divided into a *"cured"* and on the other hand a " *fatal and refractor* " group. The Kaplan-Meyer survival analysis revealed a significantly better 5-year-overall-survival rate for the *"cured"* group with 70% in comparison to the *"fatal and refractor* " group with 12% (P=0.00004). Based on these results their model was validated in the dataset from the previously mentioned study. A set of 90 genes was represented in their cDNA microarray as well as in the oligonucleotide microarray (*lymphoma-chip*) employed by Alizedah. Both patient groups could be divided in the two *"cell of origin"* groups classified by Alizedah using this 90 gene set. This *"cell of origin"* differential analysis distinction was strongly associated with clinical outcome in the dataset of Alizedah but not in the actual patient group of DLBCL. Further the predictive power of their own 13 gene based "predictor" was tested in the dataset of Alizedah. Three of these genes where represented in the *"lymphoma-chip"* of Alizedah NOR1, PDB4B and PKCβ. Significant correlation with outcome was found for NOR1(p=0,05) and PDB4B (p=0,07). Results for PKCβ were discordant by representing multiple cDNAs on the *"lymphoma-chip"*. These results suggest a significant advantage for a "supervised learning" method for the prediction of clinical outcome of disease in comparison to the "unsupervised hierarchical clustering". The same dataset of Alizedah was reanalysed by a Japanese group [Ando 2002]. "Fuzzy Neural Network^{3}" as a new statistical method to analyse prediction power of gene expression was applied in their investigation. Their model identified four genes (CD 10, AA807551, AA805611 and IRF-4) that could be used to predict prognosis with a 93% accuracy.

In 2003 a Japanese, cooperative research group published a study in Primary Central Nervous System Lymphomas (PCNL) using Filter-Array Assays [Yamanaka 2003]. Among 21 brain tumor and normal brain tissue samples six PCNLs could be clearly distinguished by "hierarchical clustering" "Fig (4)". The genes encoding for Laminin-receptor-2, thioredoxin-peroxidase and elongation-factor-1 were selected by Principal Component Analysis (PCA) as genes specific for

PCNSL. The gene expression profiles of the six PCNL samples were correlated with the clinical outcome of the corresponding patients. These six patients could be distinguished into groups on the basis of post-treatment survival, a parameter likely related to response to therapy: group I >24 month (3 cases); group II < 23month (3 cases). All six patients were treated uniformly with the same chemo-radiotherapy regimen. A set of 76 genes capable of distinguishing the treatment sensitive group from the non-sensitive to treatment group were selected by the Whitney-Mann test. Among these 76 genes, 37 genes were found up-regulated and 39 genes were found down-regulated in responders. Interestingly, ten of the 37 up-regulated genes were involved in angiogenesis, while six of the 39 down-regulated genes were involved in apoptosis. Using the selected genes related to response to chemo-radiotherapy, re-clustering was performed. The responders and non-responders could successfully be separated on the basis of subtle, differences in distribution of gene expression "Fig.(5)".This study represents another example of how genomic techniques may improve evaluation of complex neoplastic disease.

Reliability and reproducibility of array data

These examples of investigations with micrarray technique surely present landmark studies in their individual fields. They have brought a realistic hope to the scientific community that long time open questions will be answered within the near future. But besides all well founded hope and enthusiasm it cannot be overlooked that analysing and interpreting array data remain a rather complex technology. Applying these novel genomic techniques uncritically to clinical data sets and clinical trials could led to potentially problematic results and conclusions. It has to be critically taken into account that these novel techniques are still experimental. The question of validation and reproducibility remains still a major issue. A number of studies regarding these

technical and statistical problems have been reported within recent years. We will try to summarize the most influential ones of these in the following paragraph.

A fundamental problem for comparing gene expression profiles from different populations or groups (e.g. normal/disease) are large variations between individuals within the same population. It is difficult to distinguish differences in gene expression that appear associated with a specific disease from random genetic variations. Oleksiak has performed a landmark study regarding this important issue for all gene expression based studies [Oleksiak 2003]. Gene expression profiles within and among populations of the teleost fish of the genus Fundulus were analyzed in this investigation. Statistically significant differences in expression profiles between individuals within the same population for approximately 18% of 907 genes were observed. Typically, expression differed by a factor of 1.5 and often even more than a factor of 2.0. In addition to that Enard et al. found in global comparisons of mRNA-levels of chimpanzee and human brain tissue greater variations within the human population than between the human and the chimpanzee population [Enard 2002]. Both studies point out the importance to recognize the large variations between individuals within a population in study design as well as in the appropriately selected statistical analysis.

