

## Bibliography

- Affymetrix (2001). *Microarray Suite User Guide, Version 5.0*.
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P (2002). *Molecular Biology of the Cell*. Garland Publishing, New York, fourth edition.
- Alizadeh A, Eisen M, Davis R, Ma C, Lossos I, Rosenwald A, Boldrick J, Sabet H, Tran T, Yu X, Powell J, Yang L, Marti G, Moore T, Hudson J, Lu L, Lewis D, Tibshirani R, Sherlock G, Chan W, Greiner T, Weisenburger D, Armitage J, Warnke R and Staudt L (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**(6769):503–11.
- Alon U, Barkai N, Notterman D, Gish K, Ybarra S, Mack D and Levine A (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* **96**(12):6745–50.
- Altman D and Royston P (2000). What do we mean by validating a prognostic model? *Stat Med* **19**(4):453–73.
- Alwine J, Kemp D and Stark G (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* **74**(12):5350–4.
- Ambroise C and McLachlan G (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A* **99**(10):6562–6.
- Armstrong S, Staunton J, Silverman L, Pieters R, den Boer M, Minden M, Sallan S, Lander E, Golub T and Korsmeyer S (2002). Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**(1):41–47.
- Ayers M, Symmans W, Stec J, Damokosh A, Clark E, Hess K, Lecoche M, Metivier J, Booser D, Ibrahim N, Valero V, Royce M, Arun B, Whitman G, Ross J, Sneige N, Hortobagyi G and Pusztai L (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* **22**(12):2284–93.
- Barrans J, Allen P, Stamatidou D, Dzau V and Liew C (2002). Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* **160**(6):2035–43.

- Barrett MT (2005). Stacking the chips for biological discovery. *Nat Genet* **37** Suppl:S1.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB and Hanash S (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* **8**(8):816–24.
- Ben-Dor A, Bruhn L, Friedman N, Nachman I, Schummer M and Yakhini Z (2000). Tissue classification with gene expression profiles. *J Comput Biol* **7**(3-4):559–83.
- Bengio Y and Grandvalet Y (2004). No unbiased estimator of the variance of k-fold cross-validation. *J Mach Learn Res* **5**(Sep):1089–1105.
- Bezdek J (1973). *Fuzzy mathematics in pattern classification*. Ph.D. thesis, Center for Applied Mathematics, Cornell University, Ithica, N.Y.
- 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. *Proc Natl Acad Sci U S A* **98**(24):13790–5.
- Bishop C (1995). *Neural networks for pattern classificaton*. Clarendon Press.
- Bissell M and Radisky D (2001). Putting tumours in context. *Nat Rev Cancer* **1**(1):46–54.
- Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D and Sondak V (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**(6795):536–540.
- Blum A, Kalai A and Langford J (1999). Beating the hold-out: Bounds for k-fold and progressive cross-validation. In *Proceedings of the twelfth annual conference on Computational learning theory (COLT)*, pp. 203–208.
- Blum A and Rivest R (1992). Training a 3-node neural networks is np-complete. *Neural Networks* **5**(1):117–127.
- Bø T and Jonassen I (2002). New feature subset selection procedures for classification of expression profiles. *Genome Biol* **3**(4):research0017.1–research0017.11.
- 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.
- Boser B, Guyon I and Vapnik V (1992). *A training algorithm for optimal margin classifiers*. 5th Annual ACM Workshop on COLT.
- Braga-Neto U and Dougherty E (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics* **20**(3):374–80.

