future. In the following overview, we will try to focus on the current role of genomic techniques in translational cancer research. In the first part of this paper, we will summarize the most important published studies for frequent tumor entities. Furthermore, the individual methods and their clinical relevance will be critically reviewed. In the second part of the manuscript, we will deal with issues of reliability and reproducibility of reported genomic array data.

BREAST CANCER

In 2000, a Norwegian – American Cooperative Group performed an analysis of breast cancer cell lines and tumor tissue based on DNA microarrays [Perou 2000]. They could classify the tumors into subtypes distinguished by their different gene expression profiles using the "hierarchical clustering[1]" methodology based on 1753 genes independent from the histological classification. The tumors were subdivided into an ER+(estrogen-receptor-positive)/luminal like, a basal-epithelium-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, in which sufficient data were available. This 256-gene-"predictor" was correlated with clinical outcome in a cohort of 49 patients with locally-advanced breast cancer uniformly treated in a prospective clinical investigation. The basal-like subtype was associated with 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 analyzing a set of 38 samples (the "training 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 data regarding the ER-status were included within this "test set". In some of these samples, the predictive probability of the "predictor" was found to be lower. This could reflect the 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

Table 1. Breast Cancer Microarray Classification by Sorlie - 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.

Table 2. Class Prediction Studies Regarding ER-Status in Breast Cancer

Author	Patients	Technique	Statistical method	Number of genes of predictor	Training set	Test set	correct prediction (%)
West (2001)	48	cDNA microarray	Bayesian regression	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 developing the "predictor" in a set of samples and corresponding clinical data (Trainig-set) both groups could validate their "predictor" in independent set of samples and clinical data (Test set) with high accuracy.

[1] **Hieachical clustering:** A hierarchical clustering is a sequence of partitions in which each partition is nested into the next partition in the sequence

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	Cytochrome 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.

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 level 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-third of the patients, but a large number of patients will have already been cured without the application of adjuvant therapy. Taking into account that the application of adjuvant chemotherapy carries a well-defined morbidity and mortality risk, the proper selection of patients with a clear and established benefit from adjuvant chemotherapy would be extremely helpful. Based on the hypothesis that patients with a poor prognosis following surgery would mostly benefit from adjuvant therapy, a Dutch group performed a translational study testing the predictive impact of DNA Microarray data on overall survival prognosis in young females 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 analyzed by DNA microarray technique. A 70-gene prognosis "predictor" ("poor prognosis signature") for patients with a short interval to the development of 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 correlative study was performed by the same group to confirm the predictive power of this 70-gene prognostic marker in a 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 had to be 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 of remaining 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% 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

Table 4. Overall Survival and Distant Metastasis Free Survival Probability According to Prognosis Signature (Van't 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 could 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.

[2] **Artificial neural network:** A network of many very simple processors "units" or "neurons"), each possibly having a (small amount of) local memory. The units are connected by unidirectional communication channels („connections", which carry numeric (as opposed to symbolic) data. The units operate only on their local data and on the inputs they receive *via* the connections.

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 investigated tumor samples and the measurement of tumor size is critically discussed [Kunkler 2003, Kopans 2003]. However, this pioneer investigation could prove, that a prediction of clinical outcome based on microarray data may in principle 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 2003, LeChevalier 2003*]. 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 within the individual patient. Classical histopathology examination 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 adenocarcinomas 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" methodology, 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 (Table **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 overall 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 postulated, 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 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 irrelevant 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 of chemotherapy. Gene expression data obtained by a 588 gene filter arrays were

Table 5. Selected Examples of the 50 Gene Risk Index of Beer (2002)

Gene name	P	Coefficient	Comment
	(normal versus tumor-t-test)	β	
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 7°
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.

* 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).

analyzed by "hierarchical clustering" method. Remarkably, normal tissue and tumor tissue from the same patient showed more similarities and clustered nearer than normal and normal tissue 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 upregulated 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 of 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,

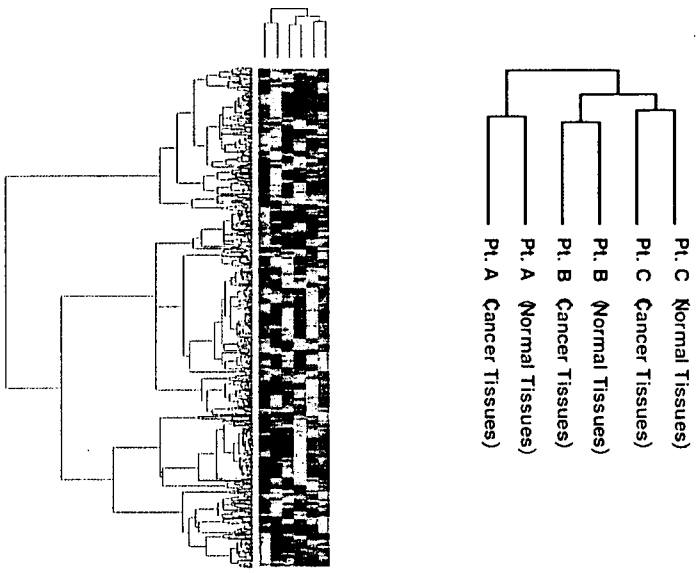


Fig (1). Clustering of gene expressions of tissues 3 from lung cancer patients (Ohira 2002). Tumor tissue and normal lung tissue was collected while surgery after neoadjuvant chemotherapy. Tumor tissue and normal tissue from the same patient show more similarities and clustered nearer than normal.

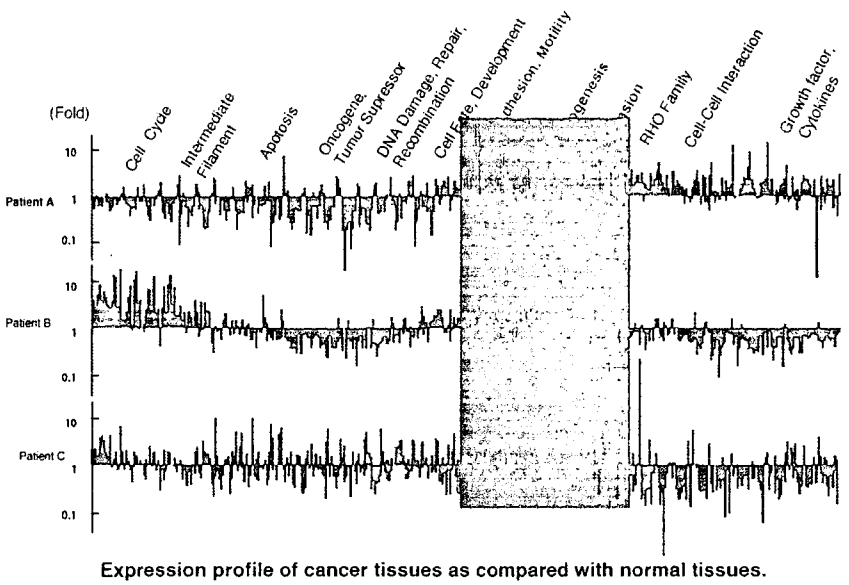


Fig. (2). Histogram of gene expression profile of lung cancer tissue. Expression profile of cancer tissues as compared with normal tissues.

Case B; increased expression of the genes related to cell cycle regulator, intermediate filaments, adhesion motility and angiogenesis in the tumor tissues. Expression of the other gene group were decreased in tumor tissue. Case C; increased expression of genes related with cell cycle adhesion were observed in the tumor tissue. Decreased expression of growth factor and cytokine related genes were also observed in tumor of Case C. Taken together, the expression profile of lung carcinoma could be characterized by the increased expression of the genes related with adhesion motility and angiogenesis.

surgery remains the cornerstone treatment approach to this disease. The development of lymph node metastasis is a well-established independent risk factor for recurrence of gastric cancer. In 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 non-cancerous gastric mucosa from 20 patients who underwent surgery were comparatively analyzed. A set of 61 genes that were commonly upregulated and 63 genes downregulated in Intestinal-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) could be identified by employing a random permutation test. Nine of these 12 genes were overexpressed and three were downregulated 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 means of the scoring system.

One year later, a Dutch group performed a comparable investigation. The molecular data of 35 gastric carcinomas were analyzed with their clinical data sets [Weiss 2003]. Microarray Comparative Genomic Hybridization (GCH), which allows to analyze accumulation of genetic changes that to a large extent 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 dissection 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 the 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, dihydro-pyrimidine 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

Table 6. Five Genes for Predicting Risk of Lymphnode Metastasis in Intestinal Gastric Cancer (Hasegawa 2002)

Title		Discriminant coefficient
DDOST	dolichyl-diphosphooligosaccharide-protein glycosyltransferase	1.87
GNS	glucosamine (N-acetyl)-6-sulfatase (Sanfilippo disease IIID)	1.26
NEDD8	neural precursor cell expressed, developmentally down-regulated 8	1.29
LOC51096	CGI-48 protein	1.36
AIM2	absent in melanoma 2	-1.54

Five genes were selected based on microarray data for predicting risk of lymph-node metastasis in intestinal gastric cancer (Hasegawa 2002).This "predictor" was validated in 9 additional independent cases. All cases were (four node positive and five node negative) were assigned to each classes.

surgically resected specimen (tumor and normal mucosa) obtained from ten patients with colorectal cancer. The correlation between several clinico-pathological factors and cancer-related genes were investigated by using complex statistical analyses including "average linkage hierarchical clustering and principal component analysis (PCA)[3]". 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 I 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-pathological factors in a small population of clinical samples from colon cancer patients.

