

**Reinhard Meister TFH Berlin**

## **Microarray Daten bei klinischen Studien Klassifikation und Validierung**

### **Vorbemerkung**

Dieser Beitrag entstand in Kooperation mit Ulrich Mansmann (LMU München) Jörn Gröne (Charité Berlin) Ina Koch (TFH Berlin) Bernd Niebank, Thomas Hamborg (ehem. TFH Berlin)

## **1 Einleitung**

Der Einsatz moderner Methoden der Molekularbiologie erlaubt unter anderem die simultane Bestimmung der Aktivität von Genen aus Gewebeproben. So ist es im Prinzip vorstellbar, dass man anhand der Aktivitätsmuster der Gene, man spricht hier von *Genexpression* Zustand und Art der Gewebe differenzieren kann. Wie häufig bei der Einführung neuer Techniken macht sich die Forschung anfangs mehr Hoffnungen auf neue Erkenntnisse, als der Einsatz unter realistischen Praxisbedingungen tatsächlich leisten kann. Wir können in diesem Beitrag nicht auf die molekularbiologischen Hintergründe eingehen, eine gute Einführung auch in die biologischen Grundlagen findet man in *Statistical Science*, Sonderheft Vol 18. (2003).

Für die statistische Methodik sind Genexpressionsdaten eine Herausforderung, weil die Anzahl der simultan bestimmten Expressionswerte im Verhältnis zur Zahl analysierter Proben sehr groß sein kann. Zur Bestimmung der Genexpression wird aufbereitetes Probenmaterial auf so genannte *Genchips*, auch *Microarrays* genannt, gegeben. Die Microarrays sind mit Abschnitten interessierender Gene besetzt. Jedes Gen des Probenmaterials kann so genannte *Messenger RNA* produzieren. Diese wird bei der Aufbereitung der Proben durch *inverse Transkription* in komplementäre cDNA umgewandelt. Nach Verstärkung mit Hilfe der *Polymerase Kettenreaktion* kann die cDNA nun an die zugehörigen Genabschnitte des Microarray binden. Dieser Vorgang heißt *Hybridisierung*. Das aufbereitete Probenmaterial wird eingefärbt, und man kann die Menge der hybridisierten cDNA anhand der Farbintensität quantitativ bestimmen. Wegen der Unterschiede in der Bindungsspezifität und möglicher weiterer systematischer Fehler werden Microarray-Daten zunächst einer Normalisierung unterzogen (Irizarry et al 2003). Die Anordnung der einzelnen Genabschnitte auf dem Genchip hat keinerlei biologische oder sonstige inhaltliche Bedeutung. Wir betrachten hier vor allem Daten, die mit Hilfe des kommerziell hergestellten Oligo-Nucleotid Arrays der Firma Affymetrix bestimmt wurden. Bisher ist die Onkologie das Hauptanwendungsgebiet der Microarray Technik in medizinischen Studien. Seit mehr als fünf Jahren werden Arbeiten zu Diagnose und Prognose von Krebserkrankungen veröffentlicht. Beispiele sind Arbeiten zu Leukämien: Golub et al (1999), Bullinger et al (2004) Lymphomen: Alizadeh et al (2000), Brustkrebs: van't Veer et al (2002), van de Vijver (2002), Huang et al (2003), sowie zum kolorektalen Karzinom: Bertucci et al (2004).

## 2 Fragestellung und Methoden für Microarrays

Wir betrachten hier Situationen, in denen Probenmaterial für zwei Gruppen von Patienten (bzw. Geweben) vorliegt. Dabei interessiert man sich für Gruppenunterschiede in der Genexpression, für die Struktur der Expression in den Gruppen und für eine Klassifikationsregel mit Hilfe der Genexpression. Wichtige methodische Ansätze sind dabei das Statistical Learning (unsupervised) zum Auffinden von Strukturen in der Genexpression, das Statistical Learning (supervised) um Phänotypen zu klassifizieren. Unterschieden zwischen Phänotypen werden mit Hilfe von Hypothesentests untersucht. Ansätze zur Auswahl von Teilgruppen interessierender Gene liefert dabei die Systembiologie.

In diesem Beitrag werden vor allem die Punkte Klassifikation und differentielle Genexpression untersucht. Wir stellen zunächst drei moderne Klassifikationsverfahren vor.

