

A. Ahr
U. Holtrich
C. Solbach
A. Scharl
K. Strebhardt
T. Karn
M. Kaufmann

Molekularbiologische Klassifikation von Hochrisiko-Patientinnen beim primären Mammakarzinom durch Genexpressionsanalysen

Gene Expression Profiling for Molecular Identification of High-Risk Breast Cancer Patients

Zusammenfassung

Fragestellung: Für viele Tumorarten existieren Subklassen, welche durch definierte genetische Marker weiter differenziert werden könnten, um eine individuelle Therapie zu ermöglichen. Mittels einer neuen molekularbiologischen Untersuchungsmethode, der DNA-Array-Technik, werden Genexpressionsprofile erstellt. Es ist zu erwarten, dass durch die Messung der gesamten zellulären transkriptionellen Aktivität Genexpressionsprofile identifiziert werden, welche mit dem klinischen Verlauf, z. B. beim Mammakarzinom, korrelieren.

Methode: Wir untersuchten mittels cDNA-Array-Hybridisierung (bis zu 45 000 Klone) in Kombination mit Real Time RT-PCR-Analysen die Genexpressionsprofile bei primären Mammakarzinomen an einem Kollektiv von 73 Gewebeproben (T1–4, N0–1, M0–1). Die Genexpressionsprofile der Mammakarzinome wurden nachfolgend für eine molekulare Tumorklassifikation mittels Gruppenanalyse herangezogen.

Ergebnisse: Durch diese Analyse („Class Discovery“) konnten Subgruppen innerhalb des Gesamtkollektivs identifiziert werden, deren Signifikanz durch Kreuzvalidierung bestätigt wird. Die Korrelation der molekularen Tumorcharakterisierung mit den klassischen klinisch/pathologischen Parametern (TNM) zeigte in der einen Gruppe (n = 16) einen Anteil von fast 90% nodal-positiver Tumorproben (14 von 16). Weiterhin auffallend war in dieser Gruppe die Akkumulation von Tumorproben mit bereits erfolgter Metastasierung zum Zeitpunkt der Primärdiagnose (25% [4 von 16] vs. 4% [3 von 57]). In einer anderen, kleine-

Abstract

Objective: Many tumors are subclassified with genetic markers to improve therapy and follow-up. A goal is to use DNA arrays as a tool to understand and classify tumors on the basis of gene expression patterns. Global determination of cellular transcriptional activity may identify gene expression signatures that predict the clinical behavior of tumors.

Methods: We performed DNA array studies and subsequent cluster analyses to establish a molecular profile of 73 breast cancer specimens (T1–4, N0–1, M0–1).

Results: Our analyses identified a group of transcriptionally related specimens. Fourteen of these 16 tumors were node positive. This group had an accumulation of patients with distant metastases at the time of diagnosis (25% vs. 4%).

Conclusions: Differentially expressed marker genes in conjunction with sample clustering algorithms provide a novel molecular classification system for breast cancers. This may help identify patients at high risk of recurrence.

Key words

DNA array · Tumor classification · Cluster analysis · Differentially expressed genes · Breast cancer

Institutsangaben

Bereich „Molekulare Gynäkologie und Geburtshilfe“ der Universitäts-Frauenklinik Frankfurt

Korrespondenzadresse

Priv.-Doz. Dr. med. A. Ahr · Universitäts-Frauenklinik Frankfurt · Theodor-Stern-Kai 7 · 60590 Frankfurt · E-mail: Ahr@em.uni-frankfurt.de

Eingang Manuskript: 11. 7. 2001 · **Eingang revidiertes Manuskript:** 8. 11. 2001 · **Akzeptiert:** 12. 11. 2001

Bibliografie

Geburtsh Frauenheilk 2001; 61: 954–963 © Georg Thieme Verlag Stuttgart · New York · ISSN 0016-5751

ren Subgruppe (n = 6) fanden sich ausschließlich hormonrezeptor-negative Mammakarzinome.

Schlussfolgerung: Mit der hier beschriebenen Gruppenanalyse kann anhand der Genexpressionsprofile ein Risikokollektiv beim Mammakarzinom identifiziert werden. Es ist zu erwarten, dass durch Erweiterung dieser multifaktoriellen Genexpressionsanalysen künftig exakte Vorhersagen des biologischen Verhaltens von Tumoren und damit patientinnenadaptierte Therapien möglich sind.

Schlüsselwörter

DNA-Chiptechnologie · Differenziell exprimierte Gene · Microarray · Mammakarzinom · Molekulare Tumorklassifikation · Cluster-Analyse

Einführung und Fragestellung

In Deutschland wird derzeit jede 9. Frau in ihrem Leben an einem Mammakarzinom erkranken. Unter der Annahme, dass es sich beim Mammakarzinom um eine systemische Erkrankung handelt, wurde die operative Radikalität in den letzten Jahrzehnten reduziert und parallel dazu die medikamentöse Systemtherapie intensiviert. Aufgrund der etablierten Prognosefaktoren (Tumorgroße, Differenzierungsgrad, Lymphknotenstatus, Alter und Hormonrezeptorstatus) wird für ca. 90% der Patientinnen eine adjuvante Chemo- und/oder Hormontherapie empfohlen (Konsensuskonferenz von St. Gallen zur adjuvanten Therapie des Mammakarzinoms, 2001 und [1]). Trotz aller Fortschritte der Therapie erleiden immer noch über die Hälfte der Patientinnen ein Rezidiv. Alle bislang in Einzeluntersuchungen analysierten Marker reichen nicht aus, um den individuellen Krankheitsverlauf (= prognostische Bedeutung) und das Ansprechen auf bestimmte medikamentöse Therapien (= prädiktive Bedeutung) zuverlässig vorherzusagen. Es besteht daher die Notwendigkeit, neue Techniken zu entwickeln, die weitergehende Klassifikationen von Mammakarzinomen zur Abschätzung des individuellen Risikos und Nutzen einer Therapie ermöglichen.

Die Klassifikation von Tumoren lässt sich in zwei Abschnitte gliedern [2]: 1. Die Definition neuer Klassen („Class Discovery“), die mit distinkten Phänotypen von Tumoren korrelieren, sowie 2. die Zuweisung von Proben zu diesen Klassen und ihre gegenseitige Abtrennung („Class Prediction“). Als Parameter einer molekularen Tumorklassifikation eignet sich in besonderer Weise die Messung der zellulären Genexpression (mRNA), da diese durch biologische Methoden einfach und verlässlich durchgeführt werden kann. Die Fertigstellung des humanen Genomprojektes [3] ermöglicht es nunmehr, mit Hilfe der DNA-Array-Technologie die Expression sämtlicher Gene in Zelllinien und Gewebeproben parallel zu analysieren [2, 4–8]. Aus diesen Expressionsprofilen sind diejenigen Gengruppen (= „informative Gene“) durch geeignete mathematische Verfahren zu filtern, die mit klinisch relevanten Eigenschaften des Tumors wie Wachstumsverhalten, metastatischem Potenzial oder Chemosensibilität korrelieren.

Ziel der vorliegenden Studie ist es, die differenziell exprimierten Gene beim Mammakarzinom mittels cDNA-Array-Hybridisierung zu identifizieren. Anhand der Genexpressionsprofile soll

anschließend eine molekularbiologische Klassifizierung durchgeführt werden, um eine zuverlässige Aussage über den individuellen Krankheitsverlauf zu erhalten.

