

## 5 Diskussion

### 5.1 Identifikation differenziell exprimierter Gene

Die Chiparray Technologie ermöglicht durch die zeiteffektive simultane Expressionsanalyse mehrerer tausend Gene, nahezu das gesamte menschliche Genom bei unterschiedlichen Erkrankungen nach differenziell exprimierten Genen zu durchsuchen. Unter Verwendung des Human Genome U133 Set von Affymetrix Inc., welches ca. 45.000 Probensätze umfasst, konnte in dieser Arbeit die Expression von etwa 33.000 bekannten menschlichen Genen gemessen werden. Die Genexpressionsanalyse der korrespondierenden Normal- und Tumorgewebe von 25 Patienten mit kolorektalem Karzinom sämtlicher UICC-Stadien führte zur Identifikation von 267 differenziell exprimierten Sequenzen (DES), deren Expression im Tumorgewebe verglichen mit den korrespondierenden Normalgeweben signifikant über- oder unterexprimiert war. Unter diesen 267 DES befanden sich 168 bekannte Gene und weitere 99 ESTs. Die Literaturrecherche der 168 bekannten differenziell exprimierten Gene ergab, dass 86 Gene bereits in Assoziation mit dem kolorektalen Karzinom oder anderen Tumorentitäten beschrieben worden sind (siehe Tabellen 15 a und b).

**Tabelle 15a**

<b>Gene in Assoziation mit dem kolorektalen Karzinom</b>	<b>Ort und Art der differenziellen Expression</b>
Amphiregulin (schwannoma-derived growth factor, AREG)	überexprimiert in IT
Antigen identified by monoclonal antibody Ki-67	überexprimiert in RT
Centromere protein A (CENPA)	überexprimiert in IT, RT
Claudin 1 (CLDN1, SEMP1)	überexprimiert in IT
c-Met proto-oncogene (hepatocyte growth factor receptor)	überexprimiert in IT, RT
CSE1 chromosome segregation 1-like (CSE1L)	überexprimiert in IT, RT
Cyclin B1 (CCNB1)	überexprimiert in IT, RT
Cyclin-dependent kinase inhibitor 3 (CDKN3)	überexprimiert in RT
Epiregulin (EREG)	überexprimiert in IT
Epithelial stromal interaction 1 (EPSTI1)	überexprimiert in IT, RT
Eukaryotic translation initiation factor 2 subunit 2 (EIF2S2)	überexprimiert in IT, RT
Hyaluronan-mediated motility receptor (RHAMM, HMMR)	überexprimiert in RT
Hypoxia-inducible protein 2 (HIG2)	überexprimiert in RT
MAC30	überexprimiert in IT, RT
Matrix metalloproteinase 12 (Macrophage elastase, MMP12)	überexprimiert in IT
Member RAS oncogene family (RAN)	überexprimiert in RT
Minichromosome maintenance deficient 2 mitotin (MCM2)	überexprimiert in IT
Non-metastatic cells 1 protein (NM23A)	überexprimiert in IT, RT
Nucleosome assembly protein 1-like 1 (NAP1L1)	überexprimiert in RT
Phosphoserine aminotransferase (PSA)	überexprimiert in IT, RT
Pituitary tumor-transforming 1 (PTTG1)	überexprimiert in IT, RT
Ribonucleotide reductase M2 (RRM2)	überexprimiert in RT
S-adenosylhomocysteine hydrolase (AHCY)	überexprimiert in IT, RT
Teratocarcinoma-derived growth factor 1 (TDGF1)	überexprimiert in IT, RT
Transforming growth factor beta-induced (TGFB1)	überexprimiert in IT, RT
Transmembrane 4 superfamily member 1 (TM4SF1)	überexprimiert in IT
Transmembrane prostate androgen induced (TMPEAI, STAG1, PMEPA1)	überexprimiert in IT
B-cell CLL/lymphoma 2 (BCL2)	unterexprimiert in IT
Carbonic anhydrase XII (CA12)	unterexprimiert in IT, RT
Dihydropyrimidine dehydrogenase (DPYD)	unterexprimiert in IT, RT
Fibroblast growth factor receptor 3 (FGFR3)	unterexprimiert in IT, RT

**Tabelle 15a weiter****Gene in Assoziation mit dem kolorektalen Karzinom**

Gene in Assoziation mit dem kolorektalen Karzinom	Ort und Art der differentiellen Expression
Fucosidase alpha-L-1 tissue (FUCA1)	unterexprimiert in RT
Growth hormone receptor (GHR)	unterexprimiert in IT
Mucin 4 (MUC4)	unterexprimiert in IT, RT
Myristoylated alanine-rich protein kinase C substrate (MARCKS)	unterexprimiert in RT
O-acetyltransferase (CAS1)	unterexprimiert in IT
3'-phosphoadenosine 5'-phosphosulfate synthase 2 (PAPSS2)	unterexprimiert in RT
p300/CBP-associated factor (PCAF)	unterexprimiert in IT, RT
Phospholipase A2 group X (PLA2G10)	unterexprimiert in IT, RT
Quinone oxidoreductase homolog (PIG3)	unterexprimiert in IT, RT
Tumor necrosis factor receptor superfamily member 6 (TNFRSF6), (Fas (CD95, Apo-1) protein is encoded by the TNFRSF6 gene)	unterexprimiert in RT
UDP glycosyltransferase 1 family Polypeptide A1 (UGT1A1)	unterexprimiert in IT