LYMPHOMA

Diffuse large B-cell lymphoma (DLBCL) represents 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

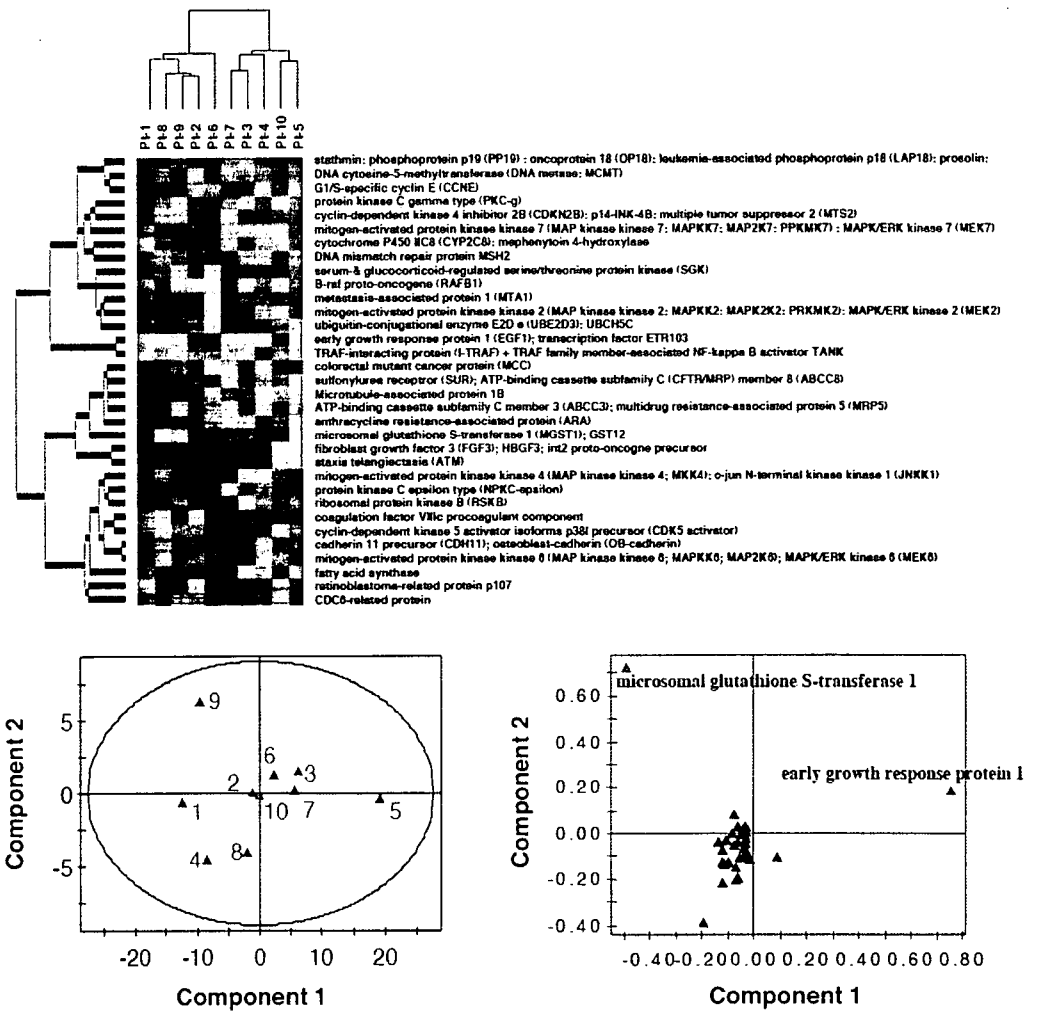


Fig. (3). (A). Average-linkage hierarchical clustering analysis of ten colorectal tumor samples on histological diagnosis. Right cluster shows the group of the well-differentiated and left shows the group of the other differentiations. (B) Principal component analysis on histological diagnosis. The numbers in blue indicate the patients with well-differentiated adenocarcinoma and the numbers in red indicate the patients with the other differentiations. The c-myc binding protein gene and the c-jun proto oncogene were identified as possible markers for histological differentiation.

[3] **Principal component analysis:** A method of analyzing multivariate data in order to express their variation in a minimum number of principal components or linear combination of the original, partially correlated variables.

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 be 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 was 76% and for patients of the activated B-like DLBCL only was found to be 16%. Two years later, Shipp and co-workers published another prognostic 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 analyzed. 58 patients with DLBCL and 19 patients with Follicular Lymphoma were clearly distinguishable by the 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 Maier survival analysis revealed a significantly better 5-year-overall-survival rate for the *"cured"* group with 70% in comparison to the *"fatal and refractory"* 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 90genes 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. Furthermore, the predictive power of their own 13 gene based "predictor" was tested in the dataset of Alizedah. Three of these genes were 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 re-analyzed by a Japanese group [Ando 2002]. "Fuzzy Neural Network[4]" as a new statistical method to analyze 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 93% accuracy.

Table 7. Model of 13 Genes Predicting Outcome in DLBCL Patients (Shipp 2002)

Genes associated with good outcome	Genes associated with poor outcome
-Dystrophin related protein 2	-H731
-3UTR of unknown protein	-Transduction like enhancer protein 1
-uncharacterised	-PDE 4 B
-Protein Kinase C gamma	-uncharacterised
-Minor / NOR 1	-Protein kinase C beta 1
-Hydroxitryptamine 2B Receptor	-Oviductal glycoprotein
-Zinc finger protein C2H2-150	

A 13-gene based "predictor" for outcome in DLBCL patients was developed based on microarray data by a supervised learning method (Shipp 2002). The expression of seven genes were associated with good and the expression of six genes was associated with poor outcome. This "predictor" was superior to "hierarchical clustering" based classification of Alizedah in predicting outcome of DLBCL patients.

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 was selected by the Whitney-Mann test. Among these 76 genes, 37 genes were found upregulated and 39 genes were found downregulated in responders. Interestingly, 10 of the 37 upregulated genes were involved in angiogenesis, while 6 of the 39 downregulated 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 treatment and prognosis evaluation of neoplastic disease.

RELIABILITY AND REPRODUCIBILITY OF ARRAY DATA

These examples of investigations with microarray technique surely represent landmark studies in their

[4] **Fuzzy neural network:** A special kind of artificial neural network.

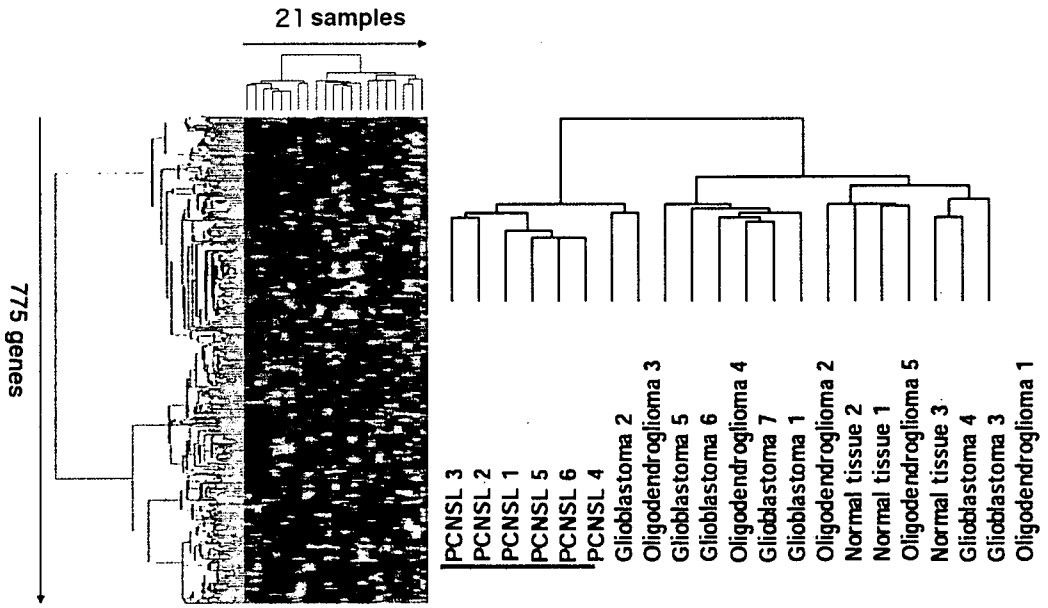


Fig. (4). Macroarray analysis of the 21 samples including PCNSCL, Glioblastoma, Oligodendroglioma and normal tissue. The phylogenetic tree obtained by application of the " clustering algorithm" shows separation of the PCNSL.

individual fields. They have brought a realistic hope to the scientific community that long time open issues will be solved within the near future. But besides all well-founded hope and enthusiasm, it cannot be overlooked that analyzing and interpreting array data remains a rather complex technology. Applying these novel genomic techniques uncritically to clinical data sets and clinical trials could lead 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 the 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 important 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

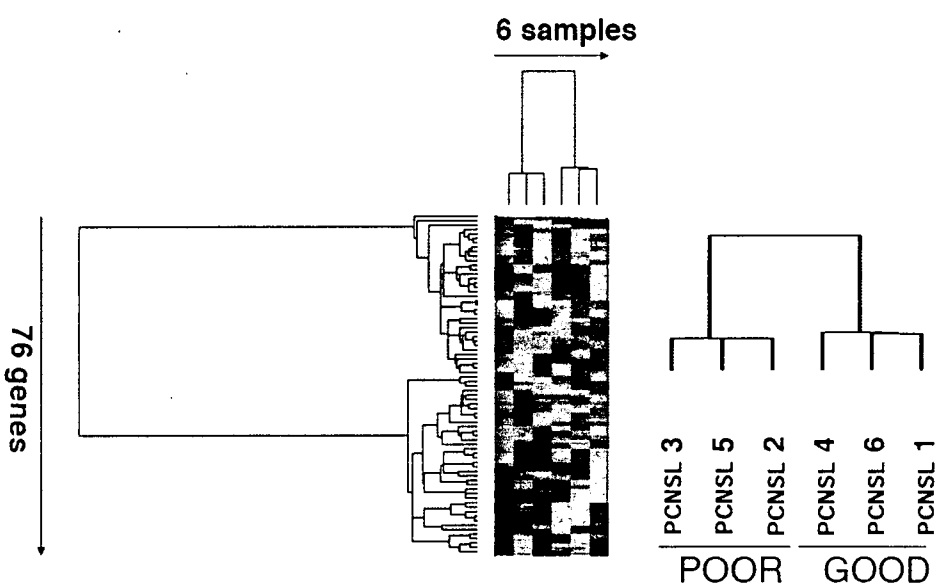


Fig. (5). Re-clustering was performed using selected genes related with response to chemo-radiotherapy.The responders (described as GOOD) and non-responders (described as POOR) were clearly separated clearly by the re-clustering method.

this investigation. Statistically, significant differences in expression profiles between individuals within the same population for approximately 18% of 907 genes were observed. Typically, expressions 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 (1_g or less) could be obtained. Then we have 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 easily 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-transferase Inhibitor, we have 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 chemo-sensitivity was used for analyzing 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 microarrays [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 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 too. An inadequate data analysis can lead

