HARVARD School of Public Health

*Department of Biostatistics*

Bio 257 Advanced Statistical Genetics

Spring 2, 2015 (2.5 credits)

Tuesday and Thursday , 3:30PM - 5:20PM

Location: FXB G12

**Instructors:** Liming Liang (course head), Nan Laird, Xihong Lin, Christoph Lange and Pete Kraft

**Course Materials:**

Will be drawn from the literature

**Prerequisites:**

Bio 227 (or equivalent) **and** either Bio231 (co-requisite) or Epi 511.

Course is intended for doctoral students; Enrollment will be limited to 15.

**Course Objectives:**

The course is a seminar style course with readings selected from the literature in areas of expertise of the participating faculty. Content may vary from year to year.

At the end of the course the student will be able to critically read foundational papers and current journal articles in statistical genetics; present sophisticated ideas to an audience of peers and engage in doctoral level research in the area.

**Session Format**

Each session of the course will focus on one topic selected by the participating faculty members. Reading material will be assigned ahead of class meeting. During class session, one student will present the assigned paper and the faculty members will coordinate discussion including important technical details, potential applications, limitation and extension of the statistical methods.

**All Tuesday will be office hours by appointment**; faculty can use their office or the classroom below

**All Thursday will be presentation of students**; location will be classroom below.

**Outcome Measures**

Each student must carefully read all assigned papers and be prepared for an in class discussion of the papers. In addition, each student is expected to prepare a 45 minute presentation of at least one paper during the course. The course grade will depend on presentations and class participation. Students are encouraged to read and discuss the papers together outside of class.

**Topics for 2015 Spring 2 and Schedule:**

Nan Laird

**Topic 1.** Estimating heritatibility and co-heritatibility from GWAS data and pleiotropy

Liming Liang

**Topic 2.** Analysis utilizing linkage disequilibrium and reference samples (Haplotyping, imputation and low pass sequencing data)

Pete Kraft and Xihong Lin

**Topic 3.** Analysis of rare variants in sequencing association studies and mediation analysis

Christoph Lange

**Topic 4.** Association analysis using family design and FBATS

**Location and Time for Class meetings:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Location** | **Activity** | **Topic and Faculty** |
| 3/24/2015 Tue | FXB FXB-G12 | Office hours | Topic 1, Nan Laird |
| 3/26/2015 Thu | FXB FXB-G12 | Office hours | Topic 1, Nan Laird |
| 3/31/2015 Tue | FXB FXB-G12 | Presentation | Topic 1, Nan Laird |
| 4/02/2015 Thu | FXB FXB-G12 | Presentation | Topic 1, Nan Laird |
| 4/07/2015 Tue | FXB FXB-G12 | Office hours | Topic 2, Liming Liang |
| 4/09/2015 Thu | FXB FXB-G12 | Office hours | Topic 2, Liming Liang |
| 4/14/2015 Tue | FXB FXB-G12 | Presentation | Topic 2, Liming Liang |
| 4/16/2015 Thu | FXB FXB-G12 | Presentation | Topic 2, Liming Liang |
| 4/21/2015 Tue | FXB FXB-G12 | Office hours | Topic 3, Pete Kraft |
| 4/23/2015 Thu | SPH2 426 (Building 2) | Office hours | Topic 4, Christoph Lange |
| 4/28/2015 Tue | FXB FXB-G12 | Office hours | Topic 3, Pete Kraft |
| 4/30/2015 Thu | FXB FXB-G12 | Office hours | Topic 4, Christoph Lange |
| 5/05/2015 Tue | SPH2 426 (Building 2) | Presentation | Topic 3, Pete Kraft, Xihong Lin |
| 5/07/2015 Thu | SPH2 426 (Building 2) | Presentation | Topic 3, Pete Kraft, Xihong Lin |
| 5/12/2015 Tue | FXB FXB-G12 | Presentation | Topic 4, Christoph Lange |
| 5/14/2015 Thu | FXB FXB-G12 | Presentation | Topic 4, Christoph Lange |

**Reading Reference for Discussion Topics:**

**Topic 1. Estimating heritatibility and co-heritatibility from GWAS data and pleiotropy**

(1) Almasy L, Blangero J. 1998. Multipoint quantitative trait linkage analysis in general pedigree. Am J Hum Genet. 62: 1198-211.