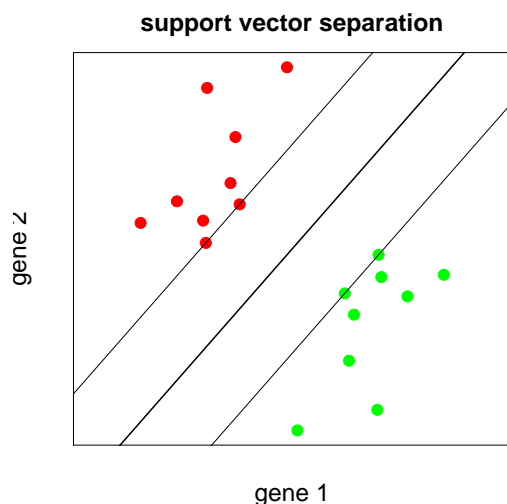
### 2.1 Datenstruktur und Notation

Wir betrachten  $N$  Proben oder Patienten. Für jede Probe liefert ein Microarray Werte der Genexpression von  $p$  Genen. Diese Daten bezeichnen wir mit  $\mathbf{x}_i$ ,  $i = 1, \dots, N$ . Für jede Probe ist außerdem Information über den Phänotyp  $y_i$ ,  $i = 1, \dots, N$  vorhanden.

### 2.2 Klassifizieren: Support Vector Machine

Die Support Vector Machine wurde von Vapnik (1998) eingeführt. Dabei wird eine trennende Hyperebene so bestimmt, dass die durch zwei Gruppen definierte Punktwolken maximalen Abstand bezüglich der Ebene haben. Konkret wird das folgende Optimierungsproblem gelöst:

$$\min(1/2) \|\boldsymbol{\beta}\|^2 \quad \text{NB: } y_i(\mathbf{x}_i\boldsymbol{\beta} + \beta_0) \geq 1 \quad i = 1, \dots, N$$



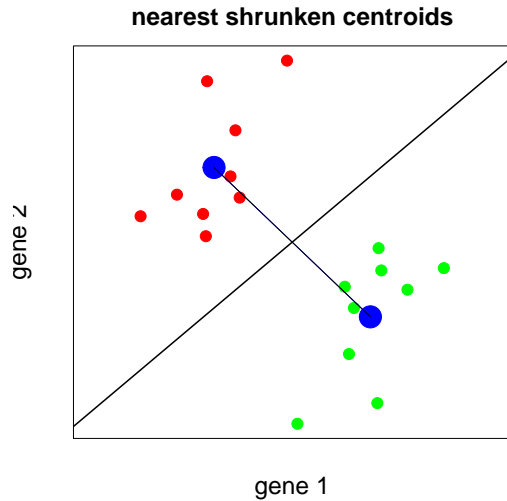
*Schemadarstellung der SVM Methode*

### 2.3 Klassifizieren: Nearest Shrunken Centroids

Hastie, Tibshirani (2003) schlagen eine Verallgemeinerung der linearen Diskriminanzfunktion vor, bei der datenabhängiges Schrumpfen (*shrinkage*) der Abstände zu stabileren Ergebnissen führen soll. Die Zuordnung zur  $k$ -ten Klasse erfolgt mit Hilfe eines modifizierten Abstandsmaßes  $\tilde{\mathbf{d}}$ , dabei werden irrelevant kleine Beiträge komponentenweise

herausgefiltert. Die Schranke  $\Delta$  kann so gewählt werden, dass der geschätzte Generalisierungsfehler minimiert wird.

$$C(\mathbf{x}) = \operatorname{argmin}_k (\|\tilde{\mathbf{d}}^k(\mathbf{x})\|^2) \quad \tilde{\mathbf{d}}_j^k(\mathbf{x}) = \operatorname{sign}(d_j^k) (|d_j^k| - \Delta)_+$$