Material und Methoden

Patientinnen und Tumorgewebe

Alle Patientinnen mit primärem Mammakarzinom wurden zwischen Juni 1997 und Oktober 1998 an der Universitäts-Frauenklinik behandelt. Von jeder Patientin wurde Nativgewebe vom Primärtumor und ggf. axillären Lymphknotenmetastasen bei – 70°C gelagert. Jede Probe wurde durch Schnellschnitt hinsichtlich des Tumorgehalts kontrolliert. Weiterhin wurde von allen Patientinnen formalinfixiertes und in Paraffin eingebettetes Material (Primärtumor, normales Drüsengewebe, Lymphknotenmetastase) asserviert. Das histologisch verifizierte normale Mammaparenchym (= Referenzgewebe) stammt von gesunden Patientinnen, welche eine Reduktionsplastik erhielten. Für die Detektion der differenziell exprimierten Gene mittels cDNA-Arrays wurde zunächst Tumorgewebe von 16 primären Mammakarzinomen (+ 2 zugehörigen axillären Lymphknotenmetastasen) untersucht (T1–4, N0–1, M0–1, GI–III, ER/PR-positiv und negativ, Histo. Typ: duktal/lobulär/tubulär, siehe auch [9]). Die nachfolgende Gruppenanalyse mittels Real Time RT-PCR-Analyse (TaqMan-Analyse) erfolgte an einem Kollektiv von 9 benignen und 73 malignen Mammagewebeproben (T1–4, N0 [n=28], N1 [n=45], M0 [n=66], M1 [n=7]), sowie weiteren Proben wie Zelllinien (n=3: MCF-7, MDA, A431), befallenen Lymphknoten (n=3), Uterusgewebe (n=2) und einem duktalem Carcinoma in situ (DCIS, n=1), die nicht in die statistischen Auswertungen Eingang fanden. Sowohl bei der Untersuchung der ersten 16, als auch bei den nachfolgenden 73 primären Mammakarzinomproben, handelt es sich um invasive duktales und tubuläre Mammakarzinome entsprechend dem T1–4-, N0–1- und M0–1-Status. Der Tumordifferenzierungsgrad des 2. Kollektivs (n=73) beinhaltet alle 3 Stufen (GI: n=13, GII: n=35, GIII: n=25) und es wurden hormonrezeptor-negative (n=19) und -positive Proben (n=54) untersucht.

Zellaufreinigung

Für vergleichende Expressionsstudien von Tumorgewebe und normalem Mammaepithel werden möglichst reine Zellpopulationen benötigt. Der Reinheitsgrad sollte mindestens bei 90% liegen [10]. Wie bereits beschrieben [9], wurde das Tumor- und Referenzgewebe, nach mechanischer Desintegration (MediMachine, DAKO, Hamburg), zweimal über immunomagnetisches Purgung mittels des monoklonalen Antikörpers HEA 125 (mAk gegen Epitheliales Glycoprotein, Miltenyi Biotec, Bergisch Gladbach) aufgereinigt. Die Zellsuspension wurde für 1 h bei 4°C mit dem mAk HEA 125 inkubiert. Anschließend erfolgte, nach Waschen mit PBS/EDTA, die Inkubation mit einem an Magnetkügelchen gekoppelten Schaf-Anti-Maus-IgG-Antikörper und die Passage durch das MACS (magnetic cell separation system, Miltenyi Biotec, Bergisch Gladbach). Der Reinheitsgrad der isolierten Epithelzellen wurde mikroskopisch verifiziert. Die Prozedur der Zellaufreinigung kann natürlich einen Einfluss auf die Expressionsmuster der Zellen ausüben. Untersuchungen zu diesem Thema zeigten jedoch, dass bei experimentell einheitlichem Vorgehen auch dieser Einfluss einheitlich ist und daher bei gleichbehandelten Proben ein relativer Vergleich zueinander möglich ist.

RNA-Isolation und cDNA-Array-Hybridisierung

Die Gesamt-RNA der aufgereinigten Tumorproben und der Referenzgewebe wurde durch die Guanidinium-Isothiocyant-Methode [11] in Kombination mit Affinitäts-Reinigungsschritten (RNeasy, Qiagen, Hilden) isoliert. Anschließend erfolgte die cDNA-Synthese mittels Reverser Transkription unter Verwendung des Nukleotids [α - 32 P]-dATP zur radioaktiven Markierung. Verfahren wurde hierbei entsprechend dem Protokoll des Herstellers der Array-Membranen (Clontech, Palo Alto, CA, USA). Die so markierten Sonden wurden in einer Konzentration von 1×10^6 dpm/ml mit den in einer Hybridisierungslösung (Clontech) befindlichen Array-Membranen inkubiert. Nach der Hybridisierung über Nacht, wurden die Nylonmembranen in $0,1 \times \text{SSC}/0,5\%$ SDS bei 68°C gewaschen und anschließend auf Röntgenfilme exponiert. Die Autoradiogramme wurden densitometrisch vermessen (Personal Densitometer, Molecular Dynamics, Sunnyvale, CA) und die Rohdaten mittels der Bildanalyse-Software ImageQuant (Molecular Dynamics) ausgewertet, in Tabellenkalkulationen transferiert und über Kalibrierungsmarker normalisiert. Alle Low-density-Arrays (Human cancer blot, 588 Gene) stammen von Clontech; die High-density-Arrays (Human GDA mit 45 000 cDNA-Klonen) wurden von GenomeSystemsInc (St. Louis, MO) bezogen (siehe auch [9]).

Real Time RT-PCR-Analyse

Die RT-PCR (Reverse Transcriptase PCR) erfolgte in Echtzeit mit dem ABI 7700 Sequenz-Detektions-System (PE-Applied Biosystems, Foster City, CA). Die PCR-Reaktionen wurden entsprechend den Herstellerangaben durchgeführt. VIC-markierte GAPDH TaqMan-Sonden dienten als interne Quantifizierungsmarker in multiplex geführten PCR-Reaktionen. Alle Bestimmungen erfolgten dreifach. Die Sequenzen der Primer und TaqMan-Sonden waren (5'-3'): ErbB2: Up AATGAGGACTTGGGCCAGC, Low CAGATACTCCTCAGCATCCACCAGG, Sonde FAM-CAGCACCTTCTACCGCTCACTGC-TAMRA; IGFBP-5: Up TACTCCCCAAGATCTTCCGGCC, Low TTCTGCGGTCCTTCTTCACTGC, Sonde FAM-CCCGCATCTCCGAGCTGAAGGC-TAMRA; MMP-14: Up TGCCGAGGGCTTCCATGGC GAC, Low GCCCTGGGAAGTAGGCATGG, Sonde FAM-AGCCGCC TCACCATCGAAGGGC-TAMRA; EGR-1: Up GTGCCGCATCTGCA TGCGCAAC, Low GCTTTTCGCTGTGTGGGTGGC, Sonde FAM-CAGCCGACGACCTCACC-TAMRA. Die cDNA für alle PCR-Reaktionen wurden durch randomisiert geprimte reverse Transkription (ProSTAR cDNA-Synthese Kit, Stratagene, La Jolla, CA) generiert.