- Breslow A (1970). Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg* **172**(5):902–8.
- Broberg P (2003). Statistical methods for ranking differentially expressed genes. *Genome Biology* **4**(6):R41.
- Brown M, Grundy W, Lin D, Cristianini N, Sugnet C, Furey T, Ares M and Haussler D (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* **97**(1):262–7.
- Brown P and Botstein D (1999). Exploring the new world of genome with dna microarrays. *Nature Genetics* **21**(Suppl. 1):33–37.
- Bullinger L, Döhner K, Bair E, Fröhling S, Schlenk R, Tibshirani R, Döhner H and Pollack J (2004). Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* **350**(16):1605–16.
- Burczynski ME, Oestreicher JL, Cahilly MJ, Mounts DP, Whitley MZ, Speicher LA and Trepicchio WL (2005). Clinical pharmacogenomics and transcriptional profiling in early phase oncology clinical trials. *Curr Mol Med* **5**(1):83–102.
- Büssow K, Konthur Z, Lueking A, Lehrach H and Walter G (2001). Protein array technology. Potential use in medical diagnostics. *Am J Pharmacogenomics* **1**(1):37–43.
- Campbell C, Li Y and Tipping M (2001). An efficient feature selection algorithm for classification of gene expression data. In *NIPS Workshop on Machine Learning Techniques for Bioinformatics, Vancouver, Canada*.
- Cario G, Stanulla M, Fine B, Teuffel O, Neuhoff N, Schrauder A, Flohr T, Schafer B, Bartram C, Welte K, Schlegelberger B and Schrappe M (2005). Distinct gene expression profiles determine molecular treatment response in childhood acute lymphoblastic leukemia. *Blood* **105**(2):821–826.
- Chang H, Sneddon J, Alizadeh A, Sood R, West R, Chi J, van de Rijn M, Botstei D and Brown P (2004). Gene expression signature of a fibroblast serum response predicts cancer progression. *PLoS* **2**(2):206–214.
- Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, Dai H, He YD, Veer LJV, Bartelink H, van de Rijn M, Brown PO and van de Vijver MJ (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102**(10):3738–43.
- Chang J, Wooten E, Tsimelzon A, Hilsenbeck S, Gutierrez M, Elledge R, Mohsin S, Osborne C, Chamness G, Allred D and O’Connell P (2003). Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* **362**(9381):362–9.
- Chatfield C (1995). Model uncertainty, data mining and statistical inference. *J R Stat Soc Series A* **158**:419–66.

- Chee M, Yang R, Hubbell E, Berno A, Huang X, Stern D, Winkler J, Lockhart D, Morris M and Fodor S (1996). Accessing genetic information with high-density DNA arrays. *Science* **274**(5287):610–4.
- Chen C, Hou M, Wang J, Chang T, Lai D, Chen Y, Hung S and Lin S (2005). Simultaneous detection of multiple mRNA markers CK19, CEA, c-Met, Her2/neu and hMAM with membrane array, an innovative technique with a great potential for breast cancer diagnosis. *Cancer Lett* .
- Cheok M, Yang W, Pui C, Downing J, Cheng C, Naeve C, Relling M and Evans W (2003). Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. *Nat Genet* **34**(1):85–90.
- Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Ritz J and Foa R (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood* **103**(7):2771–8.
- Ciro M, Bracken AP and Helin K (2003). Profiling cancer. *Curr Opin Cell Biol* **15**(2):213–20.
- Clark E, Golub T, Lander E and Hynes R (2000). Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**(6795):532–5.
- Clarke M, Greaves L and James S (1997). MeSH terms must be used in Medline searches. *BMJ* **314**(7088):1203.
- Cortes C and Vapnik V (1995). Support vector networks. *Machine Learning* **20**(3):273–297.
- Dave S, Wright G, Tan B, Rosenwald A, Gascoyne R, Chan W, Fisher R, Braziel R, Rimsza L, Grogan T, Miller T, LeBlanc M, Greiner T, Weisenburger D, Lynch J, Vose J, Armitage J, Smeland E, Kvaloy S, Holte H, Delabie J, Connors J, Lansdorp P, Ouyang Q, Lister T, Davies A, Norton A, Muller-Hermelink H, Ott G, Campo E, Montserrat E, Wilson W, Jaffe E, Simon R, Yang L, Powell J, Zhao H, Goldschmidt N, Chiorazzi M and Staudt L (2004). Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* **351**(21):2159–2169.
- Dhanasekaran S, Barrette T, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta K, Rubin M and Chinnaiyan A (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**(6849):822–6.
- Duan KB, Rajapakse JC, Wang H and Azuaje F (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans Nanobioscience* **4**(3):228–34.
- 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.