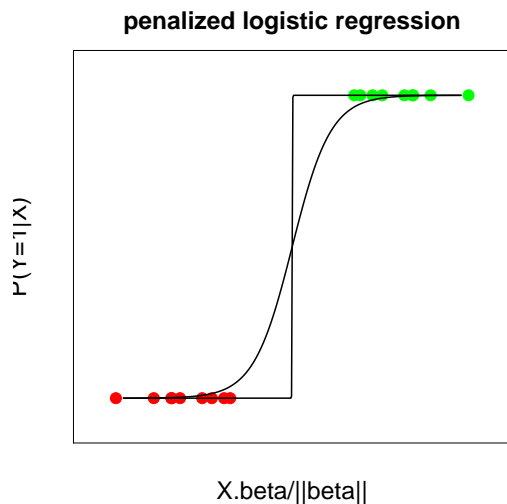


*Schemadarstellung der NSC Methode*

## 2.4 Klassifizieren: Penalized Logistic Regression

Eilers et al (2001) schlagen eine Variante der logistischen Regression vor, bei der die Optimierung durch einen Strafterm, welcher *zu große Parameterwerte* verhindert, erreicht wird.

$$P(\mathbf{y} = 1 | \mathbf{x}) = 1 / (1 + \exp(-\mathbf{x}\boldsymbol{\beta})), \quad \text{Strafterm: } (\lambda/2) \times \|\boldsymbol{\beta}\|^2$$



*Schemadarstellung der PLR Methode*

Sind die Daten trennbar, dann ergibt das Maximum der Likelihood den Wert 1, die Parameterschätzung  $\hat{\beta} = \infty$ . Durch Addition des Strafterm zur Likelihood wird Eindeutigkeit der Parameterschätzung erreicht. Dabei wird ein optimaler Wert für  $\lambda$  durch das AIC - Kriterium bestimmt.

### 3 Schätzung des Generalisierungsfehlers

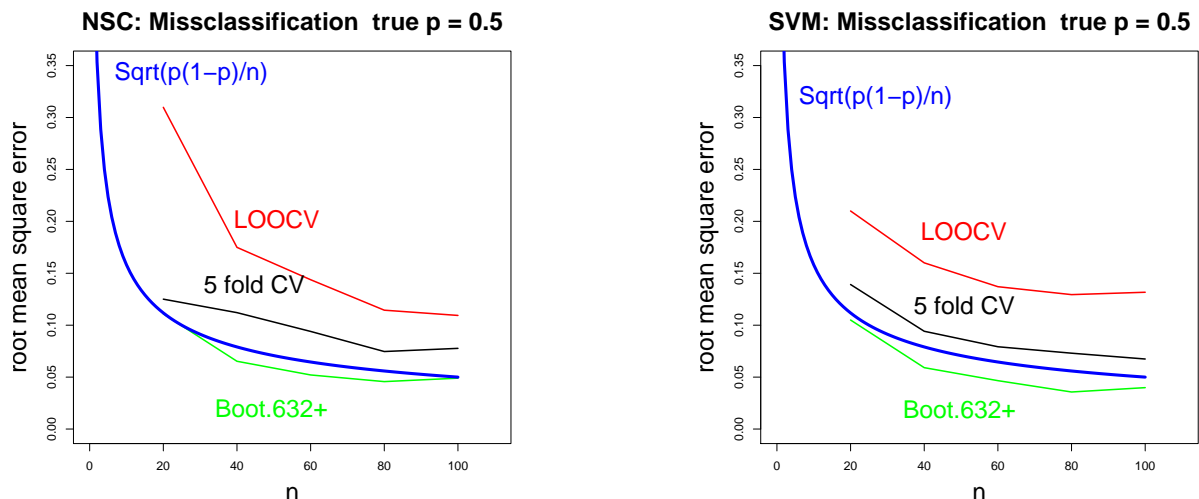
Im Gegensatz zur üblichen Anwendung statistischer Klassifikationsverfahren hat man bei Genexpressionsdaten in der Regel keine Probleme eine hervorragende Zuordnungsregel für die vorliegenden Daten herzuleiten. Dieser Effekt ist auf die enorm hohe Variablenzahl zurückzuführen. Es kommt aber nicht darauf an vorhandene Daten fehlerlos zu klassifizieren, sondern stabile und robuste Regeln zu finden, die sich auf neue Proben übertragen lassen. Ein unverzichtbarer Bestandteil dieses hochdimensionalen Verfahrens ist deshalb die Schätzung des Generalisierungsfehlers. Man wählt Klassifikationsregeln dann so aus, dass der geschätzte Generalisierungsfehler möglichst klein ist.

