Review Article

The role of gene expression profiling in early-stage non-small cell lung cancer

Wenlong Shao[1,2,3], Daoyuan Wang[2,3], Jianxing He[1,2,3]

[1]Guangdong Cardiovascular Institute, Guangdong General Hospital; [2]Department of Cardiothoracic Surgery, The First Affiliated Hospital of Guangzhou Medical College; [3]Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, Guangzhou, Guangdong Province, China

ABSTRACT For patients with identical clinical-pathological characteristics or the same stage of lung cancer, great uncertainties remain regarding how some patients will be cured while other patients will have cancer recurrence, metastasis, or death after surgical resection. Identification of patients at high risk of recurrence, those who are unlikely to respond to specific chemotherapeutic agents, is the rationale for measuring specific biochemical markers. Thus, main investigational studies nowadays are focused in identifying molecular markers of recurrence, beyond pathologic stage, after surgical treatment and factors that can predict a benefit from adjuvant chemotherapy in poor prognosis subgroups, to individualize treatments. Advances in genomics and proteomics have generated many candidate markers with potential clinical value. Gene expression profiling (GEP) by microarray or real-time quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR) can be useful in the classification or prognosis of various types of cancer, including lung cancer. A number of prognostic gene expression signatures have been reported to predict survival in non-small cell lung cancer (NSCLC). In this review, we focus on the role of GEP in early-stage NSCLC as predictive and prognostic biomarker and its potential use for a 'personalized' medicine in the years to come.

Key Words: non-small cell lung cancer; prognostic biomarker; gene expression profiling

J Thorac Dis 2010; 2: 89-99.

Introduction

The incidence of lung cancer in 2007 is estimated to be 213,380 with 160,390 deaths in the U.S. It will contribute to 31% of male and 26% of female cancer-related deaths and is the largest cause of cancer-related mortality in both men and women. Non-small-cell lung cancer (NSCLC) accounts for approximately 75% of all cases of lung cancer, which is one of the most common tumors affecting humans in the world (1). The current standard of treatment for patients with stage I NSCLC is surgical resection (2-4), despite the observation that nearly 30 to 35 percent will relapse after the initial surgery and thus have a poor prognosis (5,6), similarly, as many as 66% of stage II and 75% of stage IIIA patients will develop recurrence and die as a result of their disease within 5 years of resection (3,4), indicating that a subgroup of these patients might benefit from adjuvant chemotherapy. Thus, the ability to identify subgroups of patients more accurately may improve health outcomes across the spectrum of disease.

Although other clinical and pathologic markers have prognostic significance (7-10), the clinic-pathologic staging system has been the standard for determining NSCLC prognosis (8). But this classification scheme is probably an imprecise predictor of the prognosis of an individual patient. For patients with identical clinical-pathological characteristics or the same stage of lung cancer, great uncertainties remain regarding how some patients will be cured while other patients will have cancer recurrence, metastasis, or death after surgical resection. Identification of patients at high risk of recurrence, those who are unlikely to respond to specific chemotherapeutic agents, is the rationale for measuring specific biochemical markers. Thus, main investigational studies nowadays are focused in identifying molecular markers of recurrence, beyond pathologic stage, after surgical treatment and factors that can predict a benefit from adjuvant chemotherapy in poor prognosis subgroups, to individualize treatments. This ability to identify subgroups of patients more accurately

This paper received no research funding. No conflict of interest.
Corresponding to: Jianxing He, MD, PhD, FACS. Department of Cardiothoracic Surgery, the First Affiliated Hospital of Guangzhou Medical College, No. 151, Yanjiang Rd, Guangzhou 510120, Guangdong Province, PR China. Tel:+86-20-83337792; Fax: +86-20-83350363, Email: drjianxing.he@gmail.com

Submitted May 2, 2010. Accepted for publication May 23, 2010.
Available at www.jthoracdis.com

ISSN: 2072-1439 © 2010 Journal of Thoracic Disease. All rights reserved.

