSC Research Institute, Toronto, February 2002
- Department of Public Health Sciences, U. of Toronto, April 2001
- Department of Biostatistics, U. of North Carolina at Chapel Hill, February 2001
- Department of Statistics, Ohio State U. February 2001

Other Presentations

- Workshop on Current Trends and Challenges in Model Selection and Related Areas, Vienna, Austria, July 2008
- Department of Statistics and Actuarial Science, Simon Fraser University, April 2008
- Research Seminar and Journal Club: Statistical Methods in Genetics/Genomics, U. of Toronto, November 2007; January 2006; September 2005; April 2004; November 2003; May 2003; March 2002
- Public Health Sciences Research Day, U. of Toronto, February 2006
- Prague Stochastics 2006: Joint session of the 7th Prague Symposium on Asymptotic Statistics and the 15th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes
- American Society of Human Genetics (ASHG), Annual Meeting, 2005; 2003; 2000; 1999

- International Genetic Epidemiology Society (IGES), Annual Meeting, 2006; 2003
- Genetic Analysis Workshop (GAW) 12, 2000
- The Institute of Mathematical Statistics/International Biometric Society Eastern North American Region Regional Meeting (IMS/ENAR), 2000
- Joint Statistical Meetings (JSM), Annual Meeting, 1999

Professional Activities

Journal referee

- American Journal of Human Genetics
- Annals of Human Genetics
- BioTechniques
- BMC Genetics, proceedings of the Genetic Analysis Workshop
- European Journal of Human Genetics
- Genetics
- Genetic Epidemiology
- Human Genetics
- Human Heredity
- Journal of the American Statistical Association
- Statistical Applications in Genetics and Molecular Biology
- Statistica Sinica

Grant review

- Natural Sciences and Engineering Research Council of Canada
- Canadian Breast Cancer Foundation

Others

- Organizer and chair, invited session on “Genomics”, IMS, 2010

- Member, IGES Publication Committee, 2007
- Member, IGES Education Committee, 2006-2008
- Organizer and chair, invited session on “Statistical analysis and modeling of complex traits”, JSM 2005
- Organizer, invited session on “Statistical genetics - modeling interaction and multi-locus analyses”, ENAR 2004
- Internal grant reviewer, HSC Research Institute, Toronto

Professional Affiliations

- International Genetic Epidemiology Society (IGES), 2002-present
- American Society of Human Genetics (ASHG), 2001-present
- Statistical Society of Canada (SSC), 2001-present
- American Statistical Association (ASA), 1997-present
- Institute of Mathematical Statistics (IMS), 1997-present

Departmental Service

Dalla Lana School of Public Health, University of Toronto

- Graduate Faculty (full member), 2007-present
- Biostatistics Admission Committee (Chair), 2009
- OGS Fellowship Review Committee (member), 2008
- Biostatistics Comprehensive Exam Committee (co-Chair), 2008
- Biostatistics Admission Committee (Chair), 2008
- Graduate Faculty (associate member), 2001-2007
- Biostatistics Comprehensive Exam Committee (member), 2001-present
- Space Committee (member), 2006-present, 2003-2004
- Curriculum Committee (member), 2006-2007
- Biostatistics Comprehensive Exam Committee (Chair), 2005

- Advisory Committee (member), 2002-2003

Department of Statistics, University of Toronto

- Graduate Faculty (full member), 2005-present
- Hiring/Search Committee (member), 2008-2009

Teaching

Graduate Courses

- CHL 5224 - Statistical Genetics,
Dalla Lana School of Public Health, University of Toronto,
Winter 2010; Fall 2008; Fall 2007
(Co-instructor: Wei Xu, Princess Margaret Hospital and DLSPH)
- CHL 5210 - Categorical Data,
Dalla Lana School of Public Health, University of Toronto,
Fall 2009
(Co-instructor: Laurent Briollais, Samuel Lunenfeld Research Institute and DLSPH)
- CHL 5224 - Statistical Genetics,
Department of PHS, University of Toronto,
Fall 2006; Fall 2005; Winter 2004; Winter 2003
- STA 4315 - Computational Methods in Statistical Genetics,
Department of Statistics, University of Toronto,
Winter 2009; Winter 2006; Winter 2005
(Co-instructor: Radu Craiu, Department of Statistics)
- CHL 7001 - Directed Reading - Statistical Methods in Genetic Epidemiology,
Department of PHS, University of Toronto,
Winter 2005
- STAT 946 - Statistical Genetics,
Department of Statistics, University of Waterloo,
Winter 2004 (via video conference)
- CHL 5210 - Statistical Analysis of Qualitative Data,
Department of PHS, University of Toronto,
Winter 2004
(Co-instructor: Joseph Beyene, Department of PHS)

- CHL 5250 - Special Topics in Biostatistics, Department of PHS, University of Toronto, Winter 2003; Fall 2002
- CHL 7001 - Statistical Methods in Genetics, Department of PHS, University of Toronto, Winter 2002

Undergraduate Courses

- STAT220 - Statistical Methods and Their Applications, Department of Statistics, University of Chicago, Spring 2000; Spring 1999; Spring 1998