**Tabelle 15 b****Gene in Assoziation mit anderen Karzinomentitäten**

Gene in Assoziation mit anderen Karzinomentitäten	Ort und Art der differentiellen Expression
CDC28 protein kinase 1 (CKS1)	überexprimiert in RT
Centromere protein F (CENPF)	überexprimiert in IT, RT
Collagen type I alpha 2 (COL1A2)	überexprimiert in IT, RT
Core-binding factor, subunit beta (CBFB)	überexprimiert in IT, RT
CSE1 chromosome segregation 1-like (CSE1L)	überexprimiert in IT, RT
DNAJ domain-containing (MCJ)	überexprimiert in IT, RT
Hypothetical protein HSPC014, Probensatznummer: 217769_s_at	überexprimiert in IT, RT
Macrophage migration inhibitory factor (MIF)	überexprimiert in IT
Never in mitosis (NIMA) gene related kinase 2 (NEK2)	überexprimiert in IT, RT
Plastin 3 (PLS3)	überexprimiert in IT
Pleckstrin homology-like domain family A, member 1 (PHLDA1), Probensatznummer: 217996_at	überexprimiert in IT
Pleckstrin homology-like domain family A, member 1 (PHLDA1), Probensatznummer: 217997_at	überexprimiert in IT, RT
Prefoldin 4 (PFDN4), Probensatznummer: 205362_s_at	überexprimiert in IT
Prefoldin 4 (PFDN4), Probensatznummer: 205361_s_at	überexprimiert in IT, RT
Protein C receptor endothelial (EPCR, PROCR)	überexprimiert in IT, RT
Ribonuclease P 40kD subunit (RPP40)	überexprimiert in RT
Roundabout axon guidance receptor homolog 1 (ROBO1)	überexprimiert in IT
Serine-threonine kinase 6 (STK6)	überexprimiert in IT, RT
Stearoyl-CoA desaturase (delta-9-desaturase) (SCD), Probensatznummer: 223839_s_at	überexprimiert in RT
Stearoyl-CoA desaturase (delta-9-desaturase) (SCD), Probensatznummer: 200832_s_at	überexprimiert in IT, RT
Topoisomerase II alpha (TOP2A)	überexprimiert in IT
Transketolase (TKT)	überexprimiert in RT
Ubiquitin carboxyl-terminal esterase L3 (Ubiquitin thiolesterase, UCHL3)	überexprimiert in IT, RT
Ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1)	überexprimiert in IT, RT
Amylase alpha 1A (AMY1A)	unterexprimiert in RT
Calpain 9 (CAPN9)	unterexprimiert in IT, RT
Carbonic anhydrase IV (CA4)	unterexprimiert in IT, RT
Carnitine palmitoyltransferase 1 (CPT1B)	unterexprimiert in RT
Endothelial differentiation G-protein-coupled receptor 2 (EDG2)	unterexprimiert in IT, RT
Endothelin 3 (EDN3, ET-3)	unterexprimiert in IT, RT
Four and a half LIM domains 1 (FHL1)	unterexprimiert in IT, RT
Hematopoietic cell-specific Lyn substrate 1 (HCLS1)	unterexprimiert in RT
Hydroxyprostaglandin dehydrogenase (HPGD)	unterexprimiert in RT
Interleukin 1 receptor type II (IL1R2)	unterexprimiert in IT, RT
Metallothionein 1G (MT1G)	unterexprimiert in IT, RT
Metallothionein 1L (MT1L)	unterexprimiert in IT, RT
Metallothionein 1X (MT1X)	unterexprimiert in IT, RT
Myeloid ecotropic viral integration site 1 (MEIS1)	unterexprimiert in IT, RT
Phospholipase C epsilon 1 (PLCE1)	unterexprimiert in IT, RT
Rabconnectin-3 (RC3)	unterexprimiert in RT
RNA binding motif protein 8B (RBM8B)	unterexprimiert in IT, RT
Serum-glucocorticoid regulated kinase (SGK)	unterexprimiert in IT, RT
Sulfotransferase family 1A, member 3 (SULT1A3)	unterexprimiert in RT

**Tabelle 15 a und b** Tabellarische Übersicht über die Gene, die in dieser Arbeit als differentiell exprimiert identifiziert wurden und bereits in der Literatur in Assoziation mit dem kolorektalen Karzinom (a) bzw. anderen Tumorentitäten (b) beschrieben wurden. Einige dieser Gene waren mit verschiedenen Probensätzen auf dem Chip vertreten, die in den unterschiedlichen Tumorarealen zum Teil Expressionsunterschiede aufwiesen. Zu solchen Genen sind zusätzlich die Probensatznummern angegeben.

Auch in anderen Studien von Lin et al., Kitahara et al., Birkenkamp-Demtroder et al., Williams et al., Notterman et al. und Frederiksen et al. wurden bereits mittels der Chiparray Technologie differenziell exprimierte Gene im kolorektalen Karzinom identifiziert (88, 90-92, 106, 107). Sämtliche aufgeführten Studien verwendeten kolorektale Gewebe, die spätestens 30 Min. nach ihrer operativen Entnahme in flüssigem Stickstoff schockgefroren wurden, um die von Huang et al. beschriebene Ischämie-bedingte mRNA-Degeneration zu minimieren (110). Obwohl das methodische Vorgehen der aufgeführten Studien dem dieser Arbeit grundsätzlich ähnelt, existieren wesentliche Unterschiede, die im Folgenden erläutert werden: In der Mehrzahl der Studien, zu denen die von Birkenkamp-Demtroder et al., Williams et al., Notterman et al., und Frederiksen et al. (88, 92, 106, 107) zählen, wurde die Genexpression gesamter Gewebelöcke untersucht, ohne mittels arbeits- und zeitaufwändiger Mikrodisektion gezielt Normal- und Tumorepithelien für die Genexpressionsanalysen aus dem restlichen Bindegewebskomponenten zu extrahieren. Während Notterman et al. und Frederiksen et al. (106, 107) versuchten, die daraus resultierende Beeinflussung der Ergebnisse zu korrigieren, indem sie den generierten Datensatz um einige Muskulatur- und Bindegewebs-assoziierte Genprodukte reduzierten, ließen Birkenkamp-Demtroder et al. und Williams et al. (88, 92) diesen Aspekt bei der Datenauswertung unberücksichtigt. In den Studien von Lin et al. und Kitahara et al., die mikrodiseziertes Material für die Chiparray basierten Genexpressionsanalysen verwendeten, wurden wiederum nur sehr kleine Patientenkollektive (zwischen acht und elf Patienten) untersucht. Dies erschwerte die Datenauswertung, die sich aus statistischer Sicht, bedingt durch die aus wenigen Patientenproben generierten großen Datenmengen, ohnehin als kritisch erwies.

In dieser Arbeit wurde wie auch von anderen Autoren empfohlen (68) mikrodiseziertes Material verwendet und dennoch insgesamt 75 Gewebeproben von 25 Patienten analysiert. Dies ist im Vergleich zu den bisher erwähnten Studien ein größeres Patientenkollektiv.