Our group has focussed largely on relevant questions of gene-profiling practice within the recent years. From small tumor samples often only a small amount of RNA can be obtained. In some cases this amount is not enough to perform gene profiling assays. Amplification of RNA to a.cRNA (amplificated RNA) is one common way to enlarge the given RNA amount. Possible artificial changes of gene expression influenced by amplification have not yet been examined in detail. A two step amplification method was used in our validation experiment "Fig. (6)". By this approach 10-100 µg amplified cRNA from a small amount of total RNA (1µg or less) could be obtained .Than we focussed on the question whether differential gene expression is conserved

after amplification. The differential gene expression profiles of the PC 14 cell line and a sample of peripheral blood lymphocytes (PBL) were compared using mRNA and after amplification a.cRNA. Although the R-Ratio was lower, we could conserve significant differences in the gene expression profiles after amplification "Fig. (7)" These preliminary results suggest that a gene-profiling study could be based on only small samples with a small amount of RNA if an appropriate amplification would be performed. Another promising application of genomic techniques is to observe time or dose dependent changes of gene expression profiles in tumor tissue under the influence of a given drug application. However, repeated tumor sampling is necessary. This remains a very encumbering approach for the patient and often not possible in clinical every day practice. Therefore, we have examined, if a more easier to be performed method to obtain peripheral blood lymphocytes might be useful as a method to identify surrogate tissue for observing drug related changes in gene expression profiles. Within a clinical phase-I study with a novel Farnesyl Transferse Inhibitor we collected tumor samples and peripheral blood lymphocytes predose and on days two and eight following drug application "Fig. (8)".A cDNA filter-array assay including 775 genes chosen for predicting chemosensitivity was used for analysing gene expression profiles "Fig. (9)". Interestingly, changes in gene expression were not only observed in tumor sample but also in the PBL. In still ongoing clinical research we are currently trying to determine the role of PBL as surrogate tissue in pharmacogenomic cancer research.

In 2002, Churchill presented a basic review article about the fundamentals of experimental design for cDNA microaarays [Churchill 2002].The appropriate design of a microarray experiment is essential for the scientifically based interpretation of the results. He pointed out the importance to analyze an adequately high number of biological samples for to achieve representative, predictive and validated results. A higher number of technical replication with the

same biological sample could not lead to validation of the results in most cases. Although the optimal design of an experiment or a study is the basis for successful results the appropriate statistical analysis of the obtained data turns out to be of further importance to. An inadequate data analysis can lead to potential pitfalls. As previously shown, the lymphoma data set of Alizedah is analyzed by different groups with different statistical methods thus leading to partially different or even conflicting results. It is most important to recognize, that different purposes of studies require different methods of statistical analyses. For example, the commonly used "unsupervised hierarchical clustering" although useful for discovery subsets in a number of tumor samples within the same histological group is not appropriate to compare these amongst each other or establish a meaningful "predictor".[Simon 2003]. In 2001, Tusher et al. published a new method for analysing microarray data [Tusher 2001]. In their investigation, they focussed on the problem to identify significant changes in gene expression profiles between different functional biological states. Cluster analysis provides only little information about statistical significance and conventional *t* test is not appropriate for the thousands of data obtained within these microarray experiments. This problem led them to develop a statistical method adapted specifically for microarray analysis. This "Significance Analysis of Microarrays (SAM)" assign a score to each gene on the basis of gene expression relative to the standard deviation of repeated measurements. This method was used in the breast cancer study of Sorlie [Sorlie 2001]. Another important challenge is the integration of microarray data generated by different research groups on different array platforms. Moreau has currently summarized three major problems: (1) the efficient access and exchange of microarray data; (2) the validation and comparison of data from different platforms (cDNA and short long oligonucleotides); and (3) the integrated statiscal analysis of data sets [Moreau 2003]. Tan has reported a considerable divergence of results from three different commercial available microarray platforms analyzing the same RNA sample [Tan