Tab 1: Examples of Analytical Procedures Used for Genomic and GEP

A. Genome (DNA)
• DNA sequencing
• DNA profiling/genotyping (e.g., mutations, single-nucleotide polymorphisms, repeated sequence, etc.)
• Gene mapping
B. Transcriptome (mRNA)
• Northern blot
• Nuclease protection assay
• Reverse transcription–polymerase chain reaction (RT-PCR)
• In situ hybridization/tissue arrays
• Differential display PCR
• cDNA/expressed sequence tag (EST) libraries
• Serial analysis of gene expression (SAGE)
• DNA microarrays
C. Proteome (Protein)
• Separation: chromatography, electrophoresis, mass spectrometry
• Immunoassays: radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), Western blot, immunocytochemistry
• Function/activity assays: enzyme activity, binding affinity, etc.
• Combined techniques: high-performance liquid chromatography/mass spectrometry (LC/MS), electrophoresis/ mass spectrometry, protein tagging/mass spectrometry, tissue arrays/immunocytochemistry, protein chips/SELDI-TOF mass spectrometry
• Protein microarrays
• Direct mass imaging

may improve health outcomes across the spectrum of disease. The study of molecular factors that influence drug responsiveness is also a potentially promising approach to decrease treatment toxicity and costs by avoiding the administration of ineffective therapy to patients destined not to benefit (11).

Advances in genomics and proteomics have generated many candidate markers with potential clinical value (12). Gene expression profiling (GEP) by microarray or quantitative real-time reverse-transcriptase polymerase chain reaction (qRT-PCR) can be useful in the classification or prognosis of various types of cancer, including lung cancer (13-15). A number of prognostic gene expression signatures have been reported to predict survival in NSCLC. In this review, we will focus on the role of GEP in NSCLC as predictive and prognostic biomarker and its potential use for a 'personalized' medicine in the years to come.

Techniques for gene expression profiling

GEP (i.e., the systematic identification and characterization of those genes activated or expressed in a cell) can be conducted on different levels depending on the specific research objectives. This could involve analysis of DNA, mRNA, and/or protein as a measure of gene expression. A variety of techniques that have been used to profile the genome and to assess gene activity at the mRNA or protein level (Table 1).

Microarray-based molecular profiles

Both genetic changes within a precancerous cell and epigenetic changes in the tumor microenvironment are thought to promote tumorigenesis (16-18). In particular, it is now well accepted that alterations in the expression levels of certain genes strongly correlate with and are considered causative for cancer (19). These changes in gene expression are reflected by quantitative changes in mRNA levels. Detecting these changes was traditionally done by identifying single genes of interest and assaying mRNA expression by techniques such as PCR and Northern blotting. In 1995, Schena et al described a GEP technique adapted from Southern blotting that used strands of cDNA spotted onto a piece of glass to examine multiple mRNA expression levels at once (20). Known as a microarray, this technology was quickly developed into a tool that could be used to take a genome-wide snapshot of mRNA transcription levels within a tissue of interest in a single experiment (21).

qRT-PCR–based molecular signatures

In response to the concern that microarray-based profiles are difficult to translate into a clinical setting, several recent efforts have focused on developing qRT-PCR

–based molecular signatures. It is unlikely that every gene in the molecular profiles obtained by microarray analysis has equal relevance with respect to prognosis. Ideally, a handful of genes could be isolated that convey close to the same prognostic information as microarray-based gene signatures. The disadvantage would be that only a small number of such genes could be tested by the current gold standard assay for gene expression, qRT-PCR. However, qRT-PCR has significant advantages to microarray-based assays, including widespread availability, cost, simplicity, reproducibility, and ability to use stored paraffin-embedded versus snap-frozen tissues (22-24). In addition, the limited number of genes in qRT-PCR–based signatures allows these signatures to be validated with protein expression by immunohistochemistry (25).