Gruppenanalyse und Validierung

Für die Gruppenanalyse wurden die Expressionsdaten log-transformiert und für jede Probe zehnmal median-zentriert. Mit Hilfe des Programms CLUSTER [12] wurden die Proben aufgrund dieser Daten mittels Pearson-Korrelation (Average Linkage Clustering) gruppiert. Die berechneten relativen Abstände (1-Korrelationskoeffizient) werden grafisch durch das TREEVIEW-Programm [12] dargestellt. Die Methode der „Class Prediction“ wurde, wie von Golub et al. [2] vorgeschlagen, zur Validierung der Klassenunterscheidungen eingesetzt. Kurzgefasst wird ein PS-Wert („Prediction Strength“) für die Zuordnung einer Probe zu einer spezifischen Klasse berechnet durch $PS = (V_{win} - V_{lose}) / (V_{win} + V_{lose})$, wobei es sich bei V um den Absolutbetrag der Summe aller Wertungen („votes“) für die Klassen (hier willkürlich mit „win“ und „lose“ bezeichnet) handelt. Eine gewichtete Wertung für eine der Klassen wird für jedes Gen durch $v_i = P_1 \times (x_i -$

$(\mu_1 + \mu_2)/2)$ kalkuliert. Hierbei wird die Korrelationsmaßinheit P für ein Gen g und eine Klasse c definiert durch $P(g, c) = (\mu_1 - \mu_2) / (\sigma_1 + \sigma_2)$, wobei μ bzw. σ den Mittelwert bzw. die Standardabweichung der beiden Klassen darstellen. Zur Kreuzvalidierung wurde jeweils eine Probe aus einer Klasse entfernt, der P(g,c)-Wert für jedes Gen aufgrund der restlichen Proben der entsprechenden Klasse kalkuliert und anschließend der PS-Wert für die zuvor entfernte Probe berechnet. Dieser Vorgang wird für alle Proben der entsprechenden Klasse wiederholt, um ein Maß für die Streuung der PS-Werte zu erhalten. Zur Abschätzung der statistischen Signifikanz der Klassenabtrennung wurden in 500 Iterationen randomisierte Klassen gebildet, jeweils Kreuzvalidierungen aller Klassenmitglieder durchgeführt und die medianen PS-Werte für jede Iteration bestimmt.

Die statistischen Berechnungen für den Vergleich der molekularbiologischen mit der bisherigen TNM-Klassifikation erfolgte mit dem Statistik-Programm SSPS (Statistical Package for the Social Sciences, München). Univariate Analysen wurden mit dem χ^2 -Test berechnet und ein $p < 0,05$ als statistisch signifikant definiert.

Ergebnisse

Der experimentelle Ablauf einer molekularen Tumorklassifikation auf der Basis von Genexpressionsprofilen gliedert sich in 4 Abschnitte:

- Identifikation von „informativen“ Markergenen (= differenziell exprimierte Gene)
- Testen der Marker an größeren Kollektiven auf ihren prognostischen/prädiktiven Informationsgehalt und Erstellung von Markerprofilen
- Klassifizierung von Tumorproben anhand der Profile,
- Vergleich der neuen Klassifizierung mit der etablierten Klassifizierung.

Detektion differenziell exprimierter Gene beim Mammakarzinom

Zur Ermittlung differenziell exprimierter Gene beim Mammakarzinom wurden radioaktiv markierte cDNAs von insgesamt 16 Mammakarzinomproben und 2 Referenzgeweben gegen Low- und High-density-Arrays (Atlas™ Human Cancer cDNA Expression Arrays von Clontech und GDA 1.3 von GenomeSystemsInc) mit 588 bzw. 45 000 Genen hybridisiert und autoradiographiert. Die Autoradiogramme wurden anschließend densitometrisch vermessen, normalisiert und Genexpressionsunterschiede als Signalverhältnis von Tumor- gegen Referenzgewebe bestimmt. Die Validierung der ermittelten Array-Expressionsdaten wurde im Folgenden durch Real Time RT-PCR-Analysen von 4 zufällig ausgesuchten differenziell exprimierten Genen durchgeführt. Das Ergebnis eines Vergleichs zwischen Array und Real Time RT-PCR-Analyse ist in Abb. 1 dargestellt. Es zeigen sich nur geringe Quantifizierungsunterschiede zwischen beiden Methoden. Um falsch-positive Genexpressionsunterschiede auszuschließen, muss ein Schwellenwert (= cut-off) für die Bewertung definiert werden. Dieser Schwellenwert ergibt sich aus der Analyse verschiedener Areale der gleichen Probe. Wie in Abb. 2 gezeigt, weisen etwa ein Drittel von 30 zufällig ausgewählten Genen im Vergleich verschiedener Areale einer Gewebeprobe geringe Expres-

sionsunterschiede von 2 – 3fach auf. Daher wurden für die weiteren Analysen nur Expressionsunterschiede berücksichtigt, die signifikant über dem Schwellenwert lagen (≥ 5 fach).

Transkriptionelle Diversität von Mammakarzinomen

Die Bestimmung der transkriptionellen Diversität von Karzinomproben ist für deren weitere Charakterisierung von entscheidender Bedeutung. Aus ihr ergibt sich, wie viele Probenanalysen notwendig sind, um geeignete Marker für die Klassifizierung von Tumorkollektiven zu selektieren. In Tab. 1 wird der Anteil der differenziell exprimierten Gene beim Vergleich Mammakarzinom gegenüber Referenzgewebe dargestellt: Bei einem cut-off ≥ 5 fach wurden 6,4% der Gene differenziell exprimiert. Wird der Grenzwert auf ≥ 20 fach erhöht, finden sich nur noch 0,9% differenziell exprimierte Gene. Somit unterscheiden sich die Tumoren gegenüber normalem Mammaparenchym nur durch eine relativ geringe Anzahl an Expressionsunterschieden. Bestätigt werden diese Ergebnisse der Low-density-Arrays (588 Gene) durch die weiterführenden Untersuchungen mit High-density-Arrays, die 45 000 humane cDNA-Klone enthalten: In zwei Vergleichen identifizierten wir ca. 100 Gene mit ≥ 20 fachen Expressionsunterschieden, was einem Anteil von 0,2% der Klone mit Hybridisierungssignal gegenüber dem Hintergrund bedeutet.

Um die Zahl der Array-Analysen von Gewebeproben zu ermitteln, die ausreicht, den überwiegenden Teil der differenziell exprimierten Gene zu identifizieren, wurde die kumulative Zahl dieser Gene gegen die Anzahl der analysierten Proben aufgetragen. Wie in Abb. 3 ersichtlich, erreichen die Kurven, unabhängig des Cut-off-Wertes, nach ca. 10 Proben eine Sättigung. Aus der Abbildung ist ersichtlich, dass ca. 20 Tumورproben für eine zuverlässige Analyse ausreichen.

Für die Tumorklassifizierung werden Marker benötigt, die folgende Kriterien erfüllen: deutliche Expressionsunterschiede über einem definierten cut-off (hier ≥ 5 fach) zwischen den einzelnen Proben und die Häufigkeit der Veränderungen. Die differenziell exprimierten Gene sollten sich bei möglichst vielen Mammakarzinomen nachweisen lassen. Es wurden daher die mittels Low- und High-density-Array detektierten differenziell exprimierten Gene mit häufigen Veränderungen ausgewählt. Zusätzlich wurden einige, aktuell in der Literatur diskutierte Gene