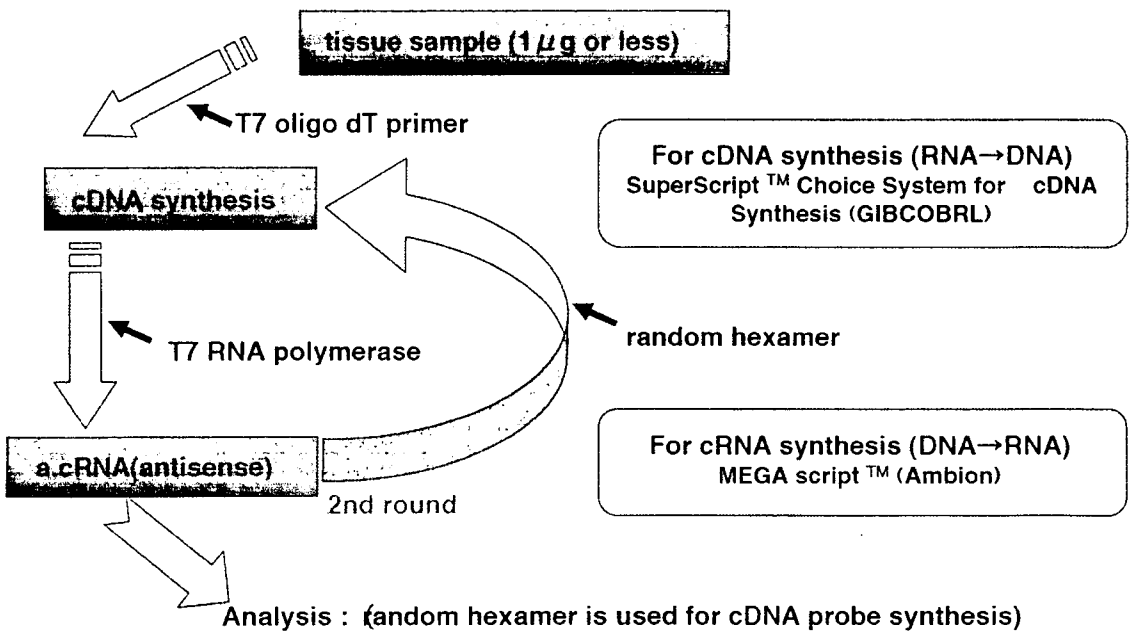


Fig. (6). Gene amplification by T7-based RNA amplification method. In a 2 step approach, first cDNA was synthesized (RNA→DNA) followed by c.a.RNA synthesis (DNA→RNA), we could purify 10-100µg RNA of amplified cRNA from small amount of total RNA (1µg or less).

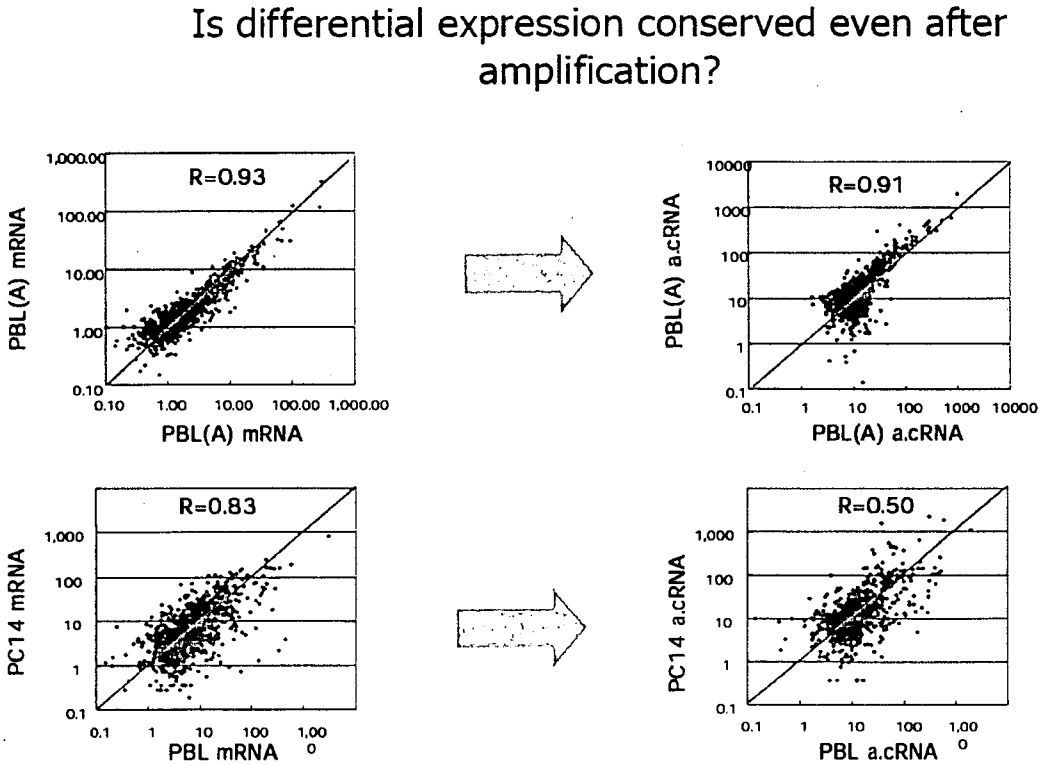


Fig. (7). Is differential expression conserved even after amplification? In order to analyze the reproducibility of the clinical samples, the gene expression profile of non-amplified and amplified samples were compared in scattered plot. Upper: gene expression data of duplicate samples of peripheral blood mononuclear cells were compared in scattered blot. High reproducibility (R=0.93) was obtained. These reproducible profiling was also observed in the amplified samples (R= 0.91). Lower: In a second experiment, we compared the differential gene expression of the PC 14 cell line and of peripheral blood mononuclear cells using mRNA and after amplification a.cRNA (amplified cRNA). Also the reproducible profiling was lower after amplification (R=0.50) than in non-amplified samples (R=0.83); we could conserve significant differences in gene expression after amplification.

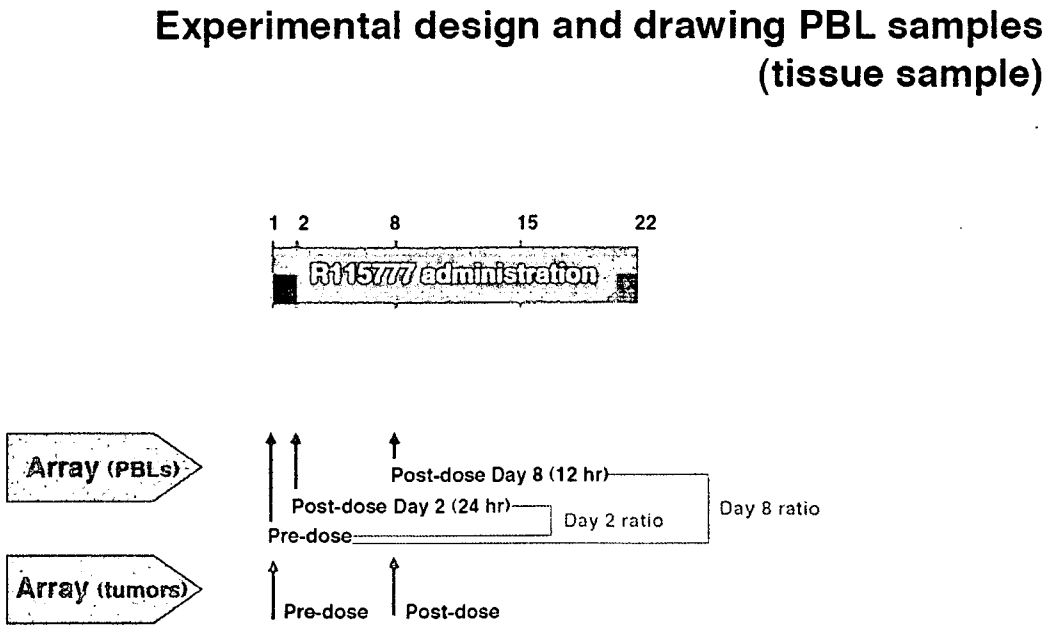


Fig. (8). Experimental design: sampling of PBL and tissue samples in correlative study in clinical phase I study of a falnesyltransferase inhibitor (FTI). Peripheral Blood Lymphocytes and tumor samples were collected predose, postdose day 2 and post dose day 8. Gene alteration after administration of FTI was analyzed for proof the pharmacodynamic effect of FTI.

1570-1603/05 $50.00+.00 ©2005 Bentham Science Publishers Ltd.

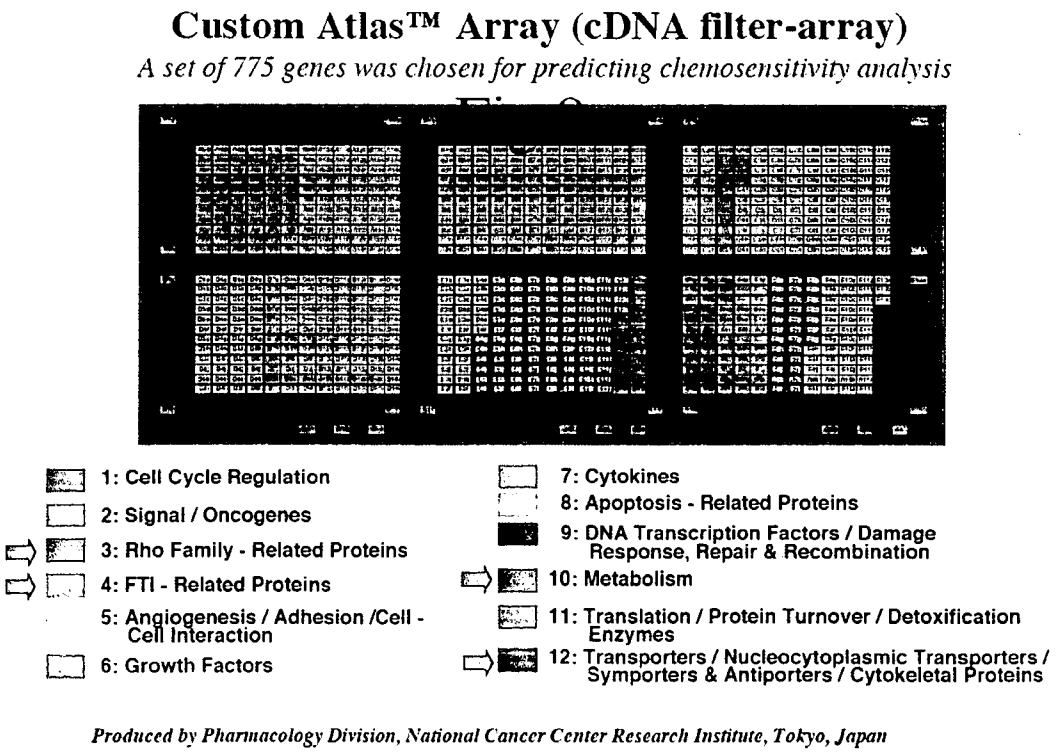


Fig. (9a). The cDNA filter-array with a set of 775 genes chosen for predicting chemosensitivity analysis.