#### 3.1 Kreuzvalidierungsverfahren

Für die Beurteilung eines Klassifikationsverfahrens ist es notwendig, die Fehlerrate möglichst unverzerrt und mit kleiner Varianz zu schätzen. Verfügt man über neue unabhängige Testdaten vom Umfang  $n$ , dann ist die relative Häufigkeit der Fehlklassifikationen ein unverzerrter Schätzer der Fehlerrate  $p$ , der mittlere quadratische Fehler beträgt  $MSE = p(1 - p)/n$ . Verfügt man nicht über neuen Daten, werden Kreuzvalidierungsverfahren eingesetzt. Die wichtigsten sind LOOCV, 5-fold-, 10-fold-CV (Stone 1974) und Boot.632+ (Efron, Tibshirani 1997). Ambroise, McLachlan (2002), Braga-Neto, Dougherty (2004) weisen in ihren Simulationsstudien zur Validierung von Microarraystudien auf gravierende Mängel vieler Arbeiten bezüglich der kritischen Bewertung des Generalisierungsfehlers hin.

#### 3.2 Simulationen zur Validierung

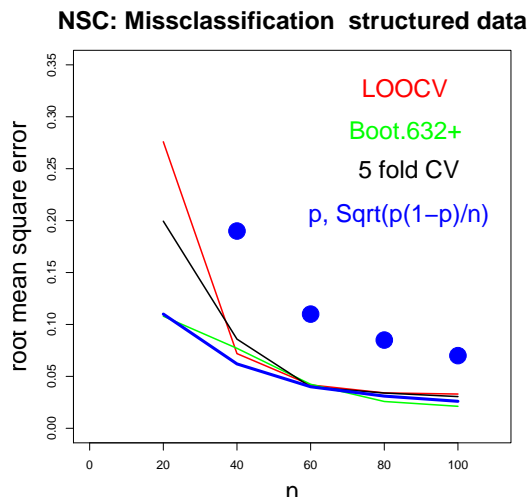
Ausgehend von der Kritik an der Zuverlässigkeit der Klassifikation haben wir (s. Hamborg 2004) zunächst untersucht, ob sich für Daten ohne Struktur, bei denen man eine Fehlklassifikationsrate von 50% bei gleichem Stichprobenumfang in den Gruppen erwarten würde, diese Rate auch mit Hilfe von Kreuzvalidierung erreichen lässt. Der *mean square error*  $MSE$  dient dabei als Gütemaß.



*Generalisierungsfehler: Support Vector Machine und Nearest Shrunken Centroids*

Unsere Simulationsergebnisse zeigen dass eine zuverlässige Schätzung des Generalisierungsfehlers möglich ist. Allerdings muss dabei unbedingt darauf geachtet werden, dass sämtliche Verfahrensschritte bei der Validierung nachvollzogen werden. Die Simulationen erfordern einen hohen Rechenaufwand, wir haben hierzu die Implementierung eines Rechnerclusters auf der Basis der R-Software `snow` (Rossini 2003) verwendet. Es wurden jeweils Daten für einen Gesamtstichprobenumfang  $n = 20, 40, \dots, 100$  für 2000 Expressionwerte teilweise auch korreliert, je Probe erzeugt. Jede Simulation wurde 1000 mal wiederholt. Unsere Simulation weist nach, dass bei einer erwarteten Fehlerrate von 50% beträchtliche Unterschiede in der Präzision der Kreuzvalidierungsverfahren bestehen. Insbesondere erscheint das leaving one out inakzeptabel. Das Verfahren von Efron et al (1997), `Boot.632+`, erreicht das Optimum. Abweichungen vom theoretischen Optimum sind Ungenauigkeit der Simulation.

Um der Verhalten der Verfahren auch unter echter Alternative, bei der eine möglichst gute Klassifikation erreicht werden sollte, zu untersuchen wurden für 40 der 2000 Gene ein Mittelwertsunterschied  $\Delta\mu = 1$  bei einer Korrelation von  $\rho = 0.8$  eingesetzt. Unter diesen Bedingungen ergaben sich folgende Simulationsergebnisse:



*Generalisierungsfehler: Nearest Shrunken Centroids bei Daten mit Struktur*

In der Grafik sind die mittleren Fehlklassifikationsraten als Punkte dargestellt. Anders als im Fall unstrukturierter Daten gelingt es praktisch mit jeder der eingesetzten Methoden, die Rate mit guter Präzision (*Root Mean Square Error*) zu schätzen. Ab einer Gesamtgruppengröße von  $n = 40$  ist nahezu das Optimum, was der möglichen Präzision beim Verwenden neuer Daten entspricht erreicht.

