# Papers for STAT 550 (DL) class, 2008

## Papers available

This list is neither necessary nor sufficient.

That is, it is a somewhat random collection compiled over the years of teaching this class. It contains some papers that were a reasonable choice for a particular student, given their background or research direction, but bear little relationship to the details of the 550 course. Not all choices on the list are reasonable for all students.

Also, additions are welcome – if you have a particular area/paper you would like to choose, feel free to suggest.

The papers are listed first, somewhat sorted by category. The references follow on the last four pages (approx.)

**I Population-related issues/studies, including LD and LD mapping**

Abecasis et al. (2005) – Using LD in association mapping.

Carlson et al. (2004) – selecting SNPs for Association Analysis.

Ceppelini et al. (1955) : the original "Gene counting" paper

Clayton (2000) – a review (as of that date) or fine-scale disequilibrium mapping

Graham and Thompson (1998) – disequilibrium mapping vi likelihoods

Falush et al. (2003) – another Stephens/Pritchard popn structure paper

Kaplan et al. (1995) – the first likelihood approach to disequilibrium mapping.

McPeek and Strahs (1999) – another diseqilibrium mapping approach

Marth et al. (2004) or Pluzhnikov et al. (2002): Inference of human history and demography from genetic data.

Pritchard et al. (2000) – the first Stephens/Pritchard popn structure paper

Storey and Tibshirani (2003) – significance for genome-wide studies.

Thompson et al. (1988) – an interesting story about LD at three tightly linked loci

Thompson and Neel (1997) – coalescent ancestry of a mutant allele and implications for disequilibrium

Voight (2006) or Li and Stephens (2003) –selection and/or recombination in the human genome.

**II Gene ibd and use in sib pair analyses etc.**

Anderson and Weir (2007) – estimating relationships in structured populations.

Blackwelder and Elston (1985) – classic paper on sib pair data.

Dudoit and Speed (2000) – a score test for linkage based on ibd in sib pairs

Karigl (1981) recursive probabilities computations for more general IBD patterns

Spielman and Ewens (1996) – the original TDT test for linkage

Spielman and Ewens (1998) – the original extension TDT to S-TDT

Suarez et al. (1978) – formalization of sib pair linkage tests

Thompson (1974) – gene IBD for multiple individuals

**III Linkage and Segregation Analysis**

Abecasis et al. (2002) – the original Merlin paper.

Boehnke (1994) – paper on limits of resolution of linkage mapping

Elston and Stewart (1971): a fundamental paper in computing likelihoods on pedigrees.

George et al. (2005) – an application of new MCMC methods to estimate lod scores on some extended pedigrees.

Lander and Botstein (1987): the basic "Homozygosity mapping" paper

Lander and Green (1987); The orginal "Mapmaker" paper, which laid methodological foundation for "Genehunter" programs.

Luo and Lin (2003)– information from using marker data in fitting genetic models.

Morton (1955) – the classic introduction of the base-10 (!!!) lod score

Ploughman and Boehnke (1989) – the basis of SIMLINK

**IV PHASE, Haplotyping, etc.**

Schouten et al. (2005) – combines haplotyping with "pedigrees" – first solid study.

Stephens and Donnelly (2003)– a comparison of alternaive computational algorithms/models for phasing

Stephens et al. (2001) – Mathhew Stephens original PHASE paper

Scheet and Stephens (2006) The FastPHASE paper

**V And more .... – not categorized yet**

Levy-Lahad, Wasco, Poorkaj, Romano, Oshima, Pettingell, Yu, Jondro, Schmidt, Wang and et al (1995)

Levy-Lahad, Wijsman, Nemens, Anderson, Goddard, Weber, Bird and Schellenberg (1995)

Meester and Sjerps (2003) – DNA Forensics – a two-stain problem.

Smith (1976) : improvements in computing likelihoods on pedigrees – easier to understand than the more general Cannings et al. papers.

Thompson (1979)

## References


Abecasis, G. R., Cherny, S. S., Cookson, W. O. and Cardon, L. R. (2002), Merlin – rapid analysis of dense genetic maps using sparse gene flow trees, *Nature Genetics* **30**, 97–101.

Abecasis, G. R., Ghosh, D. and Nichols, T. E. (2005), Linkage disequilibrium: Ancient history drives the new genetics, *Human Heredity* **59**, 118–124.

Anderson, A. D. and Weir, B. S. (2007), A maximum-likelihood method for the estimation of pairwise relatedness in structured populations, *Genetics* **176**, 421–440.

Blackwelder, W. C. and Elston, R. C. (1985), A comparison of sib-pair linkage tests for disease susceptibility loci, *Genetic Epidemiology* **2**, 85–97.

Boehnke, M. (1994), Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes, *Am J Hum Genet* **55**(2), 379–390.

Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L. and Nickerson, D. A. (2004), Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium, *American Journal of Human Genetics* **74**, 106–120.

Ceppelini, R., Siniscalco, M. and Smith, C. A. B. (1955), The estimation of gene frequencies in a random mating population, *Annals of Human Genetics* **20**, 97–115.

Clayton, D. (2000), Linkage disequilibrium mapping of disease susceptibility genes in human populstions, *International Statistical Review* **68**, 23–43.