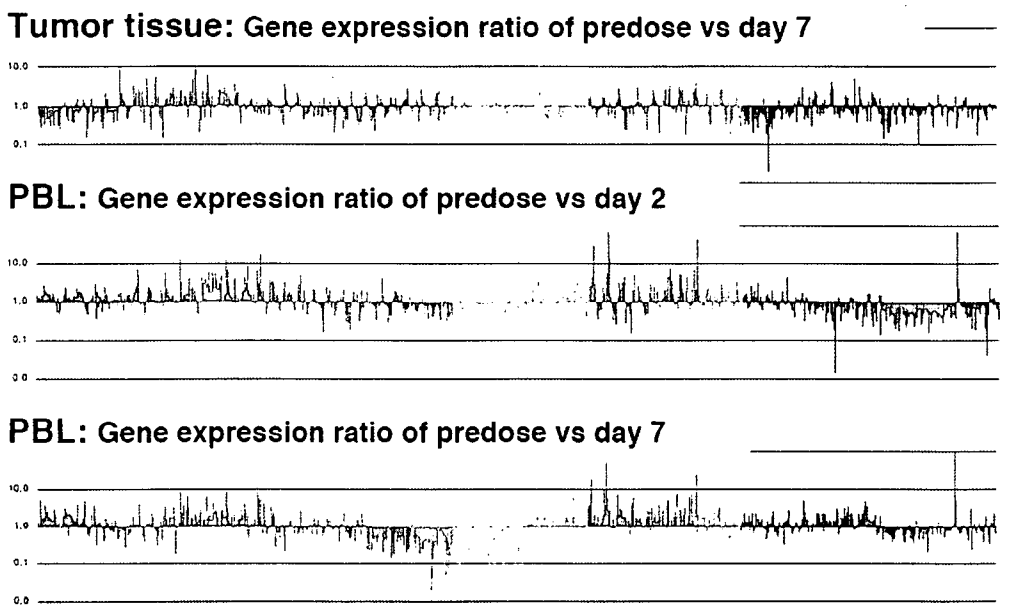


Fig. (9b). Gene expression changes of tumor tissue and PBL in the melanoma patient after administration of FTI. Specific gene groups were modulated by FTI. Changes in gene expression influenced by FTI were not only observed in the tumor samples but also in the peripheral blood lymphocytes. This findings suggest that drug modulated changes of gene expression in peripheral blood lymphocytes could be useful as surrogate markers in pharmacogenomic studies.

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 analyzing 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 statistical 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 analyzed studies correlating outcome with genetic profiles based on microarray data published from

Table 8. Analysis of Microarray Based Correlation Studies (1999-2003) by Nitzani and Ioannidis (Nitzani 2003)

Characteristic	Studies of major clinical outcomes (n=30)	Other studies (n=54)	Total (n=84)
Year of publication			
1999	1(3%)	2(4%)	3(4%)
2000	2(7%)	1(2%)	3(4%)
2001	6(20%)	18(33%)	24(29%)
2002	18(60%)	28(52%)	46(55%)
2003	3(10%)	5(9%)	8(10%)
Malignant disorder			
Haematological	9(30%)	9(17%)	18(21%)
Solid tumor	21(70%)	45(83%)	66(79%)
Median (IQR) number of samples			
Total	62(29-96)	30(18-44)	37(20-57)
Specific cancer	43(24-69)	20(13-36)	25(15-45)
Microarray type			
cDNA	19(63%)	31(57%)	50(60%)
Oligonucleotide	11(37%)	23(43%)	34(40%)
Median (IQR) number of probes	8683 (6817-18 624)	6936 (4569-12 600)	7014 (5534-12 600)
Training			
Independent	9(30%)	17(32%)	26(31%)
Dependent	8(27%)	20(37%)	28(33%)
Both	13(43%)	17(32%)	30(36%)
Validation			
Independent	3(10%)	1(2%)	4(5%)
Cross-validation	6(20%)	4(7%)	10(12%)
Both	3(10%)	5(9%)	8(10%)
None	18(60%)	44(82%)	62(74%)
Outcomes/correlates assessed			
One	9(30%)	35(65%)	44(52%)
Two to four	12(40%)	11(20%)	23(27%)
Five or more	9(30%)	8(15%)	17(20%)
Significant associations reported			
Yes	21(70%)	20(37%)	41(49%)
No	9(30%)	34(63%)	43(51%)

Microarray correlation studies focused on prediction outcome or other important clinico-pathological features were systematically analyzed by Nitzani and Ioannidis in 2003. This table shows the results of their investigations. In 70% of the studies correlating major clinical outcome with gene expression significant associations were reported. However, in only 30 percent of the major outcome focused studies cross-validation or independent validation was performed. These findings underline the need for consequent quality control and validation in microarray based clinical studies.

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 validated their DNA-microarray results by Real-Time-PCR [Chaqui 2002]. Although we can measure the level of expression of the genes of interest in a reliable way, we will miss information of the post-translational 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 western blotting. 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 epithelial 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 upregulated and 122 downregulated in the cancer-cells. In general, most of the genes that are overexpressed in the head and neck cancer cells encode for growth factors and cell structure. The underexpressed 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-microarray, 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. These differences may be explained by differing rates of translation or protein stability in the cancer cell versus their normal counterparts. Tissue microarrays (TMA) are promising approaches for validation of DNA-microarray 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 hybridization* 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 of translational cancer research. However, a consequent and critical evaluation is definitely needed. An internationally commonly accepted standardization must be established. Public microarray databases should allow critical comparisons of independent experiences within the same malignant tumor entity. [Stoeckert 2002]. Published investigations should in detail report all key features of the experimental design, the samples used, the extract preparation and labeling performed, hybridization 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 unsolved issues, genomic techniques have taken translational cancer research a significant step forward. In some cancer types such as lymphomas 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 will eventually become the backbone of translational cancer research in the upcoming future.

ACKNOWLEDGEMENTS

This work was partially supported by funds for the Third Term Comprehensive 10-Year Strategy for Cancer Control and a Grant-in-Aid for Scientific Research and for Health and Labour Science Research Grants, Research on Advanced Medical Technology, H14-Toxico-007

REFERENCES

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

Al Moustafa, A. E.; Aloui-Jamali, M. A.; Batist, G.; Hernandez-Perez, M.; Serruya, C.; Alpert, L.; Black, M. J.; Sladek, R. and Foulkes, W. D. (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), 2634-2640.

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, D. G.; Kardia, S. L.; Huang, C. C.; Giordano, T. J.; Levin, A. M.; Misek, D. E.; Lin, L.; Chen, G.; Gharib, T. G.; Thomas, D. G.; Lizyness, M. L.; Kuick, R.; Hayasaka, S.; Taylor, J. M.; Iannettoni, M. D.; Orringer, M. B. and Hanash, S. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **8**, 816-824.

Bonenkamp, J. J.; 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, C. A.; Causton, H. C.; Gaasterland, T.; Glenisson, P.; Holstege, F. C.; Kim, I. F.; Markowitz, V.; Matese, J. C.; 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, R. F.; Bonner, R. F. ; Best, C. J.; Gillespie, J. W.; Flaig, M. J.; Hewitt, S. M.; Phillips, J. L.; Krizman, D. B.; Tangria, M. A.; Ahram, M.; Linehan, W. M.; Knezevic, V. and Emmert-Buck, M. R. (2002) Postanalysis follow-up and validation of microarray experiments. *Nat Genet.* **32**(Suppl.), 509-514.

Churchill, G. A. (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, G. M.; Bontrop, R. E. 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, L. H.; Borg, A.; Ferno, M.; Peterson, C. and Meltzer, P. S. (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, S. W.; Wang, H.; Wu, C.; Wang, H.; Fuller, G. N.; Symmans, W. F.; Shmulevich, I. and 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), 1110-1122.

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. and 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, D. B. (2003) Gene-expression signatures in breast cancer. *N. Engl. J. Med.* **348**, 1715-1717.

Kunkler, I. H. (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, A. G.; Willi, N.; Koivisto, P.; Ferhle, W.; Raffeld, M.; Sauter, G. and Kallioniemi, O. P. (2002) Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. *Cancer Res.* **62**(5), 1256-1260.

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, M. F.; Churchill, G. A. and Crawford, D. L. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261-266.

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

Perou, C. M.; Sorlie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Rees, C. A.; Pollack, J. R.; Ross, D. T.; Johnsen, H.; Akslen, L. A.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, S. X.; Lonning, P. E.; Borresen-Dale, A. L.; Brown, P. O. 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 favourable outcome. *J. Clin. Oncol.* **16**, 63-69.

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

Slonim, D. K (2002) From patterns to pathways: gene expression data analysis comes of age. *Nature genetics* supplement **32**, 502-550.

Schiller, J. H.; Harrington, D.; Belani, C. P.; Langer, C.; Sandler, A.; Krook, J.; Zhu, J. and Johnson, D. H. 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, M. A.; Ross, K. N.; Tamayo, P.; Weng, A. P.; Kutok, J. L.; Aguiar, R. C.; Gaasenbeek, M.; Angelo, M.; Reich, M.; Pinkus, G. S.; Ray, T. S.; Koval, M. A.; Last, K. W.; Norton, A.; Lister, T. A.; Mesirov, J.; Neuberg, D. S.; Lander, E. S.; Aster, J. C. and Golub, T. R. (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* **8**, 68-74.