Lung developmental pathways in lung cancer

Current paradigms suggest that lung carcinomas arise from pluripotent stem and progenitor cells capable of differentiation into one or several histologic cell types. These paradigms suggest that lung tumor cell ontology is determined by the consequences of gene transcriptional activation and/or repression events that recapitulate embryonic lung development (26,27). The hypothesis that lung cancer arises from aberrant expression of genes involved in lung development is supported by gene expression studies demonstrating similarities between signatures obtained from human lung tumors and signatures characteristic of normal lung development. In an analysis of 32 NSCLC specimens and 7 normal specimens, unsupervised hierarchical analysis segregated tumors on the basis of histologic type and differentiation (28). Supervised clustering analysis of tumors identified numerous genes with known important function in embryonic lung development. Comparison of human lung tumor histology classifiers with genes temporally activated during mouse lung development reveals that genes expressed by large cell carcinoma (LCC) are similarly expressed during the early pseudoglandular and canalicular stages of lung development, while those expressed by adenocarcinoma mirror those expressed during the later terminal sac and alveolar stages. In addition to highlighting the expression of proliferation-associated genes by LCC and of differentiation-associated genes by adenocarcinoma, these results suggest a recapitulation of developmentally regulated pathways in lung tumors. In addition, Glinsky and colleagues reported that a gene signature of 'stemness' derived from BMI-1–regulated genes in normal stem cells is associated with metastasis and survival in several tumor types, including NSCLC (29). Taken together, these observations suggest that poor differentiation is linked to molecular parameters of early development representing lung stem and progenitor cell programs, and that gene signatures of these phenotypes are important for lung cancer differentiation, progression, and clinical outcome.

Predicting response to treatment by gene expression profiling

GEP has been used to predict response to treatment. The first clinical study of microarray as a predictor of benefit from chemotherapy in NSCLC used tissues from 133 patients enrolled in the JBR.10 study. JBR.10 is a North American phase III Intergroup trial led by the National Cancer Institute of Canada Clinical Trials Group (NCIC CTG), in which 482 patients with completely resected stages IB and II—excluding T3N0-NSCLC were randomly assigned to receive four cycles of adjuvant cisplatin plus vinorelbine or observation alone (30). Chemotherapy-treated patients enjoyed a significant survival advantage (HR, 0.70; P=0.03), although a significant interaction with stage was seen, with benefit limited to stage II patients. By use of a supervised analysis, a 15-gene signature that correlated with survival, and was independent of stage, histology, age, and sex was derived from patients in the observation group (HR 15.02, 95% CI 5.12–44.04; p<0.0001). In the high-risk group, treatment with vinorelbine plus cisplatin conferred significant survival benefit compared with observation alone (0.33, 0.17–0.63; p=0.0005), whereas in the low-risk group, patients who received this chemotherapy regimen had shorter survival compared with observation alone (3.67, 1.22–11.06; p=0.0133). This interaction was highly significant (p=0.0001) (31). If the 15-gene signature is validated by further testing, it may improve the current method for deciding which patients should receive adjuvant chemotherapy.

Staunton et al. used DNA microarrays to measure gene expression in the NCI-60 panel (a collection of 60 human cancer cell lines) (32-34). By combining the untreated gene expression profile of each cell line together with information about each cell lines' chemosensitivity profile, they were able to predict drug sensitivity in an independent test set of cell lines. A subsequent study by Potti et al. (35) repeated and built upon Staunton's work. Potti and colleagues used molecular profiles from cell lines to establish sensitivity to chemotherapy. The signature that predicted response to individual agents was then further validated in cell lines, but also in clinical samples from patients with other tumor types (35). The usefulness of this approach is that one tumor sample can be interrogated for response to many agents on the basis of cell-line derived signatures. For example, a relationship between

docetaxel resistance and deregulation of the PI3-kinase pathway was observed. Using a panel of 17 NSCLC cell lines a significant association was found between docetaxel resistance and sensitivity to a PI3-kinase inhibitor (LY-294002), suggesting its use as a second-line therapy (36).

For many years, we have been discovering that expression of certain genes or the presence of certain gene mutations has implications in the prognosis of NSCLC or response to specific therapy. The improved responses seen with the use of tyrosine kinase inhibitors (TKIs; eg, erlotinib or gefitinib) in patients carrying mutations in the epidermal growth factor receptor (EGFR) gene are a good example of an attempt to stratify tumors that are more sensitive to these agents (37). We know that no more than 10% of the general population will have a response to these agents; however, when only selected patients who carry gene mutations are treated, the response rate to these agents can be as high as 70% (38,39), and now we are still trying to define which are the best techniques (gene mutation analysis, FISH, or others) to detect these mutations and moving these discoveries into our clinical practice. Similar efforts to identify predictive markers for the EGFR inhibition have been undertaken in the area of proteomics (40,41).

