

Genexpressionsanalysen
zur Charakterisierung
primärer Mammatumoren
im neoadjuvanten Setting

Dissertation

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Biologie der Universität Bielefeld
vorgelegt von

Anika Tauchen

Dezember 2009

INHALTSVERZEICHNIS

Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
Glossar	IX
1 Zusammenfassung	1
2 Einleitung	5
2.1 Brustkrebsepidemiologie	7
2.2 Diagnose und Therapie des Mammakarzinoms	12
2.3 Brustkrebsforschung - wichtige Studien im Überblick	17
2.4 Prozesse in der Tumorzelle	20
2.5 Ziele der Arbeit	26
3 Material und Methoden	27
3.1 Patientinnenkollektiv und Tumorproben	27
3.1.1 Auswahlkriterien	28
3.1.2 Patientinnenbezogene Daten	29
3.1.3 Tumormaterial und Referenz	29
3.1.4 Tumorbezogene Daten	29
3.2 Themenmikroarray	31
3.2.1 Geninformation und -sequenzen	32
3.2.2 Gen-Ontologie	35
3.3 Labormaterial	35
3.3.1 Lösungen und Puffer	35
3.3.2 Chemikalien/Sonstiges	36
3.3.3 Verbrauchsmaterialien	37
3.3.4 Enzyme	37
3.3.5 Real-Time-PCR-Primer	37
3.3.6 Komplett-Systeme (Kits)	38
3.3.7 Geräte	38
3.3.8 Software	38
3.4 Methoden	38
3.4.1 Gesamt-RNA-Isolierung	39
3.4.2 RNA-Messungen	40
3.4.3 Mikroarrayhybridisierung	41
3.4.4 Quantitative Real-Time-PCR	47
3.4.5 Datenanalyse	50

4	Ergebnisse	53
4.1	Erstellung des Themenarrays	53
4.1.1	Die 70-Gen-Signatur von van't Veer et al.	54
4.1.2	Weitere brustkrebsassoziierte Gene	55
4.2	Datengrundlage der Analysen	63
4.2.1	Charakterisierung des Patientinnenkollektivs	64
4.2.2	Gewonnene Expressionsdaten und ihre Einordnung	66
4.3	RNA aus Tumorgewebe	67
4.4	Charakterisierung anhand von Genexpressionsanalysen	69
4.4.1	Gene mit gleich bleibender Expression	72
4.4.2	Gene mit variierender Expression	74
4.4.3	Identifizierung auffälliger Gengruppen	78
4.5	Klinische Parameter und Genexpressionsdaten	83
4.5.1	Rangkorrelationsanalysen	83
4.5.2	Normalized Tree Index	87
4.5.3	Detaillierte Analysen zu den Hormonrezeptoren	88
5	Diskussion	93
5.1	Studiendesign und Methoden	93
5.2	Molekulare Charakterisierung von Brusttumoren	100
5.3	Genexpressionsanalysen in der klinischen Anwendung	107
6	Ausblick	111
A	Geninformation	XI
A.1	Sequenzinformation der 70mer-Oligonukleotide	XI
A.2	Gene aus der 70-Gen-Signatur nach van't Veer	XIV
A.3	Genauswahl des Themenarrays (Literaturrecherche)	XV
B	Patientinnendaten/klinische Parameter	XXI
C	RNA-Isolierung	XXV
D	Mikroarrayexperimente	XXIX
	Literaturverzeichnis	XXXIII

ABBILDUNGSVERZEICHNIS

Abbildung 2.1	Die häufigsten Krebserkrankungen von Frauen in Deutschland (2004)	7
Abbildung 2.2	Altersspezifische Inzidenz bei Frauen in Deutschland (2004)	8
Abbildung 2.3	Exogene und endogene Faktoren	9
Abbildung 2.4	Modelle der Metastasierung	22
Abbildung 3.1	Schott Nexterion Slide E MPX 16	32
Abbildung 3.2	Vergleich der Normalisierungsmethoden	47
Abbildung 4.1	BMI der Studienpatientinnen nach Altersklassen . . .	66
Abbildung 4.2	GeneOntology(GO)-Prozesse und -Funktionen der Gene des Themenarrays	68
Abbildung 4.3	Gewonnene RNA aus Biopsien und OP-Proben im Vergleich zum Ausgangsgewicht der Gewebeproben	69
Abbildung 4.4	Gewonnene RNA in ng pro mg Tumorgewebe für verschiedene Probenarten	70
Abbildung 4.5	Heat Maps nach Funktionsklassen der analysierten Gene	71
Abbildung 4.6	Box-Plots der signifikant regulierten Gene	75
Abbildung 4.7	Verteilung der p -Werte für die gemessenen Expressionswerte M	76
Abbildung 4.8	Agglomeratives hierarchisches Clustern	79
Abbildung 4.9	Die 11 Haupt-Gencluster des hierarchischen Clusters	80
Abbildung 4.10	Ausprägung der klinischen Parameter für die resultierenden fünf Probencluster	84
Abbildung 4.11	NTI für ausgewählte klinische Parameter	87
Abbildung 4.12	Protein- und Genexpressionsnachweis des Östrogenrezeptors	89
Abbildung 4.13	Protein- und Genexpressionsnachweis des Progesteronrezeptors	90
Abbildung 4.14	Protein- und Genexpressionsnachweis für Her2-neu .	91
Abbildung 6.1	„omics“-Techniken zur Analyse biologischer Prozesse	112

TABELLENVERZEICHNIS

Tabelle 2.1	Faktoren zur Abschätzung des relativen Brustkrebsrisikos	10
Tabelle 2.2	Risikokategorien von Brustkrebspatientinnen	15
Tabelle 3.1	Immunreaktiver Score	30
Tabelle 3.2	Bewertungsschema der Her2-neu-Immunhistochemie. .	31
Tabelle 4.1	Themenarray: Gene des Lipid- und Hormonstoffwechsels	56
Tabelle 4.2	Themenarray: Zellproliferationsgene	58
Tabelle 4.3	Themenarray: Gene der Zellzyklusregulation	59
Tabelle 4.4	Themenarray: Apoptoseassoziierte Gene	60
Tabelle 4.5	Themenarray: Gene der Angiogenese	61
Tabelle 4.6	Themenarray: mit Zelladhäsion und extrazellulärer Ma- trix assoziierte Gene	61
Tabelle 4.7	Themenarray: Metalloproteinasegene	62
Tabelle 4.8	Signifikant hochregulierte Gene über alle Tumorproben	73
Tabelle 4.9	Signifikant runterregulierte Gene über alle Tumorproben	74
Tabelle 4.10	Gene mit stark variierender Expression über alle Tu- morproben	77
Tabelle 4.11	Binäre Umwandlung der klinischen Parameter	85

GLOSSAR

- ADJUVANT begleitend, unterstützend
- ALLEL Merkmalsausprägung eines Gens
- ANAMNESE Beurteilung
- ANGIOGENESE Ausbildung neuer (Blut-)Gefäße
- ANGIOINVASION Einbruch des Tumors in umliegende Blut- oder Lymphgefäße
- APOPTOSE programmierter Zelltod
- BILATERAL zweiseitig
- BILIRUBIN Abbauprodukt des Hämoglobins
- BMI *Body-Mass-Index* berechnet aus Größe und Gewicht:
$$BMI = \text{Gewicht}(kg) / (\text{Größe}(m))^2$$
- DNA *Desoxyribonucleic Acid* = Desoxyribonukleinsäure (cDNA = *complementary DNA* beschreibt aus mRNA synthetisierte (revers transkribierte) DNA)
- DUKTAL von den Milchgängen der Brustdrüse ausgehend
- ENDOGEN/INTRINSISCH von innen heraus wirkend
- ENDOKRINER DISRUPTOR Substanz mit hormoneller Wirkung
- ER *Estrogen Receptor* = Östrogenrezeptor
- EST *Expressed Sequence Tag* = transkribierte Gensequenzen , enthalten in einer cDNA-Bibliothek klonierter Fragmente zur Sequenzierung
- EXOGEN/EXTRINSISCH von außen einwirkend
- FISH Fluoreszenz *in situ* Hybridisierung
- GRADING Histologische Einteilung von Tumoren anhand des Differenzierungsgrades der Tumorzellen (G1 = gut differenziert bis G4 = entdifferenziert)
- GRAVIDITÄT Anzahl Schwangerschaften
- HEREDITÄR erblich
- HYBRIDISIERUNG Vorgang der Zusammenlagerung von komplementären, einzelsträngigen Nukleinsäuremolekülen (DNA, RNA)
- INDIKATION Voraussetzung für eine bestimmte medizinische Maßnahme
- INVASION Eindringen von Tumorzellen in das den Tumor umgebende Gewebe
- IRS immunreaktiver Score, Skala zur Bewertung des immunhistochemischen Färbenachweises für Östrogen- und Progesteronrezeptoren (0-12)
- KARNOFSKY-INDEX Skala zur Bewertung der Lebensqualität von Patienten mit bösartigen Tumoren (0-100 %)
- KARZINOM bösartiger Tumor mit Ursprung im Epithelgewebe (Deckgewebe, auch Drüsenepithel)
- LOBULÄR von den Drüsenläppchen der Brustdrüse ausgehend
- MALIGNE bösartig

MAMMA die weibliche Brust

MENARCHE/MENOPAUSE Beginn und Ende der Regelblutung der Frau

METASTASE durch Absiedelung von Tumorzellen des Primärtumors in anderen Geweben entstandener Tumorzellverband

MIKROARRAY auch Genchip, System zur Untersuchung von Genen

MITOSE Zell(kern)teilung

NEOADJUVANT begleitende Chemotherapie, vor Operation

NEOPLASIE Gewebeneubildung/Tumor

OLIGONUKLEOTID/OLIGOMER aus wenigen Nukleotiden (Grundbausteine der DNA/RNA) aufgebautes Molekül; ein 70mer besteht bspw. aus 70 Nukleotiden

PARITÄT Anzahl Geburten

PATHOMORPHOLOGIE Lehre der krankhaften Veränderungen von Geweben

PGR Progesteronrezeptor

PHÄNOTYP Erscheinungsbild

PRÄVENTION (primär, sekundär, tertiär) vorbeugende Maßnahmen (z.B. Gesundheitsförderung, Früherkennung, Therapie/Rehabilitation)

PRIMÄRTUMOR Entstehungsort/Ursprungsgeschwulst eines Tumors

PROGRESSION Prozess des Fortschreitens eines Tumors

PROLIFERATION Zellwachstum; im Tumor Wachstum des Zellverbandes

REMISSION Rückgang der Krankheitssymptome bzw. Rückbildung eines Tumors; bei einer komplett Remission ist der Tumor nicht mehr nachweisbar

REZESSIV die Merkmale rezessiver Gene treten nur dann in Erscheinung, wenn beide Allele des Gens die gleiche Ausprägung zeigen

REZIDIV Rückfall; Wiederauftreten einer Krebserkrankung

RNA *Ribonucleic Acid* = Ribonukleinsäure (Gesamt-RNA bestehend aus mRNA, rRNA, tRNA und anderen)

SPOT eine Einheit auf einem Mikroarray; repräsentiert ein Gen

STANZBIOPSIE minimal-invasive Methode zur Entnahme von Tumorgewebe

STROMA Raum zwischen Tumorgewebe und gesundem Gewebe, nicht-epithelialer Anteil des Tumors

SYSTEMISCH den ganzen Körper betreffend

TNM-KLASSIFIKATION Einteilung von Tumoren anhand ihrer Größe (T), der Anzahl betroffener Lymphknoten (N) und eventueller vorhandener Fernmetastasen (M)

TUMOR Gewebeneubildung/Neoplasie

UBIQUITÄR überall, ubiquitär exprimierte Gene sind in jedem Gewebe aktiv

UNILATERAL einseitig

WHO *World Health Organization* = Weltgesundheitsorganisation

ZYTOSTATIKUM das Zellwachstum hemmende Substanz

ZUSAMMENFASSUNG

Brustkrebs ist weltweit die häufigste Krebsneuerkrankung bei Frauen. Aufgrund bisher unzureichender prädiktiver und prognostischer Faktoren für diese sehr heterogene Erkrankung ist eine individualisierte Therapie und damit eine optimale Versorgung von Brustkrebspatientinnen nicht immer gewährleistet. Seit einigen Jahren werden hier Genexpressionsanalysen als viel versprechender Ansatz für die Routineanwendung untersucht. Ergänzend zu den etablierten Diagnosefaktoren bieten sie das Potential für eine verbesserte Charakterisierung von Brusttumoren zur Vorhersage von Krankheitsverläufen und Therapieansprechen.

Im Rahmen der vorliegenden Studie wurden die Zusammenhänge von Genexpressionsprofilen und klinischen Parametern (tumorbezogene sowie patientinnenbezogene Daten) primärer Mammakarzinome eines ausgewählten Patientinnenkollektivs unter begleitender Chemotherapie untersucht. Grundlage bietet ein im Rahmen dieser Arbeit konzipierter Themenmikroarray mit einer Auswahl brustkrebsrelevanter Gene. Schwerpunkt der Analysen ist, (i) zu untersuchen, ob eine Unterscheidung der Brusttumore anhand ihrer Genexpressionsprofile möglich ist; (ii) zu prüfen, ob diese Kategorien Rückschlüsse auf die klinischen Eigenschaften der Tumore ermöglichen; und (iii) zu identifizieren, welche Gene in den unterschiedlichen Tumoren bzgl. ihres Expressionsverhaltens besonders auffällig sind.

Auf funktionaler Ebene konnte die Expression von Genen der Steroidhormon- und Wachstumsfaktor-Signalwege als besonderes Unterscheidungsmerkmal identifiziert werden. Beide stehen in enger Verbindung miteinander und nehmen sowohl Einfluss auf die Entwicklung als auch auf das Therapieansprechen von Brusttumoren. Darüber hinaus dienten die gewonnenen Genprofile als Basis für die Kategorisierung der analysierten Tumore. Es konnten fünf Tumor-Cluster ermittelt werden, die sich wiederum in ihren klinischen Eigenschaften eindeutig voneinander unterscheiden. Dabei konnte auch beobachtet werden, dass sich Tumorproben einer Patientin, unabhän-

gig von der Art der Probennahme oder der Behandlung, in ihrem Profil grundsätzlich ähnlicher sind als Tumorproben verschiedener Patientinnen.

Eine Schlussfolgerung der vorliegenden Studie ist, dass Östrogen und der Steroid-Stoffwechsel in Bezug auf die Brustkrebserkrankung eine besondere Stellung einnehmen. Die komplexen Mechanismen, die Einfluss auf die Brustkrebserkrankung nehmen können, fordern jedoch weitere Forschungstätigkeit hinsichtlich zusätzlicher Ebenen (Genom, Proteom, Metabolom) und Endpunkte (verschiedene Gene/Genprodukte, unterschiedliche Therapie-schemata, Langzeitüberleben etc.), um langfristig das Ziel einer individualisierten Patientinnenversorgung erreichen zu können.

SUMMARY

GENE EXPRESSION ANALYSIS OF PRIMARY MAMMA TUMOURS IN A NEOADJUVANT SETTING

Breast Cancer is the most common cancer in women worldwide. Predictive and prognostic factors are not yet sufficient to allow for individualised therapies and optimal treatment for this very heterogenous disease. For several years, gene expression analysis has been studied as a useful method for routine diagnosis. It is considered to be a promising complement to established routine parameters in breast cancer characterisation for the prognosis of disease outcome and therapy response.

In this study primary mamma carcinomas from patients treated with neoadjuvant therapy were analysed. Correlations between gene expression profiles and clinical parameters (tumour- and patient-associated) were detected using a breast-cancer-related microarray, which was specifically designed within the project. The primary focus of the work was to analyse if, (i) different breast tumours could be distinguished based on their expression profiles, (ii) there are any correlations between tumour categories and clinical characteristics, and (iii) single genes could be identified that show significant expression characteristics in the analysed tumours.

On a functional level the expression of steroid hormone and growth factor pathways was identified as a significant difference between the analysed tumours. Both pathways are closely related and affect tumour progression and therapy response in breast cancer. Furthermore, the determined gene expression profiles were used for tumour clustering, resulting in five tumour-clusters associated with distinct clinical characteristics. The cluster analysis further indicated that tumour samples from one patient are more closely related to each other than to samples from different patients, independent of the method or time (before or after treatment) of the tissue sampling.

The dominant role of estrogen and steroid biosynthesis in breast cancer is one conclusion of the present study. The complex mechanisms influencing the disease require further studies concerning additional molecular levels

(genome, proteome, metabolome) and outcomes (genes/geneproducts, therapies, survival/disease-free survival) for the long-term goal of individualised breast cancer treatment.

EINLEITUNG

Das Mammakarzinom ist weltweit die mit Abstand häufigste Krebserkrankung der Frau (Boyle und Levin, 2008) und mit über 57.000 Neuerkrankungen jährlich auch die häufigste Krebsform bei Frauen in Deutschland (RKI, 2008). Die bevölkerungsmedizinisch große Relevanz des Themas Brustkrebs ist jedoch nicht allein in der Häufigkeit begründet, sondern „darüber hinaus ist Brustkrebs eine Erkrankung, deren Verlauf sich häufig über Jahre hinweg erstreckt und zu erheblichen physischen und psychischen Leiden, Funktionsverlust und Einschränkung der Lebensqualität führt.“ (BMGS, 2003, S.62). In direktem Zusammenhang hiermit steht, dass trotz der Fortschritte im Bereich der Tumordiagnostik und -therapie die Biologie des Mammakarzinoms noch teilweise ungeklärt ist und die Heterogenität der Tumoren von Brustkrebspatientinnen nach wie vor ein großes Problem für die Auswahl einer adäquaten Behandlung darstellt. Patientinnen mit der gleichen Tumorart und dem gleichen Tumorstadium können völlig unterschiedliche Krankheitsverläufe aufweisen. Eine Vorhersage der Tumorentwicklung oder eine konkrete Klassifizierung der Tumoren hinsichtlich ihres Therapieverhaltens ist mit den bisher etablierten klinischen Faktoren nicht immer möglich. Dies führt nicht selten dazu, dass Patientinnen eine Standardchemo- oder -hormontherapie erhalten, die ihren individuellen Bedürfnissen nicht gerecht wird und aufgrund von Nebenwirkungen mit erheblichen Einschränkungen der Lebensqualität verbunden sein kann. Einen wichtigen Schwerpunkt der Brustkrebsforschung stellen daher, neben Verfahren zur Früherkennung wie das Mammographie-Screening, neue Ansatzpunkte im Bereich der **Prognose** des Krankheitsverlaufs und **Prädiktion** des Therapieansprechens dar. Ein aktuell verfolgter Ansatz ist dabei die Bestimmung von prädiktiven und prognostischen Faktoren auf Genebene. Entscheidender Vorteil der Analyse auf Genebene, im Vergleich zu den bisherigen Faktoren mit gesicherter klinischer Relevanz wie der Tumorgröße und dem Lymphknoten- oder Hormonrezeptorstatus, ist die Möglichkeit der gleichzeitigen Untersuchung unter-

schiedlichster für die Entstehung, das Wachstum und die Ausbreitung eines Tumors verantwortlicher Faktoren. Mammakarzinome können große Unterschiede bezüglich ihrer metastatischen Potenz, der Zellzusammensetzung und ihrem Differenzierungsgrad aufweisen. Da angenommen werden kann, dass sich diese große phänotypische Variabilität der Brusttumoren auch auf molekularer Ebene widerspiegelt (Perou et al., 2000), ist die **Identifizierung spezifischer Genexpressionsprofile mittels Mikroarrayhybridisierung** und eine darauf basierende Klassifizierung der Tumoren ein viel versprechender Ansatz, der entscheidende Fortschritte für eine angepasste Versorgung von Brustkrebspatientinnen bewirken kann (Brennan et al., 2005).

Vor diesem Hintergrund beschäftigt sich die vorliegende Arbeit mit der molekularen Charakterisierung primärer Mammatumoren mittels Genexpressionsanalysen (Mikroarrayhybridisierung und Real-Time-PCR). Eine Besonderheit der hier durchgeführten Analysen ist, dass zusätzlich das aus Patientinnen mit Indikation zur begleitenden Chemotherapie zusammengesetzte Patientinnenkollektiv näher charakterisiert wird, um potentielle Einflussfaktoren auf die Brustkrebserkrankung mitzuerfassen. Für die Analysen kommt ein speziell für diese Fragestellung entwickelter Brustkrebs-Themenmikroarray zum Einsatz. Die dieser Arbeit zugrunde liegenden Untersuchungen wurden im Rahmen einer regionalen Studie in Kooperation mit dem Brustzentrum der Städtischen Kliniken Bielefeld (Region Ostwestfalen-Lippe, Nordosten Nordrhein-Westfalens) durchgeführt. In diese Studie eingeschlossen wurden Patientinnen, die zwischen 2006 und 2008 im Brustzentrum mit einer neoadjuvanten Chemotherapie behandelt wurden.

Um die nötige Wissensbasis für die hier vorliegenden Analysen zu schaffen, wird in diesem einführenden Kapitel Grundlegendes zur Brustkrebserkrankung erläutert. Dabei wird zunächst auf die Brustkrebsepidemiologie, insbesondere auch auf regionaler Ebene, und die Bedeutung von Risikofaktoren eingegangen. Anschließend wird ein kurzer Überblick über den aktuellen Stand von Diagnose und Therapie des Mammakarzinoms gegeben sowie einige wichtige Studien aus diesem Bereich vorgestellt. Daran schließt sich eine Einleitung in die Mechanismen der Tumorentstehung und -entwicklung auf molekularer Ebene an. Im letzten Abschnitt dieses Kapitels werden Fragestellung und Zielsetzung der Arbeit konkretisiert.

2.1 BRUSTKREBSEPIDEMIOLOGIE - HÄUFIGKEITEN UND RISIKOFAKTOREN

Das Mammakarzinom ist **weltweit die am häufigsten verbreitete Krebserkrankung der Frau**. Bei Männern tritt Brustkrebs mit unter 1 % Anteil an allen Brustkrebserkrankungen hingegen nur äußerst selten auf. In den Industrieländern hat Brustkrebs mit 26,5 % (ca. 715.000 Neuerkrankte) im Vergleich zu 18,8 % (577.000 Neuerkrankte) in den Entwicklungsländern einen deutlich höheren Anteil an allen Krebsneuerkrankungen bei Frauen (Boyle und Levin, 2008, geschätzte Daten für das Jahr 2008). Dies spiegelt sich auch in den Zahlen für Deutschland wider¹. Das Mammakarzinom ist hier mit einem Anteil von 27,8 % ebenfalls die bei weitem am häufigsten verbreitete Krebserkrankung (RKI, 2008, vgl. auch Abbildung 2.1) und liegt damit im europäischen Mittel (geschätzte 28,9 % im Jahr 2006, siehe Ferlay et al., 2007).

Auch wenn die Sterblichkeit unter den Patientinnen stetig sinkt, nimmt Brustkrebs aufgrund seiner Häufigkeit mit 17,9 % immer noch die erste Stelle der Statistik zur Mortalität von Frauen mit Krebs in Deutschland ein (RKI, 2008). Im europäischen Vergleich liegt die Mortalitätsrate in Deutschland (ebenso wie in Dänemark, den Niederlanden, Belgien und Großbritannien) etwas höher als in den übrigen EU-Ländern. Zu berücksichtigen ist bei diesen Zahlen der Einfluss des ab 2003 begonnenen Aufbaus einer flächendeckenden Reihenuntersuchung (Screening) der 50-69-jährigen Frauen in Deutschland (BMGS, 2003). Im Rahmen dieses Screenings können zusätzliche Erkrankungsfälle beobachtet werden, die zunächst zu einem Anstieg der Inzidenz führen. Mittelfristig wird durch frühzeitige Behandlungsmöglichkeiten aufgrund von Screenings jedoch ein Rückgang der Mortalitätsraten erwartet (Boyle und Levin, 2008).

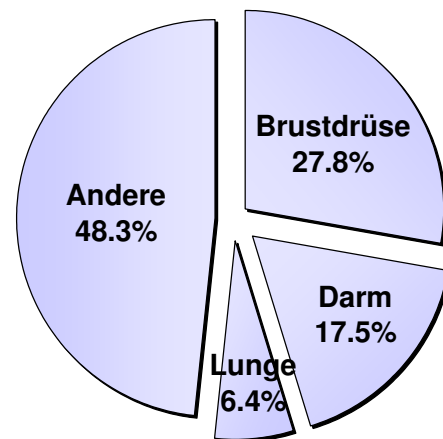


Abbildung 2.1: Prozentualer Anteil der häufigsten Krebsformen bei Frauen in Deutschland (2004).

¹ Die hier verwendeten Daten für Deutschland stützen sich, wenn nicht anders angegeben, auf die aktuellsten Schätzungen der Dachdokumentation Krebs im Robert Koch-Institut. Diese sind auf Basis der Daten aller vollzählig erfassenden epidemiologischen Krebsregister in Deutschland für das Jahr 2004 berechnet (RKI, 2008)

Das mittlere Erkrankungsalter für Krebs allgemein liegt bei 69 Jahren, für Brustkrebs liegt es mit 63 Jahren deutlich unter diesem Wert. Besonders bei einem Lebensalter zwischen 30 und 50 Jahren ist ein starker Anstieg der Brustkrebsneuerkrankungen zu beobachten (RKI, 2008, vgl. Abbildung 2.2). Bei 40 % der diagnostizierten Krebsneuerkrankungen bei unter 60-jährigen Frauen handelt es sich um Brustkrebs (Giersiepen et al., 2005).

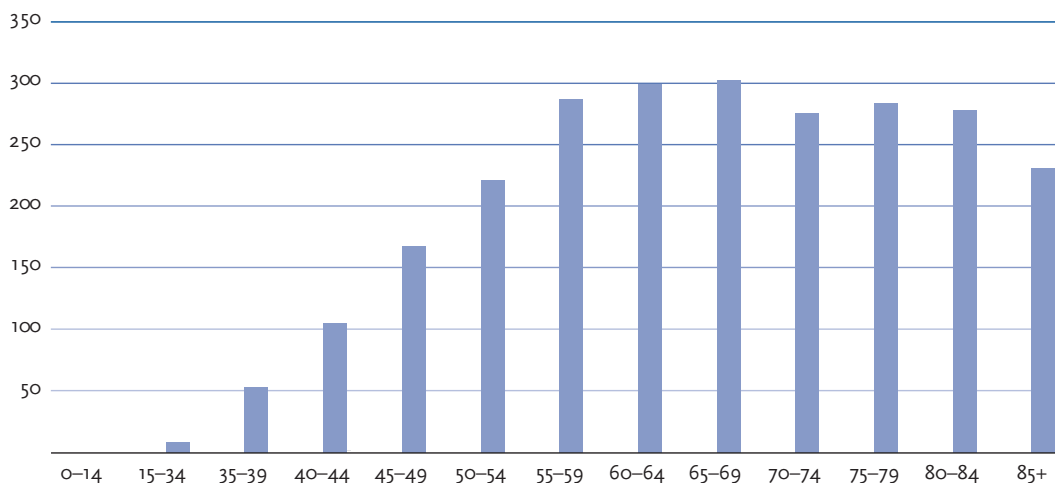


Abbildung 2.2: Altersspezifische Inzidenz bei Frauen in Deutschland für das Jahr 2004 (RKI, 2008). Angegeben ist die Anzahl der Neuerkrankungen pro 100.000 Frauen (y-Achse) in den jeweiligen Altersgruppen (x-Achse).

Regionale Daten zur Brustkrebsepidemiologie liefern in Deutschland die Krebsregister. Die dort präsentierten Zahlen können zum Teil von den deutschland- oder europaweiten Vergleichswerten abweichen, so gibt das epidemiologische Krebsregister Nordrhein-Westfalen (NRW) für den Landesteil Westfalen-Lippe im Jahr 2007 für Brustkrebs sogar einen Anteil von 35,5 % (7.538 Fälle) an allen Krebsneuerkrankungen an. Die Krebssterblichkeit liegt hier mit 16,6 % im Deutschland-Mittel und nimmt tendenziell weiter ab. Knapp 70 % der in Westfalen-Lippe diagnostizierten Mammakarzinome waren dabei aus den Milchgängen der Brustdrüse hervorgegangene duktale Karzinome, Tumoren mit Ursprung in den Drüsenläppchen (lobuläre Karzinome) kamen nur bei rund 20 % der Betroffenen Frauen vor (Epidemiologisches Krebsregister NRW, 2009). Auch die Daten der Weltgesundheitsorganisation WHO zeigen weltweit einen deutlich größeren Anteil (>80 %) an dukталen Mammakarzinomen (Boyle und Levin, 2008).

Bisherige Untersuchungen zur Identifizierung von **Risikofaktoren** für die Brustkrebserkrankung konnten bereits Zusammenhänge zwischen familiären Faktoren und dem Erkrankungsrisiko feststellen. So sind aufgetretene Brustkrebserkrankung in jungen Jahren (<40 Jahre), das gleichzeitige Vorhan-

densein von Mamma- und Ovarialkarzinom oder nachgewiesene Mutationen in den Brustkrebsgenen BRCA1 und BRCA2 in einer Familie Kennzeichen für ein hereditäres Mammakarzinom. Diese Faktoren allein reichen jedoch nicht aus, um die hohen Inzidenzraten für Brustkrebs zu erklären. Eine **genetische Disposition** ist nur für ca. 5-10 % aller Brustkrebsneuerkrankungen die Ursache (Kuschel et al., 2005), der weit größere Anteil betrifft das spontan oder **sporadisch** auftretende Mammakarzinom. Es ist daher anzunehmen, dass auch weitere Faktoren einen Einfluss auf das Brustkrebsrisiko haben können. So zeigen sich im internationalen Vergleich z.B. große Unterschiede der Inzidenz- und Mortalitätsraten, deren Ursachen noch nicht geklärt sind (vgl. hierzu auch Unterschiede der Industrie- und Entwicklungsländer). Die Beobachtung, dass bei Migranten, die von einem Land mit niedrigen Brustkrebsraten in ein Land mit hoher Inzidenz auswandern, innerhalb weniger Generationen eine Anpassung an die höheren Inzidenzraten erfolgt (Boyle und Levin, 2008), lässt darauf schließen, dass **exogene Faktoren** wie umweltbedingte oder verhaltensbedingte Faktoren eine große Rolle für das Erkrankungsrisiko spielen (Abbildung 2.3).

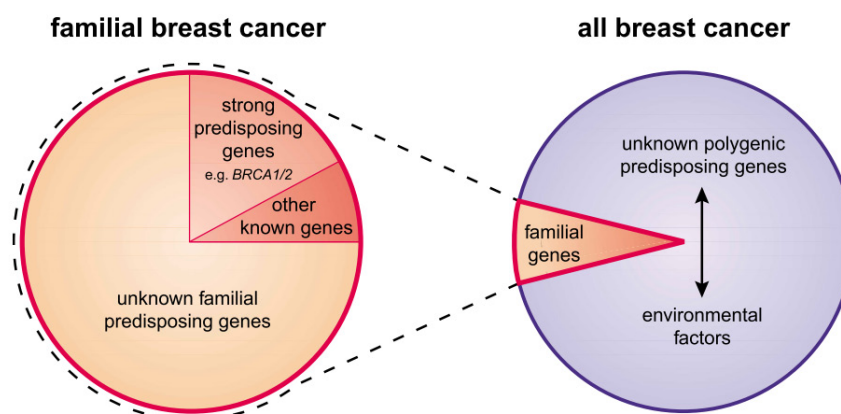


Abbildung 2.3: Bedeutung exogener und endogener Faktoren für das Brustkrebsrisiko (Balmain et al., 2003)

Unter den Risikofaktoren gibt es solche mit gesicherter Relevanz und solche, die hinsichtlich ihres kausalen Wirkzusammenhanges zur Entstehung von Brustkrebs kontrovers diskutiert werden bzw. für die ein Zusammenhang mit erhöhtem Brustkrebsrisiko noch nicht nachgewiesen werden konnte (BMGS, 2003). Im Folgenden wird ein kurzer Überblick über diese Faktoren gegeben. Zu den **Risikofaktoren mit gesicherter Relevanz** für die Brustkrebserkrankung gehört, wie bereits oben erwähnt, die **genetische Disposition**. So liegt das normale Lebenszeitrisiko für Frauen in Deutschland, an Brustkrebs zu erkranken, bei knapp 10 %. Das Lebenszeitrisiko für Trägerinnen von Muta-

tionen in den BRCA-Genen liegt im Vergleich dazu zwischen 45 % (BRCA2) und 65 % (BRCA1) (Kuschel et al., 2005). Grundsätzlich nimmt das Risiko einer Brustkrebserkrankung, wie für die meisten anderen Krebsarten auch, mit dem **Lebensalter** zu (Giersiepen et al., 2005). Zudem beeinflussen **reproduktive und hormonelle Faktoren** das Erkrankungsrisiko. So sind eine frühe Menarche, eine spät einsetzende Menopause, eine geringe Anzahl von Schwangerschaften oder Kinderlosigkeit, kein oder nur kurzes Stillen und eine späte erste Schwangerschaft mit einem erhöhten Erkrankungsrisiko assoziiert. Umgekehrt verringern eine frühe erste ausgetragene Schwangerschaft sowie mehrere Geburten das Risiko, an Brustkrebs zu erkranken (Giersiepen et al., 2005; RKI, 2008; Boyle und Levin, 2008). Diese Faktoren sind in erster Linie mit Östrogen assoziiert, das bei der Tumorentwicklung und -progression eine wichtige Rolle spielt. Frauen mit negativem Hormonrezeptorstatus (Östrogen und Progesteron, siehe hierzu auch Abschnitt 2.2) haben eine schlechtere Prognose für den Krankheitsverlauf (Osborne, 1998). Einen Vergleich des Einflusses einiger der genannten Faktoren auf das relative Risiko², an Brustkrebs zu erkranken, zeigt Tabelle 2.1.

Tabelle 2.1: Faktoren zur Abschätzung des relativen Risikos für eine Brustkrebserkrankung nach Armstrong et al. (2000).

Risikofaktor	relatives Risiko
familiäre Belastung (Verwandtschaft 1. Grades)	1,4-13,6
Alter (>/= 50 Jahre vs. <50 Jahre)	6,5
Alter bei erster Lebendgeburt (>30 Jahre vs. <20 Jahre)	1,3-2,2
Alter bei Menopause (>/= 55 Jahre vs. <55 Jahre)	1,5-2,0
Alter bei Menarche (<12 Jahre vs. >/= 14 Jahre)	1,2-1,5

Neben diesen etablierten Risikofaktoren werden **potentielle Risikofaktoren** für die Brustkrebsentstehung diskutiert. Diese betreffen vor allem den allgemeinen Lebensstil (z.B. Rauchen, Bewegungsmangel) und dabei insbesondere die Ernährung (z.B. Alkohol, Fettkonsum). Darüber hinaus wird der Einfluss von physikalisch-chemischen Umweltfaktoren (z.B. Strahlung, Umweltverschmutzung, exogene Hormone) auf das Brustkrebsrisiko untersucht. Insbesondere den Lebensstilfaktoren kommt dabei aus präventiver Sicht (**Primärprävention**) eine besondere Bedeutung zu, da sie den Einflussfaktor bilden, der sich theoretisch am einfachsten modifizieren lassen sollte (Paepke et al., 2005).

² Das relative Risiko ist der Faktor, um den sich das Risiko für eine Erkrankung unter spezifischen Voraussetzungen erhöht. Bei einem relativen Risiko von 1 unterscheidet sich das Risiko nicht vom allgemeinen Erkrankungsrisiko.

Unter den den **Lebensstil** betreffenden Risikofaktoren ist nach Untersuchungen der Weltgesundheitsorganisation (WHO) **Tabakrauch** einer der bedeutendsten für viele Krebserkrankungen (insbesondere Lungenkrebs) (Boyle und Levin, 2008). Für das Risiko einer Brustkrebserkrankung konnte jedoch kein Einfluss nachgewiesen werden. Aus früheren Studien liegen hier zum Teil widersprüchliche Ergebnisse vor (Stewart und Kleihues, 2003; Coyle, 2004), nach neuesten Erkenntnissen besteht jedoch kein Zusammenhang von Tabakkonsum und Brustkrebsrisiko (Ahern et al., 2009; Boyle und Levin, 2008). Weitere den Lebensstil betreffende mögliche Risikofaktoren, die zudem eng miteinander verknüpft sind, sind unzureichende **körperliche Aktivität** und ein hohes **Körpergewicht** (Boyle und Levin, 2008). Mit steigendem *Body-Mass-Index*³ erhöhte sich in Studien sowohl die Inzidenz als auch die Mortalitätsrate für Brustkrebs signifikant (Paepke et al., 2005). Beide Risikofaktoren spielen hauptsächlich bei Frauen nach dem Eintreten der Menopause eine Rolle (Boyle und Levin, 2008).

Insbesondere mit Hinblick auf die international auftretenden Unterschiede in der Brustkrebsinzidenz stellen auch **Ernährungsgewohnheiten und Nahrungsbestandteile** einen wichtigen Forschungsschwerpunkt dar. Untersucht werden vor allem der Einfluss fleisch- und fettreicher im Vergleich zu fleischloser Kost bzw. des Verzehrs von Fisch und Gemüse. Ein konkreter Zusammenhang zum Brustkrebsrisiko konnte bisher für einzelne Nahrungskomponenten nicht nachgewiesen werden, es gibt aber Hinweise, dass eine reduzierte Fettaufnahme das Brustkrebsrisiko verringern kann (Boyle und Levin, 2008). Insgesamt wird **Alkohol** unter den ernährungsbedingten Faktoren in der Literatur zwar durchgehend als das Brustkrebsrisiko steigernd eingeschätzt, die Studien sind aber insbesondere bzgl. Dosis-Wirkungsbeziehungen und möglicher weiterer Einflussfaktoren zu inkonsistent, um die Relevanz endgültig zu bestätigen. Daher zählt Alkoholkonsum immer noch zu den potentiellen Risikofaktoren (Tjønneland et al., 2007; Boyle und Levin, 2008; Allen et al., 2009).

Als potentiell das Brustkrebsrisiko beeinflussende **Umweltfaktoren** werden ionisierende Strahlen (Coyle, 2004), verschiedene Umweltschadstoffe (polyzyklische Kohlenwasserstoffe, organische Lösemittel) (Brody und Rudel, 2003) und Spurenelemente (z.B. Zink, Arsen, Nickel, Kadmium, Selen) (Silvera und Rohan, 2007) diskutiert. Die WHO resümiert in ihrem *World Cancer*

3 Der *Body-Mass-Index* (BMI) berechnet sich aus Größe und Gewicht, $BMI = \text{Gewicht}(kg) / (\text{Größe}(m))^2$. Nach Definition der WHO gelten Personen mit einem BMI unter 18,5 als untergewichtig, über 25 als übergewichtig und mit einem BMI über 30 als stark übergewichtig.

Report, dass ionisierende Strahlung (aus Umwelt- und anthropogenen Quellen) als einziger der hier aufgeführten Umweltfaktoren einen weitestgehend gesicherten Einfluss auf das Brustkrebsrisiko hat. Auch Umweltfaktoren die nicht als klassische Umweltschadstoffe gelten werden als Risikofaktoren diskutiert. Studien haben gezeigt, dass das Brustkrebsrisiko unter Arbeiterinnen in Nachtschichten leicht erhöht zu sein scheint (Srinivasan et al., 2008). Auch eine unzureichende Schlafdauer (regelmäßig weniger als sechs Stunden) wurde mit einem höheren Brustkrebsrisiko assoziiert (Kakizaki et al., 2008). Beide Beobachtungen werden mit dem Einfluss der Lichtexposition auf den Melatonin-Spiegel der Zellen und eine damit verbundene Störung der „biologischen Uhr“ (circadiane Rythmus) in Verbindung gebracht.

Die hier als potentielle Risikofaktoren für Brustkrebs dargestellten Faktoren sind in ihrer Wirkung nicht oder nur eingeschränkt bestätigt. Es muss jedoch berücksichtigt werden, dass viele Faktoren auch indirekt oder durch Wechselwirkungen mit anderen Faktoren Einfluss auf das Erkrankungsrisiko nehmen könnten. So können z.B. einige der diskutierten Faktoren **hormonelle Vorgänge** in den Zellen, und dadurch potentiell auch das Krebsrisiko, beeinflussen. Ein Beispiel wäre hier u.a. die Bedeutung von Übergewicht für ein erhöhtes Brustkrebsrisiko, insbesondere bei postmenopausalen Frauen, durch die Umwandlung von Testosteron im Fettgewebe zu biologisch aktivem Östrogen durch Aromatasen (Renehan et al., 2008). Oder auch die Ernährung betreffende Faktoren, wie die Rolle von Phytoöstrogenen, die z.B. in hohem Maß in der Sojabohne enthalten sind. Aufgrund ihrer östrogenen Aktivität wird ein protektiver Effekt bzgl. des Brustkrebsrisikos vermutet, der in Studien jedoch bisher nicht bestätigt werden konnte (Keinan-Boker et al., 2004). Ein weiterer umweltassoziiertes endokriner Disruptor, eine Substanz mit östrogenen Aktivität, ist z.B. Cadmium, für das jedoch eine kanzerogene Wirkung vermutet wird (Brody und Rudel, 2003; McElroy et al., 2006).

2.2 DIAGNOSE UND THERAPIE DES MAMMAKARZINOMS

Unter dem Begriff „Mammakarzinom“ werden alle bösartigen (malignen) Gewebeneubildungen (Tumoren/Neoplasien) der Brust zusammengefasst. Dabei werden verschiedene **histologische Tumortypen** unterschieden. Am häufigsten tritt das aus dem Epithelgewebe der Milchgänge entstehende duktales Karzinom auf, seltener aus den Drüsenläppchen hervorgehende lobuläre Karzinome (vgl. Abschnitt 2.1). Weiterhin unterscheidet man invasive

und nicht-invasive Mammakarzinome, abhängig davon, ob das umliegende Gewebe betroffen ist oder nicht: duktales *Carcinoma in situ* (DCIS), lobuläres *Carcinoma in situ* (LCIS), invasives duktales *Carcinoma* (IDC), invasives lobuläres *Carcinoma* (ILC). Auch Tumoren einer Art unterscheiden sich jedoch zum Teil beträchtlich in ihren biologischen Eigenschaften, wie der metastatischen Potenz, der Zellzusammensetzung und dem Differenzierungsgrad der Zellen. Brusttumoren weisen also eine große **Heterogenität** auf, die es auch hinsichtlich Diagnostik und Therapie zu berücksichtigen gilt.

Die **Früherkennung** ist der erste Schritt in der Diagnose- und Versorgungskette des Mammakarzinoms. Zunächst ist hier die Selbstuntersuchung der Brust zu nennen, deren Bedeutung vor allem auch darin zu sehen ist, dass sie das Bewusstsein der Frauen für die Notwendigkeit der Brustkrebsfrüherkennung erhöht (Sittek et al., 2005). Das klinische Standardverfahren zur Brustkrebsfrüherkennung ist die Röntgendiagnostik oder auch Mammographie, die in Deutschland im Rahmen eines Routine-Screenings eingesetzt wird und somit die wichtigste Maßnahme zur **Sekundärprävention** darstellt (vgl. Abschnitt 2.1). Weitere bildgebende Verfahren sind die auf Ultraschall basierende Sonographie und die Magnetresonanztomographie (MRT), die häufig ergänzend zur Mammographie zum Einsatz kommen. Diese Verfahren haben im Gegensatz zur Röntgendiagnostik den Vorteil, dass sie auch bei besonders dichtem Brustkrebsgewebe, wie es z.B. häufig bei jüngeren Frauen vorliegt, angewendet werden können. Die MRT hat jedoch auch Nachteile, so kann z.B. nicht zwischen gut- und bösartigen Veränderungen der Brustdrüsen unterschieden werden (Sittek et al., 2005; Kreienberg et al., 2008).

Bei auffälligem Befund werden minimal-invasive Methoden für die weitere Diagnostik eingesetzt. Hierfür wird präoperativ durch Einschließen einer Hohlnadel unter Röntgen- oder Ultraschallkontrolle (Stereotaxie) eine Probe aus dem Gewebe gestanzt und diese **Stanzbiopsie** pathomorphologisch untersucht. Liegt ein Mammakarzinom vor, werden weitere Parameter untersucht, die zur genaueren Beschreibung der Tumorart dienen. Neben dem histologischen Typ (duktal/lobulär, invasiv/nicht-invasiv) werden der TNM-Status, der Differenzierungsgrad (Grading) und der Rezeptorstatus (Hormonrezeptoren und Her2-neu) vom Pathologen in der Routinediagnostik bestimmt.

Der **TNM-Status** dient zur Klassifizierung von Tumoren anhand ihrer Größe (T=Tumorgröße) und Ausbreitung (N=Nodalstatus/Lymphknotenbefall und M=Metastasen). Sind weder ein Primärtumor noch Lymphknoten- und

Fernmetastasen nachweisbar, wird der Status T0 N0 M0 zugewiesen. Die zunehmende Ausdehnung des Primärtumors wird mit den Kategorien T1-4 beschrieben, je nach Lage und Anzahl der befallenen Lymphknoten unterscheidet man die Kategorien N1 bis N3 (Lymphknotenstatus), und das Vorhandensein von Fernmetastasen gibt der Status M1 an. Das **histologische Grading** beschreibt den Differenzierungsgrad der Tumorzellen von G1, gut differenziert, bis G3, schlecht differenziert, im Vergleich zum normalen Gewebe, als Maß für die Aggressivität des Tumors (zur Klassifikation des Mammakarzinoms und den jeweiligen Abstufungen vgl. Lebeau et al., 2005). Des Weiteren erfolgt die immunhistochemische Bestimmung der Hormonrezeptoren für Östrogen und Progesteron. Auf Basis einer Färbereaktion kann der immunreaktive Score (IRS) von 0-12 und damit der **Hormonrezeptorstatus** der Zellen von negativ bis stark positiv bestimmt werden. Eine Beschreibung der Berechnung dieses Scores folgt in Kapitel 3.1.4, Tabelle 3.1. Auch der **Her2-neu-Status** wird immunhistochemisch bestimmt. Anders als bei den Hormonrezeptoren wird die Expression des Wachstumsfaktorrezeptors Her2-neu als Score von 0 bis 3+ angegeben (siehe Kapitel 3.1.4, Tabelle 3.2). Bei einem nicht eindeutigen Färbeergebnis (2+) wird der Rezeptor zusätzlich auf Genebene analysiert. Hierfür wird eine Fluoreszenz *in situ* Hybridisierung durchgeführt, um eine eventuelle Amplifikation des Her2-neu kodierenden ERBB2-Gens im Genom nachzuweisen, die häufig ursächlich für die Überexpression des Rezeptors ist (Funke et al., 2005).

Die hier genannten Faktoren gelten als etablierte **prognostische und prädiktive Faktoren** für die Brustkrebserkrankung, d.h. sie dienen als Grundlage für die Abschätzung des individuellen Krankheitsverlaufs (TNM-Status, Grading, Hormonrezeptorstatus) bzw. des Therapieansprechens (Hormonrezeptor- und Her2-neu-Status) (Decker et al., 2009). So werden u.a. basierend auf diesen Parametern auf der zweijährlich statt findenden St. Gallen Konsensuskonferenz Risikogruppen (niedriges, mittleres und hohes Risiko) definiert, die auch für die Wahl der begleitenden Chemotherapie ausschlaggebend sind (Tabelle 2.2). Östrogen- und progesteronrezeptorpositive Tumoren gehören in die niedrige Risikogruppe und sind zudem für eine Anti-Hormontherapie geeignet. Ein positiver Her2-neu-Status geht hingegen mit einem höheren Risiko einher (vgl. Tabelle 2.2). Hier kann jedoch ergänzend zur Chemotherapie eine Therapie mit dem Antikörper *Trastuzumab* (*Herceptin*[®]) durchgeführt werden, der den Rezeptor blockiert und so zur Hemmung des Tumorwachstums bzw. einem Absterben der Krebszellen führt (Lebeau et al., 2005). Weitere immunhistochemisch zu bestimmende Prognosefaktoren werden diskutiert. Hier

sind z.B. der Proliferationsmarker Ki-67 sowie die Proteine uPA und PAI-1 zu nennen, die durch ihre Protease-Aktivität zur Metastasierungsfähigkeit der Tumorzellen beitragen können (Funke et al., 2005). Diese Faktoren wurden jedoch auch auf der zuletzt stattfindenden St. Gallen Konsensuskonferenz von 2009 aufgrund fehlender Validierungsdaten nicht als etablierte prognostische Faktoren anerkannt (Goldhirsch et al., 2009).

Tabelle 2.2: Risikokategorien der St. Gallen Konsensuskonferenz 2005 für die Behandlung von Brustkrebspatientinnen (nach Goldhirsch et al. (2005))

Faktor	Niedriges Risiko	Mittleres Risiko	Hohes Risiko
Befallene Lymphknoten	keiner	1 bis 3 oder keiner, aber 1 weiteres Kriterium	mehr als 4 oder 1 bis 3 und
Tumorgröße	T1 (max. 2 cm)	T2 bis T4 (größer als 2 cm)	
Differenzierung	G1	G2 und G3	
HER2-neu-Status	HER2-neu negativ	HER2-neu positiv	HER2-neu positiv
Alter	35 Jahre und älter		oder jünger als 35 Jahre

Die **Therapie des Mammakarzinoms** erfolgt in der Regel zunächst lokal durch Operation (brusterhaltend oder Amputation) oder auch gezielte Bestrahlung. Nach dem operativen Eingriff wird häufig eine adjuvante Chemo- oder Hormontherapie eingesetzt, da es sich bei Brustkrebs um eine **systemische** Erkrankung handelt, die den ganzen Körper und nicht nur das primär erkrankte Organ betreffen kann. Sich möglicherweise noch im Körper befindende Tumorzellen können so bekämpft werden, bevor Metastasen nachweisbar sind. **Neoadjuvante** (primär systemische) Therapien werden meist bei fortgeschrittener Erkrankung bzw. inoperablem Tumor angewandt, um diesen zunächst zu verkleinern und eine Verringerung des Tumorstadiums zu erreichen (zur Indikation der neoadjuvanten Chemotherapie vgl. Kreienberg et al., 2008).

Die Auswahl der geeigneten **Anti-Hormontherapie** hängt jeweils vom menopausalen Status der Patientin und entsprechender Hormonrezeptorpositivität der Tumoren ab. **Postmenopausal** können Aromatase-Hemmer (z.B. *Anastrozol*, *Letrozol*, *Formestan*) oder Gestagene die Östrogenproduktion unterbinden. Nach den Wechseljahren werden Östrogenvorstufen (Estron, Estradiol etc.) von der Nebenniere produziert, die dann durch das Enzym Aromatase im Muskel- und Fettgewebe zu Östrogen umgewandelt werden. Vorher findet die Östrogenproduktion hauptsächlich in den Eierstöcken (Ovarien) statt. Gestagene (Gelbkörperhormone) besetzen die Bindungsstellen des natürli-

chen Gelbkörperhormons Progesteron und senken durch die Beeinflussung verschiedener Stoffwechselschritte den Östrogenspiegel im Blut. Zudem blockieren Gestagene vermutlich die Bildung von Östrogenrezeptoren. Anti-Östrogene wie *Tamoxifen* blockieren hingegen direkt die Östrogenrezeptoren und können so das Wachstum der Tumorzellen verhindern. Sie können bei östrogenpositiven Tumoren vor oder während der Menopause eingesetzt werden. **Prämenopausal** wirken GnRH(*Gonadotropin-releasing Hormone*)-Analoga unterdrückend auf die ovarielle Östrogenproduktion (Boyle und Levin, 2008; Deutsche Krebsgesellschaft, 2009).

Für die begleitende **Chemotherapie** kommen verschiedene Therapieschemata zur Anwendung. Es werden verschiedene Zytostatika, zumeist in Kombination, über einen längeren Zeitraum und in definierten Zeitabständen (Zyklen), eingesetzt. Dabei können vier hauptsächliche Stoffklassen mit unterschiedlichen Wirkweisen unterschieden werden: Alkylanzien (z.B. *Cyclophosphamid*, binden an DNA und verhindern die Replikation der DNA), Anthrazykline (z.B. *Doxorubicin* und *Epirubicin*, diese zytostatischen Antibiotika verhindern das Aufbrechen des DNA-Stranges bei der DNA-Synthese), Antimetabolite (z.B. *Fluorouracil*, *Capecitabin*, *Methotrexat* und *Gemcitabin*, werden statt körpereigener Stoffe beim Aufbau des Erbguts verwendet und können zum Zelltod führen) und Taxane (z.B. *Paclitaxel* und *Docetaxel*, wirken auf den Spindelapparat und stören so die Mitose) (Boyle und Levin, 2008).

In den Leitlinien der Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften werden derzeit als Standardbehandlung in erster Linie anthrazyklinhaltige Kombinationen empfohlen (siehe auch (Kreienberg et al., 2008)). Gängige Kombinationen sind z.B. das TAC-Schema mit der gemeinsamen Gabe von *Docetaxel* (*Taxotere*[®]), *Doxorubicin* (*Adriamycin*) und *Cyclophosphamid* über sechs Zyklen, oder das AC-DOC-Therapieschema mit vier Zyklen *Doxorubicin* und *Cyclophosphamid*, gefolgt von vier Zyklen *Docetaxel*⁴. Eine Zykluslänge beträgt in der Regel drei bis vier Wochen.

Wie bereits erwähnt sind die Möglichkeiten der Primär- und Sekundärprävention des Mammakarzinoms begrenzt, die Bestimmung prognostischer und prädiktiver Faktoren ist daher ein wichtiger Ansatzpunkt zur Verbesserung der Versorgung von Mammakarzinompatientinnen. Allerdings berücksichtigen auch die bisher etablierten Faktoren noch nicht die Heterogenität der Tumoren und die damit verbundene, sehr individuelle Therapieempfind-

4 TAC: Docetaxel 75 mg/m²KO (Körperoberfläche), Doxorubicin 60 mg/m²KO, Cyclophosphamid 600mg/ mg/m²KO; AC-DOC: Doxorubicin 60 mg/m²KO, Cyclophosphamid 600 mg/m²KO, Docetaxel 100 mg/m²KO

lichkeit und Tumorentwicklung. Die Weiterentwicklung prognostischer und prädiktiver Faktoren, z.B. durch Genexpressionsprofile, ist daher Gegenstand aktueller Brustkrebsforschung, auf die im folgenden Abschnitt dieses Kapitels genauer eingegangen wird.

2.3 BRUSTKREBSFORSCHUNG - WICHTIGE STUDIEN IM ÜBERBLICK

Ein tieferes Verständnis der zellulären und molekularen Prozesse in der Tumorzelle ist grundlegender Gegenstand der Brustkrebsforschung, da das Zusammenspiel vieler komplexer Zellregulationsmechanismen den Krankheitsverlauf einer Brustkrebserkrankung und auch das Ansprechen oder Versagen einer Therapie bestimmt (siehe Abschnitt 2.4). In den letzten Jahren wurden daher vermehrt Genexpressionsstudien mit dem Ziel der besseren Klassifizierung von Mammatumoren als Basis für eine angepasste Versorgung der Patientinnen durchgeführt. Methodischer Schwerpunkt dieser Studien sind vor allem die Mikroarraytechnik und Expressionsanalysen mittels quantitativer Real-Time-PCR (qRT-PCR).

Mit Hilfe der **Mikroarraytechnologie** lässt sich in einem einzigen Experiment die Expression einer großen Anzahl von Genen analysieren, die potentiell in die komplexen Mechanismen der Tumorentstehung involviert sein können. So können nicht nur einzelne Markergene identifiziert, sondern auch Gensignaturen erfasst werden, die eine molekulare Charakterisierung und Klassifizierung der Tumorzellen ermöglichen. Das Grundprinzip dieser Methode ist die Hybridisierung von RNA aus Tumorzellen gegen Teilabschnitte interessanter Gensequenzen, die auf einem Träger-Substrat verankert sind (eine detaillierte Beschreibung der Methode befindet sich in Kapitel 3.4.3). Die **quantitative Real-Time-PCR** hingegen ist besonders für die Analyse einer beschränkten Anzahl von Genen geeignet. Sie wird unter anderem für Expressionsanalysen einzelner Kandidatengene oder zur Bestätigung der Ergebnisse eines Mikroarrayexperiments genutzt, ist aber auch als eigenständiger Versuchsansatz von Bedeutung. Bei dieser Methode werden Zielsequenzen der zu analysierenden Gene mittels Polymerase-Kettenreaktion vervielfältigt und gleichzeitig quantifiziert. So kann auf die Expression dieser Gene im untersuchten Gewebe rückgeschlossen werden (zum methodischen Vorgehen siehe Kapitel 3.4.4). Gemessen wird sowohl bei der Mikroarrayhybridisierung als auch der qRT-PCR die relative differentielle Expression der Gene im Tumor im Vergleich zu einer Referenz. Hierbei kann es sich um RNA

aus gesundem Gewebe, Zellkulturlinien oder einem Pool aus Tumorproben handeln. Beide Methoden sind grundsätzlich für klinische Routineuntersuchungen einsetzbar. Mikroarrays haben den Vorteil hoher Durchsatzraten und guter Standardisierbarkeit, insbesondere im Vergleich zu den immunhistochemischen Methoden. Real-Time-PCR-Analysen können auf der Grundlage identifizierter Markergene kostengünstig und mit geringen Probenmengen durchgeführt werden.