Simon, R.; Radmacher, M. D.; Dobbin, K. and McShane, L. M. (2003) Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl. Cancer Inst.* **95**, 14-18.

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

Stoeckert, C. J. Jr.; Causton, H. C. and Ball, C. A. (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, P. K.; Downey, T. J.; Spitznagel, E. L. Jr.; Xu, P.; Fu, D.; Dimitrov, D. S.; Lempicki, R. A.; Raaka, B. M. and Cam, M. C. (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, V. G.; 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, M. J.; He, Y. D.; van't Veer, L. J.; Dai, H.; Hart, A. A.; Voskuil, D. W.; Schreiber, G. J.; Peterse, J. L.; Roberts, C.; Marton, M. J.; Parrish, M.; Atsma, D.; Witteveen, A.; Glas, A.; Delahaye, L.; van der Velde, T.; Bartelink, H.; Rodenhuis, S.; Rutgers, E. T.; Friend, S. H. 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, L. J.; Dai, H.; van de Vijver, M. J.; He, Y. D.; Hart, A. A.; Mao, M.; Peterse, H. L.; van der Kooy, K.; Marton, M. J.; Witteveen, A. T.; Schreiber, G. J.; Kerkhoven, R. M.; Roberts, C.; Linsley, P. S.; Bernards, R. and Friend, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-536.

Weiss, M. M.; Kuipers, E. J.; Postma, C.; Snijders, A. M.; Siccama, I.; Pinkel, D.; Westerga, J.; Meuwissen, S. G.; Albertson, D. G. and Meijer, G. A. (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, J. A. Jr.; Marks, J. R. and Nevins, J. R. (2001) Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. USA* **98**, 11462-11467.

White, S. L.; Gharbi, S.; Bertani, M. F.; Chan, H. L.; Waterfield, M. D. and Timms, J. F. (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.

Int. J. Cancer: **116,** 36–44 (2005)
© 2005 Wiley-Liss, Inc.

Establishment of a human non-small cell lung cancer cell line resistant to gefitinib

Fumiaki Koizumi[1,3], Tatsu Shimoyama[1,4], Fumiko Taguchi[1,4], Nagahiro Saijo[2] and Kazuto Nishio[1,4]*
[1]*Shien-Lab, National Cancer Center Hospital, Tokyo, Japan*
[2]*Medical Oncology Department, National Cancer Center Hospital, Tokyo, Japan*
[3]*Investigative Treatment Division, National Cancer Center Research Institute EAST, Kashiwa, Japan*
[4]*Pharmacology Division, National Cancer Center Research Institute, Tokyo, Japan*

The epidermal growth factor receptor (EGFR) tyrosine-kinase inhibitor gefitinib (Iressa®, ZD1839) has shown promising activity preclinically and clinically. Because comparative investigations of drug-resistant sublines with their parental cells are useful approaches to identifying the mechanism of gefitinib resistance and select factors that determine sensitivity to gefitinib, we established a human non-small cell lung carcinoma subline (PC-9/ZD) that is resistant to gefitinib. PC-9/ZD cells are ~180-fold more resistant to gefitinib than their parental PC-9 cells and PC-9/ZD cells do not exhibit cross-resistance to conventional anticancer agents or other tyrosine kinase inhibitors, except AG-1478, a specific inhibitor of EGFR. PC-9/ZD cells also display significant resistance to gefitinib in a tumor-bearing animal model. To elucidate the mechanism of resistance, we characterized PC-9/ZD cells. The basal level of EGFR in PC-9 and PC-9/ZD cells was comparable. A deletion mutation was identified within the kinase domain of *EGFR* in both PC-9 and PC-9/ZD, but no difference in the sequence of *EGFR* cDNA was detected in either cell line. Increased EGFR/HER2 (and EGFR/HER3) heterodimer formations were demonstrated in PC-9/ZD cells by chemical cross-linking and immunoprecipitation analysis in cells unexposed to gefitinib. Exposure to gefitinib increased heterodimer formation in PC-9 cells, but not in PC-9/ZD cells. Gefitinib inhibits EGFR autophosphorylation in a dose-dependent manner in PC-9 cells but not in PC-9/ZD cells. A marked difference in inhibition of site-specific phosphorylation of EGFR was observed at Tyr1068 compared to other tyrosine residues (Tyr845, 992 and 1045). To elucidate the downstream signaling in the PC9/ZD cellular machinery, complex formation between EGFR and its adaptor proteins GRB2, SOS, and Shc was examined. A marked reduction in the GRB2-EGFR complex and absence of SOS-EGFR were observed in PC-9/ZD cells, even though the protein levels of GRB2 and SOS in PC-9 and PC-9/ZD cells were comparable. Expression of phosphorylated AKT was increased in PC-9 cells and inhibited by 0.02 μM gefitinib. But the inhibition was not significant in PC-9/ZD cells. These results suggest that alterations of adaptor-protein-mediated signal transduction from EGFR to AKT is a possible mechanism of the resistance to gefitinib in PC-9/ZD cells. These phenotypes including EGFR–SOS complex and heterodimer formation of HER family members are potential biomarkers for predicting resistance to gefitinib.
© *2005 Wiley-Liss, Inc.*

Key words: resistance; gefitinib; EGFR; Grb2; SOS; non-small cell lung cancer

Chemotherapy has played a central role in the treatment of patients with inoperable NSCLC for over 30 years, although its efficacy seems to be of very limited value.[1,2] Human solid tumors, including lung cancer, glioblastoma, breast cancer, prostate cancer, gastric cancer, ovarian cancer, cervical cancer and head and neck cancer, express epidermal growth factor receptor (EGFR) frequently, and elevated EGFR levels are related to disease progression, survival, stage and response to therapy.[2–10] The therapies directed at blocking EGFR function are attractive.

Interest in target-based therapy has been growing ever since the clinical efficacy of STI-571 was first demonstrated,[11–13] and small molecules and monoclonal antibodies that block activation of the EGFR and HER2 have been developed over the past few decades. The leading small-molecule EGFR tyrosine-kinase inhibitor, gefitinib (Iressa®, ZD1839), has shown excellent antitumor activity in a series of Phase I and II studies,[14,15] and Phase II international multicenter trials (Iressa Dose Evaluation in Advanced Lung Cancer (IDEAL) 1 and 2) yield an overall RR of 11.8–18.4% and overall disease control rate of 42.2–54.4% (gefitinib 250 mg/day) in patients with advanced non-small cell lung cancer (NSCLC) who had undergo at least 2 previous treatments with chemotherapy. INTACT 1 and 2 ('Iressa' NSCLC Trials Assessing Combination Therapy) have demonstrated that gefitinib does not provide improvement in survival when added to standard first line platinum-based chemotherapy *vs.* chemotherapy alone in advanced NSCLC.[16,17] Two small retrospective studies reported recently that activating mutation of *EGFR* correlate with sensitivity and clinical response to gefitinib and erlotinib.[18–20] Although information of *EGFR* mutation may enable to identify the subgroup of patients with NSCLC who will respond to gefitinib and erlotinib, it would be expected that acquired resistance would develop in such patients after treatment. The problem of acquired resistance to gefitinib might be growing, but there has been no preclinical research about the mechanism of developing resistance to gefitinib. We established resistant subline using PC-9 that is highly sensitive to gefitinib.

Establishment of drug-resistant sublines and comparative investigations with their parental cells to identify their molecular, biological and biochemical properties are useful approaches to elucidating the mechanism of the drug's action. Our study describes the establishment of a gefitinib-resistant cell line and its characterization at the cellular and subcellular levels. The PC-9/ZD cell line is the first human NSCLC cell line resistant to gefitinib ever reported. PC-9 is a lung adenocarcinoma cell line that is highly sensitive to gefitinib at its IC_{50}-value of 0.039 μM, but the PC-9/ZD subline, which has a level of EGFR expression comparable to that of PC-9 cells, is specifically resistant to gefitinib. Thus, PC-9 and PC-9/ZD cells will provide useful information about the mechanism of developing resistance to gefitinib and molecules as surrogate markers for predicting chemosensitivity to gefitinib.

Material and methods

Drugs and cells

Gefitinib(*N*-(3-chloro-4-fluorophenyl)-7-methoxy-6-[3-(morpholin-4-yl)propoxy] quinazolin-4-amine) was supplied by AstraZeneca Pharmaceuticals (Cheshire, UK). AG-1478, AG-825, K252a, staurosporin, genistein, RG-14620 and Lavendustin A were purchased from Funakoshi Co. Ltd (Tokyo, Japan).

NSCLC cell line PC-9 (derived from a patient with adenocarcinoma untreated previously) was provided by Prof. Hayata of Tokyo Medical University (Tokyo, Japan).[21] PC-9 and PC-9/ZD cells were cultured in RPMI-1640 medium (Sigma, St. Louis, MO) supplemented with 10% FBS (GIBCO-BRL, Grand Island, NY), penicillin and streptomycin (100 U/ml and 100 μg/ml, respectively; GIBCO-BRL) in a humidified atmosphere of 5%

***Correspondence to:** Shien-Lab, Medical Oncology Department, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan. Fax: +81-3-3547-5185. E-mail: knishio@gan2.res.ncc.go.jp
Received 1 July 2004; Accepted after revision 21 December 2004
DOI 10.1002/ijc.20985
Published online 10 March 2005 in Wiley InterScience (www.interscience.wiley.com).

CO_2 at 37°C. Gefitinib-resistant PC-9/ZD cells were selected from a subculture that had acquired resistance to gefitinib using the following procedure. Cultured PC-9 cells were exposed to 2.5 μg/ml *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) for 24 hr and then washed and cultured in medium containing 0.2 μM gefitinib for 7 days. After exposure to gefitinib, they were washed and cultured in drug-free medium for 14 days. When variable cells had increased, they were seeded in medium containing 0.3–0.5 μM of gefitinib on 96-well cultured plates for subcloning. After 21–28 days, the colonies were harvested and a single clone was obtained. The subcloned cells exhibited an 182-fold increase in resistance to the growth-inhibitory effect of gefitinib as determined by MTT assay, and the resistant phenotype has been stable for at least 6 months under drug-free conditions.