Zudem wurde in dieser Arbeit im Gegensatz zu sämtlichen anderen erwähnten Arbeiten die von Baisse et al. beschriebene intratumoröse Heterogenität berücksichtigt, indem mittels Laser gestützter Mikrodisektion definierte Tumorareale getrennt voneinander untersucht wurden und deren Genexpression somit anschließend verglichen werden konnte (62). Des Weiteren wurden in der vorliegenden Arbeit wie auch in den Studien von Lin et al., Kitahara et al., Williams et al. und Notterman et al. (68, 90-92, 107) korrespondierende Tumor- und Normalgewebeproben analysiert. Hingegen verwendeten Birkenkamp-Demtroder et al. und

Frederiksen et al. (88, 106) nicht-korrespondierende Gewebe für die Genexpressionsanalysen. Die Genexpressionsanalyse von korrespondierenden Normal- und Tumorgeweben ermöglicht in der späteren Datenauswertung, individuelle Normvarianten der Genexpression rechnerisch zu berücksichtigen. Folgendes Beispiel soll dies veranschaulichen: Das Gen A, welches in keinem Zusammenhang mit der Entstehung und Progression maligner Neoplasien steht, ist ohne pathologische Ursache im Normalgewebe von Patient 1 höher exprimiert als in dem von Patient 2. Analog verhält es sich im Tumorgewebe, auch hier ist das Gen A bei Patient 1 als Normvariante höher exprimiert als bei Patient 2. Wird nun zur Identifikation differenziell exprimierter Gene beispielsweise der Tumor-Normalgewebe-Quotient für das Gen A berechnet, so gibt lediglich der intraindividuelle Vergleich der korrespondierenden Tumor- und Normalgewebe wider, dass das Gen A in den Patienten 1 und 2 weder über- noch unterexprimiert ist. Der interindividuelle Vergleich hingegen würde zum Ergebnis führen, das Gen A sei im Tumorgewebe von Patient 1 höher als im Normalgewebe von Patient 2 exprimiert oder umgekehrt. Selbst der Vergleich des Tumorgewebes mit dem Median oder Mittelwert der Normalgewebe von Patient 1 und 2 könnte zur Fehlinterpretation führen.

Ein Vergleich der von den aufgeführten Studien erarbeiteten Ergebnisse mit denen dieser Arbeit gestaltete sich darüber hinaus schwierig, da unterschiedliche Chip Technologien verwendet wurden. So kamen neben den in dieser Arbeit benutzten Oligonukleotid-Arrays, die auch Birkenkamp-Demtroder et al., Notterman et al. und Frederiksen et al. verwendeten (88, 106, 107), in den Studien von Lin et al., Kitahara et al. und Williams et al. cDNA-Arrays zur Anwendung (90-92). Zusätzlich existieren zu jeder dieser Chip Technologien zahlreiche Ausführungen, die einerseits die Expression verschiedener Sequenzen untersuchen und andererseits die Expression gleicher Sequenzen (Gen oder ESTs) durch unterschiedliche Probensätze messen. In dieser Arbeit wurde mit dem Human Genome U133 Set von Affymetrix Inc. simultan die Expression von ca. 45.000 Probensätzen gemessen, welche ca. 33.000 bekannte menschliche Gene repräsentieren. Damit ist das Human Genome U133 Set von Affymetrix Inc. eines der umfangreichsten Chip Sets, die zum Zeitpunkt der Experimente verfügbar waren.

Eine weitere große Herausforderung stellt das Problem der großen Variationsbreite der Ansätze zur Datenauswertung dar, da die Ergebnisse und deren Interpretation nicht zuletzt von der Auswahl der datenanalytischen Methode abhängen. Trotz der aufgeführten methodischen Unterschiede zwischen dieser und anderen Studien und der datenanalytischen

Schwierigkeiten ließen sich Überschneidungen im Hinblick auf die identifizierten differenziell exprimierten Gene feststellen. Die Tabellen 16 a und b enthalten eine Liste von Genen, die in anderen Publikationen als differenziell exprimiert identifiziert wurden und deren Expression sich in dieser Studie ebenfalls als signifikant unterschiedlich beim Vergleich der Expressionswerte in den Normal- und Tumorgewebe erwies.