- Dunn J (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* **3**:32–57.
- Dvorak H (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med* **315**(26):1650–9.
- Dvorák Z, Pascussi J and Modrianský M (2003). Approaches to messenger RNA detection - comparison of methods. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **147**(2):131–5.
- Eberwine J (1996). Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *Biotechniques* **20**(4):584–91.
- Edén P, Ritz C, Rose C, Fernö M and Peterson C (2004). "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* **40**(12):1837–41.
- Efron B (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7**(1):1–26.
- Efron B (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association* **78**(382):316–331.
- Eisen M, Spellman P, Brown P and Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* **95**(25):14863–8.
- Endoh H, Tomida S, Yatabe Y, Konishi H, Osada H, Tajima K, Kuwano H, Takahashi T and Mitsudomi T (2004). Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. *J Clin Oncol* **22**(5):811–9.
- Evgeniou T and Poggio T (1997). Sparse representations to multiple signals. AI Memo 1619, Massachusetts Institute of Technology.
- Fernandes R and Skiena S (2002). Microarray synthesis through multiple-use PCR primer design. *Bioinformatics* **18 Suppl 1**:S128–35.
- Foss D, Baarsch M and Murtaugh M (1998). Regulation of hypoxanthine phosphoribosyl-transferase, glyceraldehyde-3-phosphate dehydrogenase and beta-actin mRNA expression in porcine immune cells and tissues. *Anim Biotechnol* **9**(1):67–78.
- Freije W, Castro-Vargas F, Fang Z, Horvath S, Cloughesy T, Liao L, Mischel P and Nelson S (2004). Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* **64**(18):6503–6510.
- Fridde C, Koga T, Rubin E and Bristow J (2000). Expression profiling reveals distinct sets of genes altered during induction and regression of cardiac hypertrophy. *Proc Natl Acad Sci U S A* **97**(12):6745–50.

- Fuller G, Hess K, Rhee C, Yung W, Sawaya R, Bruner J and Zhang W (2002). Molecular classification of human diffuse gliomas by multidimensional scaling analysis of gene expression profiles parallels morphology-based classification, correlates with survival, and reveals clinically-relevant novel glioma subsets. *Brain Pathol* **12**(1):108–116.
- Geisser S (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association* **70**(350):320–328.
- Geman S, Bienenstock E and Doursat R (1992). Neural networks and the bias/variance dilemma. *Neural Computation* **4**:1–58.
- Glinsky G, Higashiyama T and Glinskii A (2004). Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res* **10**(7):2272–83.
- Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, Coller H, Loh M, Downing J, Caligiuri M, Bloomfield C and Lander E (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**(5439):531–7.
- Griffiths AJF, Gelbart WM, Lewontin RC and Miller JH (2002). *Modern Genetic Analysis – Integrating Genes and Genomes*. W. H. Freeman and Company, second edition.
- Grzeskowiak R, Witt H, Drungowski M, Thermann R, Hennig S, Perrot A, Osterziel KJ, Klingbiel D, Scheid S, Spang R, Lehrach H and Ruiz P (2003). Expression profiling of human idiopathic dilated cardiomyopathy. *Cardiovascular Research* **59**(2):400–411.
- Guyon I and Elisseeff A (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**(Mar):1157 – 1182.
- Guyon I, Weston J, Barnhill S and Vapnik V (2002). Gene selection for cancer classification using support vector machines. *Machine Learning* **46**(1-3):389–422.
- Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W and Schoch C (2005). A global approach to the diagnosis of leukemia using gene expression profiling. *Blood* **106**(4):1189–98.
- Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, Federman S, Miller JR, Allen RE, Singer MI, Leong SP, Ljung BM, Sagebiel RW and Kashani-Sabet M (2005). The gene expression signatures of melanoma progression. *Proc Natl Acad Sci U S A* **102**(17):6092–7.
- Harries M and Smith I (2002). The development and clinical use of trastuzumab (Herceptin). *Endocr Relat Cancer* **9**(2):75–85.
- Hartemink A, Gifford D, Jaakkola T and Young R (2001). Maximum likelihood estimation of optimal scaling factors for expression array normalizations. In *SPIE BIOS*.
- Hartigan J and Wong MA (1979). A k-means clustering algorithm. *Applied Statistics* **28**:100–4.