In vitro *growth-inhibition assay*

The growth-inhibitory effects of cisplatin, carboplatin, adriamycin, irinotecan, gemcitabine, vindesine, paclitaxel, genistein, K252a, staurosporin, AG-825, AG-1478, Tyrophostin 51, RG-14620, Lavendustin A and gefitinib in PC-9 and PC-9/ZD cells were examined by using a 3-(4,5-dimethylthiazol-2-yl)- 2,5-diphenyltetrazolium bromide (MTT) assay.[22] A 180 μl volume of an exponentially growing cell suspension (6×10^3 cells/ml) was seeded into a 96-well microtiter plate, and 20 μl of various concentrations of each drug was added. After incubation for 72 hr at 37°C, 20 μl of MTT solution (5 mg/ml in PBS) was added to each well, and the plates were incubated for an additional 4 hr at 37°C. After centrifuging the plates at 200*g* for 5 min, the medium was aspirated from each well and 180 μl of DMSO was added to each well to dissolve the formazan. Optical density was measured at 562 and 630 nm with a Delta Soft ELISA analysis program interfaced with a Bio-Tek Microplate Reader (EL-340, Bio-Metallics, Princeton, NJ). Each experiment was carried out in 6 replicate wells for each drug concentration and carried out independently 3 or 4 times. The IC_{50} value was defined as the concentration needed for a 50% reduction in the absorbance calculated based on the survival curves. Percent survival was calculated as: (mean absorbance of 6 replicate wells containing drugs − mean absorbance of six replicate background wells)/(mean absorbance of 6 replicate drug-free wells − mean absorbance of 6 replicate background wells) × 100.

In vivo *growth-inhibition assays*

Experiments were carried out in accordance with the United Kingdom Coordinating Committee on Cancer Research Guidelines for the welfare of animals in experimental neoplasia (2nd ed.). Female BALB/c nude mice, 6-weeks-old, were purchased from Japan Charles River Co. Ltd (Atsugi, Japan). All mice were maintained in our laboratory under specific-pathogen-free conditions. *In vivo* experiments were scheduled to evaluate the effect of oral administration of gefitinib on pre-existing tumors. Ten days before administration, 5×10^6 PC-9 or PC-9/ZD cells were injected subcutaneously (s.c.) into the back of the mice, and gefitinib (12.5, 25 or 50 mg/kg, p.o.) was administered to the mice on Days 1–21. Tumor diameter was measured with calipers on Days 1, 4, 8, 11, 14, 19 and 22 to evaluate the effect of treatment, and tumor volume was determined by using the following equation: tumor volume = $ab^2/2$ (mm^3) (where *a* is the longest diameter of the tumor and *b* is the shortest diameter). Day "x" denotes the day on which the effect of the drugs was estimated, and Day "1" denotes the first day of treatment. All mice were sacrificed on Day 22, after measuring their tumors. We considered absence of a tumor mass on Day 22 to indicate a cure. Differences in tumor sizes between the treatment groups and control group at Day 22 were analyzed by the unpaired *t*-test. A *p*-value of <0.05 was considered statistically significant.

cDNA expression array

The gene expression profile of PC-9/ZD was assessed with an Atlas Nylon cDNA Expression Array (BD Bioscience Clontech, Palo Alto, CA). Total RNA was extracted by a single-step guanidinium thiocyanate procedure (ISOGEN, Nippon Gene, Tokyo, Japan). An Atlas Pure Total RNA Labeling System was used to isolate RNA and label probes. The materials provided with the kit were used, and the manufacturer's instructions were followed for all steps. Briefly, streptavidin-coated magnetic beads and biotinylated oligo(dT) were used to isolate poly A RNA from 50 μg of total RNA and the RNA obtained was converted into ^{32}P-labeled first-strand cDNA with MMLV reverse transcriptase. The ^{32}P-labeled cDNA fraction was purified on NucleoSpin columns and was added to the membrane on which fragments of 777 genes were spotted. Hybridization was allowed to proceed overnight at 68°C. After washing, the radiolabeled spots were visualized and quantified by BAS-2000II and Array Gauge 1.1 (Fuji Film Co., Ltd., Tokyo, Japan). The data were adjusted for the total density level of each membrane.

Quantitative real-time RT-PCR analysis

Total RNAs extracted from PC-9 cells and PC-9/ZD cells (1×10^6 cells each) were incubated with DNase I (Invitrogen, Carlsbad, CA) for 30 min. First-strand cDNA synthesis was carried out on 1 μg of RNA in 10 μl of a reaction mixture with 50 pmol of Random hexamers and 50 U of M-MLV RTase. Oligonucleotide primers for human *EGFR* were obtained from Takara (HA003051, Takara Bio Co., Tokyo, Japan). For PCR calibration, we generated a calibrator dilution series for *EGFR* cDNA in pUSEamp vector (Upstate, Charlottesville, VA) ranging from 10^8–10^2 copies/1 μl. A total of 2 μl of reverse transcriptase products was used for PCR amplification using Smart Cycler system (Takara) according to manufacturer's instructions. Absolute copy numbers were calculated back to the initial cell numbers, which were set into the RNA extraction. As a result we obtained copies/cell:ratio representing the average *EGFR* RNA amount per cell.

Immunoprecipitation and immunoblotting

The cultured cells were washed twice with ice-cold PBS, and lysed in EBC buffer (50 mM Tris-HCl, pH 8.0, 120 mM NaCl, 0.5% Nonidet P-40, 100 mM NaF, 200 mM Na orthovanadate, and 10 mg/ml each of leupeptin, aprotinin, pepstatin A and phenylmethylsulphonyl fluoride). The lysate was cleared by centrifugation at 15,000 r.p.m. for 10 min, and the protein concentration of the supernatant was measured by BCA protein assay (Pierce, Rockford, IL). The membrane was probed with antibody against EGFR (1005; Santa Cruz, Santa Cruz, CA), HER2/neu (c-18; Santa Cruz), HER3 (c-17; Santa Cruz), HER4 (c-18; Santa Cruz), PI3K (4; BD), Grb2 (81; BD), SOS1/2 (D-21; Santa Cruz), Shc (30; BD, San Jose, CA), PTEN (9552; Cell Signaling, Beverly, MA), AKT (9272; Cell Signaling), phospho-EGFR specific for Tyr 845, Tyr 992, Tyr 1045, and Tyr 1068 (2231, 2235, 2237, 2234; Cell Signaling), phospho-AKT (Ser473) (9271; Cell Signaling), phospho-Erk (9106; Cell Signaling), and phospho-Tyr (PY-20; BD) as the first antibody, and then with by horseradish-peroxidase-conjugated secondary antibody. The bands were visualized by enhanced chemiluminescence (ECL Western Blotting Detection Kit, Amersham, Piscataway, NJ). For Immunoprecipitation, 5×10^6 cells were washed, lysed in EBC buffer, and centrifuged, and the supernatants obtained (1,500 μg) were incubated at 4°C with the anti-EGFR (1005), -HER2 (c-18), and -HER3 (c-17) Ab overnight. The immunocomplexes were absorbed onto protein A/G-Sepharose beads, washed 5 times with lysate buffer, denatured, and subjected to electrophoresis on a 7.5% polyacrylamide gel.

Analysis of the genes of the HER families by direct sequencing

Total RNAs were extracted from PC-9 and PC-9/ZD cells with ISOGEN (Nippon Gene) according to manufacturer's instructions. First-strand cDNA was synthesized from 2 μg of total RNA by using 400 U of SuperScript II (Invitrogen, Carlsbad, CA). After reverse transcription with oligo (dT) primer (Invitrogen) or random primer (Invitrogen), the first-strand cDNA was amplified by PCR by using specific primers for *EGFR*, *HER2* and *HER3*. The

reaction mixture (50 µl) contained 1.25 U AmpliTaq DNA polymerase (Applied Biosystem, Foster City, CA), and amplification was carried out by 30 cycles of denaturation (95°C, 30 sec), annealing (55–59°C, 30 sec), and extension (72°C, 30 sec) with a GeneAmp PCR System 9600 (Applied Biosystem). After amplification, 5 µl of the RT-PCR products was subjected to electrophoretic analysis on a 2% agarose gel with ethidium bromide. DNA sequencing of the PCR products was carried out by the dideoxy chain termination method using the ABI PRISM 310 Genetic Analyzer (Applied Biosystem).

Chemical cross-linking

Chemical cross-linking in intact cells was carried out as described previously.[23] In brief, after 6 hr exposure to 0.2 µM gefitinib, cells were washed with PBS and incubated for 25 min at 4°C in PBS containing 1.5 mM of the nonpermeable cross-linker bis (sulfosuccinimidyl) substrate (Pierce, Rockford, IL). The reaction was terminated by adding 250 mM glycine for 5 min while rocking. Cells were washed in EBC buffer and 20 µg of protein was resolved by 5–10% gradient SDS-PAGE, and then immunoblot analyzed for EGFR, HER2, HER3 and P-Tyr.