Ausgehend von Genexpressionsanalysen konnten für die Brustkrebskrankung bereits eine Vielzahl unterschiedlicher **prognostischer Modelle** und damit einhergehende Markergene bzw. Gensignaturen identifiziert werden. Eine der ersten Arbeiten, die hier wichtige neue Erkenntnisse gebracht hat, ist die Studie von Perou et al. (2000). Die Gruppe um Perou und Sørlie zeigte in ihren Mikroarrayanalysen an 65 Tumorproben von 42 Patientinnen mittels hierarchischen Clustern vier große Gruppen von Tumoren auf, die, mit Blick auf ihre histologischen Eigenschaften, als ER⁺ / *Luminal-like*, *Basal-like*, Her2⁺ und *Normal Breast-like* bezeichnet wurden. Die Bezeichnung „Luminal“ und „Basal“ beschreibt die Ähnlichkeit des Expressionsmusters dieser Tumoren zum Genprofil von zwei ebenfalls untersuchten Zelltypen des Brustepithelgewebes, den Basal- und Luminalzellen, die sich bzgl. der in ihnen exprimierten Gene und Proteine eindeutig voneinander unterscheiden. Zudem weisen die Kategorien hinsichtlich des Östrogenrezeptorstatus Unterschiede auf, Her2⁺- und *Basal-like*-Tumoren sind z.B. rezeptornegativ, Tumoren des ER⁺ / *Luminal-like*-Typs zeigen in der Regel eine geringe Expression des Wachstumsfaktorrezeptors Her2-neu, im Gegensatz zur Her2⁺-Gruppe. Dieses „**Intrinsic Subtype Model**“ mit seiner 427 Gene umfassenden Signatur wurde von der Forschergruppe anschließend um zwei molekulare Subtypen erweitert (Sørlie et al., 2001). Sie differenzierten anhand der intrinsischen Gensignatur zusätzlich zwischen *Luminal A*- und *Luminal B*-Tumoren, die sich auch bzgl. ihrer Prognose unterschieden. Patientinnen mit *Luminal B*-Tumoren wiesen in einen signifikant schlechteren Krankheitsverlauf auf, was auch in einer weiteren Folgestudie bestätigt wurde (Sørlie et al., 2003).

Eine weitere umfangreiche Studie ist die Arbeit von van't Veer et al. (2002), die anhand eines 60mer-Oligonukleotid-Mikroarrays mit RNA bzw. ESTs von 25.000 Genen ein Set prognostisch relevanter Markergene identifizierten. Dafür wurden zunächst unüberwachte Cluster-Algorithmen (agglomeratives hierarchisches Clustern) auf 5.000 signifikant (mind. zweifach reguliert, $p < 0,01$ in mind. fünf Experimenten) regulierten Genen angewandt. Anschließend wurde mittels überwachter Verfahren ein Genprofil für Tumoren von

Patientinnen mit schlechter Prognose (Metastasenbildung innerhalb von 5 Jahren) identifiziert. Dieses Profil wurde mit jeder Probe des Tumor-Trainings-Sets validiert und so ein optimales Genset mit 70 Genen zur Unterscheidung von Tumoren mit schlechter bzw. guter Prognose ermittelt. Auf Basis dieser **70-Gen-Signatur** konnte im Rahmen der Studie eine ungünstige Prognose für die Patientinnen zuverlässiger vorhergesagt werden als mit den klassischen Faktoren (negativer Östrogenrezeptorstatus, hoher Tumordifferenzierungsgrad, Tumor größer als 2cm, Alter der Patientin <40 Jahre). In die van't Veer-Studie wurden 117 ausschließlich junge Frauen (unter 55 Jahren) mit primär auftretendem sporadischem Mammakarzinom, nahezu alle ohne systemische Behandlung (fünf Patientinnen mit adjuvanter Chemo- oder Hormontherapie), eingeschlossen. Eine interne Evaluation der Studie erfolgte anhand eines unabhängigen Testsets von 19 Tumorproben junger, lymphknotennegativer Patientinnen. Eine Validierung der Signatur folgte in einer größer angelegten Studie mit 295 Patientinnen (van de Vijver et al., 2002). Auch das TRANSBIG(*Translational Research Breast International Group*)-Consortium hat als externe Instanz die 70-Gen-Signatur in zusätzlichen Analysen (MIN-DACT(*Microarray In Node-negative Disease may Avoid Chemotherapy*)-Studie) als unabhängigen Prognosefaktor bestätigt und seine klinische Anwendbarkeit untersucht (Buyse et al., 2006). Die 70-Gen-Signatur wurde bereits in einen maßgeschneiderten, für den Hochdurchsatz geeigneten Mikroarray überführt (Glas et al., 2006) und steht unter dem Namen *MammaPrint*[®] als diagnostischer Test zur Untersuchung von Brustkrebspatientinnen unter 61 Jahren mit negativem Lymphknotenstatus zur Verfügung.

Ebenfalls bereits in der klinischen Anwendung ist der auf 21 Genen basierende „**Recurrence Score**“ von Paik et al. (2004). Dieser speziell für Patientinnen mit lymphknotennegativen, hormonrezeptorpositiven Tumoren entwickelte und unter dem Namen *Oncotype DX*[®] verfügbare Test untersucht die Expression von 16 krebsassoziierten und fünf Referenzgenen mittels qRT-PCR zur Vorhersage des Erfolgs einer Anti-Hormontherapie. Auch hier wurden externe Validierungsstudien durchgeführt, in denen insgesamt bereits fast 4.000 Patientinnen untersucht wurden. Des Weiteren sollen in den noch laufenden TAILORx(*Trial Assigning Individualized Options for Treatment (Rx)*)-Studien insgesamt mehr als 10.000 Patientinnen eingeschlossen werden, um den Nutzen des diagnostischen Tests in der Praxis zu bestätigen, und nicht nur den Erfolg einer Anti-Hormontherapie, sondern auch einer zusätzlichen Chemotherapie abzuschätzen (Paik, 2007; Sparano und Paik, 2008).

Die hier genannten Gensignaturen geben einen guten Überblick über die Bandbreite der Expressionsstudien. Sie unterscheiden sich nicht nur in der angewandten Technik (PCR, Mikroarrays, verschiedene Mikroarrayplattformen), sondern auch bzgl. des benötigten Ausgangsmaterials (frisches oder fixiertes Gewebe), der identifizierten Markergene, der beobachteten Endpunkte (Wiedererkrankungsrisiko, Therapieansprechen) und des Patientinnenkollektivs (Alter, Hormonrezeptorstatus etc.). Auf eine mögliche Darstellung weiterer Gensignaturen, wie z.B. die 76-Gen-Signatur von Wang et al. (2005) zur Vorhersage von Metastasen in lymphknotennegativem Brustkrebs, die Zellzyklus-Gensignatur nach Liu et al. (2008) oder die „Wound Response“ Signatur von Chang et al. (2005), wird daher an dieser Stelle verzichtet, da sie über den Fokus der Arbeit hinausgehen. Im folgenden Kapitel soll nun auf die molekularen Prozesse eingegangen werden, die für die Tumorentwicklung von Bedeutung sind und sich z.T. in den Genprofilen der verschiedenen Untersuchungsansätze widerspiegeln.

2.4 PROZESSE IN DER TUMORZELLE

Charakteristische Kennzeichen von Krebszellen sind unkontrollierte Vermehrung, Veränderungen in Stoffwechsel- und Differenzierungsvorgängen und genomische Instabilität. Hanahan und Weinberg (2000) führen sechs grundlegende Charakteristika an, deren Zusammenwirken eine Tumorentstehung ermöglicht:

- ▷ Zellwachstum unabhängig von exogener Stimulation durch Signalmoleküle (Eingreifen in die Signalkaskade z.B. über Deregulation von Rezeptoren oder Nachahmung von Wachstumsfaktoren)
- ▷ Nicht-Ansprechen auf Anti-Proliferationssignale durch Veränderung zellzyklusregulierender Komponenten
- ▷ Deregulation des programmierten Zelltods
- ▷ Ungehinderte Zellproliferation
- ▷ Fähigkeit zur Angiogenese
- ▷ Gewebe-Invasion und Metastasierung (z.B. durch Veränderung von Faktoren der Zell-Zell- und Zell-Matrix-Interaktion)

Kennzeichen bösartiger Tumoren ist insbesondere die Fähigkeit, nicht nur auf den Entstehungsort (Primärtumor) beschränkt zu bleiben, sondern in umliegendes Gewebe einzudringen (**Invasion**) oder entfernte Organe zu

befallen (**Metastasierung**). Nach dem derzeitigen Kenntnisstand handelt es sich hierbei um mehrstufige Prozesse, bei denen Wechselwirkungen zum angrenzenden Gewebe entscheidende Bedeutung haben. Der Einfluss dieser Stroma-Zellen ist auch Basis der „*Seed and Soil*“-Hypothese zur Metastasierung, die bereits Ende des 19. Jahrhunderts (1889) von Stephen Paget aufgestellt wurde (vgl. Dell, 2006). Sie geht davon aus, dass die Ausbreitung von Tumorzellen (*Seed*) über das Blut- oder Lymphsystem im Körper nicht zufällig erfolgt, sondern dass manche Gewebe einen geeigneteren „Nährboden“ (*Soil*) für Metastasen bieten als andere. Im **Mehrstufen-Prozess der Metastasierung** (*Metastatic Cascade*) stellen Tumor-Stroma-Interaktionen einen wichtigen Aspekt dar, es werden jedoch etliche weitere Schritte zur Tumor-Progression angenommen. Zunächst wächst der Primärtumor (Proliferation), was mit einem Auflockern der Zell-Zell- und Zell-Matrix-Verbindungen bzw. dem Degradieren der extrazellulären Matrix einhergeht. Daraufhin kann die Invasion des Tumors in angrenzendes Gewebe erfolgen. Der Tumor kann sich als Ganzes ausdehnen und umliegendes Gewebe verdrängen oder Zellgruppen bzw. einzelne Zellen können sich vom Tumor ablösen und abwandern. Treffen diese Zellen auf Blut- oder Lymphgefäße, können sie über den entsprechenden Kreislauf auch entfernte Organe erreichen, sich dort zusätzlich anlagern und zur Bildung von Metastasen führen. Metastasierung ist kein linearer Prozess. Jeder der hier genannten Schritte ist von einer Vielzahl von Faktoren beeinflusst und setzt bestimmte Eigenschaften der Tumorzellen voraus bzw. selektiert Zellen anhand konkreter Merkmale, die zum Durchlaufen dieser Prozesse notwendig sind. Beispielsweise geht Tumorwachstum häufig auch mit Angiogenese, der Bildung neuer Gefäßsysteme, einher. Das Eindringen von Tumorzellen in Blutgefäße ist daher nicht nur von ihrem „zufälligen“ Aufeinandertreffen abhängig, sondern wird zusätzlich durch durch den Tumor verursachte Veränderungen im Gewebe gefördert (Schulz, 2005).

Eine zentrale und bisher ungeklärte Frage ist, was den Prozess der Metastasierung antreibt. Hierzu gibt es zwei grundlegende Hypothesen, die Hypothese des klonalen Ursprungs und die Stammzell-Hypothese (Abbildung 2.4).

Anhänger der **Hypothese des klonalen Ursprungs** gehen davon aus, dass im Laufe der Tumorentwicklung innerhalb eines Tumors (heterogene) Zell-Subklone entstehen, die metastatische Potenz besitzen. Kennzeichen dieser Zellen ist ihre genomische Instabilität. Durch Anhäufung zufälliger Mutationen, die einen selektiven Vorteil bedeuten können, entstehen Zellen, die schließlich eine Metastase bilden können. Dieser Prozess ist durch eine fort-

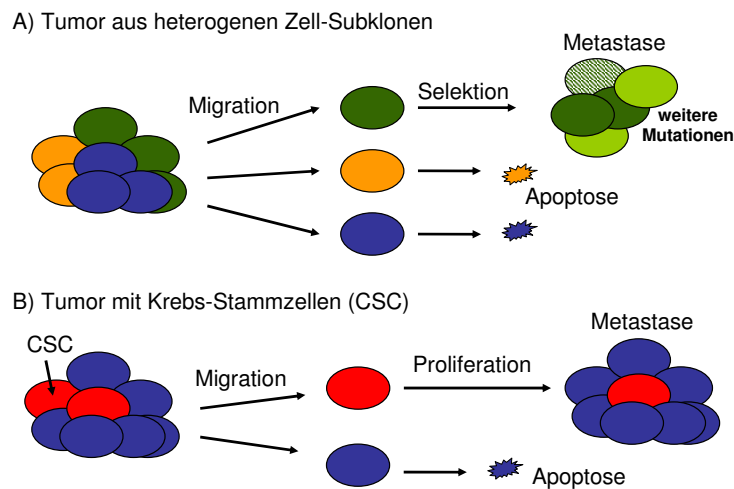


Abbildung 2.4: Dargestellt sind zwei Modelle zur Metastasierung von Tumoren, die Hypothese des klonalen Ursprungs (A) und die Stammzell-Hypothese (B).

laufende Veränderung der Subklone durch Mutationen charakterisiert und spricht für die große Heterogenität innerhalb eines Tumors (**intratumorale Heterogenität**) (Allred et al., 2008; Polyak, 2008). Das **Stammzell-Modell** hingegen geht davon aus, dass nur wenige Zellen aus dem Tumor eine inhärente metastatische Potenz besitzen. Diese Zellen weisen mit der Fähigkeit zur Selbsterneuerung oder der Wachstumsförderung anderer Zellen sogenannte Stammzell-Charakteristika auf. Nur aus Tumor-Stammzellen können Metastasen hervorgehen. Für diese Hypothese spricht, insbesondere mit Blick auf die Brustkrebserkrankung, dass es möglich ist, mittels Genexpressionsanalysen die metastatische Potenz von primären (Brust-)Tumoren vorherzusagen (vgl. hierzu die Vorhersage einer schlechten Prognose von van't Veer et al., 2002); siehe auch Abschnitt 2.3). Dies lässt darauf schließen, dass die metastatische Potenz des Primärtumors bereits in der frühen Phase der Tumorentwicklung festgelegt ist. Des Weiteren ist diese Hypothese eine mögliche Erklärung für die Beobachtung, dass sich Primärtumor und Metastasen eines Tumors in ihren Genprofilen ähnlicher sind als zwei Primärtumoren (gleichen Tumortyps) unterschiedlicher Patientinnen (**intertumorale Heterogenität**) (Schulz, 2005; Li et al., 2007; Allred et al., 2008; Polyak, 2008). Letztendlich schließen sich die Theorien nicht aus und können grundsätzlich beide zum Verständnis der Metastasierungs-Mechanismen beitragen (Allred et al., 2008).

Als sicher gilt, dass **Veränderungen im genetischen Programm** der Zellen die Tumorentstehung fördern. Krebszellen können sowohl Veränderungen in einzelnen Genen (Genmutationen wie z.B. Punktmutation, Insertion, Deletion) als auch strukturelle- (Chromosomenmutationen wie Translokation,

Inversion und Genamplifikation) und numerische chromosomale Veränderungen (Genommutationen) aufweisen. Der Einfluss dieser genetischen Veränderungen auf die Krebsentwicklung ist vielfältig und abhängig von der jeweiligen Funktion der betroffenen Gene. Schulz (2005) unterscheidet zwischen folgenden Kategorien von „Krebsgenen“:

- ▷ Tumorsuppressor- und Onkogene: diese Gene sind in der Regel direkt an der Tumorentstehung beteiligt
- ▷ Risikomodulierende Gene: ihre Bedeutung ist oft abhängig von Umweltfaktoren, z.B. der Exposition gegenüber Karzinogenen; sie beeinflussen Tumorentstehung und -wachstum indirekt z.B. über die Immunantwort oder Hormon- und Wachstumsfaktor-Level
- ▷ „Ausführende“ Gene: sie spielen häufig erst in der späteren Phase der Tumorentwicklung (Zellinvasion und Metastasierung) eine Rolle und werden durch Onkogene oder Tumorsuppressorgene reguliert

Onkogene entstehen durch genetische Veränderungen wie z.B. Mutationen und/oder Genamplifikation in normalen Genen, sogenannten Proto-Onkogenen, die zu einer gesteigerten Aktivität bzw. Überexpression dieses Gens führen. Die meisten Onkogene sind in Mechanismen der Zellzyklusregulation involviert. Es können eine Vielzahl von Molekülen betroffen sein, sezernierte Proteine, Transmembranrezeptoren, Proteinkinasen, Wachstums- oder Transkriptionsfaktoren etc. Hierzu gehört z.B. auch ERBB2 als ein Gen der ERBB-Wachstumsfaktorrezeptorfamilie, das in Brusttumoren eine besondere Rolle spielt (Schulz, 2005, siehe auch Abschnitt 2.2). Im Gegensatz zu der Über-Aktivierung von Onkogenen trägt im Falle von **Tumorsuppressorgenen** der Funktionsverlust zum Entstehen oder Fortschreiten eines Tumors bei. Hier sind die Auswirkungen genetischer Veränderungen nur dann sichtbar, wenn beide Genkopien ausgeschaltet sind (*Loss of Heterozygosity*), da Tumorsuppressorgene, anders als Onkogene, zumeist rezessiv sind. Tumorsuppressorgene lassen sich in zwei große Kategorien einteilen, die „**Caretaker**“ und die „**Gatekeeper**“. Caretaker codieren für Proteine, die das Genom stabilisieren und z.B. an DNA-Reparaturmechanismen beteiligt sind. Mutationen in diesen Genen führen häufig zu weiteren genetischen Veränderungen in anderen Genen, z.B. Gatekeeper-Genen. Gatekeeper-Gene verhindern direkt unkontrolliertes Wachstum von Zellen oder sind für den programmierten Zelltod, die Apoptose, verantwortlich. Das für den Transkriptionsfaktor p53 codierende TP53-Gen ist ein Tumorsuppressorgen, das sehr häufig in Tumoren deaktiviert ist. Es reguliert Gene der Zellzyklus-

regulation, Apoptose, Angiogenese und der DNA-Reparatur und kann so direkt die Entstehung und das Wachstum von Tumoren begünstigen. Diese Kategorisierung der Tumorsuppressorgene ist jedoch nicht ausschließend, einige relevante Gene (so z.B. die BRCA-Gene in Brustkrebs) lassen sich nicht nur einer, sondern beiden Klassen zuordnen (Oliveira et al., 2005).

Häufig in Tumoren veränderte Gene werden allgemein auch als **krebsskritische Gene** bezeichnet. In der Regel lösen nicht Mutationen in einzelnen Genen Krebs aus, sondern fördern Defekte in mehreren Genen die Entstehung invasiver Tumoren. Betroffen sind vor allem Gene, die mit den wichtigsten Stoffwechselprozessen einer normalen Zelle (Proliferation, Differenzierung, Überleben) assoziiert sind und durch Wechselwirkungen auch weitere Gene beeinflussen können (Vogelstein und Kinzler, 2004). Es wird daher auch von **Krebs-Pathways** gesprochen, also von Stoffwechselwegen, die bei der Entwicklung einer Vielzahl von Krebserkrankungen eine Rolle spielen, wie z.B. der in zahlreichen Krebsarten deregulierte MAP-Kinase-Signalweg (Regulation von Zellwachstum, -differenzierung, Apoptose) (Schulz, 2005).

Andere häufig betroffene Stoffwechselwege sind spezifisch für bestimmte Krebsarten. Für die Brustkrebskrankung (und weitere hormonabhängige Karzinome) z.B. der Östrogenstoffwechsel, der auch eng mit dem MAP-Kinase-Signalweg verknüpft ist und ebenfalls Prozesse der Differenzierung, Angiogenese, Apoptose etc. beeinflusst. Östrogen kann dabei genomisch wirken und an Rezeptoren im Zytoplasma binden, was zu deren Dimerisierung und Anlagerung an EREs (*Estrogen Responsive Elements* - spezifische DNA-Erkennungssequenzen) im Zellkern führt. Diese EREs dienen als Bindungsstelle in der Kontrollregion bestimmter Zielgene, die somit reguliert werden. Alternativ binden Östrogen-Vorstufen an membranständige Rezeptoren, die Signale werden über nacheinander geschaltete Kinasen in den Zellkern geleitet und führen dort zur Aktivierung des MAPK-Signaltransduktionsweges (Chen et al., 2008).

Um den Aspekt der Krebs-Pathways auch in den hier vorgenommenen Analysen berücksichtigen zu können, wurde zur funktionalen Einordnung der untersuchten Gene auf Gen-Ontologien zurückgegriffen. Diese werden im nachfolgenden Exkurs kurz erläutert. Auf die verschiedenen (Stoffwechsel-)Prozesse und beispielhafte Vertreter, die Einfluss auf die Brusttumorentwicklung nehmen können, wird im Ergebnisteil dieser Arbeit (Kapitel 4.1.2) näher eingegangen.

EXKURS: GEN-ONTOLOGIEN

Ontologien dienen allgemein dazu, Wissen zu repräsentieren und zu formalisieren. Gen-Ontologien stellen die Struktur zur Verfügung, die eine Kategorisierung von Genen bzw. Genprodukten anhand ihrer biologischen Funktion ermöglicht. Gen-Ontologien liefern hierfür ein kontrolliertes und vereinheitlichtes Vokabular, das auf der Basis von Expertenwissen entwickelt wird. Sie sind speziesunabhängig und ermöglichen die (computergestützte) Verarbeitung und Interpretation komplexer Daten, wie sie zum Beispiel in Mikroarrayanalysen generiert werden. Hierfür müssen nicht nur die Kategorien (Ontologien) selbst festgelegt, sondern auch Hierarchien und Vernetzungen abgebildet werden. Erst in einem weiteren Schritt erfolgt die Annotation von Genen oder Genprodukten, also die Zuordnung zu den erstellten Kategorien.

Es gibt eine Vielzahl unterschiedlicher Strategien zur Repräsentation von Genfunktionen. Thomas et al. (2007) beschreiben zwei komplementäre Ansätze zur Abbildung biochemischer Prozesse und Interaktionen (Pathways). Beide benutzen ein ähnliches Vokabular und bieten Werkzeuge zur Datenanalyse, unterscheiden sich jedoch in Aspekten der Gen-Annotation: Das Modell der **GeneOntology**(GO)-Initiative (Ashburner et al., 2000) enthält drei Hauptkomponenten der Wissensrepräsentation: die molekulare Funktion, assoziierte biologische Prozesse und die Zellkomponenten, in denen diese stattfinden. Die molekulare Funktion ist als konkrete Funktion oder Aktivität eines Genproduktes definiert und unterscheidet sich insofern von den biologischen Prozessen, dass diese eine höhere funktionale Ebene beschreiben, in die einzelne oder auch eine Reihe von molekularen Funktionen involviert sein können (Thomas et al., 2007). Das **PANTHER**(*Protein Analysis Through Evolutionary Relationships*)-System (Thomas et al., 2003) bietet eine GO-ähnliche Repräsentation mit der Möglichkeit, Gene molekularen Funktionen und biologischen Prozessen zuzuordnen. Darüber hinaus beinhaltet sie jedoch auch eine Pathway-Annotation. Beide Projekte bieten u.a. eine webbasierte Datenbank an, an die Abfragen von Ontologien und Annotationsdaten gerichtet werden können, oder auch Werkzeuge zur Visualisierung dieser Daten (URLs: <http://www.geneontology.org/>, <http://www.pantherdb.org/>).

2.5 ZIELE DER ARBEIT

Die vorliegende Arbeit „Genexpressionsanalysen zur Charakterisierung primärer Mammatumoren im neoadjuvanten Setting“ beschäftigt sich mit der Unterscheidung von Tumoren anhand der Expressionsmuster ausgewählter Brustkrebsrelevanter Gene. Hierfür wurden Tumor-Gewebeproben von Brustkrebspatientinnen vor und während einer begleitenden Chemotherapie mit Mikroarraytechnik (und ergänzender Real-Time-PCR) analysiert. Der für die Experimente eingesetzte Mikroarray wurde eigens für diese Analysen hergestellt. Ziel der durchgeführten Untersuchungen ist dabei die Beantwortung folgender Hauptforschungsfragen und damit verbundener konkretisierender Teilfragen:

1. Lassen sich die untersuchten Brusttumoren anhand ihrer Genprofile signifikant voneinander unterscheiden und Gruppen zuordnen?
 - a) Können signifikant exprimierte Gene identifiziert werden?
 - b) Welchen funktionalen Gruppen/Stoffwechselwegen lassen sich diese Gene ggf. zuordnen?
2. Wenn sich die Tumoren verschiedenen Kategorien zuordnen lassen, können diese dann mit klinischen Parametern in Beziehung gesetzt werden?
 - a) Sind bestimmte Tumoreigenschaften mit den Genprofilen assoziiert?
 - b) Gibt es z.B. Korrelationen zu einzelnen klinischen Parametern (etablierte/diskutierte Faktoren) oder auch zu klinischen Profilen (z.B. Risikogruppen, Tumorsubtypen; vgl. Kapitel 2.3)?
 - c) Geben die klinischen Eigenschaften der Tumoren einer auf Grundlage des Genprofils erstellten Kategorie Hinweise auf die Prognose der Patientinnen?
 - d) Unterscheiden sich Tumorproben einer Patientin hinsichtlich ihres Genprofils im Laufe der Chemotherapie?

Nach der experimentellen Bearbeitung dieser Fragen werden die Ergebnisse im aktuellen Forschungsrahmen dargestellt und erläutert. Zudem wird prinzipiell über die Anwendung von Genexpressionsanalysen in Tumordiagnostik und Therapieentscheidung und die damit verbundenen möglichen Vor- und Nachteile diskutiert.

MATERIAL UND METHODEN

Im diesem Kapitel werden die in dieser Arbeit angewandten Methoden und alle hierfür verwendeten Materialien aufgeführt. Zunächst werden das verwendete Gewebe näher charakterisiert, die zur Verfügung stehenden zugehörigen klinischen Daten kurz erläutert und die zu analysierenden Gene beschrieben. Da es sich hierbei um die eigentlichen Untersuchungsgegenstände handelt werden diese nicht mit dem zusätzlichen Material zusammen aufgelistet, sondern getrennt betrachtet.

3.1 PATIENTINNENKOLLEKTIV UND TUMORPROBEN

Im Rahmen der vorliegenden Studie wurden Tumorproben von Patientinnen analysiert, die sich im Zeitraum von 2006-2008 im Brustzentrum der Städtischen Kliniken Bielefeld in Behandlung befanden. Bei allen Patientinnen lag ein primäres Mammakarzinom vor, das mittels primär systemischer (neoadjuvanter) Chemotherapie therapiert wurde. Von jeder Patientin wurden Gewebeproben an bis zu drei verschiedenen Zeitpunkten entnommen. Eine erste Stanzbiopsie (entnommen unter Ultraschallkontrolle) wurde jeweils im Rahmen der Routinediagnostik beim Erstbesuch durchgeführt. Eine zweite Stanze wurde in der Regel nach dem ersten Zyklus der Chemotherapie entnommen. Im Falle einer operativen Entfernung des Tumors stand noch ein Operationspräparat für die Analysen zur Verfügung.

Zur Unterscheidung der einzelnen Tumorproben und Identifizierung der Patientinnen wurden interne Bezeichnungen vergeben. Die Patientinnen wurden zunächst durchnummeriert, zur Kennzeichnung der unterschiedlichen Proben einer Patientin wird zusätzlich eine Zahlen-Buchstaben-Kombination verwendet, wobei durch Buchstaben Proben unterschiedlicher Zeitpunkte und durch angehängte Zahlen mehrere Proben eines Zeitpunktes unterschied-

den werden: 5-C-2 beschreibt z.B. die zweite Probe des dritten Entnahmezeitpunktes der Patientin 5.

Für die späteren Analysen werden dieser Bezeichnung noch Informationen über die Probennummer (die in der Klinik vergebene, 3- bis 4-stellige, fortlaufende Nummer der Gewebeprobe) und Art der Tumorprobe (BP-0 oder BP-1 = Biopsie vor bzw. nach Chemotherapie; OP = Operationspräparat) als Identifikation hinzugefügt: 5-C-2_1035_OP (vgl. „EMMA-Value“ in Anhang D).

3.1.1 Auswahlkriterien

Folgende Einschlusskriterien mussten für die Studienteilnahme erfüllt werden:

- ▷ Schriftliche Einwilligungserklärung
- ▷ Unilaterales oder bilaterales primäres Mammakarzinom mit Indikation zu primär systemischer Therapie¹
- ▷ Keine Kontra-Indikation gegen primär systemische Therapie wie eine schwere vorliegende Allgemeinerkrankung oder Allergie gegen verwendete Medikamente
- ▷ Alter über 18 Jahre
- ▷ Karnofsky-Index² größer 80 %
- ▷ Normale Leberfunktion, Bilirubin (Abbauprodukt des Hämoglobins) im Normbereich (unter 21 $\mu\text{mol/L}$ Serum)
- ▷ negativer Schwangerschaftstest bei prämenopausalen Frauen

Ausgeschlossen wurden Patientinnen mit schweren Grunderkrankungen oder vorausgegangenen Chemotherapien sowie Frauen in der Stillzeit.

1 Grundlage für die Indikation zur systemischen Therapie waren die Leitlinien der St. Gallen Konsensus Konferenz 2005. Diese empfehlen eine (neo-)adjuvante Chemotherapie für alle Patientinnen der hohen oder mittleren Risikokategorie. Für die Einteilung der Risikogruppen sowie den Behandlungsempfehlungen siehe Goldhirsch et al. (2005) und Abschnitt 2.2 der Einleitung

2 Der Karnofsky-Index (*Karnofsky Performance Status Scale*) nach Schag et al. (1984) dient zur Bewertung der Lebensqualität von Patienten mit bösartigen Tumoren und soll einen gewissen Grad der Standardisierung dieses abstrakten Faktors ermöglichen. Die Skala reicht von maximal 100 % (keinerlei Einschränkungen durch die Erkrankung) in 10er-Schritten bis zu 0 % (Tod).

3.1.2 *Patientinnenbezogene Daten*

Soweit möglich wurden folgende patientinnenbezogene Daten erfasst:

- ▷ Angaben zu Person und Herkunft: Alter, Beruf, Geburtsland, Volksgruppenzugehörigkeit
- ▷ Hormonabhängige/reproduktive Faktoren: Gravidität (Schwangerschaften), Parität (Geburten), Alter bei Geburt des ersten Kindes, Stillen und Zeitraum des Stillens, Menarche, Menopause
- ▷ Lebensstilfaktoren: Rauchen, Alkoholkonsum, *Body-Mass-Index* (BMI), Besonderheiten der Ernährung, Umfang sportlicher Betätigung, Schlafverhalten
- ▷ Klinische Anamnese: familiärer Brustkrebs, vorhergegangene Krebserkrankung, Allgemeinerkrankungen, Medikamente, Eigenmedikation

Eine Zusammenfassung der Merkmale des Patientinnenkollektivs liefert der Ergebnisteil dieser Arbeit (4.2). In Abschnitt B des Anhangs sind die in die Analysen eingegangenen Tumor- und Patientinnendaten dokumentiert.

3.1.3 *Tumormaterial und Referenz*

Das untersuchte Tumormaterial umfasst sowohl Stanzzyylinder als auch OP-Präparate. Das Gewebe wurde jeweils sofort nach Entnahme in den Städtischen Kliniken Bielefeld in flüssigem Stickstoff (-196 °C) eingefroren und gelagert. Die weitere Lagerung an der Universität fand in -80 °C Gefrierschränken statt, der Transport erfolgte dabei auf Trockeneis (-79 °C). Als Referenz zum Tumormaterial diente die Universal Human Referenz-RNA von Stratagene. Diese stellt einen Pool aus zehn verschiedenen Tumorzelllinien dar, wobei die Herstellerfirma eine gleich bleibende Qualität der RNA garantiert (Novoradovskaya et al., 2004). Alle Experimente der vorliegenden Arbeit wurden mit einem Pool dieser RNA durchgeführt.

3.1.4 *Tumorbezogene Daten*

Neben den patientinnenbezogenen Daten stehen noch eine Reihe weiterer tumorbezogener Daten zur Verfügung, die nachfolgend kurz aufgelistet werden.

- ▷ Tumorlokalisation (Quadrant der Brust) und Histologie (duktal, lobulär)

- ▷ Tumorgröße, Lymphknotenbefall und Fernmetastasierung (TNM-Klassifikation)
- ▷ Differenzierung der Tumorzellen (Grading)
- ▷ Status für den Progesteron- und Östrogenrezeptor sowie den Wachstumsfaktorrezeptor Her2-neu
- ▷ Chemoregime, Herceptinbehandlung

Die immunhistochemische Bestimmung des Hormonrezeptorstatus und der eventuelle Nachweis einer Genamplifikation des Wachstumsfaktorrezeptors Her2-neu mittels Fluoreszenz *in situ* Hybridisierung (FISH) wurde in der Pathologie des Städtischen Klinikums Bielefeld durchgeführt. Die dort benutzten Methoden werden nachfolgend kurz dargestellt, auf alle weiteren tumorbezogenen Daten wird in Kapitel 2.2 eingegangen.

HISTOPATHOLOGIE Der Antikörpernachweis erfolgte mit dem BenchMark IHC/ISH Färbemodul von Ventana Medical Systems unter Verwendung des BMK iVIEW DAB Paraffin Detection Kits (Ventana Medical Systems) für die Progesteron- und Östrogenrezeptoren und HercepTest® (Dako, Genentech) für den Her2-neu-Nachweis. Die Auswertung der Antikörperfärbung für die Progesteron- und Östrogenrezeptoren erfolgt semiquantitativ. Angegeben wird der immunreaktive Score (IRS) nach Remmele und Stegner (1987), der als Produkt der Ergebnisse für die Anzahl positiver Zellen und die Färbeintensität einen Wert zwischen 0 und 12 liefert, wie in Tabelle 3.1 dargestellt (Tabelle angepasst nach Kreienberg et al., 2004, aktualisierte Fassung 2008).

Tabelle 3.1: Immunreaktiver Score nach Remmele und Stegner (1987) berechnet aus dem Prozentsatz angefärbter Zellkerne multipliziert mit der Färbeintensität.

Prozentsatz positiver Zellkerne	X Färbeintensität
Keine positiven Kerne 0 Punkte	Keine Färbereaktion 0 Punkte
< 10 % positive Kerne 1 Punkt	Schwache Färbereaktion 1 Punkt
10-50 % positive Kerne 2 Punkte	Mäßige Färbereaktion 2 Punkte
51-80 % positive Kerne 3 Punkte	Starke Färbereaktion 3 Punkte
> 80 % positive Kerne 4 Punkte	

Die Auswertung des immunhistochemischen Nachweises mittels HercepTest® zur Bestimmung des Her2-neu-Status ist in Tabelle 3.2 dargestellt.

Bei einem schwach positivem Farbeergebnis wird zusätzlich ein Nachweis mittels FISH und dem Inform® Her2-neu plus Sondenkit durchgeführt. Von einer Genamplifikation und einer damit verbundenen Her2-neu-

Tabelle 3.2: Bewertungsschema der Her2-neu-Immunhistochemie.

HercepTest Score	Reaktionsmuster	Bewertung
0+	Keine Färbereaktion oder <10 % der invasiven Tumorzellen mit Markierung der Zellmembran	Negativ
1+	>10 % der invasiven Tumorzellen mit schwacher inkompletter Markierung der Zellmembran	Negativ
2+	>10 % der invasiven Tumorzellen mit zirkulärer Markierung der Zellmembran; Färbintensität gering bis mittelgradig	Schwach positiv (geringe HER2-neu-Überexpression)
3+	>10 % der invasiven Tumorzellen mit zirkulärer Markierung der Zellmembran; Färbintensität stark	Stark positiv (starke HER2-neu-Überexpression)

Überexpression wird bei Verwendung dieses Sondenkits bei mehr als vier Signalen pro Zelle ausgegangen (Kreienberg et al., 2004).

3.2 THEMENMIKROARRAY

Bei dem für die vorliegende Arbeit verwendeten Mikroarray handelt es sich um einen Brustkrebs-Themenarray. Er enthält ein Set brustkrebsrelevanter Gene, das speziell für diese Anwendung zusammengestellt wurde.

Die Auswahl der Gene erfolgte auf Basis bereits veröffentlichter Studien. Hauptsächlich ist hier die Studie von van't Veer et al. (2002), publiziert in *Nature*, zu nennen (van't Veer et al., 2002). Dieses Genset wurde basierend auf Ergebnissen weiterer Studien ergänzt. Die Ergebnisse der hierzu erfolgten Literaturrecherchen und weitere Informationen zum Themenarray finden sich im Ergebnisteil unter 4.1. Eine Übersicht über die analysierten Gene und ihre biologische Funktion ist in Anhang A dargestellt.

Als Trägermaterial zur Herstellung der Mikroarrays wurde der Nexterion® Slide E MPX 16 von Schott Nexterion verwendet. Die Epoxy-Oberfläche erlaubt kovalentes und direktes Binden von Molekülen, wie z.B. Oligonukleotiden, ein ultrahydrophobes Gitter unterteilt die Oberfläche des Objektträgers in 16 individuelle Kammern. Bedingt durch den Druckprozess und aus

Gründen der Handhabbarkeit wurden für alle durchgeführten Experimente jedoch nur 12 Felder des Arrays benutzt. Die 12 Raster des Themenarrays bestehen jeweils aus 24 Spalten mal 27 Zeilen und wurden mit einer Nadel in zwei Durchgängen mit dem MicroGrid II 610 (Genomic Solutions) erstellt (Herstellung der Mikroarrays durch Prof. Dr. Anke Becker, Universität Freiburg). Jedes dieser Raster beinhaltet 216 verschiedene Elemente, die in drei Replikaten mit je 40 nM Oligonukleotide in 3 x SSC/1,5 M Betain gedruckt wurden (648 *Spots* pro Feld). Im Rahmen der vorliegenden Arbeit wurden Mikroarrays aus zwei verschiedenen Druck-Serien eingesetzt, die Objektträger wurden jeweils durchnummeriert und mit der entsprechenden Seriennummer versehen (Mc2001-Mc2008 S01; Mc2009-Mc2016 S02). Die Felder jedes Mikroarrays werden zur Unterscheidung der einzelnen Experimente mit Spalte A Felder 1-6 und Spalte B, 1-6 benannt (vgl. Abbildung 3.1).

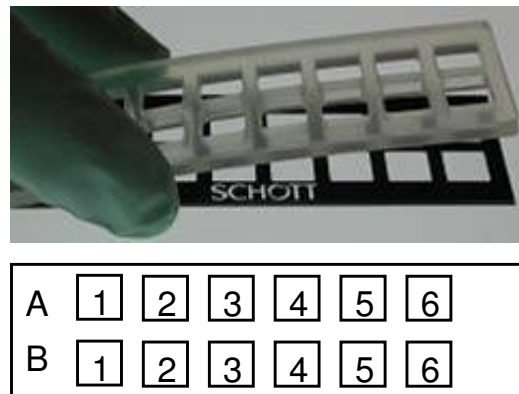


Abbildung 3.1: Unterteilung des Schott Nexterion Slides E MPX 16.

3.2.1 Geninformation und -sequenzen

Die 216 verschiedenen Elemente des Mikroarrays repräsentieren brustkrebsrelevante Gene und verschiedene Kontroll-Elemente, die jeweils in drei Replikaten pro Raster vorliegen:

- ▷ „Alien“-Oligonukleotide (Neomycin Phosphotransferase II (nptII) aus dem Transposon Tn5 und Gentamicin Acetyltransferase-3-I(AAC(3)-I) (aacC1)) des Transposons Tn1696 aus *Sinorhizobium meliloti*, jeweils dreimal Puffer des Druckvorganges und Leer-Kontrollen (*Empty Spots*) als interne Negativkontrollen
- ▷ Filamin Beta (FLNB), Hydroxymethylbilane Synthase (HMBS), Beta-2-Microglobulin (B2M), Lysyl-tRNA Synthetase (KARS), Aktin Beta und

Non-POU domain containing, octamer-binding (NONO) als ubiquitär exprimierte humane Haushaltsgene

- ▷ 163 Gene aus verschiedenen Studien, z.B:
 - ABC-Transporter, die eine Rolle bei Chemotherapieresistenzen spielen
 - Wachstumsfaktoren (z.B. *Insulin-like Growth Factors*) und ihre Rezeptoren
 - Zykline als weitere Gene der Zellzyklusregulation
 - Apoptose-Gene (z.B. TP53, BCL2)
 - Zelladhäsions- und extrazelluläre Matrix-Proteine (z.B. Cadherine), die eine wichtige Rolle für Tumorwachstum und Metastasierung spielen
 - Gene des Östrogenstoffwechsels, z.B. Steroidhormone die einen Einfluss auf Tumorentwicklung und Progression haben können
 - Angiogenesefaktoren wie z.B. Angiopoietin, VEGF (*Vascular Endothelial Growth Factor*) und PDGF (*Platelet-derived Growth Factor*), die für die Bildung neuer Blutgefäße und damit für die Versorgung des Tumors essentiell sind
 - verschiedene Matrix-Metalloproteinasen, die ebenfalls in Metastasierungsprozesse involviert sind
- ▷ 19 cDNA-Klone (van't Veer et al., 2002) in jeweils zwei Orientierungen, um sicher zu gehen, dass sich die Sequenz auf dem kodierenden Sense-Strang befindet

Grundlage für die Auswahl der 70mer-Oligonukleotide der ausgewählten Brustkrebsrelevanten Gene waren GenBank Sequenzinformationen (NCBI Reference Sequences mRNA; URL: <http://www.ncbi.nlm.nih.gov>; Zugriffsdatum: Januar 2007). Auf Basis dieser Zielsequenzen wurden mittels OligoDesigner (Daniel Paarmann, unveröffentlichte Software) repräsentative 70mere der Gene identifiziert. Die Spezifität der Oligonukleotide (*Uniqueness*) muss gewährleistet sein, um spätere Kreuzhybridisierungen auf dem Array zu vermeiden, daher erfolgt der Abgleich mit einer Referenzsequenz. Hierfür wurde die komplette Sequenz des menschlichen Genoms aus der TIGR Gene Index Database (The Institute for Genomic Research, Release 16; URL: <http://www.tigr.org/tdb/tgi>; Zugriffsdatum: Januar 2007; aktuelle URL: <http://compbio.dfci.harvard.edu/tgi/>) im OligoDesigner verwendet (Analysen durchgeführt von Burkard Linke, Bioinformatics Resource

Facility, CeBiTec, Universität Bielefeld). Für die Identifizierung möglicher Sequenzabschnitte wurden folgende Kriterien angelegt:

- ▷ Schmelztemperatur T_m der Oligonukleotide, berechnet anhand des GC-Gehalts nach Howley et al. (1979):

$$T_m = \frac{81,5 + (16,6 \times \log[Na^+] \times (A_G + A_C) / \hat{A} - 500)}{\hat{A}} \quad (3.1)$$

$$= 78 \pm 5 \text{ } ^\circ\text{C}$$

wobei $[Na^+] = 0,1M$ ist, A_x jeweils die Anzahl der Basen Adenin (A), Cytosin (C), Guanin (G) oder Thymin (T) und $\hat{A} = A_A + A_C + A_G + A_T$ die Summe aller Anzahlen der Basen angibt

- ▷ Wenn möglich, wird pro Sequenz alle 250 Basen ein geeignetes 70mer angeboten
- ▷ Automatische Entfernung von polyA-/polyT-Sequenzen
- ▷ Mindestabstand vom 5'- und 3'-Ende: 50 Basen
- ▷ Maximal zulässige Übereinstimmung zu anderen Sequenzen: 75 %
- ▷ Maximale Anzahl benachbarter Basen wie bei einem anderem Gen: 14-18 Nukleotide, wobei die maximal zulässige Gesamtübereinstimmung der Sequenzen nicht größer als 70 % sein darf und zwischen identischen Teilsequenzen mindestens 9 Nukleotide Abstand sein müssen
- ▷ Größter Stamm mit Haarnadelstruktur: 8 Basen

Die endgültige Auswahl der anhand dieser Kriterien vorgeschlagenen möglichen Sequenzabschnitte erfolgte per Hand. Soweit möglich wurden Oligonukleotide ausgewählt, die alle Filterkriterien erfüllten. Konnten nicht alle Kriterien eingehalten werden, wurden Sequenzen bevorzugt, die Uniqueness zeigten bzw. sich eher im Bereich des 3'-Endes des Gens befinden (bevorzugter Abstand 200 bp). Schmelztemperatur und Bildung von Haarnadelstrukturen wurden erst in zweiter Priorität berücksichtigt.

Für insgesamt 74 der 206 gedruckten Oligonukleotide konnte kein eindeutiger Oligonukleotid-Status nachgewiesen werden (darunter 16 cDNA-Klone). In vielen Fällen handelt es sich dabei um Mitglieder größerer Genfamilien, weshalb hier keine eindeutigen Sequenzabschnitte im Genom identifiziert werden konnten. Die Oligonukleotide wurden durchnummeriert und es wurde ihnen ein interner Identifikationsname zugeordnet. Dieser Reporter Name „R:“ setzt sich aus der Beschreibung des Gensets „McBiSet2“ und der

Nummer (1-206) zusammen. Eine vollständige Liste der Sequenzen befindet sich im Anhang A.1.

3.2.2 *Gen-Ontologie*

Als Grundlage für eine funktionelle Einordnung der Gene werden diese zunächst anhand von bestehenden Ontologien Klassen zugeordnet. Hierfür wurden alle GenBank-Accession-Nummern mit dem PANTHER(*Protein Analysis Through Evolutionary Relationships*)-Index (Thomas et al., 2003) und dem GeneOntology-Datenset (Ashburner et al., 2000), zwei sich ergänzende Ontologien die speziesunabhängig molekulare Funktionen, assoziierte biologische Prozesse und Wirkorte (Zellkomponenten) möglicher Genprodukte beschreiben, abgeglichen. Die Annotation (Zuordnung) von Genen erfolgt literaturgestützt bzw. anhand von Homologien. Da die direkte Eingabe von GenBank-Accession-Nummern in GeneOntology nicht möglich ist, wurde für die Erstellung einer GO-Mapping-Datei, wie sie zur Verwendung in der Clustersoftware Genesis (siehe 3.4.5 und 3.3.8) notwendig ist, die SOURCE Batch Search Funktion benutzt (URL: <http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch>; letzter Zugriff: Februar 2009). Der Zugriff aus dem Programm Genesis erfolgt auf eine lokal hinterlegte GO-Datenbank mit dem Stand von Oktober 2007.

Für die Suche in der PANTHER-Datenbank können die Accessions direkt auf der Internetseite (URL: <http://www.pantherdb.org/genes/batchIdSearch.jsp>; letzter Zugriff: Februar 2009) eingegeben werden. Die Ergebnisse dieser Datenbankvergleiche sind beispielhaft für die biologischen Prozesse im Anhang A.3 wiedergegeben.

3.3 LABORMATERIAL

3.3.1 *Lösungen und Puffer*

Für die Herstellung der Lösungen wird, wenn nicht anders angegeben, deionisiertes Wasser verwendet. Die Abkürzung RT steht für Raumtemperatur.

Lösung	Verwendung	Zusammensetzung
Lysepuffer	RNA-Isolierung	14,3 M β -Mercaptoethanol 10 μ l in 1 ml RLT-Puffer
Erststrang-Puffer (5x, Invitrogen)	cDNA-Synthese	250 mM Tris-HCl, pH 8,3 (RT)/375 mM KCl/15 mM $MgCl_2$
Stopp-Lösung	cDNA-Synthese	0,5 M NaOH/50 mM EDTA
Neutralisierungslösung	cDNA-Synthese	1 M Tris-HCl, pH 7,5
Stopp-Lösung	Ligation (Farbstoffdendrimere)	0,5 M EDTA, pH 8,0
TE-Puffer (1x)	Ligation (Farbstoffdendrimere)	10 mM Tris-HCl, pH 8,0/1 mM EDTA
Spülpuffer 1 (250 ml)	Mikroarray Prozessierung	0,1 % Triton X-100 15 Minuten bei 80 °C lösen, danach auf RT abkühlen lassen
Spülpuffer 2 (500 ml)	Mikroarray Prozessierung	1 mM HCl (aus 32 % HCl)
Spülpuffer 3 (250 ml)	Mikroarray Prozessierung	100 mM KCl (aus 1 M KCl)
Blockingpuffer (1x)	Mikroarray Prozessierung	15 ml H_2O + 4,7 μ l 32 % HCl auf 50 °C vorwärmen 5 Minuten vor Gebrauch 5 ml 4x Blocking Lösung (Nexterion Block E Solution) hinzugeben
20x SSC	Hybridisierung	0,3 M tri-Natriumcitrat/3 M Natriumchlorid
Waschpuffer 0 (100 ml)	cDNA Hybridisierung	2x SSC
Waschpuffer 1 (500 ml)	cDNA Hybridisierung	2x SSC/0,2 % SDS (42 °C)
Waschpuffer 2 (500 ml)	cDNA Hybridisierung	0,2x SSC/0,1 % SDS (RT)
Waschpuffer 3 (500 ml)	cDNA Hybridisierung	0,2x SSC (RT)
Waschpuffer 4 (250 ml)	cDNA Hybridisierung	0,05x SSC (21 °C)
Waschpuffer I (500 ml)	Hybridisierung 3DNA	2x SSC/0,2 % SDS (65 °C)
Waschpuffer II (250 ml)	Hybridisierung 3DNA	2x SSC (RT)
Waschpuffer III (250 ml)	Hybridisierung 3DNA	0,2x SSC

3.3.2 Chemikalien/Sonstiges

Chemikalie	Hersteller
β -Mercaptoethanol	Merck
Ethanol Rotipuran > 99,8 %, p.a.	Roth
flüssiger Stickstoff	Linde
HCl (32 %)	Merck
Nexterion Block E Solution (Blocking Lösung)	Peqlab
RNase-ZAP	Sigma
SDS	Merck
Triton X-100	Sigma
NaOH	Chemikalienlager, Uni Bielefeld
HCl	Chemikalienlager, Uni Bielefeld
EDTA	Sigma
Tris	MP Biomedicals
Lachs-Sperma DNA 5 ng/ μ l	Amersham Biosciences
Universal Human Referenz-RNA	Stratagene

3.3.3 Verbrauchsmaterialien

Material	Hersteller
Autoklavierbare Entsorgungsbeutel	Roth
Filterspitzen (10 μ l, 100 μ l, 1000 μ l)	Biozym
LifterSlips, Deckgläser (24x60 mm) mit Abstandhalter (0,05 mm)	Nunc
PP-Tube 14 ml steril („BlueCap“)	Greiner Bio-one
PP-Tube 50 ml steril („Greiner Röhrchen“)	Greiner Bio-one
Reaktionsgefäße 3810X	Eppendorf
Ribolyser-Tubes (Fast Protein Blue)	Bio 101 Systems, QBioGene
Hybridisierungskammer	Biologie Werkstatt, Universität Bielefeld
Pipette 2 μ l	Eppendorf
Pipetten (10 μ l, 100 μ l, 1000 μ l)	BioHit
Kryostifte	Nalge Nunc International
-80 °C Kühlblock (Quick Freeze)	Roth
Mörser, Pistill, Pinzette, Mikrospatel	Chemiekalienlager, Universität Bielefeld

3.3.4 Enzyme

Enzym	Hersteller
RNase-free DNaseI [1500 Kunitz units ³]	Qiagen
RNase-out [40 U/ μ l]	Invitrogen
BioScript [200 U/ μ l]	Bioline, Invitrogen
Terminal Deoxynucleotidyl Transferase (TdT)	Genisphere
T4 DNA Ligase	Genisphere

3.3.5 Real-Time-PCR-Primer

Primerpaar	Contig	Produktgröße	Hersteller
RT ² PCR-Primer-Set für ERBB2 Human	NM_004448	185 bp	Biomol
RT ² PCR-Primer-Set für ESR1 Human	NM_000125	162 bp	Biomol
RT ² PCR-Primer-Set für PGR Human	NM_000926	105 bp	Biomol
RT ² PCR-Primer-Set für RPL7 Human	NM_000971	106 bp	Biomol
RT ² PCR-Primer-Set für HMBS Human	NM_000190	102 bp	Biomol

³ Eine Kunitz unit bezeichnet die Menge an DNaseI, die einen Anstieg von 0,0001 pro Minute pro Milliliter polymerisierter DNA-Lösung bei 25 °C in A260 bewirkt (Kunitz, 1950)

3.3.6 Komplett-Systeme (Kits)

Kit	Hersteller
RNeasy Fibrous Tissue Mini Kit	Qiagen
MinElute Cleanup Kit	Qiagen
3DNA Array 900MPX Kit (Cy3/Cy5)	Genisphere (Implen)
MinElute PCR Purification Kit	Qiagen
RNA 6000 Nano/Pico LabChip Kit	Agilent
SensiMix One-Step Kit	Quantace

3.3.7 Geräte

Gerät	Hersteller
Bioanalyzer 2001	Agilent Technologies
Heizblock Thermomixer Comfort	Eppendorf
Heizblock Bioer	Biogen
Hybridisierungsschrank HB-1D	Techne
NanoDrop ND-1000 Spektrophotometer	Peqlab
Ribolyser Fast Prep FP120	Qbiogene
Scanner LS-Reloaded	Tecan
Spotter MicroGrid II 610	Genomic Solutions (Biorobotics)
Vakuumrotationsverdampfer (Speed-Vac) SPD 111V	Savant
Waage CP153	Sartorius
Wasserbad TE-10D Tempunit	Techne
Zentrifuge 5417 R	Eppendorf
Zentrifuge Multifuge 3L-R	Heraeus
Opticon	Bio-Rad

3.3.8 Software

Programm	Hersteller
EMMA 2.8, ArrayLIMS http://www.cebitec.uni-bielefeld.de/groups/brf/software/	Michael Dondrup, Bioninformatics Research Facility, CeBiTec Universität Bielefeld
Genesis 1.7.2	Alexander Sturn, Technische Universität Graz
Imagene [®] 6.0	BioDiscovery
SPSS für Windows 16.0	SPSS Inc.
Excel 2002	Microsoft [®]
Access 2002	Microsoft [®]

3.4 METHODEN

Ziel der in dieser Arbeit verwendeten Methoden ist die Bestimmung des Expressionsstatus ausgewählter Gene in Brusttumoren. Dies geschieht jeweils

durch die Analyse der RNA aus Tumorzellen im Vergleich zu einer Referenz-RNA.

Vorrangig wurden die Genexpressionsanalysen mittels Mikroarraytechnik durchgeführt. Zusätzlich wurde, wenn ausreichend RNA für mehrfach Analysen gewonnen werden konnte, für ausgewählte Gene eine Real-Time-PCR durchgeführt. Eine kurze Beschreibung der Methoden, der generelle Ablauf der Experimente und die einzelnen Schritte vom Aufschluss der Tumorzellen bis zur Auswertung der gewonnenen Daten werden im Folgenden dargestellt.

3.4.1 *Gesamt-RNA-Isolierung*

Zunächst wurden alle Tumorproben vor der weiteren Verwendung gewogen. Das Leergewicht der Kryoröhrchen wurde vorher bestimmt und auf den Röhrchen vermerkt.

Die Isolierung der RNA erfolgt mit dem Rneasy Fibrous Tissue Mini Kit von Qiagen. Hierbei wird Gesamt-RNA gewonnen. Da Moleküle >200 Nucleotiden isoliert werden, kommt es zu einer Anreicherung der kodierenden mRNA, die kleineren nicht-kodierenden RNAs (tRNA, rRNAs etc.) werden nicht aufgereinigt.

Der Zellaufschluss erfolgt mechanisch. Das Vorgehen unterscheidet sich hier für OP-Proben und Biopsien: Biopsien werden mit Hilfe eines vorgekühlten Pistills im Kryoröhrchen in einem -80° -Kühlblock zerkleinert und mit Spatel in vorgekühlte Ribolyser-Tubes mit je $500 \mu\text{l}$ Lysepuffer überführt. Die meist größeren OP-Proben werden in einen Plastikbeutel transferiert und auf einem gekühlten Metallblock mittels eines Hammers rasch zertrümmert. Die zerkleinerten Tumorstücke werden anschließend in einen mit flüssigem Stickstoff gefüllten Mörser gegeben, ggf. weiter zerkleinert und von dort mit Hilfe eines Spatels in die vorbereiteten Ribolyser-Tubes überführt.