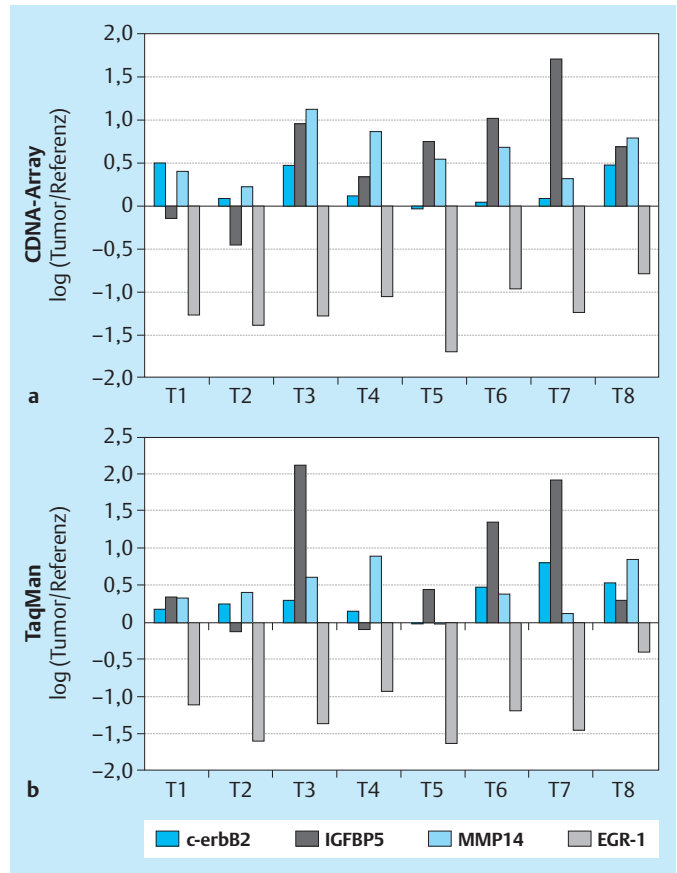


Abb. 1a und b Vergleich der Ergebnisse von cDNA-Array-Analysen und Real Time RT-PCR (TaqMan). a Verhältnis der durch cDNA-Array-Analysen ermittelten Expressionshöhen von vier willkürlich ausgewählten differenziell exprimierten Genen bei acht Tumорproben (Kollektiv siehe Ahr et al. 2000, Tumорproben #1–#8) im Vergleich zu benignen duktalem Mammapithelzellen (= Referenzgewebe). b Die gleichen Proben und Gene wie in (a) wurden mittels Real Time RT-PCR (TaqMan-Analyse) untersucht. Wie in (a) sind auch hier die Verhältnisse der Expressionshöhen von Tumor, gegenüber normalen Epithelzellen dargestellt.

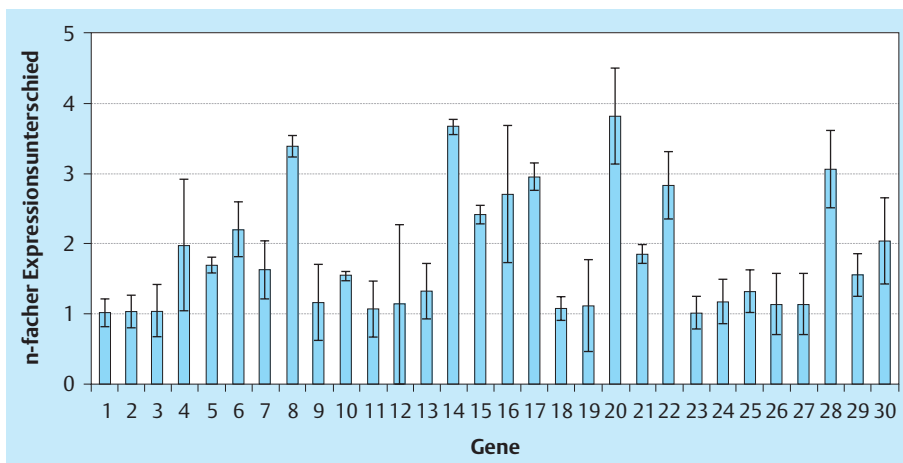


Abb. 2 Fluktuation der Genexpression innerhalb von Gewebeproben. Verifikation von Genexpressionsunterschieden, die durch cDNA-Array-Analysen verschiedener Bereiche derselben Gewebeprobe erhalten wurden. Real Time RT-PCR wurde eingesetzt, um die Expressionsunterschiede von 30 zufällig ausgewählten humanen Genen zu bestimmen. Beispielhaft wird die Variabilität (mit Standardabweichung) in der Genexpression zwischen zwei Bereichen einer Gewebeprobe gezeigt.

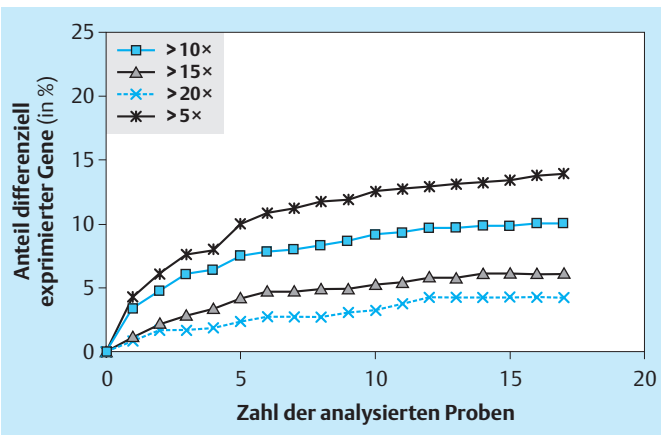


Abb. 3 Analyse der transkriptionellen Diversität von Mammakarzinomproben. Eine Serie von randomisiert ausgewählten Mammakarzinomproben wurde durch cDNA-Array-Analysen untersucht und die kumulative Zahl der detektierten Genexpressionsunterschiede gegen die Anzahl untersuchter Proben aufgetragen. Die Ergebnisse für vier verschiedene Schwellenwerte bezüglich der Expressionsunterschiede sind gezeigt.

Tab. 1 Differenziell exprimierte Gene beim Mammakarzinom. Mammakarzinomproben sowie benigne duktales Mammapithelzellen wurden mit cDNA-Arrays (588 Gene) analysiert. Unterschiede in der Genexpression wurden als Signalverhältnis von Tumor- zu Normalzellen bestimmt. Der Anteil differenziell exprimierter Gene ist für verschiedene Schwellenwerte angegeben. Die Ergebnisse stellen Mittelwerte von 7 Analysen dar

Genexpressionsdifferenz	Anteil unterschiedlich exprimierter Gene
≥ 5 fach	6,4% (SD \pm 1,27)
≥ 10 fach	2,7% (SD \pm 0,92)
≥ 15 fach	1,5% (SD \pm 0,55)
≥ 20 fach	0,9% (SD \pm 0,29)

SD = Standardabweichung

im Zusammenhang mit dem Mammakarzinom in die Analyse mit einbezogen. Diese so genannten „informativen“ Gene, wie in Tab. 2 dargestellt, stellen ein Anfangspanel dar und können in Zukunft durch zusätzliche Gene erweitert werden.

Molekulare Tumorklassifizierung durch Genexpressionsprofile

Die mittels Array-Hybridisierung gewonnenen Genexpressionsdaten können zu Gruppierungen von Tumoren genutzt werden, welche ihre biologischen Eigenschaften reflektieren [12–15]. Wir analysierten die Genexpressionsprofile bei 73 malignen Mammagewebebeben mittels des Klassifizierungsprogramms CLUSTER [12]. Der Algorithmus des Programms sortiert die Gewebeproben entsprechend ihrer Expressionsprofile. Das Ergebnis der Analyse wird graphisch mit Hilfe des Programms TREEVIEW [12] dargestellt.