- Hastie T, Tibshirani R and Friedman J (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York.
- Holleman A, Cheok MH, den Boer ML, Yang W, Veerman AJ, Kazemier KM, Pei D, Cheng C, Pui CH, Relling MV, Janka-Schaub GE, Pieters R and Evans WE (2004). Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. *N Engl J Med* **351**(6):533–42.
- Hoyert DL, Kung HC and Smith BL (2005). Deaths: preliminary data for 2003. *Natl Vital Stat Rep* **53**(15):1–48.
- Hu, Kapoor, Zhang, Hamilton and Coombes (2005). Analysis of dose-response effects on gene expression data with comparison of two microarray platforms. *Bioinformatics* **21**(17):3524–9.
- Hua J, Xiong Z, Lowey J, Suh E and Dougherty ER (2005). Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* **21**(8):1509–15.
- 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.
- Hughes G (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans Information Theory* **14**:55–63.
- Hwang J, Allen P, Tseng G, Lam C, Fananapazir L, Dzau V and Liew C (2002). Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* **10**(1):31–44.
- Ihaka R and Gentleman R (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**(3):299–314.
- 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):249–64.
- Irizarry R, Wu Z and Jaffee H (2006). Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* **22**(7):789–94.
- Jäger J, Weichenhan D, Ivandic B and Spang R (2005). Early diagnostic marker panel determination for microarray based clinical studies. *SAGMB* **4**(1):Art 9.
- Jain A and Zongker D (1997). Feature selection - evaluation, application and small sample performance. *IEEE Trans Pattern Anal Machine Intell* **19**:153–158.

- Jiang C, Lu H, Vincent K, Shankara S, Belanger A, Cheng S, Akita G, Kelly R, Goldberg M and Gregory R (2002). Gene expression profiles in human cardiac cells subjected to hypoxia or expressing a hybrid form of HIF-1 alpha. *Physiol Genomics* **8**(1):23–32.
- Jones J, Otu H, Spentzos D, Kolia S, Inan M, Beecken W, Fellbaum C, Gu X, Joseph M, Pantuck A, Jonas D and Libermann T (2005). Gene signatures of progression and metastasis in renal cell cancer. *Clin Cancer Res* **11**(16):5730–9.
- Kaynak B, von Heydebreck A, Mebus S, Seelow D, Hennig S, Vogel J, Sperling HP, Pregla R, Alexi-Meskishvili V, Hetzer R, Lange PE, Vingron M, Lehrach H and Sperling S (2003). Genome-wide array analysis of normal and malformed human hearts. *Circulation* **107**(19):2467–74.
- Kohavi R (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, pp. 1137–1145.
- Kohavi R and John G (1997). Wrappers for feature subset selection. *Artificial Intelligence* **97**(1-2):273–324.
- Kostka D and Spang R (2004). Finding disease specific alterations in the co-expression of genes. *Bioinformatics* **20 Suppl 1**:I194–I199.
- Kothapalli R, Yoder SJ, Mane S and Loughran TP (2002). Microarray results: how accurate are they? *BMC Bioinformatics* **3**:22.
- Krishnapuram B, Carin L and Hartemink A (2004). *Kernel methods in computational biology*, chapter Joint feature selection and classifier design. MIT Press.
- Kroll T and Wölfel S (2002). Ranking: a closer look on globalisation methods for normalisation of gene expression arrays. *Nucleic Acids Res* **30**(11):e50.
- Kudo M, Somol P, Pudil P, Shimbo M and Sklansky J (2000). Comparison of classifier-specific feature selection algorithms. In *SSPR/SPR*, pp. 677–686.
- Lachenbruch P and Mickey M (1968). Estimation of error rates in discriminant analysis. *Technometrics* **10**(1):1–11.
- Lai C, Reinder M and Wessels L (2005). "random subspace method for multivariate feature selection". In *Eleventh annual conference of the Advanced School for Computing and Imaging*.
- Lai C, Reinders M, Van't Veer L and Wessels L (2006). A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets. *BMC Bioinformatics* **7**(1):235.
- Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD and Pollack JR (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* **101**(3):811–6.