Results

Sensitivity of PC-9/ZD cells to cytotoxic agents and tyrosine kinase inhibitors

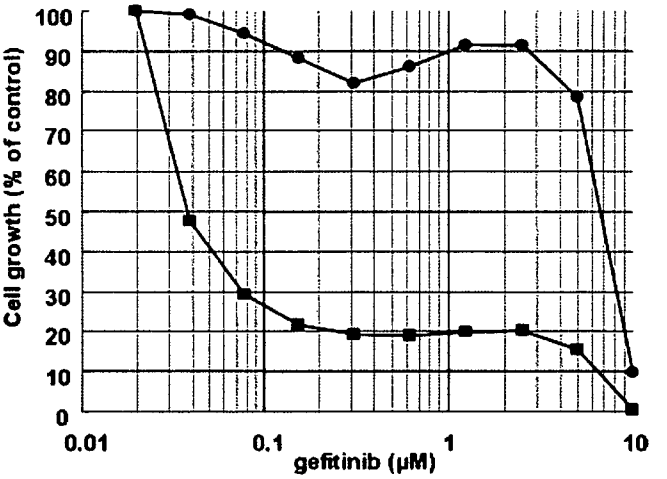
No significant difference between PC-9 and PC-9/ZD cells was observed in *in vitro* cell growth (doubling time of 20.3 hr and 21.4 hr, respectively) and microscopic morphology. Figure 1 shows the growth-inhibitory effect of gefitinib on the parent PC-9 cell line and its resistant subline, PC-9/ZD. The IC_{50}-value of gefitinib in PC-9 cells was 0.039 µM, as compared to 7.1 µM in PC-9/ZD cells (182-fold resistance). PC-9/ZD cells exhibited no cross-resistance to other conventional anticancer agents, including cisplatin, carboplatin, adriamycin, vindesine, paclitaxel and irinotecan. We also examined the growth-inhibitory effect of the EGFR tyrosine kinase inhibitors AG-1478, RG-14620 and Lavendustin A and other tyrosine kinase inhibitors in PC-9 and PC-9/ZD cells. PC-9/ZD cells show cross-resistance to AG1478, but not to all of the tyrosine kinase inhibitors (Tables I, II). It is likely that PC-9/ZD would also be resistant to EGFR-targeted quinazoline derivatives including gefitinib and erlotinib.[20]

PC-9/ZD cells show significant resistance to gefitinib in an in vivo *model*

To ascertain whether the resistance of PC-9/ZD occurs *in vivo*, we investigated the growth-inhibitory effect of gefitinib on PC-9 cells and PC-9/ZD cells in a xenotransplanted model. There was no significant difference in the size of the of PC-9 and PC-9/ZD cell tumor masses in nude mice before the start of gefitinib injection. Figure 2 shows the growth-inhibition curve of PC-9 (Fig. 2*a*) and PC-9/ZD (Fig. 2*b*) cells *in vivo* during the observation period. The PC-9 tumor masses decreased markedly in volume at all doses of gefitinib. In the 50 mg/kg/day p.o. group, the PC-9 masses were eradicated in all mice and did not regrow within the observation period. Growth of the PC-9/ZD masses, on the other hand, was inhibited by gefitinib administration in a dose-dependent manner, but significant tumor reduction was observed only in the 25 and 50 mg/kg/day groups, and the PC-9/ZD masses were not eradicated even in 50 mg/kg/day group. These results clearly demonstrate the significant *in vivo* resistance of PC-9/ZD cells to gefitinib.

Expression of HER family members and related molecules in PC-9 and PC-9/ZD cells

We examined the gene expression and protein levels of HER family members and related molecules by cDNA expression array (followed by confirmation using RT-PCR, data not shown) and immunoblotting. The ratios of the protein expression levels of PC-9 cells to PC-9/ZD cells almost paralleled the expression levels of their genes (Fig. 3*a*). The basal level of EGFR was comparable or slightly higher in PC-9/ZD cells (Fig. 3*a,b*), whereas the HER3 and AKT levels were lower in resistant cells.



	PC-9	PC-9/ZD
IC_{50} value (µM)	0.039± 0.002	7.1± 0.06
Doubling time (hr)	20.3	21.0

FIGURE 1 – Growth-inhibitory effect of gefitinib on PC-9 and PC-9/ZD cells determined by MTT assay. The cells were exposed to the concentrations of gefitinib indicated for 72 hr. The growth–inhibition curves of PC-9 (■) and PC-9/ZD (●) are shown. Doubling time was determined by MTT assay.

TABLE I – CHEMOSENSITIVITY TO OTHER ANTICANCER DRUGS

Drug	IC_{50} values (µM)[1]		RR[2] 1.6
	PC-9	PC-9/ZD	
Cisplatin	1.9 ± 0.7	3.1 ± 1.5	2.0
Carboplatin	25 ± 21	49 ± 23	1.3
Adriamycin	0.16 ± 0.13	0.20 ± 0.15	2.2
Irinotecan	15 ± 10	32 ± 11	1.5
Etoposide	4.5 ± 1.5	6.6 ± 1.3	1.5
Gemcitabine	18 ± 1.5	27 ± 1.5	0.7
Vindesine	0.0046 ± 0.0004	0.0032 ± 0.0009	1.2
Paclitaxel	0.0041 ± 0.0011	0.0048 ± 0.0004	1.6

[1]As assessed by MTT assay in PC-9 and PC-9/ZD cells. Values are the mean ± SD of >3 independent experiments.–[2]Relative resistance value (IC_{50} of resistant cells/IC_{50} of parental cells).

We carried out quantitative RT-PCR to measure the copy numbers of *EGFR*. Estimated transcript levels of *EGFR* were 786.3 and 712.1 copies/cell for PC-9 cells and PC-9/ZD cells, respectively (Fig. 3*d*). Relative ratio of *EGFR* expression levels in PC-9 cells and PC-9/ZD cells is 1.104. Microarray analysis using CodeLink Bioarray (Amersham Bio, Piscataway, NJ) confirmed equivalent gene expression of *EGFR* with ratio of 1.002 between PC-9 and PC-9/ZD cells (data not shown).

Expression of PI3K, Grb2, SOS, and Shc, the adaptor proteins of EGFR, and PTEN was almost the same in PC-9 and PC-9/ZD cells, and no change in the protein levels was observed after exposure to gefitinib (data not shown). The relative densitometric units of each protein are shown in Figure 3*c*. These results suggest that the difference in protein levels of EGFR, HER2, and related proteins can not explain the high resistance of PC-9/ZD cells to gefitinib.

Sequence of HER family member in PC-9/ZD cells

Several reports suggest that the resistance to receptor tyrosine kinase inhibitor STI-571 is partially due to mutations in the

TABLE II – CHEMOSENSITIVITY TO PROTEIN KINASE INHIBITORS[1]

Inhibitor	Target	IC_{50} values (μM)		RR[2]
		PC-9	PC-9/ZD	
AG-1478	EGFR	0.052 ± 0.02	6.0 ± 0.8	117
RG-14620	EGFR	13 ± 1.0	13 ± 2.5	1.0
Lavendustin A	EGFR	20 ± 4.6	27 ± 2.6	1.3
Genistein	TK	18 ± 1.5	27 ± 1.5	1.5
K252a	PKC	0.47 ± 0.17	0.63 ± 0.04	1.3
Staurosporin	PKC	0.0036 ± 0.0019	0.004 ± 0.0014	1.1
AG-825	HER2	>50	>50	

[1]Assessed by MTT assay in PC-9 and PC-9/ZD cells. Values are the mean ± SD of >3 independent experiments.–[2]Relative resistance value (IC_{50} of resistant cells/IC_{50} of parental cells).

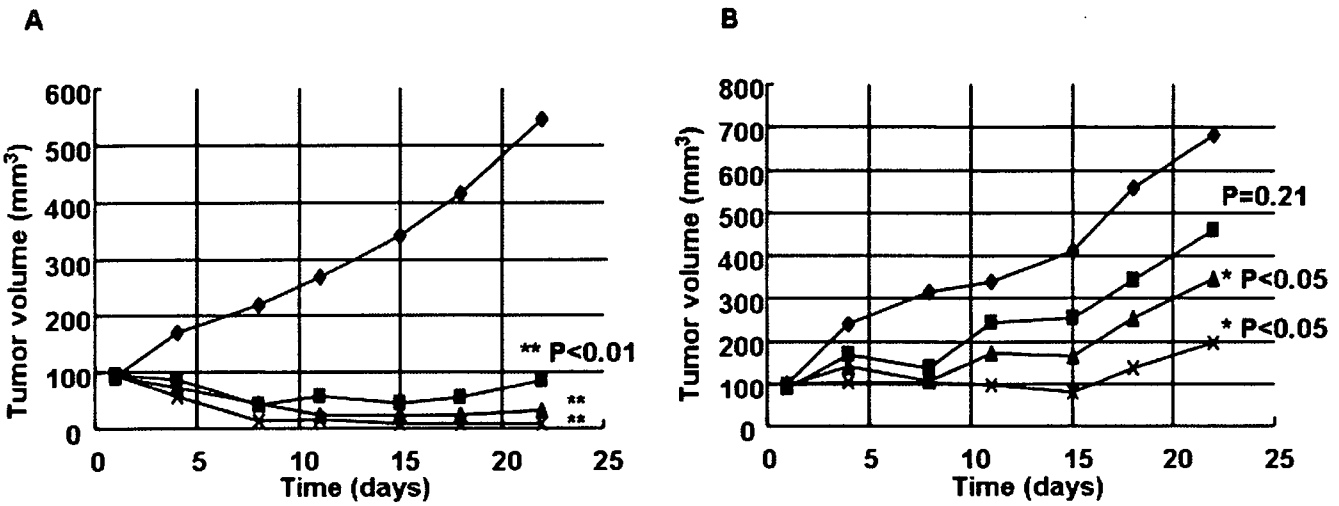


FIGURE 2 – Growth-inhibitory effect of gefitinib on PC-9 and PC-9/ZD cells xenotransplanted into nude mice. Ten days before gefitinib administration, 5×10^6 PC-9 (*a*) or PC-9/ZD (*b*) cells were injected s.c. into the back of mice. The mice were divided into 4 groups (◆, control group; ■, 12.5 mg/Kg group; ▲, 25 mg/Kg group; ×, 50 mg/Kg group). Gefitinib was administered p.o. to the tumor-inoculated mice on Days 1–21. Each group consisted of 6 mice. The statistical analysis was carried out by using the unpaired *t*-test.

ATP-binding site of the Bcr-Abl, the target of the drug.[24–27] We analyzed the sequences of the cDNAs of *EGFR*, *HER2*, and *HER3*, but found no differences in their sequences between PC-9 and PC-9/ZD cells. We did detect a deleted position of *EGFR* in both cell lines that results in deletion of 5 amino acids (Glu722, Leu723, Arg724, Glu725, and Ala726) (Fig. 4). Our findings indicate that the deletion does not directly contribute to the cellular resistance.

Inhibitory effect of gefitinib on autophosphorylation of EGFR in PC-9/ZD cells

Phosphorylation of EGFR is necessary for EGFR-mediated intracellular signaling. Although the EGFR phosphorylation levels of tumors were thought to be correlated with sensitivity to gefitinib, the basal level of phosphorylated EGFR in PC-9 and PC-9/ZD cells is almost the same. Gefitinib inhibited EGFR autophosphorylation in a dose-dependent manner and completely inhibited its phosphorylation at 0.2–2 μM in PC-9 cells (Fig. 5*a*), but its inhibitory effect on autophosphorylation of EGFR in PC-9/ZD cells was less than in PC-9 cells (Fig. 5*a*). Because each phosphorylation site of EGFR has a different role in the activation of downstream signaling molecules, we examined the inhibitory effect of gefitinib on site-specific phosphorylation of EGFR. Phosphorylation of several different EGFR tyrosine residues (Tyr845, Tyr992 and Tyr1068) was dose-dependently inhibited by gefitinib in PC-9 cells, whereas no clear inhibitory effects of gefitinib on phosphorylation at Tyr 845 and Tyr1068 residues in PC-9/ZD cells was observed (Fig. 5*b*,*c*,*e*). The most marked difference of in inhibition between the cells was observed at Tyr1068 (Fig. 5*e*). Tyr1045 showed resistance to inhibition of autophosphorylation by gefitinib in both PC-9 and PC-9/ZD cells (Fig. 5*d*).