Tab. 2 Molekulare Tumorklassifizierung durch Genexpressionsprofile. Auflistung der häufig alterierten differenziell exprimierten Gene beim primären Mammakarzinom, die für die Klassifizierung verwendet wurden. Ausgewählt wurden differenziell exprimierte Gene der Low- und High-density-Arrays bzw. zusätzlich in die Analyse aufgenommene Gene der Literatur

Gen	accession-number	Low-density-Array	High-density-Array	Literatur
<i>IGFBP-6</i>	M62402	X		
<i>ErbB2</i>	M11730	X		
<i>Ki67</i>	NM_002417	X		
<i>PLK</i>	NM_005030			[7, 25]
<i>EST-HSM800383</i>	AL050276		X	
<i>Egr-1</i>	X52541	X		
<i>AKAP450</i>	AJ131693		X	
<i>ZNF217</i>	AF041259		X	
<i>FGFR 4</i>	X59932	X		
<i>Beta-act.</i>	X00351	X		
<i>BRCA2</i>	U43746			[19]
<i>GADD45</i>	M60974			[14, 22]
<i>EB1</i>	U24166	X		
<i>PDGF-assoc.-prot.</i>	U41745	X		
<i>Estrogen receptor</i>	X03635			[12]
<i>Progesteron receptor</i>	X51730			[12]
<i>BAD</i>	U66879	X		
<i>IGFBP2</i>	M35410	X		
<i>EST-AI817868</i>	R85813		X	
<i>Semaphorin-V</i>	U33920	X		
<i>STAT5 b</i>	U47686	X		
<i>IGFBP-5</i>	M65062	X		
<i>Tpl-2/Cot</i>	NM_005204		X	
<i>TNFR 1</i>	M33294	X		
<i>EST-HS465208</i>	T89015		X	
<i>VEGFR</i>	X51602	X		
<i>CD24</i>	L33930	X		
<i>Stat1</i>	M97935	X		
<i>DDR2</i>	X74764			[10, 24]
<i>THBS2</i>	L12350	X		
<i>MMP-14</i>	D26512	X		
<i>NDKA/nm23</i>	X17620	X		
<i>EST-AA928060</i>	AA029434		X	
<i>HEK2</i>	X75208	X		
<i>HTK</i>	U07695	X		
<i>EST-AA679039</i>	N54493		X	
<i>Estrogen receptor-beta-cx</i>	AB006589			[9]
<i>FEZ1</i>	AF123659			[8]
<i>MRP6</i>	AF076622		X	
<i>RGC1</i>	X78817	X		
<i>RhoC</i>	L25081	X		

Wie in Abb. 4 dargestellt, konnte in dem vorliegenden Probenkollektiv von 73 primären Mammakarzinomen, 9 normalen Mammaparenchymproben, 3 Zelllinien und weiteren Gewebeproben (siehe Material und Methoden) durch diese Analyse vier Hauptgruppen (I–IV) identifiziert werden. Die größte Gruppe (III) teilt sich in zwei Untergruppen, welche aus Tumorproben mit einer relativen transkriptionellen Homogenität (A, n = 26) besteht und einer kleineren Gruppe von Tumorproben mit einer geringeren Genexpressionshomologie (B, n = 6). Die geringere Distanz der Arme des Dendrogramms von A gegenüber B stellt die stärkere transkriptionelle Übereinstimmung dar. Die Gruppe A kann bei weiterer Untersuchung in die Gruppe A1 und A2 unterteilt werden. Diese Unterteilung wird auch sichtbar in der Matrix (Abb. 4, links), wenn einzelne Gene auf ihre transkriptionelle Aktivität untersucht werden. In Gruppe A2 wird im Vergleich zu A1 sowohl der Östrogen- und Progesteronrezeptor, als auch das proapoptische BAD-Gen und IGFBP2 geringer exprimiert. Eine weitere Auffälligkeit von A ist, dass sowohl in allen Tumorproben von A1 als auch von A2 eine Down-Regulation des erst kürzlich beschriebenen ER- β -cx-Rezeptors [16] stattgefunden hat. In Gruppe B sind alle Tumorproben hormonrezeptor-negativ (sowohl Östrogen-, als auch Progesteronrezeptor-negativ).

Die Korrelation der Genexpressionsdaten mit den klassischen histomorphologischen Parametern (TNM) zeigte interessanterweise bei dem Vergleich der Gruppe A (n = 26) und dem Restkollektiv (n = 47) eine signifikante Korrelation mit dem Nodal- und Metastasenstatus (siehe Tab. 3). In der Gruppe A fanden sich signifikant häufiger nodal-positive Mammakarzinome (20/26 vs. 25/47; p = 0,046) und signifikant mehr primär metastasierte Mammakarzinome zum Zeitpunkt der Primärdiagnose (Gruppe A: 5/26 [19%] gegenüber Restkollektiv: 2/45 [4%]; p = 0,038). Bei weiterer Differenzierung von Gruppe A in A1 (n = 16) und A2 (n = 10) zeigt sich, dass 14 von 16 Patientinnen in Gruppe A1 einen positiven Lymphknotenstatus aufwiesen im Vergleich zu 53% im restlichen Tumorkollektiv (Gruppe A1: 14/16 [88%] vs. Gruppe A: 20/26 [77%] vs. A2: 6/10 [60%] vs. Restkollektiv: 25/47 (53%); p = 0,015, siehe auch Tab. 3). Dargestellt ist der positive Lymphknotenstatus anhand der magentafarbenen Punkte am Dendrogramm in Abb. 4. Weiterhin auffallend ist in Gruppe A1 die signifikante Anhäufung von primär metastasierten Mammakarzinomen im Vergleich zum Restkollektiv (A1: 4/16 [25%] vs. Restkollektiv: 2/47 [4%]; p = 0,015) und hormonrezeptor-positiven Mammakarzinomen (15/16 [94%] vs. 35/47 [74%]; p = 0,01). Im Gegensatz dazu finden sich in Gruppe A2 vermehrt hormonrezeptor-negative Mammakarzinome (A2: 6/10 [69%] vs. Restkollektiv: 12/47 [26%]; p = 0,034). In Abb. 4 sind die Tumorproben mit bereits erfolgter Metastasierung zum Zeitpunkt der Primärdiagnose des Mammakarzinoms durch schwarze Pfeile markiert.

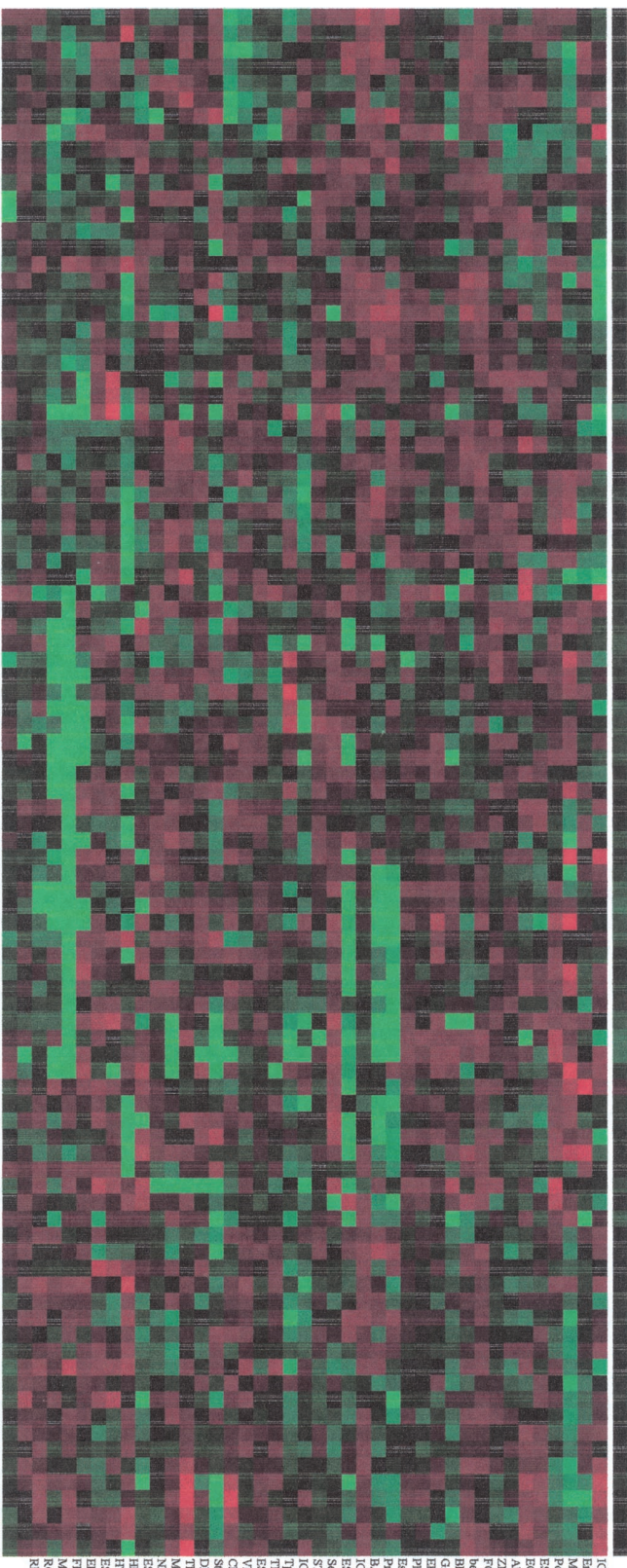
Eine weitere molekularbiologisch klassifizierte Gruppe ist Gruppe B, welche aus 6 primären Mammakarzinomen besteht. In dieser Gruppe konnte keine Korrelation mit dem Lymphknotenstatus, Grading oder Metastasenstatus im Vergleich zum Gesamtkollektiv nachgewiesen werden, jedoch wiesen sämtliche Proben dieser Gruppe einen negativen Hormonrezeptor-Status auf (6/6 [100%] vs. 13/54 [26%]; p < 0,001).