Das weitere Vorgehen ist für alle Tumorproben identisch: Die Zellen werden im Ribolyser bei einer Geschwindigkeit von $6,5 \text{ m/s}$ je viermal für 30 Sekunden lysiert und nach jedem Disruptionsschritt drei Minuten lang auf Eis gekühlt. Nach der Zelllyse und Homogenisation werden die Ribolyser-Tubes bei 10.000 g und 4°C drei Minuten lang zentrifugiert. Der Überstand wird vollständig in Eppendorf-Reaktionsgefäße ($1,5 \text{ ml}$) überführt und erneut drei Minuten zentrifugiert. Der resultierende Überstand (ca. $300 \mu\text{l}$) wird in eine zweite Serie Reaktionsgefäße überführt und in 2-fachen Volumen Rnase-freiem Wasser aufgenommen. Anschließend erfolgt ein Protein-Verdau

durch Zugabe von 10 μl Proteinase K bei 55 °C für 10 Minuten. Danach wird 3 Minuten bei 8.000 g und 19 °C zentrifugiert und der Überstand in neue Eppendorf-Gefäße pipettiert, bevor die Aufreinigung über MinElute Säulen (MinElute Cleanup Kit, Qiagen) erfolgt. Zunächst wird der Probe 0,5x Volumen Ethanol zugegeben und so die RNA an die Silika-Gel-Matrix der Säule gebunden. Alle weiteren Waschschrte und der DNase-Verdau erfolgen nach dem Rneasy Fibrous Tissue Mini Kit Protokoll (Qiagen Handbuch Nov. 2006). Die Gesamt-RNA wird in 15 μl Rnase-freiem Wasser eluiert und bei -80 °C eingefroren.

3.4.2 RNA Konzentrations- und Qualitätsbestimmung

Qualitätskontrolle und Quantifizierung der Gesamt-RNA erfolgen mittels ND-1000 NanoDrop Spektrophotometer (Pepqlab) und Bioanalyzer 2001 (Agilent Technologies). Zur Durchführung der Messungen wird jeweils nur 1 μl Probe benötigt, es wird nach Herstellerangaben vorgegangen.

Das NanoDrop Spektrophotometer dient zur Konzentrationsbestimmung bis zu 3.000 ng/ μl . Zudem wird die Reinheit der RNA als 260/230- und 260/280-Absorptionsverhältnis angezeigt. Eine Gesamt-RNA Probe von hoher Qualität weist ein 260/230-Verhältnis von >1,8 auf (keine bzw. nur geringe Verunreinigung mit organischen Komponenten). Ein 260/280-Verhältnis von ca. 2,0 zeigt an, dass die Nukleinsäureprobe weitestgehend frei von Proteinen ist.

Der Bioanalyzer ermöglicht zusätzlich die gelelektrophoretische Auftrennung der Gesamt-RNA im Chip-Format. Das RNA 6000 Nano LabChip® Kit ist für RNA-Konzentrationsgehalte von 25-500 ng/ μl geeignet und das RNA 6000 Pico Kit für Konzentrationen von 50-5.000 pg/ μl . Neben der Konzentrationsbestimmung wird die Integrität der Gesamt-RNA bestimmt (*RNA Integrity Number*, RIN). Die Integrität der Gesamt-RNA wird anhand des Ratios der ribosomalen RNA (28S/18S-Verhältnis für intakte RNA 2:1) sowie unter Berücksichtigung des gesamten Elektrohoreselaufs und evtl. vorhandener Degradationsprodukte bestimmt. Eine RIN von 1 zeigt degradierte RNA an, eine RIN von 10 steht für Gesamt-RNA mit hervorragender Qualität.

3.4.3 Mikroarrayhybridisierung

Das Prinzip der Hybridisierung beruht auf der Zusammenlagerung komplementärer einzelsträngiger Nukleinsäuremoleküle. Mikroarrays bestehen aus einem Trägermaterial auf dem für ein bestimmtes Gen repräsentative Sequenzen angeheftet sind. Durch die Bindung von DNA-Transkripten aus dem zu untersuchenden Zellmaterial kann die Expression bestimmter Gene, z.B. in einem Gewebe, nachgewiesen werden. Eine erfolgte Anlagerung der Moleküle wird dabei über den Nachweis von in die Ziel-cDNA eingebauten Fluoreszenzfarbstoffen gemessen. Zu den hier eingesetzten Mikroarrays siehe Abschnitt 3.2.

Von der reversen Transkription der Tumor- und Referenz-RNA bis hin zur Fluoreszenzmarkierung mittels Farbstoff-Dendrimeren erfolgen alle Schritte mit dem 3DNA Array 900MPX Kit von Genisphere. Notwendige Aufreinigungsschritte der cDNA werden mit dem MinElute PCR Purification Kit (Qiagen) durchgeführt. Das 3DNA Array 900MPX Kit ist insbesondere für geringe Mengen oder auch zum Teil degradierte RNA geeignet.

3.4.3.1 cDNA-Synthese und Fluoreszenzmarkierung

Zunächst wird die cDNA, an die eine spezifische Erkennungssequenz gekoppelt wurde, auf den Mikroarray hybridisiert. Erst in einem zweiten Hybridisierungsschritt werden die Farbstoffmoleküle an die bereits auf dem Array hybridisierten cDNA-Proben angelagert. Bei den Farbstoffmolekülen (3DNA-Moleküle) handelt es sich um Dendrimer-Komplexe, die aus ca. 850 Fluorophoren (Cy3 bzw. Cy5) pro Molekül bestehen. Jede cDNA wird jeweils von nur einem dieser Farbstoffkomplexe detektiert, die ausreichende Signalintensität wird hier nicht über Amplifikation der Sequenz, sondern durch die Vielzahl der gebundenen Fluorophore erreicht. Diese Methode der Fluoreszenzmarkierung hat den Vorteil, dass die Markierungseffizienz nicht von der Basenzusammensetzung oder der Länge des Transkriptes abhängt. Die einzelnen Schritte der Methode sind nachfolgend dargestellt, soweit nicht anders angegeben wird nach dem Protokoll zum 3DNA Array 900MPX Kit (Genisphere Manual, April 2006) vorgegangen. Alle Waschschriffe finden bei Raumtemperatur statt, sofern nicht anders angegeben.

CDNA SYNTHESE Als Ausgangsmenge für die cDNA-Synthese werden jeweils 1 µg der RNA (Tumor und Referenz) eingesetzt. Als Enzym für die

reverse Transkription der RNA-Proben wird BioScript Reverse Transkriptase verwendet. Als Primer dienen Random(Zufalls)- und Oligo-dT-Primer in Kombination. Dies hat den Vorteil, dass die Reaktion sowohl vom PolyA-Schwanz des 3'-Endes starten kann, als auch durch Binden der Zufalls-Primer innerhalb der Sequenz, wodurch auch teilweise degradierte RNA vervielfältigt werden kann.

CDNA AUFREINIGUNG Alle Aufreinigungsschritte der cDNA erfolgen wie im MinElute PCR Purification Kit Protokoll (Qiagen Handbuch, März 2006) angegeben.

TAILING REAKTION Mit Hilfe der Terminalen Deoxynukleotidyl Transferase (TdT) werden dTTPs an das 3'-Ende der cDNA angefügt (*Tailing*). Die TdT ist eine template-unabhängige Polymerase, für die Reaktion sind keine Primer notwendig.

HERSTELLUNG VON TAGGED CDNA In einem separaten Ligationsschritt wird mit Hilfe eines „Brücken-Oligonukleotids“ eine spezifische Erkennungssequenz an den polyT-Schwanz der cDNA gebunden (*Tagged cDNA*). In diesem Schritt werden jeweils unterschiedliche Erkennungssequenzen an die Tumor- bzw. Referenz-cDNA ligiert, so dass an diese später die unterschiedlichen Farbstoffkomplexe (Cy3 oder Cy5) binden können.

AUFREINIGUNG DER TAGGED CDNA Alle Aufreinigungsschritte der mit der Erkennungssequenz versehenen cDNA erfolgen nach dem MinElute PCR Purification Kit Protokoll (Qiagen Handbuch, März 2006).

CDNA HYBRIDISIERUNG Die Proben für ein Experiment, jeweils aus Referenz- und Tumor-cDNA bestehend, werden vor der Hybridisierung kombiniert und das Ausgangsvolumen von 20 μ l im Vakuumrotationsverdampfer auf max. 10 μ l verringert. Die Hybridisierung erfolgt auf dem vorbereiteten Nexterion Slide E MPX 16 (für die Mikroarray Prozessierung siehe 3.4.3.2). Pro Experiment wird ein Volumen von 30 μ l auf ein Feld des Mikroarrays aufgetragen. Hierfür werden die Proben mit nukleasefreiem Wasser auf 12,5 μ l aufgefüllt, 0,5 μ l Lachs-Sperma DNA und 2 μ l LNA dT Blocker hinzugegeben und alles in 15 μ l Hybridisierungspuffer gelöst. Als Hybridisierungspuffer wird der im 3DNA Array 900MPX Kit enthaltene formamidhaltige Puffer

(Gefäß 7) verwendet, der auch bei relativ niedrigen Temperaturen (hier 45 °C) eine hohe Stringenz erreicht. Unspezifische Bindungen an den Hintergrund der Objektträger werden durch die Zugabe der genomischen Lachs-Sperma DNA reduziert. Nach Inkubation des Hybridisierungsmixes bei ca. 78 °C für 10 Minuten wird er auf den Mikroarray aufgetragen. Die einzelnen Kammern der auf dem Objektträger aufgebracht Silikonstruktur werden zusätzlich mit einem mitgelieferten Klebestreifen abgedichtet. Der Mikroarray wird in eine speziell angefertigte, dicht verschlossene Hybridisierungskammer überführt und bei 45 °C für ca. 16 Stunden (über Nacht) im Wasserbad inkubiert.

WASCHSCHRITTE NACH CDNA HYBRIDISIERUNG Nach der Inkubationszeit wird zunächst der auf der Silikonstruktur klebende Dichtstreifen entfernt und die einzelnen Kammern mit auf 45 °C vorgewärmten 2x SSC (Waschpuffer 0) vorgewaschen. Im ersten Schritt werden 70 µl der Waschlösung in jede Kammer pipetiert. Die Lösung wird abpipetiert und die Kammer sofort wieder mit 100 µl 2xSSC befüllt. Dieser Waschschrift wird für jede Kammer drei mal wiederholt. Anschließend wird die Silikonstruktur entfernt und der Objektträger zügig in einen Waschbehälter mit 250 ml 2x SSC (Waschpuffer 1, 45 °C) überführt und kurz gespült. Für alle weiteren Waschschriffe wird der Objektträger aufrecht in eine Halterung gestellt und während des Waschens in jeweils 250 ml der entsprechenden Lösung geschüttelt und dabei leicht auf und ab bewegt, um Luftblasen zu vermeiden.

- ▷ Waschpuffer 1 (2x SSC, 0.2 %SDS) bei 45 °C, 5 Minuten
- ▷ Waschpuffer 2 (0,2x SSC, 0.1 %SDS), 1 Minute
- ▷ Waschpuffer 3 (0,2x SSC), 1 Minute
- ▷ Waschpuffer 3 (0,2x SSC), 1 Minute
- ▷ Waschpuffer 4 (0,05x SSC) bei 21 °C, 1 Minute

Der letzte Waschschrift ist entscheidend für die Spezifität der Bindungen auf dem Mikroarray, je höher die Waschttemperatur und je geringer die Konzentration der Salzlösung, desto höher ist die Stringenz. Zuletzt wird der Objektträger in einer Multifuge (Heraeus) bei 1.200 U/Min für 3 Minuten trocken zentrifugiert.

3DNA HYBRIDISIERUNG Die Hybridisierung der Fluoreszenzfarbstoffkomplexe an die auf dem Array gebundene cDNA erfolgt in einem Volumen

von 60 μ l. Alle weiteren Schritte bis zur Detektion der Fluoreszenzsignale werden im Dunkeln (ohne direkte Lichteinstrahlung) durchgeführt, um ein Ausbleichen der Farbstoffe zu vermeiden. Die Vorbereitung des Hybridisierungsansatzes erfolgt nach Protokollangaben. Verwendet wird wiederum der formamidhaltige Hybridisierungspuffer (Gefäß 7). Die Hybridisierungslösung wird auf den Objektträger aufgetragen, mit einem LifterSlip abgedeckt und der Mikroarray in die Hybridisierungskammer gelegt. Um ein Austrocknen des Mikroarrays zu verhindern werden jeweils 30 μ l 2x SSC in die dafür vorgesehenen Vertiefungen an den Enden der Hybridisierungskammer pipetiert, bevor die verschlossene Kammer für 4 Stunden bei 50 °C im Wasserbad inkubiert wird.

WASCHSCHITTE NACH 3DNA HYBRIDISIERUNG Zunächst wird der LifterSlip in 2x SSC, 0,2 % SDS bei 65 °C vom Objektträger gespült. Die weiteren Waschschrirte erfolgen, wie bereits für das Waschen nach der cDNA Hybridisierung beschrieben, unter Schütteln in jeweils 250 ml der entsprechenden Lösung.

- ▷ Waschpuffer I (2x SSC, 0,2 % SDS) bei 65 °C, 15 Minuten
- ▷ Waschpuffer II (2x SSC), 10-15 Minuten
- ▷ Waschpuffer III (0,2x SSC), 10-15 Minuten

Vor dem Scannen des Mikroarrays wird dieser in einer Multifuge (Heraeus) bei 1.200 U/Min für 3 Minuten trocken zentrifugiert.

3.4.3.2 *Mikroarray Prozessierung*

Um evtl. fehlende oder schlecht geprintete Spots auf dem Mikroarray zu identifizieren wird der Objektträger vor der Prozessierung leer, d.h. ohne Hybridisierungssonden, gescannt. Damit wird eine Fehlinterpretation dieser Spots vermieden. Ungebundene Nukleinsäuremoleküle und Pufferrückstände werden durch die Mikroarray Prozessierung unmittelbar vor der Hybridisierung entfernt. Wie auch bei den Waschschrirten nach der Hybridisierung wird der Objektträger aufrecht in die mit Spülpuffer gefüllten Container gestellt und während des Waschens geschüttelt, auch hier finden die Waschschrirte, falls nicht anders angegeben, bei Raumtemperatur statt:

- ▷ Spülpuffer 1 (0,1 % Triton X-100), 5 Minuten
- ▷ Spülpuffer 2 (1 mM HCL), 2 Minuten
- ▷ Spülpuffer 2 (1 mM HCL), 2 Minuten

- ▷ Spülpuffer 3 (100 mM KCL), 10 Minuten
- ▷ deionisiertes H_2O , 1 Minute
- ▷ in vorgewärmten 1x Blockingpuffer bei 50 °C liegend im Glascontainer für 15 Minuten schütteln
- ▷ deionisiertes H_2O , 1 Minute

Objektträger bei 1.200 U/Min. für 3 Minuten trocken zentrifugieren (Multifuge, Heraeus).

3.4.3.3 Datengenerierung

Zur Abtastung der Fluoreszenzsignale dient der Tecan Laser Scanner Reloaded. Das empfangene Signal wird als dimensionsloser Quotient in Graustufen gemessen, die Sättigung liegt bei 65.535. Die Auflösung des Scanners beträgt 10 μm /Pixel. Die verwendeten Fluoreszenzfarbstoffe Cy3 und Cy5 haben Anregungswellenlängen von 532 bzw. 635 nm. Mittels automatischer Verstärkungsregelung des *Photo Multipliers (Automatic Gain Control)* werden die Abtastintensitäten für die beiden Fluoreszenzfarbkanäle aufeinander abgestimmt. Hierfür wird zunächst nur ein Feld des Mikroarrays (Feld A3, vgl. Abbildung 3.1) gescannt und die Intensitäten so eingestellt, dass höchstens einzelne Spots im Sättigungsbereich liegen. Anschließend wird der gesamte Mikroarray mit den ermittelten Einstellungen abgetastet und für beide Kanäle jeweils ein eigenes Graustufenbild erstellt.

Das Programm ImaGene 6.0 wird zur Spot-Lokalisierung und Intensitätsbestimmung von Signalen und Hintergrund auf den Mikroarrays verwendet. Hier muss zunächst ein Raster (*Grid*) erstellt und für die verwendeten Mikroarrays angepasst werden, um die Spots zu definieren. Das Raster des Themenarrays besteht aus 24 Spalten mal 27 Zeilen, die Segmentierung in Signal und Hintergrund erfolgt automatisch. Die verwendeten Einstellungen hierfür sind eine Flexibilität von bis zu 5 Pixeln für Abweichungen des zu detektierenden Spots vom angelegten *Grid* und ein Spot-Durchmesser zwischen 10 und 13 Pixeln. Extrem unregelmäßige Spots (z.B. *Donut-Effekt*) sind nicht zugelassen und werden, ebenso wie leere oder negative Spots wenn sie in beiden Farbkanälen auffällig sind, markiert. Funktioniert die automatische Erkennung nicht zuverlässig, werden Leer-Kontrollen nachträglich per Hand markiert.

Die Software ermöglicht es außerdem, die einzelne Gene repräsentierenden Gensequenzen den entsprechenden Spots auf dem Array zuzuordnen,

hierfür muss eine Layout-Datei geladen werden. Anhand der beiden beim Scannen erstellten Graustufenbilder können dann die Intensitäten der Signale bestimmt werden. Angegeben werden jeweils arithmetisches Mittel, Median und Modalwert des Signals für einen Spot, berechnet aus der Gesamtheit der segmentierten Pixel. Sämtliche Rohdaten und Informationen zu den Experimenten (Beschreibung der Proben und Methoden sowie der Scan-Einstellungen) werden in der Datenbank ArrayLIMS (*Laboratory Information Management System*) abgelegt.

3.4.3.4 Vorverarbeitung der Mikroarraydaten

Vor der eigentlichen Datenanalyse (Abschnitt 3.4.5) ist eine Vorverarbeitung der Daten notwendig, um technische Varianzen auszugleichen. Zum einen können, z.B. bedingt durch Oberflächenmerkmale des Mikroarrays, unspezifische Signale entstehen, die durch einen Abgleich der Signale mit den Hintergrundintensitäten korrigiert werden können. Hierfür wird für jeden Spot die Differenz zu einem globalen Hintergrundwert, der jeweils für den gesamten Mikroarray ermittelt wird, berechnet. Diese Methode ist nur dann anzuwenden, wenn keine systematischen Abweichungen, wie z.B. Helligkeitsgradienten, auf dem Array existieren. Zum anderen ist es notwendig Unterschiede zwischen den beiden Farbkanälen zu korrigieren, die z.B. durch Konzentrationsschwankungen der cDNA oder unterschiedliche Eigenschaften der Farbstoffe, wie eine höhere Ozonsensitivität des Farbstoffes Cy5 (Fare et al., 2003), auftreten können. Da es sich hierbei um nicht-lineare Abhängigkeiten handelt wird eine Lowess-Normalisierung (*Locally Weighted Scatterplot Smoothing*, Cleveland und Devlin, 1988) durchgeführt. Hierfür wird eine intensitätsabhängige Regressionsfunktion c zur Korrektur berechnet und das ursprünglich berechnete Verhältnis der beiden Farbkanäle gemäß Formel 3.4 normalisiert.

M_i beschreibt das Verhältnis der Intensitäten der beiden Farbkanäle ($Ch1, Ch2$) eines Spots i (differentielle Expression des entsprechenden Gens)

$$M_i = \log_2 Ch1_i - \log_2 Ch2_i \quad (3.2)$$

A_i gibt die mittlere Intensität eines Spots i an

$$A_i = \frac{1}{2} \times (\log_2 Ch1_i + \log_2 Ch2_i) \quad (3.3)$$

Die Lowess-Regression verschiebt den M-Wert eines Datenpaares ($M_i; A_i$) um die Größe $c(A_i)$. Diese Funktion c wird dabei als lokal-gewichtete lineare Regression nach Cleveland und Devlin (1988) ermittelt. Die Hintergrundkorrektur (bg) erfolgt in einem direkt integrierten Schritt:

$$M_i^{Lowess} = \log_2 (Ch1_i - Ch1_{bg}) - \log_2 (Ch2_i - Ch2_{bg}) - c(A_i) \quad (3.4)$$

Grundvoraussetzung für eine Lowess-Normalisierung ist eine Gleichverteilung von hoch- und runterregulierten Genen sowie keine differentielle Expression für die Mehrheit der analysierten Gene. Für Themenarrays ist dies in der Regel nicht der Fall, allerdings gehen z.B. Wit und MacClure (2004) davon aus, dass die Verteilung der Daten grundsätzlich vernachlässigt werden kann. Ein Vergleich verschiedener Normalisierungsmethoden hat zudem für die hier vorliegenden Daten gezeigt, dass die Lowess-Normalisierung die zuverlässigsten Ergebnisse liefert (siehe Abbildung 3.2). Für eine genauere Darstellung der Normalisierungsmethoden vergleiche Diplomarbeit von Mentz (2006).

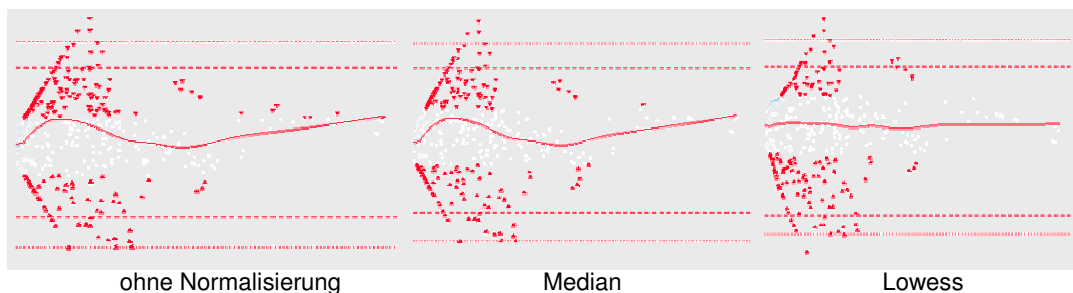


Abbildung 3.2: Ergebnisse verschiedener Normalisierungsmethoden für einen Beispieldatensatz.

Alle Schritte der Datenvorverarbeitung wurden mit der EMMA-Software (Dondrup et al., 2003, 2009) für jedes einzelne Experiment durchgeführt, hierfür wurde jeweils der Mittelwert der Intensitäten herangezogen.

3.4.4 *Quantitative Real-Time-PCR*

Parallel zur Mikroarrayhybridisierung wird für ausgewählte Tumorproben (siehe auch Ergebnisteil 4.5.3) die Expression der Rezeptorgene PGR und ESR1 sowie ERBB2 mittels quantitativer Real-Time-Polymerase-Kettenreaktion (qRT-PCR) bestimmt. Die Methode der qRT-PCR unterscheidet sich von der einfachen Polymerase-Kettenreaktion durch eine parallele

Quantifizierung der vervielfältigten Nukleinsäuren, die z.B. über die Bindung eines Fluoreszenzfarbstoffes an die doppelsträngige DNA ermöglicht wird. Wie bei einer Standard-PCR findet die Vervielfältigung der Sequenz zyklweise in drei Schritten statt: bei der Denaturierung wird zunächst die doppelsträngige DNA aufgeschmolzen, dann erfolgt die spezifische Anlagerung der Primer an die DNA (Annealing) und schließlich die Vervollständigung des DNA-Stranges durch die DNA-Polymerase (Elongation). Die Menge an entstandenen Produkten kann nach jedem Zyklus über die Zunahme der Fluoreszenz gemessen werden.

Da für die vorliegende Untersuchung RNA als Ausgangsmaterial vorliegt muss zunächst eine reverse Transkription der RNA in cDNA erfolgen. Für die Quantifizierung wird das SensiMix One-Step Kit von Quantace verwendet, die gesamte Reaktion findet hier in einem Ansatz statt. Nach der reversen Transkription wird die Hot-Start-Taq-DNA-Polymerase aktiviert und anschließend laufen mehrere Zyklen bestehend aus den drei Schritten Denaturierung, Annealing und Elongation ab. Die Markierung der PCR-Produkte erfolgt durch die Bindung des Fluoreszenzfarbstoffes Sybr[®] Green. Die verwendeten Primerpaare für den Nachweis der Hormonrezeptoren sowie zweier ubiquitär exprimierten Gene zur Normalisierung (siehe Vorverarbeitung der qRT-PCR-Daten) stammen von der Firma Biomol (siehe Tabelle unter 3.3.5). Die Reaktionstemperaturen werden an die Schmelztemperatur der Primer und das Temperaturoptimum der verwendeten DNA-Polymerase angepasst. Die Zusammensetzung des Reaktionsansatzes sowie Zyklenzahl und Ablauf des PCR-Programms sind nachfolgender Auflistung zu entnehmen:

Reaktionsansatz (je 25 μ l):

2x SensiMix One-Step 12,5 μ l
50x SYBR Green Lösung 0,5 μ l
RNase Inhibitor (10 U/ μ l) 0,5 μ l
Primerpaar (je 10 μ M) 1 μ l
RNA Probe (200 ng) mit H₂O 10,5 μ l

qRT-PCR-Programm:

Reverse Transkription 50 °C, 30 min.
Enzymaktivierung 95 °C, 15 min.
40 Zyklen mit:
Denaturierung 95 °C, 20 sek.
Annealing 55 °C, 30 sek.
Elongation 72 °C, 30 sek.

Die Reaktionen finden in einem Opticon Thermocycler statt, die Fluoreszenzmessung erfolgt dabei nach jedem Zyklus.

Zusätzlich wird nach der Amplifikation noch eine Schmelzkurvenanalyse durchgeführt, so können die Fragmentlängen bestimmt und damit spezifische PCR-Produkte von unspezifischen wie z.B. Primer-Dimeren, unterschieden

werden. Hierfür wird die Schmelztemperatur der doppelsträngigen DNA in einem Intervall von 55 °C bis 90 °C durch Freisetzung von Sybr[®] Green und damit einer Abnahme der Fluoreszenz gemessen (1 °C Intervalle, halten der Temperatur für je 10 Sek.).

Nach Beendigung des Programms wird eine Temperatur von 10 °C gehalten.

3.4.4.1 Vorverarbeitung der qRT-PCR-Daten

Die Messung der Real-Time-PCR liefert für jede Reaktion einen *Ct*-Wert (*Cycle Threshold* = Schwellenwert-Zyklus). Dieser Wert beschreibt den Zyklus, in dem die Fluoreszenz erstmals signifikant über den Hintergrund ansteigt (exponentielle Phase der Kettenreaktion). Da es sich um eine log-lineare Funktion handelt (der *Ct* ist umgekehrt proportional zum Logarithmus der Kopienzahl), sollte eine statistische Darstellung auf Basis der Rohdaten vermieden werden (Livak, 2001). Es muss zumindest immer eine lineare Umformung erfolgen, z.B. eine Umformung der Art 2^{-} .

Ebenso wie bei der Mikroarrayanalyse wird hier die Genexpression relativ quantifiziert. Mittels $2^{-\Delta\Delta Ct}$ -Methode wird die n-fache Expression von Experiment (Tumorprobe) zu Referenz (Zelllinien) bestimmt. Bei dieser Methode wird für jede Probe von einer gleichen, optimalen Effizienz (100 %) der Amplifikationsreaktion ausgegangen, also von einer Verdopplung des Produktes in jedem Zyklus (Livak, 2001). In einem ersten Schritt werden Varianzen durch Normalisierung mittels Referenzgenen, z.B. ubiquitär exprimierten Haushaltsgenen (Huggett et al., 2005) ausgeglichen. In dieser Arbeit wurden die Hydroxymethylbilane-Synthase (HMBS; Vandesompele et al., 2002; Szabo et al., 2004) und das Gen des ribosomalen Proteins L7 (RPL7; Hsiao et al., 2001) herangezogen.

Berechnet wird die Differenz der Schwellenwert-Zyklen *Ct* für ein untersuchtes Gen *t* zu einem Haushaltsgen *h* in einer Probe (ΔCt). Dies wird für Experiment und Referenz durchgeführt und so der $\Delta\Delta Ct$ -Wert für ein Gen *t* ermittelt (siehe Formel 3.5).

$$\Delta\Delta Ct_t = (Ct_t - Ct_h)_{Experiment} - (Ct_t - Ct_h)_{Referenz} \quad (3.5)$$

3.4.5 Datenanalyse

3.4.5.1 Differentielle Genexpression

Zur Überprüfung der differentiellen Expression eines Gens aus Tumorgewebe im Vergleich zur Referenz wird das Signifikanzniveau mittels t-Teststatistik bestimmt. Es wird berechnet, wie wahrscheinlich es ist, dass die beobachteten Unterschiede der Expression zufällig auftreten. Hier wird als ein Verfahren zur statistischen Hypothesenprüfung der **Student's t-Test** verwendet, es handelt sich dabei um einen parametrischen Test der eine Normalverteilung der Daten voraussetzt. Es wird geprüft, ob die Mittelwerte zweier Stichproben signifikant verschieden sind, d.h. die Nullhypothese H_0 , dass es kein Unterschied zwischen den Stichproben gibt, abgelehnt werden kann. Der berechnete p-Wert gibt an, mit welcher Irrtumswahrscheinlichkeit die Nullhypothese verworfen werden kann. Je kleiner dieser Wert, desto höher das Konfidenzlevel. In der Regel betrachtet man Konfidenzlevel von 95 % (bzw. 99 %) als (hoch) signifikant. Als Maße für die Streuung der Variablen um ihren (arithmetischen) Mittelwert werden Standardabweichung und Varianz (Quadrat der Standardabweichung) eingesetzt. Mit Hilfe des Konfidenzintervalls kann außerdem auf Basis der Stichprobe abgeschätzt werden, wie gut der angegebene Mittelwert die Daten repräsentiert. Berechnet wird der Wertebereich, in dem sich der tatsächliche Mittelwert mit einer bestimmten Wahrscheinlichkeit (z.B. 95 %) befindet. Der t-Test nach Student ist in der EMMA-Software (Dondrup et al., 2003) implementiert. Dort können die Mikroarray-Genexpressionsdaten direkt ausgewertet und die Ergebnisse exportiert werden.

Betrachtet man die Expression eines Gens über alle Tumorproben kann häufig nicht von einer Normalverteilung der Daten ausgegangen werden. Eine gute Möglichkeit, um die Verteilung der Expressionsdaten zu visualisieren, bieten **Box-Plots** (*Box-and-Whisker-Plots*). Hierfür werden Minimum, unteres Quartil (25 %-Perzentil), Median (50 %), oberes Quartil (75 %) und Maximum der Messwerte berechnet. Die zentrale Box, in der sich die Hälfte aller Messwerte befinden, ergibt sich durch den Abstand der Quartile und beinhaltet den mit einer Linie markierten Median. Die *Whiskers* erstrecken sich von den Grenzen dieser Box aus zum Minimum und Maximum, Extremwerte (Ausreißer) der Messwerte können zusätzlich als separate Punkte dargestellt werden. Ausreißer sind dabei definiert als Messwerte, die außerhalb des anderthalbfachen Interquartilsabstand (Differenz zwischen dem 25 %- und 75 %-Perzentil) vom

oberen oder unteren Quartil entfernt liegen. Dieser Interquartilsabstand kann, vergleichbar zur Standardabweichung als Streuungsmaß normalverteilter Variablen, auch als Maß für die Streuung der Messpunkte berechnet werden. Der Vorteil der Methode liegt in der Robustheit gegenüber Ausreißern. Box-Plots wurden mit der Statistik-Software SPSS erstellt.

3.4.5.2 Identifizierung von Clustern

Eine einfache Form der grafischen Darstellung von Mikroarraydaten ist die *Heat Map*. Die Daten werden als zweidimensionale Karte mit Hilfe von farblich markierten Boxen dargestellt. In der Regel repräsentieren die Spalten dieser Karte Experimente und die Zeilen einzelne Gene. Gene die in einer Probe runterreguliert sind werden hier grün, hoch regulierte Gene rot dargestellt. Die Farbintensität gibt Aufschluss über die Stärke der Expression. Nicht-regulierte Gene erscheinen schwarz. Diese Art der Repräsentation von Expressionsdaten gibt einen ersten visuellen Eindruck der Ergebnisse. Darauf aufbauend ist die Identifizierung von Ähnlichkeiten in Genexpressionsmustern und daraus resultierende Gruppierungen von Genen und Experimenten eine grundlegende Strategie bei der Auswertung von Mikroarraydaten.

Als unüberwachtes Lernverfahren wurde hier das **hierarchische Clustern** eingesetzt. Es beruht darauf, den Abstand (die Ähnlichkeit) zwischen Gruppen von Datenpunkten zu berechnen. Als Ergebnis erhält man ein Dendrogramm (Cluster-Baum), in dem Gene (oder Proben) mit ähnlichen Expressionsmustern räumlich näher beieinander positioniert und verbunden sind. Häufig verwendete Distanz- bzw. Ähnlichkeitsmaße zur Bestimmung des Abstandes zwischen zwei Punkten sind der euklidische Abstand und die Pearson Korrelation. Pearson berücksichtigt ähnliche Trends in den Daten, unabhängig von der Ausprägung der Messwerte. Mit der euklidischen Distanz werden hingegen Daten mit ähnlichen absoluten Werten zusammengefasst. Bei dem am häufigsten eingesetzten agglomerativen Verfahren oder auch *Bottom-up* Ansatz wird zunächst jedes einzelne Objekt als Cluster aufgefasst und die entstehenden neuen Cluster immer weiter zu größeren Gruppen zusammengefasst. Die Zusammenführung der einzelnen beobachteten ähnlichen Gruppen erfolgt hier anhand des Mittelwerts der paarweisen Distanzen (*Average Linkage*) zweier Cluster. Die hier dargestellten Clusterverfahren wurden mit dem Programm Genesis (Sturn et al., 2002) durchgeführt.

3.4.5.3 Korrelationsanalysen

Um die Beziehung zwischen klinischen Parametern und den Ergebnissen der Genexpressionsanalysen näher zu untersuchen wird ihre Korrelation bestimmt. Dabei muss berücksichtigt werden, dass die zu untersuchenden Daten unterschiedliches Skalenniveau haben. Einerseits liegen z.B. aus den Expressionsanalysen intervallskalierte Daten vor, andererseits sind ein großer Teil der klinischen Daten ordinalskaliert. Grundsätzlich ist zu beachten, dass eine beobachtete Korrelation von Variablen nicht direkt kausal interpretiert werden darf. Für Korrelationsanalysen zwischen ordinalskalierten Daten (z.B. klinischen Parametern) wird die **Rangkorrelation nach Spearman** (ρ) verwendet. Sie liefert hier ein robustes Maß, welches auch bei kleineren Stichproben angewandt werden kann. Die Werte für ρ liegen zwischen +1 (positive Korrelation) und -1 (negative Korrelation), ein ρ von 0 zeigt, dass die Daten unkorreliert sind. Auch hier kann die Signifikanz berechnet werden, der p-Wert gibt an, ob sich der beobachtete Wert für ρ signifikant von 0 unterscheidet, d.h. ob eine signifikante Korrelation vorliegt. Äquivalent wird für Vergleiche von Daten mit ungleichen Skalen (z.B. Genexpressionsdaten mit klinischen Parametern) das universellere τ **nach Kendall** berechnet. Zusammenhänge zwischen rein intervallskalierten Daten (z.B. Genexpressionsdaten) können mit dem Korrelationskoeffizienten nach Pearson bestimmt werden. Für die Korrelationsbestimmungen wird SPSS verwendet.

Der **Normalized Tree Index** (NTI) bietet im Vergleich zu den bisher genannten Methoden der Korrelationsanalyse den Vorteil, dass Genexpressionsdaten in ihrer Gesamtheit erfasst werden können. Der NTI wurde an der Technischen Fakultät der Universität Bielefeld von Christian Martin entwickelt (Publikation eingereicht bei *J Biomed Inform*) und ist eine Erweiterung des Tree Index (Martin et al., 2000). Der NTI detektiert Korrelationen zwischen externen Kennzeichen (z.B. klinischen Daten) und bereits geclusterten Mikroarraydaten. Hierfür wird zunächst mittels agglomerativem hierarchischem Clustern ein Cluster-Baum erstellt und dieser in einem zweiten Schritt durch den NTI evaluiert. Für jede klinische Variable wird ihre Korrelation zum Genexpressionsprofil gemessen. Ein NTI von 0 zeigt, dass die Daten unkorreliert sind, ein NTI von 1 bedeutet eine vollständige positive Korrelation, der p-Wert gibt die Signifikanz der beobachteten Abhängigkeit an.

ERGEBNISSE

Der Ergebnisteil dieser Arbeit beinhaltet auf Basis von Literaturrecherchen gewonnene sowie experimentell ermittelte Daten. Zunächst folgt die Erläuterung und Charakterisierung des zugrunde liegenden Materials des hergestellten Brustkrebs-Themenarrays (Abschnitt 4.1 und 4.2) sowie der in die Studie eingeschlossenen Patientinnen und ihrer Tumoren (4.2). Dies dient als Grundlage für die spätere Einordnung der experimentellen Daten. Anschließend wird kurz auf die aus der Aufarbeitung des Tumormaterials resultierende Ergebnisse eingegangen (Abschnitt 4.3), um dann im Hauptteil die Genexpressionsdaten aus Mikroarrayhybridisierung und Polymerase-Kettenreaktion im Kontext auszuwerten (ab 4.4).

4.1 ERSTELLUNG DES THEMENARRAYS

Der in dieser Arbeit verwendete Mikroarray (Brustkrebs-Themenarray) wurde auf Basis einer Literaturrecherche zusammengestellt. Die auf dem zunächst 108 Oligonukleotide umfassenden Mikroarray repräsentierten Gene wurden in publizierten Genexpressionsstudien als brustkrebsrelevant identifiziert oder sind an Stoffwechselwegen beteiligt, die für Tumorentstehung und Tumorwachstum bzw. Metastasierung von Bedeutung sind (siehe hierzu Kapitel 2.4).

Nach einem Pre-Test¹ mit Tumorproben aus OP, der im Rahmen einer Diplomarbeit durchgeführt wurde (Mentz, 2006), wurde der Brustkrebs-Themenarray auf insgesamt 163 Gene und um zusätzliche Kontrollelemente (siehe auch Kapitel 3.2) erweitert. Auf dieses erweiterte Bielefelder Mammakarzinom-Genset wird hier als „McBiSet2“ Bezug genommen.

¹ Dieser Pre-Test diente der Entwicklung und Optimierung der Hybridisierungsmethoden sowie einer ersten Evaluation des Mikroarrays und der Datenvorverarbeitung.

Die detaillierten Ergebnisse der Literaturrecherche und die daraus resultierende Auswahl der Gene werden nachfolgend dargestellt und beispielhaft erläutert. Alle hier aufgeführten Gene sind durch 70mer-Oligonukleotide auf dem Themenarray repräsentiert.

4.1.1 Die 70-Gen-Signatur von van't Veer et al.

Die Grundlage des Themenarrays bildet die 70-Gen-Signatur aus einer Studie von van't Veer et al. (2002). Die Amsterdamer Gruppe identifizierte anhand eines 60mer-Oligonukleotid-Mikroarrays ein Set prognostisch relevanter Markergene zur Unterscheidung von Patientinnen mit schlechter oder guter Prognose anhand des rezidivfreien 5-Jahres-Überlebens.

68 Gene bzw. ESTs (*Expressed Sequence Tags*) dieser Signatur wurden in die Auswahl übernommen. Von diesen 68 Genen sind 19 Klone aus der cDNA-Bibliothek von Ewing und Green (2000). Die entsprechenden Oligonukleotide wurden jeweils in beiden Orientierungen auf den Themenarray gedruckt, da nicht eindeutig war, ob die vorliegenden Sequenzen dem kodierenden Sense-Strang entsprechen. Zwar wurden alle Sequenzen, die in Orientierung des komplementären Anti-Sense-Strangs vorhergesagt wurden, zunächst umgeschrieben, allerdings lagen zu diesem Zeitpunkt nur lückenhafte Sequenzinformationen vor, und die Vorhersage mit der von Ewing und Green (2000, Zusatzmaterial) angewandten Methode ermöglichte nur eine Genauigkeit von ca. 90 %.

Unter den restlichen 49 Genen befinden sich nur wenige Gene, die auch in anderen klinischen Studien auffällig wurden (van't Veer et al., 2002). Funktionell interessant sind hier z.B. Gene der Zellzyklusregulation (Cyclin E2, PRC1, KNTC2) und Zellproliferation (VEGFA, FLT1, ESM1), einzelne Metalloproteinasen (MMP9) sowie weitere in Protein Metabolismus und Transport (PITRM1, EXOC7, RAB6B) involvierte Gene. Für zwölf der von van't Veer und Mitarbeitern identifizierten Gene ist in den gängigen Ontologien keine Funktion bekannt (siehe Abschnitt 4.2.2).

Zwei Gene der 70-Gen-Signatur wurden auf dem Themenarray nicht berücksichtigt. Dabei handelt es sich um das Gen HCA64, da es mit einem bereits vorhandenen Gen (PECI) übereinstimmte, und eine Transkriptvariante von IGFBP5, da es sich um veraltete Sequenzinformation handelte. Eine vollständige Übersicht über alle auf dem Array repräsentierten Gene des van't Veer-Sets ist dem Anhang A.2 zu entnehmen.

4.1.2 Weitere Brustkrebsassoziierte Gene

Das in Abschnitt 4.1.1 beschriebene Grundset wurde für das auf dem Themenarray repräsentierte McBiSet2-Genset um weitere funktionell mit Brustkrebs assoziierte Gene ergänzt (vgl. hierzu auch Kapitel 2.4 der Einleitung). Besonders berücksichtigt wurden dabei Gene des Lipid- und Hormonstoffwechsels (Abschnitt 4.1.2.1), Angiogenese- (4.1.2.3) und Apoptosefaktoren sowie weitere Gene der Zellzyklusregulation (4.1.2.2). Die Genfamilie der ABC-Transporter, deren Mitglieder als Membranproteine aktiv Stoffe aus der Zelle transportieren können und daher bei Chemotherapieresistenzen eine Rolle spielen, sind ebenso durch Vertreter auf dem Array repräsentiert wie Zelladhäsions- und extrazelluläre Matrix-Proteine (Abschnitt 4.1.2.4), die für Tumorwachstum und Metastasierung von Bedeutung sind. Die folgenden Abschnitte geben einen kurzen Überblick über die verwendeten Quellen und nennen explizit die ausgewählten Gene (siehe auch Anhang A.3).

4.1.2.1 Lipid-/Hormonstoffwechsel

Der hormonelle Status hat einen großen Einfluss auf die Brustkrebserkrankung. Steroidhormone spielen bei der Tumorentwicklung und -progression eine wichtige Rolle. Ein negativer Östrogen- und Progesteronrezeptorstatus geht mit einer schlechteren Prognose für den Krankheitsverlauf einher (Osborne, 1998). Die östrogenrezeptorkodierenden Gene ESR1 und ESR2 sowie das für den Progesteronrezeptor kodierende PGR sind auf dem Themenarray vorhanden. Yoshimura et al. (2004) konnten für verschiedene Enzyme der Östradiol-Synthese auf mRNA-Ebene Expressionsunterschiede aufzeigen und einen Zusammenhang zu den krankheitsfreien Überlebensjahren und weiteren klinischen Parametern nachweisen. In weiteren Analysen von Kristensen et al. (2005) wurde ebenfalls die Expression verschiedener Gene des Östradiol-Stoffwechsels und zugehöriger Signalkaskaden untersucht, und eine Klassifizierung erarbeitet, die über die Unterscheidung von östrogenrezeptornegativen und -positiven Tumoren hinaus geht. Zu den auf dem Themenarray repräsentierten Genen aus diesen Studien gehören unter anderem verschiedene Vertreter der Zytochrom-P450-Familie und 17β -Hydroxysteroid-Dehydrogenasen (HSD17B1, HSD17B2). Auch Monooxygenasen, die Reaktionen bei der Synthese von Cholesterol, Steroiden und anderen Lipiden katalysieren, wurden dem Genset hinzugefügt.

Da eine gesteigerte Lipidsynthese mit einem aggressiveren Verlauf der Tumorerkrankung assoziiert wird (Kuhajda, 2000), wurden weiterhin Gene des Lipidstoffwechsels zum McBiSet2-Geneset hinzu genommen. Zudem wiesen Lupu und Menendez (2006) und Menendez et al. (2004) einen Zusammenhang zwischen der Fettsäure-Synthase(FASN)-Aktivität und der genomischen und nicht-genomischen Östrogenwirkung nach und sehen in der Hemmung von FASN einen neuen Ansatzpunkt in der Brustkrebstherapie. Auch das dieses Enzym kodierende Gen ist durch Oligonukleotide auf dem Array repräsentiert. Eine Übersicht über die Gene des Themenarrays, die maßgeblich in Prozesse des Lipid-/Hormonstoffwechsels (und im besonderen des Östrogenstoffwechsels) involviert sind, zeigt Tabelle 4.1.

Tabelle 4.1: Gene der Lipidsynthese und assoziierten Prozessen sowie in den Östrogenstoffwechsel involvierte Gene des Themenarrays.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_54	NM_176895	PPAP2A	phosphatidic acid phosphatase type 2A
R:McBiSet2_79	NM_198837	ACACA	acetyl-Coenzyme A carboxylase alpha
R:McBiSet2_89	NM_004104	FASN	fatty acid synthase
R:McBiSet2_105	NM_006117	PECI	peroxisomal D3,D2-enoyl-CoA isomerase
R:McBiSet2_13	NM_000926	PGR	progesterone receptor
R:McBiSet2_20	NM_000499	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
R:McBiSet2_21	NM_000104	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
R:McBiSet2_23	NM_031226	CYP19A1	cytochrome P450, family 19, subfamily A, polypeptide 1
R:McBiSet2_55	NM_005420	SULT1E1	sulfotransferase family 1E, estrogen-preferring, member 1
R:McBiSet2_97	NM_000413	HSD17B1	hydroxysteroid (17-beta) dehydrogenase 1
R:McBiSet2_98	NM_002153	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2
R:McBiSet2_99	NM_000963	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
R:McBiSet2_124	NM_004176	SREBF1	sterol regulatory element binding transcription factor 1
R:McBiSet2_127	NM_000351	STS	steroid sulfatase (microsomal), arylsulfatase C, isozyme S
R:McBiSet2_145	NM_032644	PPARA	homo sapiens peroxisome proliferative activated receptor, alpha
R:McBiSet2_151	NM_000125	ESR1	estrogen receptor 1
R:McBiSet2_154	NM_007310	COMT	catechol-O-methyltransferase
R:McBiSet2_168	NM_001040	SHBG	sex hormone-binding globulin
R:McBiSet2_184	NM_001437	ESR2	estrogen receptor 2 (ER beta)
R:McBiSet2_187	NM_002456	MUC1	homo sapiens mucin 1, cell surface associated

Im Rahmen der vorliegenden Studie wurde der Östradiol-Stoffwechsel nicht nur auf Transkriptionsebene analysiert. Zusätzlich wurden im Rahmen einer Diplomarbeit Bartneck (2006) relevante Metabolite mittels GC-MS (Gaschromatographie-Massenspektrometrie) untersucht und eine Reihe weiterer hydrophober Metabolite identifiziert (z.B. Glycerol und verschiedene Methylsäureester). Unter Berücksichtigung des Fokus der vorliegenden Arbeit wird auf diese Analysen hier nicht näher eingegangen. Die in diesem

Zusammenhang erweiterte Genauswahl des Brustkrebs-Themenarrays beinhaltet in grundlegende Stoffwechselprozesse (z.B. Citratzyklus und Glykolyse) involvierte Vertreter wie das Gen der Citrat-Synthase, verschiedene Dehydrogenasen, die Transketolase TKT und andere.

4.1.2.2 Zellzyklusregulation - Zellproliferation und programmierter Zelltod

Die Zellzyklusregulation ist in der Tumorzelle in der Regel gestört, und die Balance zwischen Zellproliferation und programmiertem Zelltod nicht mehr gegeben (Meteoglu et al., 2005). Dies kann zu unkontrolliertem Wachstum der Zellen und somit zum Fortschreiten bzw. der Entstehung eines Tumors führen.

Der Zellzyklus in Eukaryonten wird im Wesentlichen durch das Zusammenspiel von Zyklinen (CCN), zyklinabhängigen Kinasen (CDKs) und Inhibitoren (CDKN) reguliert. Sutherland und Musgrove (2004) beschreiben insbesondere die Rolle von D- und E-Zyklinen (wie CCND1, CCNE1 und CCNE2) in Brusttumoren, deren Überexpression gesteigerte Zellproliferation zur Folge hat. Diese und weitere Vertreter des Zyklin/CDK- und Retinoblastoma-Protein(Rb)-Stoffwechselweges, wie das Onkogen CDK4, verschiedene CDK-Inhibitoren (CDKN1A, CDKN1B, CDKN2A) und das Tumorsuppressorgen RB1, wurden ebenfalls dem McBiSet2-Genset hinzugefügt.

Liu et al. (2008) hatten mit ihrer Studie zum Ziel, ein Genexpressionsprofil für die Brustkrebsprognose zu finden, das Gene bereits bekannter und brustkrebsrelevanter biologischer Prozesse einschloss. Sie identifizierten eine Zellzyklus-Gensignatur als prädiktiven Marker, die von ihnen als aussagekräftiger als die 70-Gen-Signatur nach van't Veer et al. (2002) eingestuft wurde. Neben Zyklinen (CCNA und CCNE) gehört hierzu auch MKI-67, das den bekannten Proliferationsmarker Ki-67 (vgl. Kapitel 2.2) codiert sowie verschiedene Wachstumsfaktoren und ihre Rezeptoren (EGFR, IGF1, IGF2, IGFBP4 und IGFBP5). Her2-neu, kodiert durch das Gen ERBB2, gehört zur EGF-Wachstumsfaktorrezeptor-Familie und ist in Brusttumoren häufig überexprimiert. Der Proteinnachweis ist ein etablierter prognostischer und prädiktiver Faktor für Brustkrebs (Cianfrocca und Goldstein, 2004) (vgl. Kapitel 2.2). Alle genannten Gene sind auf dem hier vorgestellten Bielefelder Brustkrebs-Themenarray vertreten.

Die Brustkrebsgene BRCA1 und BRCA2 wurden zunächst in Zusammenhang mit hereditärem Brustkrebs identifiziert. Mutationen dieser Gene sind mit einem erhöhten Risiko für die Brustkrebserkrankung assoziiert

(siehe Kapitel 2.1). Da sie in lebenswichtige Zellfunktionen wie DNA-Reperaturmechanismen und Zellzyklusregulation involviert sind, spielen sie aber auch für die Entstehung des sporadischen Brustkrebs eine große Rolle (Venkitaraman, 2002).

Die Gene der Zellproliferation bzw. allgemein der Zellzyklusregulation des Themenarrays sind den Tabellen 4.2 und 4.3 zu entnehmen. Auf die Mechanismen des programmierten Zelltods (Apoptose) und die in diesem Zusammenhang in den Analysen berücksichtigten Gene wird nachfolgend gesondert eingegangen.

Tabelle 4.2: Gene des Gensets, die in Prozesse der Zellproliferation involviert sind.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_2	NM_002417	MKI67	antigen identified by monoclonal antibody Ki-67
R:McBiSet2_3	NM_002417	MKI67	antigen identified by monoclonal antibody Ki-67
R:McBiSet2_4	NM_002417	MKI67	antigen identified by monoclonal antibody Ki-67
R:McBiSet2_9	NM_000612	IGF2	insulin-like growth factor 2 (somatomedin A)
R:McBiSet2_27	NM_005247	FGF3	fibroblast growth factor 3 (murine mammary tumor virus integration site (v-int-2) oncogene homolog)
R:McBiSet2_28	NM_000075	CDK4	cyclin-dependent kinase 4
R:McBiSet2_38	NM_000618	IGF1	insulin-like growth factor 1 (somatomedin C)
R:McBiSet2_61	NM_002019	FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
R:McBiSet2_107	NM_000599	IGFBP5	insulin-like growth factor binding protein 5
R:McBiSet2_110	NM_000077	CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
R:McBiSet2_132	NM_001005915	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
R:McBiSet2_149	NM_001552	IGFBP4	insulin-like growth factor binding protein 4

APOPTOSE Für den programmierten Zelltod gibt es zwei grundlegende Mechanismen. Der intrinsische Signalweg geht mit einer strukturellen und funktionalen Veränderung der Mitochondrien einher. Mitglieder der BCL2-Genfamilie sind hier die wichtigsten Anti-Apoptose-Regulatoren, die für ein Einleiten der Apoptose ausgeschaltet sein müssen. Survivin (BIRC5) ist ein weiterer Inhibitor der Apoptose. Pro-Apoptose-Faktoren sind z.B. das BCL2-assoziierte Protein BAX das als entsprechender Gegenspieler fungiert (Schulz, 2005). Dieser intrinsische Signalweg wird häufig durch die Expression des Transkriptionsfaktors TP53 in Folge von DNA-Schäden oder anderen die Zelle schädigenden Ereignissen induziert. Bekannte Zielgene dieses Transkriptionsfaktors sind u.a. das bereits genannte BAX sowie PUMA (BBC3), die ihrerseits entsprechend den programmierten Zelltod einleiten. Auch das

Tabelle 4.3: Gene des Gensets, die in allgemeine Prozesse der Zellzyklusregulation involviert sind.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_7	NM_000321	RB1	retinoblastoma 1 (including osteosarcoma)
R:McBiSet2_29	NM_001237	CCNA2	cyclin A2
R:McBiSet2_30	NM_057182	CCNE1	cyclin E1
R:McBiSet2_31	NM_057749	CCNE2	cyclin E2
R:McBiSet2_33	NM_004064	CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
R:McBiSet2_44	NM_145888	KLK10	kallikrein-related peptidase 10
R:McBiSet2_76	NM_198433	AURKA	aurora kinase A
R:McBiSet2_85	NM_201283	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
R:McBiSet2_86	NM_014781	RB1CC1	RB1-inducible coiled-coil 1
R:McBiSet2_109	NM_000127	EXT1	exostoses (multiple) 1
R:McBiSet2_131	NM_004448	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
R:McBiSet2_148	NM_000598	IGFBP3	insulin-like growth factor binding protein 3
R:McBiSet2_153	NM_007294	BRCA1	breast cancer 1, early onset
R:McBiSet2_175	NM_053056	CCND1	cyclin D1
R:McBiSet2_179	NM_000059	BRCA2	breast cancer 2, early onset
R:McBiSet2_183	NM_018401	STK32B	serine/threonine kinase 32B

Onkogen MYC kann Apoptose-Mechanismen beeinflussen und dabei ebenfalls, z.B. durch Aktivierung der TP53-Expression, die Apoptose herbeiführen (Lowe et al., 2004).

Der zweite Weg wird über extrinsische Signale induziert. TNF-Rezeptoren an der Zelloberfläche und der Hauptbindungsfaktor TNF α (*Tumour Necrosis Factor*), der u.a. bei zellulären Stress-Antworten ausgeschüttet wird, sind hier der Startpunkt der Signalkaskade (Schulz, 2005).

In beiden Fällen endet die Signalübertragung mit der Aktivierung von Kaspasen (Proteasen) und führt schließlich zum Absterben der Zelle. Die hier genannten Gen-Vertreter der beiden Signalwege sowie einige weitere apoptoseassoziierte Gene sind der McBiSet2-Genauswahl hinzugefügt worden (Tabelle 4.4).

4.1.2.3 Angiogenese

Die Bildung neuer Blutgefäße ist für die Versorgung des Tumors essentiell. Angiogenesefaktoren spielen daher für das Tumorwachstum eine große Rolle (Kerbel, 2008). Einige Faktoren wie z.B. Angiopoietin, VEGFA und VEGFB (*Vascular Endothelial Growth Factors*) sowie PDGF (*Platelet-derived Growth Factors*) sind auf dem Brustkrebs-Themenarray vertreten (siehe auch Tabelle 4.5). Die Hemmung der Angiogenese durch eine Anti-VEGFA-Therapie ist ein

Tabelle 4.4: Auswahl der mit Apoptose-Mechanismen assoziierten Gene des Themenarrays.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_14	NM_005235	ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
R:McBiSet2_32	NM_078467	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
R:McBiSet2_49	NM_014417	BBC3	BCL2 binding component 3
R:McBiSet2_51	NM_000594	TNF	tumor necrosis factor (TNF superfamily, member 2)
R:McBiSet2_69	NM_004324	BAX	BCL2-associated X protein
R:McBiSet2_70	NM_003998	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (p105)
R:McBiSet2_75	NM_000582	SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)
R:McBiSet2_116	NM_033512	TSPYL5	TSPY-like 5
R:McBiSet2_143	NM_001012270	BIRC5	baculoviral IAP repeat-containing 5 (survivin)
R:McBiSet2_146	NM_005163	AKT1	v-akt murine thymoma viral oncogene homolog 1
R:McBiSet2_162	NM_002467	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
R:McBiSet2_163	NM_000633	BCL2	B-cell CLL/lymphoma 2
R:McBiSet2_166	NM_000314	PTEN	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)
R:McBiSet2_173	NM_000362	TIMP3	TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)
R:McBiSet2_178	NM_000546	TP53	tumor protein p53 (Li-Fraumeni syndrome)

viel versprechender neuer Ansatzpunkt der Brustkrebstherapie (Ferrara und Kerbel, 2005).