**Tabelle 16 a**

Literatur	Gen / EST	Probensatz- nummer	Median ± Standardabweichung (Min- Max)		Wilcoxonstest	
			IT	RT	IT	RT
Birkenkamp- Demtroder et al., Cancer Res, 2002	General transcription factor IIIA (GTF3A)	201338_x_at	2,20 ± 0,87 (0,72-4,00)	2,07 ± 0,91 (0,87-4,40)	<0,0001	<0,0001
	General transcription factor IIIA (GTF3A)	215091_s_at	2,42 ± 0,98 (0,77-4,03)	2,33 ± 0,93 (0,60-4,18)	<0,0001	<0,0001
	Melanoma growth stimulating activity alpha (GRO1 oncogene)	204470_at	1,55 ± 3,36 (0,37-15,8)	1,45 ± 5,50 (0,24-22,27)	0,0064	0,0006
	S-adenosylhomocysteine hydrolase (AHCY)	200903_s_at	2,34 ± 1,77 (0,26-6,51)	2,39 ± 2,03 (0,62- 7,76)	<0,0001	<0,0001
Kitahara et al., Cancer Res, 2001	Interferon induced transmembrane protein 1 (IFITM1)	201601_x_at	4,12 ± 3,45 (0,28-12,00)	4,13 ± 3,44 (0,45-14,82)	<0,0001	<0,0001
	Transforming growth factor, beta-induced (TGFBI)	201506_at	8,35 ± 8,34 (0,06-36,99)	8,00 ± 6,06 (1,02-22,25)	<0,0001	<0,0001
Lin et al., Oncogene, 2002	CD81 antigen (CD81)	200675_at	2,23 ± 1,97 (0,27-7,63)	2,48 ± 1,49 (0,42-5,32)	<0,0001	<0,0001
	DEK oncogene (DEK)	200934_at	1,37 ± 0,45 (0,62-2,43)	1,38 ± 0,51 (0,43-2,57)	0,0002	<0,0001
	High-mobility group protein 2 (HMG2)	208808_s_at	1,25 ± 0,80 (0,52-4,09)	1,25 ± 0,65 (0,43-3,40)	0,0111	0,0005
	High-mobility group protein isoforms I and Y (HMG1Y)	206074_s_at	1,30 ± 0,72 (0,53-3,15)	1,35 ± 0,70 (0,43-3,02)	0,0421	0,0046
	IMP (inosine monophosphate) dehydrogenase 2 (IMPDH2)	201892_s_at	1,67 ± 0,82 (0,64-4,62)	1,43 ± 0,90 (0,49-4,72)	<0,0001	<0,0001
	Macrophage myristoylated alanine-rich C kinase substrate (MACMARCKS)	200644_at	1,49 ± 1,42 (0,19-7,45)	1,31 ± 1,24 (0,30-6,70)	0,0606	0,0172
	Member RAS oncogene family (RAN)	200749_at	1,95 ± 1,53 (0,21-6,81)	1,88 ± 1,27 (0,44-5,48)	<0,0001	<0,0001
	Member RAS oncogene family (RAN)	200750_s_at	2,44 ± 1,97 (0,16-7,25)	2,57 ± 1,30 (0,89-5,74)	0,0001	<0,0001
	Non-metastatic cells 1, protein (NM23A) expressed in (NME1)	201577_at	3,53 ± 2,59 (0,21-9,28)	2,55 ± 2,79 (0,95-10,91)	<0,0001	<0,0001
	Nucleophosmin (nucleolar phosphoprotein B23, numatrin, NPM1)	221691_x_at	3,36 ± 1,92 (0,20-7,48)	2,37 ± 1,65 (0,91-6,65)	<0,0001	<0,0001
Ornithine decarboxylase 1 (ODC1)	200790_at	1,89 ± 0,92 (0,27-4,28)	1,62 ± 0,85 (0,55-3,69)	0,0006	<0,0001	
Notterman et al., Cancer Res, 2002	Cell division cycle 25B (CDC25B)	201853_s_at	1,74 ± 1,86 (0,41-8,31)	1,77 ± 1,96 (0,60-7,82)	0,0002	<0,0001
	General transcription factor IIIA (GTF3A)	201338_x_at	2,20 ± 0,87 (0,72-4,00)	2,07 ± 0,91 (0,87-4,40)	<0,0001	<0,0001
	General transcription factor IIIA (GTF3A)	215091_s_at	2,42 ± 0,98 (0,77-4,03)	2,33 ± 0,93 (0,60-4,18)	<0,0001	<0,0001
	GRO3 oncogene (GRO3)	207850_at	1,90 ± 3,50 (0,45-13,85)	2,01 ± 7,89 (0,54-38,35)	0,0002	<0,0001
Williams et al., Clin Cancer Res, 2003	Baculoviral IAP repeat- containing 5 (survivin, BIRC5)	202095_s_at	1,79 ± 1,60 (0,72-6,96)	1,82 ± 1,94 (0,64-8,40)	<0,0001	<0,0001
	Collagen, type III, alpha 1 (COL3A1)	201852_x_at	1,74 ± 2,38 (0,16-10,57)	1,46 ± 1,29 (0,26-5,38)	0,0297	0,1332
	Collagen, type III, alpha 1 (COL3A1)	215076_s_at	1,60 ± 3,69 (0,29-14,63)	1,76 ± 2,06 (0,22-7,99)	0,0076	0,0818
	High-mobility group protein isoforms I and Y (HMG1Y)	206074_s_at	1,30 ± 0,72 (0,53-3,15)	1,35 ± 0,70 (0,43-3,02)	0,0421	0,0046
	Melanoma growth stimulating activity alpha (GRO1 oncogene)	204470_at	1,55 ± 3,36 (0,37-15,80)	1,45 ± 5,50 (0,24-22,27)	0,0064	0,0006
	Prostate differentiation factor (PLAB)	221577_x_at	2,53 ± 3,53 (0,25-17,89)	2,18 ± 1,98 (0,30-8,82)	0,0001	<0,0001

**Tabelle 16 b**

Literatur	Gen / EST	Probensatz- nummer	Median $\pm$ Standardabweichung (Min- Max)		Wilcoxonstest	
			IT	RT	IT	RT
Birkenkamp- Dentroder et al., Cancer Res, 2002	Carbonic anhydrase IV (CA4)	206209_s_at	0,07 $\pm$ 4,12 (0,02-16,36)	0,07 $\pm$ 0,46 (0,01-1,76)	<0,0001	<0,0001
	D component of complement (Adipsin)	205382_s_at	0,54 $\pm$ 2,51 (0,13-12,17)	0,50 $\pm$ 1,27 (0,13-6,05)	0,0100	0,0002
Kitahara et al., Cancer Res, 2001	Carbonic anhydrase I (CA1)	205950_s_at	0,05 $\pm$ 13,63 (0,02-64,92)	0,05 $\pm$ 0,53 (0,02-1,70)	<0,0001	<0,0001
	Fatty acid binding protein 1(FABP1)	205892_s_at	0,34 $\pm$ 5,49 (0,02-27,38)	0,42 $\pm$ 4,58 (0,03-23,27)	0,0006	<0,0001
	MAX interacting protein 1 (MXI1)	202364_at	0,78 $\pm$ 0,39 (0,07-1,96)	0,73 $\pm$ 0,40 (0,16-1,97)	0,0339	0,0122
Lin et al., Oncogene, 2002	Mothers against decapentaplegic homolog 2 (MADH2)	203075_at	0,97 $\pm$ 0,42 (0,18-1,91)	0,85 $\pm$ 0,34 (0,15-1,48)	0,0152	0,0012
	Rho GTPase activating protein 5 (ARHGAP5)	217936_at	0,94 $\pm$ 0,53 (0,31-3,3)	0,91 $\pm$ 0,44 (0,29-2,50)	0,0681	0,0439
Notterman et al., Cancer Res, 2002	SPARC-like 1 (Mast9, Hevin)	200795_at	0,85 $\pm$ 0,69 (0,12-3,32)	0,55 $\pm$ 0,47 (0,05-2,05)	0,0137	0,0001
Williams et al., Clin Cancer Res, 2003	A disintegrin-like and metalloprotease with thrombospondin type 1, motif 1 (ADAMTS1)	222162_s_at	0,66 $\pm$ 1,10 (0,14-4,50)	0,62 $\pm$ 0,30 (0,15-1,13)	0,0026	<0,0001
	Calpain 2 (CAPN2)	208683_at	0,71 $\pm$ 0,55 (0,23-2,73)	0,55 $\pm$ 0,35 (0,17-1,88)	<0,0001	<0,0001
	Oxysterol binding protein-like 1A (OSBPL1A)	208158_s_at	0,67 $\pm$ 0,45 (0,16-2,39)	0,63 $\pm$ 0,29 (0,16-1,07)	<0,0001	<0,0001
	RNA Helicase-related protein (RNAHP)	210524_x_at	0,20 $\pm$ 3,20 (0,08-10,83)	0,19 $\pm$ 0,38 (0,05-1,44)	0,0001	<0,0001
Zhang et al., Science, 1997	Carbonic anhydrase II (CA2)	209301_at	0,07 $\pm$ 5,25 (0,03-24,12)	0,07 $\pm$ 0,63 (0,03-2,69)	<0,0001	<0,0001
	Cytokeratin 20	213953_at	0,62 $\pm$ 6,84 (0,18-34,86)	0,52 $\pm$ 3,26 (0,15-16,73)	0,0015	0,0002

**Tabelle 16 a und b** Tabellarische Übersicht über Gene, die in anderen Chiparray basierten Studien beim kolorektalen Karzinom als differenziell über- (a) und unterexprimiert (b) identifiziert wurden, und den zu diesen Genen in dieser Arbeit generierten Daten. Einige dieser Gene waren mit verschiedenen Probensätzen auf dem Chip vertreten, daher sind zusätzlich die Probensatznummern angegeben.