Recently, Altorki et al. examined safety and efficacy of short-term, preoperative pazopanib (an oral angiogenesis inhibitor targeting VEGFR, platelet-derived growth factor receptor, and c-kit; 800 mg/d for 2 to 6 weeks) monotherapy in 35 patients with operable stage I/II NSCLC, and gene-expression profiling was performed on 77 pre- and post-treatment lung samples from 34 patients. They found that several target genes were dysregulated after pazopanib treatment, validating target-specific response and indicating a persistent pazopanib effect on lung cancer tissue (42). In further study, they conducted a broad profiling of cytokine and angiogenic factors (CAF) to investigate the relationship between baseline CAF levels, CAF changes during treatment, and tumor shrinkage. Plasma samples were collected before treatment and on the last day of therapy from 33 patients with early-stage NSCLC. Levels of 31 CAFs were measured by suspension bead multiplex assays or ELISA and correlated with change in tumor volume. Pazopanib therapy was associated with significant changes of eight CAFs; sVEGFR2 showed the largest decrease, whereas placental growth factor underwent the largest increase. Increases were also observed in stromal cell-derived factor-1alpha, IP-10, cutaneous T-cell-attracting chemokine, monokine induced by IFN-gamma, tumor necrosis factor-related apoptosis-inducing ligand, and IFN-alpha. Posttreatment changes in plasma sVEGFR2 and interleukin (IL)-4 significantly correlated with tumor shrinkage. Baseline levels of 11 CAFs significantly correlated with tumor shrinkage, with IL-12 showing the strongest association. Using multivariate classification, a baseline CAF signature consisting of hepatocyte growth factor and IL-12 was associated with tumor response to pazopanib and identified responding patients with 81% accuracy. These data suggest that CAF profiling may be useful for identifying patients likely to benefit from pazopanib, and merit further investigation in clinical trials (43).

Predicting survival and recurrence by gene expression profiling

GEP has been used to predict response to treatment and patients' outcome (13,31,44-68). Beer et al. analyzed the genetic profile in 86 patients with primary lung adenocarcinoma, and found that the genes most associated with survival were identified to create a risk index based on the top 50 genes that separated patients into high-risk and low-risk groups. When applying this risk predictor to a test data set of 62 stage I patients from another study, they were able to predict survival with statistical significance difference (P=0.006) (51). This study also identified certain patients with stage I along with stage III disease with poor prognosis based on gene profile. This demonstrated the ability for GEP to identify a patient with poor prognosis that is independent of the stage at the time of diagnosis.

Guo et al. devised a computational model system that redicted the clinical outcome of individual patients based on their GEP. A 37-gene signature was created, and the authors studied a cohort of 86 patients diagnosed with lung adenocarcinoma. The gene signature was then applied to predict the survival of the other 84 patients with adenocarcinoma. The predictive accuracy of the gene signature was 96%. The cluster analysis, using the 37-gene signature, aggregated the test patient samples into 3 groups with good (mean survival, 66.9 months), moderate (mean survival, 27.6 months), and poor (mean survival, 22.4 months) prognoses (Kaplan-Meier analysis; $P < .0005$; log-rank test) (Fig 1). Notably, when the results were reviewed, all patients who had grouped together in cluster 1 (good prognosis) had stage I disease, with N0 lymph node status (no metastasis) and smaller tumor size (T1 or T2) (63).

Landmark studies such as the one conducted by Potti et al. from Duke University have identified GEP, which predicted the risk of recurrence following surgery from a cohort of patients with early-stage NSCLC (52). The accuracy was > 70%. The investigators were also able to identify a subgroup of patients with stage IA disease who were at high risk for recurrence, with a very poor survival, and who might be suitable for adjuvant chemotherapy. This