Dudoit, S. and Speed, T. P. (2000), A score test for linkage analysis of qualitative and quantitative traits based on identity by descent on sib pairs, *Biostatistics* **1**, 1–26.

Elston, R. C. and Stewart, J. (1971), A general model for the analysis of pedigree data, *Human Heredity* **21**, 523–542.

Falush, D., Stephens, M. and Pritchard, J. K. (2003), Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics* **164**, 1567–1587.

George, A. W., Wijsman, E. M. and Thompson, E. A. (2005), MCMC multilocus lod scores: Application of a new approach, *Human Heredity* **59**, 98–108.

Graham, J. and Thompson, E. A. (1998), Disequilibrium likelihoods for fine-scale mapping of a rare allele, *American Journal of Human Genetics* **63**, 1517–1530.

Kaplan, N. L., Hill, W. G. and Weir, B. S. (1995), Likelihood methods for locating disease genes in nonequilibrium populations, *American Journal of Human Genetics* **56**(1), 18–32.

Karigl, G. (1981), A recursive algorithm for the calculation of gene identity coefficients, *Annals of Human Genetics* **45**, 299–305.

Lander, E. S. and Botstein, D. (1987), Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children, *Science* **236**, 1567–1570.

Lander, E. S. and Green, P. (1987), Construction of multilocus genetic linkage maps in humans., *Proceedings of the National Academy of Sciences (USA)* **84**(8), 2363–2367.

Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D. M., Oshima, J., Pettingell, W. H., Yu, C. E., Jondro, P. D., Schmidt, S. D., Wang, K. and et al (1995), Candidate gene for the Chromosome 1 familial Alzheimer's disease locus, *Science* **269**(5226), 973–977.

Levy-Lahad, E., Wijsman, E. M., Nemens, E., Anderson, L., Goddard, K. A., Weber, J. L., Bird, T. D. and Schellenberg, G. D. (1995), Familial Alzheimer's disease locus on Chromosome 1, *Science* **269**(5226), 970–973.

Li, N. and Stephens, M. (2003), Modelling linkage disequilibrium, and identifying recombination hotspots using SNP data, *Genetics* **165**, 2213–2233.

Luo, Y. and Lin, S. (2003), Information gain for genetic parameter estimation with incorporation of marker data, *Biometrics* **59**, 393–401.

Marth, G. T., Czabarka, E., Murval, J. and Sherry, S. T. (2004), The allele frequency spectrum in genome-wide human variation reveals signals of differential demographic history, *Genetics* **166**, 1699–1712.

McPeek, M. S. and Strahs, A. (1999), Assessment of linkage diseqilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping, *American Journal of Human Genetics* **65**, 858–875.

Meester and Sjerps (2003), *Biometrics* .

Morton, N. E. (1955), Sequential tests for the detection of linkage, *American Journal of Human Genetics* **7**, 277–318.

Ploughman, L. M. and Boehnke, M. (1989), Estimating the power of a proposed linkage study for a complex genetic trait, *American Journal of Human Genetics* **44**, 543–551.

Pluzhnikov, A., DiRienzo, A. and Hudson, R. R. (2002), Inferences about human demography based on multilocus analyses of non-coding sequences, *Genetics* **161**, 1209–1281.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000), Inference of population structure using multilocus genotype data, *Genetics* **155**, 945–959.

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, *American Journal of Human Genetics* **78**, 629–644.

Schouten, M. T., Williams, C. K. I. and Haley, C. S. (2005), The impact of using related individuals for haplotype reconstruction in population studies, *Genetics* **171**, (Nov).

Smith, C. A. B. (1976), The use of matrices in calculating Mendelian probabilities, *Annals of Human Genetics* **40**, 37–54.

Spielman, R. S. and Ewens, W. J. (1996), The TDT and other family-based tests for linkage disequilibrium, *American Journal of Human Genetics* **59**, 983–989.

Spielman, R. S. and Ewens, W. J. (1998), A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test, *American Journal of Human Genetics* **62**, 450–458.

Stephens, M. and Donnelly, P. (2003), A comparison of Bayesian methods for haplotype reconstruction, *American Journal of Human Genetics* **73**, 1162–1169.

Stephens, M., Smith, N. and Donnelly, P. (2001), A new statistical method for haplotype reconstruction from population data, *American Journal of Human Genetics* **68**, 978–989.

Storey, J. D. and Tibshirani, R. (2003), Statistical significance for genome-wide studies, *Proceedings of the National Academy of Sciences* **100**, 9440–9445.

Suarez, B. K., Rice, J. and Reich, T. (1978), The generalized sib pair IBD distribution: its use in the detection of linkage, *Annals of Human Genetics* **42**, 87–94.

Thompson, E. A. (1974), Gene identities and multiple relationships, *Biometrics* **30**, 667–680.

Thompson, E. A. (1979), Ancestral inference III: The ancestral structure of the population of Tristan da Cunha, *Annals of Human Genetics* **43**, 167–176.

Thompson, E. A., Deeb, S., Walker, D. and Motulsky, A. G. (1988), The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes, *American Journal of Human Genetics* **42**, 113–124.

Thompson, E. A. and Neel, J. V. (1997), Allelic disequilibrium and allele frequency distribution as a function of social and demographic history, *American Journal of Human Genetics* **60**, 197–204.

Voight (2006), A map of recent positive selection in the human genome, *PLoS Biol.* **4**, e72.