- Lee J, Lee J, Park M and Song S (2005). An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics Data Analysis* **48**:869–85.
- Li C and Wong W (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98**(1):31–6.
- Li J, Pankratz M and Johnson J (2002). Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. *Toxicol Sci* **69**(2):383–90.
- Li W (2005). How many genes are needed for early detection of breast cancer, based on gene expression patterns in peripheral blood cells? *Breast Cancer Res* **7**(5):E5.
- Li W and Yang Y (2002). How many genes are needed for a discriminant microarray data analysis. In *Methods of Microarray Data Analysis*, pp. 137–150. Kluwer Academic.
- Lipshutz R, Morris D, Chee M, Hubbell E, Kozal M, Shah N, Shen N, Yang R and Fodor S (1995). Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**(3):442–7.
- Liu H and Motoda H (1998). *Feature Selection for knowledge discovery and data mining*. Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061, 2nd edition.
- Loi S, Desmedt C, Cardoso F, Piccart M and Sotiriou C (2005). Breast cancer gene expression profiling: clinical trial and practice implications. *Pharmacogenomics* **6**(1):49–58.
- Lomuto M, Calabrese P and Giuliani A (2004). Prognostic signs in melanoma: state of the art. *J Eur Acad Dermatol Venereol* **18**(3):291–300.
- Lossos I, Czerwinski D, Alizadeh A, Wechser M, Tibshirani R, Botstein D and Levy R (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* **350**(18):1828–1837.
- Massart-Leën A and Massart D (1981). The use of clustering techniques in the elucidation or confirmation of metabolic pathways. Application to the branched-chain fatty acids present in the milk fat of lactating goats. *Biochem J* **196**(2):611–8.
- Metz C (1978). Basic principles of ROC analysis. *Sem Nuc Med* **8**(4):283–298.
- Mocellin S, Rossi C, Pilati P, Nitti D and Marincola F (2003). Quantitative real-time PCR: a powerful ally in cancer research. *Trends Mol Med* **9**(5):189–95.
- Mukherjee S, Tamayo P, Rogers S, Rifkin R, Engle A, Campbell C, Golub T and Mesirov J (2003). Estimating dataset size requirements for classifying DNA microarray data. *J Comput Biol* **10**(2):119–42.
- Müller P, Parmigiani G, Robert C and Rousseau J (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *Journal of the American Statistical Association* **99**(468):990–1001.