2003]. The most common application of microarray technology is the prediction of clinical outcome in cancer. The most important reports have been referred to within in the first part of our review. Nitzani and Ioannidis systematically analysed studies correlating outcome with genetic profiles based on microarray data published from 1995-2003 [Nitzani 2003] (Table 8). They concluded that the predictive performance of this new technique was variable and in many cases molecular classifications were not subject to an appropriate validation. Of note is, that they found out that only in 30% of the studies with major clinical implication, an appropriate cross-validation or independent validation check was performed.

Another substantial open issue in the proceedings of DNA-microarray techniques in translational cancer research is the lack of information of the concrete biological function of encoding proteins. Most investigators have been validated their DNA-microarray results by Real-Time-PCR [Chaqui 2002] Although we could measure the level of expression of the genes of interest in a reliable way, we miss information of the postranslational protein modifications, time course of protein expression, conditions of protein synthesis, cellular location of the protein, activation of the protein and interaction with other molecules. Therefore, more and more authors combine in their investigations DNA-techniques as DNA microarray and Real-Time-PCR with Non-DNA techniques such as tissue-array, immunohistochemistry and westernblotting. White and co-workers published a remarkable study focused on the correlation between mRNA and protein expression [White 2004]. This British group performed a microarray analysis to compare transcription in response to the ErbB-2 receptor tyrosine kinase activity in a model of a mammary luminal epithel cell system. They compared the differences of mRNA expression with changes at protein level using a parallel proteomic strategy employing two-dimensional difference gel electrophoresis (2-D-DIGE) and quantification of multiple immunoblotting experiments. Interestingly, they found a high correlation between transcription and translation for the subset of

genes studied. Moustafa and colleagues include immunoblotting in the validation of their DNA microarray experiment [Moustafa 2002]. To identify genes involved in head and neck cancer they compared the gene expression profile in matched primary normal epithelial cells and primary head and neck cancer cells from the same patients employing a cDNA microarray consisting of 12530 genes human genes. They found significant changes in the expression of 213 genes. 91 genes were found up-regulated and 122 down-regulated in the cancer-cells. In general, most of the genes that are over-expressed in the head and neck cancer cells encode for growth factors and cell structure. The under-expressed genes are involved in cell-cell adhesion and motility, apoptosis and metabolism. To validate their results at protein level they investigated the expression of nine selected genes from the cell-cell adhesion and motility group by immunoblotting and Reverse Transcriptase-PCR. They found in three of the four cell line pairs consisting results of DNA-micrarray, Reverse Transcriptase–PCR and immunoblotting. However, in one sample they found conflicting results between the protein and the mRNA expression of E-cadherin and γ-catenin. This differences may be explained by differing rates of translation or protein stability in the cancer cell versus their normal counterparts. Tissue microarrays (TMA) are an promising approach in validation of DNA-micrarray results [Chaqui 2002, Hao 2004, Mousses 2002]. A TMA is a slide with dozens to hundred predefined microscopic sections of tissue. This makes it feasible for an investigator to measure DNA, mRNA and protein expression in a large number of samples, providing enough statistical power for meaningful analysis. Immunohistochemistry is the most common method applied to TMAs, but *in situ hybridisation* is increasingly used. In spite of many clear advantages for TMAs in the validation of microarray results this technique is not without any limitations. The critical issues involve sensitivity and lack of quantification.

Summary

In summary, these new techniques will play an important role in future translational cancer research. However, a consequent and critical evaluation is urgently needed. An internationally commonly accepted standardisation must be established. Public microarray databases should allow critical comparisons of independent experience within the same malignant clinical entity. [Stoeckert 2002]. Published investigations should in detail report all key features of the experimental design, the samples used, the extract preparation and labelling performed, hybridisation procedures and variables employed, measurement data and specifications generated. Future studies should generally be performed on the basis of the recommendations proposed by the Microarray Gene Expression Database Group (MEGD) [Brazma 2001].