4.1.2.4 Zelladhäsion/extrazelluläre Matrix

Eine entscheidende Rolle für Tumorwachstum und Metastasierung spielen Zelleigenschaften wie die Fähigkeit zu Invasion und Migration. Auf molekularer Ebene sind hierfür Mechanismen der Zell-Zell-/Zell-Matrix-Interaktion und Degradierungsprozesse der extrazellulären Matrix von Bedeutung (Cairns et al., 2003) (vgl. Kapitel 2.4). Die als Prognosefaktor für die Brustkrebserkrankung diskutierte Protease uPA (PLAU-Gen) und ihr Inhibitor PAI-1 (codiert durch SERPINE1) sind z.B. in solche Degradierungsprozesse involviert und auf dem Themenarray repräsentiert. Weitere wichtige Proteinfamilien sind Integrine, Cadherine und Metalloproteinasen. Integrine vermitteln Zell-Matrix-Interaktionen und beeinflussen die Angiogenese. Cadherine sind in Zell-Zell-Kontakte involviert. Aggressive bzw. invasive Tumoren weisen häufig eine reduzierte oder fehlende E-Cadherin(CDH1)-Expression auf (Cairns et al., 2003). CDH1 ebenso wie einige andere Gene dieser Familien wurden daher ebenfalls in das Genset aufgenommen (Tabelle 4.6).

Tabelle 4.5: Gene der Angiogenese, die im McBiSet2-Genset enthalten sind.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_15	NM_002253	KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
R:McBiSet2_26	NM_033016	PDGFB	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)
R:McBiSet2_40	NM_139013	MAPK14	mitogen-activated protein kinase 14
R:McBiSet2_41	NM_139049	MAPK8	mitogen-activated protein kinase 8
R:McBiSet2_42	NM_001146	ANGPT1	angiopoietin 1
R:McBiSet2_61	NM_002019	FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
R:McBiSet2_81	NM_003377	VEGFB	vascular endothelial growth factor B
R:McBiSet2_88	NM_005238	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
R:McBiSet2_102	NM_001147	ANGPT2	angiopoietin 2
R:McBiSet2_113	NM_000603	NOS3	nitric oxide synthase 3 (endothelial cell)
R:McBiSet2_136	NM_006755	TALDO1	transaldolase 1
R:McBiSet2_144	NM_006206	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide
R:McBiSet2_157	NM_002609	PDGFRB	platelet-derived growth factor receptor, beta polypeptide
R:McBiSet2_160	NM_003376	VEGFA	vascular endothelial growth factor A
R:McBiSet2_176	NM_033023	PDGFA	platelet-derived growth factor alpha polypeptide

Tabelle 4.6: Zelladhäsion vermittelnde Gen-Vertreter und mit der extrazellulären Matrix assoziierte Gene des Themenarrays.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_6	NM_000602	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1
R:McBiSet2_22	NM_000425	L1CAM	L1 cell adhesion molecule
R:McBiSet2_25	NM_004360	CDH1	cadherin 1, type 1, E-cadherin (epithelial)
R:McBiSet2_34	NM_001846	COL4A2	collagen, type IV, alpha 2
R:McBiSet2_35	NM_003882	WISP1	WNT1 inducible signaling pathway protein 1
R:McBiSet2_36	NM_033666	ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
R:McBiSet2_52	NM_007115	TNFAIP6	tumor necrosis factor, alpha-induced protein 6
R:McBiSet2_125	NM_002658	PLAU	plasminogen activator, urokinase
R:McBiSet2_133	NM_018695	ERBB2IP	erb2 interacting protein
R:McBiSet2_155	NM_015640	SERPBI	SERPINE1 mRNA binding protein 1
R:McBiSet2_159	NM_002295	RPSA	ribosomal protein SA

MATRIX-METALLOPROTEINASEN UND ADAMS Matrix-Metalloproteinasen (MMPs) degradieren verschiedene extrazelluläre Matrix-Komponenten und sind an Prozessen der Gewebeerneuerung wie Wundheilung, Angiogenese und Tumorerkrankung beteiligt. Die Bedeutung von Matrix-Metalloproteinasen als unabhängige Prognosefaktoren für die Brustkrebskrankung ist noch nicht gesichert. Genexpressionsstudien lassen aber vermuten, dass die Überexpression verschiedener MMPs mit einer schlechteren Prognose einhergeht (Têtu et al., 2006; Jinga et al., 2006).

ADAMs (*a Disintegrin and Metalloproteinase*) sind Zellmembran-Proteine, die ebenfalls eine Metalloproteinase-Domäne besitzen und sowohl Adhäsions- (Bindung von Integrinen) als auch Proteasewirkung aufweisen. Lendeckel et al. (2005) konnten für ADAM 9, 12 und 17 eine erhöhte mRNA-Expression in Brusttumoren nachweisen. Auch O’Shea et al. (2003) untersuchten die Expression von ADAM 9 auf Protein und Genebene und konnten einen Zusammenhang zu gesicherten Prognosefaktoren wie Nodalstatus und Her2-neu-Expression herstellen. Daher wurden diese Gene und weitere ausgewählte Matrix-Metalloproteinase-Gene dem Brustkrebs-Themenarray hinzugefügt (siehe Tabelle 4.7).

Tabelle 4.7: Metalloproteinasen (Matrix-Metalloproteinasen MMP, Disintegrin und Metalloproteinasen ADAMs) des McBiSet2-Gensets.

Reporter Name	NCBI Accession	Gen Symbol	Gen Name
R:McBiSet2_17	NM_004995	MMP14	matrix metallopeptidase 14 (membrane-inserted)
R:McBiSet2_18	NM_002422	MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
R:McBiSet2_129	NM_003816	ADAM9	ADAM metallopeptidase domain 9 (meltrin gamma)
R:McBiSet2_137	NM_005940	MMP11	matrix metallopeptidase 11 (stromelysin 3)
R:McBiSet2_164	NM_003474	ADAM12	ADAM metallopeptidase domain 12 (meltrin alpha)
R:McBiSet2_165	NM_003183	ADAM17	ADAM metallopeptidase domain 17 (tumor necrosis factor, alpha, converting enzyme)
R:McBiSet2_167	NM_002422	MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
R:McBiSet2_169	NM_003254	TIMP1	TIMP metallopeptidase inhibitor 1
R:McBiSet2_171	NM_004994	MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)
R:McBiSet2_172	NM_004530	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
R:McBiSet2_173	NM_000362	TIMP3	TIMP metallopeptidase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)

Darüber hinaus spielen auch Zell-Umwelt-Interaktionen eine wichtige Rolle. Hier sind eine Reihe risikomodulierender Gene bekannt, die z.B. auf die Immunantwort oder den Hormonspiegel wirken bzw. direkte Transportfunktion haben (Schulz, 2005). So haben z.B. Glutathione-S-transferasen (GSTM1 und GSTM3), deren Expression hier untersucht wurde, einen Einfluss auf die Sensitivität der Zelle gegenüber Karzinogenen. ABC-Transporter wie ABCB1, ABCC1 und ABCG2 sind für die Vermittlung von Chemotherapieresistenzen von Bedeutung und ebenfalls auf dem Themenarray vorhanden.

Viele der in diesem Kapitel genannten Gene sind in mehrere der dargestellten Funktionsklassen involviert. Eine klare Trennung der oft ineinander greifenden oder voneinander abhängigen Stoffwechselwege und Signalkaskaden ist nicht möglich. Die hier aufgegriffenen Funktionen geben daher nur

einen Überblick über die wichtigsten Prozesse (siehe auch Kapitel 2.4) und benennen einige Vertreter der beteiligten Gene bzw. Genprodukte.

4.2 DATENGRUNDLAGE DER ANALYSEN

In diesem Abschnitt werden alle für die weiteren Analysen genutzten Daten kurz beschrieben. Hierzu gehören sowohl patientinnen- und tumorbezogene Daten als auch Zusatzinformationen zu den analysierten Genen sowie eine Übersicht über die durchgeführten Analysen.

Genexpressionsanalysen wurden an insgesamt 94 Tumorproben von 56 Brustkrebspatientinnen vorgenommen. Je nach vorhandener RNA-Menge (vgl. hierzu auch Abschnitt 4.3) wurden mittels Mikroarrayhybridisierung ein bis drei Replikate pro Probe analysiert. Insgesamt handelte es sich um 166 Experimente auf 15 Mikroarrays (Übersicht über die Experimente in Anhang D). Zusätzlich zu den Mikroarrayexperimenten konnte an 31 Tumorproben von 26 der 56 Brustkrebspatientinnen die Expression ausgewählter Gene mittels quantitativer Real-Time-PCR (qRT-PCR) gemessen werden.

Nicht alle durchgeführten Experimente konnten in die Auswertung einbezogen werden, die Analysen der Tumorproben von 16 der 56 Patientinnen konnten nur eingeschränkt ausgewertet werden. Sieben der eingeschlossenen Patientinnen wurden nachträglich aus der Studie genommen, da sie entweder nicht für eine neoadjuvante Chemotherapie geeignet waren oder aus persönlichen Gründen die weitere Studienteilnahme ablehnten. Neun weitere Patientinnen erfüllten zwar alle Kriterien für die Anwendung einer neoadjuvanten Therapie und hatten ihr Einverständnis erklärt, wurden aber aus organisatorischen Gründen nicht initial in die Studie eingeschlossen. Auch wenn die Tumorproben dieser Patientinnen im Rahmen der Genexpressionsanalysen berücksichtigt wurden, sind nicht alle begleitenden Daten vollständig verfügbar. Zusätzlich kam es bei einzelnen Experimenten zu technischen Ausfällen, so dass für die Analysen letztlich 77 Tumorproben von 40 Studienpatientinnen zur Verfügung stehen:

- ▷ 8 Patientinnen mit vor und nach neoadjuvanter Chemotherapie entnommenen Tumorstanzen sowie OP-Präparat
- ▷ 6 Patientinnen mit Primär-Stanze (vor Therapiebeginn) und OP-Präparat
- ▷ 5 Patientinnen mit Primär-Stanze und einer Stanze nach Therapiebeginn

- ▷ 3 Patientinnen mit ausschließlich nach Beginn der Therapie entnommenem Tumormaterial (Stanze und OP)

Bei den übrigen Tumorproben handelt es sich jeweils um Einzelproben der Patientinnen. Eine Übersicht über alle Proben ist dem Anhang C zu entnehmen.

Im folgenden Abschnitt werden die wichtigsten vorliegenden klinischen Parameter für die 40 in die Analysen eingegangenen Studienpatientinnen zusammengefasst (siehe auch Anhang B). Dabei handelt es sich in erster Linie um die etablierten Prognosefaktoren. Zudem wurden grundlegende patientinnenbezogene Daten erfasst und solche, die mit Bezug zu potentiellen Risikofaktoren für die Brustkrebserkrankung von Interesse sind (vgl. Kapitel 2.1).

4.2.1 Charakterisierung des Patientinnenkollektivs

RISIKOTYPISIERUNG DER TUMOREN Bei den Tumoren des Patientinnenkollektivs handelt es sich fast ausschließlich um invasive duktale Karzinome. Fernmetastasen liegen bei keiner der Patientinnen vor. Gut die Hälfte der Patientinnen hat jedoch befallene Lymphknoten. Alle Tumoren weisen undifferenzierte bis mäßig differenzierte Zellen auf (Grade 2 und 3) und sind mehrheitlich größer als 2 cm (mindestens T2). Eine Überexpression oder Amplifikation des Her2-neu-Rezeptors konnte für etwa ein Drittel der Tumoren festgestellt werden. Alle Patientinnen der Studie können damit nach den St. Gallen Kriterien von 2005 (Goldhirsch et al., 2005) der **hohen oder mittleren Risikogruppe** zugeordnet werden (siehe Kapitel 2.2, Tabelle 2.2).

Bezüglich des Rezeptorstatus der Patientinnen weisen sieben Patientinnen einen *Triple-negativ*-Tumortyp (keine Expression von Östrogen-, Progesteron- und Her2-neu-Rezeptoren) auf. Dieser geht mit einer höheren Aggressivität der Erkrankung und damit mit einer schlechteren Prognose einher (Reis-Filho und Tutt, 2008).

KLINISCHE ANAMNESE UND THERAPIE Sieben Patientinnen haben eine familiäre Vorbelastung und eine Patientin hatte bereits eine vorangegangene Krebserkrankung.

Alle Patientinnen erhielten eine taxanhaltige **neoadjuvante Chemotherapie**. Den meisten (25 von 40) wurde dabei das AC-DOC- und sieben Patientinnen das TAC-Therapieschema verabreicht (zu den Therapieschemata siehe

Kapitel 2.2). Die weiteren acht Patientinnen erhielten verschiedene andere Therapieschemata.

PERSONENBEZOGENE DATEN Die Patientinnen stammen bis auf zwei Ausnahmen aus Europa und waren zum Zeitpunkt der Diagnose zwischen 29 und 74 Jahre alt. Das Durchschnittsalter der Patientinnen liegt bei 50 Jahren und damit 13 Jahre unter dem mittleren Erkrankungsalter für Brustkrebs bei Frauen in Deutschland (RKI, 2008).

HORMONABHÄNGIGE UND REPRODUKTIVE FAKTOREN Bei 16 Patientinnen (40 %) ist die Erkrankung postmenopausal aufgetreten. Fünf der insgesamt 32 Patientinnen mit ausgetragenen Schwangerschaften waren bei der Geburt über 30 Jahre alt. Keine der Patientinnen weist eine frühe Menarche (vor dem elften Lebensjahr) und nur eine Patientin eine späte Menopause (nach dem 55. Lebensjahr) auf.

LEBENSSTILFAKTOREN Knapp die Hälfte (18 von 40) der Patientinnen raucht bzw. hat geraucht. Regelmäßigen Alkoholkonsum (mehrmals pro Woche) geben neun der Patientinnen an.

Ein Vergleich des *Body-Mass-Index* (BMI) der Patientinnen mit der Altersverteilung in Deutschland (Datenquelle <http://www.gbe-bund.de>; BMI nach Altersverteilung für Frauen in Deutschland 2005) zeigt, dass 25 von 38 Patientinnen (66 %), für die der BMI bekannt ist, über dem Durchschnitts-BMI ihrer Altersklasse liegen (Abbildung 4.1). Im Mittel liegt der BMI dieser Studienpatientinnen drei Index-Punkte über dem Durchschnittswert ihrer jeweiligen Altersklasse. Nach WHO-Kriterien (siehe Kapitel 3.1.2) sind 13 Patientinnen übergewichtig (BMI>25) und weitere 12 Patientinnen stark übergewichtig (BMI>30). Übergewicht ist vor allem bei Frauen in der Menopause ein Risikofaktor. Dies trifft auf sieben der 25 Patientinnen mit einem erhöhten BMI zu.

Ein Drittel der Patientinnen gibt an, regelmäßig Sport zu treiben (Zwei Stunden pro Woche und mehr). Keine der Patientinnen schläft weniger als sechs Stunden pro Nacht.

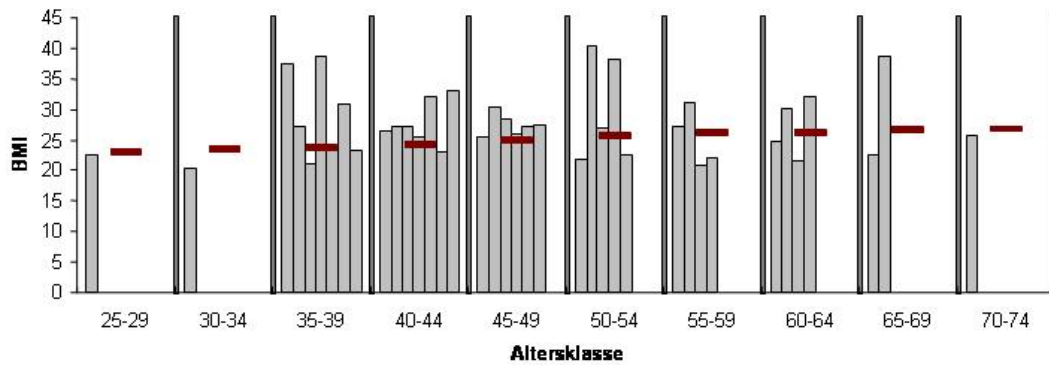


Abbildung 4.1: Dargestellt ist der *Body-Mass-Index* (BMI) der Studienpatientinnen nach Altersklassen. Der rote Balken markiert den durchschnittlichen BMI für Frauen der jeweiligen Altersklasse in Deutschland.

4.2.2 Gewonnene Expressionsdaten und ihre Einordnung

Expressionsanalysen wurden in erster Linie mittels Mikroarraytechnik durchgeführt. Eine Überprüfung der Ergebnisse erfolgte für ausgewählte Gene und Tumorproben zusätzlich durch quantitative Real-Time-PCR (qRT-PCR). Die Auswertung der Expressionsdaten erfolgt jeweils anhand der normalisierten Daten (Beschreibung der Vorverarbeitungsschritte unter 3.4.3.4 und 3.4.4.1). Um die Expressionsausprägung für ein Gen zu bestimmen werden die *M*-Werte (Kapitel 3.4.3.4) aus den Mikroarrayexperimenten für jedes Gen über alle Replikate gemittelt, die Standardabweichung berechnet und mittels t-Test (siehe 3.4.5.1) bestimmt, ob ein Gen in einer Tumorprobe im Vergleich zur Referenz signifikant reguliert ist (ein *p*-Wert von $< 0,05$ entspricht einem Signifikanzlevel von 95 %). Die Mittelwerte aus dem Signifikanztest sind die Grundlage weiterer Analysen.

Die Basis der qRT-PCR-Auswertung bilden die Berechnungen aus der $\Delta\Delta Ct$ -Methode (Kapitel 3.4.4.1), wobei jeweils über zwei technische Replikate gemittelt wurde.

Für eine funktionelle Einordnung der analysierten Gene wurden neben Literaturrecherchen (Abschnitt 4.1) auch Datenbankabfragen auf zwei Ontologien (PANTHER-Index und GeneOntology-Datenset, vgl. Exkurs in Kapitel 2.4 und Methodik in 3.2.2) vorgenommen. Diese Zuordnung zu Funktionsklassen dient u.a. als Grundlage für die an die Expressionsanalysen anschließende Auswertung. Aus Clusteranalysen resultierende Gen-Cluster können so auch auf funktionaler Ebene verglichen und eingeordnet werden. Abbildung 4.2 zeigt die Zuordnung der Gene des McBiSet2-Gensets zu biologischen Prozes-

sen und molekularen Funktionen und ihre Verteilung im Vergleich zu allen in der GO-Datenbank vorhandenen Einträgen Die Ergebnisse der Datenbankabfragen bzgl. zugrunde liegender biologischer Prozesse (GeneOntology und PANTHER) können im Anhang A.3 eingesehen werden.

4.3 RNA AUS TUMORGEWEBE - SCHWANKUNGEN DER QUALITÄT UND QUANTITÄT

Die Isolierung der Gesamt-RNA aus Tumorgewebe erfolgte wie in Kapitel 3.4.1 beschrieben. Zur Durchführung der Mikroarrayhybridisierungen wird ein Minimum von 1 μg Gesamt-RNA (Konz. 200 ng/ μl) benötigt. Die Konzentration sämtlicher Isolate wurde daher zunächst mit dem NanoDrop Spektrophotometer bestimmt. Für alle für die Analysen geeigneten Proben wurde zusätzliche eine Quantitäts- und Qualitätsbestimmung mittels Agilent Bioanalyser durchgeführt. Für einen Großteil der bearbeiteten Proben liegt außerdem das Probengewicht vor. Die Ergebnisse der einzelnen Messungen sind dem Anhang C zu entnehmen.

58 (von 155 bearbeiteten) Tumorproben konnten aufgrund zu geringer RNA-Konzentrationen und unzureichender Qualität der RNA nicht analysiert werden. RNA aus weiteren 27 Tumorproben stand nur für eine Hybridisierungsreaktion (ohne Replikat) zur Verfügung. 77 der analysierten Tumorproben stammen dabei von Studienpatientinnen, die weiteren Proben stammen von Patientinnen, die nicht in die Studie eingeschlossen wurden (siehe auch Abschnitt 4.2).

Um Ursachen für diese hohe Ausfallquote von 37 % zu finden wurde zunächst die Abhängigkeit der RNA-Ausbeute vom Gesamtgewicht der Tumorprobe untersucht. Anhand der Messergebnisse konnte eine Korrelation ausgeschlossen werden. Die Menge der gewonnenen RNA (Konzentrationsbestimmung mittels NanoDrop) steigt nicht mit dem Probengewicht an (Abbildung 4.3).

Ein weiterer möglicher Einflussfaktor für die RNA-Gewinnung ist die Tumorbehandlung. Chemotherapeutika wirken auf die Zellaktivität und somit auch auf die Transkription in der Zelle (vgl. Kapitel 2.2). Die zu überprüfende Hypothese war hier, ob grundsätzlich aus Probenmaterial nach einer Therapie mit Chemotherapeutika weniger RNA gewonnen werden kann. Wie Abbildung 4.4 zu entnehmen ist konnte statistisch (t-Test) kein direkter Zusammenhang zwischen Behandlung des Tumors und der RNA-Ausbeute

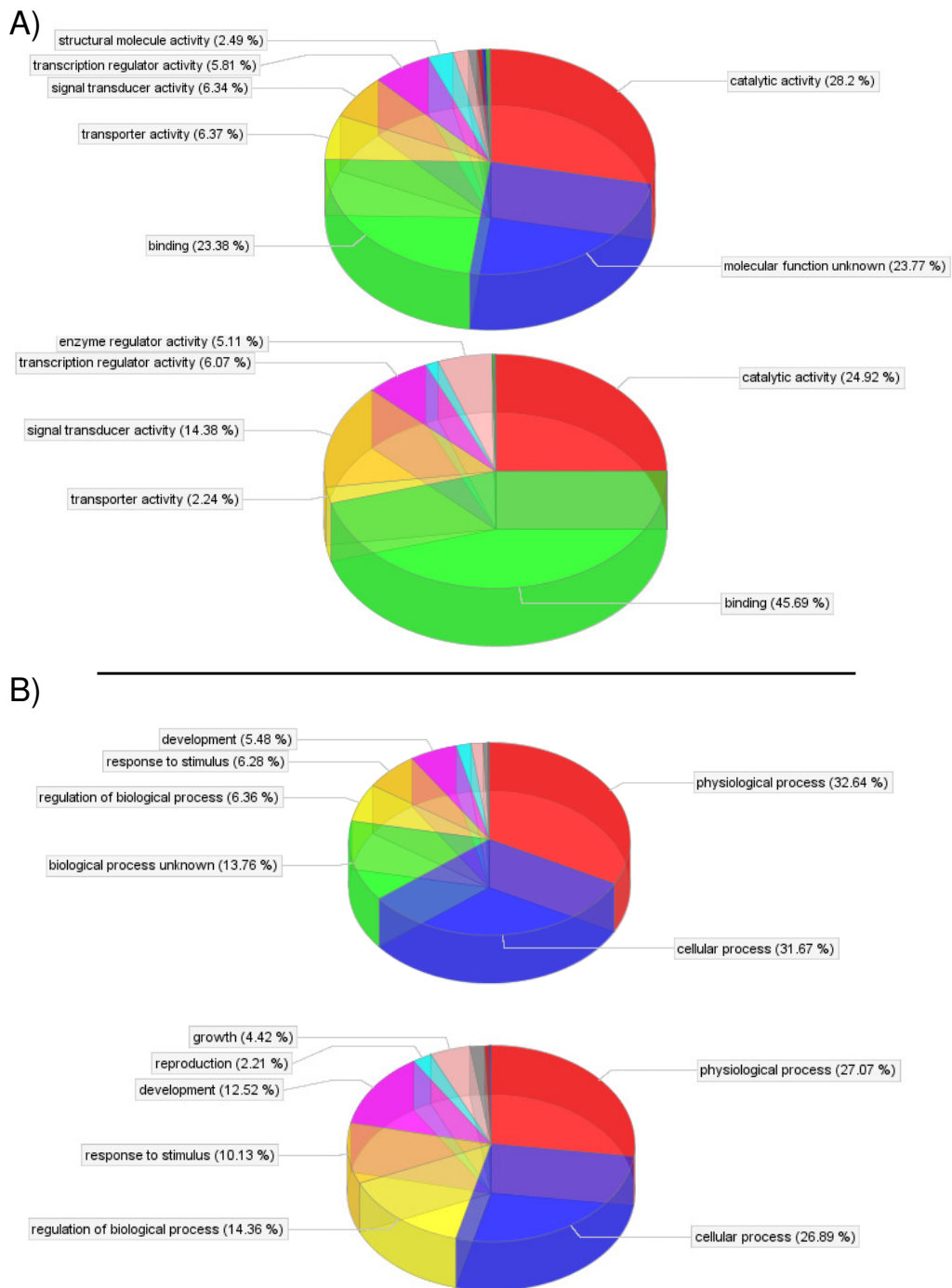


Abbildung 4.2: Übersicht über alle verfügbaren GeneOntology(GO)-Prozesse und -Funktionen der Gene des Themenarrays. Dargestellt ist jeweils zuerst die gesamte Verteilung aller Einträge der GO-Datenbank und im Vergleich darunter die Zuordnung für das ausgewählte Geneset. Abbildung A zeigt die molekularen Funktionen und Abbildung B die übergeordneten biologischen Prozesse.

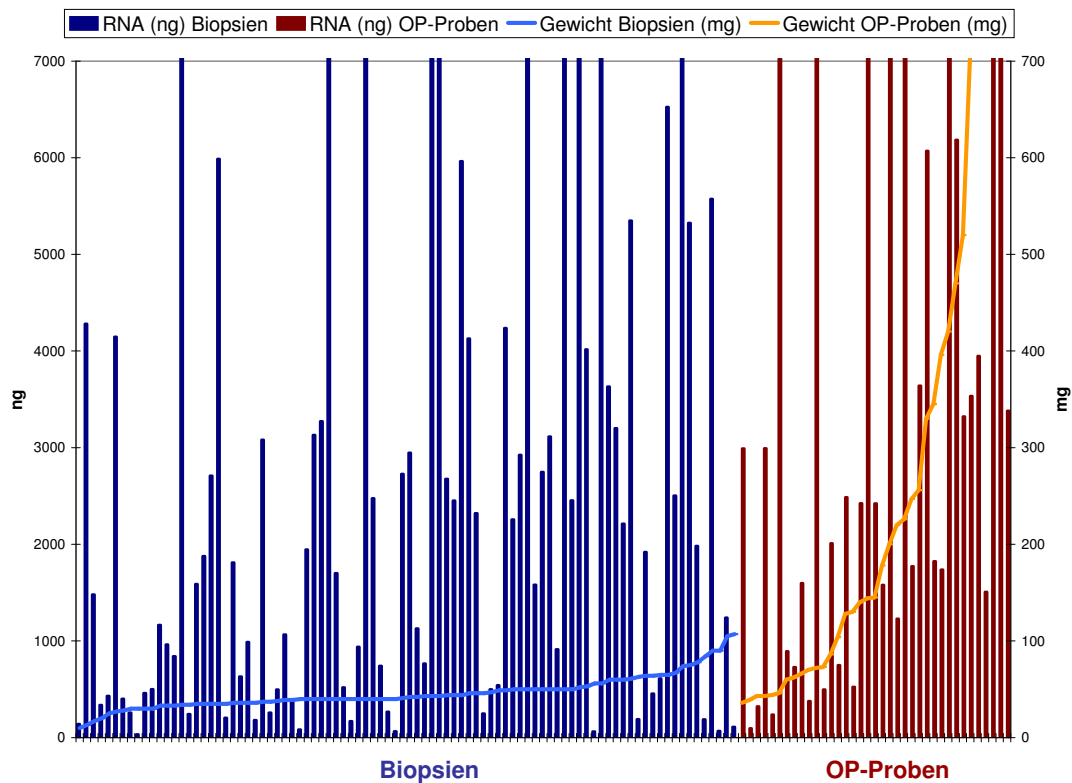


Abbildung 4.3: Gewonnene RNA in ng aus Biopsien und OP-Proben jeweils im Vergleich zum Ausgangsgewicht der Gewebeproben in mg.

festgestellt werden. Signifikante Unterschiede ($p < 0,05$) in der RNA-Ausbeute zeigen sich lediglich beim Vergleich von vor Chemotherapie entnommenen Biopsien zu OP-Proben ($p = 0,003$). Insgesamt konnte aus Biopsien mehr RNA isoliert werden als aus OP-Präparaten (Abbildung 4.4).

Zusammengefasst konnte für die hier untersuchten Faktoren kein eindeutiger Zusammenhang mit der Qualität und Quantität der gewonnenen RNA festgestellt werden.

4.4 CHARAKTERISIERUNG VON BRUSTTUMOREN ANHAND VON GEN-EXPRESSIONSANALYSEN

Auf Basis der gewonnenen Expressionsdaten sollen in den folgenden Abschnitten Gene bzw. **Gengruppen** identifiziert werden, die in dem analysierten Mammakarzinomgewebe auffällig sind.

Grundlage für die Visualisierung der Genexpression ist zunächst die Darstellung als Heat Map (vgl. Kapitel 3.4.5.2). Die Expression der analysierten Gene wird innerhalb ihrer in Abschnitt 4.1.2 definierten Funktionsklassen dargestellt. Die differentielle Expression der Gene über alle Experimente

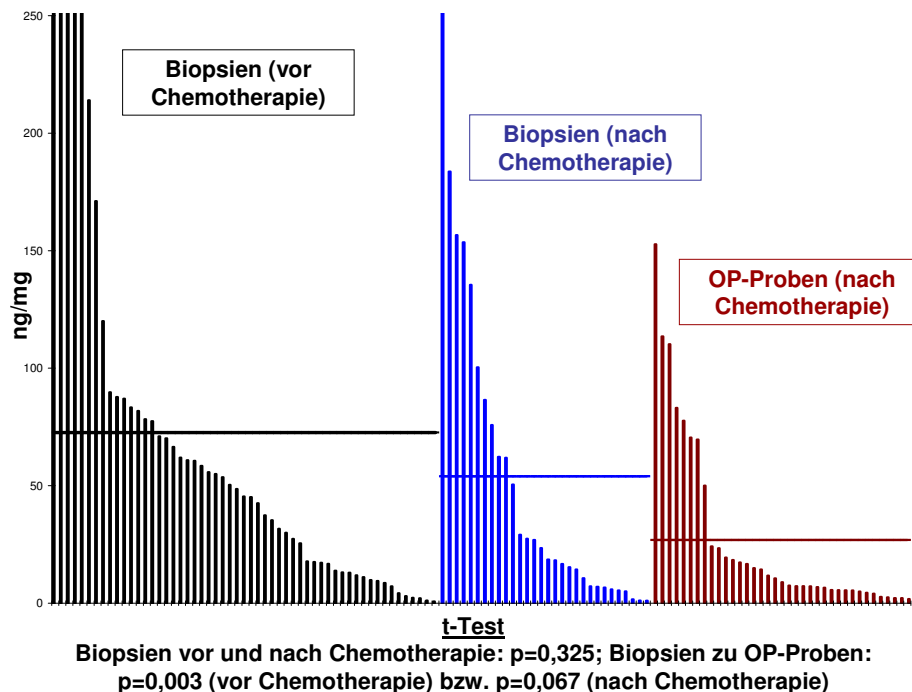


Abbildung 4.4: Gewonnene RNA in ng pro mg Tumorgewebe für die verschiedenen Probenarten. Dargestellt sind zusätzlich die durchschnittlich gewonnenen RNA-Mengen pro Probenart (waagrechte Linien) sowie im Text die Ergebnisse aus dem t-Test.

wird durch rote und grüne Boxen repräsentiert, wie in Abbildung 4.5 zu sehen ist. Die Analyse erfolgt zunächst rein visuell auf Basis dieser grafischen Darstellung und ist daher als rein qualitativ anzusehen.

Für die Gene der **Angiogenese** kann keine eindeutige Tendenz über alle Tumorproben beobachtet werden. Einige Gene sind in einer Vielzahl der Tumoren nicht reguliert, es können jedoch auch Unterschiede in der Regulation zwischen verschiedenen Tumoren beobachtet werden.

Auch die mit **Apoptose** assoziierten Gene zeigen Variationen in ihrer Expression. Die Ausprägung ihrer Expression ist bei diesen Genen jedoch insgesamt stärker als in der Gruppe der Angiogenese-Gene.

Gene der **Zellzyklusregulation** sind im Vergleich zur Referenz-RNA tendenziell weniger stark exprimiert.

In **Lipid- und Hormonstoffwechsel** involvierte Vertreter ebenso wie mit **Zelladhäsion und extrazellulärer Matrix** assoziierte Gene sind im Gegensatz dazu überwiegend hochreguliert.

Die analysierten **cDNA**-Sequenzen zeigen insgesamt eine eher geringere Expression und weisen häufiger fehlende Messwerte auf.

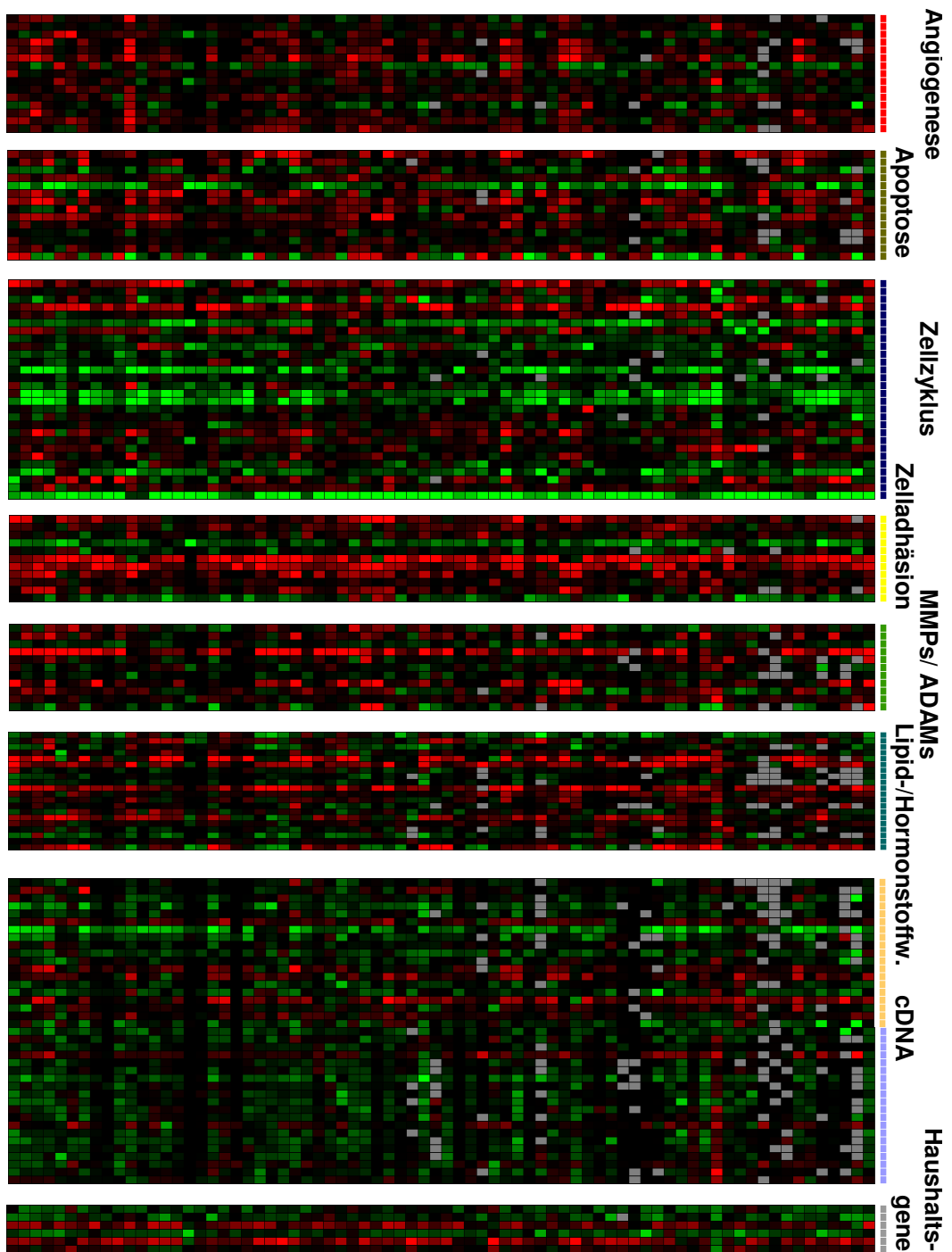


Abbildung 4.5: Heat Maps der Expressionsdaten nach Funktionsklassen der analysierten Gene. Runterregulierte Gene werden grün, hochregulierte Gene rot dargestellt. Die Farbintensität repräsentiert die Stärke der Expression. Nicht-regulierte Gene erscheinen schwarz, fehlende Messwerte grau.

Betrachtet man die Expression der **Haushaltsgene** fällt auf, dass nur zwei (ACTB und KARS) der insgesamt sechs Gene als nicht-regulierte Referenzgene geeignet sind. Die anderen vier sind im Vergleich zur verwendeten Referenz über- (z.B. B2M, siehe auch Tabelle 4.8) oder unterrepräsentiert.

In den folgenden detaillierten Analysen sollen zum einen einzelne Gene identifiziert werden, die unabhängig von Tumorart oder anderen individuellen Unterschieden in der Mehrheit der Tumoren stark reguliert sind (Abschnitt 4.4.1). So sollen Hinweise auf biologische Prozesse gewonnen werden, die grundsätzlich bei der Tumorentwicklung eine Rolle spielen. Zum anderen sind Gene interessant, die auf Transkriptionsebene Unterschiede zwischen den einzelnen Tumoren aufweisen (Abschnitt 4.4.2). Diese Unterscheidung verschiedener Brusttumoren anhand ihrer Genprofile ist eine grundlegende Fragestellung dieser Arbeit.

4.4.1 *Gene mit gleich bleibender Tendenz der differentiellen Expression in allen Tumorproben*

Die Identifizierung von Genen mit gleich bleibender Tendenz der Expression in den untersuchten Brusttumoren - unabhängig von individuellen Unterschieden - basiert auf allen Messungen, die sowohl einen großen als auch signifikanten Unterschied in der Expression im Vergleich zur Referenz-RNA aufweisen:

Für jedes Gene wird die mittlere Expression durch Berechnung des arithmetischen Mittelwertes \bar{M} der M -Werte (siehe Formel 3.2) über alle Tumorproben bestimmt. Dabei werden für jede Tumorprobe nur die gemessenen M -Werte berücksichtigt, die im t-Test über alle Replikate eine signifikante Regulation (5 %-Signifikanzniveau) im Vergleich zur Referenz zeigten (vgl. Abschnitt 4.2.2). Mit Hilfe des Konfidenzintervalls kann dann unter Berücksichtigung der Anzahl Proben (n), für die $p < 0,05$ zutrifft, die Zuverlässigkeit des ermittelten Mittelwertes eingeschätzt werden.

Filterkriterien gleiche Expression: Es wird nach Genen gefiltert, die in mehr als der Hälfte aller analysierten Tumorproben von Studienpatientinnen ($n > 38$) unter Berücksichtigung des 95 %-Konfidenzintervalls (K) mindestens zweifach hoch- oder runterreguliert sind: $|\bar{M}| - 1/2K < -1$ oder $|\bar{M}| - 1/2K > 1$.

Die resultierenden Gene sind in den Tabellen 4.8 (hochregulierte Gene) und 4.9 (runterregulierte Gene) dargestellt.

Tabelle 4.8: Signifikant hochregulierte Gene über alle Tumorproben. Angegeben sind jeweils der arithmetische Mittelwert \bar{M} und die halbe Breite des 95 %-Konfidenzintervalls K .

Reporter Name	Gen Name	\bar{M}	$\pm 1/2K$	Gen Funktion
R:McBiSet2_21	cytochrome P450, family 1, subfamily B, polypeptide 1;CYP1B1	2.46	0.35	Lipid-/Hormonstoffwechsel
R:McBiSet2_25	cadherin 1, type 1, E-cadherin (epithelial);CDH1	2.81	0.24	Zelladhäsion
R:McBiSet2_34	collagen, type IV, alpha 2;COL4A2	2.23	0.27	Zelladhäsion
R:McBiSet2_73	Homo sapiens beta-2-microglobulin (B2M), Housekeeping gene	1.40	0.24	Haushaltsgen
R:McBiSet2_89	fatty acid synthase;FASN	1.32	0.29	Lipid-/Hormonstoffwechsel
R:McBiSet2_107	insulin-like growth factor binding protein 5;IGFBP5	1.45	0.39	Zellzyklus
R:McBiSet2_131	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian);ERBB2	2.59	0.45	Zellzyklus
R:McBiSet2_137	matrix metalloproteinase 11 (stromelysin 3);MMP11	3.00	0.47	MMP

Um einen besseren Eindruck über die Verteilung der Expressionsdaten der identifizierten Gene zu bekommen werden die signifikanten M -Werte der in den Tabellen 4.8 und 4.9 aufgeführten Gene zusätzlich mittels Box-Plots grafisch dargestellt (Abbildung 4.6). Die starke Tendenz der Expressionsausprägung der herausgefilterten Gene wird hier bestätigt. Bis zu 75 % der gemessenen M -Werte eines Gens liegen jeweils über bzw. unter Null. Der Vergleich von arithmetischem Mittelwert und für die Box-Plots verwendeten Median-Werte zeigt, dass beide Maße die Daten gut repräsentieren. Auch wenn zum Teil Extremwerte auftreten verzerren diese die mittlere Verteilung der signifikanten Messwerte kaum.

Werden die Daten nach den selben Kriterien (mindestens 2-fach reguliert, $M < -1$ oder $M > 1$, in mehr als der Hälfte aller analysierten Tumorproben, $n > 38$) statt auf Basis des arithmetischen Mittelwertes anhand des **Medians** gefiltert, ergibt dies mit einer Ausnahme die identischen Gene. Als signifikant hoch reguliert wird hier zusätzlich das Gen IGFBP4 erkannt, dessen arithmetischer Mittelwert knapp unter 1 liegt.

Insgesamt konnten 22 Gene als in der Mehrheit der Tumorproben stark reguliert identifiziert werden. Unter den acht Überrepräsentierten (Tabelle 4.8) sind Gene, deren Produkte mit der extrazellulären Matrix assoziiert sind, wie etwa **zelladhäsion**-vermittelnde Proteine und eine **Matrix-Metalloproteinase** sowie in den **Lipidstoffwechsel** involvierte Gene. Diese Funktionsklassen konnten bereits in der Heat Map als in ihrer Tendenz hochreguliert beobachtet werden. Die zwei mit dem Zellzyklus assoziierten hochregulierten Gene beeinflussen ebenfalls Prozesse der Zellmigration, wo-

Tabelle 4.9: Signifikant runterregulierte Gene über alle Tumorproben. Angegeben sind jeweils der arithmetische Mittelwert \bar{M} und die halbe Breite des 95 %-Konfidenzintervalls K .

Reporter Name	Gen Name	\bar{M}	$\pm 1/2K$	Gen Funktion
R:McBiSet2_2	antigen identified by monoclonal antibody Ki-67;MKI67	-2.06	0.30	Zellzyklus
R:McBiSet2_3	antigen identified by monoclonal antibody Ki-67;MKI67	-2.21	0.32	Zellzyklus
R:McBiSet2_9	insulin-like growth factor 2 (somatomedin A);IGF2	-2.85	0.25	Zellzyklus
R:McBiSet2_10	telomerase reverse transcriptase;TERT	-2.31	0.35	nucleobase, nucleoside, nucleotide and nucleic acid
R:McBiSet2_29	cyclin A2;CCNA2	-2.12	0.33	Zellzyklus
R:McBiSet2_37	topoisomerase (DNA) II alpha 170kDa;TOP2A	-2.43	0.36	cell cycle (GO:0007049)
R:McBiSet2_101	transketolase (Wernicke-Korsakoff syndrome);TKT	-1.16	0.11	carbohydrate metabolism (GO:0005975)
R:McBiSet2_115	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in);TFF1	-2.25	0.65	cell proliferation (GO:0030154); cell differentiation
R:McBiSet2_138	cDNA_204_new	-1.91	0.35	cDNA neu
R:McBiSet2_141	solute carrier family 2 (facilitated glucose transporter), member 3;SLC2A3	-2.62	0.23	carbohydrate metabolism (GO:0005975)
R:McBiSet2_142	nucleolar and spindle associated protein 1;NUSAP1	-1.52	0.22	biological_process unknown (GO:0000004)
R:McBiSet2_148	insulin-like growth factor binding protein 3;IGFBP3	-1.43	0.18	Zellzyklus
R:McBiSet2_159	ribosomal protein SA;RPSA	-1.28	0.17	Zelladhäsion
R:McBiSet2_162	v-myc myelocytomatosis viral oncogene homolog (avian);MYC	-1.58	0.34	Apoptose

bei das Tumorsuppressorgen IGFBP5 ebenso wie das zusätzlich identifizierte IGFBP4, Zellmigration und -proliferation hemmen und das Onkogen ERBB2 Zelladhäsionsprozesse fördert. Auffällig ist, dass sich auch ein **Haushaltsgen** unter den signifikant über alle Tumorproben hochregulierten Genen befindet. Die 14 stark herunterregulierten Gene (Tabelle 4.9) kodieren in erster Linie für Proteine, die in **Zellzyklusregulation** und **Kohlenhydratstoffwechsel** involviert sind.

4.4.2 Gene mit stark variierender differentieller Expression über alle Tumorproben

Gene, die in ihrer Expression starke Variationen zwischen den untersuchten Tumoren aufweisen, können Hinweise auf ein unterschiedliches Verhalten oder auch eine unterschiedliche Biologie der Tumoren geben. Daher wird nach Genen mit einer möglichst großen Streuung der für sie gemessenen Expression M gesucht. Hierfür wird die Bestimmung von Quartilen und ihren

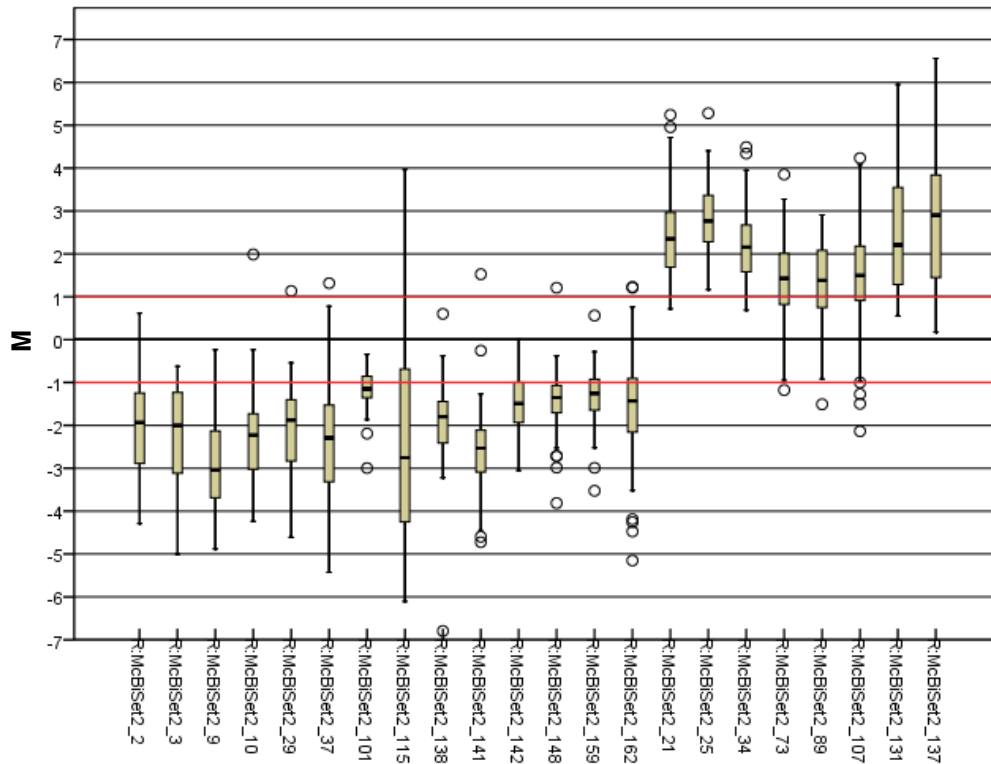


Abbildung 4.6: Box-Plots der signifikant regulierten Gene. Dargestellt ist die Verteilung der M -Werte für die Gene mit gleich bleibender Tendenz ihrer differentiellen Expression in allen Tumorproben. Die roten Linien kennzeichnen den gesetzten Schwellenwert für M bei -1 und 1.

Abstände herangezogen. Die Visualisierung mittels Box-Plots ermöglicht zudem die Darstellung von Höchst- und Extremwerten (vgl. Methode in Kapitel 3.4.5.1).

Wie bereits im ersten Teil der Genexpressionsanalysen beschrieben (Abschnitt 4.4.1) werden auch hier zunächst die signifikanten Messwerte betrachtet (5 %-Signifikanzniveau: $p < 0,05$). Für jedes Gen wird der zentrale Expressionswert, der Median von M , über alle Tumorproben bestimmt.

Filterkriterien variierende Expression, Schritt 1: Als Filterkriterien wurde ein Interquartilsabstand $QA = |25\% \text{ Quartil} - 75\% \text{ Quartil}| > 2$ und signifikante Messwerte für mindestens ein Viertel der analysierten Proben ($n > 20$) angenommen. Der Schwellenwert für den Stichprobenumfang kann hier geringer gewählt werden als in Abschnitt 4.4.1, da zwar keine Einzelmessungen berücksichtigt, jedoch auch die Gene identifiziert werden sollen, die nur in einigen Tumoren auffällig sind. Der große Quartilsabstand stellt sicher, dass Gene ausgewählt werden, die in ihrer Expression mindestens einen vierfachen Unterschied zwischen verschiedenen Tumorproben aufweisen.

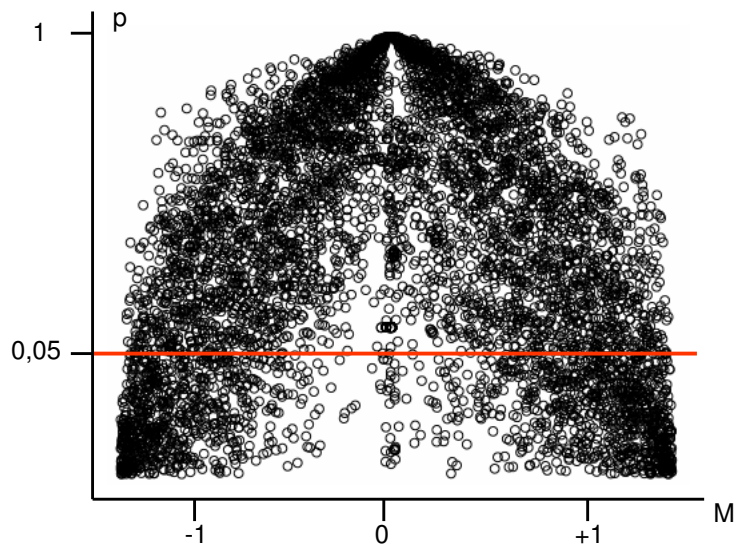


Abbildung 4.7: Die Grafik zeigt exemplarisch die Verteilung der p -Werte für die gemessenen Expressionswerte M . Bei unveränderter Expression ($M = 0$), also keinem Unterschied zwischen den beiden Stichproben (Referenz und Tumorprobe), liefert der t-Test ein großes p , da die Nullhypothese anzunehmen ist.

Bei der Betrachtung der Streubreite der Expression nur über alle signifikanten Messwerte ($p < 0,05$), werden möglicherweise nicht alle interessanten Gene erfasst: Für im Vergleich zur Referenz nicht-regulierte Gene ist der p -Wert in der Regel größer als 0,05, da die Nullhypothese (keine Regulation) zutrifft. Nur sehr wenige Gene zeigen geringe und dennoch hoch signifikante Unterschiede in ihrer Expression im Vergleich zur Referenz (vgl. Abbildung 4.7). Je geringer die zu detektierenden Unterschiede sind, desto größer müsste die Stichprobe sein, um dennoch signifikante Werte zu erhalten. Des Weiteren ist zu berücksichtigen, dass geringe Expressionsunterschiede aufgrund zu geringer Signalintensitäten häufig gar nicht erst detektiert werden.

Filterkriterien variierende Expression, Schritt 2: Da nicht nur Gene berücksichtigt werden sollen, die in den verschiedenen Tumorproben hoch- und runterreguliert sind, sondern auch solche, die in einigen Proben eine starke Regulation (hoch oder runter) aufweisen, in anderen jedoch nicht reguliert sind, wird das Verfahren zur Überprüfung zusätzlich auf dem ungefilterten Datensatz durchgeführt.

Filterkriterien variierende Expression, Schritt 3: Zuletzt werden aus dieser Auswahl diejenigen Gene entfernt, bei denen 75 % der Messwerte über 1 oder unter -1 liegen, obwohl sie einen großen Quartilsabstand aufweisen. So

können Überschneidungen mit den in Abschnitt 4.4.1 identifizierten Genen weitestgehend vermieden werden.

Nur ein Gen, das Gen TFF1, wird mit beiden Ansätzen herausgefiltert. Es ist zwar in der Mehrheit der Proben signifikant herunterreguliert, zeigt jedoch in einigen wenigen Proben eine deutliche Überexpression und ist deshalb auch ein interessantes Kandidatengen zur Unterscheidung der Tumoren.

Alle aus dem Filterprozess resultierenden Gene können in Tabelle 4.10 eingesehen werden und werden nachfolgend kurz beschrieben.

Tabelle 4.10: Gene mit stark unterschiedlicher Expression über alle Tumorproben. Angegeben sind jeweils 25 %, 50 % und 75 % Quartil für alle Gene mit einem Interquartilsabstand der gemittelten (Median) *M*-Werte von $QA = |25\% \text{ Quartil} - 75\% \text{ Quartil}| > 2$.

Reporter Name	Gen Name	25 %	50 %	75 %	QA	Gen Funktion
R:McBiSet2_47	glutathione S-transferase M1;GSTM1	-0.78	0.92	1.97	2.75	defense response (GO:0006952)
R:McBiSet2_75	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1);SPP1	-1.87	-0.48	1.47	3.34	Apoptose
R:McBiSet2_85	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian);EGFR	-0.26	1.63	2.95	3.21	Zellzyklus
R:McBiSet2_110	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4);CDKN2A	-1.52	-0.48	1.58	3.10	Zellzyklus
R:McBiSet2_115	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in);TFF1	-4.29	-2.75	-0.62	3.67	cell proliferation (GO:0030154); cell differentiation
R:McBiSet2_151	estrogen receptor 1;ESR1	-0.10	1.29	2.83	2.93	Lipid-/Hormonstoffwechsel
R:McBiSet2_175	cyclin D1;CCND1	-1.30	-0.76	0.86	2.16	Zellzyklus
R:McBiSet2_187	Homo sapiens mucin 1, cell surface associated;MUC1	-1.10	1.62	2.82	3.92	Lipid-/Hormonstoffwechsel

Das für eine Glutathion-S-Transferase kodierende Gen GSTM1 weist in den hier untersuchten Tumoren große Schwankungen in seiner Expression auf. Das Protein ist im Zytoplasma der Zelle lokalisiert und spielt dort eine wichtige Rolle bei Entgiftungsprozessen.

Das Gen SPP1 kodiert das Glykoprotein Osteopontin (OPN), das in Prozesse der Angiogenese, Apoptose sowie Entzündungsprozesse involviert ist. OPN kann mit einer Reihe von Faktoren wie dem Wachstumsfaktorrezeptor EGFR und anderen membranständigen Rezeptoren interagieren, ebenso mit sezernierten Proteinen wie z.B. Matrix-Metalloproteinasen.