- Mulvihill ER, Jaeger J, Sengupta R, Ruzzo WL, Reimer C, Lukito S and Schwartz SM (2004). Atherosclerotic plaque smooth muscle cells have a distinct phenotype. *Arteriosclerosis Thrombosis and Vascular Biology* **24**(7):1283–9.
- Naef F, Hacker C, Patil N and Magnasco M (2002). Empirical characterization of the expression ratio noise structure in high-density oligonucleotide arrays. *Genome Biology* **3**(4):research0018.
- Napoli C, Lerman L, Sica V, Lerman A, Tajana G and de Nigris F (2003). Microarray analysis: a novel research tool for cardiovascular scientists and physicians. *Heart* **89**(6):597–604.
- Neuvians T, Gashaw I, Sauer C, von Ostau C, Kliesch S, Bergmann M, Häcker A and Grobholz R (2005). Standardization strategy for quantitative PCR in human seminoma and normal testis. *J Biotechnol* **117**(2):163–71.
- Notterman D, Alon U, Sierk A and Levine A (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* **61**(7):3124–30.
- Novoradovskaya N, Whitfield M, Basehore L, Novoradovsky A, Pesich R, Usary J, Karaca M, Wong W, Aprelikova O, Fero M, Perou C, Botstein D and Braman J (2004). Universal Reference RNA as a standard for microarray experiments. *BMC Genomics* **5**(1):20.
- Ntzani E and Ioannidis J (2003). Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* **362**(9394):1439–44.
- Nutt C, Mani D, Betensky R, Tamayo P, Cairncross J, Ladd C, Pohl U, Hartmann C, McLaughlin M, Batchelor T, Black P, von Deimling A, Pomeroy S, Golub T and Louis D (2003). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* **63**(7):1602–1607.
- Olson J (2004). Application of microarray profiling to clinical trials in cancer. *Surgery* **136**(3):519–23.
- Pan W (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**(4):546–554.
- Pan W, Lin J and Le CT (2002). How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. *Genome Biol* **3**(5):research0022.
- Park J, Kerbel R, Kelloff G, Barrett J, Chabner B, Parkinson D, Peck J, Ruddon R, Sigman C and Slamon D (2004). Rationale for biomarkers and surrogate end points in mechanism-driven oncology drug development. *Clin Cancer Res* **10**(11):3885–96.
- Park PJ, Pagano M and Bonetti M (2001). A nonparametric scoring algorithm for identifying informative genes from microarray data. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 6, pp. 52–63.

- Pease A, Solas D, Sullivan E, Cronin M, Holmes C and Fodor S (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**(11):5022–6.
- Pepe M, Longton G, Anderson G and Schummer M (2003). Selecting differentially expressed genes from microarray experiments. *Biometrics* **59**(1):133–142.
- Perou C, Brown P and Botstein D (2000). Tumour classification using gene expression patterns from dna microarrays. In *New technologies for life sciences: A Trends Guide*, volume 12, pp. 67–76. Elsevier Science.
- Pfaffl MW, Tichopad A, Prgomet C and Neuvians TP (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper–excel-based tool using pair-wise correlations. *Biotechnol Lett* **26**(6):509–15.
- Poe M, Blake J, Boulton D, Gammon M, Sigal N, Wu J and Zweerink H (1991). Human cytotoxic lymphocyte granzyme B. Its purification from granules and the characterization of substrate and inhibitor specificity. *J Biol Chem* **266**(1):98–103.
- Poulsen C, Borup R, Nielsen F, Borregaard N, Hansen M, Grønbaek K, Møller M and Ralfkiaer E (2005). Microarray-based classification of diffuse large B-cell lymphoma. *Eur J Haematol* **74**(6):453–65.
- Pudil P, Novovic J and Kittler J (1994). Floating search methods in feature selection. *Pattern Recognition Lett* **15**:1119–1125.
- Quinlan J (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, CA, USA.
- Raetz EA and Moos PJ (2004). Impact of microarray technology in clinical oncology. *CANCER INVESTIGATION* **22**(2):312–320.
- Rakotomamonjy A (2003). Variable selection using svm-based criteria. *Journal of Machine Learning Research* **3**(Mar):1357–1370.
- Reid JF, Lusa L, De Cecco L, Coradini D, Veneroni S, Daidone MG, Gariboldi M and Pierotti MA (2005). Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst* **97**(12):927–30.
- Reimers M (2005). Statistical analysis of microarray data. *Addict Biol* **10**(1):23–35.
- Robison J, Perreard L and Bernard P (2004). State of the science: molecular classifications of breast cancer for clinical diagnostics. *Clin Biochem* **37**(7):572–8.
- Rocke D and Durbin B (2001). A model for measurement error for gene expression arrays. *J Comput Biol* **8**:557–569.