Auch die Tatsache, dass von den insgesamt 267 Sequenzen, die in dieser Arbeit als differenziell exprimiert identifiziert wurden, in der Literatur bereits 86 als assoziiert mit dem kolorektalen Karzinom oder anderen Tumorentitäten beschrieben waren (siehe Tabellen 15 a und b), unterstreicht die Plausibilität der Ergebnisse.

## 5.2 Chiparray basierte Klassifikation von Geweben

Ein entscheidendes Ziel von Chiparray basierten Genexpressionsanalysen ist die Identifikation charakteristischer Expressionsprofile, sogenannter Signaturen, oder auch molekularer Profile, die Tumore über die etablierten histopathologischen Merkmale hinaus in prognostisch und therapeutisch relevante molekularbiologische Subgruppen einteilen. Der erste Schritt in diese Richtung besteht in der Entwicklung von Genexpressions basierten

Klassifizierungsverfahren, die Gewebe nach herkömmlichen histopathologischen Parametern einzuteilen vermögen.

In dieser Arbeit gelang es, die 75 kolorektalen Normal- und Tumorgewebe der 25 untersuchten Patienten nach bestimmten Gewebemerkmale zu klassifizieren. Hierzu wurde die Clusteranalyse, ein mathematisches Klassifizierungsverfahren, angewandt, welches einander ähnliche Objekte in Klassen (Cluster) zusammenfasst bzw. sich stark unterscheidende Objekte in verschiedene Klassen gruppiert. Wie es den Abbildungen 14 a bis c, 15, 21 und 27 zu entnehmen ist, teilten sämtliche durchgeführten Clusteranalysen die Gewebeproben der 25 Patienten in zwei deutlich voneinander getrennte Hauptgruppen, wobei die eine überwiegend Tumorgewebeproben (Tumorgewebecluster) enthielt, die andere hingegen vor allem Normalgewebeproben (Normalgewebecluster). Ansonsten ließen sich trotz Anwendung verschiedener Cluster-Algorithmen innerhalb der Tumorgewebecluster keine anderen Tumorstadien- (T-, N-, M-, UICC-Stadien) oder Herkunfts spezifischen (orale bzw. aborale Gewebeproben) Subgruppen identifizieren (siehe Abbildungen 16 bis 19 und 22 bis 25). Die Clusteranalyse der in der Tumorinvasionsfront identifizierten differenziell exprimierten Sequenzen ergab allerdings innerhalb des Normalgewebeclusters eine durch das Dendrogramm separierte Gruppe, welche fünf der insgesamt sechs Gewebeproben enthielt, die aus oral, d.h. proximal des Kolon descendens, gelegenen Dickdarmabschnitten stammten, sowie zwei weitere Gewebeproben aus distal des Kolon transversum gelegenen Darmabschnitten (siehe Abbildung 20). Die innerhalb dieses „Subclusters“ befindlichen zwei Proben aboraler Darmabschnitte stammten von den Patienten 1 und 33. Der Patient 1 litt unter einem im Sigma gelegenen Karzinom, der Patient 33 wies ein hoch sitzendes Rektumkarzinom auf.

Genexpressionsunterschiede zwischen diesen entwicklungsgeschichtlich unterschiedlichen Darmabschnitten sind in der Literatur bereits beschrieben worden, so zeigten z.B. verschiedene Studien, dass Blutgruppenantigene des ABO-Systems im rechten Hemikolon, nie jedoch im linken Hemikolon exprimiert werden, außer in dort befindlichen malignen Neoplasien (121-123).

Die Aussagekraft dieses Subclusters von Normalepithelien aus oral gelegenen Darmabschnitten ist trotz bekannter Genexpressionsunterschiede wegen der kleinen Fallzahl und der insgesamt drei „fehl“-gruppierten Proben sehr fraglich. Zudem gruppierten sich die Normalgewebeproben oral gelegener Dickdarmabschnitte in der beschriebenen Formation

ausschließlich unter Verwendung eines bestimmten Cluster-Algorithmus (Pearson Korrelationseffizient und Complete Link Methode).

Andere klinisch relevante molekulare Signaturen für das kolorektale Karzinom sind aber bereits in Studien beschrieben worden (91, 105-107). So klassifizierten Frederiksen et al. (106) 20 kolorektale Karzinome und fünf kolorektale Normalgewebe anhand ihrer Genexpressionsprofile in Normal- und Tumorgewebe sowie Tumorstadien abhängig in Dukes´B und Dukes´C Karzinome mit einer Fehlerquote < 20 %. Dazu wurden 124 bzw. 200 der insgesamt etwa 6800 gemessenen Sequenzen verwendet. Die entsprechende Klassifikation von Dukes´A und Dukes´D schlug jedoch fehl.

Lin et al. (91) gelang es, neun kolorektale Adenome und weitere elf kolorektale Karzinome basierend auf ihren Expressionsprofilen in zwei separate Gruppen zu klassifizieren. Hierzu wurde eine Auswahl von 771 der insgesamt ca. 23.000 gemessenen Sequenzen eingesetzt.

Notterman et al. (107) erarbeitete anhand der Chiparray generierten Expressionsdaten eine Signatur, die erfolgreich 18 kolorektale Karzinome, vier Adenome und die korrespondierenden Normalgewebe in separate Gruppen eines Dendrogramms einzuteilen vermochte. Diese Signatur umfasste 1096 Gene und ESTs, die von den insgesamt ca. 6000 gemessenen Sequenzen ausgewählt wurden.