Die Tumorgröße und der histologische Subtyp zeigte innerhalb der Gruppenbildung keine statistisch signifikante Korrelation

(p > 0,05). Unter den übrigen Tumoren war bei dieser molekularbiologischen Analyse keine weitere Gruppenbildung zu beobachten. Die Referenzgewebe sind entsprechend ihrer Genexpressionsprofile zusammengruppiert (Abb. 4).

Die von Golub et al. [2] vorgeschlagene Methode der „Class Prediction“ wurde herangezogen, um die ermittelten Klassenunterscheidungen zu validieren und die Konsistenz der Gruppen zu überprüfen. In Abb. 5a ist die Verteilung von PS-Werten („Prediction Strength“ nach [2]) einer Kreuzvalidierung der Proben von Gruppe A gegenüber den nicht zu A oder B gehörenden malignen Proben gezeigt. Der mediane PS-Wert von 0,73 ist durch einen vertikalen Balken symbolisiert. Die beiden anderen Verteilungswolken zeigen PS-Werte, die zufälligen Klasseneinteilungen entsprechen (Median 0,21 bzw. 0,31). Insgesamt wurden 500 solche zufälligen Klasseneinteilungen der Proben analysiert, um die statistische Signifikanz der Klasse A zu evaluieren. Abb. 5b zeigt die Verteilung der medianen PS-Werte. Der höchste beobachtete mediane PS-Wert unter diesen 500 zufälligen Iterationen lag bei 0,52. Der mediane PS-Wert für die Unterscheidung zwischen Subgruppe A1 und A2 liegt bei 0,54 (nicht gezeigt) und auch hier liegen die PS-Werte der meisten Proben über dem von Golub et al. [2] vorgeschlagenen Schwellenwert von 0,3.

Abb. 4 Klassenbildung von primären Mammakarzinomen aufgrund von Expressionsprofilen. Variation der Genexpression von 41 differenziell exprimierten Genen in 73 primären Mammakarzinomproben. Die Daten werden in einem Matrixformat dargestellt: jede Spalte repräsentiert ein Gen und jede Zeile eine Gewebeprobe. Der Vergleich Probe/Gen wird dargestellt als Ratio der Abweichung der transkriptionellen Diversität jedes Gens zu der medianen Abweichung der transkriptionellen Abweichung aller Gewebeproben und wird durch das entsprechende Farbfeld in der Matrix repräsentiert. Die grünen Quadrate entsprechen transkriptionellen Leveln unterhalb des Median, schwarze Quadrate entsprechen einem transkriptionellen Level gleich dem Median und rote Quadrate zeigen einen transkriptionellen Level oberhalb des Medians an. Die Farbintensität reflektiert die Höhe der Ratio relativ zu dem Median für jeden Vergleich Probe/Gen. Das Dendrogramm auf der rechten Seite der Abbildung zeigt die Ähnlichkeit der Gewebeproben in Bezug auf ihre Genexpressionsmuster. Tumorproben mit positivem Nodalstatus (n = 45) sind durch eine magenta- und nodal-negative Probe (n = 28) durch einen schwarzen Punkt gekennzeichnet. Die Pfeile stellen Mammakarzinomproben mit bereits erfolgter Metastasierung zum Zeitpunkt der primären Diagnose dar (n = 7). Zu beachten ist die Gruppenbildung von Proben in den Klassen A1 und A2 (Balken in magenta und violett). Der Prozentsatz von nodal-positiven Patientinnen in Klasse A1 liegt mit 88% (14 von 16) signifikant höher als im restlichen Tumorkollektiv (53%, 30 von 57). Die Anhäufung von Patientinnen mit Fernmetastasen zum Zeitpunkt der Primärdiagnose in der Klasse A1 bzw. A ist an den schwarzen Pfeilen zu erkennen. Insgesamt finden sich 4 Tumorproben mit Metastasierung in der Gruppe A1 (25%, 4 von 16) bzw. 5 Tumorproben in Gruppe A (19%, 5 von 26), im Vergleich zum Restkollektiv mit 4% (2 von 47). Eine zweite Gruppe von Mammakarzinomen wurde in Klasse B (schwarzer Balken) identifiziert, die ausschließlich hormonrezeptor-negative Tumorproben beinhaltet. Die gutartigen Mammaproben sind durch gelbe Balken hervorgehoben.



IGFBP-6	M62902
ERBB2	M11730
MAK67	NM_002417
Pol-alkelinase	NM_005030
ES1-HSN/800033	AL050276
RCR-1	X52541
AKA490	AKA490
ZNF217	A111093
PRK4	A081259
PRK4	U14174
PRK4	X00351
RRC42	U43746
GADD154	M60974
IBL1	U24166
PDGF-abc - prot.	U41745
Estrogen receptor	X03635
Pregnen. receptor	X51720
BAD	U66879
IGFBP2	M38410
ES1-AS1/7888	R38813
Springbrunh-V	U33920
STRE12	U33920
STRE12	M65902
TN1.2/Cc	NM_002204
TNFR1	M33294
ES1-HSN/65208	T89015
VEGFR	X51602
CD24	L33930
Shal1	M97935
DDR2	X74764
THRS2	L12350
NAPRT4	D26512
NKX2-1/mx23	X17620
HER2	X17620
HER2	X75208
HTK	U07055
ES1-AA679039	N54693
BR-beta-ex	AB005889
FEZ1	AF124659
MFB6	AF076622
RGC1	X78817
RHC	L29081

