like group, an ERB-B2-(erythroblastic leukemia viral oncogene homolog 2) group and a normal-breast-like group. The same group published within the following year a further investigation on this subject [Sorlie 2001]. This time, in a larger number of tumors at least six subtypes could be differentiated on the basis of their gene expression patterns using an intrinsic set of 457 genes (Table 1).The previous (ER+)/luminal-like group could be subdivided into three prognostic groups. In addition, a list of 256 genes related to clinical outcome was developed using a special supervised data-analysis (SAM: significance analysis of microarray, this technique will be explained more detailed in the second part of this article) derived from a clinical data set of 76 carcinomas, where sufficient data were available. This 256-gene-"predictor" was correlated with clinical outcome in a subcohort of 49 patients with locally-advanced breast cancer uniformly treated in a prospective clinical study. The basal-like subtype was associated with a poorer prognosis regarding overall as well as relapse-free survival (Table 1). Interestingly, significant differences could be observed in the clinical outcome between two different estrogen–receptor-positive groups. West and co-workers could demonstrate the potential usefulness of DNA microarray analysis to discriminate different breast cancer patients on the basis of their ER (estrogen-receptor) status (Table 2)[West 2001]. A 100-gene "predictor" to estimate estrogen receptor status was developed analysing a set of 38 samples (the "trainings set"). Then in a second step, the "predictor" was validated in a set of nine independent tissue samples (the "test set") (Table 3). In some samples, the "predictor" could adequately predict clinical ER-status. Five samples with conflicting results of immunohistochemistry and immunoblotting regarding the ER-status were included within this "test set". In some of these samples, the predictive probability of the "predictor" was found lower. This could be reflect a heterogeneity of expression within the individual tumor. Furthermore, impressive data regarding an analysis of estrogen-receptor status by calculating gene expression profiles were published by a Swedish group within the same year

(Table 2) [Gruvberger 2001]. A 100-gene "predictor" for ER status was developed in a "trainings set" of 47 tumors based on microarray data using "artificial neural network[2]"(Table 3). The "predictor" was validated in 11 independent samples of a "test set". All 11 tumors could be classified correctly by this 100-gene marker. Interestingly, even without the top discriminator genes, including ER itself, the "artificial neural network" could adequately predict ER-status.

Adjuvant cancer therapy is well established in the treatment of breast cancer. Chemotherapy or hormonal therapy is able to reduce the risk of disease dissemination in one-thirds of the patients, but a large number of patients would have already been cured without the application of adjuvant therapy. Taking into account that application of adjuvant chemotherapy carries a well-defined morbidity and mortality risk, a proper selection of patients with an established benefit from adjuvant chemotherapy would be extremely helpful. Based on the hypothesis, that patients with a poor prognosis following surgery would get benefit most from adjuvant therapy a Dutch group performed a translational study testing the predictive impact of DNA Microarray data on overall survival prognosis in young female with primary-lymph–node-negative breast cancer [van 't Veer 2002]. In this study, tumor tissue from 117 young patients with primary lymph-node-negative breast cancer was analysed by DNA microarray technique. A 70-gene prognosis "predictor" ("poor prognosis signature") for patients with a short interval to development distant metastasis was established by supervised classification of the gene expression profiling. This "poor prognosis signature" included genes regulating cell cycle, invasion, metastasis and angiogenesis. A second correlation study was undertaken by the same group to confirm the predictive power of this 70-gene prognostic marker in an larger and less homogeneous group of patients [van de Vijver 2002]. A series of 295 consecutive patients with stage I and II breast cancer, who underwent surgery, were included into this study. All patients included were younger than 53 years. In this investigation patients with-lymph-node-negative (151 pts) and lymph-node-positive