### 3.3 Klassifikationsfehler Beispiel

Die Schätzung der Generalisierungsfehlers ist vor allem ein wichtiges Instrument, um die Relevanz von Studienergebnissen einzuschätzen. Teilweise sind unsere Erkenntnisse noch nicht ausreichend, und die Messdaten noch nicht präzise genug, um zu medizinisch relevanten validen Ergebnissen zu gelangen.

In einer Studie zum Kolonkarzinom (Gröne, Mansmann, Meister et al) erhielten wir folgende Ergebnisse bei der Klassifikation von Mikro-Schnitten von Tumorgewebe in zwei unterschiedliche Stadien (UICC II und III).

Methode	Ergebnisse		
	Fehler %	Vorauswahl	Validierung
Shrunken Centroid	30	5000 Gene	CV
Support Vector	20	50–70 Gene	CV
Logistic Regression	17	15000 Gene	shrinkage

## 4 Globale Tests auf differentielle Genexpression

Im Rahmen von Studien zur Klassifikation auf der Basis von Genexpressionsdaten ist der Einsatz statistischer Hypothesentests ein nützliches Hilfsmittel. Wegen der großen Anzahl von Genen und dem damit verbundenen Problem des simultanen Testens, sind insbesondere solche Tests von Interesse, die eine globale Hypothese bezüglich des Verhaltens einer ganzen Gruppe von Genen prüfen. Interessant ist insbesondere die Analyse so genannter Signaturen, das sind Gruppen von Genen, die als Basis einer Klassifikation dienen. Globale Tests können unter anderem zum

- Check von Signaturen
- Vergleich mit Signaturen anderen Arbeiten
- Prüfung auf Unterschiede zwischen Phänotypen in Pathways

eingesetzt werden.

### 4.1 Die Ansätze von Goeman und Mansmann & Meister

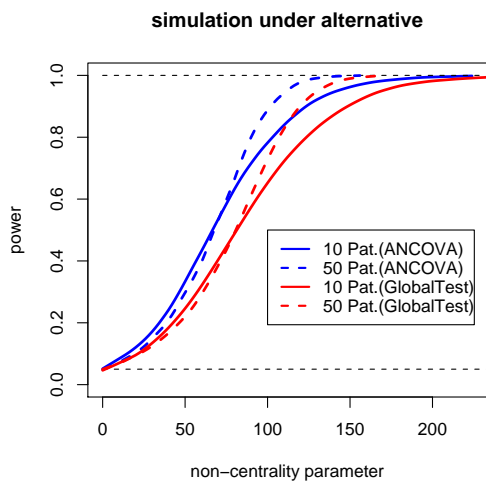
Mit den Bezeichnungen  $Y$  für Phänotyp und Daten für die Genexpressionsdaten können wir zwei aktuelle Ansätze beschreiben.

- $H_0 : P(Y = 1|\text{Daten}) = P(Y = 2|\text{Daten})$   
Technik: Random Coefficient Generalized Linear Model  
Goeman et al (2004) R-package globaltest
- $H'_0 : P(\text{Daten}|Y = 1) = P(\text{Daten}|Y = 2)$   
Technik: Mittelwertvergleich  
Mansmann, Meister (2005), R-package GlobalAncova

Beide Hypothesen sind äquivalent (Bayes Theorem). Das Prinzip des ANCOVA Ansatzes wird im Folgenden erklärt.