- Roepman P, Wessels L, Kettelarij N, Kemmeren P, Miles A, Lijnzaad P, Tilanus M, Koole R, Hordijk G, van der Vliet P, Reinders M, Slootweg P and Holstege F (2005). An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* **37**(2):182–186.
- Rosenwald A, Wright G, Chan W, Connors J, Campo E, Fisher R, Gascoyne R, Muller-Hermelink H, Smeland E, Giltnane J, Hurt E, Zhao H, Averett L, Yang L, Wilson W, Jaffe E, Simon R, Klausner R, Powell J, Duffey P, Longo D, Greiner T, Weisenburger D, Sanger W, Dave B, Lynch J, Vose J, Armitage J, Montserrat E, Lopez-Guillermo A, Grogan T, Miller T, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T and Staudt L (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* **346**(25):1937–1947.
- Ross D, Scherf U, Eisen M, Perou C, Rees C, Spellman P, Iyer V, Jeffrey S, Van de Rijn M, Waltham M, Pergamenschikov A, Lee J, Lashkari D, Shalon D, Myers T, Weinstein J, Botstein D and Brown P (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* **24**(3):227–35.
- Ross M, Mahfouz R, Onciu M, Liu H, Zhou X, Song G, Shurtleff S, Pounds S, Cheng C, Ma J, Ribeiro R, Rubnitz J, Girtman K, Williams W, Raimondi S, Liang D, Shih L, Pui C and Downing JR (2004). Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* **104**(12):3679–87.
- Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, Liu HC, Mahfouz R, Raimondi SC, Lenny N, Patel A and Downing JR (2003). Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102**(8):2951–2959.
- Roth V (2004). The generalized lasso. *IEEE Transactions on Neural Networks* **15**(1):16–28.
- Ruschhaupt M, Huber W, Poustka A and Mansmann U (2004). A compendium to ensure computational reproducibility in high-dimensional classification tasks. *Statistical Applications in Genetics and Molecular Biology* **3**(1):37.
- Saiki R, Scharf S, Faloona F, Mullis K, Horn G and Erlich H (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**:1350–1354.
- Sandvik A, Alsberg B, Nørsett K, Yadetie F, Waldum H and Lægreid A (2005). Gene expression analysis and clinical diagnosis. *Clin Chim Acta* Epub ahead of print.
- 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 (2005). twilight; a Bioconductor package for estimating the local false discovery rate. *Bioinformatics* **21**(12):2921–2922.

- Schmidt U and Begley C (2003). Cancer diagnosis and microarrays. *Int J Biochem Cell Biol* **35**(2):119–24.
- Schmittgen T and Zakrajsek B (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J Biochem Biophys Methods* **46**(1-2):69–81.
- Schneider J, Bitterlich N and Schulze G (2005). Improved sensitivity in the diagnosis of gastro-intestinal tumors by fuzzy logic-based tumor marker profiles including the tumor M2-PK. *Anticancer Res* **25**(3A):1507–15.
- Schulze A and Downward J (2001). Navigating gene expression using microarrays—a technology review. *Nat Cell Biol* **3**(8):E190–5.
- Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neubergh DS, Lander ES, Aster JC and Golub TR (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine* **8**(1):68–74.
- Simon R (2003). Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. *Br J Cancer* **89**(9):1599–604.
- 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.
- Slonim D (2001). Transcriptional profiling in cancer: the path to clinical pharmacogenomics. *Pharmacogenomics* **2**(2):123–36.
- Smyth GK and Speed T (2003). Normalization of cDNA microarray data. *Methods* **31**(4):265–73.
- Sørli T, Perou C, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen M, van de Rijn M, Jeffrey S, Thorsen T, Quist H, Matese J, Brown P, Botstein D, Eystein Lønning P and Børresen-Dale A (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**(19):10869–74.
- Sotiriou C, Neo S, McShane L, Korn E, Long P, Jazaeri A, Martiat P, Fox S, Harris A and Liu E (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* **100**(18):10393–8.
- Spellman P, Sherlock G, Zhang M, Iyer V, Anders K, Eisen M, Brown P, Botstein D and Futcher B (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**(12):3273–97.
- Stewart S, MacIntyre K, Hole D, Capewell S and McMurray J (2001). More ‘malignant’ than cancer? Five-year survival following a first admission for heart failure. *Eur J Heart Fail* **3**(3):315–22.