(144 pts) disease were analyzed. A group of 180 patients with "poor prognosis signature" could be separated from 115 patients with "good-prognosis-signature" (Table 4). The overall 10-year survival rate was found to be 54.6 percent within the "poor" and 94.5 percent in the "good prognosis group". The probability to remain free from distant metastases within 10 years was found to be 50.6 percent in the group with "poor-prognosis- signature" and 85.2 percent in the group with "good–prognosis-signature". In comparison to the "good prognosis group" the estimated hazard ratio for developing distant metastasis in the group with "poor-prognosis-signature" was 5.1(95 percent confidence interval, 2.9 to 9.0; $P< 0.001$). This microarray based prognosis profile was identified as a strong independent factor in predicting disease outcome. Interestingly, the prognostic profile did not depend on lymph–node status, well-known to be one of the classical prognostic factors. On the other hand, the positive-estrogen–receptor status, the other classical predictor, was strongly correlated with a "good prognosis signature" profile. In spite of these impressive results, there has been important criticism towards these findings: An insufficient analysis of intratumoral representativeness of the tumor samples and the measurement of tumor size is critically discussed [Kunkler 2003, Kopans 2003]. However, this pioneer investigation has proven, that -in principle- a prediction of clinical outcome based on microarray data may be possible.

Lung Cancer

Lung Cancer is still the leading cause of cancer-related deaths in the industrialized world [Jemal 2002]. Physicians treating patients with this disease often face difficult decisions to be made within all clinical disease stages. For example, the issue of adjuvant chemotherapy following complete resection is not conclusively solved with conflicting results published [Scagliotti 2003; International Adjuvant Lung Cancer Trial Cooperative Group 2004]. A reliable method to predict

patient prognosis following surgical treatment of early stage lung cancer could be most helpful to estimate the benefit of adjuvant chemotherapy for an individual patient. Classical histopathology is definitely insufficient for this decision making. Beer published in 2002 a translational research study correlating gene expression profile data with overall survival in patients with early stages lung adenocarcinoma after surgery [Beer 2002]. Tumor tissue of 86 primary adenocarcimas of the lung, including 67 stage I and 19 stage III tumors, was investigated. As an internal control 10 non-neoplastic lung tissue samples were analyzed. Using "hierarchical clustering" methology, three different patient groups with association of cluster and stage (P=0,030) or tumor differentiation (P=0,01) could be differentiated. All ten non-neoplastic tissue samples could be clustered within the same patient group. In addition, the authors could derive a 50-gene-risk-index by identifying survival related genes using univariate Cox analysis (Tab.5).When calculating the 50-gene-risk-index and grouping the patients based on the results, significant differences in overall survival between the individual groups could be identified. Grouped "high-risk" and "low-risk" stage I adenocarcinomas differed significantly between each other (p=0.003), whereas low- and high risk stage III tumors did not. The robustness of the 50-gene-risk-index in predicting overall survival in early stage lung adenocarcinoma was tested in an independent data set of 84 tumor samples and related to the survival. A high and a low risk group could be separated (P=0.003). Interestingly, among the 62 stage I tumors in this analyzed population, high- and low-risk groups could be observed differing significantly (P=0.006) in their overall survival duration. In conclusion, the authors postulate, that the identification of a high risk group within stage I lung cancer patients would lead to the consideration of a postoperative adjuvant intervention for this group. In 2003 a Japanese group has published another important cDNA microarray based study regarding lung cancer [Kikuchi 2003]. A set of 37 tumor tissue samples of non-small cell lung cancer patients were analyzed. To avoid investigation on non relevant

tissue they only analyzed cancer cells selected by laser capture microdissection .The most frequent NSCLC-subtypes adenocarcinoma and squamous cell carcinoma could be easily distinguished by applying a clustering algorithm to the expression data results. To explore gene expression in post-chemotherapeutic lung cancer tissue, a small pilot-study using RNA filter-array was performed in our institution [Ohira 2002]. Lung and normal tissue from three patients who underwent neoadjuvant therapy prior to surgery were collected following the end chemotherapy. Gene expression data obtained by a 588 gene filter arrays were analysed by "hierarchical clustering" method. Remarkably, normal tissue and tumor tissue from the same patient showed more similarities and clustered nearer than normal and normal or tumor and tumor samples from different patients "Fig. (1)".On the other hand, groups of genes significantly differed in expression profiles between normal and malignant tissue. Especially angiogenesis and invasion related genes were up-regulated in the tumor samples "Fig (2)" These results suggest that molecules involved in angiogenesis are suitable targets for novel drugs administered following chemotherapy. This early study is one example for how genomic techniques could help to discover new candidates for target based therapeutics in the future.