Auch einige mit dem Zellzyklus assoziierte Gene weisen in den Brusttumoren eine stark unterschiedliche Regulation auf. So z.B. der im Zusammenhang mit SPP1 genannte *Epidermal Growth Factor Receptor* EGFR oder das Tumor-

suppressorgen CDKN2A, ein Inhibitor der zyklinabhängigen Kinase CDK4. Auch das Gen CCND1, das ein CDK4 regulierendes Zyklin kodiert, zeigt große Expressionsunterschiede.

Die Transkription des TFF1-Gens wird durch Östrogen induziert. Sowohl TFF1 als auch das den Östrogenrezeptor Alpha kodierende Gen ESR1 weisen in dem untersuchten Gewebe eine stark variierende Expression auf. Ein weiteres mit dem Östrogenrezeptor assoziiertes Protein, ein Transmembranprotein kodiert durch MUC1, ist ebenfalls unter den gefilterten Genen.

In den anschließenden Clusteranalysen werden die hier erhaltenen Gene besonders berücksichtigt. So können noch weitere Gene, die durch die gewählten Filterkriterien nicht erfasst wurden, aber dennoch ein ähnliches Expressionsmuster aufweisen, identifiziert werden.

4.4.3 Identifizierung auffälliger Gengruppen

Bisher wurde die Expression einzelner Gene betrachtet, in diesem Kapitel sollen nun mittels Clusteranalysen (zur Methode siehe Kapitel 3.4.5.2) weitere Gengruppen (Cluster) identifiziert werden, die Auffälligkeiten in den Tumoren zeigen.

Auf Basis der ungefilterten Expressiondaten für alle Studienpatientinnen wurde ein hierarchisches agglomeratives Clusterverfahren angewandt. Acht Tumorproben (26-B_1092A_OP, 27-AA_290_BP-0, 29-B-2_1065_BP-1, 38-C_1094_OP, 40-C_1104A_OP, 54-C_585_OP, 55-A-2_326_BP-0, 69-B-2_400_BP-1) wurden aus den Analysen ausgeschlossen, da hier für mehr als ein Drittel der Gene keine Werte vorlagen. In die weiteren Analysen sind demnach 69 Proben von 37 Patientinnen eingegangen.

BENACHBARTE GENE IM CLUSTER Für die Analyse von Genen mit gleichem Verhalten in ihrer Expression über alle Tumorproben, unabhängig von den absoluten Messwerten, wird die **Pearson Korrelation** als Ähnlichkeitsmaß zur Abstandsberechnung der Gene verwendet. Anhand des resultierenden Dendrogramms (Abbildung 4.8) können auf Genebene 11 größere Cluster identifiziert werden (Abbildung 4.9 A).

Innerhalb dieser Cluster werden die stark variierenden Gene aus Abschnitt 4.4.2 betrachtet und ihre unmittelbare Nachbarschaft im Cluster-Baum analysiert. So werden weitere auffällige Gene gefunden, die durch die strikten Filterkriterien bisher unberücksichtigt blieben. Die identifizierten Gene

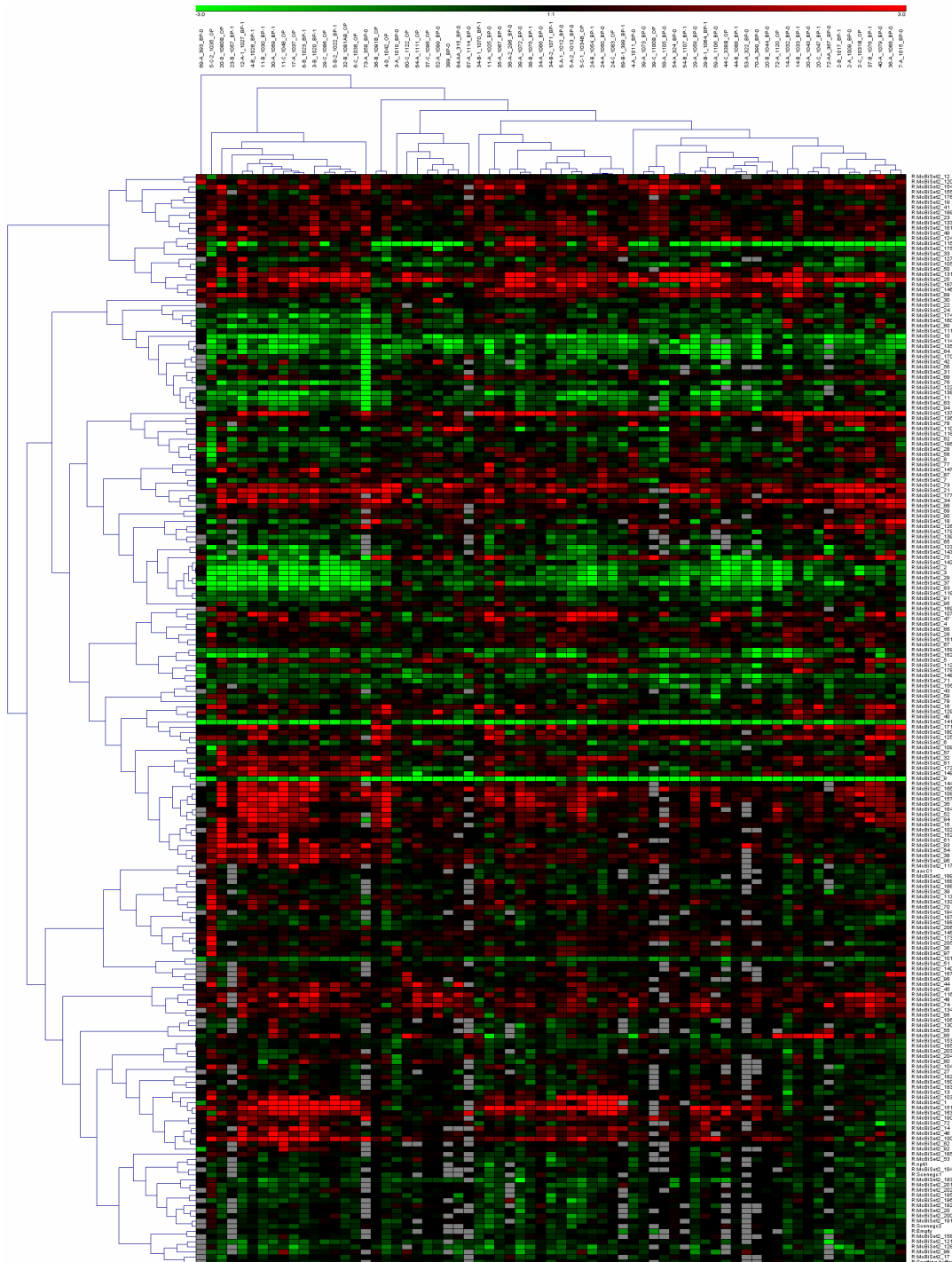


Abbildung 4.8: Agglomeratives hierarchisches Clustern (Distanzmaß: Pearson Korrelation). Diese Gesamtdarstellung zeigt die resultierenden Cluster auf Gen- (links) und Probenebene (oben). Die Genexpression ist farblich kodiert, wobei die Intensität die Stärke der Expression im Vergleich zur Referenz wieder gibt: grün=runterreguliert, rot=hochreguliert, schwarz=nicht-reguliert, grau=fehlende Werte.

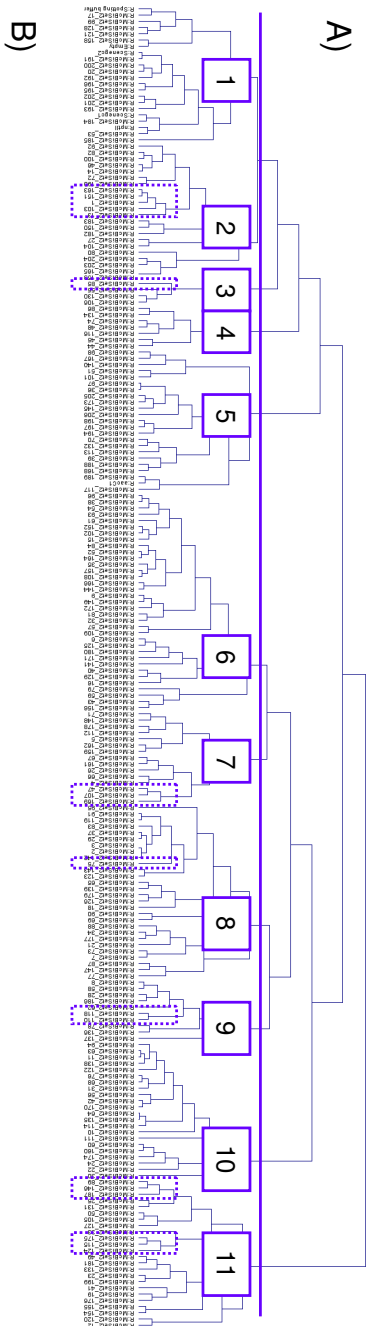


Abbildung 4.9: Teil A der Abbildung zeigt das mittels agglomerativen hierarchischen Clustern erhaltene Dendrogramm für die 11 Haupt-Gencluster aus
 Abbildung 4.8. Ausschnitt B zeigt die direkt benachbarten Gene der in Kapitel 4.4.2 als stark variierend in ihrer Expression identifizierten
 Gene (rot markiert).

werden nachfolgend kurz beschrieben, ihre Lokalisation im Cluster-Baum ist zusätzlich in Abbildung 4.9 B dargestellt.

- ▷ Das Gen ESR1 ist im 2. Cluster lokalisiert und zeigt eine ähnliche Expression wie der Anti-Apoptose-Regulator BCL2, das Gen für den mit dem EGF verwandten Wachstumsfaktorliganden SCUBE2, das N-Acetyltransferase2-Gen NAT2 und das für den Progesteronrezeptor kodierende Gen PGR.
- ▷ In Cluster 3 befindet sich das Gen EGFR. Es können auf unterster hierarchischer Ebene im Cluster keine assoziierten Gene gefunden werden. Insgesamt beinhaltet dieses Cluster nur vier Gene, wobei EGFR auch mit Abstand die höchsten Expressionswerte aufweist.
- ▷ GSTM1 kodiert für die Glutathion-S-Transferase-1 und liegt im Cluster-Baum im 7. Cluster, neben dem als stark hochreguliert identifiziertem IGFBP5 und dem Metalloproteinase-Inhibitor-Gen TIMP1.
- ▷ Das Osteopontin kodierende Gen SPP1 ist, ebenso wie EGFR ohne direkt benachbarte Gene, in Cluster 8 lokalisiert.
- ▷ Das Gen des zyklinabhängigen Kinase-Inhibitors 2A (CDKN2A) zeigt eine ähnliche Expression wie ACTB, das als ubiquitär exprimierte Haushaltsgen dem Themenarray hinzugefügt wurde. Beide Gene sind in Cluster 9 des resultierenden Dendrogramms zu finden.
- ▷ Als weitere östrogenassoziierte Gene sind MUC1 und das tendenziell hochregulierte Gen FASN in Gruppe 11 des Clusters zu finden, ebenso das Onkogen AKT1. In einem weiteren Unterzweig des Clusters finden sich das Gen CCND1 als ein Mitglied der Zyklin-Familie, das durch Östrogen induzierbare Gen TFF1 und SREBF1, ein Transkriptionsfaktor der in die Steroidhormon-Biosynthese involviert ist.

KORRELATIONEN ZU WEITEREN GENEN Um zusätzlich eine Aussage über die Signifikanz der Zusammenhänge auf Expressionsebene machen zu können wurde ausgehend von den acht stark variierenden Genen die **Pearson-Korrelation** zu weiteren Genen bestimmt. Dargestellt sind hier die Ergebnisse mit einem Korrelationskoeffizienten größer 0,5 oder kleiner -0,5 und 5 %-Signifikanzniveau ($p < 0,05$):

ESR1 zeigt zu allen seinen im Cluster direkt benachbarten Genen, BCL2, SCUBE2, NAT2 und PGR auch eine signifikante Korrelation der Expressionsergebnisse. Zusätzlich konnte eine Korrelation zu ERBB4, ein Gen desselben Clusters, das bei den bisherigen Analysen jedoch noch nicht berücksichtigt

wurde, beobachtet werden. Auch zu den im Cluster-Baum weit entfernten Genen GSTM1 (Cluster 7) und TFF1 (Cluster 11) konnte eine signifikante Ähnlichkeit hinsichtlich des Expressionsverhaltens festgestellt werden.

Das Gen für den Trefoil-Faktor-1 (TFF1) in Cluster 11 wiederum zeigt von seinem Expressionsmuster her signifikante Korrelationen zu dem im Cluster direkt benachbarten CCND1-Gen, aber auch zu dem in Cluster 2 lokalisierten Gen SCUBE2.

SPP1 hat im Cluster-Baum auf unterster hierarchischer Ebene in Cluster 8 keine direkt benachbarten Gene, zeigt aber signifikante Zusammenhänge bezüglich der Expression mit den im selben Cluster lokalisierten PRC1 und dem signifikant runterregulierten Gen MKI-67.

Unter Verwendung der **euklidischen Distanz** zur Berechnung der Abstände zwischen den Genen werden Gene mit ähnlicher Expressionsstärke zusammengefasst. Hier ergeben sich drei große Cluster. Zwei der drei Cluster sind stark auffällig, da sie nur stark herunter- bzw. hochregulierte Gene enthalten. Diese Cluster beinhalten unter anderem alle die bereits in Abschnitt 4.4.1 als über alle Tumorproben gleich reguliert identifizierten Gene.

Insgesamt fallen bei der Betrachtung der Gene, die in ihrer Expression starke Variationen zwischen den untersuchten Tumoren aufweisen, die beiden Hormonrezeptorgene PGR und ESR1 als Gene mit besonders hoher klinischer Relevanz (etablierte prognostische Faktoren, siehe Kapitel 2.2) auf. Aber z.B. auch das für den wichtigen klinischen Proliferationsmarker Ki-67 kodierende Gen MKI-67, das in diesen Analysen stark unterexprimiert ist.

Des Weiteren zeigt sich EGFR (oder auch ERBB1) als Gen mit ausgeprägter variierender Expression. Weitere Gene, deren Produkte mit diesem Rezeptor interagieren (SPP1, SCUBE2), zeigten in den Analysen Zusammenhänge auf Expressionsebene. Für verwandte Rezeptorgene wie ERBB4 konnten ebenfalls eindeutige Korrelationen nachgewiesen werden. Das klinisch hauptsächlich aufgrund seiner Bedeutung als etablierter prädiktiver Faktor relevante Onkogen ERBB2 ist hingegen in den untersuchten Tumoren stark hochreguliert und zeigt keine signifikanten Expressionsunterschiede zwischen einzelnen Tumoren.

4.5 ZUSAMMENFÜHRUNG VON KLINISCHEN PARAMETERN UND GENEXPRESSIONSDATEN

Bisher wurden die Expressionsdaten isoliert betrachtet. Eine weitere Zielsetzung dieser Arbeit ist es, mögliche Zusammenhänge zwischen der Genexpression in Tumoren und ihren klinischen Eigenschaften aufzuzeigen. Hierfür wird nun die **Gruppierung der Tumorproben** auf Basis der Ähnlichkeit ihrer Expressionsmuster untersucht und die Eigenschaften der Tumoren genauer beleuchtet.

Betrachtet man die Gruppierung der Tumorproben auf Basis der exprimierten Gene (Abbildung 4.8) können fünf große Cluster beobachtet werden. Betrachtet man die Tumorproben der einzelnen Cluster genauer, fällt auf, dass sich die verschiedenen Proben einer Patientin unabhängig von Entnahmezeitpunkt und -art (vor oder nach Therapie; Biopsie oder OP-Präparat) im Cluster häufig näher und damit ähnlicher sind, als z.B. Proben gleicher Probenart verschiedener Patientinnen.

Ausgewählte (sowohl tumor- als auch patientinnenbezogene) klinische Parameter der gruppierten Tumorproben werden in Abbildung 4.10 dargestellt. Da es sich um völlig unterschiedliche Parameter mit variierender Maßeinteilung handelt wurden sie für die Analysen in ein Binärformat transformiert, um so eine Vergleichbarkeit zu erreichen. Hierfür wurden tumorbiologische Erkenntnisse als Grundlage genutzt. So wird zum Beispiel nach positivem oder negativem Hormonrezeptorstatus, nach dem Vorhandensein von Metastasen oder jüngeren und älteren Patientinnen (zu den Risiko- und Prognosefaktoren für die Brustkrebserkrankung siehe Kapitel 2) unterschieden. Bei den Hormonrezeptoren werden zudem zwei unterschiedliche Schwellenwerte gesetzt, um einen Informationsverlust zu vermeiden. Einmal wurde der Schwellenwert für einen positiven Rezeptorstatus relativ hoch (IRS=9; ER/PGR pos.), im anderen Fall deutlich niedriger (IRS=3; ER/PGR neg.) angesetzt. Die genauen Bedingungen für die binäre Umwandlung sind Tabelle 4.11 zu entnehmen.

4.5.1 Rangkorrelationsanalysen

Bei der Analyse der Ausprägung der **klinischen Parameter in Abhängigkeit der Proben-Cluster** 1-5 (Abbildung 4.10) fallen einige klinische Faktoren auf, die mit den Clustergrenzen übereinzustimmen scheinen (z.B. ER-Status,



Abbildung 4.10: Klinische Parameter und ihre Ausprägung dargestellt für die resultierenden fünf Probencluster, grün=0 und orange=1. Die zugrunde liegende Umwandlung der Werte in ein Binärformat ist in Tabelle 4.11 einzusehen.

Tabelle 4.11: Bedingungen der binären Umwandlung für die untersuchten klinischen Parameter.

Parameter	Bedingung für 0	Bedingung für 1
Alter	<40	>=40
Menopause	nein	ja
BMI	<30	>=30
Histo	lobulär	duktal
T	<3	>=3
N	0	>0
M	0	>0
Grade	2	3
PGR (pos.)	<9	>=9
PGR (neg.)	<3	>=3
ER (pos.)	<9	>=9
ER (neg.)	<3	>=3
Her2-neu	<3	3
Para (Geburten)	0	>0
Erstgeburt (Alter)	<30	>=30
Stillen	nein	ja
Rauchen	nein	ja
Alkohol	nein	ja
Sport	<2h/Woche	>=2h/Woche
Chemoregime	TAC	AC DOC
Chemo	vor Therapiebeginn	nach Therapiebeginn
familiärer Brustkrebs	nein	ja

Histologie oder Metastasierung). Um genauere Aussagen darüber machen zu können, ob die beobachtete Ausprägung der Parameter innerhalb eines Clusters von der über alle Patientinnen abweicht, wurde für jeden Parameter die Rangkorrelation nach Spearman bestimmt. Zusätzlich wurden die Tumoren anhand des Status der Hormon- und Wachstumsfaktorrezeptoren (PGR, ER und Her2-neu) ihrem jeweiligen molekularen Subtyp (*Luminal A/B*, *Basal-like*, $Her2^+ / ER^-$; siehe Kapitel 2.3 der Einleitung) zugeordnet und ebenfalls die Korrelation zu den Probenclustern bestimmt. Für die tumorbezogenen Parameter Grading, Östrogenrezeptorstatus (ER) und Progesteronrezeptorstatus (PGR) ergibt sich eine statistisch signifikante Korrelation zu der auf Genexpressions-Basis ermittelten Ähnlichkeit der Proben. Diese Unterschiede spiegeln sich auch in den molekularen Tumorsubtypen wider.

Auf Clusterebene bedeutet dies konkret:

- ▷ Cluster 2 beinhaltet Proben von Tumoren mit kaum differenzierten Zellen (Grade 3) und negativem Hormonrezeptorstatus. Diese Faktoren geben Hinweise für eine eher ungünstige Prognose. Dies bestätigt auch

die positive Korrelation mit dem Tumortyp *Basal-like*, der weitestgehend einem *Triple-negativ*-Tumortyp (ER^{-} , PGR^{-} , $Her2\text{-neu}^{-}$) entspricht.

- ▷ Die in Cluster 3 zusammengefassten Tumoren weisen mäßig differenzierte Zellen (Grade 2) sowie einen positiven Östrogenrezeptorstatus auf und wären somit mit einer besseren Prognose des Krankheitsverlaufs assoziiert. Die Tumorproben dieses Clusters weisen signifikant häufiger als in anderen Clustern den *Luminal A*-Tumortyp auf.
- ▷ Cluster 5 enthält wiederum Proben von Tumoren mit einer voraussichtlich schlechteren Prognose, da beide Hormonrezeptoren einen negativen Status aufweisen. Dieses Cluster zeigt eine signifikante Korrelation zum $Her2^{+}/ER^{-}$ -Tumortyp.
- ▷ Die Cluster 1 und 4 zeigen keine signifikanten Korrelationen zu den klinischen Tumordaten.

Für die auf den ersten Blick auffallenden Parameter Histologie und Metastasierung kann keine Korrelation gefunden werden. Alle hier untersuchten Patientinnen sind metastasiefrei und der Großteil weist einen duktalem Tumortyp auf (vgl. Risikotypisierung der Tumoren in Abschnitt 4.2.1). Der rein visuell zunächst vermutete Zusammenhang zu den resultierenden Clustern ist aufgrund der zu geringen Fallzahlen in jeweils einer der Kategorien demnach nicht gegeben.

Auch die Abhängigkeiten der **klinischen Parameter untereinander** wurden mittels Korrelationsanalysen untersucht. Für die Hormonrezeptoren ergibt sich eine stark positive Korrelation. Das Grading ist signifikant mit dem Östrogenrezeptorstatus verknüpft, dabei ist der Tumordifferenzierungsgrad 3 mit einem negativen Rezeptorstatus assoziiert. Die Tumorsubtypen *Luminal A*, *Luminal B* und *Basal-like* unterscheiden sich signifikant voneinander, Tumoren des Typs $Her2^{+}/ER^{-}$ zeigen keine signifikanten Unterschiede zu den anderen Typen, in diese Gruppe fallen nur vier der untersuchten Tumoren (Signifikanzniveau 1 %).

Korrelationsanalysen der binär umgewandelten **klinischen Parameter mit einzelnen Genen** (M -Werte aus den Mikroarrayhybridisierungen) mittels Kendall's τ konnten, neben der positiven Korrelation der auf Protein- und Genebene untersuchten Faktoren ER, PGR und Her2-neu, weitere interessante signifikante Zusammenhänge ($|\tau| > 0,4$ und $p < 0,05$) aufzeigen. So ist das Grading z.B. negativ mit der Expression von SCUBE2 korreliert ($\tau = -0,41$) und der Status des Progesteronrezeptors PGR und die SREBF1-Expression zeigen eine positive Korrelation ($\tau = 0,46$). Der Östrogenrezeptorstatus

weist unter anderem mit ERBB4 und NAT2 positive Zusammenhänge, mit EGFR jedoch eine gegenläufige Tendenz auf. Hier können zwar insgesamt die höchsten Werte für τ beobachtet werden, sie liegen jedoch nur knapp über 0,5 (positive Korrelationen) oder unter 0,5 (negative Korrelation). Berücksichtigt wurden für den Östrogenrezeptor jeweils nur die Ergebnisse, die mit beiden Formen der binären Umwandlung (positiv und negativ, vgl. Tabelle 4.11) ein signifikantes Ergebnis ergaben.

4.5.2 Normalized Tree Index

Der Normalized Tree Index (NTI) bietet die Möglichkeit die Genexpressionsdaten in ihrer Gesamtheit zu erfassen und Korrelationen zu einzelnen klinischen Parametern aufzuzeigen. Auch hier wird mit bereits geclusterten Expressionsdaten gearbeitet. Abbildung 4.11 zeigt die Ergebnisse des NTI für ausgewählte klinische Parameter.

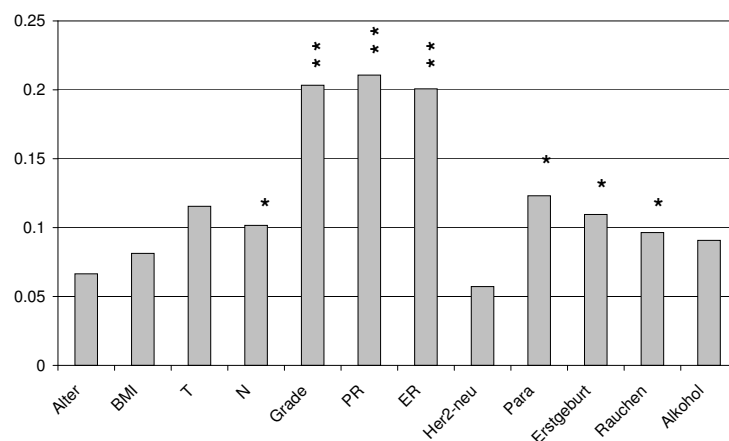


Abbildung 4.11: Ergebnisse des Normalized Tree Index (NTI) für ausgewählte klinische Parameter. Signifikante Werte sind jeweils mit Stern markiert (*=5 %-Signifikanzniveau; **=1 %-Signifikanzniveau).

Wie bereits im vorherigen Abschnitt zeigen auch hier die Proteinexpression der Hormonrezeptoren und das Grading der Tumoren die größte Korrelation zum Expressionsprofil der untersuchten Gene.

Auch der Nodalstatus, die Anzahl der ausgetragenen Schwangerschaften und das Alter der Patientin bei Erstgeburt sowie die Information über den Tabakkonsum zeigen signifikante Ergebnisse. Allerdings ist zu beachten, dass die berechnete Korrelation mit einem Maximum um 0,2 (1=vollständige positive Korrelation) insgesamt sehr gering ist.

In den bisherigen Ergebnissen zur Betrachtung der klinischen Parameter in Abhängigkeit zu den Genexpressionsdaten fällt immer wieder die große Relevanz der Hormonrezeptoren auf. Diese werden daher in weiteren Analysen genauer untersucht.

4.5.3 Detaillierte Analysen zu den Hormonrezeptoren

Für die detaillierte Analyse des Östrogen- (ER) und Progesteronrezeptors (PGR) werden die Ergebnisse des immunhistochemischen Proteinnachweises (durchgeführt in den Städtischen Kliniken, vgl. Kapitel 3.1.4) und der Genexpressionsmessungen mittels qRT-PCR und Mikroarrayhybridisierungen verglichen. Die zugrunde liegenden Daten aus den Expressionsanalysen sind in Abschnitt 4.2.2 näher beschrieben. Zur Normalisierung der PCR-Daten wurden die ubiquitär exprimierte Gene HMBS und RPL7 verwendet (Kapitel 3.4.4.1). Da sie sich in den Tumorproben unterschiedlich verhalten, werden jeweils beide Varianten in die Ergebnisse aufgenommen. RPL7 ist in den untersuchten Proben in der Regel stärker exprimiert. Daher zeigt sich für alle Proben mit negativ regulierter Genexpression ein größerer Effekt bei Normalisierung mit RPL7. Im Gegensatz dazu ergeben sich bei Normalisierung mit HMBS tendenziell höhere Werte für hochregulierte Gene.

Für den Östrogenrezeptor kann, insbesondere bei den stark rezeptorpositiven und -negativen Tumorproben, eine sehr gute Übereinstimmung von Genexpressionsdaten und klinischen Parametern beobachtet werden (Abbildung 4.12). Nur zwei Proben (25-AAA-1_263_BP-0 und 34-B-1_1070_BP-1) weisen trotz stark positivem Proteinnachweis in PCR und Hybridisierung schwach positive Werte für das ESR1-Gen auf.

Alle Proben, für die im Proteinnachweis keine Östrogenrezeptoren nachgewiesen wurden, zeigen auch auf Transkriptionsebene keine Expression des für diesen Rezeptor kodierenden ESR1-Gens. Im intermediären Bereich (immunreaktiver Score IRS 2-8) können weniger eindeutige Aussagen getroffen werden. Hier zeigen nur zwei Proben (4-A_1011_BP-0 und 54-A_324_B-0) schwach positive Genexpression in den Mikroarrayexperimenten, alle anderen Proben zeigen keine bzw. eine negative Regulation des Gens.

Abbildung 4.13 zeigt die Ergebnisse aller Analysen für den Progesteronrezeptor, kodiert durch das Gen PGR. Grundsätzlich fällt hier auf, dass mittels Mikroarraytechnik nur sehr geringe Expressionsunterschiede detektiert wurden. Für Tumorproben mit stark positivem Rezeptorstatus (IRS > 8) kann

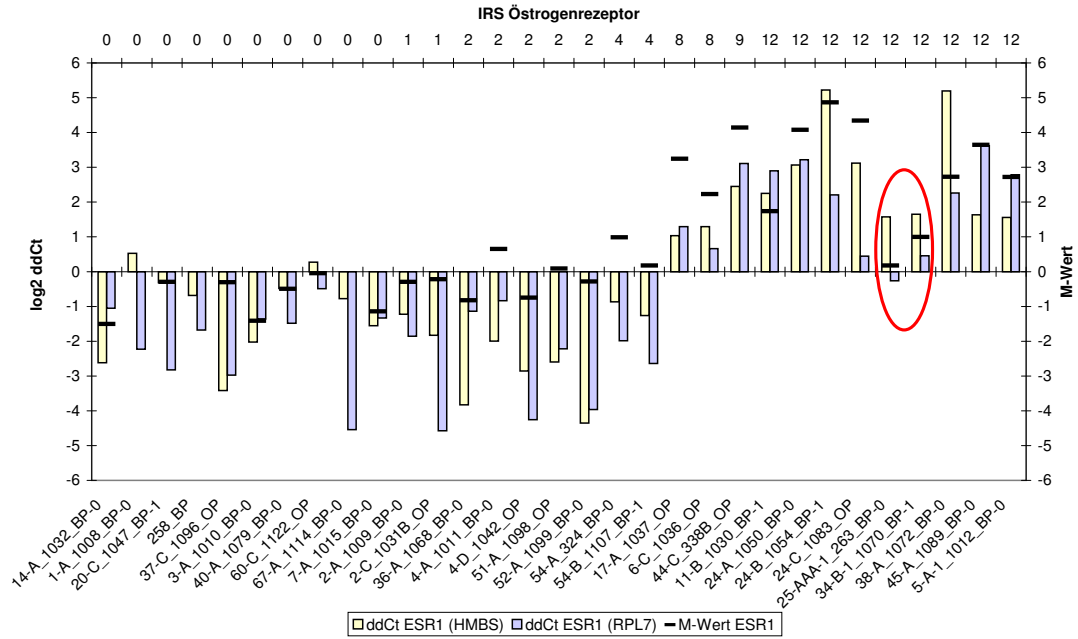


Abbildung 4.12: Dargestellt sind die Nachweisergebnisse des Östrogenrezeptors ER auf Protein- und Transkriptionsebene. Auf der y-Achse ist die Genexpression, ermittelt mit qRT-PCR (\log_2 ddCt) und der M-Wert aus den Mikroarrayhybridisierungen aufgetragen. Die x-Achse zeigt die Tumorproben nach aufsteigendem immunreaktivem Score (IRS) aus dem Proteinnachweis. Auffallende bzw. abweichende Werte sind durch rote Kreise hervorgehoben.

jedoch auch hier eine gute Übereinstimmung der Daten festgestellt werden. Im Proteinnachweis negative bzw. schwach positive Proben zeigen auf Transkriptionsebene teilweise Abweichungen von diesen Ergebnissen. Insgesamt können aufgrund der geringen Ausprägung der differentiellen Expression hier aber keine zuverlässigen Aussagen getroffen werden.

Die vergleichende Analyse von Protein- und Genexpression wird außerdem für den Wachstumsfaktorrezeptor Her2-neu, kodiert durch das Gen ERBB2, durchgeführt. Der Proteinnachweis von Her2-neu ist ebenfalls Standard in der klinischen Routine. Her2-neu zeigt, wie die beiden Hormonrezeptoren, in den vorangegangenen Untersuchungen eine positive Korrelation zwischen Genexpression und immunhistochemischen Proteinnachweis (siehe Korrelationsanalyse mit Kendall's τ , Abschnitt 4.5.1). Für neun der zehn immunhistochemisch Her2-neu-positiv (IRS=3) getesteten Tumorproben konnte das Ergebnis auch mittels qRT-PCR bestätigt werden. Eine Probe (4-D_1042_OP) zeigt auf Genexpressionsebene (in Mikroarrayhybridisierung und qRT-PCR) keine Übereinstimmung mit dem klinischen Ergebnis. Auffällig ist hier, dass eine weitere Probe dieser Patientin (4-A_1011_BP-0) in beiden Analysen eine eindeutige Überexpression des ERBB2-Gens aufweist (Überblick der Ergebnis-

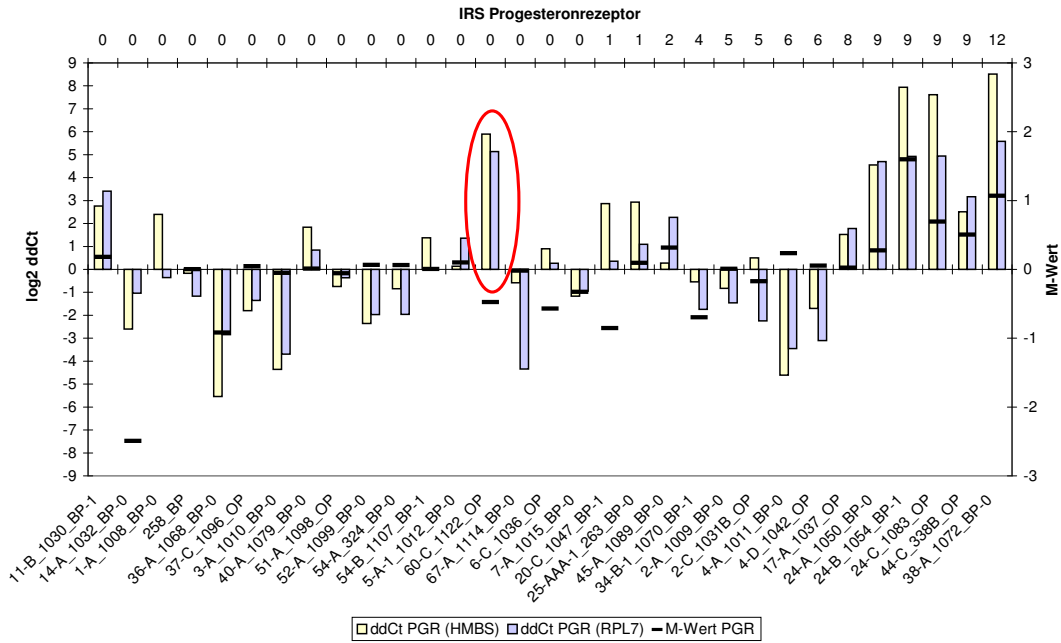


Abbildung 4.13: Dargestellt sind die Nachweisergebnisse des Progesteronrezeptors PGR auf Protein- und Transkriptionsebene. Auf der y-Achse ist die Genexpression, ermittelt mit qRT-PCR (\log_2 ddCt) und der M-Wert aus den Mikroarrayhybridisierungen aufgetragen. Die x-Achse zeigt die Tumorproben nach aufsteigendem immunreaktivem Score (IRS) aus dem Proteinnachweis. Auffallende bzw. abweichende Werte sind durch rote Kreise hervorgehoben.

se in Abbildung 4.14). Auch für den Her2-neu-Rezeptor kann bei Betrachtung der Tumorproben mit schwach positivem oder negativem Status keine Korrelation zur Genexpression gefunden werden. Die mittels Mikroarraytechnik bestimmte Genexpression ist auffällig, die Werte liegen durchgängig im stark positiven Bereich.

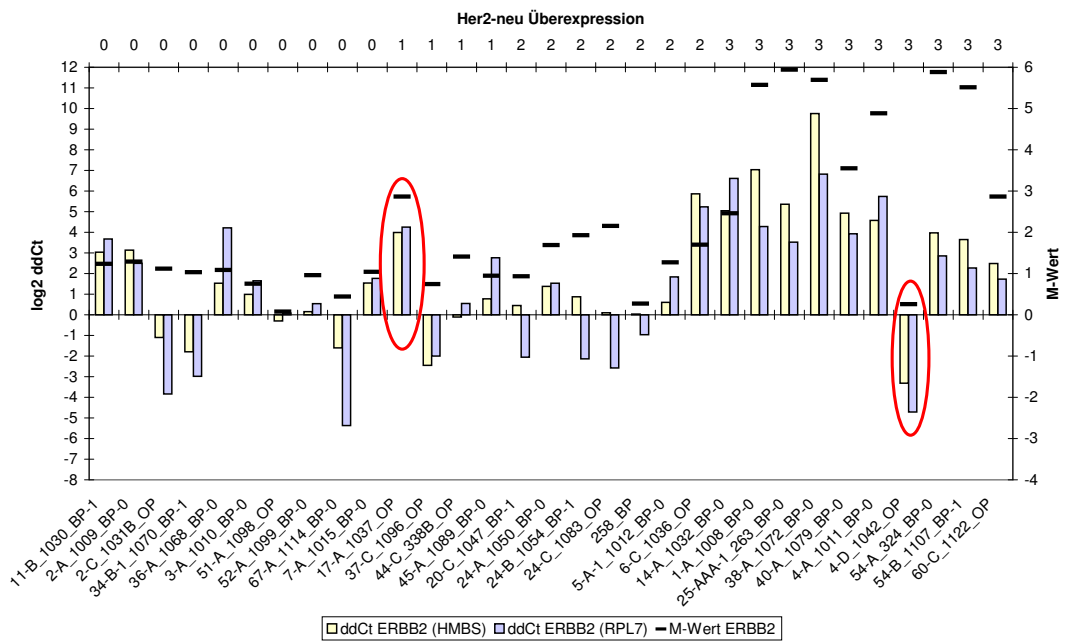


Abbildung 4.14: Dargestellt sind die Nachweisergebnisse des Wachstumsfaktorrezeptors Her2-neu auf Protein- und Transkriptionsebene. Auf den y-Achsen ist jeweils die Genexpression ermittelt mit qRT-PCR (\log_2 ddCt) und der M-Wert aus den Mikroarrayhybridisierungen aufgetragen. Die x-Achse zeigt die Tumorproben nach steigender Überexpression des Proteins. Abweichungen zwischen Gen- und Proteinexpression sind durch rote Kreise hervorgehoben.

DISKUSSION

Die Diskussion der Ergebnisse dieser Arbeit gliedert sich in drei Hauptteile. Zunächst werden in Abschnitt 5.1 die Durchführung der Studie und die benutzten Methoden kritisch beleuchtet. Im daran anschließenden Kapitel (5.2) werden die Ergebnisse der Genexpressionsanalysen zusammenfassend dargestellt und Zusammenhänge diskutiert. Kapitel 5.3 geht auf die Anwendung von Genexpressionsanalysen in der klinischen Routine ein und beschäftigt sich mit der Frage des zusätzlichen Nutzens im Vergleich zu den etablierten prädiktiven und prognostischen Faktoren für die Brustkrebserkrankung. Dieses Kapitel leitet zum Ausblick (Kapitel 6) über.

5.1 STUDIENDESIGN UND METHODEN

Die **Altersverteilung** der Patientinnen der hier vorgestellten Studie unterscheidet sich mit einem Durchschnittsalter von 50 Jahren deutlich vom mittleren Erkrankungsalter für Brustkrebs bei Frauen in Deutschland, das bei 63 Jahren liegt (RKI, 2008). Dieses niedrigere Durchschnittsalter ist durch das Studiendesign bedingt, da Patientinnen im neoadjuvanten Setting untersucht wurden. Dies beinhaltet, dass es sich in erster Linie um Hochrisiko-Patientinnen mit großen, oft rezeptornegativen Tumoren handelt (siehe Ergebnisse der St. Gallen Konsensus Konferenz 2005, Goldhirsch et al., 2005). Durch die neoadjuvante Chemotherapie sollen die Tumoren zunächst verkleinert und dadurch operierbar gemacht werden. Im höheren Alter auftretende Tumoren wachsen häufig langsamer, weshalb hier seltener (neo-)adjuvante Chemotherapien zum Einsatz kommen. Zudem sind Wirkung und besonders auch mögliche Nebenwirkungen von Chemotherapeutika bei älteren Patientinnen bisher nicht ausreichend erforscht (Loibl et al., 2008), auch wenn Studien zunehmend zeigen, dass die Wirksamkeit einer Therapie nicht vom Alter der Patientinnen abhängt (Bernardi et al., 2008).

Ein Ziel der Studie war es, zusätzlich zu den klassischen tumorbezogenen Daten (TNM-Status, Grading etc.) ein möglichst umfassendes Bild jeder Patientin zu bekommen. Zu diesem Zweck wurden weitere personenbezogene Daten (Herkunft, Beruf), Informationen zum Lebensstil (z.B. Alkoholkonsum, Rauchen, Schlaf- und Sportverhalten, *Body-Mass-Index* BMI) und den reproduktiven Faktoren (Anzahl der Geburten und Schwangerschaften, Menarche und Menopause etc.) erhoben (vgl. Kapitel 4.2.1 zur Charakterisierung des Patientinnenkollektivs). Aufgrund des insgesamt sehr geringen Datenumfanges und daraus resultierender fehlender (wie z.B. bei der Herkunft) oder zu stark variierender (z.B. Beruf) Daten wurden nur ausgewählte **patientinnenbezogene Parameter** in die Auswertung eingeschlossen. Insgesamt lassen sich die rein qualitativen verhaltensassoziierten Faktoren (z.B. Ernährungs- und Bewegungsverhalten) schwer kategorisieren, während die Berücksichtigung echter Messwerte (wie BMI und Alter) verlässlichere Aussagen zulässt. Auffällig ist hier zum Beispiel ein überdurchschnittlich häufig auftretender hoher bis sehr hoher BMI bei den Patientinnen dieser Studie. Dass ein erhöhter BMI mit einem erhöhten Lebenszeitrisko, an Brustkrebs zu erkranken, und auch mit einer schlechteren Prognose einhergeht, wurde bereits in verschiedenen internationalen Studien gezeigt (Renehan et al., 2008; Majed et al., 2009). Adipositas konnte als unabhängiger Prognosefaktor bestätigt werden. Zusätzlich wird hier auch ein Zusammenhang mit hormonellen Mechanismen vermutet (vgl. Majed et al., 2008). Ein weiterer signifikanter Zusammenhang der patientinnenbezogenen Daten mit der Genexpression kann nur für die Parameter ausgetragene Schwangerschaften und Tabakkonsum beobachtet werden, wobei die Ausprägung hier sehr gering ist (Kapitel 4.5.2 des Ergebnisteils). Reproduktive Faktoren spielen, zumeist im direkten Zusammenhang mit Östrogen, eine wichtige Rolle bei der Brustkrebserkrankung (GEKID, 2006). Bezüglich des Einflusses von Tabakkonsum auf das Erkrankungsrisiko liegen widersprüchliche Ergebnisse vor (vgl. auch Stewart und Kleihues, 2003). Aktuelle Studien gehen jedoch davon aus, dass keine kausalen Zusammenhänge bestehen (Ahern et al., 2009).

Der Einfluss der Chemotherapie auf die Genexpression in Tumorzellen konnte in dieser Studie nicht adäquat untersucht werden. Zum einen wurden unterschiedliche Therapieschemata angewandt, so dass keine Vergleichbarkeit über alle teilnehmenden Patientinnen gegeben war, zum anderen standen aufgrund einer Ausfallquote von 37 % bei der RNA-Isolierung (Kapitel 4.3) letztlich nur die Tumorproben von acht Patientinnen für einen solchen Vergleich zur Verfügung. Generell scheint aber der Expressionsunterschied in

Tumorzellen unterschiedlicher Patientinnen deutlich größer zu sein, als in Tumorproben einer Patientin, die zu unterschiedlichen Zeiten einer Therapie entnommen wurden. Das spiegelt die Clusterbildung der hier analysierten Gewebeproben deutlich wider (vgl. Kapitel 4.5) und wurde auch bereits in anderen Studien beobachtet (Perou et al., 2000).

Insgesamt ist wegen der Heterogenität der analysierten Tumoren der **Stichprobenumfang** der Studie mit 40 Patientinnen noch zu gering, um empirisch gesicherte Aussagen abzuleiten. Aufgrund des Ausscheidens von Patientinnen und des insgesamt sehr hohen Betreuungsaufwandes einer solchen Studie war es bei dem Patientinnenaufkommen der Städtischen Kliniken Bielefeld nicht möglich, eine größere Anzahl von Patientinnen im vorgesehenen Zeitraum zu untersuchen. Weitere Brustzentren wurden jedoch nicht in die Studie einbezogen, da die Herkunft des Tumormaterials eine zusätzliche Variable bedeutet hätte, die sich möglicherweise in den Analysen widerspiegelt.

Aufgrund der geringen Stabilität von RNA ist die **Probennahme** ein wichtiger Schritt der Analysen. Zunächst muss das sofortige Einfrieren des Tumormaterials gewährleistet sein, um ein Degradieren der RNA zu verhindern. Daher wurden die Gewebeproben direkt im Operationssaal bzw. Behandlungszimmer in flüssigen Stickstoff überführt. Zusätzlich wurde für die Bearbeitung der RNA (reverse Transkription in cDNA) von vornherein eine Methode gewählt, die auch für teilweise degradierte RNA geeignet ist (siehe Kapitel 3.4.3). Es wurden für alle Genexpressionsanalysen sowohl Operationspräparate als auch Stanzbiopsien verwendet. Ein generelles Problem bei der Analyse von Biopsien ist, dass der Tumoranteil der Gewebeprobe stark variieren kann. So weisen Cleator et al. (2006) darauf hin, dass der Anteil der den Tumor umgebenden Stroma-Zellen in einer Tumorprobe Einfluss auf das Genprofil haben kann und gerade weniger stark ausgeprägte Expressionsunterschiede so möglicherweise nicht detektiert werden. Eine Bestimmung des Anteils an Tumorzellen ist hier aus praktischen Gründen (Notwendigkeit des sofortigen Einfrierens der Proben) nicht möglich. Dass Stanzbiopsien jedoch grundsätzlich für Expressionsanalysen geeignet sind und in ihren Eigenschaften den ganzen Tumor repräsentieren, belegen Zanetti-Dällenbach et al. (2006) in ihren Untersuchungen.

Die Bearbeitung von Stanzbiopsien erfordert zusätzliche methodische Anpassungen, um zuverlässige Analysen durchführen zu können. Die aus Stanzbiopsien gewonnene RNA reicht häufig für die Genexpressionsanalysen nicht aus (vgl. auch Sotiriou et al., 2002; Rody et al., 2006), so dass in der

Regel eine Amplifikation des Materials oder, wie hier durchgeführt, des zu detektierenden Signals (siehe Abschnitt 3.4.3 des Methodenteils) erfolgen muss. Die aus den Tumorproben gewonnene RNA unterliegt häufig nicht nur Schwankungen bzgl. der Quantität, sondern auch der Qualität. Probengewicht oder Tumorbehandlung scheinen darauf zunächst keinen direkten Einfluss zu haben, obwohl in den vorliegenden Analysen die Tendenz beobachtet werden konnte, dass aus den Stanzbiopsien, die vor einer Chemotherapie entnommen wurden, die größten **RNA-Ausbeuten** erzielt werden (Kapitel 4.3). Weitere Faktoren, die die RNA-Gewinnung aus einer Tumorprobe beeinflussen können, sind z.B. die Zellzusammensetzung oder eingeschlossene Blutgefäße in der Gewebeprobe. Eine Bestimmung der Zelltypen ist jedoch, ebenso wie die Bestimmung des Tumoranteils, praktisch nicht durchführbar. Aus Gründen der Vergleichbarkeit wurde für Stanzbiopsien und Operationspräparate die gleiche Methode zur RNA-Isolierung verwendet. Die grundsätzlich schlechtere Ausbeute bei OP-Präparaten, trotz der größeren Gewebemenge, könnte darauf zurückzuführen sein, dass die Methodik auf die eigentlich schwieriger zu handhabenden Biopsien angepasst wurde. Ein Einfluss der Chemotherapie ist nicht unwahrscheinlich, was mit der Beobachtung übereinstimmen würde, dass das Tumorgewebe nach Chemotherapie häufig sehr weich und der Tumor, nach Angabe der behandelnden Ärztin, oft kaum noch zu ertasten ist. Das erschwert zusätzlich die Probenentnahme, so dass zum Teil auch weniger Material entnommen werden kann. Zudem greifen die während der Chemotherapie verabreichten Zytostatika erheblich in den Zellzyklus ein und nehmen somit auch Einfluss auf die RNA-Synthese der Zellen (vgl. auch Kapitel 2.2).

Für die Analyse der Expression der untersuchten Gene wird bei den Mikroarrayexperimenten die differentielle Expression im Tumorgewebe im Vergleich zu einer **Referenz** gemessen. Es wird eine kommerzielle humane Referenz-RNA eingesetzt, für die der Hersteller Stratagene eine konstante Qualität garantiert. Um weitere Varianzen auszuschließen, wurde ein großer Pool von cDNA der Referenz hergestellt, der für alle Experimente eingesetzt wurde. Auch bei Perou et al. (2000) findet eine Mischung von RNA aus verschiedenen Zelllinien als Referenz Verwendung, in späteren Studien (Hu et al., 2006) wird ebenfalls die benannte kommerzielle Referenz-RNA eingesetzt. Andere Studien hingegen verwenden als Referenz einen Pool der RNA aller untersuchten Tumorproben (vgl. z.B. van't Veer et al., 2002). Diese Variante eignet sich jedoch hauptsächlich für Studien mit einem sehr großen Stichpro-

benumfang und wurde daher für die vorliegende Studie nicht in Erwägung gezogen.

Für die Auswahl der zu analysierenden Gene wurde auf etablierte Expressionsstudien zurück gegriffen. Dabei sollte das **Genset** mit Hinblick auf die Anwendung in der klinischen Routine nicht zu groß gewählt werden, aber dennoch, durch die Auswahl funktionell interessanter Gene, im Rahmen dieser Studie Einblicke in die Tumorbiologie ermöglichen. Die schließlich analysierten 163 Gene (vgl. Kapitel 4.1) stellen einen guten Kompromiss zwischen der Berücksichtigung aktueller Expressionsstudien und einer sinnvollen Handhabbarkeit der Ergebnisse dar. Die Auswahl eines einzigen auf wenige Gene reduzierten und für alle Analysen optimalen Brustkrebs-Gensets ist dabei nicht möglich, wie auch Ein-Dor et al. (2005) und Ioannidis (2007) feststellen. Wichtiger ist es, einige in tumorrelevante Stoffwechselprozesse involvierte Gen-Vertreter zu berücksichtigen und so aussagekräftige Profile zu erhalten (Ioannidis, 2007).

Zu den Kontrollelementen des Themenarrays gehören, neben Negativkontrollen, auch nicht regulierte Referenzgene. Dabei handelt es sich um **Haushaltsgene**, die als ubiquitär exprimiert gelten. Von den hier verwendeten sechs Haushaltsgenen zeigten jedoch nur ACTB und KARS keine auffällige differentielle Regulation (Ergebnisse in Abschnitt 4.4). Haushaltsgene sind auch in anderen Studien bereits als differentiell exprimiert und daher als für die Normalisierung in Mikroarrayexperimenten nur bedingt geeignet aufgefallen. Lee et al. (2002) beobachteten in ihren Analysen zwölf gängiger Kontrollgene starke Variationen der Expression und konnten keine weiteren Säugetier-Gene identifizieren, die die Eigenschaften stabiler, gewebeunabhängiger Expression aufwiesen. Auch Tricarico et al. (2002) halten die Verwendung einzelner Haushaltsgene als Kontrollen in Studien mit Gewebe-Biopsien generell für schwierig. Bei der Analyse von zehn gängigen Haushaltsgenen in Biopsien aus normalem und malignem Brustgewebe konnten sie große Schwankungen der Expression zwischen verschiedenen Patientinnen, aber auch bei Proben eines Individuums, beobachten. Die hier analysierten Referenzgene wurden daher letztendlich nicht für eine Normalisierung verwendet.

Mit Hilfe von Mikroarrays können zahlreiche Analysen parallel und in einem gut handhabbaren Format durchgeführt werden. Gleichzeitig werden mit dieser Technik aber auch sehr große Datenmengen produziert. Von den Rohdaten bis zum Ergebnis eines Mikroarrayexperiments sind daher mehrere Schritte der Datentransformation und -reduzierung notwendig, die immer auch potentielle Fehlerquellen bergen (Owzar et al., 2008). Es existieren

zahlreiche Methoden zur Datenanalyse, die häufig auf ganz bestimmte Fragestellungen und Ausgangsdaten ausgerichtet sind. Wie umfangreich dieses Thema ist, schlägt sich auch in der verfügbaren Literatur nieder. So ist z.B. im Dezember 2002 bereits zum zweiten Mal eine komplette Ergänzungsausgabe von Nature Genetics erschienen, die sich ausschließlich mit der Mikroarray-technologie und ihren Möglichkeiten, Herausforderungen und Nachteilen beschäftigt (The Chipping Forecast II, Dezember 2002, Ausgabe 32 Nr.4s, Nature Genetics, URL: <http://www.nature.com/ng/journal/v32/n4s/index.html>).

Die Analyse der Mikroarraydaten erfolgte in dieser Arbeit nach Standardmethoden. Die **differentielle Genexpression** wurde mittels t-Test nach Student ermittelt. Dieser statistische Test setzt eine Normalverteilung der Daten voraus, ebenso ist der Stichprobenumfang von Bedeutung. Trotz der zahlreichen alternativen Methoden (vgl. z.B. Slonim, 2002) wurde hier auf den etablierten t-Test zurückgegriffen, da der zusätzliche Nutzen alternativer Methoden nicht gesichert ist: *„By evaluating all test methods it was clearly found that there is no prove for any advantage for any of them.[...] As a result, there can be one simple conclusion: to be very careful about untestified advantages of any new analysis method“* (Dondrup, 2007, Seite 170). Um systematische Abweichungen (z.B. Unterschiede in der RNA-Qualität, in der Markierungseffizienz oder Farbstoffunterschiede) auszugleichen und die einzelnen Experimente vergleichbar zu machen, wurde eine **Normalisierung** der Expressionsdaten mittels LOWESS vorgenommen (vgl. Kapitel 3.4.3.4), wie sie z.B. auch von Quackenbush (2002) empfohlen wird. Auch bei der quantitativen Real-Time-PCR (qRT-PCR) erfolgt vor der Bestimmung der differentiellen Expression der untersuchten Gene im Tumor im Vergleich zur Referenz eine Normalisierung der Expressionsdaten. Hierfür werden in diesem Fall Haushaltsgene herangezogen. Hier ergibt sich ebenfalls die Schwierigkeit der Wahl geeigneter ubiquitär exprimierter Gene. Daher wurde in den Analysen auf zwei verschiedene Gene (HMBS und RPL7) zurückgegriffen und die Ergebnisse verglichen. Wie in Kapitel 4.5.3 des Ergebnisteils dargestellt, zeigen die beiden Gene Unterschiede in ihrer Expression, da jedoch nur relative Veränderungen und keine absoluten Expressionswerte gemessen werden sollen, ist dieses Vorgehen für die Analysen ausreichend. Aufgrund der tendenziell stärkeren Expression von RPL7 ergibt sich für alle negativ regulierten Gene ein größerer Effekt bei Normalisierung mit RPL7 und für hochregulierte Gene bei HMBS, die Richtung der Expression (über- oder unterexprimiert im Vergleich zur Referenz) wird dadurch in der Regel nicht beeinflusst.