In spite of many issues to be solved genomic techniques have taken translational cancer research a significant step forward. In some lymphoma and solid tumors more detailed and biologically relevant risk classifications could be developed using these novel techniques. For several anticancer agents significant knowledge about mechanisms of action and resistance could be gained. As a consequence to this, genomic techniques are awaited to become the backbone of translational cancer research in the future.

References

Alizadeh, A. A.; Eisen, M. B.; Davis, R. E.; Ma, C.; Lossos, I. S.; Rosenwald, A.; Boldrick, JC.; Sabet, H.; Tran, T.; Yu, X.; Powell, JI.; Yang, L.; Marti, GE.; Moore, T.; Hudson, J. Jr.; Lu, L.; Lewis, DB.; Tibshirani, R.; Sherlock, G.; Chan, WC.; Greiner, TC.; Weisenburger, DD.; Armitage, JO.; Warnke, R.; Levy, R.; Wilson, W.; Grever, MR.; Byrd, JC.; Botstein, D.; Brown, PO. and Staudt, LM. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature **403**, 503-511.

Al Moustafa AE, Aloui- Jamali MA, Batist G, Hernandez-Perez M, Serruya C, Alpert L, Black MJ, Sladek R, Foulkes WD. (2002) Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous cell carcinoma cells. Oncogene **21** (17):2634-40.

Ando, T.; Suguro, M.; Hanai, T.; Kobayashi, T.; Honda, H. and Seto, M. (2002) Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma. Jpn. J. Cancer Res. **93**, 1207-1212.

Beer, DG.; Kardia, SL.; Huang, CC.; Giordano, TJ.; Levin, AM.; Misek, DE.; Lin, L.; Chen, G.; Gharib, TG.; Thomas, DG.; Lizyness, ML.; Kuick, R.; Hayasaka, S.; Taylor, JM.; Iannettoni, MD.; Orringer, MB. and Hanash, S. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat. Med. **8**, 816-824.

Bonenkamp, JJ.; Hermans, J.; Sasako, M.; and van de Velde, C. J. (1999) Extended lymph-node dissection for gastric cancer. Dutch Gastric Cancer Group. N. Engl. J. Med. **340**, 908-914.

Brazma, A.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, CA.; Causton, HC.; Gaasterland, T.; Glenisson, P.; Holstege, FC.; Kim, IF.; Markowitz, V.; Matese, JC.; Parkinson, H.; Robinson, A.; Sarkans, U.; Schulze-Kremer, S.; Stewart, J.; Taylor, R.; Vilo, J.and Vingron, M. (2001) Minimum information about a

microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. **29**, 365-371.

Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangria MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR. (2002) Postanalysis follow-up and validation of microarray experiments. Nat Genet. **32** Suppl: 509-14.

Le Chevalier, T. for the IALT Investigators (2003): Results of the Randomized International Adjuvant Lung Cancer Trial (IALT): cisplatin-based chemotherapy (CT) vs no CT in 1867 patients (pts) with resected non-small cell lung cancer (NSCLC). Proc. Am. Soc. Clin. Oncol. **22**, page 2, (abstr 6)

Churchill GA.(2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet. **32**, Suppl:490-495.

Cuschieri, A.; Weeden, S.; Fielding, J.; Bancewicz, J.; Craven, J.; Joypaul, V.; Sydes, M. and Fayers, P. (1999) Patient survival after D1 and D2 resections for gastric cancer: long-term results of the MRC randomized surgical trial. Surgical Co-operative Group. Br. J. Cancer **79**, 1522-1530.

Enard, W.; Khaitovich, P.; Klose, J.; Zollner, S.; Heissig, F.; Giavalisco, P.; Nieselt-Struwe, K.; Muchmore, E.; Varki, A.; Ravid, R.; Doxiadis, GM.; Bontrop, RE. and Paabo, S. (2002) Intra- and interspecific variation in primate gene expression patterns. Science **296**, 340-343.

Gruvberger, S.; Ringner, M.; Chen, Y.; Panavally, S.; Saal, LH.; Borg, A.; Ferno, M.; Peterson, C. and Meltzer, PS. (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. **61**, 5979-5984.