41 differenziell exprimierte Gene
= „informative Gene“

Gewebebeobachten – 73 primäre Mammakarzinome und Referenzgewebe

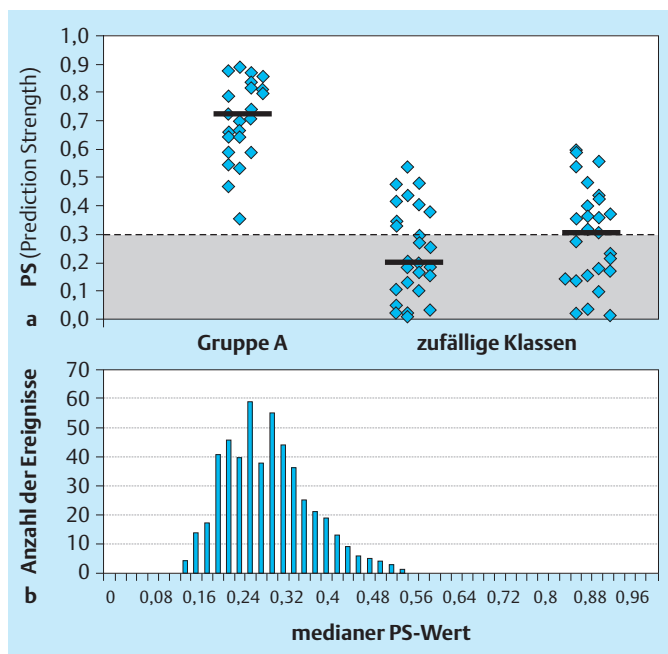


Abb. 5a und b Validierung der Klassenbildung durch „Class-Prediction“. **a** Die Verteilung der PS-Werte in Kreuzvalidierungsanalysen ist dargestellt. Die erste Streukurve zeigt das Ergebnis einer Kreuzvalidierung der Klasse A mit einem medianen PS-Wert von 0,73 (vertikaler Balken). Die beiden rechten Streukurven zeigen Ergebnisse von Kreuzvalidierungen zufällig gebildeter Klassen (Median 0,21 bzw. 0,31). **b** Insgesamt 500 solche randomisierte Klassen wurden aus den Proben gebildet und durch Kreuzvalidierung untersucht. Die Grafik zeigt die Verteilung der erhaltenen medianen PS-Werte.

Diskussion

Seit einigen Jahren werden differenzielle Genexpressionsanalysen mittels verschiedener molekularbiologischer Techniken durchgeführt. Die Analyse pathologisch veränderter Genaktivität kann zur Aufklärung von Krankheitsmechanismen, Verbesserung der Diagnostik und zur Identifizierung neuer therapeutischer Angriffspunkte beitragen. In den vergangenen Monaten rückte die DNA-Array-Technologie zunehmend in das Zentrum des wissenschaftlichen Interesses. Neue Automatisierungstechniken und die Identifizierung aller humanen Gene im Rahmen der Genomprojekte ermöglichen die Herstellung hochkomplexer Nukleinsäure-Biochips, mittels derer die komplexen transkriptionellen Änderungen im Rahmen der malignen Zelltransformation analysierbar werden.

Das Hauptziel unserer Studie war es, differenziell exprimierte Gene beim Mammakarzinom zu identifizieren und diese nachfolgend für eine molekulare Klassifizierung von Tumoren heranzuziehen. Es wurden zunächst systematische Genexpressionsanalysen bei Mammakarzinomen durchgeführt, die Aufschluss geben sollten über die transkriptionelle Diversität und die Zahl der Gewebeprobe, die untersucht werden müssen, um die Mehrzahl der differenziell exprimierten Gene zu detektieren. Die Genexpressionsanalyse von Mammakarzinomgewebe und dem korrespondierenden Referenzgewebe durch cDNA-Array-Hybridisierung zeigte, dass die meisten Gene mit nahezu gleicher Expressionsstärke exprimiert werden. Vergleichbar niedrige Werte der transkriptionellen Diversität wurden auch bei der Analyse anderer Tumoren beobachtet [10]. Zhang et al. [10] untersuchten die Genexpressionsunterschiede zwischen primären Kolonkarzinom- und normalen Kolonepithelzellen mittels SA-GE-Analyse und beschrieben einen Anteil von ca. 2% transkrip-

Tab. 3 Korrelation der molekularbiologischen Klassenbildung anhand der Genexpressionsdaten mit den klassischen histomorphologischen Parametern (TNM-Klassifikation)

	Gesamtkollektiv n = 73	Kollektiv ohne A n = 47	Gruppe A n = 26	Gruppe A1 n = 16	Gruppe A2 n = 10	Gruppe B n = 6
Tumorgröße T1 vs. T2	12 vs. 61 (16%)	7 vs. 40 (15%)	5 vs. 21 (19%)	3 vs. 13 (19%)	2 vs. 8 (20%)	0 vs. 6 (0%)
			p = 0,632	p = 0,716	p = 0,688	p = 0,311
Nodalstatus N1 vs. N0	45 vs. 28 (61%)	25 vs. 22 (53%)	20 vs. 6 (77%)	14 vs. 2 (88%)	6 vs. 4 (60%)	2 vs. 4 (33%)
			p = 0,046	p = 0,015	p = 0,695	p = 0,360
Metastase M1 vs. M0	7 vs. 66 (10%)	2 vs. 45 (4%)	5 vs. 21 (19%)	4 vs. 12 (25%)	1 vs. 9 (10%)	0 vs. 6 (0%)
			p = 0,038	p = 0,015	p = 0,46	p = 0,607
Grading GIII vs. GI/II	25 vs. 48 (34%)	13 vs. 34 (28%)	12 vs. 14 (46%)	4 vs. 12 (25%)	8 vs. 10 (80%)	3 vs. 3 (50%)
			p = 0,111	p = 0,836	p = 0,196	p = 0,262
Hormonrezeptor ER/PR- vs. ER/PR+	19 vs. 54 (26%)	12 vs. 35 (26%)	7 vs. 19 (27%)	1 vs. 15 (6%)	6 vs. 4 (60%)	6 vs. 0 (100%)
			p = 0,897	p = 0,01	p = 0,034	p < 0,001
Histologie nicht duktal vs. duktal	15 vs. 58 (21%)	11 vs. 36 (23%)	4 vs. 22 (15%)	2 vs. 14 (13%)	2 vs. 8 (20%)	0 vs. 6 (0%)
			p = 0,417	p = 0,352	p = 0,816	p = 0,184

p-Wert < 0,05 = statistisch signifikant; Korrelation der Gruppen A, A1 und A2 jeweils gegen Gesamtkollektiv ohne Gruppe A und der Gruppe B gegen Gesamtkollektiv ohne Gruppe B

tionell alterierten Genen. Unsere Ergebnisse lassen den Schluss zu, dass die Untersuchung von 20 Tumoren ausreicht, um die meisten der transkriptionell veränderten Gene in Mammakarzinomen zu detektieren. Somit dürften die hier beschriebenen High-density-Array-Analysen mit 45 000 Genen (rund 50% des humanen Genoms) bislang einen Bruchteil von etwa 10–20% aller differenziell exprimierten Gene identifiziert haben. Diese Markergene wurden in weitergehenden Analysen für die molekulare Klassifizierung eines größeren Mammakarzinomkollektivs herangezogen.

Die derzeit eingesetzten histomorphologischen/klinischen Parameter der Tumorklassifikation sind nicht ausreichend, um zwischen Subtypen mit unterschiedlichem klinischen Verlauf zu unterscheiden. Es wird angenommen, dass der Vergleich von Tumorgenexpressionsprofilen eine äußerst genaue Beschreibung des zellulären Status quo zulässt und es ermöglichen wird, diese Subtypen zu detektieren und in einer klinisch sinnvollen Weise zu klassifizieren [17].

Für die molekulare Klassifizierung von Tumorproben („Class Discovery“) auf Basis von Genexpressionsprofilen eignen sich sowohl die so genannten „unsupervised“ Methoden wie „Hierarchisches Clustering“ z. B. mit Hilfe der hier verwendeten Pearson-Korrelation [12], „k-Means Clustering“ [13] oder „SOMs“ („Self-Organizing Maps“ [14]) als auch „supervised“ Methoden wie „SVMs“ („Support Vector Machines“ [15]). Hierarchisches Clustering wurde bereits erfolgreich zur Probenklassifizierung anhand von Genexpressionsprofilen eingesetzt [4, 6]. Alizadeh et al. [4] konnten distinkte, bisher unbekannt Typen von B-Zell-Lymphomen identifizieren, die sich in ihrem klinischen Verlauf unterscheiden. Golub et al. [2] verwendeten SOMs, um mittels DNA-Array-Analysen das Krankheitsbild der akuten Leukämie zu differenzieren. Es gelang ihnen, anhand von 50 „informativen Genen“ korrekt die unterschiedliche Behandlungsregimen erfordernden Subtypen ALL und AML zu klassifizieren, aber auch neue Subgruppen zu identifizieren. Neueste Daten von Scherf et al. [8] und Ross et al. [7] belegen, dass durch Kombination von Klassenbildung aufgrund von Chemotherapie-Sensibilität und Genexpression auch neue Ansätze zur Therapieauswahl möglich werden.