Zusätzlich zu den durch Mikroarray und qRT-PCR ermittelten Genexpressionsdaten spielt für die Auswertung auch die in den Städtischen Kliniken Bielefeld gemessene Proteinexpression der Östrogen-, Progesteron- und Her2-neu-Rezeptoren eine wichtige Rolle. Für den **immunhistochemischen Proteinnachweis** ergeben sich weitere potentielle Fehlerquellen, die es zu berücksichtigen gilt. Der immunreaktive Score (IRS) bzw. der Her2-neu-Status eines Tumorgewebes wird auf Basis einer Färbereaktion ermittelt. Es handelt sich also um eine subjektive Beurteilung, die je nach durchführender Person variieren kann. Auch dieses spricht für die Beschränkung auf eine Klinik als Studienort, um die Varianz möglichst gering zu halten (vgl. hierzu den Absatz zum Stichprobenumfang auf Seite 95). So ist in den Städtischen Kliniken Bielefeld die Bestimmung des IRS in der Regel auf maximal zwei Personen beschränkt. Um die Proteinexpressionsdaten ebenso wie die weiteren erhobenen patientinnen- und tumorbezogenen Daten untereinander, aber auch zu den Genexpressionsdaten in Beziehung setzen zu können, ist eine Vereinheitlichung der Datenformate notwendig. Es wurde daher eine **binäre Umwandlung** der Parameter auf Basis tumorbiologischer Grundlagen und Erkenntnisse bezüglich Risikofaktoren der Brustkrebserkrankung vorgenommen (zu den Umwandlungsparametern siehe Tabelle 4.11 in Kapitel 4.5). Grundsätzlich ist eine solche Datenumformung mit einem Informationsverlust gekoppelt. Es ist allerdings auch zu beachten, dass eine (gerichtete) vorangehende Datenreduktion den Informationsgehalt beobachteter Zusammenhänge deutlich vergrößern kann. Um möglichst eindeutige Kategorien zu schaffen, wurde hier mit der Umwandlung in ein binäres Datenformat gearbeitet. So soll eine Überinterpretation bzw. -differenzierung der Ergebnisse vermieden werden. Dem subjektiven Charakter der immunhistochemischen Nachweise wird damit begegnet, dass der Umformung zwei unterschiedliche Varianten zugrunde gelegt werden, um möglichst zuverlässige Ergebnisse zu erhalten. Zunächst wird nur nach eindeutig positiv getesteten Proben ausgewählt (mehr als 50 % positive Kerne und eine starke Färbereaktion, $IRS \geq 9$; vgl. auch van't Veer et al., 2002), alternativ wird nach eindeutig negativen Proben mit weniger als 10 % gefärbten Kernen und einer höchstens mäßigen Färbeintensität (IRS unter 3) gefiltert.

Die Anwendung von **Filterkriterien** zur Auswertung von Genexpressionsprofilen ist, neben Clusteranalysen, eine gängige Praxis (Quackenbush, 2002; Cho et al., 2004). Bei der großen Datenmenge die in Mikroarrayexperimenten generiert wird, sind Filterkriterien notwendig, um die Anzahl näher zu analysierender Gene einzuschränken und die Ergebnisse auswertbar zu

machen. Beschränken sich die Analysen darauf, z.B. nur Gene mit extrem ausgeprägter Expression zu identifizieren, ist gleichzeitig unvermeidbar, dass weitere interessante Gene mit den angelegten Filterkriterien evtl. nicht identifiziert werden. Auch wenn diese Methode Einschränkungen unterliegt, da nie alle potentiell interessanten Faktoren berücksichtigt werden, kann dennoch durch Anwendung verschiedener Filtersets (siehe hierzu Ergebnisteil Abschnitt 4.4.2 und 4.4.3) ein gutes Ergebnis erzielt werden. Dies wird in den vorliegenden Analysen durch Anwendung zusätzlicher Methoden, wie z.B. dem Normalized Tree Index (Kapitel 4.5.2) oder den Korrelationsanalysen (4.5.1), die diese Analysen bestätigen und ergänzen, deutlich.

Die Ergebnisse dieser Analysen werden im nachfolgenden Abschnitt zusammenfassend diskutiert.

5.2 MOLEKULARE CHARAKTERISIERUNG VON BRUSTTUMOREN

Die Analyse zur Charakterisierung der Tumoren gliederte sich in zwei Hauptteile, die auf die Fragestellung der Arbeit zur Unterscheidung von Brusttumoren anhand ihrer Genprofile abzielen. Im ersten Teil wurden einzelne Gene und Gengruppen des mittels Mikroarraytechnik gewonnenen Expressionsprofils analysiert und ihre biologische Funktion näher beleuchtet (vgl. Kapitel 4.4). Anschließend wurden diese Profile als Ganzes betrachtet und in Relation zu den klinischen (tumor- und patientinnenbezogenen) Parametern gesetzt (Kapitel 4.5). Damit wurde die Frage beantwortet, welche Gemeinsamkeiten die in den Analysen als ähnlich identifizierten Tumoren auf Expressionsebene haben, welche klinischen Eigenschaften hiermit möglicherweise im Zusammenhang stehen und wie sie sich von den anderen Tumoren unterscheiden. Die Ergebnisse werden in diesem Kapitel zusammenhängend interpretiert und diskutiert.

Zur Unterscheidung von Brusttumoren anhand ihrer Genprofile sind insbesondere die Gene von Interesse, die in ihrer Expression starke Variationen zwischen den untersuchten Tumoren aufweisen. Sie können Hinweise auf eine unterschiedliche Biologie der Tumoren geben. In allen durchgeführten Analysen fallen, sowohl auf Gen- als auch auf Proteinebene, der **Östrogen-(ESR1-Gen) und Progesteronrezeptor (PGR-Gen)** auf, deren klinische Relevanz unbestritten ist (vgl. Samaan et al., 1981; Goldhirsch et al., 2005; Decker et al., 2009, u.a.). In verschiedenen Genexpressionsanalysen wurde die Ausprägung des Östrogenrezeptors (ER) alpha bereits als ein Hauptkriterium

zur Unterscheidung von Profilen erkannt (Loi et al., 2005; Sotiriou et al., 2006; Sørlie et al., 2006, u.a.). Gerade auch zur Vorhersage des Ansprechens von Hormontherapien in Brustkrebspatientinnen konnten östrogenabhängige Gensignaturen identifiziert werden (Chanrion et al., 2008). Gruvberger et al. (2001) vertreten sogar die Meinung, dass es sich bei hormonrezeptornegativem und -positivem Brustkrebs um zwei unterschiedliche Erkrankungen handelt. Beide Hormonrezeptor-Gene zeigen in dieser Studie in den unterschiedlichen Tumorproben große Variationen in ihrer zum Teil sehr stark ausgeprägten Expression. Der Vergleich von Gen- und Proteinexpression (vgl. Abschnitt 4.5.3) liefert im eindeutig rezeptorpositiven Fall immer eine Übereinstimmung der Daten, auch der negative Rezeptorstatus kann für den Östrogenrezeptor in den Genexpressionsanalysen bestätigt werden. Im intermediären Bereich (immunreaktiver Score IRS 2-8) trifft dies für beide Rezeptoren jedoch nicht zu. Dies ist wahrscheinlich darauf zurückzuführen, dass eine nicht besonders stark ausgeprägte Expression in beiden Nachweisen (Gen- und Proteinebene) nicht immer eindeutig bestimmt werden kann. Dies ist bedingt durch die Subjektivität des Färbenachweises (Proteinnachweismethode, siehe auch Abschnitt 5.1), bzw. dadurch, dass in den Genexpressionsanalysen nicht die absolute, sondern die relative Expression im Vergleich zu einer Referenz bestimmt wird. Für den Progesteronrezeptor kann auch bei einem negativen Proteinnachweis keine eindeutige Aussage bezüglich der Übereinstimmung auf Genebene getroffen werden, da die in den Experimenten gemessene differentielle Expression des PGR-Gens generell sehr gering ausgeprägt ist.

Neben ESR1 gehört auch **EGFR (ERBB1)**, das für einen Vertreter der Familie der Wachstumsfaktorrezeptoren codiert, zu den Genen mit stark ausgeprägten Expressionsunterschieden zwischen den einzelnen Tumoren. Der Östrogenrezeptorstatus (Proteinebene) und die EGFR-Genexpression sind hier negativ korreliert (vgl. Kapitel 4.5.1 des Ergebnisteils), was sich mit den Ergebnissen verschiedener anderer Expressionsstudien deckt (Jeffrey et al., 2002). Auf Proteinebene ist der *Basal-like*-Tumortyp, der weitestgehend einem *Triple-negativ*-Tumortyp (ER^{-} , PGR^{-} , $Her2-neu^{-}$) entspricht und mit einer schlechteren Prognose einhergeht, mit einer EGFR-Überexpression und einem negativem Hormonrezeptorstatus assoziiert (Cheang et al., 2008).

Nahezu alle in dieser Studie als auffällig identifizierten Gene hängen funktional mit dem Östrogenrezeptor oder dem *Epidermal-Growth-Factor*-Rezeptor EGFR (bzw. beiden) zusammen. Die Expression dieser Gene und der mit ih-

nen assoziierten Faktoren ist damit das auffälligste Unterscheidungsmerkmal in den Profilen der hier analysierten Tumoren. Funktional stehen Steroid- und Wachstumsfaktor-Signalwege in enger Verbindung miteinander und nehmen sowohl Einfluss auf die Entwicklung, als auch auf das Therapieansprechen von Brusttumoren (Silva und Shupnik, 2007).

So sind z.B. die Produkte des ebenfalls als stark variierend identifizierten Gens **SPP1** (Osteopontin) und auch **SCUBE2** (*Signal Peptide-CUB-epidermal Growth Factor-like Domain-containing Protein 2*) potentielle Bindungspartner des EGF-Rezeptors. In den vorliegenden Analysen zeigt SCUBE2 bzgl. der Expression eine signifikante Korrelation mit ESR1, was frühere Ergebnisse im Rahmen dieser Studie bestätigt (vgl. Diplomarbeit Mentz, 2006). Cheng et al. (2009) zeigten für SCUBE2 auf Proteinebene signifikante Korrelationen der Expression mit dem Progesteronrezeptorstatus sowie dem krankheitsfreien Überleben der Patientinnen und gehen daher von einer besseren Prognose für diese Patientinnen aus. Die in den hier durchgeführten Analysen beobachtete signifikante negative Korrelation von SCUBE2-Expression und dem Tumorgrad würde ebenfalls für eine bessere Prognose bei SCUBE2-hochregulierten Tumoren sprechen. Osteopontin (OPN) ist in zahlreiche Stoffwechselwege involviert und somit an Prozessen der Angiogenese, Apoptose und Entzündungsprozessen beteiligt. Beispielsweise interagiert es neben EGFR mit weiteren membranständigen Rezeptoren (z.B. Integrinen), aber auch sezernierten Proteinen (wie Matrix-Metalloproteinasen, dem Urokinase Plasminogen Aktivator PLAU etc.), die mit Brustkrebs assoziiert werden (Tuck et al., 2007). SPP1 zeigt in seinem Expressionsverhalten signifikante Ähnlichkeit zu dem Gen MKI-67, das hier tendenziell herunterreguliert ist und für einen wichtigen Proliferationsmarker (Ki-67) kodiert. In Brustkrebspatientinnen ist eine Überexpression des OPN-Glykoproteins mit einer schlechten Prognose assoziiert. So zeigen Ribeiro-Silva und da Costa (2008) eine negative Korrelation von OPN zu den Hormonrezeptoren ER und PGR und positive Korrelation zum Her2-neu-Status, der mit einer schlechteren Prognose einhergeht. Wang et al. (2008) konnten eine signifikant höhere OPN-Proteinexpression in *Triple-negativ*-Brusttumoren und somit auch einen Zusammenhang zur Tumoraggressivität nachweisen.

ERBB2 codiert den Her2-neu-Rezeptor, der als etablierter prognostischer und prädiktiver Faktor einen hohen Stellenwert in der klinischen Routine einnimmt (Goldhirsch et al., 2005). Er bildet zusammen mit ERBB1, ERBB3 oder ERBB4 funktionelle Heterodimere und ist in ein komplexes Netzwerk

von Signalwegen eingebunden (Olayioye, 2001). In den hier durchgeführten Expressionsanalysen zeigt ERBB2 durchgängig eine starke Überexpression, es können keine signifikanten Unterschiede zwischen einzelnen Tumoren beobachtet werden. Allerdings zeigt auch die Korrelationsanalyse von Gen- und Proteinexpression für diesen Rezeptor nur eingeschränkt Übereinstimmung. Die Genexpressionsergebnisse aus PCR und Mikroarrayhybridisierung stimmen nur für Tumoren mit stark positivem Rezeptorstatus überein (siehe auch Kapitel 4.5.3). Bei Betrachtung der Tumorproben mit schwach positivem Her2-neu-Status konnte keine Korrelation der Genexpression gefunden werden. Möglicherweise ist dieses Gen in der verwendeten Referenz herunterreguliert, was im Mikroarrayexperiment zur Detektion einer generell positiven Expression in der Tumorprobe führt. Eine weitere Ursache könnte sein, dass das Gen auf dem Mikroarray nicht gut repräsentiert ist, z.B. die verwendete Oligonukleotid-Sequenz nicht eindeutig ist.

Generell ist zu beachten, dass mit der qRT-PCR relative Expressionsunterschiede im Vergleich zu einer Referenz aus verschiedenen Zelllinien und keine absoluten Unterschiede gemessen wurden. Gerade sehr gering ausgeprägte differentielle Genexpression kann daher falsch interpretiert bzw. teilweise nicht detektiert werden. Außerdem ist es möglich, dass bei den Fällen mit schwach positivem Rezeptorstatus (2+), die im FISH-Test negativ waren, die Gen- bzw. Proteinexpression gesteigert ist, ohne dass eine Amplifikation im Genom vorliegt. Ein dritter Vertreter dieser Gen-Familie, ERBB4, zeigt zwar keine auffällige Expression in den Analysen, ist jedoch signifikant mit der ESR1-Expression korreliert, und zwar auf Protein- und Genebene. Die Gene liegen in dem aus dem hierarchischen Clustern resultierenden Dendrogramm nah benachbart vor (Abbildung 4.9 A in Kapitel 4.4.3) und zeigen in den Korrelationsanalysen klinischer Parameter mit einzelnen Genen signifikante Zusammenhänge (Kapitel 4.5.1).

Ebenfalls im ESR1-Cluster befinden sich die Gene **BCL2** und **NAT2**. BCL2 codiert für einen Anti-Apoptose-Regulator und ist ein viel versprechender Kandidat zur Vorhersage der Chemosensitivität von Brusttumoren. Eine Überexpression ist mit dem Auftreten von Resistenzen (u.a. gegen das in der Brustkrebstherapie verwendete Anthrazyklin *Doxorubicin*) assoziiert worden (Sekine et al., 2009). Ein Zusammenhang von ESR1- und BCL2-Expression wurde bereits in früheren Analysen mit dem entwickelten Brustkrebsthemenarray beobachtet (Mentz, 2006). NAT2 ist in den Analysen auf Protein- und Genebene mit der ESR1-Expression korreliert und, ebenso wie das gleichfalls signifikant differentiell exprimierte Gen **GSTM1**,

in Entgiftungsprozesse der Zelle involviert. GSTM1 zeigt hier ebenfalls signifikante Ähnlichkeit zur ESR1-Expression, obwohl es in einem anderen Cluster lokalisiert ist. Für BCL2 und NAT2 sind Gen-Umwelt-Interaktionen mit Einfluss auf (Brust-)Krebserkrankungen bekannt: Polymorphismen dieser Gene konnten Firozi et al. (2002) signifikant mit dem Vorhandensein von aromatischen DNA-Addukten in Verbindung bringen, insbesondere unter Rauchern (vgl. auch Thier et al., 2003). Kovalent verknüpfte Komplexe aus DNA und aromatischen Verbindungen wie z.B. Benzo(a)pyren (u.a. aus Zigarettenrauch) stehen in Verdacht krebserregend zu sein (Veglia et al., 2003, 2008). In einem anderen Zusammenhang, der Untersuchung selbstberichteter Chemikalien-Empfindlichkeit (*Self-reported Chemical-related Sensitivity*), wurden vergleichbare Ergebnisse bzgl. des Auftretens von Polymorphismen dieser Gene erlangt (Schnakenberg et al., 2007). Ein weiteres mit dem Östrogenstoffwechsel assoziiertes und mit EGFR interagierendes Gen ist **MUC1** (Zaretsky et al., 2006; Pochampalli et al., 2007), das in den Analysen als stark differentiell exprimiert auffällt. Das Gen wird mit der Entwicklung von Chemoresistenzen in Verbindung gebracht. Pitroda et al. (2009) haben daher eine MUC1-induzierte 38 Gene umfassende Gensignatur zur Vorhersage des Überlebens von Tamoxifen-behandelten Brustkrebspatientinnen entwickelt. Dieses Genset enthält mit Cholesterol- und Fettsäuresynthese assoziierte Gene, darunter unter anderem auch das in der vorliegenden Arbeit als tendenziell hochreguliert identifizierte **FASN**, das durch MUC1 aktiviert wird. Die Überexpression von FASN ist in Brusttumoren (und anderen Karzinomen) mit einer schlechten Prognose assoziiert (Kuhajda, 2000; Menendez et al., 2004). Es ist ebenfalls eng mit der Östrogen-Signalkaskade verknüpft (Lupu und Menendez, 2006). Das Onkogen **AKT1** findet sich in den hier durchgeführten Analysen im selben Cluster wie die beiden Gene FASN und MUC1. Es wird durch Wachstumsfaktoren aktiviert und inhibiert die Apoptose (Ju et al., 2007), zeigt hier jedoch keine ausgeprägten Expressionseigenschaften.

Weitere in diesen Analysen auffällige und mit der Fettsäuresynthese assoziierte Gene sind der *Sterol Regulatory Element Binding Transcription Factor 1* (**SREBF1**) und der *Trefoil Factor 1* (**TFF1**). Die Proteinexpression des Progesteronrezeptors PGR und die SREBF1-Genexpression zeigen in dieser Arbeit eine positive Korrelation, mit den angelegten Filterkriterien zeigt das Gen sonst aber keine weiteren Auffälligkeiten. Die Transkription des TFF1-Gens wird durch Östrogen induziert (Carroll und Brown, 2006) und weist in dem untersuchten Gewebe eine variierende Expression auf. Die Expression von TFF1 ist stark mit der von SCUBE2 korreliert, das sich in einem Cluster mit

ESR1 befindet, sowie mit **CCND1**. Zyklin D1 (CCND1) und das *Insulin-like Growth Factor Binding Protein* kodierende **IGFBP5** sind weitere östrogenregulierte Gene (Yoshida et al., 2004). IGFBP5 weist in den Tumoren eine Überexpression auf, wohingegen Cyclin D1 stark unterschiedlich reguliert ist. CCND1 kodiert ein CDK4 regulierendes Zyklin. In Genexpressionsanalysen konnte eine starke Assoziation der CCND1-Expression mit dem Östrogenrezeptorstatus in Brustkrebs beobachtet werden (Gruvberger et al., 2001). Ein Inhibitor der zyklinabhängigen Kinase CDK4, das Tumorsuppressorgen CDKN2A, weist in den Brusttumoren ebenfalls eine stark unterschiedliche Regulation auf.

Durchgeführt wurden die Mikroarrayanalysen mit einer Kombination aus etablierter Gensignatur (70-Gen-Signatur nach van't Veer et al., 2002), ergänzt durch funktionell mit Brustkrebs assoziierte Gene (Kapitel 4.1). Insgesamt konnte die von van't Veer et al. (2002) beobachtete signifikante Expression der Gene nicht bestätigt werden. Die untersuchten cDNA-Sequenzen wiesen größtenteils keine auffällige Expression auf, und fünf der als signifikant hochreguliert identifizierten Gene sind in der van't Veer-Studie negativ reguliert. Unter den in der Mehrheit der Tumorproben stark hochregulierten Genen befinden sich zwei Gene aus dem van't Veer-Set (IGFBP5 und COL4A2), die auch dort eine Überexpression aufweisen. Unter den in ihrer Expression stark variierenden Genen ist nur das MUC1-Gen aus der 70-Gen-Signatur, dieses weist dort allerdings durchgängig eine starke Überexpression auf. Insgesamt zeigen die Ergebnisse daher kaum Übereinstimmung, was beispielsweise auf die unterschiedliche Auswahl des Patientinnenkollektivs und die Verwendung einer abweichenden Referenz zurückzuführen sein könnte. Die Analyse der weiteren brustkrebsassoziierten Gene lässt jedoch funktionelle Zusammenhänge erkennen.

Funktionell spielen bei den aus den Analysen resultierenden Genen **östrogenabhängige** (Creighton et al., 2006; Oh et al., 2006), **zellzyklusassoziierte** (Liu et al., 2008) und die **Chemosensitivität** bzw. -resistenz beeinflussende Gene (Sekine et al., 2009) die Hauptrollen. In diesem Zusammenhang ist allerdings zu beachten, dass diese Prozesse wiederum eng miteinander verknüpft sein können. Insbesondere die Östrogen-Signalwege sind sehr komplex und beeinflussen direkt und indirekt viele relevante Prozesse in der Zelle (Hall et al., 2001; Silva und Shupnik, 2007; Bourdeau et al., 2008, u.a.).

Die Tumoren der analysierten Brustkrebspatientinnen lassen sich auf Basis der gewonnenen Genprofile in fünf Tumor-Cluster aufteilen (Kapitel 4.5).

Die Tumoren aus drei dieser Cluster lassen sich anhand ihrer klinischen Eigenschaften eindeutig voneinander unterscheiden. Zwei Cluster (Cluster 2 und 5, vgl. Kapitel 4.5.1) lassen aufgrund der Eigenschaften der Tumoren (negativer Hormonrezeptorstatus, höhere EGFR-Genexpression etc.) eine schlechtere Prognose des Krankheitsverlaufs erwarten. Die Tumoren des dritten Clusters weisen hingegen Kennzeichen des *Luminal A*-Tumortyps auf, der mit einer besseren Prognose assoziiert wird. Dem widerspricht zunächst die Beobachtung, dass Gene der Fettsäuresynthese wie FASN, die mit einer schlechteren Prognose assoziiert werden, in diesem Cluster tendenziell höher reguliert sind. Sørli et al. (2006) konnten in ihren Analysen auf drei unabhängigen Mikroarrayplattformen jedoch ebenfalls eine höhere Expression von FASN in Tumoren des *Luminal A*-Typs nachweisen und führen dies darauf zurück, dass die gegenseitige Beeinflussung von Fettsäure- und Steroidhormon-Biosynthese auch Mechanismen der Zellproliferation und Apoptose in diesem Tumorsubtyp beeinflussen.

Auch in Bezug zu den vorhandenen personenbezogenen Daten lassen sich die Genexpressionsdaten analysieren. Grundsätzlich ist hier jedoch zu beachten, dass die untersuchten Fallzahlen zu gering sind, um für Brusttumoren allgemein gültige Aussagen abzuleiten. Zudem gibt es methodische Einschränkungen (vgl. Kapitel 5.1), die die Ergebnisse beeinflussen und eine mögliche Fehlinterpretation zu Folge haben könnten. So ergeben die Analysen eine Korrelation von Genprofil und rauchenden/nicht rauchenden Patientinnen. Zwar zeigen Studien keinen kausalen Zusammenhang zwischen dem Erkrankungsrisiko für Brustkrebs und Tabakkonsum (siehe Kapitel 2.1), auffällig ist jedoch, dass unter den hier als signifikant exprimiert identifizierten Genen auch solche zu finden sind, die Einfluss auf die Chemosensitivität der Zellen haben können (z.B. GSTM1, vgl. auch Kapitel 4.1.2). Ein weiterer Punkt ist die große Relevanz des Hormon-/Lipidstoffwechsels und der überdurchschnittlich hohe BMI der untersuchten Patientinnen. Campa et al. (2009) untersuchten in ihren Experimenten den Einfluss genetischer Variationen von Genen der Lipidsynthese. Sie konnten zwar keinen Zusammenhang zum Brustkrebsrisiko, aber dennoch eine Korrelation von BMI und bestimmten Polymorphismen des FASN-Gens aufzeigen. Die hier dargestellten Ergebnisse sind demnach nicht ursächlich zu interpretieren, können aber doch Hinweise auf wichtige Prozesse in der (Brust-)Tumorzelle geben.

Nachdem die Ergebnisse der vorliegenden Analysen im Detail diskutiert wurden, sollen im nächsten Abschnitt Vor- und Nachteile von Genexpressionsanalysen als Werkzeug für die Tumordiagnostik betrachtet werden.

5.3 GENEXPRESSIONSANALYSEN IN DER KLINISCHEN ANWENDUNG

Methoden der Genexpressionsanalyse finden in der (Brust-)Krebsforschung breite Anwendung. Mikroarray- und qRT-PCR-basierte Tests werden immer häufiger in der klinischen Diagnostik eingesetzt (*MammaPrint*[®] und *OncotypeDX*[®], siehe Kapitel 2.3). Dennoch sind sich Forscher und Kliniker in der Diskussion um den (zusätzlichen) Nutzen von Genprofilen für die Prognose des Krankheitsverlaufs (prognostisch) und der Vorhersage des Therapieansprechens (prädiktiv) längst nicht einig. Zum einen werden Genprofile als neue Chance zur Prognose (van't Veer et al., 2002; van de Vijver et al., 2002, und weitere) und als prädiktive Faktoren (Chang et al., 2003; Pawitan et al., 2005) gesehen, zum anderen besteht die Kritik, dass die relativ teure und aufwendige Technik keine Zusatzinformationen zu den bisher verwendeten Routine-Markern (TNM-Status, Hormonrezeptorstatus und Her2-neu-Status) bereit hält oder zumindest die bisherigen Studien nicht ausreichen, um zuverlässige Antworten hierüber zu finden (Edén et al., 2004; Reid et al., 2005; Reis-Filho et al., 2006, etc.). Die Argumente beider Seiten sollen im folgenden Abschnitt vor dem Hintergrund der hier vorgestellten Studie beleuchtet und abschließend diskutiert werden.

Ein Hauptkritikpunkt ist zunächst die geringe Studiengröße und die sich daraus ergebende mangelnde Reproduzierbarkeit bisheriger Studien. Michiels et al. (2005) analysierten sieben Studien und konnten in ihren Berechnungen größtenteils die Zuverlässigkeit der Vorhersage mit den dort identifizierten Gensets nicht bestätigen. Ihre Ergebnisse zeigten, dass die Selektion bestimmter Gene zur Prognose stark vom analysierten Patientinnenkollektiv abhängt und erst durch die Analyse deutlich größerer Trainingssets eine zuverlässigere Tumorklassifizierung erreicht werden kann. Auch Ioannidis (2005, S.455) stimmt in seinem Kommentar der Forderung nach deutlich größeren Stichproben zu und fordert zusätzlich jeweils eine Validierung der Daten durch mehrere völlig unabhängige Forschergruppen: *„However, we should aim for many independent studies with a total of several thousand patients, a hundred-fold more than the current standard. If we truly believe that microarrays and molecular research in general are important, we should not settle for less“*. Für die in der vorliegenden Arbeit durchgeführten Analysen wurde auf die etablierte Gensignatur von van't Veer et al. (2002) zurückgegriffen. Auch hier wichen die Beobachtungen stark von den publizierten Ergebnissen ab (siehe Abschnitt 5.2), was ihre Abhängigkeit vom Patientinnenkollektiv und der Methodik (Bezugsgrößen, Datenverarbeitung etc.) nicht nur im Hinblick auf

die Genselektion, sondern auch auf die gemessene Genexpression deutlich macht. In einer neueren Publikation gehen Ioannidis et al. (2009) noch einen Schritt weiter und bemängeln die grundsätzlich schlechte Nachvollziehbarkeit von Genexpressionsanalysen durch fehlende Datenverfügbarkeit sowie unvollständige Angaben zur Auswertung in den publizierten Ergebnissen. Sie fordern mehr Transparenz bzgl. Studiendesign und Datenverarbeitung für eine bessere Reproduzierbarkeit der Methoden.

Des Weiteren wird von vielen die große Anzahl verschiedener Gensets zur Brustkrebsprognose bzw. die geringe Schnittmenge, die aus den verschiedenen Expressionstudien resultiert, als problematisch angesehen (Liu et al., 2008). So zeigen Ein-Dor et al. (2005) am Beispiel der Amsterdamer Signatur (van't Veer et al., 2002), dass das resultierende Genset nicht eindeutig ist, sondern selbst aus der Analyse eines einzigen Datensatzes zahlreiche unterschiedliche Genlisten resultieren können. Mögliche Erklärungen hierfür sind zum einen die Komplexität der Erkrankung mit einer großen Anzahl involvierter Gene und die Auswahl der Gene eines Sets auf Basis eines relativ kleinen homogenen Patientinnenkollektivs. Zudem wird durch das schrittweise Vorgehen zur Identifizierung relevanter Gene, wie sie in den meisten Studien vorgenommen wird, das Genset bereits durch die Auswahl der ersten Gene (in der Regel die mit der größten Korrelation zum untersuchten Endpunkt) beeinflusst (Jenssen und Hovig, 2005). Um dieser Problematik zu begegnen, wurde in den hier durchgeführten Analysen nicht nur auf eine Gensignatur zurückgegriffen, sondern zusätzlich eine Literaturrecherche zur Auswahl weiterer Gene durchgeführt. Grundsätzlich wurde aber davon ausgegangen, dass die unterschiedlichen Gensets trotz ihrer geringen Schnittmenge ein ähnliches Potential zur Vorhersage der Prognose haben und die Betrachtung von Genprofilen im Vergleich zu einzelnen molekularen Markern auf Gen- oder Proteinebene einen Informationsgewinn bedeutet (Wirapati et al., 2008).

Es lässt sich schlussfolgern, dass Genprofile, nicht isoliert betrachtet, sondern vielmehr als Ergänzung zu den etablierten Prognosefaktoren, ein großes Potential für die Charakterisierung von Brusttumoren bereit halten. Daher fordert Ioannidis (2007) mehr Studien, die das Potential klassischer Faktoren mit dem von etablierten Faktoren in Kombination mit Genprofilen vergleichen. Sun et al. (2007) zeigen die parallele Berücksichtigung klinischer und genetischer Informationen als zukunftsweisenden Weg für die Brustkrebsforschung auf. Dabei ist weit weniger die Identifizierung des einen, optimalen Gensets, als vielmehr die Betrachtung mehrerer, in relevante Stoffwechsel-

prozesse involvierter Gene, von entscheidender Bedeutung (Ein-Dor et al., 2005; Ioannidis, 2007), wie auch in dieser Studie berücksichtigt.

Genprofile nicht nur zur Prognose sondern auch zur Vorhersage des Therapieansprechens nutzen zu können, ist ein weiterer wichtiger Schritt zur Verbesserung der Versorgung von Brustkrebspatientinnen. Hierfür sind zusätzliche Studien nötig, die neben der Information über die genbasierte Klassifizierung der Tumortypen auch Therapieschemata und Langzeit-Überlebensraten berücksichtigen (Mullan und Millikan, 2007). Cleator und Ashworth (2004) sind der Meinung, dass die Anwendung von Genprofilen als prädiktive Faktoren durch die Identifizierung von Patientinnen mit Tumoren, die nicht auf eine bestimmte Therapie ansprechen, das Potential zur Vermeidung von Überversorgung hat. Sie geben aber auch zu bedenken, dass zu einer wirklichen Verbesserung des Überlebens und der Lebensqualität der Patientinnen nicht nur die Anwendung individualisierter Therapien, sondern auch die Entwicklung neuer Chemotherapeutika mit deutlich geringeren Nebenwirkungen im Vordergrund steht.

AUSBLICK

Wie bereits in dieser Arbeit angesprochen wurde, ist die **Heterogenität** der Brusttumoren ein offensichtliches Problem für die Diagnose und Therapie des Mammakarzinoms. Genexpressionsanalysen haben sich daher als ein neuer Ansatz zur besseren Charakterisierung der Erkrankung in der Brustkrebsforschung etabliert und zu umfangreichen Wissensfortschritten geführt. So werden auf Basis von Expressionsprofilen fünf verschiedene Tumortypen (*Luminal A/B*, *Basal-like*, *Her2⁺/ER⁻*, *Normal-like*) unterschieden, die konkrete Hinweise auf die Prognose der Patientinnen geben können. Damit werden nicht nur bekannte klinische Unterschiede der Tumoren auf molekularer Ebene bestätigt, wie z.B. die Relevanz des Östrogenrezeptors, sondern darüber hinaus noch weitere Erkenntnisse erlangt, wie die Unterschiede zwischen den Subtypen hormonrezeptorpositiver Tumoren (*Luminal A* und *B*). Gerade die Analyse von Gen-Vertretern relevanter Zell-Stoffwechselwege kann weitere Einblicke in die noch längst nicht vollständig verstandene Biologie der Tumoren ermöglichen. Die Analyse auf Genebene ist dabei nur ein erster Schritt der molekularen Tumorcharakterisierung. Für umfangreichere Erkenntnisse der Tumorbiologie müssen auch Analysen auf weiteren Ebenen berücksichtigt werden. Neben den bisher erwähnten **Transcriptomic-Techniken** wie Mikroarrays und quantitativer Real-Time-PCR spielen daher weitere „omics“-Techniken wie **Genomics** (Methylierung, SNPS etc.), **Proteomics** (Proteinexpression, -interaktion) und **Metabolomics** (Enzymaktivität usw.) eine wichtige Rolle (für eine Übersicht verschiedener Methoden und ihrer Ansatzpunkte siehe Abbildung 6.1).

Die Analyse der transkribierten DNA in einer Tumorzelle gibt Aufschluss über die Regulation einzelner Gene. Wechselwirkungen mit anderen Genen, Veränderungen im Genom oder der Proteinaktivität bleiben mit reinen Transkriptionsanalysen jedoch unentdeckt. Wie bereits in dieser Arbeit angesprochen, sind die biologischen Prozesse, die zu Entstehung und

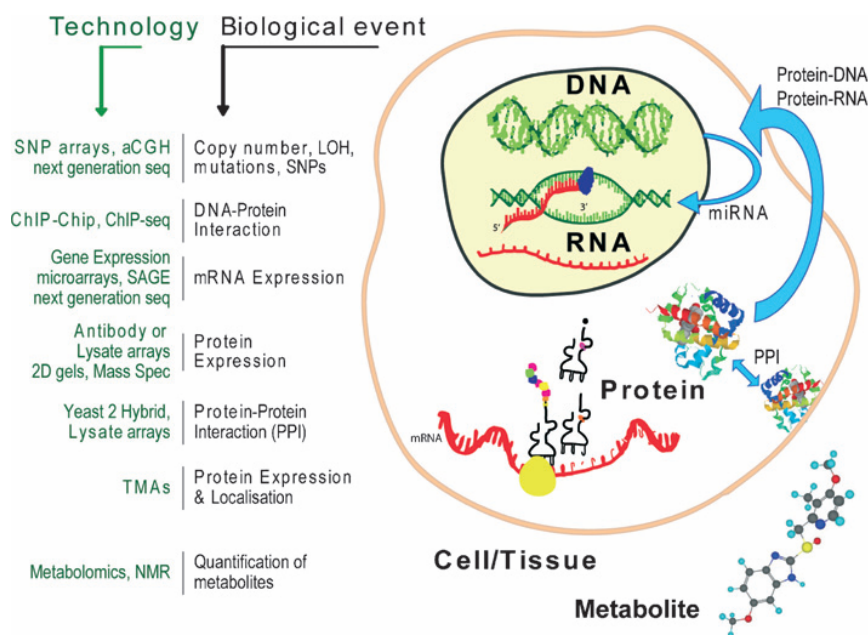


Abbildung 6.1: „omics“-Techniken zur Analyse biologischer Prozesse. Quelle: Culhane und Howlin (2007).

Fortschreiten eines Tumors beitragen, sehr komplex. Eine Reihe von Stoffwechselprozessen ist so eng mit der Tumorentstehung verknüpft, dass von „Krebs-Pathways“ gesprochen wird. Eine wichtige Rolle spielen hier Transkriptionsfaktoren wie z.B. das in vielen Tumoren veränderte p53 Protein. Transkriptionsfaktoren steuern die Aktivierung weiterer Gene u.a. durch direkte Anlagerung an spezifische Bindestellen in der DNA. Das Auffinden solcher Bindestellen im Genom kann z.B. durch die **ChIP-Chip-Technik**, einer Kombination aus Chromatin-Immunopräzipitation (ChIP) und klassischer Mikroarray-Technologie (Chip-Technik), erfolgen. Das zunächst *in vivo* an die DNA gebundene Protein wird mitsamt der DNA-Zielsequenz isoliert, das Protein anschließend entfernt (Chromatin-Immunopräzipitation) und die Zielsequenz mittels Mikroarrayhybridisierung analysiert.

Transkriptions- und Translationsprozesse in der Zelle können zusätzlich durch epigenetische Veränderungen, also Veränderungen, die nicht in der DNA-Sequenz festgelegt sind, wie z.B. Methylierung, beeinflusst werden. Untersuchungen mittels ChIP-Chip-Technik können hier neue Erkenntnisse bzgl. der Transkription zugrunde liegender Prozesse aufzeigen und bieten damit auch wichtige neue Anhaltspunkt für zielgerichtete Therapien für die Brustkrebserkrankung. Mit diesem Fokus hat sich auch die St. Gallen Konsensuskonferenz 2009 (Goldhirsch et al., 2009) mit den Ergebnissen solcher Analysen beschäftigt. So haben z.B. Lupien et al. (2008) die Rolle des Transkriptionsfaktors FOXA1 (dessen Aktivität durch Methylierungsprozesse

beeinflusst wird) im Steroidrezeptor-Cistrom in hormonabhängigen Tumoren untersucht (vgl. auch Lupien und Brown, 2009). Das „Cistrom“ (in Anlehnung an *cis*-Elemente, die DNA-Bindestellen der Transkriptionsfaktoren oder auch *trans*-Elemente) beinhaltet dabei alle miteinander verknüpften und in Wechselwirkung stehenden Faktoren eines bestimmten Stoffwechselweges.

Eine häufig beobachtete genomische Veränderung bestimmter Gene in Tumorzellen ist die Amplifikation (siehe z.B. ERBB2 in Brustkrebstumoren, Kapitel 2.2). Diese Amplifikation muss sich jedoch nicht zwangsläufig auch in einer gesteigerten Transkription des entsprechenden Gens äußern und kann daher auf RNA-Ebene nicht nachgewiesen werden. Genomics-Techniken wie die **vergleichende Genom-Hybridisierung (CGH)** ermöglichen hier Analysen auf DNA-Ebene, die ebenfalls im Chip-Format durchgeführt werden können (aCGH).

Eine bisher noch nicht erwähnte Anwendung der Transcriptomic-Technik, die großes Potential für die Brustkrebsforschung bietet, ist die Untersuchung von **microRNA**. Anders als bei der klassischen Genexpressionsanalyse wird hier die Expression nicht-kodierender microRNAs untersucht, kleine RNAs die die (posttranskriptionale) Genregulation beeinflussen. Im Zusammenhang mit Brustkrebs wurden bereits microRNAs identifiziert, deren Fehlen mit einer höheren Metastasierungspotenz der Tumoren einher zu gehen scheint (Goldhirsch et al., 2009).

In der Krebsforschung kommen aufgrund ihrer guten Eignung für die klinische Routine (hohe Durchsatzraten bei relativ geringen Kosten) die verschiedensten Chip- oder Array-Techniken zur Anwendung. Neben den bisher aufgeführten DNA-Mikroarrays sind für die Analyse auf Proteinebene auch **Gewebe-Mikroarrays** (*Tissue-Microarrays*; TMA) interessant. Diese ermöglichen die parallele immunhistochemische Analyse mehrerer Tumoren. Aber auch Proteinexpressionsanalysen mittels **Massenspektrometrie** an Blut- oder Serumproben werden zur Identifizierung neuer Tumormarker durchgeführt (Abramovitz und Leyland-Jones, 2006). Metabolomanalysen beschreiben nicht nur das Proteinmuster von Tumorzellen, sondern sämtliche Charakteristika eines Stoffwechselweges, wie z.B. die Aktivität der beteiligten Enzyme, entstehende Zwischenprodukte und mögliche Interaktionspartner. Methodisch kommen hier ebenfalls Massenspektrometrie und **Kernspinresonanzspektroskopie** (NMR) zur Anwendung. Untersuchungen an Blut- oder Serumproben haben grundsätzlich den Vorteil, dass das Probenmaterial nicht-invasiv gewonnen werden kann und so eine deutlich geringere Belastung für die Patientin entsteht.

Die hier vorgestellten Techniken zeigen nur einen Ausschnitt der aktuellen Entwicklung der molekularen Krebsforschung. Die Methoden dienen nicht nur zur Identifizierung neuer Tumormarker, sondern tragen zu einem besseren Verständnis der Tumorbiologie und der in der Tumorzelle stattfindenden Prozesse bei. Dadurch können gezielt neue Ansatzpunkte für Therapien gefunden oder bereits bestehende Therapeutika verbessert werden. Insgesamt lässt sich schlussfolgern, dass durch die **Kombination aller verfügbaren Techniken** auch bestmögliche Erkenntnisse erreicht werden könnten. Ein Zusammenführen der Ergebnisse aus Analysen mit unterschiedlichsten Methoden und Untersuchungsebenen könnte somit zur besseren Diagnose und gezielteren Therapie von Brustkrebspatientinnen beitragen. *„The ultimate goal is to be able to apply a systems biology approach to the information gleaned from the combination of these techniques [Genomics, Transcriptomics, Proteomics] in order to select the best treatment strategy, monitor its effectiveness and make changes as rapidly as possible where needed to achieve the optimal therapeutic results for the patient.“* (Abramovitz und Leyland-Jones, 2006). Die oftmals mangelnde Vergleichbarkeit verschiedener Studien und fehlende Standardisierung der Methoden (vgl. Diskussion der Genexpressionsanalysen für die klinische Anwendung in Kapitel 5.3) schränkt diese gewünschte Kombination von Techniken allerdings weiterhin ein.

Darüber hinaus ist auch zu beachten, dass nicht nur Prozesse innerhalb der Zelle das Krebsrisiko und die Tumorentwicklung beeinflussen, sondern auch **externe Faktoren** Einfluss auf diese Prozesse nehmen. Für wirklich umfassende Studien zur Brustkrebserkrankung wäre daher zusätzlich die Berücksichtigung möglichst vieler patientinnenbezogener Daten von Interesse. So könnten neben dem Alter der Patientinnen auch Informationen zu Herkunft, Beruf und Lebensstilfaktoren wie Ernährungs- und Bewegungsverhalten Aussagen darüber ermöglichen, warum sich ein Tumor in einer bestimmten Art und Weise (weiter-)entwickelt. Dieser Ansatz wurde in der vorliegenden Studie zwar grundsätzlich berücksichtigt, konnte jedoch nicht in vollem Umfang genutzt werden. Die Ergebnisse machen deutlich, dass dieser Vielzahl von Parametern mit sehr großem Stichprobenumfang und standardisierten Datenerhebungen begegnet werden muss. Hier besteht weiterer Forschungsbedarf, um diese oft qualitativ erhobenen verhaltens- und umgebungsassoziierten Faktoren zu vereinheitlichen und sie vor dem Hintergrund der Genexpressionsprofile systematisiert auswertbar zu machen.

Bisherige Studien haben gezeigt, dass Genexpressionsanalysen grundsätzlich das Potential haben, die Behandlung von Brustkrebspatientinnen hin zu ei-

ner **individualisierten**, stark patientinnenbezogenen **Therapie** zu verändern. Die Erwartungen sind so hoch, dass sich bereits klinische Anwendungen aus dieser Forschung ergeben haben, wie *MammaPrint*[®] und *OncotypeDX*[®]. Der auf Mikroarraytechnologie basierende Test *MammaPrint*[®] ermöglicht die Vorhersage des Metastasierungsrisikos nach einer Brustkrebserkrankung anhand eines Genprofils. *OncotypeDX*[®] misst die Expression bestimmter Gene mittels quantitativer Real-Time-PCR zur Vorhersage des Rückfallrisikos und des Therapieansprechens. Beide Anwendungen wurden bereits in Studien validiert (TAILORx, TRANSBIG). Eine Therapieentscheidung anhand dieser expressionsbasierten klinischen Tests erfolgt bisher aber weiterhin nur in Studien.

Für die Übertragbarkeit von Methoden zur Genexpressionsanalyse in die klinische Routine spielen nicht nur weitere, groß angelegte Validierungsstudien eine entscheidende Rolle, vielmehr muss auch die Patientin mit ihren Bedürfnissen in den Fokus der Aufmerksamkeit gerückt werden. Bei der Identifizierung spezifischer Gensignaturen für Mammatumoren handelt es sich zwar um einen Ansatz zur Optimierung der Versorgung von sich bereits in Behandlung befindender Krebspatientinnen, womit die Gefahr einer Überforderung der Patientin durch die Vorhersage einer möglicherweise schlechten Prognose geringer einzuschätzen ist, als aus einer „gesunden“ Ausgangssituation heraus. Dennoch können aus den nur eingeschränkten Handlungskonsequenzen, die diese Tests mit sich bringen, Unsicherheiten entstehen: Die Durchführung der Genexpressionsanalysen schürt bei den Betroffenen Hoffnungen auf eine bessere, individualisierte Therapie, die so (noch) nicht zur Verfügung steht.

Den notwendigen Prozess der Verknüpfung von molekularbiologischen Ergebnissen, der klinischen Praxis und Patientenbedürfnissen können gesundheitswissenschaftliche Ansätze vorantreiben. Auch wenn konkrete Umsetzungen noch weitestgehend fehlen, sind erste Bestrebungen in diese Richtung erkennbar, z.B. im Arbeitsfeld der Public Health Genomik, in dem genom-basiertes Wissen in die öffentlich koordinierte Gesundheitsprävention- und versorgung integriert werden soll (PHGEN, 2009). Speziell unter präventiven Aspekten wird die Bedeutung der fachdisziplinübergreifenden, ganzheitlichen Krebsforschung auch von der WHO betont: *„...genomics is likely to play an increasingly important role in cancer prevention, and its new technologies will undoubtedly become integrated into epidemiological and public health practice.“* WHO (2002, S.54).



GENINFORMATION

A.1 SEQUENZINFORMATION DER 70MER-OLIGONUKLEOTIDE

Reporter Name	Länge Gesamtsequenz (bp)	Entfernung 3'-Ende (bp)	Temperatur (°C)	Filter-Kriterien erfüllt	Oligo-Sequenz
R:McBiSet2_1	3737	2253	78	ja	ggggtctctcagatgtcactctggtcattcacctcttcagatgtcaccacatcaggacaagtgttaa
R:McBiSet2_2	12684	4561	78	ja	acacaaaaatctccctgcaaatctccccaccagaactaacagactgccacagacacaaagagatgcc
R:McBiSet2_3	-	-	78	ja	atcaggctcacactcaggaatcactgactctggcaagccactaaaatccctgcgaatctccccacta
R:McBiSet2_4	-	-	80	ja	agcctaagggtggcagaggagagcggaggcagaagagtgcaaggttctcatgcaaatcagaaggga
R:McBiSet2_5	9463	4149	80	ja	taaatgcatgtccctggatcaggccctgggtgacccaagagccatgtcccaagtgatgacatgaa
R:McBiSet2_6	2876	426	78	nein	cctaccaggacagaacttcccattacagggctgactcacagccgattggtgactcactcaatgtgt
R:McBiSet2_7	4772	572	76	nein	acagctgcattgaaaaagaggcgtctctccccctccacacaaagggtgattaaactctfvtg
R:McBiSet2_8	1703	509	77	nein	agagcaacatcctaattggctcagtcacgtttggcaatagtcaccgctgtctgtcttaatgcatgg
R:McBiSet2_9	5165	5014	86	ja	ttctgtgtctccccgccccctctfvgcgtctggccccgccccctctctcccccaacctctct
R:McBiSet2_10	4018	168	82	nein	gacctgagaaggacctggagctctgggaaattggagtgaccaaaagggtgctctgacacagcgag
R:McBiSet2_11	1431	378	78	nein	aagaccacttgagcagttgctggaaggctggcatttccatcatatagacctgccccctcagatgag
R:McBiSet2_12	3935	435	78	ja	ctggccgtttgttctgactctggtacacaggggactcccagggttaacaaatcagtgaaatgcc
R:McBiSet2_13	5000	3596	81	ja	tttctccctctggagacggggaggagaaaagggtgctccatgctcatgactgagctgaaggcaaa
R:McBiSet2_14	11941	1645	77	ja	gtccaaccacttgcctttactgtatatttccccaatgctgagatattccccagcctatagcc
R:McBiSet2_15	5830	799	78	ja	gagctgtgtgctgttctgactcctaagagatctccacagacctgactgtctctctgccaagc
R:McBiSet2_16	1912	162	78	nein	gaaatgaggtctctggccctgggtgccccttaagcctgtgacagttggcctgtgaaatattgtata
R:McBiSet2_17	3558	245	82	nein	cattgagctctctgggcccaaaagggtggtccatggtaccgggactgggagagtgagaccagt
R:McBiSet2_18	1973	482	74	nein	ttgaaatgaaaaacatggtgtgagtcacaaagggtgtttctgaaagactgtctatttctcagtc
R:McBiSet2_19	4099	70	77	ja	taacatggcggaggagggttaactggcaggcatggcaaggtgcaatgtaaaagtaacagctgtt
R:McBiSet2_20	2602	479	78	ja	atactggcacagaggtgctcactgctgactgagcaactgacacatgacccatgggtctaggacaca
R:McBiSet2_21	5128	1817	76	ja	cctgtggaagcagagaataactaagctgtgtctgccaatgaaatggaaatgcaatgaaactctgg
R:McBiSet2_22	4525	319	80	nein	tggctgggactgggaacagaactcgggtgcccccactctgctcttcttctgccaactctgctcca
R:McBiSet2_23	3116	2951	78	ja	ggggggaaaaatccgcacacacaaagcaacttggaaatccctgtggactaaattggccccctgag
R:McBiSet2_24	638	548	78	ja	ggtttattgtntgggaggggttggatggcacaatgacagccacagcctttaaattgggagccccgtgnc
R:McBiSet2_25	4828	3420	77	ja	ccagtgaaacacagatggcatttggaaacagcaaaaggctggatfttgaggccagcagcagatcttc
R:McBiSet2_26	2378	1640	78	ja	accctggagcttagggcattggcagagaggtgtggtggcagggatttaataatgattgtctgtat
R:McBiSet2_27	1548	405	83	ja	ataactggagccctcactcactgactgactgagactggctccagctggagggcagctgacactagct
R:McBiSet2_28	1474	392	78	nein	gccttgcagctctgagcactcttatctacaagagatgaggtaatccggagtgacaaatgagtggc
R:McBiSet2_29	1682	524	78	nein	ttgcatcagttatgctgagctgcttctttagcactctacacagctcaggcaaaagctggcctg
R:McBiSet2_30	1787	361	78	nein	aggttttttaaataagttgggtcaagtacaccagccactccagacacagctgctgctcccagctgctg
R:McBiSet2_31	2748	514	74	nein	aaggtcttctaaatctcctcattgtgagataaggacagtgctcaaaagtataaagcttaactgacc
R:McBiSet2_32	2281	528	78	nein	tgagtggtgttactctgtgtttaggggtatagatggggagtagatcttctagggagagacactggc
R:McBiSet2_33	2422	967	74	ja	gagaagcaatttggggaagggttgaattgtttcttaagatgtaatgctcttcagagacagctg
R:McBiSet2_34	6276	278	78	nein	caactcgggtaggtcattcttggccacatccctatagaccactgggttggcaaaactcaggcagaag
R:McBiSet2_35	2819	555	78	ja	ccccgttctaggacacatctattgcagagactattcacagccttctgtctgaccaaattggcca
R:McBiSet2_36	3399	461	77	ja	ccagtggtccccagggaagcacaagatggacacccctgtatatactatagattccatttagaggac
R:McBiSet2_37	5698	798	74	ja	agtgaccatctatggcattgttctctctgctgtgtgtttgagctgcttcttctgtctt
R:McBiSet2_38	7260	1866	78	ja	gtgaaagtattgccccctctgcaagacacataaagctcagatctgtagggggcaacacaaaagg
R:McBiSet2_39	496	403	78	nein	acacagcaccagtagcagcactgtgccacagagtaagcactcaataatgacgacggttactaagt
R:McBiSet2_40	1312	156	76	ja	gccgaagatgaaacttgcgaatgtatttgggtcccaatccccgggtaagtgaccatatactctac
R:McBiSet2_41	1412	324	78	nein	aaggaaagtattgactgggagagacacaaagaatggagttatccggggcagccccctctttaggtg
R:McBiSet2_42	4338	1449	73	ja	ggagttatgtaagtgaagtaaaagcaagctacttggagcagcttattacagatctactgtcagtg
R:McBiSet2_43	3635	358	78	nein	ttctagcggcttggcagatagcctacaagcttggcactcctgccccctgtgtgtgctgactgtt
R:McBiSet2_44	1443	70	76	ja	atgggtattgtcatgtaaggcttaacacagtggtgtggtgagttctgactaaaggttacctgtgtgta
R:McBiSet2_45	7940	590	78	nein	tgggaagacagccctctcaacagctcaactcagttctcttaaacctctgactatgggatgctca
R:McBiSet2_46	575	458	78	ja	ttcaactgctaaatcttctgagggctgtttgtgtgtgaccacgaaggccccctgtgtaggacagca

Reporter Name	Länge Gesamtsequenz (bp)	Entfernung 3'-Ende (bp)	Temperatur (° C)	Filter-Kriterium erfüllt	Oligo-Sequenz
R:McBiSet2_189	496	70	78	ja	ggctctggaggagtagttatcagttgaactgagttatccacagtggaaggtaaccctccctgcctag
R:McBiSet2_190	575	520	78	ja	tcttcaattggattcagcgtttctggtctctcacacactctctggtgggggattcaggtttgfgc
R:McBiSet2_191	443	228	76	ja	aatgggtgagaatgacactggaagtataaagtgtattcaaacagggttgaactgcataggaccgc
R:McBiSet2_192	525	354	78	ja	ccctcgaagcctctgtactcttttctaaftctctccagccactgcagatctagfctcaatgg
R:McBiSet2_193	439	385	78	nein	tctgacaltggactgagactctgccaacagctttggcatcggtctcttttctgtctgaggaa
R:McBiSet2_194	400	232	77	ja	ggcaagctttccacgtctctctttatccacaatccatgtacgagtagattccagtcagaactcc
R:McBiSet2_195	426	308	78	ja	aagtcctctgctccgcttttggggatccctctcggagcgagaagctactaaaatccaccat
R:McBiSet2_196	510	183	73	nein	attgcagtactatgtctttgagtgagtaacacaagaatfttccctcccttttggcaagttcac
R:McBiSet2_197	453	70	81	ja	tccagcaccgctcccaacggccccattactgtctctcttctccacagatagacagacaag
R:McBiSet2_198	1056	433	78	ja	acatggaaactgtggtgctctctctccaggtctctcatgtgccccaccactgcagttgataaaata
R:McBiSet2_199	561	421	78	nein	tgccctgatacacaaactccagcttctgtatcgcttaaaaagggggcatgggagcatgacc
R:McBiSet2_200	566	308	78	nein	taaaaagggtgctgagctctcagggaggtctcctaaagacctagcaagtctgagcaacagcatacc
R:McBiSet2_201	533	290	75	ja	caaggggaaaaatactcttagcaagatcaagaattcggcagcactgcaactctgttaccacag
R:McBiSet2_202	517	425	78	nein	gaaatcattgggtgaaactctggtggtgattctgtctgagaagacagcagctgcaacagcctcc
R:McBiSet2_203	580	70	77	ja	tcacaactgaggaacagctgctggttaccgattgtattctcttatatactgactgngtagcaggga
R:McBiSet2_204	566	379	78	nein	taatgcaaaaggagggggaaactgacacaggaacctgcccaagaactagctggcagagatgctagg
R:McBiSet2_205	392	70	82	ja	tggcccacagtaagctgctgctggcagtaagttagctctgaggtgtgacacccaagacacacagggc
R:McBiSet2_206	472	220	82	ja	agttgaccagctcttgaacagctgactccccagagctggggccaggggagcctgctgtgagggttaa