Hao X, Sun B, Hu L, Lahdesmaki H, Dunmire V, Feng Y, Zhang SW, Wang H, Wu C, Wang H, Fuller GN, Symmans WF, Shmulevich I, Zhang W. (2004) Differential gene and protein

expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. Cancer 100 (6):1110-22

Hasegawa, S.; Furukawa, Y.; Li, M.; Satoh, S.; Kato, T.; Watanabe, T.; Katagiri, T.; Tsunoda, T.; Yamaoka, Y. and Nakamura, Y. (2002) Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes. Cancer Res. 62, 7012-7017.

International Lymphoma Study Group (1997) A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. Blood 89, 3909-3918.

Jemal, A.; Thomas, A.; Murray, T. and Thun, M. (2002) Cancer statistics, 2002. CA. Cancer J. Clin. 52, 23-47.

Kikuchi, T.; Daigo, Y.; Katagiri, T.; Tsunoda, T.; Okada, K.; Kakiuchi, S.; Zembutsu, H.; Furukawa, Y.; Kawamura, M.; Kobayashi, K.; Imai, K. andd Nakamura, Y. (2003). Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogene. 22, 2192-2205.

Kopans, DB. (2003) Gene-expression signatures in breast cancer. N. Engl. J. Med. 348, 1715-1717.

Kunkler, IH. (2003) Gene-expression signatures in breast cancer. N. Engl. J. Med. 348, 1715-1717.

Moreau, Y.; Aerts, S.; De Moor, B.; De Strooper, B. and Dabrowski, M. (2003) Comparison and meta-analysis of microarray data: from the bench to the computer desk. Trends Genet. 19, 570-577.

Mountain, C. F. (1997) Revisions in the International System For Staging Lung Cancer. Chest **111**, 1710-1717.

Mousses S, Bubendorf L, Wagner U, Hostetter G, Kononen J, Cornelison R, Goldberger N, Elkahoun AG, Willi N, Koivisto P, Ferhle W, Raffeld M, Sauter G, Kallioniemi OP. (2002) Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. Cancer Res. **62** (5):1256-60.

Ohira, T.; Akutagawa, S.; Usuda, J.; Nakamura, T.; Hirano, T.; Tsuboi, M.; Nishio, K.; Taguchi, F.; Ikeda, N.; Nakamura, H.; Konaka, C.; Saijo, N. and Kato, H. (2002). Up-regulated gene expression of angiogenesis factors in post-chemotherapeutic lung cancer tissues determined by cDNA macroarray. Oncol. Rep. **9**, 723-728.

Oleksiak, MF.; Churchill, GA. and Crawford, DL. (2002) Variation in gene expression within and among natural populations. Nat. Genet. **32**, 261-266.

Parkin, DM. (2001) Global cancer statistics in the year 2000. Lancet. Oncol. **2**, 533-543.

Perou, CM.; Sorlie, T.; Eisen, MB.; van de Rijn, M.; Jeffrey, SS.; Rees, CA.; Pollack, JR.; Ross, DT.; Johnsen, H.; Akslen, LA.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, SX.; Lonning, PE.; Borresen-Dale, AL.; Brown, PO. and Botstein, D. (2000) Molecular portraits of human breast tumors. Nature **406**, 747-752.

Popat, U.; Przepiork, D.; Champlin, R.; Pugh, W.; Amin, K.; Mehra, R.; Rodriguez, J.; Giralt, S.; Romaguera, J.; Rodriguez, A.; Preti, A.; Andersson, B.; Khouri, I.; Claxton, D.; de Lima, M.; Donato, M.; Anderlini, P.; Gajewski, J.; Cabanillas, F. and van Besien, K. (1998) High-dose chemotherapy for relapsed and refractory diffuse large B-cell lymphoma: mediastinal localization predicts for a favorable outcome. J. Clin. Oncol. **16**, 63-69.

Scagliotti, GV. and Novello, S. (2003) The dream is almost over...don't worry, look ahead. Lung Cancer **40**, 187-190.

Slonim, DK (2002) From patterns to pathways: gene expression data analysis comes of age. Nature genetics supplement **32,** 502-50

Schiller, JH.; Harrington, D.; Belani, CP.; Langer, C.; Sandler, A.; Krook, J.; Zhu, J. and Johnson, DH. Eastern Cooperative Oncology Group. (2002) Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N. Engl. J. Med. **346,** 92-98.