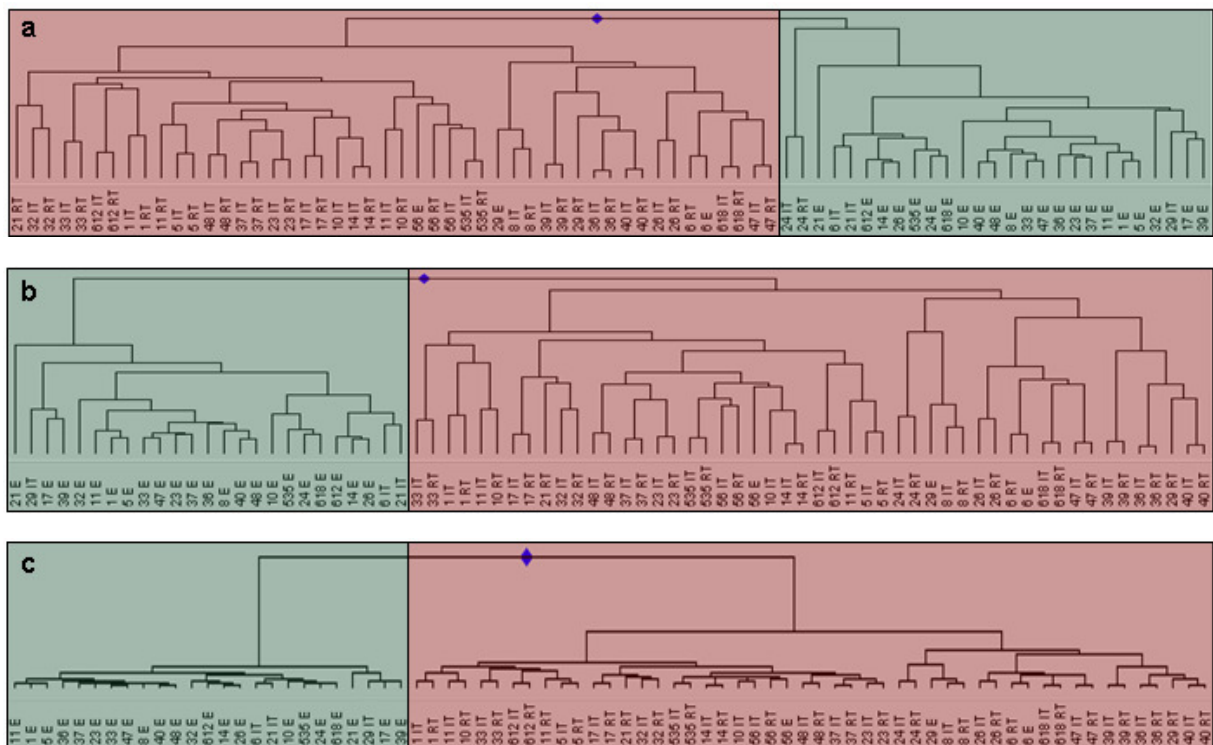
Alon et al. (105) beschrieb in seiner Studie ein Klassifikationsverfahren, welches 40 kolorektale Karzinome, 22 kolorektale Normalgewebe und darüber hinaus Zelllinien sowie „in-vivo“ Gewebe in separate Gruppen einteilte.

Molekulare Signaturen sind jedoch kritisch zu hinterfragen, da ein grundsätzlich schwer zu beherrschendes Problem sämtlicher Chiparray basierter Genexpressionsstudien der Umfang der untersuchten Patientenkollektive darstellt, der durch die hohen Kosten der Chiparray Experimente stark limitiert ist. Insbesondere bei der Suche nach Eigenschafts spezifischen charakteristischen Expressionsprofilen sind die im Verhältnis zu der Menge der generierten Daten geringen Fallzahlen problematisch, da durch beliebige Datenreduktion nahezu für jedes Gewebemerkmale eine spezifische diskriminierende molekulare Signatur identifiziert werden kann. Es ist daher von zentraler Bedeutung, sich dieses mathematischen Phänomens, welches auch als „*Overfitting*“ bezeichnet wird, bei der Datenauswertung bewusst zu sein, so dass nur solche Schritte zur Datenreduktion vorgenommen werden, die auch aus medizinischer und biologischer Sicht sinnvoll erscheinen. Für die Aussagekraft der in dieser Arbeit identifizierten charakteristischen Expressionsprofile, die Gewebe anhand der



identifizierten DES entsprechend ihrer Dignität in Tumor- und Normalgewebegruppen einteilen, spricht, dass die Clusteranalyse sämtlicher gemessener Probensätze, die in 80 % der untersuchten Proben detektiert waren, nämlich 11.903 Probensätze, ebenfalls zur Aufteilung der Proben in die zwei beschriebenen Hauptgruppen führt (siehe Abbildung 14 a bis c). Es kann also davon ausgegangen werden, dass diese durchgeführte Datenreduktion keine folgenschwere Verzerrungen der Daten verursachte.

Neben dem statistischen Missverhältnis zwischen den untersuchten Proben und den gemessenen Merkmalen birgt auch die große Variationsbreite, die Chiparray generierten Expressionsdaten zu verarbeiten und zu analysieren, ein nicht zu vernachlässigendes Problem, da beispielsweise die Ergebnisse einer Clusteranalyse sowohl von der Methode der Datennormalisierung als auch von der Auswahl der Cluster-Algorithmen abhängt. Für die Gültigkeit der in dieser Arbeit identifizierten Signaturen spricht, dass sie die kolorektalen Gewebe weitgehend unabhängig von dem angewandten Cluster-Algorithmus in Normal- und Tumorgewebecluster einteilen, das heißt, mit nur unerheblich veränderter Trennschärfe bei Anwendung unterschiedlicher Algorithmen. Ein Vergleich der Clusteranalyseergebnisse unter Anwendung verschiedener Cluster-Algorithmen für die 11.903 detektierten Sequenzen ist Abbildung 14 a bis c zu entnehmen. Die Abbildung 28 a bis c hingegen liefert den Vergleich der Clusteranalyseergebnisse der identifizierten DES unter Verwendung verschiedener Cluster-Algorithmen. Aus den genannten Abbildungen geht hervor, dass die unterschiedlichen Cluster-Algorithmen in geringem Maße die Trennschärfe sowie die Größe der gebildeten Subgruppen beeinflussen. Die Bildung der zwei Hauptgruppen, das Normal- und das Tumorgewebecluster, zeigte sich jedoch unabhängig vom angewendeten Cluster-Algorithmus stabil. Anders verhielt es sich jedoch mit der oben erwähnten Signatur zur Einteilung der kolorektalen Gewebe nach oral und aboral gelegener Lokalisation. Diese Einteilung der Patientenproben gelang nur unter Verwendung eines bestimmten Algorithmus (Pearson Korrelationseffizient und Complete Link Methode).