A.2 GENE AUS DER 70-GEN-SIGNATUR NACH VAN'T VEER

Reporter Name	Gen Name	Accession
R:McBiSet2_1	signal peptide, CUB domain, EGF-like 2;SCUBE2	NM_020974
R:McBiSet2_8	ubiquitin carboxyl-terminal hydrolase L5;UCHL5	NM_015984
R:McBiSet2_11	centromere protein A;CENPA	NM_001809
R:McBiSet2_16	small EDRK-rich factor 1A (telomeric);SERF1A	NM_021967
R:McBiSet2_31	cyclin E2;CCNE2	NM_057749
R:McBiSet2_34	collagen, type IV, alpha 2;COL4A2	NM_001846
R:McBiSet2_35	WNT1 inducible signaling pathway protein 1;WISP1	NM_003882
R:McBiSet2_49	BCL2 binding component 3;BBC3	NM_014417
R:McBiSet2_50	aldehyde dehydrogenase 4 family, member A1;ALDH4A1	NM_003748
R:McBiSet2_57	CDC42 binding protein kinase alpha (DMPK-like);CDC42BPA	NM_003607
R:McBiSet2_60	replication factor C (activator 1) 4, 37kDa;RFC4	NM_002916
R:McBiSet2_61	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor);FLT1	NM_002019
R:McBiSet2_63	origin recognition complex, subunit 6 like (yeast);ORC6L	NM_014321
R:McBiSet2_64	diaphanous homolog 3 (Drosophila);DIAPH3	NM_030932
R:McBiSet2_65	minichromosome maintenance deficient 6 homolog (S. cerevisiae);MCM6	NM_005915
R:McBiSet2_66	acyltransferase like 2;AYTL2	NM_024830
R:McBiSet2_78	HRAS-like suppressor;HRASLS	NM_020386
R:McBiSet2_82	glutathione S-transferase M3 (brain);GSTM3	NM_000849
R:McBiSet2_83	protein regulator of cytokinesis 1;PRC1	NM_003981
R:McBiSet2_87	pitriylsin metalloproteinase 1;PITRM1	NM_014889
R:McBiSet2_91	chromosome 16 open reading frame 61;C16orf61	NM_020188
R:McBiSet2_95	deoxycytidine kinase;DCK	NM_000788
R:McBiSet2_96	fibroblast growth factor 18;FGF18	NM_003862
R:McBiSet2_100	transforming growth factor, beta 3;TGFB3	NM_003239
R:McBiSet2_104	guanine nucleotide binding protein (G protein), alpha z polypeptide;GNAZ	NM_002073
R:McBiSet2_105	peroxisomal D3,D2-enoyl-CoA isomerase;PECI	NM_006117
R:McBiSet2_107	insulin-like growth factor binding protein 5;IGFBP5	NM_000599
R:McBiSet2_108	membrane-spanning 4-domains, subfamily A, member 7;MS4A7	NM_021201
R:McBiSet2_109	exostoses (multiple) 1;EXT1	NM_000127
R:McBiSet2_112	3-oxoacid CoA transferase 1;OXCT1	NM_000436
R:McBiSet2_116	TSPY-like 5;TSPYL5	NM_033512
R:McBiSet2_121	wi32g10.x1 NCL_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2392002 3'	cDNA_clone
R:McBiSet2_123	kinetochore associated 2;KNCT2	NM_006101
R:McBiSet2_126	guanine monophosphate synthetase;GMP5	NM_003875
R:McBiSet2_135	neuromedin U;NMU	NM_006681
R:McBiSet2_141	solute carrier family 2 (facilitated glucose transporter), member 3;SLC2A3	NM_006931
R:McBiSet2_142	nucleolar and spindle associated protein 1;NUSAP1	NM_016359
R:McBiSet2_147	exocyst complex component 7;EXOC7	NM_015219
R:McBiSet2_160	vascular endothelial growth factor A;VEGFA	NM_003376
R:McBiSet2_161	adaptor-related protein complex 2, beta 1 subunit;AP2B1	NM_001282
R:McBiSet2_170	G protein-coupled receptor 126;GPR126	NM_020455
R:McBiSet2_171	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase);MMP9	NM_004994
R:McBiSet2_174	denticleless homolog (Drosophila);DTL	NM_016448
R:McBiSet2_177	A kinase (PRKA) anchor protein 2;AKAP2	NM_001004065

Reporter Name	Gen Name	Accession
R:McBiSet2_180	endothelial cell-specific molecule 1;ESM1	NM_007036
R:McBiSet2_181	PREDICTED: Homo sapiens KIAA1442 protein (KIAA1442)	XM_044921
R:McBiSet2_182	chromosome 20 open reading frame 46;C20orf46	NM_018354
R:McBiSet2_183	serine/threonine kinase 32B;STK32B	NM_018401
R:McBiSet2_185	RAB6B, member RAS oncogene family;RAB6B	NM_016577
R:McBiSet2_188	gi 1405982 gb W90004.1 W90004 zh69b12.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:417311 3', mRNA sequence	cDNA_188_old
R:McBiSet2_189	gi 2059103 gb AA404325.1 AA404325 zw36g12.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:772198 3', mRNA sequence	cDNA_189_old
R:McBiSet2_190	gi 2270312 gb AA528243.1 AA528243 nh92b12.s1 NCL_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:965951 3', mRNA sequence	cDNA_190_old
R:McBiSet2_191	gi 2325568 gb AA555029.1 AA555029 nl16c07.s1 NCL_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1030476 3', mRNA sequence	cDNA_191_old
R:McBiSet2_192	gi 2908673 gb AA834945.1 AA834945 aj40a09.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1392760 3', mRNA sequence	cDNA_192_old
R:McBiSet2_193	gi 3179134 gb AA992378.1 AA992378 ot37g04.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1618998 3' similar to TR:O08808 O08808 P140MDIA.; mRNA sequence	cDNA_193_old
R:McBiSet2_194	gi 3807291 gb AI224578.1 AI224578 qw96b10.x1 NCL_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:1998907 3', mRNA sequence	cDNA_194_old
R:McBiSet2_195	gi 3921501 gb AI283268.1 AI283268 qk50h02.x1 NCL_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1872435 3', mRNA sequence	cDNA_195_old
R:McBiSet2_196	gi 4187271 gb AI377418.1 AI377418 tc35c11.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:2066612 3', mRNA sequence	cDNA_196_old
R:McBiSet2_197	gi 4486424 gb AI554061.1 AI554061 te49f03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2090045 3', mRNA sequence	cDNA_197_old
R:McBiSet2_198	gi 4569857 gb AI583960.1 AI583960 ts08g12.x1 NCL_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2228038 3', mRNA sequence	cDNA_198_old
R:McBiSet2_199	gi 4971660 gb AI694320.1 AI694320 wd45d09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2331089 3', mRNA sequence	cDNA_199_old
R:McBiSet2_200	gi 5109405 gb AI741117.1 AI741117 wg19b08.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2365527 3', mRNA sequence	cDNA_200_old
R:McBiSet2_201	gi 5424546 gb AI813331.1 AI813331 wj33e08.x1 NCL_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2404646 3', mRNA sequence	cDNA_201_old
R:McBiSet2_202	gi 5436816 gb AI817737.1 AI817737 wk25f07.x1 NCL_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2413381 3', mRNA sequence	cDNA_202_old
R:McBiSet2_203	gi 5637887 gb AI918032.1 AI918032 tt84f12.x1 NCL_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2248271 3', mRNA sequence	cDNA_203_old
R:McBiSet2_204	gi 5839063 gb AI992158.1 AI992158 wt75c08.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone IMAGE:2513294 3', mRNA sequence	cDNA_204_old
R:McBiSet2_205	gi 5863678 gb AW014921.1 AW014921 UI-H-BI0-aaf-e-11-0-UI.s1 NCL_CGAP_Sub1 Homo sapiens cDNA clone IMAGE:2709213 3', mRNA sequence	cDNA_205_old
R:McBiSet2_206	gi 5878414 gb AW024884.1 AW024884 wu92c03.x1 NCL_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:2527492 3', mRNA sequence	cDNA_206_old

A.3 GENAUWAHL DES THEMENARRAYS AUF BASIS VON LITERATUR-RECHERCHEN

Reporter Name	Gen Name	NCBI Accession	PANTHER (Biol. Process)	GeneOntology (Biol. Process)
R:McBiSet2_2	antigen identified by monoclonal antibody Ki-67;MKI67	NM_002417	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_3	antigen identified by monoclonal antibody Ki-67;MKI67	NM_002417	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_4	antigen identified by monoclonal antibody Ki-67;MKI67	NM_002417	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_5	Homo sapiens filamin B, beta (actin binding protein 278) (FLNB), Housekeeping gene	NM_001457	Cell structure and motility (BP00285)	
R:McBiSet2_6	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1;SERPINE1	NM_000602	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_7	retinoblastoma 1 (including osteosarcoma);RB1	NM_000321	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_9	insulin-like growth factor 2 (somatomedin A);IGF2	NM_000612	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_10	telomerase reverse transcriptase;TERT	NM_198253	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_12	amplified in breast cancer 1;unassigned	NM_022070	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)

Reporter Name	Gen Name	NCBI Accession	PANTHER (Biol. Process)	GeneOntology (Biol. Process)
R:McBiSet2_13	progesterone receptor;PGR	NM_000926	Neuronal activities (BP00166)	No match
R:McBiSet2_14	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian);ERBB4	NM_005235	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_15	kinase insert domain receptor (a type III receptor tyrosine kinase);KDR	NM_002253	Cell proliferation and differentiation (BP00224)	development (GO:0007275)
R:McBiSet2_17	matrix metalloproteinase 14 (membrane-inserted);MMP14	NM_004995	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_18	matrix metalloproteinase 3 (stromelysin 1, progelatinase);MMP3	NM_002422	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_19	histone deacetylase 6;HDAC6	NM_006044	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_20	cytochrome P450, family 1, subfamily A, polypeptide 1;CYP1A1	NM_000499	Electron transport (BP00076)	electron transport (GO:0006118)
R:McBiSet2_21	cytochrome P450, family 1, subfamily B, polypeptide 1;CYP1B1	NM_000104	Electron transport (BP00076)	electron transport (GO:0006118)
R:McBiSet2_22	L1 cell adhesion molecule;L1CAM	NM_000425	Cell adhesion (BP00124)	cell adhesion (GO:0007155)
R:McBiSet2_23	cytochrome P450, family 19, subfamily A, polypeptide 1;CYP19A1	NM_031226	Electron transport (BP00076)	electron transport (GO:0006118)
R:McBiSet2_25	cadherin 1, type 1, E-cadherin (epithelial);CDH1	NM_004360	Cell adhesion (BP00124)	cell adhesion (GO:0007155)
R:McBiSet2_26	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog);PDGFB	NM_033016	Cell proliferation and differentiation (BP00224)	PARENT: cell proliferation (GO:0030154); cell differentiation (GO:0008283)
R:McBiSet2_27	fibroblast growth factor 3 (murine mammary tumor virus integration site (v-int-2) oncogene homolog);FGF3	NM_005247	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_28	cyclin-dependent kinase 4;CDK4	NM_000075	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_29	cyclin A2;CCNA2	NM_001237	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_30	cyclin E1;CCNE1	NM_057182	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_32	cyclin-dependent kinase inhibitor 1A (p21, Cip1);CDKN1A	NM_078467	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_33	cyclin-dependent kinase inhibitor 1B (p27, Kip1);CDKN1B	NM_004064	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_36	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12);ITGB1	NM_033666	Cell adhesion (BP00124)	cell adhesion (GO:0007155)
R:McBiSet2_37	topoisomerase (DNA) II alpha 170kDa;TOP2A	NM_001067	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_38	insulin-like growth factor 1 (somatomedin C);IGF1	NM_000618	Cell proliferation and differentiation (BP00224)	development (GO:0007275)
R:McBiSet2_40	mitogen-activated protein kinase 14;MAPK14	NM_139013	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_41	mitogen-activated protein kinase 8;MAPK8	NM_139049	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_42	angiopoietin 1;ANGPT1	NM_001146	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_43	F-box protein 31;FBXO31	NM_024735	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_44	kallikrein-related peptidase 10;KLK10	NM_145888	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_45	nuclear receptor co-repressor 1;NCOR1	NM_006311	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_47	glutathione S-transferase M1;GSTM1	NM_000561	Immunity and defense (BP00148)	defense response (GO:0006952)
R:McBiSet2_51	tumor necrosis factor (TNF superfamily, member 2);TNF	NM_000594	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_52	tumor necrosis factor, alpha-induced protein 6;TNFAIP6	NM_007115	Signal transduction (BP00102)	signal transduction (GO:0007165)
R:McBiSet2_54	phosphatidic acid phosphatase type 2A;PPAP2A	NM_176895	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_55	sulfotransferase family 1E, estrogen-preferring, member 1;SULT1E1	NM_005420	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_56	transmembrane protein with EGF-like and two follistatin-like domains 1;TMEFF1	NM_003692	Oncogenesis (BP00281)	No match
R:McBiSet2_58	metadherin;MTDH	NM_178812	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)

Reporter Name	Gen Name	NCBI Accession	PANTHER (Biol. Process)	GeneOntology (Biol. Process)
R:McBiSet2_62	nuclear receptor coactivator 3;NCOA3	NM_006534	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_67	SMAD family member 4;SMAD4	NM_005359	Cell proliferation and differentiation (BP00224)	development (GO:0007275)
R:McBiSet2_68	IMP (inosine monophosphate) dehydrogenase 1;IMPDH1	NM_183243	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_69	BCL2-associated X protein;BAX	NM_004324	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_70	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (p105);NFKB1	NM_003998	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_71	Homo sapiens non-POU domain containing, octamer-binding (NONO), Housekeeping gene	NM_007363	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	
R:McBiSet2_72	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4;NFATC4	NM_004554	Developmental processes (BP00193)	defense response (GO:0006952)
R:McBiSet2_73	Homo sapiens beta-2-microglobulin (B2M), Housekeeping gene	NM_004048	Immunity and defense (BP00148)	
R:McBiSet2_75	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1);SPP1	NM_000582	Cell adhesion (BP00124)	cell adhesion (GO:0007155)
R:McBiSet2_76	aurora kinase A;AURKA	NM_198433	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_77	citrate synthase;CS	NM_004077	Amino acid metabolism (BP00013)	amino acid metabolism (GO:0006520)
R:McBiSet2_79	acetyl-Coenzyme A carboxylase alpha;ACACA	NM_198837	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_81	vascular endothelial growth factor B;VEGFB	NM_003377	Cell proliferation and differentiation (BP00224)	development (GO:0007275)
R:McBiSet2_84	mannose receptor, C type 2;MRC2	NM_006039	Intracellular protein traffic (BP00125)	intracellular protein transport (GO:0006886)
R:McBiSet2_85	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian);EGFR	NM_201283	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_86	RB1-inducible coiled-coil 1;RB1CC1	NM_014781	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_88	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian);ETS1	NM_005238	Cell proliferation and differentiation (BP00224)	development (GO:0007275)
R:McBiSet2_89	fatty acid synthase;FASN	NM_004104	Amino acid metabolism (BP00013)	amino acid metabolism (GO:0006520)
R:McBiSet2_92	ATP-binding cassette, sub-family B (MDR/TAP), member 1;ABCB1	NM_000927	Transport (BP00141)	transport (GO:0006810)
R:McBiSet2_93	thyroid hormone responsive (SPOT14 homolog, rat);THRSP	NM_003251	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_97	hydroxysteroid (17-beta) dehydrogenase 1;HSD17B1	NM_000413	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_98	hydroxysteroid (17-beta) dehydrogenase 2;HSD17B2	NM_002153	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_99	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase);PTGS2	NM_000963	Cell proliferation and differentiation (BP00224)	defense response (GO:0006952)
R:McBiSet2_101	transketolase (Wernicke-Korsakoff syndrome);TKT	NM_001064	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_102	angiopoietin 2;ANGPT2	NM_001147	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_103	N-acetyltransferase 2 (arylamine N-acetyltransferase);NAT2	NM_000015	Other metabolism (BP00289)	PARENT: metabolism (GO:0008152)
R:McBiSet2_110	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4);CDKN2A	NM_000077	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_111	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase;CAD	NM_004341	Amino acid metabolism (BP00013)	amino acid metabolism (GO:0006520)
R:McBiSet2_113	nitric oxide synthase 3 (endothelial cell);NOS3	NM_000603	Electron transport (BP00076)	electron transport (GO:0006118)
R:McBiSet2_114	hyaluronan synthase 2;HAS2	NM_005328	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_115	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in);TFF1	NM_003225	Cell proliferation and differentiation (BP00224)	PARENT: cell proliferation (GO:0030154); cell differentiation (GO:0008283)

Reporter Name	Gen Name	NCBI Accession	PANTHER (Biol. Process)	GeneOntology (Biol. Process)
R:McBiSet2_118	Homo sapiens actin, beta (ACTB), Housekeeping gene	NM_001101	Intracellular protein traffic (BP00125), Cell structure and motility (BP00285), Cell cycle (BP00203), Transport (BP00141)	
R:McBiSet2_119	Homo sapiens lysyl-tRNA synthetase (KARS), Housekeeping gene	NM_005548	Protein metabolism and modification (BP00060)	
R:McBiSet2_120	phosphofructokinase, liver;PFKL	NM_002626	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_124	sterol regulatory element binding transcription factor 1;SREBF1	NM_004176	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_125	plasminogen activator, urokinase;PLAU	NM_002658	Cell structure and motility (BP00285)	defense response (GO:0006952)
R:McBiSet2_127	steroid sulfatase (microsomal), arylsulfatase C, isozyme S;STS	NM_000351	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_129	ADAM metallopeptidase domain 9 (meltrin gamma);ADAM9	NM_003816	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_131	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian);ERBB2	NM_004448	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_132	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian);ERBB3	NM_001005915	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_133	erb2 interacting protein;ERBB2IP	NM_018695	Immunity and defense (BP00148)	defense response (GO:0006952)
R:McBiSet2_136	transaldolase 1;TALDO1	NM_006755	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_137	matrix metallopeptidase 11 (stromelysin 3);MMP11	NM_005940	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_143	baculoviral IAP repeat-containing 5 (survivin);BIRC5	NM_001012270	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_144	platelet-derived growth factor receptor, alpha polypeptide;PDGFRA	NM_006206	Cell proliferation and differentiation (BP00224)	defense response (GO:0006952)
R:McBiSet2_145	Homo sapiens peroxisome proliferative activated receptor, alpha (PPARA)	NM_032644		
R:McBiSet2_146	v-akt murine thymoma viral oncogene homolog 1;AKT1	NM_005163	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_148	insulin-like growth factor binding protein 3;IGFBP3	NM_000598	Homeostasis (BP00267)	cell homeostasis (GO:0019725)
R:McBiSet2_149	insulin-like growth factor binding protein 4;IGFBP4	NM_001552	Homeostasis (BP00267)	cell homeostasis (GO:0019725)
R:McBiSet2_150	ATP-binding cassette, sub-family G (WHITE), member 2;ABCG2	NM_004827	Transport (BP00141)	transport (GO:0006810)
R:McBiSet2_151	estrogen receptor 1;ESR1	NM_000125	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_152	peroxisome proliferator-activated receptor gamma;PPARG	NM_015869	Carbohydrate metabolism (BP00001)	carbohydrate metabolism (GO:0005975)
R:McBiSet2_153	breast cancer 1, early onset;BRCA1	NM_007294	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_154	catechol-O-methyltransferase;COMT	NM_007310	Neuronal activities (BP00166)	PARENT: metabolism (GO:0008152)
R:McBiSet2_155	SERPINE1 mRNA binding protein 1;SERBP1	NM_015640	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_156	Homo sapiens hydroxymethylbilane synthase (HMBS), Housekeeping gene	NM_001024382	Coenzyme and prosthetic group metabolism (BP00081)	
R:McBiSet2_157	platelet-derived growth factor receptor, beta polypeptide;PDGFRB	NM_002609	Cell proliferation and differentiation (BP00224)	defense response (GO:0006952)
R:McBiSet2_158	dihydroorotate dehydrogenase;DHODH	NM_001361	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)
R:McBiSet2_159	ribosomal protein SA;RPSA	NM_002295	Biological process unclassified (BP00216)	biological_process unknown (GO:0000004)
R:McBiSet2_162	v-myc myelocytomatosis viral oncogene homolog (avian);MYC	NM_002467	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_163	B-cell CLL/lymphoma 2;BCL2	NM_000633	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_164	ADAM metallopeptidase domain 12 (meltrin alpha);ADAM12	NM_003474	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_165	ADAM metallopeptidase domain 17 (tumor necrosis factor, alpha, converting enzyme);ADAM17	NM_003183	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_166	phosphatase and tensin homolog (mutated in multiple advanced cancers 1);PTEN	NM_000314	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_167	matrix metallopeptidase 3 (stromelysin 1, progelatinase);MMP3	NM_002422	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)

Reporter Name	Gen Name	NCBI Accession	PANTHER (Biol. Process)	GeneOntology (Biol. Process)
R:McBiSet2_168	sex hormone-binding globulin,SHBG	NM_001040	Lipid, fatty acid and steroid metabolism (BP00019)	lipid metabolism (GO:0006629)
R:McBiSet2_169	TIMP metalloproteinase inhibitor 1;TIMP1	NM_003254	Developmental processes (BP00193)	development (GO:0007275)
R:McBiSet2_172	matrix metalloproteinase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase);MMP2	NM_004530	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_173	TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory);TIMP3	NM_000362	Protein metabolism and modification (BP00060)	protein metabolism (GO:0019538)
R:McBiSet2_175	cyclin D1;CCND1	NM_053056	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_176	platelet-derived growth factor alpha polypeptide;PDGFA	NM_033023	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_178	tumor protein p53 (Li-Fraumeni syndrome);TP53	NM_000546	Apoptosis (BP00179)	apoptosis (GO:0006915)
R:McBiSet2_179	breast cancer 2, early onset;BRCA2	NM_000059	Nucleoside, nucleotide and nucleic acid metabolism (BP00031)	No match
R:McBiSet2_184	estrogen receptor 2 (ER beta);ESR2	NM_001437	Cell cycle (BP00203)	cell cycle (GO:0007049)
R:McBiSet2_186	ATP-binding cassette, sub-family C (CFTR/MRP), member 1;ABCC1	NM_004996	Immunity and defense (BP00148)	defense response (GO:0006952)
R:McBiSet2_187	Homo sapiens mucin 1, cell surface associated (MUC1)	NM_001457		

PATIENTINNENDATEN/KLINISCHE PARAMETER

EMMA-Value	Alter	Meno	BMI	Histo	T	N	M	Grade	PGR	ER	Her2	FISH	Grav	Para	Erstgeburt	Stillen	Rauchen	Alkohol	Sport	Chemo	fam. Brustkrebs
11-A_1025_BP-0	41	nein	27	dukktal	2	0	0	2	0	12	0	0	2	2	21	ja	ja	nein	0	AC/DOC	ja
11-B_1030_BP-1	41	nein	27	dukktal	2	0	0	2	0	12	0	0	2	2	21	ja	ja	nein	0	AC/DOC	ja
11-C_1048_OP	41	nein	27	dukktal	2	0	0	2	0	12	0	0	2	2	21	ja	ja	nein	0	AC/DOC	ja
12-A-1_1027_BP-1	52	nein	27										2	2	23	ja	ja	ja	0		nein
14-A_1032_BP-0	57	ja	21	dukktal	2	0	0	2	0	0	3	0	2	2	30	nein	ja	ja	2	TAC	nein
14-B_1033_BP-1	57	ja	21	dukktal	2	0	0	2	0	0	3	0	2	2	30	nein	ja	ja	2	TAC	nein
17-A_1037_OP	37	nein	27	dukktal	2	0	0	2	8	8	1	0	0	0			nein			FEC	nein
20-A_1043_BP-0	65	50	22	lobulär	2	1	0	3	1	0	2	0	1	1	27	nein	ja	ja	4	AC/DOC	ja
20-B_1044_BP-0	65	50	22	lobulär	2	1	0	3	1	0	2	0	1	1	27	nein	ja	ja	4	AC/DOC	ja
20-C_1047_BP-1	65	50	22	lobulär	2	1	0	3	1	0	2	0	1	1	27	nein	ja	ja	4	AC/DOC	ja
20-D_1080B_OP	65	50	22	lobulär	2	1	0	3	1	0	2	0	1	1	27	nein	ja	ja	4	AC/DOC	ja

EMMA-Value	Alter	Meno	BMI	Histo	T	N	M	Grade	PGR	ER	Her2	FISH	Grav	Para	Erstgeburt	Stillen	Rauchen	Alkohol	Sport	Chemo	fam. Brustkrebs
23-B_1057_BP-1	48	nein	26	lobulär	2	1	0	2	12	12	1	1	1	1	28	ja	nein	ja	0	AC DOC	nein
24-A_1050_BP-0	70	50	26	duktal	4	0	0	2	9	12	2	0	2	2	30	ja	nein	nein	7	FEC	ja
24-B_1054_BP-1	70	50	26	duktal	4	0	0	2	9	12	2	0	2	2	30	ja	nein	nein	7	FEC	ja
24-C_1083_OP	70	50	26	duktal	4	0	0	2	9	12	2	0	2	2	30	ja	nein	nein	7	FEC	ja
29-A_1058_BP-0	58	55	22	duktal	2	0	0	3	9	9	1	0	2	2	27	ja	nein	ja	5	TAC	nein
29-B-1_1064_BP-1	58	55	22	duktal	2	0	0	3	9	9	1	0	2	2	27	ja	nein	ja	5	TAC	nein
29-C_1086_OP	58	55	22	duktal	2	0	0	3	9	9	1	0	2	2	27	ja	nein	ja	5	TAC	nein
2-A_1009_BP-0	51	ja	22	duktal	2	0	0	3	5	1	0	0	0	0			nein	nein	0	TAC	nein
2-B_1017_BP-1	51	ja	22	duktal	2	0	0	3	5	1	0	0	0	0			nein	nein	0	TAC	nein
2-C_1031B_OP	51	ja	22	duktal	2	0	0	3	5	1	0	0	0	0			nein	nein	0	TAC	nein
30-A_1059_BP-1	52	51	38	duktal	2	0	0	3	12	12	2	1	2	2	20	nein	nein	nein	1	AC DOC	ja
32-B_1081AB_OP	49	nein	28	duktal	2	0	0	2	12	4	3	7	3	3	21	nein	nein	ja	0	GEPAR4	nein
34-A_1066_BP-0	56	49	27	duktal	2	1	0	2	4	12	0	1	1	1	30	ja	ja	ja	0	AC DOC	nein
34-B-1_1070_BP-1	56	49	27	duktal	2	1	0	2	4	12	0	1	1	1	30	ja	ja	ja	0	AC DOC	nein
34-B-2_1071_BP-1	56	49	27	duktal	2	1	0	2	4	12	0	1	1	1	30	ja	ja	ja	0	AC DOC	nein
35-A_1067_BP-0	42	nein	32	duktal	2	1	0	2	6	12	1	4	4	4	22	ja	nein	nein	0	AC DOC	nein
36-A_1068_BP-0	38	nein	39	duktal	2	1	0	3	0	2	0	0	3	3	25	ja	ja	ja	0	AC DOC	nein
36-B_1091B_OP	38	nein	39	duktal	2	1	0	3	0	2	0	0	3	3	25	ja	ja	ja	0	AC DOC	nein
37-B_1076_BP-1	29	nein	23	duktal	2	0	0	3	0	0	1	0	0	0			nein	ja	3	AC DOC	nein
37-C_1096_OP	29	nein	23	duktal	2	0	0	3	0	0	1	0	0	0			nein	ja	3	AC DOC	nein
389_BP-0	74	ja		duktal	2	0	0	3	2	2	1	4	4	4		ja				AC DOC	ja
38-A_1072_BP-0	39	nein	31	lobulär	2	1	0	3	12	12	3	2	2	2	22	nein	ja	ja	0	AC DOC	nein
38-A-2_296_BP-0	39	nein	31	lobulär	2	1	0	3	12	12	3	2	2	2	22	nein	ja	ja	0	AC DOC	nein
38-B_1075_BP-1	39	nein	31	lobulär	2	1	0	3	12	12	3	2	2	2	22	nein	ja	ja	0	AC DOC	nein
39-A_1073_BP-0	41	nein	25	duktal	1	0	0	2	0	0	0	3	3	3	23	ja	nein	ja	1	AC DOC	nein
39-C_1100B_OP	41	nein	25	duktal	1	0	0	2	0	0	0	3	3	3	23	ja	nein	ja	1	AC DOC	nein
3-A_1010_BP-0	48	47	29	duktal	2	1	0	3	0	0	0	0	6	2	23	ja	nein	ja	0	TAC	nein

EMMA-Value	Alter	Meno	BMI	Histo	T	N	M	Grade	PGR	ER	Her2	FISH	Grav	Para	Erstgeburt	Stillen	Rauchen	Alkohol	Sport	Chemo	fam. Brustkrebs
3-B_1020_BP-1	48	47	29	duktal	2	1	0	3	0	0	0	0	6	2	23	ja	nein	ja	0	TAC	nein
40-A_1079_BP-0	51	nein	40	duktal	2	1	0	3	0	0	3	3	3	3	21	ja	ja	ja	0	AC DOC	nein
44-B_1088_BP-1	38	nein	23	duktal	2	1	0	3	9	9	1	3	2	2	32	ja	nein	ja	1	AC DOC	nein
44-C_338B_OP	38	nein	23	duktal	2	1	0	3	9	9	1	3	2	2	32	ja	nein	ja	1	AC DOC	nein
4-A_1011_BP-0	66	ja	39	duktal	2	0	0	3	6	2	3	2	2	2	26	ja	nein	ja	1	GEPAR4	nein
4-B_1026_BP-1	66	ja	39	duktal	2	0	0	3	6	2	3	2	2	2	26	ja	nein	ja	1	GEPAR4	nein
4-D_1042_OP	66	ja	39	duktal	2	0	0	3	6	2	3	2	2	2	26	ja	nein	ja	1	GEPAR4	nein
52-A_1099_BP-0	32	nein	20	duktal	2	1		3	0	2	0	0	0	0			nein	ja	2	AC DOC	nein
53-A_322_BP-0	63	50	21	duktal	2	1	0	2	4	12	0	2	2	2	27	nein	ja	ja	7	AC DOC	ja
54-A_324_BP-0	60	42	25	duktal	2	1		2	0	4	3	4	2	2	19	ja	ja	ja	0	AC DOC	nein
54-B_1107_BP-1	60	42	25	duktal	2	1		2	0	4	3	4	2	2	19	ja	ja	ja	0	AC DOC	nein
58-A_1105_BP-0	39	nein	23	duktal	1	0		3	6	6	3	0	2	2	35						nein
59-A_1106_BP-0	40	nein	26	duktal	2		0	2	12	12	3	3	3	3	34	ja	nein	nein	0	AC DOC	nein
5-A-1_1012_BP-0	57	50	31	duktal	3	1	0		0	12	2	0	0	0			ja	ja	2	TAC	nein
5-A-2_1013_BP-0	57	50	31	duktal	1	0	0	2	0	12	2	0	0	0			ja	ja	2	TAC	nein
5-B-2_1022_BP-1	57	50	31														ja	ja	2		nein
5-C-1_1034B_OP	57	50	31	duktal	1	0	0	2	0	12	2	0	0	0			ja	ja	2	TAC	nein
5-C-2_1035_OP	57	50	31	duktal	3	1	0		0	12	2	0	0	0			ja	ja	2	TAC	nein
60-C_1122_OP	64	46	32	duktal	4		0	3	0	0	3	1	1	1	22	ja	nein	nein	0	AC DOC	nein
64-A_1111_OP	61			duktal	2	2		3	0	0	0										
64-AA_315_BP-0	61			duktal	2	2		3	0	0	0										
67-A_1114_BP-0	40	nein	27	duktal	2	1	0	3	0	0	0	1	1	1	38	ja	ja	ja	0	AC DOC	nein
69-A_393_BP-0	36	nein	37	duktal	2	1		3	6	12	0	1	1	1	24	ja	nein	ja	2	AC DOC	nein
69-B-1_399_BP-1	36	nein	37	duktal	2	1		3	6	12	0	1	1	1	24	ja	nein	ja	2	AC DOC	nein
6-B_1023_BP-1	49	nein	27		3	1	0	2	0	8	2	0	1	1	33	nein	nein	ja	0	TAC	ja
6-C_1036_OP	49	nein	27		3	1	0	2	0	8	2	0	1	1	33	nein	nein	ja	0	TAC	ja
70-A_390_BP-0	45	nein	25	duktal	2	1		2	9	2	2	0	0	0			ja	ja	0	AC DOC	nein

EMMA-Value	Alter	Meno	BMI	Histo	T	N	M	Grade	PGR	ER	Her2	FISH	Grav	Para	Erstgeburt	Stillen	Rauchen	Alkohol	Sport	Chemo	fam. Brustkrebs
72-A_1120_OP	61	57	30	duktal	3	1	3	3	0	0	0		3	3	17	ja	ja	nein	0	AC/DOC	nein
72-AA_367_BP-0	61	57	30	duktal	3	1	3	3	0	0	0		3	3	17	ja	ja	nein	0	AC/DOC	nein
73-A_358_BP-0	43	nein	23	duktal	2	1	2	2	8	12	3		3	2			ja	ja	0		nein
7-A_1015_BP-0	44	nein	33	duktal	2	1	3	3	0	0	0		2	2	26	nein	ja	nein	7	AC/DOC	nein



RNA-ISOLIERUNG: QUALITÄTS- UND QUANTITÄTSBESTIMMUNG DER GESAMT-RNA

Proben-nr.	Eppi-ID	ID-intern	Proben-art	Datum	Gewicht (mg)	NanoDrop		Bioanalyzer			Anzahl Hybs	Real-Time	in Studie
						ng/ μ l	260/230	ng/ μ l	28S/18S	RIN			
257	257	low-risk1		26.03.2008	60	246.18	1.91	278	3.1	6.9	2	nein	nein
258	258		0	01.07.2008	56	551.07	1.43	347	1.8	6.9	2	ja	nein
263	263	25-AAA-1	0	26.03.2008	90	428.65	2.09	496	0.6	3.9	2	ja	nein
264	264	25-AAA-2	0	26.03.2008	90	55.44	0.36	24	0	3.6	0	nein	nein
276	276		2	31.10.2008	124	39.08	1.15				0	nein	ja
281	281_A		2	22.07.2008	717	23.90	0.28				0	nein	ja
281	281_B		2	22.07.2008	717	31.16	0.32				0	nein	ja
281	281_C		2	22.07.2008	717	172.98	1.68				0	nein	ja
281	281_D		2	22.07.2008	717	94.41	1.86				0	nein	ja
281	281_PoolAB		2	22.07.2008	717	30.15	0.33				0	nein	ja
281	281_PoolCD		2	22.07.2008	717	121.65	1.77	148	2.1	8.2	0	nein	ja
282	282_A		2	22.07.2008	345	53.76	0.17				0	nein	ja
282	282_B		2	22.07.2008	345	52.67	0.53				0	nein	ja
282	282_C		2	22.07.2008	345	36.92	1.52				0	nein	ja
282	282_Pool		2	22.07.2008	345	46.75	0.3	41	1.8	7.8	0	nein	ja
285	285	25-AA	1	26.05.2008	107	8.44	0.54	7	0	NaN	0	nein	ja
290	290	27-AA	0	18.09.2007	67	192.60	1.45	222	0.7	3.8	2	nein	ja
296	296	38-A-2	0	27.03.2008	60	279.38	1.33	449	0.9	5.8	2	nein	ja
298	298	51-AA	0	26.05.2008	105	95.43	0.9	85	1.3	5.9	1	nein	ja
299	299_A		2	27.03.2008	520	59.83	0.69				0	nein	nein
299	299_B		2	27.03.2008	520	124.35	1.73				0	nein	nein
299	299_C		2	27.03.2008	520	31.97	1.01				0	nein	nein
299	299_Pool		2	27.03.2008	520	85.15	1.16	101	2.3	7.7	0	nein	nein
300	300	low-risk2	0	26.03.2008	50	173.63	1.38	174	1.3	2.9	2	nein	nein
315	315	64-AA	0	26.05.2008	78	152.48	1.95	246	1.5	6.2	1	nein	ja
322	322	53-A	0	18.09.2007	35	122.15	1.17	99	2.2	7.4	1	nein	ja
323	323	34-C	2	19.09.2007	39	7.35	1.22	9	0	NaN	0	nein	ja
324	324	54-A	0	28.11.2007	34	1500.48	2.11	1407	1.7	7	2	ja	ja
326	326	55-A-1	0	26.05.2008	39	81.92	1.42	74	1.8	4.7	1	nein	ja
329	329	low-risk3		26.03.2008	60	170.16	0.77	179	0.9	4.2	2	nein	nein
338	338_A	44-C	2	01.07.2008	144	352.99	1.57	469	0.7	5.4	2	nein	ja
338	338_B	44-C	2	01.07.2008	144	507.10	2.03	367	0.3	4.8	2	ja	ja
339	339	low-risk4		26.03.2008	50	188.72	1.55	192	0.8	4.9	2	nein	nein
340	340	60-A	0	28.11.2007	40	6.25	0.63	14	0	1.1	0	nein	ja
342	342		0	01.07.2008	56	4.75	0.07				0	nein	ja
343	343		0	01.07.2008	84	14.30	0.56				0	nein	ja
358	358	73-A	0	01.07.2008	17	113.80	1.16	165	1.9	6.8	1	nein	ja
366	366	60-B	1	27.03.2008	30	2.46	0.09	1	0	NaN	0	nein	ja
367	367	72-AA	0	26.05.2008	44	205.85	0.76	254	1.1	4.8	2	nein	ja
368	368	63-A	0	27.03.2008	36	139.36	1.99	489	1.2	3.9	1	nein	nein
370	370		0	01.07.2008	43	58.85	0.4				0	nein	nein
375	375		0	01.07.2008	34	18.66	0.6				0	nein	ja
386	386		0	01.07.2008	20	25.78	0.54				0	nein	ja
389	389		0	26.05.2008	35	144.32	1.38	184	1.6	6.4	2	nein	ja
390	390	70-A	0	01.07.2008	50	224.87	1.91	345	2	6.1	2	nein	ja
393	393	69-A	0	01.07.2008	35	208.40	1.92	234	0.9	4.5	2	nein	ja
399	399	69-B-1	1	26.05.2008	50	239.42	2.03	320	2.3	7.1	2	nein	ja
400	400	69-B-2	1	26.05.2008	42	86.85	0.89	89	0.9	4.7	1	nein	ja

Proben-nr.	Eppi-ID	ID-intern	Proben-art	Datum	Gewicht (mg)	NanoDrop		Bioanalytiker			Anzahl Hybs	Real-Time	in Studie
						ng/µl	260/230	ng/µl	28S/18S	RIN			
410	410	74-AA	0	31.10.2008	60	164.79	1.98				0	nein	ja
425	425	59-B	2	01.07.2008		27.27	0.79				0	nein	ja
573	573_A	63-B	2	31.10.2008	270	1021.86	2.04				0	nein	nein
573	573_B	63-B	2	31.10.2008	270	461.00	1.28				0	nein	nein
583	583	53-B	2	26.05.2008	43	24.75	0.52	2	0	NaN	0	nein	ja
585	585	54-C	2	26.05.2008	86	154.54	1.96	213	1.4	6	2	nein	ja
601	601	69-C	2	31.10.2008	74	70.80	0.82				0	nein	ja
602	602	70-C	2	31.10.2008	74	21.17	0.39				0	nein	ja
1008	1008	1-A	0	01.06.2006		546.92	2.02	991	0	N/A	2	ja	nein
1009	1009	2-A	0	01.06.2006		4101.00	1.85	6976	0	5.2	3	ja	ja
1010	1010	3-A	0	01.06.2006		1105.00	2.19	2592.5	0	4.1	2	ja	ja
1011	1011	4-A	0	01.06.2006		3005.00	2.03	4086	0	4.6	3	ja	ja
1012	1012	5-A-1	0	01.06.2006		381.00	2.1	564	0	N/A	2	ja	ja
1013	1013	5-A-2	0	01.06.2006		202.62	1.6	286	0	N/A	2	nein	ja
1014	1014	6-A	0	01.06.2006		41.39	1.24	13	0	N/A	0	nein	ja
1015	1015	7-A	0	01.06.2006		1368.00	2.16	2181	0	N/A	3	ja	ja
1016	1016	8-A	0	02.06.2006		74.33	2	77	0	N/A	1	nein	nein
1017	1017	2-B	1	02.06.2006		275.25	2.05	342	0	N/A	2	nein	ja
1018	1018	9-A	0	02.06.2006		-0.74	-	2	0	N/A	0	nein	nein
1019	1019	7-B	1	02.06.2006		29.62	0.78	80	0	2.4	0	nein	ja
1020	1020	3-B	1	02.06.2006		116.01	1.74	111	1.9	4.4	1	nein	ja
1021	1021	5-B-1	1	02.06.2006		55.99	0.57	0	3.8	3.9	0	nein	ja
1022	1022	5-B-2	1	02.06.2006		88.04	1.12	0	3	3.5	1	nein	ja
1023	1023	6-B	1	24.07.2006		89.27	1.87	0	0.2	2.6	1	nein	ja
1024	1024	10-A	0	24.07.2006		20.06	1.23	18	0	1	0	nein	nein
1025	1025	11-A	0	24.07.2006		199.00	2.09	166	0	2.4	2	nein	ja
1026	1026	4-B	1	24.07.2006		150.48	1.71	127	0	2.7	1	nein	ja
1027	1027	12-A-1	1	28.07.2006		87.68	1.75	0			1	nein	ja
1028	1028	12-A-2	1	28.07.2006		62.26	1.55	0	0	N/A	0	nein	ja
1029	1029	13-A	2	28.07.2006		9.55	0.52	0	0	2.1	0	nein	ja
1030	1030	11-B	1	28.07.2006		378.86	2	437	0.4	3.3	2	ja	ja
1031	1031_A	2-C	2	05.12.2006	420	406.18	1.7	479	0.9	9.5	3	nein	ja
1031	1031_B	2-C	2	05.12.2006	420	1870.00	2.11	2940	2.3	6.8	3	ja	ja
1032	1032	14-A	0	05.12.2006	50	823.05	2.19	1364	3.6	7.6	2	ja	ja
1033	1033	14-B	1	05.12.2006	53	308.94	1.75	359	3.3	3.3	2	nein	ja
1034	1034_A	5-C-1	2	05.12.2006	330	160.02	1.64	174	3.3	8.5	2	nein	ja
1034	1034_B	5-C-1	2	05.12.2006	330	306.99	1.94	462	4.8	8.5	2	nein	ja
1035	1035	5-C-2	2	05.12.2006	60	68.66	1.65	0	0	N/A	1	nein	ja
1036	1036_A	6-C	2	05.12.2006	470	273.51	1.73	352	3.4	N/A	2	ja	ja
1036	1036_B	6-C	2	05.12.2006	470	206.51	1.81	262	3.7	N/A	2	ja	ja
1036	1036_Pool	6-C	2	05.12.2006	470	237.84	1.73				2	ja	ja
1037	1037_A	17-A	2	05.12.2006	810	1175.45	2.12	1197	3	8.7	2	ja	ja
1037	1037_B	17-A	2	05.12.2006	810	884.76	2.16	970	1.9	9	2	ja	ja
1037	1037_C	17-A	2	05.12.2006	810	1150.00	2.11	1206	2	8.9	2	ja	ja
1037	1037_Pool	17-A	2	05.12.2006	810	891.64	2.05				2	ja	ja
1038	1038	4-C	2	15.02.2007	104	57.62	1.62	0	2.4	5.1	0	nein	ja
1039	1039_A	12-B-1	2	17.01.2007	720	55.45	1.28	0	4.3	5.5	0	nein	ja
1039	1039_B	12-B-1	2	17.01.2007	720	60.42	1.16	0	0	N/A	0	nein	ja
1040	1040_A	12-B-2	2	17.01.2007	1280	61.88	1.5	0	4.3	5.1	0	nein	ja
1040	1040_B	12-B-2	2	17.01.2007	1280	28.38	1.29	0	0	2.9	0	nein	ja
1040	1040_C	12-B-2	2	17.01.2007	1280	96.46	0.62	151	0	2.4	0	nein	ja
1040	1040_D	12-B-2	2	17.01.2007	1280	33.93	0.99	0	0	4.7	0	nein	ja
1040	1040_E	12-B-2	2	17.01.2007	1280	39.45	1.49	0	0	4	0	nein	ja
1041	1041	3-C	2	05.12.2006	62	55.99	0.67	0	2.3	N/A	0	nein	ja
1042	1042	4-D	2	17.01.2007	200	770.17	2.12	990	6.6	N/A	2	ja	ja
1043	1043	20-A	0	05.12.2006	13	329.30	1.05	356	3.2	6.7	2	nein	ja
1044	1044	20-B	0	02.02.2007	64	147.58	1.24	201	3	8.5	1	nein	ja
1045	1045	21-A	0	02.02.2007	50	121.64	1.99	132	0	3.5	1	nein	nein
1046	1046	22-A	0	02.02.2007	38	38.19	1.55	0	2.4	4.1	0	nein	nein
1047	1047	20-C	1	02.02.2007	65	501.89	1.99	734	4.3	8.1	2	ja	ja
1048	1048	11-C	2	17.01.2007	220	94.55	1.47	54	6.1	N/A	1	nein	ja
1049	1049	23-A	0	02.02.2007	39	29.31	0.21	0	3.7	N/A	0	nein	ja
1050	1050	24-A	0	02.02.2007	43	872.56	2.17	1334	5.9	N/A	2	ja	ja
1051	1051	25-A	2	17.01.2007	70	28.86	0.39	0	6.7	N/A	0	nein	ja
1052	1052	26-A	0	02.02.2007	49	41.64	1.2	0	3.3	4.2	0	nein	ja
1053	1053	27-A	1	02.02.2007	47	38.34	0.33	0	18.6	N/A	0	nein	ja
1054	1054	24-B	1	15.02.2007	52	1242.73	2.19	1170	5.4	8.8	2	ja	ja
1055	1055_A	14-C	2	22.03.2007	710	22.49	1.3	13	0	N/A	0	nein	ja
1055	1055_B	14-C	2	22.03.2007	710	47.10	0.13	20	3.1	N/A	0	nein	ja

Proben-nr.	Eppi-ID	ID-intern	Proben-art	Datum	Gewicht (mg)	NanoDrop		Bioanalyzer			Anzahl Hybs	Real-Time	in Studie
						ng/µl	260/230	ng/µl	28S/18S	RIN			
1055	1055_C	14-C	2	22.03.2007	710	24.49	0.49	18	3.3	N/A	0	nein	ja
1055	1055_D	14-C	2	22.03.2007	710	63.20	0.81	32	3.7	N/A	0	nein	ja
1055	1055_E	14-C	2	22.03.2007	710	51.93	0.96	30	3.6	N/A	0	nein	ja
1055	1055_F	14-C	2	22.03.2007	710	62.57	1.56	48	2.5	8.3	0	nein	ja
1056	1056	28-A	1	15.02.2007	35	15.75	0.36	0	3.6	4.4	0	nein	ja
1057	1057	23-B	1	15.02.2007	36	75.96	0.87	22	0	N/A	1	nein	ja
1058	1058	29-A	0	15.02.2007	46	317.52	1.61	319	4.2	N/A	2	nein	ja
1059	1059	30-A	1	15.02.2007	33	73.93	1.71	60	3.3	N/A	1	nein	ja
1060	1060	31-A	0	15.02.2007	40	149.57	1.96	124	4.8	N/A	2	nein	nein
1061	1061	32-A-1	1	15.02.2007	28	30.87	1.23	0	2.7	N/A	0	nein	ja
1062	1062	32-A-2	1	21.03.2007	30	35.30	0.87	12	0	1	0	nein	ja
1063	1063	33-A	0	21.03.2007	40	240.79	1.92	237	6.5	N/A		nein	nein
1064	1064	29-B-1	1	21.03.2007	40	72.18	0.17	66	2.6	N/A	1	nein	ja
1065	1065	29-B-2	1	21.03.2007	50	70.11	0.78	23	0	4.3	1	nein	ja
1066	1066	34-A	0	21.03.2007	50	211.36	0.68	273	4.6	N/A	2	nein	ja
1067	1067	35-A	0	21.03.2007	40	251.66	1.97	272	3.6	N/A	2	nein	ja
1068	1068	36-A	0	21.03.2007	35	460.56	1.97	603	4.2	8.7	2	ja	ja
1069	1069	37-A	0	21.03.2007	36	48.57	1.43	34	4.2	N/A	0	nein	ja
1070	1070	34-B-1	1	22.03.2007	40	564.98	1.82	725	4	8.5	2	ja	ja
1071	1071	34-B-2	1	22.03.2007	44	458.57	1.44	545	3	9.1	2	nein	ja
1072	1072	38-A	0	08.05.2007	40	785.55	1.4	917	2.7	8.3	2	ja	ja
1073	1073	39-A	0	08.05.2007	40	130.78	1.94	136	2	3.1	2	nein	ja
1074	1074	35-B	1	08.05.2007	64	34.98	0.38	20	0	1	0	nein	ja
1075	1075	38-B	1	08.05.2007	27	318.92	1.69	278	4.5	8.2	2	nein	ja
1076	1076	37-B	1	29.05.2007	46	178.46	1.64	139	2	6.5	2	nein	ja
1077	1077	39-B	1	29.05.2007	46	19.13	0.63	14	0	N/A	0	nein	ja
1078	1078_A	27-B	2	29.05.2007	396	35.68	1.18	44	1.6	N/A	0	nein	ja
1078	1078_B	27-B	2	29.05.2007	396	46.02	1.57				0	nein	ja
1078	1078_C	27-B	2	29.05.2007	396	51.83	0.24				0	nein	ja
1079	1079	40-A	0	29.05.2007	61	411.40	1.74	403	0	N/A	2	ja	ja
1080	1080_A	20-D	2	29.05.2007	178	51.24	0.45	37	0	5.4	1	nein	ja
1080	1080_B	20-D	2	29.05.2007	178	70.21	1.37	69	0.8	N/A	1	nein	ja
1081	1081_A	32-B	2	29.05.2007	128	110.63	1.47	125	0.1	4.3	2	nein	ja
1081	1081_B	32-B	2	29.05.2007	128	80.54	1.1	47	0	N/A	2	nein	ja
1082	1082	43-A	0	18.09.2007	44	188.57	0.58	267	1.3	4.8	2	nein	nein
1083	1083_A	24-C	2	17.09.2007	226	931.01	2.2				3	nein	ja
1083	1083_B	24-C	2	17.09.2007	226	1670.29	1.97				3	nein	ja
1083	1083_Pool	24-C	2	17.09.2007	226	1972.83	2.05	2264	0	8.6	3	ja	ja
1084	1084	44-A	0	04.07.2007	40	39.81	0.17	30	0.1	4.3	0	nein	ja
1085	1085	40-B	1	04.07.2007	30	38.53	1.02	40	2.9	6	0	nein	ja
1086	1086_A	29-C	2	17.09.2007	256	176.52	1.91				1	nein	ja
1086	1086_B	29-C	2	17.09.2007	256	110.49	1.75				1	nein	ja
1086	1086_Pool	29-C	2	17.09.2007	256	140.01	1.8	159	2.1	8.2	1	nein	ja
1087	1087	30-B	2	17.09.2007	130	40.28	0.35	37	1.8	7.4	0	nein	ja
1088	1088	44-B	1	04.07.2007	40	190.38	1.89	200	1	5.7	2	nein	ja
1089	1089	45-A	0	04.07.2007	75	409.46	0.84	437	1.5	4.9	2	ja	nein
1090	1090	46-A	0	18.09.2007	40	12.93	0.26	5	0	N/A	0	nein	nein
1091	1091_A	36-B	2	17.09.2007	145	29.11	1.34	33	1.7	7.1	2	nein	ja
1091	1091_B	36-B	2	17.09.2007	145	157.11	1.78	194	1.7	6.9	2	nein	ja
1092	1092_A	26-B	2	17.09.2007	247	86.70	1.5	100	1.7	6.4	1	nein	ja
1092	1092_B	26-B	2	17.09.2007	247	49.57	0.16	57	1.8	6.5	1	nein	ja
1093	1093	49-A	0	18.09.2007	63	14.44	0.88	9	1.7	N/A	0	nein	ja
1094	1094	38-C	2	17.09.2007	66	122.77	1.2	120	1.8	7.4	1	nein	ja
1095	1095	50-A	2	18.09.2007	44	18.33	0.09	19	1.2	5.8	0	nein	ja
1096	1096	37-C	2	18.09.2007	72	610.10	1.52	782	2.6	7.1	2	ja	ja
1098	1098_A	51-A	2	19.09.2007	858	214.95	2.03				2	ja	ja
1098	1098_B	51-A	2	19.09.2007	858	710.18	2.02				2	ja	ja
1098	1098_C	51-A	2	19.09.2007	858	101.85	1.45				2	ja	ja
1098	1098_D	51-A	2	19.09.2007	858	863.78	2.11				2	ja	ja
1098	1098_E	51-A	2	19.09.2007	858	563.52	2.1				2	ja	ja
1098	1098_Pool	51-A	2	19.09.2007	858	692.71	2.04	848	1.8	6.9	2	ja	ja
1099	1099	52-A	0	18.09.2007	74	683.27	1.88	1004	2.6	6.7	2	ja	ja
1100	1100_A	39-C	2	28.11.2007	43	8.22	0.31				2	nein	ja
1100	1100_B	39-C	2	28.11.2007	43	222.04	1.02	271	2.7	5.1	2	nein	ja
1101	1101	49-B	1	28.11.2007	40	57.08	0.7	52	3.1	4.8	0	nein	ja
1102	1102	55-A	0	28.11.2007	10	10.60	0.45	7	0	NaN	0	nein	ja
1103	1103	56-A	1	28.11.2007	40	20.48	0.44	1	0	NaN	0	nein	ja
1104	1104_A	40-C	2	28.11.2007	140	56.59	1.68	67	0	2.3	1	nein	ja
1104	1104_B	40-C	2	28.11.2007	140	129.77	0.62				1	nein	ja
1105	1105	58-A	0	28.11.2007	37	237.07	0.71	341	0.6	4.4	2	nein	ja

Proben-nr.	Eppi-ID	ID-intern	Proben-art	Datum	Gewicht (mg)	NanoDrop		Bioanalyzer			Anzahl Hybs	Real-Time	in Studie
						ng/ μ l	260/230	ng/ μ l	28S/18S	RIN			
1106	1106	59-A	0	28.11.2007	42	226.64	1.1	297	3.1	7.3	2	nein	ja
1107	1107	54-B	1	28.11.2007	50	602.50	1.96	722	2.4	7.8	2	ja	ja
1108	1108	58-B	1	27.03.2008	40	4.93	0.03	3	0	NaN	0	nein	ja
1111	1111	64-A	2	27.03.2008	36	230.08	1.56	280	1.1	6.2	2	nein	ja
1112	1112	65-A	0	27.03.2008	33	89.55	1.28	213	1.5	5.8	1	nein	ja
1113	1113	66-A	0	27.03.2008	30	19.79	0.73	6	0	NaN	0	nein	ja
1114	1114	67-A	0	26.05.2008	43	887.92	0.82	933	2.5	7.8	2	ja	ja
1115	1115	68-A	0	26.05.2008	25	33.02	0.44	26	2.1	5.4	0	nein	ja
1116	1116	52-B	2	26.05.2008	73	38.23	1.07	28	2.1	4.9	0	nein	ja
1117	1117	67-B	1	26.05.2008	37	19.85	0.85	15	1.2	4.1	0	nein	ja
1119	1119	68-B	1	26.05.2008	36	13.82	0.31	9	0	NaN	0	nein	ja
1120	1120	72-A	1	26.05.2008	49	325.85	2.03	300	0.9	5	2	nein	ja
1121	1121	70-B	0	26.05.2008	65	47.00	0.58	66	0	2.6	0	nein	ja
1122	1122	60-C	2	18.06.2008	46	540.38	1.74	399	2.2	7.8	2	ja	ja
1123	1123	71-A	0	18.06.2008	41	209.67	1.43	313	2.4	8.2	2	nein	ja
1124	1124	74-A		22.07.2008	33	64.72	0.62	65	1.5	7.4	0	nein	ja