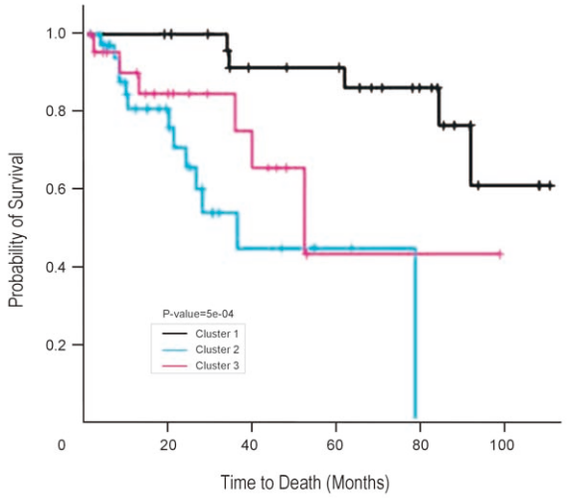

Fig 1. Kaplan-Meier survival analysis of three clusters of patients. Average survival time of patients in cluster 1, 66.9 months; average survival time of patients in cluster 2, 22.4 months; average survival time of patients in cluster 3, 27.6 months (P<0.0005, log--rank test) (ref 63).

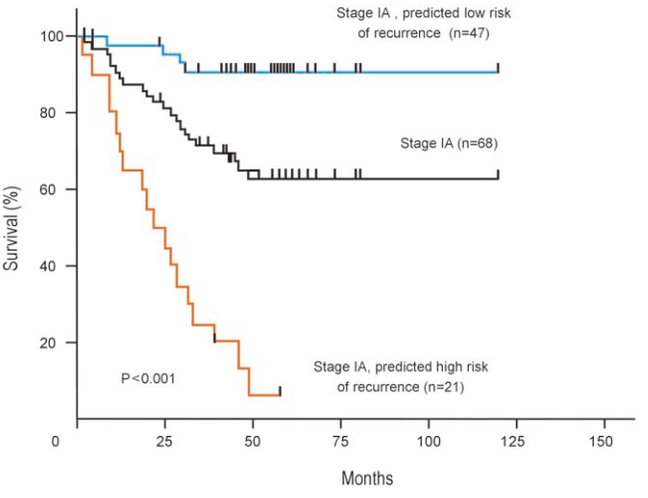

Fig 2. Kaplan–Meier survival estimates for a group of patients with stage IA disease from the Duke, ACOSOG, and CALGB cohorts and the subgroups predicted to have either a high probability (>0.5) or a low probability (≤0.5) of recurrence. P values were obtained with the use of a log-rank test. Tick marks indicate patients whose data were censored by the time of last follow-up or owing to death (ref 52).

is clinically relevant when the current standard of care for patients with stage IA disease is just clinical observation (no adjuvant chemotherapy is offered) because of a 70% chance of 5-year survival. This genetic strategy was then validated in two separate cohorts from multicenter cooperative group trials: 25 patients from the American College of Surgeons Oncology Group Z0030 study and 84 from the prospective CALGB 9761 trial, this genomic strategy had an overall predictive accuracy of 72 and 79%, respectively. This gene expression profile also was applied to 68 patients with stage IA disease, who are not usually candidates for adjuvant chemotherapy. Kaplan-Meier survival curves were generated for the group as a whole and for the subgroups predicted to be at high or low risk for recurrence by the lung metagene model. Although the survival rate for the group was approximately 70% at 4 years, the survival rate for those predicted to be at low risk was 90% and less than 10% for those predicted to be at high risk, thus identifying the subgroup of patients with stage IA NSCLC at high risk of recurrence, who might benefit from adjuvant chemotherapy (Fig 2).

In another important study from Taiwan University (13), authors examined the expression of multiple genes associated with invasive activity in frozen specimens of lung-cancer tissue from 125 randomly selected patients who underwent surgical resection of NSCLC and not received adjuvant chemotherapy, to identify a gene signature that is correlated with clinical outcome.

Sixteen genes were initially identified by analyzing microarray data and then confirmed by RT-PCR. From these, the authors further identified five genes that were significantly associated with survival. The levels of expression of these five genes were used to construct a decision tree to classify patients as having a high-risk gene signature or a low-risk gene signature. The five selected genes were: dual-specificity phosphatase 6 (DUSP6), monocyte-to-macrophage differentiation-associated protein (MMD), signal transducer and activator of transcription 1 (STAT1), v-erb b2 avian erythroblastic leukemia viral oncogene homolog 3 (ERBB3), and lymphocyte-specific protein tyrosine kinase (LCK).