Gastrointestinal Cancer

Gastric cancer is still the fourth leading cause of cancer in the world [Parkin 2001]. Due to the lack of sufficient systemic control induced by current anti-neoplastic agents surgery remains the cornerstone treatment approach in this disease. The development of lymph node metastasis is a well established independent risk factor for recurrence of gastric cancer. In the recent years, two independent research groups have established microarray based risk factor scoring systems for the development of lymph-node metastasis in gastric cancer. In 2002, a Japanese group published

a study in patients with Intestinal Type Gastric Cancer [Hasegawa 2002]. Primary gastric cancer and corresponding noncancerous gastric mucosa from 20 patients who underwent surgery were comparatively analyzed. A set of 61 genes that were commonly up-regulated and 63 genes down-regulated in Intestal-Type Gastric Cancer in more than 75% of the cases could be identified. In a second step the expression profiles of nine cases with and cases without lymph-node metastasis were compared. In this approach, 12 genes that were differentially expressed (P< 0,01) were identified by employing a random permutation test. Nine of these 12 genes were over-expressed and three were down-regulated in node positive tumors. By use of a "stepwise discriminant analysis" five independent "predictors" were identified among these 12 genes (Table 6). The predictive scoring system was confirmed in nine independent additional tumor tissue samples. All nine cases (four node positive and nine node negative) were correctly assigned to each class by the means of the scoring system.

One year later, a Dutch group performed a comparable investigation. The molecular data of 35 gastric carcimomas were analysed with their clinical data sets [Weiss 2003]. Microarray Comparative Genomic Hybridization (GCH), which allow to analyze accumulation of genetic changes that to a large extend occur on a chromosomal level was applied to their approach. Three different groups could be distinguished by "hierarchical clustering" of the microarray CGH results. For each cluster they could define a signature of 204 genes by using a "leave one out" cross validation. Each cluster was analyzed for correlation with clinico-pathological data. The lymph–node status and the overall survival were criteria with significant differences between the individual groups. In one group significantly less lymph-node positive cases (40%) were found than in the other one (83% in comparison to 88%). Patients who belong to the former group were found to have a significantly longer survival duration (P=0,019).

Both cited studies focussed on a clinically most important issue. The discussion about the inclusion of extended lymph node disssection into surgery for gastric cancer remains rather controversial [Bonenkamp 1999,Cuschieri 1999]. The increase of overall morbidity and mortality associated with this treatment strategy has to be well balanced against the benefit for lymph-node positive patients following this intervention. Predicting the overall risk of lymph-node metastasis by microarray techniques in an individual patient could be a reasonable strategy to select patients for this kind of therapy in the future and would have major implications for clinical practice in this entity.

In 2003 Suganuma published a study focussed on possible chemoresistance-related genes in gastric cancer [Suganuma 2003]. Tumor samples and corresponding normal mucosa from 35 patients with advanced gastric cancer were differentially examined. The in vitro sensitivity of cells from each dissociated tumor sample against cisplatin, 5-flourouracil, mitomycin C and doxorubicin was measured by MTT (4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. The results obtained from in-vitro cytotoxicity testing assays were correlated with the results of cDNA microarray analysis of corresponding tissue samples. In the case of cisplatin, "hierarchical clustering" could successfully distinguish "sensitive" and "resistant" tumors from each other. A set of 23 potential "cisplatin-resistance-related-genes" could be selected by this method. The latter group included vascular permeability factor, two membrane-transporting subunits and retinoblastoma-binding protein. In a further selection based on strong criteria, metallothionein IG and heparin-binding-epidermal-growth-factor-like-growth factor were also identified as candidates for cisplatin-resistance-related genes. Within this approach dihydropyrimidine dehydrogenase and HB-EGF-like growth factor were suggested to be 5-FU resistance-related genes. In this innovative approach the authors could demonstrate that DNA-microarray could be useful to investigate drug-resistance and would be a means to understand