D

ÜBERSICHT DER MIKROARRAYEXPERIMENTE

Slide-ID	Slide-Feld	Tumorprobe	Replik	EMMA-Value
Mc2001S01	Mc2001S01_A3	1009	1009I_BP0	2-A_1009_BP-0
Mc2001S01	Mc2001S01_A5	1011	1011I_BP0	4-A_1011_BP-0
Mc2001S01	Mc2001S01_B2	1015	1015I_BP0	7-A_1015_BP-0
Mc2001S01	Mc2001S01_B4	1031_B	1031BI_OP	2-C_1031B_OP
Mc2001S01	Mc2001S01_A1		YellowI	
Mc2001S01	Mc2001S01_B6		YellowII	YellowII
Mc2002S01	Mc2002S01_A1	1009	1009II_BP0	2-A_1009_BP-0
Mc2002S01	Mc2002S01_A2	1010	1010I_BP0	3-A_1010_BP-0
Mc2002S01	Mc2002S01_A3	1011	1011II_BP0	4-A_1011_BP-0
Mc2002S01	Mc2002S01_A4	1012	1012I_BP0	5-A-1_1012_BP-0
Mc2002S01	Mc2002S01_A5	1013	1013I_BP0	5-A-2_1013_BP-0
Mc2002S01	Mc2002S01_A6	1015	1015II_BP0	7-A_1015_BP-0
Mc2002S01	Mc2002S01_B1	1017	1017I_BP1	2-B_1017_BP-1
Mc2002S01	Mc2002S01_B2	1025	1025I_BP0	11-A_1025_BP-0
Mc2002S01	Mc2002S01_B3	1030	1030I_BP1	11-B_1030_BP-1
Mc2002S01	Mc2002S01_B4	1031_B	1031BII_OP	2-C_1031B_OP
Mc2002S01	Mc2002S01_B5	1032	1032I_BP0	14-A_1032_BP-0
Mc2002S01	Mc2002S01_B6	1033	1033I_BP1	14-B_1033_BP-1
Mc2003S01	Mc2003S01_A1	1034_B	1034BI_OP	5-C-1_1034B_OP
Mc2003S01	Mc2003S01_A2	1036_Pool	1036I_OP	6-C_1036_OP
Mc2003S01	Mc2003S01_A3	1037_Pool	1037I_OP	17-A_1037_OP
Mc2003S01	Mc2003S01_A4	1042	1042I_OP	4-D_1042_OP
Mc2003S01	Mc2003S01_A5	1047	1047I_BP1	20-C_1047_BP-1
Mc2003S01	Mc2003S01_A6	1050	1050I_BP0	24-A_1050_BP-0
Mc2003S01	Mc2003S01_B1	1054	1054I_BP1	24-B_1054_BP-1
Mc2003S01	Mc2003S01_B2	1058	1058I_BP0	29-A_1058_BP-0
Mc2003S01	Mc2003S01_B3	1060	1060I_BP0	31-A_1060_BP-0
Mc2003S01	Mc2003S01_B4	1067	1067I_BP0	35-A_1067_BP-0
Mc2003S01	Mc2003S01_B5	1068	1068I_BP0	36-A_1068_BP-0
Mc2003S01	Mc2003S01_B6	1070	1070I_BP1	34-B-1_1070_BP-1
Mc2004S01	Mc2004S01_A1	1016	1016_BP0	8-A_1016_BP-0
Mc2004S01	Mc2004S01_A2	1020	1020_BP1	3-B_1020_BP-1
Mc2004S01	Mc2004S01_A3	1023	1023_BP1	6-B_1023_BP-1
Mc2004S01	Mc2004S01_A4	1026	1026_BP1	4-B_1026_BP-1
Mc2004S01	Mc2004S01_A5	1027	1027_BP1	12-A-1_1027_BP-1
Mc2004S01	Mc2004S01_A6	1035	1035_OP	5-C-2_1035_OP
Mc2004S01	Mc2004S01_B1	1045	1045_BP0	21-A_1045_BP-0
Mc2004S01	Mc2004S01_B2	1048	1048_OP	11-C_1048_OP
Mc2004S01	Mc2004S01_B3	1059	1059_BP1	30-A_1059_BP-1
Mc2004S01	Mc2004S01_B4	1071	1071I_BP1	34-B-2_1071_BP-1
Mc2004S01	Mc2004S01_B5	1072	1072I_BP0	38-A_1072_BP-0
Mc2004S01	Mc2004S01_B6	1075	1075I_BP1	38-B_1075_BP-1
Mc2005S01	Mc2005S01_A1	1009	1009III_BP0	2-A_1009_BP-0
Mc2005S01	Mc2005S01_A2	1010	1010II_BP0	3-A_1010_BP-0
Mc2005S01	Mc2005S01_A3	1011	1011III_BP0	4-A_1011_BP-0
Mc2005S01	Mc2005S01_A4	1012	1012II_BP0	5-A-1_1012_BP-0
Mc2005S01	Mc2005S01_A5	1013	1013II_BP0	5-A-2_1013_BP-0
Mc2005S01	Mc2005S01_A6	1015	1015III_BP0	7-A_1015_BP-0
Mc2005S01	Mc2005S01_B1	1017	1017II_BP1	2-B_1017_BP-1
Mc2005S01	Mc2005S01_B2	1025	1025II_BP0	N.A.
Mc2005S01	Mc2005S01_B3	1030	1030II_BP1	11-B_1030_BP-1
Mc2005S01	Mc2005S01_B4	1031_B	1031BIII_OP	2-C_1031B_OP
Mc2005S01	Mc2005S01_B6	1073	1073I_BP0	39-A_1073_BP-0
Mc2005S01	Mc2005S01_B5		YellowIII	YellowIII

Slide-ID	Slide-Feld	Tumorprobe	Replikat	EMMA-Value
Mc2006S01	Mc2006S01_A1	1032	1032II_BP0	14-A_1032_BP-0
Mc2006S01	Mc2006S01_A2	1033	1033II_BP1	14-B_1033_BP-1
Mc2006S01	Mc2006S01_A3	1034_B	1034BII_OP	5-C-1_1034B_OP
Mc2006S01	Mc2006S01_A4	1036_Pool	1036II_OP	6-C_1036_OP
Mc2006S01	Mc2006S01_A5	1037_Pool	1037II_OP	17-A_1037_OP
Mc2006S01	Mc2006S01_A6	1042	1042II_OP	4-D_1042_OP
Mc2006S01	Mc2006S01_B1	1047	1047II_BP1	20-C_1047_BP-1
Mc2006S01	Mc2006S01_B2	1050	1050II_BP0	24-A_1050_BP-0
Mc2006S01	Mc2006S01_B3	1054	1054II_BP1	24-B_1054_BP-1
Mc2006S01	Mc2006S01_B4	1058	1058II_BP0	29-A_1058_BP-0
Mc2006S01	Mc2006S01_B5	1076	1076I_BP1	37-B_1076_BP-1
Mc2006S01	Mc2006S01_B6	1079	1079I_BP0	40-A_1079_BP-0
Mc2007S01	Mc2007S01_A1	1060	1060II_BP0	31-A_1060_BP-0
Mc2007S01	Mc2007S01_A2	1067	1067II_BP0	35-A_1067_BP-0
Mc2007S01	Mc2007S01_A3	1068	1068II_BP0	36-A_1068_BP-0
Mc2007S01	Mc2007S01_A4	1070	1070II_BP1	34-B-1_1070_BP-1
Mc2007S01	Mc2007S01_A5	1071	1071II_BP1	34-B-2_1071_BP-1
Mc2007S01	Mc2007S01_A6	1072	1072II_BP0	38-A_1072_BP-0
Mc2007S01	Mc2007S01_B1	1075	1075II_BP1	38-B_1075_BP-1
Mc2007S01	Mc2007S01_B2	1076	1076II_BP1	37-B_1076_BP-1
Mc2007S01	Mc2007S01_B3	1079	1079II_BP0	40-A_1079_BP-0
Mc2007S01	Mc2007S01_B4	1080_B	1080B_OP	20-D_1080B_OP
Mc2007S01	Mc2007S01_B5	1081_A	1081A_OP	32-B_1081AB_OP
Mc2007S01	Mc2007S01_B6		YellowIV	YellowIV
Mc2008S01	Mc2008S01_A4	1083_Pool	1083III_OP	24-C_1083_OP
Mc2008S01	Mc2008S01_A5	1088	1088II_BP1	44-B_1088_BP-1
Mc2008S01	Mc2008S01_A6	1089	1089II_BP0	45-A_1089_BP-0
Mc2008S01	Mc2008S01_B1	1091_B	1091BII_OP	36-B_1091B_OP
Mc2008S01	Mc2008S01_B2	1092_A	1092A_OP	26-B_1092A_OP
Mc2008S01	Mc2008S01_B3	1094	1094_OP	38-C_1094_OP
Mc2008S01	Mc2008S01_B4	1096	1096II_OP	37-C_1096_OP
Mc2008S01	Mc2008S01_B5	1098_Pool	1098II_OP	51-A_1098_OP
Mc2008S01	Mc2008S01_B6	1099	1099II_BP0	52-A_1099_BP-0
Mc2008S01	Mc2008S01_A2	290	290II_BP	27-AA_290_BP
Mc2008S01	Mc2008S01_A3	322	322_BP	53-A_322_BP
Mc2008S01	Mc2008S01_A1		YellowV	YellowV
Mc2009S02	Mc2009S02_B5	1022	1022_BP1	5-B-2_1022_BP-1
Mc2009S02	Mc2009S02_A2	1043	1043I_BP0	20-A_1043_BP-0
Mc2009S02	Mc2009S02_A3	1044	1044_BP0	20-B_1044_BP-0
Mc2009S02	Mc2009S02_A4	1057	1057_BP1	23-B_1057_BP-1
Mc2009S02	Mc2009S02_A5	1064	1064_BP1	29-B-1_1064_BP-1
Mc2009S02	Mc2009S02_A6	1065	1065_BP1	29-B-2_1065_BP-1
Mc2009S02	Mc2009S02_B1	1066	1066I_BP0	34-A_1066_BP-0
Mc2009S02	Mc2009S02_B2	1073	1073II_BP0	39-A_1073_BP-0
Mc2009S02	Mc2009S02_B3	1081_B	1081B_OP	32-B_1081AB_OP
Mc2009S02	Mc2009S02_B4	1082	1082I_BP0	43-A_1082_BP-0
Mc2009S02	Mc2009S02_A1	1083_Pool	1083I_OP	N.A.
Mc2009S02	Mc2009S02_B6	1086_Pool	1086_OP	29-C_1086_OP
Mc2010S02	Mc2010S02_A3	1043	1043II_BP0	20-A_1043_BP-0
Mc2010S02	Mc2010S02_A4	1066	1066II_BP0	34-A_1066_BP-0
Mc2010S02	Mc2010S02_A5	1082	1082II_BP0	43-A_1082_BP-0
Mc2010S02	Mc2010S02_A6	1083_Pool	1083II_OP	24-C_1083_OP
Mc2010S02	Mc2010S02_B1	1088	1088I_BP1	44-B_1088_BP-1
Mc2010S02	Mc2010S02_B2	1089	1089I_BP0	45-A_1089_BP-0
Mc2010S02	Mc2010S02_B3	1091_B	1091BI_OP	36-B_1091B_OP
Mc2010S02	Mc2010S02_B4	1096	1096I_OP	37-C_1096_OP
Mc2010S02	Mc2010S02_B5	1098_Pool	1098I_OP	51-A_1098_OP
Mc2010S02	Mc2010S02_B6	1099	1099I_BP0	52-A_1099_BP-0
Mc2010S02	Mc2010S02_A2	290	290I_BP	27-AA_290_BP
Mc2010S02	Mc2010S02_A1		YellowVI	YellowVI
Mc2011S02	Mc2011S02_A1	257	257I	N.A.
Mc2011S02	Mc2011S02_A2	263	263I_BP-0	25-AAA-1_263_BP-0
Mc2011S02	Mc2011S02_A3	296	296I_BP-0	38-A-2_296_BP-0
Mc2011S02	Mc2011S02_A4	300	300I	N.A.
Mc2011S02	Mc2011S02_A5	315	315_BP-0	64-AA_315_BP-0
Mc2011S02	Mc2011S02_A6	324	324I_BP-0	54-A_324_BP-0
Mc2011S02	Mc2011S02_B1	326	326_BP-0	55-A-2_326_BP-0
Mc2011S02	Mc2011S02_B2	329	329I	N.A.
Mc2011S02	Mc2011S02_B3	339	339I	N.A.
Mc2011S02	Mc2011S02_B4	368	368_BP-0	63-A_368_BP-0
Mc2011S02	Mc2011S02_B5	389	389I	389
Mc2011S02	Mc2011S02_B6	399	399I_BP-1	69-B-1_399_BP-1
Mc2012S02	Mc2012S02_A1	1008	1008I_BP-0	1-A_1008_BP-0
Mc2012S02	Mc2012S02_A2	1104_A	1104A_OP	40-C_1104A_OP

Slide-ID	Slide-Feld	Tumorprobe	Replikat	EMMA-Value
Mc2012S02	Mc2012S02_A3	1107	1107I_BP-1	54-B_1107_BP-1
Mc2012S02	Mc2012S02_A4	1111	1111I_OP	64-A_1111_OP
Mc2012S02	Mc2012S02_A5	1120	1120I_OP	72-A_1120_OP
Mc2012S02	Mc2012S02_A6	1122	1122I_OP	60-C_1122_OP
Mc2012S02	Mc2012S02_B1	1123	1123I_BP-0	71-A_1123_BP-0
Mc2012S02	Mc2012S02_B3	257	257II	N.A.
Mc2012S02	Mc2012S02_B4	263	263II_BP-0	25-AAA-1_263_BP-0
Mc2012S02	Mc2012S02_B5	296	296II_BP-0	38-A-2_296_BP-0
Mc2012S02	Mc2012S02_B6	300	300II	N.A.
Mc2012S02	Mc2012S02_B2	585	585I_OP	54-C_585_OP
Mc2013S02	Mc2013S02_B3	1100_B	1100BI_OP	39-C_1100B_OP
Mc2013S02	Mc2013S02_B4	1105	1105I_BP-0	58-A_1105_BP-0
Mc2013S02	Mc2013S02_B5	1106	1106I_BP-0	59-A_1106_BP-0
Mc2013S02	Mc2013S02_B6	1114	1114I_BP-0	67-A_1114_BP-0
Mc2013S02	Mc2013S02_A1	258	258I_BP	258_BP
Mc2013S02	Mc2013S02_A2	298	298_BP-0	51-AA_298_BP-0
Mc2013S02	Mc2013S02_A3	338_B	338BI_OP	44-C_338B_OP
Mc2013S02	Mc2013S02_A4	358	358_BP-0	73-A_358_BP-0
Mc2013S02	Mc2013S02_A5	367	367I_BP-0	72-AA_367_BP-0
Mc2013S02	Mc2013S02_A6	390	390I_BP-0	70-A_390_BP-0
Mc2013S02	Mc2013S02_B1	393	393I_BP-0	69-A_393_BP-0
Mc2013S02	Mc2013S02_B2	400	400_BP-1	69-B-2_400_BP-1
Mc2014S02	Mc2014S02_A2	1008	1008II_BP-0	1-A_1008_BP-0
Mc2014S02	Mc2014S02_A3	1100_B	1100BII_OP	39-C_1100B_OP
Mc2014S02	Mc2014S02_A4	1105	1105II_BP-0	58-A_1105_BP-0
Mc2014S02	Mc2014S02_A5	1106	1106II_BP-0	59-A_1106_BP-0
Mc2014S02	Mc2014S02_A6	1107	1107II_BP-1	54-B_1107_BP-1
Mc2014S02	Mc2014S02_B1	1111	1111II_OP	64-A_1111_OP
Mc2014S02	Mc2014S02_B2	1112	1112_BP-0	65-A_1112_BP-0
Mc2014S02	Mc2014S02_B3	1114	1114II_BP-0	67-A_1114_BP-0
Mc2014S02	Mc2014S02_B4	1120	1120II_OP	72-A_1120_OP
Mc2014S02	Mc2014S02_B5	1122	1122II_OP	60-C_1122_OP
Mc2014S02	Mc2014S02_B6	1123	1123II_BP-0	71-A_1123_BP-0
Mc2014S02	Mc2014S02_A1	585	585II_OP	54-C_585_OP
Mc2015S02	Mc2015S02_A1	258	258II_BP	258_BP
Mc2015S02	Mc2015S02_A2	324	324II_BP-0	54-A_324_BP-0
Mc2015S02	Mc2015S02_A3	329	329II	N.A.
Mc2015S02	Mc2015S02_A4	338_B	338BII_OP	44-C_338B_OP
Mc2015S02	Mc2015S02_A5	339	339II	N.A.
Mc2015S02	Mc2015S02_B1	367	367II_BP-0	72-AA_367_BP-0
Mc2015S02	Mc2015S02_B2	389	389II	389
Mc2015S02	Mc2015S02_B3	390	390II_BP-0	70-A_390_BP-0
Mc2015S02	Mc2015S02_B4	393	393II_BP0	69-A_393_BP-0
Mc2015S02	Mc2015S02_B5	399	399II_BP-1	69-B-1_399_BP-1

LITERATURVERZEICHNIS

Abramovitz M. und Leyland-Jones B., **2006**. A systems approach to clinical oncology: focus on breast cancer. *Proteome Sci*, 4:5.

Ahern T., Lash T., Egan K. und Baron J., **2009**. Lifetime tobacco smoke exposure and breast cancer incidence. *Cancer Causes Control*, 20(10):1837–1844.

Allen N. E., Beral V., Casabonne D., Kan S. W., Reeves G. K., Brown A. und Green J., **2009**. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst*, 101(5):296–305.

Allred D. C., Wu Y., Mao S., Nagtegaal I. D., Lee S., Perou C. M., Mohsin S. K., O'Connell P., Tsimelzon A. und Medina D., **2008**. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clin Cancer Res*, 14(2):370–378.

Armstrong K., Eisen A. und Weber B., **2000**. Assessing the risk of breast cancer. *N Engl J Med*, 342(8):564–571.

Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis A. P., Dolinski K., Dwight S. S., Eppig J. T., Harris M. A., Hill D. P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J. C., Richardson J. E., Ringwald M., Rubin G. M. und Sherlock G., **2000**. Gene ontology: tool for the unification of biology. *Nat Genet*, 25(1):25–29.

Balmain A., Gray J. und Ponder B., **2003**. The genetics and genomics of cancer. *Nat Genet*, 33 Suppl:238–244.

Bartneck M., **2006**. Metabolite profiling and gene expression profiling of human primary mamma carcinomas. Diplomarbeit, Universität Bielefeld, Fakultät für Biologie, Lehrstuhl der Genetik.

Bernardi D., Errante D., Galligioni E., Crivellari D., Bianco A., Salvagno L. und Fentiman I. S., **2008**. Treatment of breast cancer in older women. *Acta Oncol*, 47(2):187–198.

- BMGS, Hrsg., **2003**. *Gesundheitsziele.de Forum zur Entwicklung und Umsetzung von Gesundheitszielen in Deutschland (Bericht). Auszug der Ergebnisse von AG 5, Brustkrebs: Mortalität vermindern, Lebensqualität erhöhen*. Bundesministerium für Gesundheit und Soziale Sicherung, Berlin.
- Bourdeau V., Deschênes J., Laperrière D., Aid M., White J. H. und Mader S., **2008**. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Res*, 36(1):76–93.
- Boyle P. und Levin B., Hrg., **2008**. *World Cancer Report*. IARC Press, Lyon.
- Brennan D. J., O'Brien S. L., Fagan A., Culhane A. C., Higgins D. G., Duffy M. J. und Gallagher W. M., **2005**. Application of DNA microarray technology in determining breast cancer prognosis and therapeutic response. *Expert Opin Biol Ther*, 5(8):1069–1083.
- Brody J. G. und Rudel R. A., **2003**. Environmental pollutants and breast cancer. *Environ Health Perspect*, 111(8):1007–1019.
- Buyse M., Loi S., van't Veer L., Viale G., Delorenzi M., Glas A. M., d'Assignies M. S., Bergh J., Lidereau R., Ellis P., Harris A., Bogaerts J., Therasse P., Floore A., Amakrane M., Piette F., Rutgers E., Sotiriou C., Cardoso F., Piccart M. J. und TRANSBIG Consortium, **2006**. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*, 98(17):1183–1192.
- Cairns R. A., Khokha R. und Hill R. P., **2003**. Molecular mechanisms of tumor invasion and metastasis: an integrated view. *Curr Mol Med*, 3(7):659–671.
- Campa D., McKay J., Sinilnikova O., Hüsing A., Vogel U., Hansen R., Overvad K., Witt P., Clavel-Chapelon F., Boutron-Ruault M.-C., Chajes V., Rohrmann S., Chang-Claude J., Boeing H., Fisher E., Trichopoulou A., Trichopoulos D., Palli D., Villarini A., Sacerdote C., Mattiello A., Tumino R., Peeters P., van Gils C., de Mesquita H. B. B., Lund E., Chirlaque M., Sala N., Suarez L., Barricarte A., Dorronsoro M., Sánchez M.-J., Lenner P., Hallmans G., Tsilidis K., Bingham S., Khaw K.-T., Gallo V., Norat T., Riboli E., Rinaldi S., Lenoir G., Tavtigian S., Canzian F. und Kaaks R., **2009**. Genetic variation in genes of the fatty acid synthesis pathway and breast cancer risk. *Breast Cancer Res Treat*, 118(3):565–574.
- Carroll J. S. und Brown M., **2006**. Estrogen receptor target gene: an evolving concept. *Mol Endocrinol*, 20(8):1707–1714.

- Chang H. Y., Nuyten D. S. A., Sneddon J. B., Hastie T., Tibshirani R., Sørlie T., Dai H., He Y. D., van't Veer L. J., Bartelink H., van de Rijn M., Brown P. O. und van de Vijver M. J., **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):3738–3743.
- Chang J. C., Wooten E. C., Tsimelzon A., Hilsenbeck S. G., Gutierrez M. C., Elledge R., Mohsin S., Osborne C. K., Chamness G. C., Allred D. C. und O'Connell P., **2003**. Gene expression profiling for the prediction of the therapeutic response to docetaxel in patients with breast cancer. *Lancet*, 362(9381):362–369.
- Chanrion M., Negre V., Fontaine H., Salvetat N., Bibeau F., Grogan G. M., Mauriac L., Katsaros D., Molina F., Theillet C. und Darbon J.-M., **2008**. A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res*, 14(6):1744–1752.
- Cheang M. C. U., Voduc D., Bajdik C., Leung S., McKinney S., Chia S. K., Perou C. M. und Nielsen T. O., **2008**. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res*, 14(5):1368–1376.
- Chen G. G., Zeng Q. und Tse G. M., **2008**. Estrogen and its receptors in cancer. *Med Res Rev*, .
- Cheng C.-J., Lin Y.-C., Tsai M.-T., Chen C.-S., Hsieh M.-C., Chen C.-L. und Yang R.-B., **2009**. SCUBE2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer. *Cancer Res*, 69(8):3634–3641.
- Cho J.-H., Lee D., Park J. H. und Lee I.-B., **2004**. Gene selection and classification from microarray data using kernel machine. *FEBS Lett*, 571(1-3):93–98.
- Cianfrocca M. und Goldstein L. J., **2004**. Prognostic and predictive factors in early-stage breast cancer. *Oncologist*, 9(6):606–616.
- Cleator S. und Ashworth A., **2004**. Molecular profiling of breast cancer: clinical implications. *Br J Cancer*, 90(6):1120–1124.
- Cleator S., Powles T., Dexter T., Fulford L., Mackay A., Smith I., Valgeirsson H., Ashworth A. und Dowsett M., **2006**. The effect of the stromal component of breast tumours on prediction of clinical outcome using gene expression microarray analysis. *Breast Cancer Res*, 8(3):R32.

- Cleveland W. S. und Devlin S. J., **1988**. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J Am Stat Assoc*, 83(403):596–610.
- Coyle Y. M., **2004**. The effect of environment on breast cancer risk. *Breast Cancer Res Treat*, 84(3):273–288.
- Creighton C. J., Cordero K. E., Larios J. M., Miller R. S., Johnson M. D., Chinnaiyan A. M., Lippman M. E. und Rae J. M., **2006**. Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors. *Genome Biol*, 7(4):R28.
- Culhane A. und Howlin J., **2007**. Molecular profiling of breast cancer: transcriptomic studies and beyond. *Cell Mol Life Sci*, 64(24):3185–3200.
- Decker T., Hungermann D. und Böcker W., **2009**. Prognostische und prädiktive Faktoren invasiver Mammakarzinome : Update 2009. *Pathologe*, 30(1):49–55.
- Dell H., **2006**. Nature Milestones in Cancer: Observations from a ploughman. *Nat Rev Cancer*, 6 Suppl:S7.
- Deutsche Krebsgesellschaft, **2009**. Brustkrebs - Antihormontherapie. http://www.krebsgesellschaft.de/pat_ka_brustkrebs_therapie_hormontherapie_107725.html. Letzter Zugriff November 2009.
- Dondrup M., **2007**. *EMMA 2 - A MAGE-compliant system for the analysis of microarray data in integrated functional genomics*. Dissertation, Universität Bielefeld.
- Dondrup M., Albaum S. P., Griebel T., Henckel K., Jünemann S., Kahlke T., Kleindt C. K., Küster H., Linke B., Mertens D., Mittard-Runte V., Neuweger H., Runte K. J., Tauch A., Tille F., Pühler A. und Goesmann A., **2009**. EMMA 2 - A MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics*, 10:50.
- Dondrup M., Goesmann A., Bartels D., Kalinowski J., Krause L., Linke B., Rupp O., Sczyrba A., Pühler A. und Meyer F., **2003**. EMMA: a platform for consistent storage and efficient analysis of microarray data. *J Biotechnol*, 106(2-3):135–146.
- Edén P., Ritz C., Rose C., Fernö M. und Peterson C., **2004**. "Good Old"clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer*, 40(12):1837–1841.

- Ein-Dor L., Kela I., Getz G., Givol D. und Domany E., **2005**. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics*, 21(2):171–178.
- Epidemiologisches Krebsregister NRW, Hrsg., **2009**. *Epidemiologisches Krebsregister NRW - Report 2009*. Epidemiologisches Krebsregister NRW gGmbH, Münster.
- Ewing B. und Green P., **2000**. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet*, 25(2):232–234.
- Fare T. L., Coffey E. M., Dai H., He Y. D., Kessler D. A., Kilian K. A., Koch J. E., LeProust E., Marton M. J., Meyer M. R., Stoughton R. B., Tokiwa G. Y. und Wang Y., **2003**. Effects of atmospheric ozone on microarray data quality. *Anal Chem*, 75(17):4672–4675.
- Ferlay J., Autier P., Boniol M., Heanue M., Colombet M. und Boyle P., **2007**. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol*, 18(3):581–592.
- Ferrara N. und Kerbel R. S., **2005**. Angiogenesis as a therapeutic target. *Nature*, 438(7070):967–974.
- Firozi P. F., Bondy M. L., Sahin A. A., Chang P., Lukmanji F., Singletary E. S., Hassan M. M. und Li D., **2002**. Aromatic DNA adducts and polymorphisms of CYP1A1, NAT2, and GSTM1 in breast cancer. *Carcinogenesis*, 23(2):301–306.
- Funke I., Eiermann W., Engel J., Harbeck N., Janni W., Lebeau A., Permanetter W., Rack B., Untch M. und C. Wolf, **2005**. *Prognostische und prädiktive Faktoren beim primären Mammakarzinom*, Seiten 80–86. In Sauer (2005), 10. Edition.
- GEKID, Hrsg., **2006**. *Krebs in Deutschland. Häufigkeiten und Trends*. Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. in Zusammenarbeit mit dem Robert Koch-Institut, Saarbrücken, 5. Edition.
- Giersiepen K., Heitmann C., Janhsen K. und Lange C., **2005**. *Gesundheitsberichterstattung des Bundes. Heft 25 Brustkrebs*. Robert Koch-Institut, Berlin.
- Glas A., Floore A., Delahaye L., Witteveen A., Pover R., Bakx N., Lahti-Domenici J., Bruinsma T., Warmoes M., Bernardis R., Wessels L. und van 't Veer L., **2006**. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics*, 7(1):278.

- Goldhirsch A., Glick J. H., Gelber R. D., Coates A. S., Thürlimann B., Senn H.-J. und members P., **2005**. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol*, 16(10):1569–1583.
- Goldhirsch A., Ingle J. N., Gelber R. D., Coates A. S., Thürlimann B., Senn H.-J. und members P., **2009**. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol*, 20(8):1319–1329.
- Gruvberger S., Ringnér M., Chen Y., Panavally S., Saal L. H., A B., Fernö M., Peterson C. und Meltzer P. S., **2001**. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res*, 61(16):5979–5984.
- Hall J. M., Couse J. F. und Korach K. S., **2001**. The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J Biol Chem*, 276(40):36869–36872.
- Hanahan D. und Weinberg R. A., **2000**. The hallmarks of cancer. *Cell*, 100(1):57–70.
- Howley P. M., Israel M. A., Law M. F. und Martin M. A., **1979**. A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *J Biol Chem*, 254(11):4876–4883.
- Hsiao L. L., Dangond F., Yoshida T., Hong R., Jensen R. V., Misra J., Dillon W., Lee K. F., Clark K. E., Haverty P., Weng Z., Mutter G. L., Frosch M. P., Macdonald M. E., Milford E. L., Crum C. P., Bueno R., Pratt R. E., Mahadevappa M., Warrington J. A., Stephanopoulos G., Stephanopoulos G. und Gullans S. R., **2001**. A compendium of gene expression in normal human tissues. *Physiol Genomics*, 7(2):97–104.
- Hu Z., Fan C., Oh D., Marron J., He X., Qaqish B., Livasy C., Carey L., Reynolds E., Dressler L., Nobel A., Parker J., Ewend M., Sawyer L., Wu J., Liu Y., Nanda R., Tretiakova M., Orrico A. R., Dreher D., Palazzo J., Perreard L., Nelson E., Mone M., Hansen H., Mullins M., Quackenbush J., Ellis M., Olopade O., Bernard P. und Perou C., **2006**. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, 7(1):96.
- Huggett J., Dheda K., Bustin S. und Zumla A., **2005**. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun*, 6(4):279–284.

- Ioannidis J. P. A., **2005**. Microarrays and molecular research: noise discovery? *Lancet*, 365(9458):454–455.
- Ioannidis J. P. A., **2007**. Is molecular profiling ready for use in clinical decision making? *Oncologist*, 12(3):301–311.
- Ioannidis J. P. A., Allison D. B., Ball C. A., Coulibaly I., Cui X., Culhane A. C., Falchi M., Furlanello C., Game L., Jurman G., Mangion J., Mehta T., Nitzberg M., Page G. P., Petretto E. und van Noort V., **2009**. Repeatability of published microarray gene expression analyses. *Nat Genet*, 41(2):149–155.
- Jeffrey S. S., Fero M. J., Borresen-Dale A. L. und Botstein D., **2002**. Expression array technology in the diagnosis and treatment of breast cancer. *Mol Intervent*, 2(2):101–109.
- Jenssen T.-K. und Hovig E., **2005**. Gene-expression profiling in breast cancer. *Lancet*, 365(9460):634–635.
- Jinga D. C., Blidaru A., Condrea I., Ardeleanu C., Dragomir C., Szegli G., Stefanescu M. und Matache C., **2006**. MMP-9 and MMP-2 gelatinases and TIMP-1 and TIMP-2 inhibitors in breast cancer: correlations with prognostic factors. *J Cell Mol Med*, 10(2):499–510.
- Ju X., Katiyar S., Wang C., Liu M., Jiao X., Li S., Zhou J., Turner J., Lisanti M. P., Russell R. G., Mueller S. C., Ojeifo J., Chen W. S., Hay N. und Pestell R. G., **2007**. Akt1 governs breast cancer progression in vivo. *Proc Natl Acad Sci U S A*, 104(18):7438–7443.
- Kakizaki M., Kuriyama S., Sone T., Ohmori-Matsuda K., Hozawa A., Nakaya N., Fukudo S. und Tsuji I., **2008**. Sleep duration and the risk of breast cancer: the Ohsaki Cohort Study. *Br J Cancer*, 99(9):1502–1505.
- Keinan-Boker L., van Der Schouw Y. T., Grobbee D. E. und Peeters P. H. M., **2004**. Dietary phytoestrogens and breast cancer risk. *Am J Clin Nutr*, 79(2):282–288.
- Kerbel R. S., **2008**. Tumor angiogenesis. *N Engl J Med*, 358(19):2039–2049.
- Kreienberg R., Kopp I., Albert U., Bartsch H. H., Beckmann M., Berg D., Bick U., du Bois A., Budach W., Dunst J., Engel J., Ernst B., Geraedts M., Henschler U., Hölzel D., Jackisch C., König K., Kreipe H., Kühn T., Lebeau A., Leinung S., Link H., Lück H.-J., Madjar H., Maiwald A., Maiwald G., Marschner N., Marx M., von Minckwitz G., Naß-Griegoleit I., Possinger K.,

- Reiter A., Sauerbrei W., Schlake W., Schmutzler R., Schreer I., Schulte H., Schulz K.-D., Souchon R., Thomssen C., Untch M., Wagner U., Weis J. und Zemmler T., **2008**. *Interdisziplinäre S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms*. Zuckschwerdt Verlag, München.
- Kreienberg R., Kopp I., Lorenz W., Budach W., Dunst J., Lebeau A., H.-J. L., v. Minckwitz G., Possinger K., Sauerbrei W., Sauerland S., Souchon R., Thomssen C., Untch M., Volm T., Weis J., Schmitt-Reißer B., Koller M. und Heilmann V., **2004**. *Diagnostik, Therapie und Nachsorge des Mammakarzinoms der Frau. S3-Leitlinie Mammakarzinom*. Deutsche Krebsgesellschaft e. V. - Informationszentrum für Standards in der Onkologie (ISTO), Frankfurt.
- Kristensen V. N., Sørlie T., Geisler J., Langerød A., Yoshimura N., Kåresen R., Harada N., Lønning P. E. und Børresen-Dale A.-L., **2005**. Gene expression profiling of breast cancer in relation to estrogen receptor status and estrogen-metabolizing enzymes: clinical implications. *Clin Cancer Res*, 11(2 Pt 2):878s–883s.
- Kuhajda F. P., **2000**. Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology. *Nutrition*, 16(3):202–208.
- Kuschel B., Rühl I. M., Ditsch N., Schwarz-Böger U., A.Vodermeier, Böttcher B., Plattner B. und Untch M., **2005**. *Das hereditäre Mammakarzinom*, Seiten 86–94. In Sauer (2005), 10. Edition.
- Lebeau A., Högel B., Nährig J. und Permanetter W., **2005**. *Pathomorphologie des Mammakarzinoms*, Seiten 28–46. In Sauer (2005), 10. Edition.
- Lee P. D., Sladek R., Greenwood C. M. T. und Hudson T. J., **2002**. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res*, 12(2):292–297.
- Lendeckel U., Kohl J., Arndt M., Carl-McGrath S., Donat H. und Röcken C., **2005**. Increased expression of ADAM family members in human breast cancer and breast cancer cell lines. *J Cancer Res Clin Oncol*, 131(1):41–48.
- Li F., Tiede B., Massagué J. und Kang Y., **2007**. Beyond tumorigenesis: cancer stem cells in metastasis. *Cell Res*, 17(1):3–14.
- Liu J., Campen A., Huang S., Peng S.-B., Ye X., Palakal M., Dunker A. K., Xia Y. und Li S., **2008**. Identification of a gene signature in cell cycle pathway for breast cancer prognosis using gene expression profiling data. *BMC Med Genomics*, 1(1):39.

- Livak K. J.; Schmittgen T. D., **2001**. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4):402–408.
- Loi S., Desmedt C., Cardoso F., Piccart M. und Sotiriou C., **2005**. Breast cancer gene expression profiling: clinical trial and practice implications. *Pharmacogenomics*, 6(1):49–58.
- Loibl S., von Minckwitz G., Harbeck N., Janni W., Elling D., Kaufmann M., Eggemann H., Nekljudova V., Sommer H., Kiechle M. und Kümmel S., **2008**. Clinical feasibility of (neo)adjuvant taxane-based chemotherapy in older patients: analysis of >4,500 patients from four German randomized breast cancer trials. *Breast Cancer Res*, 10(5):R77.
- Lowe S. W., Cepero E. und Evan G., **2004**. Intrinsic tumour suppression. *Nature*, 432(7015):307–315.
- Lupien M. und Brown M., **2009**. Cistromics of hormone-dependent cancer. *Endocr Relat Cancer*, 16(2):381–389.
- Lupien M., Eeckhoute J., Meyer C. A., Wang Q., Zhang Y., Li W., Carroll J. S., Liu X. S. und Brown M., **2008**. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, 132(6):958–970.
- Lupu R. und Menendez J. A., **2006**. Targeting fatty acid synthase in breast and endometrial cancer: An alternative to selective estrogen receptor modulators? *Endocrinology*, 147(9):4056–4066.
- Majed B., Moreau T., Senouci K., Salmon R. J., Fourquet A. und Asselain B., **2008**. Is obesity an independent prognosis factor in woman breast cancer? *Breast Cancer Res Treat*, 111(2):329–342.
- Majed B., Senouci K. und Asselain B., **2009**. Shortened survival and more metastasis recurrences among overweight breast cancer patients. *Breast J*, 15(5):557–559.
- Martin K. J., Kritzman B. M., Price L. M., Koh B., Kwan C. P., Zhang X., Mackay A., O'Hare M. J., Kaelin C. M., Mutter G. L., Pardee A. B. und Sager R., **2000**. Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res*, 60(8):2232–2238.
- McElroy J. A., Shafer M. M., Trentham-Dietz A., Hampton J. M. und Newcomb P. A., **2006**. Cadmium exposure and breast cancer risk. *J Natl Cancer Inst*, 98(12):869–873.

- Menendez J. A., Oza B. P., Atlas E., Verma V. A., Mehmi I. und Lupu R., **2004**. Inhibition of tumor-associated fatty acid synthase activity antagonizes estradiol- and tamoxifen-induced agonist transactivation of estrogen receptor (ER) in human endometrial adenocarcinoma cells. *Oncogene*, 23(28):4945–4958.
- Mentz A., **2006**. Identifizierung von spezifischen Genexpressionsmustern in primären Brusttumoren. Diplomarbeit, Universität Bielefeld, Fakultät für Biologie, Lehrstuhl der Genetik.
- Meteoglu I., Dikicioglu E., Erkus M., Culhaci N., Kacar F., Ozkara E. und Uyar M., **2005**. Breast carcinogenesis. Transition from hyperplasia to invasive lesions. *Saudi Med J*, 26(12):1889–1896.
- Michiels S., Koscielny S. und Hill C., **2005**. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*, 365(9458):488–492.
- Mullan P. und Millikan R., **2007**. Molecular subtyping of breast cancer: opportunities for new therapeutic approaches. *Cell Mol Life Sci*, 64(24):3219–3232.
- Novoradovskaya N., Whitfield M. L., Basehore L. S., Novoradovsky A., Pesich R., Usary J., Karaca M., Wong W. K., Aprelikova O., Fero M., Perou C. M., Botstein D. und Braman J., **2004**. Universal Reference RNA as a standard for microarray experiments. *BMC Genomics*, 5(1):20.
- Oh D. S., Troester M. A., Usary J., Hu Z., He X., Fan C., Wu J., Carey L. A. und Perou C. M., **2006**. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol*, 24(11):1656–1664.
- Olayioye M. A., **2001**. Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. *Breast Cancer Res*, 3(6):385–389.
- Oliveira A. M., Ross J. S. und Fletcher J. A., **2005**. Tumor suppressor genes in breast cancer: the gatekeepers and the caretakers. *Am J Clin Pathol*, 124 Suppl:S16–S28.
- Osborne C. K., **1998**. Steroid hormone receptors in breast cancer management. *Breast Cancer Res Treat*, 51(3):227–238.

- O'Shea C., McKie N., Buggy Y., Duggan C., Hill A. D. K., McDermott E., O'Higgins N. und Duffy M. J., **2003**. Expression of ADAM-9 mRNA and protein in human breast cancer. *Int J Cancer*, 105(6):754–761.
- Owzar K., Barry W. T., Jung S.-H., Sohn I. und George S. L., **2008**. Statistical challenges in preprocessing in microarray experiments in cancer. *Clin Cancer Res*, 14(19):5959–5966.
- Paepke S., Waal J. C. D., Gießelmann B., Heywang-Köbrunner S., Kiechle M., Sitttek H., M.Untch und C.Wolf, **2005**. *Mammakarzinom-Prävention*, Seiten 12–27. In Sauer (2005), 10. Edition.
- Paik S., **2007**. Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. *Oncologist*, 12(6):631–635.
- Paik S., Shak S., Tang G., Kim C., Baker J., Cronin M., Baehner F. L., Walker M. G., Watson D., Park T., Hiller W., Fisher E. R., Wickerham D. L., Bryant J. und Wolmark N., **2004**. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*, 351(27):2817–2826.
- Pawitan Y., Bjöhle J., Amler L., Borg A.-L., Egyhazi S., Hall P., Han X., Holmberg L., Huang F., Klaar S., Liu E. T., Miller L., Nordgren H., Ploner A., Sandelin K., Shaw P. M., Smeds J., Skoog L., Wedrén S. und Bergh J., **2005**. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res*, 7(6):R953–R964.
- Perou C. M., Sorlie T., Eisen M. B., van de Rijn M., Jeffrey S. S., Rees C. A., Pollack J. R., Ross D. T., Johnsen H., Akslen L. A., Fluge O., Pergamenschikov A., Williams C., Zhu S. X., Lonning P. E., Borresen-Dale A. L. und Brown P., **2000**. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–752.
- PHGEN, **2009**. Public Health Genomics European Network (PHGEN). <http://www.phgen.nrw.de/>. Letzter Zugriff Dezember 2009.
- Pitroda S. P., Khodarev N. N., Beckett M. A., Kufe D. W. und Weichselbaum R. R., **2009**. MUC1-induced alterations in a lipid metabolic gene network predict response of human breast cancers to tamoxifen treatment. *Proc Natl Acad Sci U S A*, 106(14):5837–5841.

- Pochampalli M. R., Bitler B. G. und Schroeder J. A., **2007**. Transforming growth factor alpha dependent cancer progression is modulated by Muc1. *Cancer Res*, 67(14):6591–6598.
- Polyak K., **2008**. Is breast tumor progression really linear? *Clin Cancer Res*, 14(2):339–341.
- Quackenbush J., **2002**. Microarray data normalization and transformation. *Nat Genet*, 32 Suppl:496–501.
- Reid J. F., Lusa L., Cecco L. D., Coradini D., Veneroni S., Daidone M. G., Gariboldi M. und Pierotti M. A., **2005**. Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst*, 97(12):927–930.
- Reis-Filho J. S. und Tutt A. N. J., **2008**. Triple negative tumours: a critical review. *Histopathology*, 52(1):108–118.
- Reis-Filho J. S., Westbury C. und Pierga J.-Y., **2006**. The impact of expression profiling on prognostic and predictive testing in breast cancer. *J Clin Pathol*, 59(3):225–231.
- Remmele W. und Stegner H., **1987**. Vorschlag zur einheitlichen Definition eines immunreaktiven Scores (IRS) für den immunhistochemischen Östrogenrezeptor- Nachweis (ER-ICA) im Mammkarzinomgewebe. *Pathologie*, 8(3):138–140.
- Renehan A. G., Tyson M., Egger M., Heller R. F. und Zwahlen M., **2008**. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*, 371(9612):569–578.
- Ribeiro-Silva A. und da Costa J. P. O., **2008**. Osteopontin expression according to molecular profile of invasive breast cancer: a clinicopathological and immunohistochemical study. *Int J Biol Markers*, 23(3):154–160.
- RKI, Hrsg., **2008**. *Krebs in Deutschland 2003-2004. Häufigkeiten und Trends*. Robert Koch-Institut und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., Berlin, 6. Edition.
- Rody A., Karn T., Gätje R., Kourtis K., Minckwitz G., Loibl S., Munnes M., Ruckhäberle E., Holtrich U., Kaufmann M. und Ahr A., **2006**. Gene expression profiles of breast cancer obtained from core cut biopsies before neoadjuvant docetaxel, adriamycin, and cyclophosphamide chemotherapy

- correlate with routine prognostic markers and could be used to identify predictive signatures. *Zentralbl Gynakol*, 128(2):76–81.
- Samaan N. A., Buzdar A. U., Aldinger K. A., Schultz P. N., Yang K. P., Romsdahl M. M. und Martin R., **1981**. Estrogen receptor: a prognostic factor in breast cancer. *Cancer*, 47(3):554–560.
- Sauer H., Hrsg., **2005**. *Manual Mammakarzinome: Empfehlungen zur Diagnostik, Therapie und Nachsorge*. Tumorzentrum München, München, 10. Edition.
- Schag C. C., Heinrich R. L. und Ganz P. A., **1984**. Karnofsky performance status revisited: reliability, validity, and guidelines. *J Clin Oncol*, 2(3):187–193.
- Schnakenberg E., Fabig K.-R., Stanulla M., Strobl N., Lustig M., Fabig N. und Schloot W., **2007**. A cross-sectional study of self-reported chemical-related sensitivity is associated with gene variants of drug-metabolizing enzymes. *Environ Health*, 6:6.
- Schulz W., **2005**. *Molecular biology of human cancers. An advanced student's textbook*. Springer, Dordrecht.
- Sekine I., Shimizu C., Nishio K., Saijo N. und Tamura T., **2009**. A literature review of molecular markers predictive of clinical response to cytotoxic chemotherapy in patients with breast cancer. *Int J Clin Oncol*, 14(2):112–119.
- Silva C. M. und Shupnik M. A., **2007**. Integration of steroid and growth factor pathways in breast cancer: focus on signal transducers and activators of transcription and their potential role in resistance. *Mol Endocrinol*, 21(7):1499–1512.
- Silvera S. A. N. und Rohan T. E., **2007**. Trace elements and cancer risk: a review of the epidemiologic evidence. *Cancer Causes Control*, 18(1):7–27.
- Sitteck H., de Waal J. C., Engel J., Hellemann H., Heywang-Köbrunner S., Keßler M., Mahl G., Perlet C., Sommer H., Strauss A., Strigl R. und Untch M., **2005**. *Mammographie-Screening, bildgebende und minimal-invasive Diagnostik*, Seiten 28–46. In Sauer (2005), 10. Edition.
- Slonim D. K., **2002**. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet*, 32 Suppl:502–508.

- Sotiriou C., Khanna C., Jazaeri A. A., Petersen D. und Liu E. T., **2002**. Core biopsies can be used to distinguish differences in expression profiling by cDNA microarrays. *J Mol Diagn*, 4(1):30–36.
- Sotiriou C., Wirapati P., Loi S., Harris A., Fox S., Smeds J., Nordgren H., Farmer P., Praz V., Haibe-Kains B., Desmedt C., Larsimont D., Cardoso F., Peterse H., Nuyten D., Buyse M., de Vijver M. J. V., Bergh J., Piccart M. und Delorenzi M., **2006**. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*, 98(4):262–272.
- Sparano J. A. und Paik S., **2008**. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol*, 26(5):721–728.
- Srinivasan V., Spence D. W., Pandi-Perumal S. R., Trakht I., Esquifino A. I., Cardinali D. P. und Maestroni G. J., **2008**. Melatonin, environmental light, and breast cancer. *Breast Cancer Res Treat*, 108(3):339–350.
- Sørliie T., Perou C. M., Tibshirani R., Aas T., Geisler S., Johnsen H., Hastie T., Eisen M. B., van de Rijn M., Jeffrey S. S., Thorsen T., Quist H., Matese J. C., Brown P. O., Botstein D., Eystein Lonning P. und Børresen-Dale A. L., **2001**. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98(19):10869–10874.
- Sørliie T., Tibshirani R., Parker J., Hastie T., Marron J. S., Nobel A., Deng S., Johnsen H., Pesich R., Geisler S., Demeter J., Perou C. M., Lønning P. E., Brown P. O., Børresen-Dale A.-L. und Botstein D., **2003**. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*, 100(14):8418–8423.
- Sørliie T., Wang Y., Xiao C., Johnsen H., Naume B., Samaha R. R. und Børresen-Dale A.-L., **2006**. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics*, 7:127.
- Stewart B. W. und Kleihues P., Hrg., **2003**. *World Cancer Report*. IARC Press, Lyon.
- Sturn A., Quackenbush J. und Trajanoski Z., **2002**. Genesis: cluster analysis of microarray data. *Bioinformatics*, 18(1):207–208.

- Sun Y., Goodison S., Li J., Liu L. und Farmerie W., **2007**. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics*, 23(1):30–37.
- Sutherland R. L. und Musgrove E. A., **2004**. Cyclins and breast cancer. *J Mammary Gland Biol Neoplasia*, 9(1):95–104.
- Szabo A., Perou C. M., Karaca M., Perreard L., Palais R., Quackenbush J. F. und Bernard P. S., **2004**. Statistical modeling for selecting housekeeper genes. *Genome Biol*, 5(8):R59.
- Thier R., Brüning T., Roos P. H., Rihs H.-P., Golka K., Ko Y. und Bolt H. M., **2003**. Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes. *Int J Hyg Environ Health*, 206(3):149–171.
- Thomas P. D., Campbell M. J., Kejariwal A., Mi H., Karlak B., Daverman R., Diemer K., Muruganujan A. und Narechania A., **2003**. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*, 13(9):2129–2141.
- Thomas P. D., Mi H. und Lewis S., **2007**. Ontology annotation: mapping genomic regions to biological function. *Curr Opin Chem Biol*, 11(1):4–11.
- Tjønneland A., Christensen J., Olsen A., Stripp C., Thomsen B. L., Overvad K., Peeters P. H. M., van Gils C. H., de Mesquita H. B. B., Ocké M. C., Thiebaut A., Fournier A., Clavel-Chapelon F., Berrino F., Palli D., Tumino R., Panico S., Vineis P., Agudo A., Ardanaz E., Martinez-Garcia C., Amiano P., Navarro C., Quirós J. R., Key T. J., Reeves G., Khaw K.-T., Bingham S., Trichopoulou A., Trichopoulos D., Naska A., Nagel G., Chang-Claude J., Boeing H., Lahmann P. H., Manjer J., Wirfält E., Hallmans G., Johansson I., Lund E., Skeie G., Hjartåker A., Ferrari P., Slimani N., Kaaks R. und Riboli E., **2007**. Alcohol intake and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control*, 18(4):361–373.
- Tricarico C., Pinzani P., Bianchi S., Paglierani M., Distante V., Pazzagli M., Bustin S. A. und Orlando C., **2002**. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem*, 309(2):293–300.

- Têtu B., Brisson J., Wang C. S., Lapointe H., Beaudry G., Blanchette C. und Trudel D., **2006**. The influence of MMP-14, TIMP-2 and MMP-2 expression on breast cancer prognosis. *Breast Cancer Res*, 8(3):R28.
- Tuck A. B., Chambers A. F. und Allan A. L., **2007**. Osteopontin overexpression in breast cancer: knowledge gained and possible implications for clinical management. *J Cell Biochem*, 102(4):859–868.
- van de Vijver M. J., He Y. D., van't Veer L. J., Dai H., Hart A. A. M., Voskuil D. W., Schreiber G. J., Peterse J. L., Roberts C., Marton M. J., Parrish M., Atsma D., Witteveen A., Glas A., Delahaye L., van der Velde T., Bartelink H., Rodenhuis S., Rutgers E. T., Friend S. H. und Bernards R., **2002**. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347(25):1999–2009.
- Vandesompele J., Preter K. D., Pattyn F., Poppe B., Roy N. V., Paepe A. D. und Speleman F., **2002**. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 3(7):research0034.
- van't Veer L. J., Dai H., van de Vijver M. J., He Y. D., Hart A. A. M., Mao M., Peterse H. L., van der Kooy K. und Marton M. J. and Witteveen A. T., **2002**. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536.
- Veglia F., Loft S., Matullo G., Peluso M., Munnia A., Perera F., Phillips D. H., Tang D., Autrup H., Raaschou-Nielsen O., Tjønneland A. und Vineis P., **2008**. DNA adducts and cancer risk in prospective studies: a pooled analysis and a meta-analysis. *Carcinogenesis*, 29(5):932–936.
- Veglia F., Matullo G. und Vineis P., **2003**. Bulky DNA adducts and risk of cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*, 12(2):157–160.
- Venkitaraman A. R., **2002**. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*, 108(2):171–182.
- Vogelstein B. und Kinzler K. W., **2004**. Cancer genes and the pathways they control. *Nat Med*, 10(8):789–799.
- Wang X., Chao L., Ma G., Chen L., Tian B., Zang Y. und Sun J., **2008**. Increased expression of osteopontin in patients with triple-negative breast cancer. *Eur J Clin Invest*, 38(6):438–446.

- Wang Y., Klijn J. G. M., Zhang Y., Sieuwerts A. M., Look M. P., Yang F., Talantov D., Timmermans M., van Gelder M. E. M., Yu J., Jatkoe T., Berns E. M. J. J., Atkins D. und Foekens J. A., **2005**. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, 365(9460):671–679.
- WHO, Hrsg., **2002**. *Genomics and World Health*. WHO Advisory Committee on Health Research, Genf.
- Wirapati P., Sotiriou C., Kunkel S., Farmer P., Pradervand S., Haibe-Kains B., Desmedt C., Ignatiadis M., Sengstag T., Schütz F., Goldstein D. R., Piccart M. und Delorenzi M., **2008**. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res*, 10(4):R65.
- Wit E. und MacClure J., **2004**. *Statistics for microarrays*. Wiley, Chichester.
- Yoshida N., Omoto Y., Inoue A., Eguchi H., Kobayashi Y., Kurosumi M., Saji S., Suemasu K., Okazaki T., Nakachi K., Fujita T. und Hayashi S., **2004**. Prediction of prognosis of estrogen receptor-positive breast cancer with combination of selected estrogen-regulated genes. *Cancer Sci*, 95(6):496–502.
- Yoshimura N., Harada N., Bukholm I., Kåresen R., Børresen-Dale A.-L. und Kristensen V. N., **2004**. Intratumoural mRNA expression of genes from the oestradiol metabolic pathway and clinical and histopathological parameters of breast cancer. *Breast Cancer Res*, 6(2):R46–R55.
- Zanetti-Dällenbach R., Vuaroqueaux V., Wight E., Labuhn M., Singer G., Urban P., Eppenberger U., Holzgreve W. und Eppenberger-Castori S., **2006**. Comparison of gene expression profiles in core biopsies and corresponding surgical breast cancer samples. *Breast Cancer Res*, 8(4):R51.
- Zaretsky J., Barnea I., Aylon Y., Gorivodsky M., Wreschner D. und Keydar I., **2006**. MUC1 gene overexpressed in breast cancer: structure and transcriptional activity of the MUC1 promoter and role of estrogen receptor alpha (ERalpha) in regulation of the MUC1 gene expression. *Mol Cancer*, 5(1):57.

Erklärung

Mit der Dissertation mit dem Titel „Genexpressionsanalysen zur Charakterisierung primärer Mammatumoren im neoadjuvanten Setting“ beantrage ich beim Promotionsausschuss der Fakultät für Biologie der Universität Bielefeld die Eröffnung des Promotionsverfahrens nach §5 der Promotionsordnung in der Fassung vom 3. Juni 2002.

Ich versichere, dass ich diese Arbeit selbst angefertigt und nur angegebene Quellen und Hilfsmittel verwendet habe. Alle aus der Literatur entnommenen Zitate sind als solche kenntlich gemacht. Die Arbeit ist gedruckt auf alterungsbeständigem Papier nach DIN ISO 9706.

Weiterhin erkläre ich, dass die hier genannte Dissertation weder vollständig noch in Auszügen einer anderen Fakultät mit dem Ziel vorgelegt worden ist, einen akademischen Titel zu erwerben. Ich bewerbe mich hiermit erstmalig um den Doktorgrad der Naturwissenschaften der Universität Bielefeld.

Birmingham (UK), den 9. Dezember 2009

Anika Tauchen

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die maßgeblich zu der erfolgreichen Fertigstellung dieser Arbeit beigetragen haben.

Zunächst gilt mein Dank Anke Becker, die mich während der gesamten Projektlaufzeit sehr gut betreut hat und mir nicht nur methodisch, sondern auch in Fragen des Projektmanagements immer mit hilfreichen Ratschlägen zur Seite stand. Frau Hornberg möchte ich ganz herzlich für ihre großartige Unterstützung und ganz besonders für die Entbindung von meinen Pflichten danken. Diese „Schonzeit“ hat erheblich zum erfolgreichen Abschluss meiner Arbeit beigetragen. Auch Frau Eichenlaub-Ritter möchte ich an dieser Stelle nochmal recht herzlich für ihre Bereitschaft danken, meine Arbeit zu begutachten.

Ich habe mich während meiner Doktorarbeit in meinen beiden Arbeitsgruppen, der AG7 (Gesundheitswissenschaften) und bei den Exopols (Biologie) sehr wohlfühlt. Allen Kollegen daher ein großer Dank für die gute Zusammenarbeit und die angenehme Zeit! Ganz besonders möchte ich mich bei Birgit, Ella und Eva bedanken. Nicht nur für die absolut besten Labortipps, sondern gerade auch für das besonders schöne „Drumherum“.

Christian Martin möchte ich für die hilfreichen Gespräche und die Bereitstellung und Anpassung des *Tree Index* danken.

Bei unseren Kooperationspartnern der Städtischen Kliniken Bielefeld möchte ich mich besonders bei Prof. Volz und Patricia Faas-Fehervary für die gute Zusammenarbeit bedanken. Ebenso danke ich allen Beteiligten des Klinikums, die durch ihre Bereitschaft zur Mitarbeit unserem Projekt zum Erfolg verholfen haben. Und nicht zuletzt ein herzliches Dankeschön an alle Patientinnen, die ihre Bereitschaft zur Teilnahme an dieser Studie erklärt haben.

Ich freue mich, diesen wichtigen Abschnitt meines Lebens nun hinter mir zu lassen und gemeinsam mit Freunden und Familie neue Wege zu gehen. Ohne Sarah wäre diese Zeit der „schlimmsten Schritte“ wohl nie zu Ende gegangen, was sind wir für ein tolles Team! Marc war mein größter Antrieb, sowohl zu Beginn als auch zum Ende dieser Arbeit. Ich freue mich auf alle weiteren gemeinsamen aufregenden und großartigen Projekte.

Danke!