Shipp, MA.; Ross, KN.; Tamayo, P.; Weng, AP.; Kutok, JL.; Aguiar, RC.; Gaasenbeek, M.; Angelo, M.; Reich, M.; Pinkus, GS.; Ray, TS.; Koval, MA.; Last, KW.; Norton, A.; Lister, TA.; Mesirov, J.; Neuberg, DS.; Lander, ES.; Aster, JC. and Golub, TR. (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat. Med. **8,** 68-74.

Simon, R.; Radmacher, MD.; Dobbin, K. and McShane, LM. (2003) Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J. Natl. Cancer Inst. **95,** 14-8.

Sorlie, T.; Perou, CM.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, MB.; van de Rijn, M.; Jeffrey, SS.; Thorsen, T.; Quist, H.; Matese, JC.; Brown, PO.; Botstein, D.; Eystein, Lonning. P. and Borresen-Dale, AL. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U S A. **98,** 10869-10874.

Stoeckert, CJ. Jr.; Causton, HC. and Ball, CA. (2002) Microarray databases: standards and ontologies. Nat. Genet. **32,** Suppl:469-473.

Suganuma, K.; Kubota, T.; Saikawa, Y.; Abe, S.; Otani, Y.; Furukawa, T.; Kumai, K.; Hasegawa, H.; Watanabe, M.; Kitajima, M.; Nakayama, H. and Okabe, H. (2003) Possible chemoresistance-related genes for gastric cancer detected by cDNA microarray. Cancer Sci. **94,** 355-359.

Tan, PK.; Downey, TJ.; Spitznagel, EL. Jr.: Xu, P.; Fu, D.; Dimitrov, DS.; Lempicki, RA.; Raaka, BM. and Cam, MC. (2003) Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. **31**, 5676-5684.

Tsunoda, T.; Koh, Y.; Koizumi, F.; Tsukiyama, S.; Ueda, H.; Taguchi, F.; Yamaue, H.; Saijo, N. and Nishio, K. (2003). Differential gene expression profiles and identification of the genes relevant to clinicopathologic factors in colorectal cancer selected by cDNA array method in combination with principal component analysis. Int. J. Oncol. **23**, 49-59.

Tusher, VG.; Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA. **98**, 5116-5121.

van de Vijver, MJ.; He, YD.; van't Veer, LJ.; Dai, H.; Hart, AA.; Voskuil, DW.; Schreiber, GJ.; Peterse, JL.; Roberts, C.; Marton, MJ.; Parrish, M.; Atsma, D.; Witteveen, A.; Glas, A.; Delahaye, L.; van der Velde, T.; Bartelink, H.; Rodenhuis, S.; Rutgers, ET.; Friend, SH. and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. **347**, 1999-2009.

van 't Veer, LJ.; Dai, H.; van de Vijver, MJ.; He, YD.; Hart, AA.; Mao, M.; Peterse, HL.; van der Kooy, K.; Marton, MJ.; Witteveen, AT.; Schreiber, GJ.; Kerkhoven, RM.; Roberts, C.; Linsley, PS.; Bernards, R. and Friend, SH. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature **415**, 530-536.

Weiss, MM.; Kuipers, EJ.; Postma, C.; Snijders, AM.; Siccama, I.; Pinkel, D.; Westerga, J.; Meuwissen, SG.; Albertson, DG. and Meijer, GA. (2003) Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene. **22**, 1872-1879.

West, M.; Blanchette, C.; Dressman, H.; Huang, E.: Ishida, S.; Spang, R.; Zuzan, H.; Olson, JA. Jr.; Marks, JR. and Nevins, JR. (2001) Predicting the clinical status of human breast cancer by using gene expression profiles. Proc. Natl. Acad. Sci. USA. **98**, 11462-11467.

White SL, Gharbi S, Bertani MF, Chan H-L, Waterfield MD, Timms JF. (2004) Cellular responses to ErbB-2 overexpression in human mammary luminal epithelial cells: comparison of mRNA and protein expression. Br J Cancer 90: 173-181

(please add chapter "breast cancer")

Table 1 The breast cancer microarray classification by Sorlie is based on a intrinsic set of 457 genes.