**Abbildung 28** Dendrogramme der differentiellen Expressionsprofile der Normal- (E) und der Tumorgewebeproben (IT und RT) erstellt durch die Clusteranalyse mit den 267 identifizierten differenziell exprimierten Sequenzen aus der Tumordinvasionfront (IT) und der zentralen Tumorregion (RT). Die Normalgewebecluster sind grün, die Tumorgewebecluster rot hinterlegt. Die Clusteranalyseergebnisse unterscheiden sich lediglich durch die Anwendung unterschiedlicher Cluster-Algorithmen: Pearson Korrelationseffizienten und Complete Link Methode (a), Pearson Korrelationseffizienten und Wards Methode (b), Euklidische Distanz und Wards Methode (c).

Unter Berücksichtigung der aufgeführten Methoden-kritischen Aspekte von Chiparray basierten Genexpressionsanalysen stellt sich insbesondere dann die Frage, wie vertrauenswürdig die Ergebnisse der Clusteranalysen sind, wenn es um die Beurteilung von „Ausreißern“ geht. So umfassten Cluster neben den Proben, die das Cluster definieren, auch Proben, die sich erwartungsgemäß in ein anderes Cluster hätten gruppieren müssen. Es ordneten sich beispielsweise bei der Clusteranalyse der differenziell exprimierten Sequenzen aus der Tumordinvasionsfront die Tumorgewebeproben 6 IT und 21 IT in das Normalgewebecluster siehe Abbildung 15. Nun ist zu überlegen, ob dieser Umstand auf eine besondere Eigenschaft dieser Proben zurückzuführen ist, oder es sich lediglich um den versehentlichen Tausch der Probenlabel handelte. Die Tumorgewebeprobe 6 IT stammt von einem Patienten mit einem T2N0M0

(UICC-Stadium I) Karzinom, welches sich als mäßig differenziert (G2) zeigte. Der Patient, von dem die Probe 21 IT stammte, wies ein mäßig differenziertes (G2) T3N0M1 (UICC-Stadium IV) Karzinom auf. Ansonsten waren zu keiner der beiden Proben histologische Auffälligkeiten, wie beispielsweise adenomatöse Herde oder ausgeprägte lymphozytäre Infiltration, beschrieben worden. Von klinisch-histologischer Seite her lässt sich demnach die unerwartete Cluster Zuordnung nicht erklären, zumal sich die Tumorproben der zentralen Tumorregionen, 6 RT und 21 RT, erwartungsgemäß ins Tumorgewebecluster einordneten (siehe Abbildung 21). Für die kolorektalen Proben des Patienten 6 ergab sich nun folgende auffällige Konstellation: Die Normalgewebeprobe 6 E gruppierte sich in Abbildung 21 ins Tumorgewebecluster, und zwar direkt neben die Probe 6 RT, was eigentlich von der Probe 6 IT zu erwarten gewesen wäre. Dies spricht also eher für eine unbeabsichtigte Vertauschung der Probenlabel 6 E gegen 6 IT, da es recht unwahrscheinlich scheint, dass das Tumorgewebe (6 IT) bezüglich der Genexpression dem Normalgewebe der anderen untersuchten Patienten ähnlicher ist, als den anderen Tumorgeweben, gleichzeitig aber das Normalgewebe (6 E) wiederum den Tumorgeweben der anderen Patienten stärker gleicht. Für die Annahme der vertauschten Label spricht des weiteren, dass sich in Abbildung 27 die Tumorprobe 6 IT ins grün hinterlegte Normalgewebecluster gruppierte, während sich abermals die Normalgewebeprobe 6 E direkt neben der korrespondierenden Tumorprobe aus der zentralen Tumorregion 6 RT innerhalb des rot hinterlegten Tumorgewebeclusters einordnete. Die Recherche der Laborbücher lieferte allerdings, bis auf dass die jeweils korrespondierenden Proben 6 E, 6 IT und 6 RT gemeinsam isoliert und aufbereitet wurden, keinen weiteren Anhalt für eine Verwechslung der Proben 6 IT und 6 N.

Für die Ausreißer 29 IT und 29 N, den Proben eines Patienten mit einem mäßig differenzierten (G2) T3N1M0 (UICC-Stadium III) Karzinom, zeigte sich zunächst eine ähnliche Fehlverteilungskonstellation: In Abbildung 15 gruppierte sich die Normalgewebeprobe 29 E ins Tumorgewebecluster, umgekehrt ordnete sich die Tumorgewebeprobe 29 IT ins Normalgewebecluster. Die Abbildung 21 zeigt zudem, dass sich die Tumorgewebeprobe 29 RT erwartungsgemäß ins Tumorgewebecluster einordnete, während sich die Normalgewebeprobe 29 E abermals ins Tumorgewebecluster eingliederte. Zudem zeigt das Dendrogramm aus Abbildung 27, dass sich die Proben 29 RT und 29 E innerhalb des Tumorgewebeclusters recht dicht, allerdings nicht direkt nebeneinander einsortierten, wohingegen sich die Probe 29 IT ins Normalgewebecluster einordnete. Auch

die Gewebeproben 29 E, 29 IT, 29 RT sind gemeinsam isoliert und aufgearbeitet worden. Zusammenfassend ist für die Ausreißer 29 IT und 29 E eine Verwechslung der Probenlabel 29 E und 29 IT ebenfalls möglich.