(2) Manolio, TA et al. 2009 Finding the missing heritability of complex diseases. Nature 46 page 747.

(3) Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010 Jul 42(7): 565-9.

(4) Yang et al. Genome Partitioning of Genetic Variation for Complex Traits using Common SNPs. Nature Genetics, 2011 43(6). Page 517.

(5) Lee SH, Yang J, Goddard ME, Visscher PM Wray NR (2012) Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012 Oct 28(19): 2540-2542.

(6) Klei, Luca, Devlin, and Roeder. Pleiotrophy and principal components of heritability combine to increase power for association analysis. 2014, AJHG 94, 662-676.

(7) Zhou, Cho, Lange, Lutz, Silverman and Laird. Integrating Multiple Correlated Phenotypes for Genetic Analysis by Maximizing Heritability. 2015. To appear in Human Heredity.

(8) Yang J, Zaitlen NA, Goddard ME, Visscher PM and Price AL (2013) Mixed model association methods: advantages and pitfalls. Nat Genet. 2014 Feb;46(2):100-6.

(9) Aschard, H. et al.  Maximizing the Power of a Principle Component Analysis of Correlated Phenotypes in Genome Wide Association Studies. 2014. AJHG 94, page 662.

**Topic 2. Analysis utilizing linkage disequilibrium and reference samples (Haplotyping, imputation and low pass sequencing data)**

(1) Stephens, M., and Donnelly, P. (2003). [A comparison of Bayesian methods for haplotype reconstruction from population genotype data.](http://stephenslab.uchicago.edu/MSpapers/Stephens2003a.pdf) *American Journal of Human Genetics,* 73:1162-1169. (PHASE package)

(2) Scheet, P and Stephens, M (2006). [A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.](http://stephenslab.uchicago.edu/MSpapers/Scheet2006.pdf) *Am J Hum Genet (fastPHASE package)*

(3) J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly (2007) A new multipoint method for genome-wide association studies via imputation of genotypes. Nature Genetics 39: 906-913 (IMPUTE package)

(4**) Li Y,** Willer CJ, Ding J, Scheet P, Abecasis GR. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic Epidemiology 34(8): 816-834. (MACH package)

(5) S R Browning and B L Browning (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084-97. (BEAGLE package)

(6) Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale B, Daly M, ARRA Autism Sequencing Collaboration, Sklar P, Sullivan PF, Bergen S, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Purcell SM, Haas DW, Liang L, Sunyaev S, Patterson N, de Bakker PIW, Reich D, Price AL (2012) Extremely low-coverage sequencing enables cost effective GWAS. *Nature Genetics* 20;44(6):631-5.

(7) C Wang, X Zhan, J Bragg-Gresham, HM Kang, D Stambolian, E Chew, K Branham, J Heckenlively, The FUSION Study, RS Fulton, RK Wilson, ER Mardis, X Lin, A Swaroop, S Zöllner, GR Abecasis (2014) Ancestry estimation and control of population stratification for sequence-based association studies. Nature Genetics, 46: 409-415

**Topic 3. Analysis of rare variants in sequencing association studies and mediation analysis**

(1) Lee, S., Abecasis, G., Boehnke, M., Lin, X. (2014) Rare-variant association analysis: Study designs and statistical tests. Am J Human Genetics, 95(1):5-23. PMCID: PMC4085641.

(2) [Li B](http://www.ncbi.nlm.nih.gov/pubmed/?term=Li%20B%5BAuthor%5D&cauthor=true&cauthor_uid=18691683), [Leal SM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Leal%20SM%5BAuthor%5D&cauthor=true&cauthor_uid=18691683). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. [Am J Hum Genet.](http://www.ncbi.nlm.nih.gov/pubmed/?term=Li+and+Leal+%282008%29+Am+J+Hum+Genet+83%3A311-321) 2008 Sep;83(3):311-21. doi: 10.1016/j.ajhg.2008.06.024.