Complex formation of EGFR and its adaptor proteins

Tyr1068 of EGFR is the tyrosine that is most resistant to inhibition of autophosphorylation by gefitinib in PC-9/ZD cells. Because the Tyr 1068 is a direct binding site for the GRB2/SH2 domain, and its phosphorylation is related to the complex formation of EGFR-adaptor proteins and their signaling, we examined complex formation between EGFR and the adaptor proteins GRB2, SOS, Shc, and PI3K by immunoprecipitation. The level of expression of these proteins in PC-9 and PC-9/ZD cells were similar (Fig. 3*a*). A smaller amount of EGFR-GRB2 complex was observed in PC-9/ZD cells and no EGFR-SOS complex was detected at all (Fig. 6). The amount of HER2- or HER3-GRB2 complex in PC-9 and PC-9/ZD cells was similar, and no decreases in complex formation were observed after exposure to gefitinib. A decreased amount of HER2-SOS complex and inability to detect HER3-SOS complex were also observed in PC-9/ZD cells. HER2-PI3K complex increased in PC-9/ZD. There are no significant differences in complex formation between SHC and EGFR, HER2, or HER3 between PC-9 and PC-9/ZD cells. These results suggest that GRB2-SOS-mediated signaling may be inactivated in PC-9/ZD cells.

Heterodimerization of HER family member in PC-9/ZD cells

Dimerization of members of the HER family is essential for activation of their catalytic activity and their signaling. We examined the effect of gefitinib on the dimerization of HER family members by immunoblotting, immunoprecipitation and chemical cross-linking analysis (Figs. 3*a*, 5*a*, 7*a*). The expression levels of EGFR and HER2 were similar and the HER3 level was lower in PC-9/ZD cells by immunoblotting (Fig. 3*a*). A chemical cross-

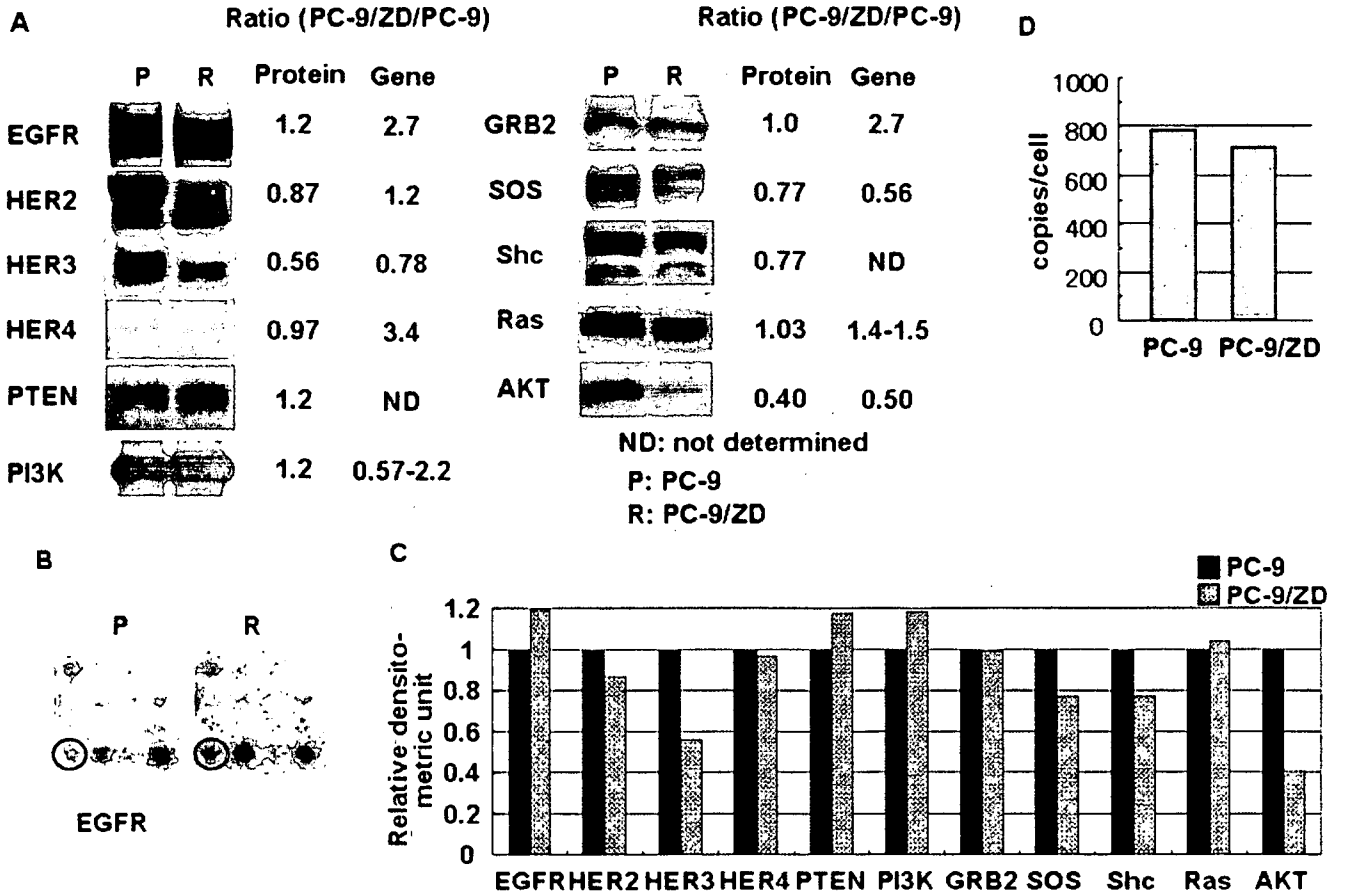


FIGURE 3 – Expresssion of HER family members and related molecules in PC-9 (P) and PC-9/ZD (R) cells. (*a*) Western blot analysis; a 20 μg sample of total cell lysates was separated by SDS-PAGE, transferred to a PVDF membrane, and incubated with a specific anti-human antibody as the first antibody and then with horseradish peroxidase-conjugated secondary antibody. The ratios of the levels of expression of proteins and genes in PC-9 cells to the levels in PC-9/ZD cells are shown. (*b*) cDNA expression array; Poly A RNA was converted into ^{32}P-labeled first-strand cDNA with MMLV reverse transcriptase. The ^{32}P-labeled cDNA fraction was hybridized to the membrane on which fragments of 777 genes were spotted. The close-up view shows *EGFR* mRNA expression. (*c*) Each band was quantified by a densitometry and with NIH image software. The levels of protein expression are shown in a graph. (*d*) Absolute amounts of *EGFR* transcripts of PC-9 cells and PC-9/ZD cells measured by real-time quantitative RT-PCR. The values were calculated back to the initial cell numbers for RNA extraction in Material and Methods.

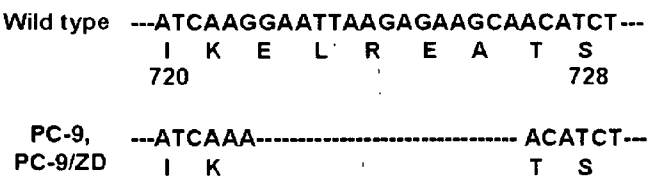


FIGURE 4 – Detection of a deleted position of EGFR. Direct sequencing of a PC-9 and PC-9/ZD-derived, amplified cDNA fragment containing the ATP-binding site of EGFR. *Top*, wild-type EGFR; *bottom*, PC-9 and PC-9/ZD.

linking assay showed that in the absence of gefitinib the amount of high molecular weight complexes (~400 kDa) that are recognized by anti-EGFR antibody (EGFR dimers), including formations of homodimers and heterodimers (EGFR-EGFR, EGFR-HER2 or EGFR-HER3), was almost the same in PC-9 and PC-9/ZD cells, whereas HER2 dimerization detected by anti-HER2 antibody was remarkably lower in PC-9/ZD cells (Fig. 7*a*). Increased EGFR/HER2 (and EGFR/HER3) heterodimer formation was detected in PC-9/ZD cells by immunoprecipitation analysis (Fig. 5*a*). The proportion of EGFR heterodimer to homodimer is increased significantly in PC-9/ZD (Fig. 7*b*). When exposed to gefitinib at a concentration of 0.2 μM for 6 hr the amount of dimer-formation increased similarly in PC-9 and PC-9/ZD cells (Fig. 7*a*), whereas marked induction of hetero-dimerization of EGFR-HER2 was observed only in PC-9 cells (Fig. 5*a*). These results suggest that a difference in hetero- or homo-dimerization is a possible determinant factor of gefitinib sensitivity.

AKT and MAPK pathways in PC-9/ZD cells

Because phosphorylation at Tyr 1068 of EGFR plays an important role for transduction of the signal to downstream of MAPK and AKT pathway,[28,29] we examined the difference between PC-9 and PC-9/ZD cells in downstream signaling. The basal level of phosphorylated AKT is higher in PC-9 cells than in PC-9/ZD cells, and although gefitinib inhibited AKT phosphorylation in a dose-dependent manner (Fig. 8*a*), the inhibitory effect of gefitinib on phosphorylation of AKT in PC-9/ZD cells was significantly less than in PC-9 cells (Fig. 8*a*). This difference in the inhibitory effect of gefitinib on AKT phosphorylation between PC-9 and PC-9/ZD cells is very similar to the difference in effect on EGFR autophosphorylation. No inhibition of phosphorylation of MAPK by gefitinib was observed in either cell line (Fig. 8*b*). These results suggest that downregulation of activated AKT is closely correlated with the cellular sensitivity to gefitinib, but that inhibition of the MAPK pathway does not contribute to drug sensitivity.