- Stone M (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B Methodological* **36**(2):111–147.
- Storey JD (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* **31**(6):2033–2035.
- Student (1908). The probable error of a mean. *Biometrika* **6**:1–25.
- Taback B, Morton D, O’Day S, Nguyen D, Nakayama T and Hoon D (2001). The clinical utility of multimarker RT-PCR in the detection of occult metastasis in patients with melanoma. *Recent Results Cancer Res* **158**:78–92.
- Tan, Naiman, Xu, Winslow and Geman (2005). Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics* **21**(20):3896–904.
- Tibshirani R (1996). Regression and shrinkage via the lasso. *Journal of the Royal Statistical Society* **58**(1):267–288.
- Tibshirani R, Hastie T, Narasimhan B and Chu G (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**(10):6567–72.
- Tsai CA, Wang SJ, Chen DT and Chen JJ (2005). Sample size for gene expression microarray experiments. *Bioinformatics* **21**(8):1502–8.
- Tuma R (2004). A big trial for a new technology: TransBIG Project takes microarrays into clinical trials. *J Natl Cancer Inst* **96**(9):648–9.
- 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* **98**(9):5116–5121.
- Ueno S, Ohki R, Hashimoto T, Takizawa T, Takeuchi K, Yamashita Y, Ota J, Choi Y, Wada T, Koinuma K, Yamamoto K, Ikeda U, Shimada K and Mano H (2003). DNA microarray analysis of in vivo progression mechanism of heart failure. *Biochem Biophys Res Commun* **307**(4):771–7.
- Vapnik V (1998). *Statistical Learning Theory*. Wiley, N.Y.
- van ’t Veer L, Dai H, van de Vijver M, He Y, Hart A, Mao M, Peterse H, van der Kooy K, Marton M, Witteveen A, Schreiber G, Kerkhoven R, Roberts C, Linsley P, Bernards R and Friend S (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**(6871):530–6.
- van de Vijver M, He Y, van’t Veer L, Dai H, Hart A, Voskuil D, Schreiber G, Peterse J, Roberts C, Marton M, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers E, Friend S and Bernards R (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**(25):1999–2009.

- Wadlow R and Ramaswamy S (2005). DNA microarrays in clinical cancer research. *Curr Mol Med* **5**(1):111–20.
- Wang A, Doyle M and Mark D (1989). Quantitation of mRNA by the polymerase chain reaction. *Proc Natl Acad Sci U S A* **86**(24):9717–21.
- Wang D, Fan J, Siao C, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris M, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson T, Lipshutz R, Chee M and Lander E (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**(5366):1077–82.
- West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson J, Marks J and Nevins J (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* **98**(20):11462–7.
- Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T and Vapnik V (2001). Feature selection for svms. In *Proceedings of Advances in Neural Information Processing System*, volume 13. MIT Press.
- Wilcoxon F (1945). Individual comparisons by ranking methods. *Biometrics* **1**:80–83.
- Willenbrock H, Juncker A, Schmiegelow K, Knudsen S and Ryder LP (2004). Prediction of immunophenotype, treatment response, and relapse in childhood acute lymphoblastic leukemia using DNA microarrays. *Leukemia* **18**(7):1270–1277.
- Xing E, Jordan M and Karp R (2001). Feature selection for high-dimensional genomic microarray data. In *In the Proceedings fo the Eighteenth International Conference on Machine Learning*.
- Xu M and Setiono R (2003). Gene selection for cancer classification using a hybrid of univariate and multivariate feature selection methods. *Applied Genomics and Proteomics* **2**(2):79–91.
- Yang Y, Buckley M, Dudoit S and Speed T (2000). Comparison of methods for image analysis on cDNA microarray data. Technical report.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J and Speed TP (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucl Acids Res* **30**(4):e15.
- 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.
- Yu L and Liu H (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research* **5**:1205–1224.

Zhang P (1993). Model selection via multifold cross validation. *The Annals of Statistics* **21**(1):299–313.

Zien A, Fluck J, Zimmer R and Lengauer T (2003). Microarrays: How many do you need? *Journal of Computational Biology* **10**(3-4):653–667.