- Ein Gen:  $H_0 : \mu_1 = \mu_2$
- Alle Gene:  $H_0 : \mu_{1j} = \mu_{2j}, j = 1, \dots, N_{\text{genes}}$
- Ansätze
  - multiples Testen: Dudoit et al (2002)
  - Kovarianzanalyse zur Untersuchung von  $\sum(\mu_{1j} - \mu_{2j})^2$   
Modell:  $y_{ijk} = \mu_{ij} + \sum \beta_{jl}x_{lik} + e_{ijk}$   
Kovariableneffekte genspezifisch
  - Verteilung unter  $H_0$  aus Permutationen

Wir haben die Gütefunktion der beiden Tests mit Hilfe einer Simulationsstudie für ausgewählte Datenstrukturen unter der Alternative untersucht (Niebank 2004, Diplomarbeit TFH Berlin). Beide Tests halten das Niveau, unser Verfahren (Mansmann, Meister 2005) erzielte dabei stets eine bessere Power als der Ansatz von Goeman (2004). Die Unterschiede sind allerdings häufig nicht gravierend. Unser Ansatz zum globalen Mittelwertvergleich läßt sich auf Phänotypen, deren Erwartungswert je Gen als Linearkombination von Einflußgrößen darstellbar ist erweitern. Die globale Hypothese testet auf identische Erwartungswerte der Phänotypen je Gen, Haupteffekt des Phänotyps und die Wechselwirkung mit dem Gen-Effekt werden gleichzeitig getestet. Wegen der möglichen Korrelationen zwischen Genexpressionswerten eines Patienten, wird ein Permutationstest verwendet.



*Vergleich der Gütefunktion des Goeman Ansatzes und des ANCOVA Permutationstests*

Die Abbildung zeigt die erreichte Güte in Abhängigkeit vom Nichtzentralitätsparameter der Zählerquadratsumme der Teststatistik. Unsere Untersuchungen ergaben *bei unkorrelierten Daten* gleiche Power der beiden Ansätze und volle Effizienz bezüglich der klassischen Normaltheorie. Die Abbildung zeigt die Power bei 100 Genen, von denen zwei Blöcken vom Umfang 20 Genen korreliert  $\rho = 0.9$  und differentiell exprimiert sind.

## 4.2 Anwendungsbeispiel

In einer Studie zum kolorektalen Karzinom untersuchen Gröne et al verschiedene Pathways auf differentielle Genexpression. Mit dem ANCOVA Ansatz können wir für den Cytochrome p53 - Signalling Pathway signifikante Unterschiede zwischen den Tumorstadien UICC II und UICC III nachweisen. Ferner läßt die Adjustierung auf zwei Kovariable, die für sich allein keinen signifikanten Unterschied zwischen den Tumorstadien aufwiesen, den Effekt deutlicher hervortreten.

	SS	DF	MS	F.value	p.value.perm
Total	4740.37616	1619	2.9279655	NA	NA
Genes adjusted	4502.83829	44	102.3372338	NA	NA
GroupXGenes	14.72441	45	0.3272091	2.246857	0.0121
Residual	222.81346	1530	0.1456297	NA	NA

Table1: ANCOVA table, sequential decomposition of SSQ, not adjusted for covariates.

	SS	DF	MS	F.value	p.value	perm
Total	4740.37616	1619	2.9279655	NA	NA	NA
Genes adjusted	4516.40461	134	33.7045120	NA	NA	NA
GroupXGenes	18.07929	45	0.4017619	2.809902	0.0019	
Residual	205.89226	1440	0.1429807	NA	NA	NA

Table2: ANCOVA table, sequential decomposition of SSQ, adjusted for sex and location of the tumour.

## 5 Diskussion

Die Hoffnungen auf neue medizinische Erkenntnisse, die man mit Hilfe der Microarray Technik erhalten kann sind sehr groß und zeigen sich in der Vielzahl von Studien und Publikationen zum Thema. Für die Statistik bedeutet die hohe Dimension des Merkmalsvektors der Daten eine wirkliche Herausforderung. Es wird vor allem darauf ankommen, durch Validierung der Ergebnisse mehr Zuverlässigkeit und Reproduzierbarkeit zu erzielen.

- Microarray Daten ermöglichen neuartige, medizinisch relevante Diagnosen.
- Vermutlich lassen sich die zugrunde liegenden Klassifikationsregeln zuverlässig und effizient validieren.
- Zuverlässige Schätzung des Generalisierungsfehlers liefern ein modifizierter Bootstrap-Ansatz bzw mehrfach Kreuzvalidierungsverfahren.
- Globale Tests ergänzen Klassifikation durch die Überprüfung von Gensignaturen.