Correlation of microarray classification with overall survival prognosis (Sorlie 2001) (n=49; p<0,01)	
Subtype	Prognosis
ER+/luminal like Typ A	good
ER+/luminal like Typ B	intermediate
ER+/luminal like Typ C	intermediate
Basal like	poor
ERB-B2	poor
Normal like	intermediate

The estrogen receptor positive ER+/luminal like group is subdivided into three subtypes. Correlation with overall survival reveals a poor prognosis for the Basal like and ERB-B2 group. Interestingly different prognosis for patients was found within the three estrogen receptor positive (ER+) groups.

(please ad chapter "breast cancer")

Table 2 Class prediction studies regarding ER-Status in Breast Cancer

Author	Patients	Techniuque	Statistical method	Number of genes of predictor	Training s set	Test set	correct predict (%)
West (2001)	48	cDNA microarray	Bayesianregression	100	38	9	100
Gruvberger (2001)	58	cDNA microarray	Artificial Neural Network	100	47	11	100

"Predictors" for estrogen receptor status based on microarray data were established by to different groups in 2001. Both "predictors"include 100 genes. After develop the "predictor" in a set of samples and corresponding clinical data (Trainigs-set) both groups could validate their "predictor" in independent set of samples and clinical data (Test set) with high accuracy.

(please add chapter "breast cancer")

Table 3 Top 5 ranked genes for prediction ER-Status

Rank	West 2001	Gruvberger 2001
1	Trefoil factor 1 (ps2)	Estrogen Receptor 1
2	Estrogen receptor	Trefoil factor 3
3	Cytochrom P450	GATA Bindind protein 3
4	Trefoil factor 3	ESTs
5	Estrogen like growth factor	Calgranulin A

West and Gruvberger established in 2001 independently "predictors" for estrogen-receptor status in breast cancer based on microarray data. The five genes with strongest correlation of expression and ER-status of the 100 gene "predictors" by West and Gruvberger are listed in this table. Both "predictors" show similarities. Beside the estrogen receptor itself the trefoil factor 3 is find within the five top ranked genes in both studies.

(please add chapter "breast cancer")

Table 4 Overall survival and distant metastasis free survival probability according the prognosis signature (vant Vijver 2002)

Group	No. of patients	Overall survival(%)		free of distant metastasis (%)	
		5YR	10YR	5YR	10YR
Poor prognosis signature	180	74.1	54.6	60.5	50.6
Good prognosis signature	115	97.4	94.5	94.7	85.2

A 70 gene prognostic marker ("predictor") was tested by van t Vijver in a series of 295 consecutive patients with stage I and II breast cancer who underwent surgery. They good distinguish 180 patients with poor prognosis (Poor prognosis signature) from 115 patients with good prognosis (Good prognosis signature) regarding to overall survival and distant metastasis free survival.

(please add chapter "lung cancer")

Table 5 Selected examples of the 50 gene risk index of Beer (2002)

Gene name	P (normal versus tumor t-test)	Coefficient β	Comment
Caspase 4	0,56	0,0022	apoptosis-related cysteine protease
LAMB 1	0,14	0,0027	Laminin β 1
BMP 2	0,54	0,0044	Bone morhogenetic protein 2
CDC 6	1,31E-05	0,0124	cell division cycle 6
Serpine 1	2,89E-03	0,0008	Serine (or cysteine) proteinase inhibitor (clade E)
ERBB2	0,04	0,0013	v-erb -b2 (Receptor)
PDE7A	0,12	- 0,0187	Phosphodiesterase 7a
PLGL	0,04	- 0,0011	Plasminogen like

The 50-gene-risk index was validated in an independent set of 84 tumor samples and corresponding A positive coefficient β is associated with poorer outcome. A 50 gene risk index ("predictor") for lung adenocarcinomas was established in a microarray based correlation study (Beer 2002). Selected examples for interesting genes of this risk index were shown in this table. The coefficient β shows the relation of gene expression and outcome. A positive coefficient β is associated with poorer outcome. This 50 survival data. Among the 62 stage I tumors including this set they could identify a high and a low risk group which differ significant in survival.