Students

PhD/Post-Doctoral Fellow (PDF) Supervision

- Baisong Huang, PhD in Biostatistics (07/2007-). Statistical Genetics.
- Lizhen Xu, PhD in Statistics (07/2007-; Co-Supervisor: Radu Craiu). Statistics and Statistical Genetics.
- Laura Faye, PhD in Biostatistics (07/2007-; Co-Supervisor: Shelley Bull). Statistical Genetics.
- Yun Joo Yoo, PDF in Statistical Genetics (10/2007-06/2009; Co-Supervisors: Shelley Bull, Radu Craiu). “Design and analysis of genome-wide studies of complex diseases and traits”.
- Sophia Lee, PhD in Biostatistics (10/2003-11/2007; Co-Supervisor: Shelley Bull). “Random Forests for multi-locus quantitative trait linkage analysis”.
- Longyang Wu, PDF in Statistical Genetics (02/2003-08/2006; Co-Supervisor: Shelley Bull). “Statistical methods to improve the reliability of results from genome-wide studies of complex disease and quantitative traits.”

Research Assistant and Project-Specific Student Supervision

- Elif Acar, PhD in Statistics (04/2009-). “Non-parametric tests for imputed SNPs incorporating genotype uncertainty”
- Apostolos Dimitromanolakis, PhD in Computer Science (11/2008-). “Upgrading PREST for fast and efficient detection of pedigree errors within and across families and cryptic relatedness among cases and controls allowing for high-through genotype data”

- Andreea Cojocaru, Research Associate (10/2008-; Primary Supervisor: Lisa Strug). "Genome-wide linkage and association studies for CF modifier genes"
- Wen Li, Research Associate (10/2008-; Primary Supervisor: Lisa Strug). "Genome-wide linkage and association studies for CF modifier genes"
- Apostolos Dimitromanolakis (10/2008-10/2009; Co-supervisor: Shelley Bull). "Implementation of bootstrap bias reduction methods for genome-wide association data"
- Ye Li, PhD in Biostatistics (09/2008-04/2009). "Non-parametric tests for imputed SNPs incorporating genotype uncertainty"
- Sebastian Song, PhD in Statistics (07/2008-10/2008). "Upgrading PREST for fast and efficient detection of pedigree errors within and across families and cryptic relatedness among cases and controls allowing for high-through genotype data"
- Ted Chiang, Programmer (02/2008-07/2008). "Upgrading PREST for fast and efficient detection of pedigree errors within and across families and cryptic relatedness among cases and controls allowing for high-through genotype data"
- Lam Opal Huang, MSc in Biostatistics (09/2007-09/2008). "Some statistical issues in genome-wide association of common alleles with long-term diabetic complications"
- Laura Faye, PhD in Biostatistics (05/2007-07/2007; Co-Supervisor: Shelley Bull). "Reducing selection bias: efficiency and robustness of parametric and non-parametric effect estimation"
- Baisong Huang, PhD in Biostatistics (05/2006-06/2007). Genetic Analysis Workshop (GAW) 15: "The multiplicity problem in linkage analysis of gene expression data - the power of differentiating *cis* and *trans* regulators"
- Jag Rangrej, Research Assistant (01/2006-04/2007; Primary Supervisor: Celia Greenwood). "Optimal selection of markers for validation from genome-wide association studies"
- Hadassa Brunschwig, MSc in Statistics (01/2005-04/2005). "Directed reading in Statistical Methods in Genetic Epidemiology"
- Haijiang Steven Shi, Computing Assistant (Summer 2004; Primary Supervisor: Shelley Bull). GAW 14: "Resampling methods to reduce the selection bias in genetic effect estimation in genome-wide scans"
- William Wen, Programmer (05/2002-11/2004). Methods implementation as software STEPC - STatistical Explanation for Positional Cloning.
- Lisa Strug, PhD in Biostatistics (05/2002-12/2003; Primary Supervisor: Mary Corey). GAW 13: "The genetics of cross-sectional and longitudinal BMI"

- Michelle Liu, PhD in Biostatistics (05/2002-12/2003; Primary Supervisor: Andrew Paterson). GAW 13: “Transmission ratio distortion in families from the Framingham Heart Study”

Committee Member with Extended Role
(Substantial contributions with training activity on the level of co-supervision)

- Lucia Mirea, PhD in Biostatistics (10/2005-). Statistical Genetics
- Joanna Biernacka, PhD in Biostatistics (09/2001-11/2004). “Statistical methods for studying two linked disease genes”

Committee Member

- Merav Yarkoni-Abitbul, PhD in Dentistry (01/2008-). “Expression profiling of chronic pain genes in mice using a whole genome approach”
- Yan Lu, MSc in Dentistry (05/2005-06/2006). “Polymorphisms in Catechol-O-Methyl-transferase gene affect chronic post-mastectomy pain syndrome”
- Juan Pablo Lewinger, PhD in Statistics (09/2001-02/2004). “Family-based nonparametric tests of linkage and association”

Thesis Examination Committee Member

- Daniel Pu, PhD in Statistics, Department of Mathematics and Statistics York University. Thesis title: “Dependency network in multivariate analysis” (External thesis examiner, May 2009)