## Literaturangaben

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* , 403:503-511.(2000)
2. Affymetrix [<http://www.affymetrix.com>]
3. Ambroise C and McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci* 99(10): 6562-6 (2002)
4. Bertucci F, Salas S, Eysteris S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene*. 23: 1377-91, 2004.
5. Braga-Neto, U., Dougherty, E. Is cross-validation valid for small-sample microarray classification? *Bioinformatics*, Vol. 20, No. 3, 374-380.(2004)
6. Breiman, L. Random Forests *Machine Learning J.* 45: 5-32.(2001)
7. Breiman L, Cutler A, Liaw A, Wiener M randomForest: Breiman's random forest for classification and regression. Fortran original by Leo Breiman and Adele Cutler and R port by Andy Liaw and Matthew Wiener. R package 4.3-3 <http://statwww.berkeley.edu/users/breiman/RandomForests>



8. Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. Bullinger L, Döhner K, Bair E, Fröhling S, Schlenk RF, Tibshirani R, Döhner H, Pollack JR. Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. *N Engl J Med*. Volume 350:1605-1616 (2004)
9. Dudoit S, Shaffer JP, Boldrick JC, Multiple Hypothesis Testing in Microarray Experiments. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 110.  
<http://www.bepress.com/ucbbiostat/paper110> (2002)
10. Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. *JASA* 92: 548-560. (1997)
11. Eilers PH, Boer JM, Van Ommen GJ, Van Houwelingen HC. Classification of microarray data with penalized logistic regression. *Proceedings of SPIE volume 4266: progress in biomedical optics and imaging* 2:187-198 (2001)
12. Fredriksen CM, Knudsen S, Laurberg S, Orntoft TF. (2003) Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 129: 263-271. (2003)
13. Gentleman R and Carey V. Bioconductor. *R News* 2(1):11-16, (2002)
14. Goeman JJ, Van de Geer SA, De Kort F, van Houwelingen JC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, 20: 93-99 (2004)
15. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286:531-537.
16. Hamborg T. Evaluierung der Generalisierungsfähigkeit moderner Klassifikationsverfahren für Genexpressionsdaten mit einem Rechnercluster. Diplomarbeit TFH Berlin 2004
17. Hastie, T., Tibshirani, R., Friedman, J. *The Elements of Statistical Learning*. Springer, New York. (2001)
18. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR, Huang AT. Gene expression predictors of breast cancer outcomes. *The Lancet* 361:1590-1596 (2003).
19. Irizarry, R.A., Gautier, L., Cope, L.M. (2003) *The Analysis of Gene Expression Data: Methods and Software*, chapter 4. Springer Verlag.
20. Mansmann U, Meister R. Testing differential gene expression in functional groups: Goeman's global test versus an ANCOVA approach. *Methods of Information in Medicine* (2005, in press)
21. Niebank B. Zur Güte globaler Tests auf differentielle Genexpression - Simulationen mit Hilfe eines Rechnernetzwerkes. Diplomarbeit TFH Berlin 2004
22. Rossini A, Tierney L, Li N Simple Parallel Statistical Computing in R Paper 193, UW Biostatistics working paper series.  
<http://www.bepress.com/uwbiostat/paper193/>(2003)

23. Ruschhaupt M, Huber W, Poustka A, Mansmann U. A compendium to ensure computational reproducibility in high-dimensional classification tasks. *Statistical Applications in Genetics and Molecular Biology*: Vol. 3: No. 1, Article 37. (2004) <http://www.bepress.com/sagmb/vol3/iss1/art37>
24. Stone M. Cross-validated choice and assessment of statistical predictions. *J.Roy.Statist.Soc.* 36: 111-147. (1974)
25. Tibshirani RJ and Efron B. Pre-validation and inference in microarrays *Statistical Applications in Genetics and Molecular Biology* Vol.1, Article 1 (2002)
26. Tibshirani R, Hastie T, Narasimhan B, Chu G. Class prediction by nearest shrunken centroids, with application to DNA microarrays. *Statistical Science* 18:104-117 (2003)
27. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT et al.: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, 415:530-536.
28. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* Vol.347:1999-2009 (2002)
29. West M. Bayesian factor regression models in the Large p, Small n" paradigm *Bayesian Statistics*, 7, 723-732 (2003)