### 5.3 Intratumoröse Genexpressionsunterschiede

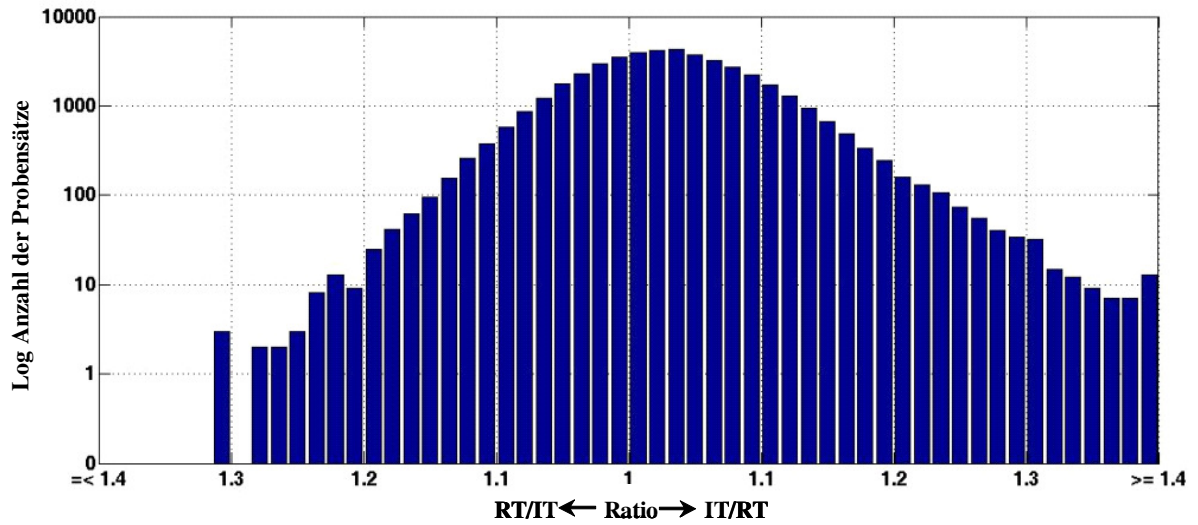
In dieser Arbeit wurde die von Baisse et al. für das kolorektale Karzinom beschriebene Heterogenität der Genexpression innerhalb eines Tumors berücksichtigt, indem die Genexpression in unterschiedlichen Tumorarealen eines kolorektalen Tumorpräparates getrennt voneinander untersucht wurde (62). Hierzu wurden die Epithelien definierter Tumorregionen, nämlich der Tumordinvasionsfront, welche ans Normalgewebe angrenzt, und zentral gelegener Tumorgebiete, mittels Laser gestützter Mikrodissektion separat aus dem restlichen Tumorgewebe entnommen (siehe Abbildung 7).

Betrachtet man nun die Expressionunterschiede zwischen diesen zwei Tumorarealen, so fanden sich von den 267 identifizierten differenziell exprimierten Sequenzen 133 in beiden Tumorregionen wieder, während sich 134 Sequenzen nur in einer der beiden Tumorregionen als differenziell exprimiert erwiesen (siehe Tabelle 8). Auf den ersten Blick scheint aufgrund der nur zu etwa 50 % (133/267) übereinstimmenden Genexpression die intratumoröse Heterogenität unerwartet groß. Eine nicht unerhebliche Ursache hierfür liegt in der ausgewählten Methode, die Expressionsdaten zu verarbeiten und zu analysieren. So wurden in dieser Arbeit die Bedingungen für die Definition „differenziell exprimierte Sequenz“ recht streng festgelegt, was jedoch aufgrund der Fülle der generierten Daten erforderlich war. So war es beispielsweise möglich, dass ein Gen bzw. ein EST, welches in der Tumordinvasionsfront als differenziell überexprimiert identifiziert wurde, sich ebenfalls in zentral gelegenen Tumorregionen als gegenüber dem Normalgewebe signifikant überexprimiert erwies, aber der entsprechende Probensatz in weniger als 80 % der untersuchten Proben detektiert war und somit die Bedingungen für die Definition „differenziell exprimierte Sequenz“ in dieser Arbeit nicht erfüllte. Eine andere datenanalytische Ursache für das eben geschilderte Phänomen könnte darin bestehen, dass eine differenziell überexprimierte Sequenz der Tumordinvasionsfront zwar auch in den zentral gelegenen Tumorregionen signifikant überexprimiert war, aber in der Rangfolge, die für die

unterschiedlichen Tumorregionen separat generiert wurde, oberhalb des arbiträr gewählten „*Cut-offs*“ der besten 100 Rangfolgenplätze lag.

Zur besseren Beurteilung der intratumorösen Heterogenität der Genexpression erfolgte die Auswertung der Clusteranalysen. Hierbei zeigte sich nun, dass sich die untersuchten Tumorregionen tatsächlich stärker bezüglich ihrer Genexpression ähneln als es zunächst erschien. So ergab die Clusteranalyse unter Verwendung der 267 identifizierten differenziell exprimierten Sequenzen das in Abbildung 27 dargestellte Dendrogramm. Das Dendrogramm teilte die 75 untersuchten Proben in zwei Hauptgruppen auf, wobei die eine Hauptgruppe (Tumorgewebecluster) 70 der insgesamt 75 Tumorgewebeproben enthielt, die andere (Normalgewebecluster) hingegen 22 der insgesamt 25 Normalgewebeproben. Innerhalb des Tumorgewebeclusters fiel auf, dass die Proben der Tumorinvasionsfront (IT) und der zentralen Tumorregionen (RT) keine eigenständigen Subgruppen bildeten, sondern dass sich die unterschiedlichen Tumorareale des gleichen Patienten in 20 der 25 Patienten direkt nebeneinander einordneten. Ein ähnliches Resultat allerdings mit reduzierter Normalgewebe-Tumorgewebe-Trennschärfe, ergab die Clusteranalyse sämtlicher Sequenzen, deren Probensätze in mindestens 80 % der insgesamt 75 untersuchten Proben detektiert waren (siehe Abbildung 14 a bis c). Hier gliederten sich sogar bei 23 der 25 Patienten die IT und RT Proben des gleichen Patienten innerhalb des Tumorgewebeclusters direkt nebeneinander.

Dies verstärkt die Annahme, dass sich unterschiedliche Tumorregionen eines Tumors bezüglich der Genexpression trotz bestehender Unterschiede, wie der Vergleich der identifizierten differenziell exprimierten Gene ergab, stärker ähneln als die gleichen Tumorregionen im interindividuellen Vergleich. Ein weiterer Anhalt hierfür bietet Abbildung 29, aus der hervorgeht, dass die meisten Sequenzen in IT und RT etwa gleich stark exprimiert sind.



**Abbildung 29** Vergleich der Expressionswerte von Probensätzen in der Tumordinvasionsfront (IT) und zentraler Tumoranteile (RT). Auf der x-Achse sind die Quotienten der Expressionswerte der Tumordinvasionsfront (IT) und zentraler Tumorregionen (RT) abzulesen, auf der y-Achse, die logarithmierte Anzahl der Probensätze, das heißt der Gene und ESTs, die den definierten Tumorgewebequotienten aufweisen. Die Mehrzahl der Sequenzen sind im Tumor- und Normalgewebe gleich oder ähnlich exprimiert, so dass der Tumor-Normalgewebe-Quotient etwa 1 beträgt.