

Bibliography

Affymetrix (2001). *Microarray Suite User Guide, Version 5.0*. Affymetrix Inc., Santa Clara, CA, USA.

Affymetrix (2002). *Statistical algorithms description document*. Affymetrix Inc., Santa Clara, CA, USA.

Allison DB, Gadbury GL, Heo M, Fernández JR, Lee CK, Prolla TA and Weindruch R (2002). A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics Data Analysis* **39**(1):1–20.

Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**(1):289–300.

Benjamini Y and Hochberg Y (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25**(1):60–83.

Bhattacharjee A, Richards W, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark E, Lander E, Wong W, Johnson B, Golub T, Sugarbaker D and Meyerson M (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences USA* **98**(24):13790–13795.

Bolstad B, Irizarry R, Astrand M and Speed T (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**(2):185–193.

Broberg P (2003). Statistical methods for ranking differentially expressed genes. *Genome Biology* **4**(6):R41.

- Broberg P (2005). A comparative review of estimates of the proportion unchanged genes and the false discovery rate. *BMC Bioinformatics* **6**(1):199.
- Cui X, Hwang GJ, Qiu J, Blades NJ and Churchill GA (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6**(1):59–75.
- Dalmasso C, Broët P and Moreau T (2005). A simple procedure for estimating the false discovery rate. *Bioinformatics* **21**(5):660–668.
- Do KA, Müller P and Tang F (2005). A Bayesian mixture model for differential gene expression. *Journal of the Royal Statistical Society Series C* **54**(3):627–644.
- Dudoit S, van der Laan MJ and Pollard KS (2004). Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology* **3**(1):Article 13.
- Dudoit S, Yang YH, Callow MJ and Speed TP (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12**:111–139.
- Efron B (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association* **99**(465):96–104.
- Efron B, Tibshirani R, Storey JD and Tusher V (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* **96**(456):1151–1160.
- Ferkingstad E, Langaas M and Lindquist B (2003). Estimating the proportion of true null hypotheses, with application to DNA microarray data. Preprint Statistics 4, Norwegian University of Science and Technology, Trondheim.
- Finner H and Roters M (2001). On the false discovery rate and expected type I errors. *Biometrical Journal* **43**(8):985–1005.
- Gao X (2006). Construction of null statistics in permutation-based multiple testing for multi-factorial microarray experiments. *Bioinformatics* **22**(12):1486–1494.
- Ge Y, Dudoit S and Speed TP (2003). Resampling-based multiple testing for microarray data analysis. *Test* **12**(1):1–77.

- Genovese C and Wassermann L (2004). A stochastic process approach to false discovery rates. *Annals of Statistics* **32**(3):1035–1061.
- Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J and Zhang J (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**(10):R80.
- Guo X and Pan W (2005). Using weighted permutation scores to detect differential gene expression with microarray data. *Journal of Bioinformatics and Computational Biology* **3**(4):989–1006.
- Huang E, Cheng S, Dressman H, Pittman J, Tsou M, Horng C, Bild A, Iversen E, Liao M, Chen C, West M, Nevins J and Huang A (2003). Gene expression predictors of breast cancer outcomes. *Lancet* **361**(9369):1590–1596.
- Huber W, von Heydebreck A, Sültmann H, Poustka A and Vingron M (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**(Suppl 1):96–104.
- Huber W, von Heydebreck A, Sültmann H, Poustka A and Vingron M (2003). Parameter estimation for the calibration and variance stabilization of microarray data. *Statistical Applications in Genetics and Molecular Biology* **2**(1):Article 3.
- Irizarry R, Bolstad B, Collin F, Cope L, Hobbs B and Speed T (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research* **31**(4):e15.
- Irizarry R, Hobbs B, Collin F, Beazer-Barclay Y, Antonellis K, Scherf U and Speed T (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**(2):249–264.
- Jain N, Thatte J, Braciale T, Ley K, O’Connell M and Lee JK (2003). Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* **19**(15):1945–1951.
- Klebanov L, Jordan C and Yakovlev A (2006). A new type of stochastic dependence revealed in gene expression data. *Statistical Applications in Genetics and Molecular Biology* **5**(1):Article 7.

- Klebanov L and Yakovlev A (2006). Treating expression levels of different genes as a sample in microarray data analysis: is it worth a risk? *Statistical Applications in Genetics and Molecular Biology* **5**(1):Article 9.
- Kooperberg C, Sipione S, LeBlanc M, Strand AD, Cattaneo E and Olson JM (2002). Evaluating test statistics to select interesting genes in microarray experiments. *Human Molecular Genetics* **11**(19):2223–2232.
- van der Laan MJ, Dudoit S and Pollard KS (2004). Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology* **3**(1):Article 14.
- Langaas M, Lindqvist B and Ferkingstad E (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society Series B* **67**(4):555–572.
- Lee MLT, Gray RJ, Björkbacka H and Freeman MW (2005). Generalized rank tests for replicated microarray data. *Statistical Applications in Genetics and Molecular Biology* **4**(1):Article 3.
- Liao J, Lin Y, Selvanayagam ZE and Shih WJ (2004). A mixture model for estimating the local false discovery rate in DNA microarray analysis. *Bioinformatics* **20**(16):2694–2701.
- Martin DE, Demougin P, Hall MN and Bellis M (2004). Rank difference analysis of microarrays (RDAM), a novel approach to statistical analysis of microarray expression profiling data. *BMC Bioinformatics* **5**(1):148.
- Meinshausen N and Bühlmann P (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika* **92**(4):893–907.
- Meinshausen N and Rice J (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Annals of Statistics* **34**(1):373–393.
- Mosig M, Lipkin E, Khutoreskaya G, Tchourzyna E, Soller M and Friedmann A (2001). A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* **157**(4):1683–1698.