(3) Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. and 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.

(4) Barnett, I. J., Lee, S., and Lin, X. (2013) [Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies](http://www.hsph.harvard.edu/xlin/pub/CEPSKATmanu+supp.pdf). Genetic Epidemiology 37.2, 142-151

(5) Lee, S., Wu, M., Lin, X. (2012) Optimal tests for rare variant effects in sequencing association studies. Biostatistics, 13(4):762-775. PMCID: PMC3440237.

(6) Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. Am J Hum Genet. 2011 Sep 9;89(3):354-67. doi: 10.1016/j.ajhg.2011.07.015.

(7) [Zuk O](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Zuk%20O%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Schaffner SF](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Schaffner%20SF%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Samocha K](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Samocha%20K%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Do R](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Do%20R%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Hechter E](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Hechter%20E%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Kathiresan S](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Kathiresan%20S%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Daly MJ](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Daly%20MJ%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Neale BM](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Neale%20BM%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Sunyaev SR](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Sunyaev%20SR%5BAuthor%5D&cauthor=true&cauthor_uid=24443550), [Lander ES](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=Lander%20ES%5BAuthor%5D&cauthor=true&cauthor_uid=24443550). Searching for missing heritability: designing rare variant association studies. [Proc Natl Acad Sci U S A.](http://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/pubmed/?term=PNAS+%5Bta%5D+Neale+Lander) 2014 Jan 28;111(4):E455-64. doi: 10.1073/pnas.1322563111. Epub 2014 Jan 17.

(8) VanderWeele, T., Asomaning, K., Christiani, D. C. and Lin, X. (2012). Effect of genetic variants on 15q25.1 on lung cancer for smokers but not through smoking. American Journal of Epidemiology,175.10, 1013-1020.

(9) Huang YT, Vanderweele TJ, Lin X. JOINT ANALYSIS OF SNP AND GENE EXPRESSION DATA IN GENETIC ASSOCIATION STUDIES OF COMPLEX DISEASES. Ann Appl Stat. 2014 Mar 1;8(1):352-376.

**Topic 4. Association analysis using family design and FBATS**

(1)    Spielman, Richard S., and Warren J. Ewens. "The TDT and other family-based tests for linkage disequilibrium and association." American journal of human genetics 59.5 (1996): 983.

(2)    Laird, Nan M., and Christoph Lange. "Family-based designs in the age of large-scale gene-association studies." Nature Reviews Genetics 7.5 (2006): 385-394.

(3)    [Thornton T](http://www.ncbi.nlm.nih.gov/pubmed/?term=Thornton%20T%5BAuthor%5D&cauthor=true&cauthor_uid=17668381), [McPeek MS](http://www.ncbi.nlm.nih.gov/pubmed/?term=McPeek%20MS%5BAuthor%5D&cauthor=true&cauthor_uid=17668381). Case-control association testing with related individuals: a more powerful quasi-likelihood score test. [Am J Hum Genet.](http://www.ncbi.nlm.nih.gov/pubmed/?term=Case-Control+Association+Testing+with+Related+Individuals%3A+A+More+Powerful+Quasi-Likelihood+Score+Test) 2007 Aug;81(2):321-37. Epub 2007 Jul 10.

(4)    Horvath, Steve, et al. "Family‐based tests for associating haplotypes with general phenotype data: Application to asthma genetics." Genetic epidemiology 26.1 (2004): 61-69.

(5)    Van Steen, Kristel, et al. "Genomic screening and replication using the same data set in family-based association testing." Nature genetics 37.7 (2005): 683-691.

(6)    Epstein, Michael P., et al. "A Statistical Approach for Rare-Variant Association Testing in Affected Sibships." The American Journal of Human Genetics (2015).

(7)    Preston, Mark D., and Frank Dudbridge. "Utilising Family‐Based Designs for Detecting Rare Variant Disease Associations." Annals of human genetics 78.2 (2014): 129-140.