In der vorliegenden Studie konnten an einem Kollektiv aus 73 malignen Mammagewebeproben (T1–4, N0–1, M0–1) anhand der Genexpressionsmuster Subgruppen identifiziert werden, welche sich durch ihre TNM-Parameter unterscheiden. In der Gruppe A1 fanden sich im Vergleich zum restlichen Tumorkollektiv vermehrt nodal-positive Mammakarzinome (88% vs. 53%). Auffallend war hier vor allem eine Akkumulation von Tumorproben mit bereits erfolgter distanter Metastasierung zum Zeitpunkt der Primärdiagnose (4 von 7). Patientinnen dieser Gruppe repräsentieren demzufolge ein Hochrisikokollektiv, deren Tumoren ein biologisches Verhalten zeigen, das eine frühe Disseminierung vitaler, proliferationsprofizienter Tumorzellen in die Körperperipherie ermöglicht. Wir versuchen derzeit, durch Einbindung weiterer „informativer“, differentiell exprimierter Gene, den Vorhersagewert in der Gruppenanalyse zu erhöhen. Da bislang nur ein Bruchteil aller in Mammakarzinomen differenziell exprimierter Gene in die Gruppenanalyse Eingang fanden, ist zu erwarten, dass sich in allernächster Zukunft ein größerer Teil von Karzinomen über „Class Discovery“ und „Class Prediction“ klinisch charakterisieren lässt.

Zusammengefasst lassen sich mit der neuen cDNA-Array-Technologie die transkriptionellen Genaktivitäten von bereits beschriebenen, aber auch von weitestgehend unbekannt Markern, beim Mammakarzinom qualitativ und quantitativ untersuchen. Der Vorteil und das Neue der vorgestellten cDNA-Array-Analysen liegt in der multiplexen und korrelativen Genexpressionsanalyse, die in Verbindung mit „Class Discovery und Prediction“ ein einzigartiges Werkzeug zur molekular-pathologischen Beschreibung von Erkrankungen repräsentiert. Somit steht nicht mehr, wie bisher, der Einzelmarker in Korrelation zum klinischen Verlauf im Vordergrund der wissenschaftlichen Untersuchungen, sondern das Genexpressionsprofil eines Tumors aus einer Vielzahl von Genprodukten (hier 45 000 Expressionssignale).

Fazit für die Praxis

Es ist durch die Massen-Genexpressionsanalyse erstmals möglich, die zellulären interaktiven Regulationsmechanismen (Genüberexpression oder -verlust) auf RNA-Ebene parallel darzustellen. Der Vergleich verschiedener Mammakarzinomproben lässt erkennen, dass ihre Genexpressionsmuster sehr ähnlich sind und sich nur bezüglich relativ weniger Markergene unterscheiden. Durch die Gruppierung der Tumoren anhand der Genexpressionsprofile, ist es möglich, Hochrisiko-Mammakarzinompatientinnen zu identifizieren. In welchem Maße dies eine zuverlässige prädiktive und/oder prognostische Aussage erlaubt, wird zur Zeit weiter untersucht.

Literatur

- Ahr A, Holtrich U, Karn T, Solbach C, Göhring U-J, Scharl A, Strebhardt K, Kaufmann M. Detektion differentiell exprimierter Gene beim primären Mammakarzinom mittels cDNA-Array-Hybridisierung. *Geburtsh Frauenheilk* 2000; 60: 412–417
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabel H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503–511
- Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussier D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 2000; 97: 262–267
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 1979; 18: 5294–5299
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; 95: 14,863–14,868
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286: 531–537
- Holtrich U, Wolf G, Brauning A, Karn T, Bohme B, Rubsamen-Waigmann H, Strebhardt K. Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors. *Proc Natl Acad Sci USA* 1994; 91: 1736–1740
- Ishii H, Baffa R, Numata SI, Murakumo Y, Rattan S, Inoue H, Mori M, Fidanza V, Aider H, Croce CM. The FEZ1 gene at chromosome Bp22 encodes a leucine-zipper protein, and its expression is altered in multiple human tumors. *Proc Natl Acad Sci USA* 1999; 96: 3928–3933
- Jarvinen TA, Pelto-Huikko M, Holli K, Isola J. Estrogen receptor beta is coexpressed with ERalpha and PR and associated with nodal status, grade, and proliferation rate in breast cancer. *Am J Pathol* 2000; 156: 29–35

- ¹⁰ Kam T, Holtrich U, Brauning A, Bohme B, Wolf G, Rubsamen-Waigmann H, Strebhardt K. Structure, expression and chromosomal mapping of TKT from man and mouse: a new subclass of receptor tyrosine kinases with a factor VIII-like domain. *Oncogene* 1993; 8: 3433–3440
- ¹¹ Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci USA* 2000; 97: 1778–1783
- ¹² Kaufmann M, Jonat W. Therapie des primären Mammakarzinoms. *Dt Ärzteblatt* 1998; 95: 1977–1980
- ¹³ Kibertis PA. Letter to the editor. *Science* 2000; 287: 933
- ¹⁴ Li S, Ting NS, Zheng L, Chen PL, Ziv Y, Shilon Y, Lee EY, Lee WH. Functional link of BRCA1 and ataxia teleangiectasia gene product in DNA damage response. *Nature* 2000; 406: 210–215
- ¹⁵ Ogawa S, Inoue S, Watanabe T, Orimo A, Hosoi T, Ouchi Y, Muramatsu M. Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor oestrogen action in human. *Nucleic Acids Res* 1998; 26: 3505–3512
- ¹⁶ Perou CM, Sorlie T, Eisen MB, van de Rijan M, Heffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature* 2000; 406: 747–752
- ¹⁷ Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000; 24: 227–235
- ¹⁸ Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, Weinstein JN. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000; 24: 236–244
- ¹⁹ Schmutzler RK, Bierhoff E, Werkhausen T, Fimmers R, Speiser P, Kubista E, Krebs D, Zeilinger R, Wiestler OD, von Deimling A. Genomic deletions in the BRCA1, BRCA2 and TP53 regions associate with low expression of the estrogen receptor in sporadic breast carcinoma. *Int J Cancer* 1997; 74: 322–325
- ²⁰ Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999; 96: 2907–2912
- ²¹ Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999; 22: 281–285
- ²² Vairapandi M, Azam N, Balliet AG, Hoffman B, Liebermann DA. Characterization of MyD118, Gadd45, and proliferating cell nuclear antigen (PCNA) interacting domains. PCNA impedes MyD118 AND Gadd45-mediated negative growth control. *J Biol Chem* 2000; 275: 16,810–16,819
- ²³ Venter JC, et al. The sequence of the human genome. *Science* 2001; 291: 1304–1351
- ²⁴ Vogel W. Discoidin domain receptors: structural relations and functional implications. *FASEB J* 1999; 13: 77–82
- ²⁵ Wolf G, Elez R, Doermer A, Holtrich U, Ackermann H, Stutte HJ, Altmannsberger HM, Rubsamen-Waigmann H, Strebhardt K. Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene* 1997; 14: 543–549
- ²⁶ Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene expression profiles in normal and cancer cells. *Science* 1997; 276: 1268–1272