- Nettleton D and Hwang JG (2003). Estimating the number of false null hypothesis when conducting many tests. Technical Report 9, Department of Statistics & Statistical Laboratory, Iowa State University.
- Neuhäuser M and Lam FC (2004). Nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. In *Second Asia-Pacific Bioinformatics Conference (APBC 2004)*, pp. 139–143. Australian Computer Society.
- Pan W (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**(4):546–554.
- Pepe M, Longton G, Anderson G and Schummer M (2003). Selecting differentially expressed genes from microarray experiments. *Biometrics* **59**(1):133–142.
- Pollard KS, Birkner MD, van der Laan MJ and Dudoit S (2005). Test statistics null distributions in multiple testing: simulation studies and applications to genomics. Technical Report 184, U.C. Berkeley Division of Biostatistics Working Paper Series.
- Pollard KS and van der Laan MJ (2003). Multiple testing for gene expression data: an investigation of null distributions with consequences for the permutation test. In *Proceedings of the 2003 International MultiConference in Computer Science and Engineering*, pp. 3–9.
- Pollard KS and van der Laan MJ (2004). Choice of a null distribution in resampling-based multiple testing. *Journal of Statistical Planning and Inference* **125**:85–100.
- Pounds S and Cheng C (2004). Improving false discovery rate estimation. *Bioinformatics* **20**(11):1737–1745.
- Pounds S and Morris SW (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics* **19**(10):1236–1242.
- Qiu X, Klebanov L and Yakovlev A (2005). Correlation between gene expression levels and limitations of the empirical Bayes methodology for finding differentially expressed genes. *Statistical Applications in Genetics and Molecular Biology* **4**(1):Article 34.

- Qiu X, Xiao Y, Gordon A and Yakovlev A (2006). Assessing stability of gene selection in microarray data analysis. *BMC Bioinformatics* 7:50.
- R Development Core Team (2005). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Scheid S and Spang R (2004). A stochastic downhill search algorithm for estimating the local false discovery rate. *IEEE Transactions on Computational Biology and Bioinformatics* 1(3):98–108.
- Scheid S and Spang R (2006). Permutation filtering: a novel concept for significance analysis of large-scale genomic data. In A Apostolico, C Guerra, S Istrail, P Pevzner and M Waterman (eds.), *Research in Computational Molecular Biology: 10th Annual International Conference, Proceedings of RECOMB 2006, Venice, Italy, April 2-5, 2006*, volume 3909 of *Lecture Notes in Computer Science*, pp. 338–347. Springer, Heidelberg.
- Scheid S and Spang R (2007). Compensating for unknown confounders in microarray data analysis using filtered permutations. *Accepted for Journal of Computational Biology*.
- Schweder T and Spjøtvoll E (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* 69(3):493–502.
- Singh D, Febbo P, Ross K, Jackson D, Manola J, Ladd C, Tamayo P, Renshaw A, D’Amico A, Richie J, Lander E, Loda M, Kantoff P, Golub T and Sellers W (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1(2):203–9.
- Smyth GK (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3(1):Article 3.
- Storey J, Taylor J and Siegmund D (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B* 66(1):187–205.
- Storey JD (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B* 64(3):479–498.

- Storey JD (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Annals of Statistics* **31**(6):2013–2035.
- Storey JD, Dai JY and Leek JT (2005). The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Technical Report 260, UW Biostatistics Working Paper Series, University of Washington.
- Storey JD and Tibshirani R (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences USA* **100**(16):9440–9445.
- Tsai CA, miin Hsueh H and Chen JJ (2003). Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* **59**(4):1071–1081.
- Tusher VG, Tibshirani R and Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences USA* **98**(9):5116–5121.
- Wu B (2005). Differential gene expression detection using penalized linear regression models: the improved SAM statistics. *Bioinformatics* **21**(8):1565–1571.
- Xie Y, Pan W and Khodursky AB (2005). A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics* **21**(23):4280–4288.
- Yeoh E, Ross M, Shurtleff S, Williams W, Patel D, Mahfouz R, Behm F, Raimondi S, Relling M, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui C, Evans W, Naeve C, Wong L and Downing J (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**(2):133–143.
- Zhao Y and Pan W (2003). Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics* **19**(9):1046–1054.