some of the complex mechanisms behind it. Although this study is only based on a limited data set this method could in principal be an important step for an individualized and customized cancer therapy, in the future.

Moreover, in 2003 a Japanese group published a correlative study based on gene expression profiling in patients with colorectal cancer [Tsunoda 2003]. To clarify the regulatory factors in this malignant disease, differential gene expression profiles were analyzed by filter-array in surgically resected specimen (tumor and normal mucosa) obtained from ten patients with colorectal cancer. The correlation between several clinico-pathologic factors and cancer-related genes were investigated by using complex statistical analyses including "average linkage hierarchical clustering and principal component analysis (PCA)[4]". As an example, the c-myc- binding- protein and the c-jun-proto-oncogene were both identified as possible correlative markers for histological differentiation and overall clinical prognosis "Fig.(3)".The early-growth-response-protein 1 was selected to play an important role in the progression of clinical stage. The authors concluded, that PCA was identified as an appropriate method to select candidate genes relevant to predict clinico-pathologic factors in a small population of clinical samples from colon cancer patients.

Lymphoma:

Diffuse large B-cell lymphoma (DLBCL) presents the most common subtype among the lymphoid neoplasm in adults (International Lymphoma Study Group 1997). Less than 50% of the patients are currently cured with standard combination chemotherapy [Popat 1998]. In 2000 the first large gene expression profiling study was performed in this disease .[Alizedah 2000]. Within this study, different types of diffuse large B-cell lymphoma patients could be identified by gene expression profiles. A specific microarray assay for lymphoma was designed. This "Lymphoma-

chip" included genes preferentially expressed in lymphoid cells and genes with known or suspected roles in cancer development and immunology. Using this array 96 normal and different malignant lymphocyte samples were comparatively analyzed. All DLBCL samples could clearly be separated from normal lymphocyte samples and other lymphomas. Focused on genes related to a separate stage of B-Cell differentiation and activation, a germinal-centre-B-like DLBCL and an in vitro-activated-B-like DLBCL could distinguished among the group with diffuse large B-cell lymphoma by "hierarchical clustering" analysis. This clustering was based on the hypothesis that DLBCL derives from normal B-cells within the germinal centres (GC). As a consequence to this, overall survival and clustering to one of these two groups were correlated. Patients belonging to the GC-B-like-group had a significantly higher five-year survival rate than patients belonging to the activated-B-cell-DLBCL group (P< 0,01). The average five-year survival rate for all patients was 52%, for patients of the GC B-like DLBCL 76% and for patients of the activated B-like DLBCL only found to be 16%. Two years later, Shipp and co-workers published another prognostically predictive score for DLBCL based on gene expression profiling [Shipp 2002]. Contrary to the latter work an alternative strategy that was independent of an *a priori* hypothesis was employed. A novel "supervised learning method" was applied. Tumor samples from 77 patients were analysed. 58 patients with DLBCL and 19 patients with Follicular Lymphoma were clearly distinguishable by use of this method. Clinical outcome prediction in DLBCL patients based on gene expression profiling data was the further purpose of this study. The long-term follow up was available for all DLBCL patients. While 32 patients were eventually cured, 26 patients turned out to have fatal or refractory lymphoma disease. An "outcome predictor" was designed using a "supervised learning classification approach" (weighted voting algorithm and cross validation test). The highest accuracy was obtained using a "predictor" set of 13 genes (Table 7). Using this "predictor" set the DLBCL collective could be

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.