The authors identified 59 patients with high-risk gene signatures and 42 with low-risk gene signatures, according to gene expression as measured with RT-PCR and decision-tree analysis. The five-gene signature was strongly associated with OS (sensitivity 98%; specificity 93%; positive predictive value 95%; negative predictive value 98%; and overall accuracy 96%). The presence of a high-risk five-gene signature in the NSCLC tumors was associated with an increased risk of recurrence and decreased OS. With a median follow-up of 20 months, the patients with a high-risk gene signature had a shorter median OS than the patients with a low-risk gene signature (20 months versus 40 months, P<0.001). The high-risk gene signature was associated with a median RFS of 13 months, whereas the low-risk gene signature was associated with a

median RFS of 29 months (P=0.002).

According to multivariate regression analysis, the high-risk five-gene signature, tumor stage III and older age were significantly associated with death from any cause among the 101 patients, and the high-risk five-gene signature and tumor stage III were significantly associated with recurrence of cancer as well (HR for the high-risk signature versus the low-risk signature, 1.92; 95% CI, 1.06 –3.46; P=0.03). In a subgroup analysis of 59 patients with stage I or II disease, those with a high-risk gene signature had a shorter OS and a shorter RFS than those with a low-risk gene signature. These results were validated in an independent cohort of 60 patients with NSCLC and with the use of a set of published microarray data from 86 patients from a Western population with NSCLC.

The identification of five genes that are closely associated with the outcomes in patients with NSCLC has clinical implications. Patients who have tumors with a high-risk gene signature could benefit from a cisplatin-based adjuvant chemotherapy, whereas those with a low-risk gene signature could be spared what may be unnecessary treatment. Prospective, large scale, multicenter studies are necessary to test this idea. These five genes that can predict the clinical outcome in patients with NSCLC may also reveal targets for the development of therapy for lung cancer. STAT1 causes arrested growth and apoptosis in many types of cancer cells by inducing the expression of p21WAF1 and caspase (53,54). MMD is preferentially expressed in mature macrophages (55). Some studies have shown that macrophage activation promotes cancer metastasis (56), although the function of the MMD protein is unknown. DUSP6 inactivates extracellular signal-regulated kinase 2 (also known as mitogen-activated protein kinase 1), resulting in tumor suppression and apoptosis (57). ERBB3, a member of the epidermal growth factor receptor family of tyrosine kinases, can shorten cell survival (58). LCK, a member of the Src family of protein tyrosine kinases, is expressed mainly in T cells and is one of the first signaling molecules downstream of the T-cell receptor. It plays a key role not only in the differentiation and activation of T cells but also in the induction of apoptosis (59). In addition, LCK is expressed in many cancers and regulates the mobility of cancer cells (60,61).

Bianchi et al. proposed a qRT-PCR–based 10-gene molecular signature for adenocarcinoma (46). They selected 49 unbiased genes based on a meta-analysis of previously published adenocarcinoma microarray data and combined this with a biased set of 31 additional genes selected from the literature demonstrated to either be important for tumorigenesis and/or to represent prognostic lung cancer markers. These 80 genes were tested on a training cohort of stage I adenocarcinoma patients using a leaveone-out validation model yielding a 10-gene signature. In two separate validation cohorts of stage I adenocarcinoma patients, this 10-gene signature was more accurate than stage (IA vs. IB), age, sex, differentiation, or presence of a K-ras mutation in predicting survival. In addition, it also demonstrated differences in survival when applied to separate cohorts of stage IA and stage IB patients with adenocarcinoma but, similar to the findings by Chen et al. (13), did not demonstrate significant predictive differences in stage II or III adenocarcinomas.

Lau et al. proposed a qRT-PCR–based 3-gene signature for NSCLC (45). One hundred twenty-eight candidate genes were identified using data from 7 previous microarray based profiling studies and assayed by qRT-PCR in 147 frozen NSCLC samples. Using a statistical method based on concordance index and risk scores, a 3-gene signature (STX1A, CCR7, and HIF1A) was developed. When applied to their own training cohort as well as to two cohorts from previously published microarray data sets, they demonstrated a statistically significant difference in survival between patients with stage I NSCLC classified as having either good or poor prognosis. In agreement with the above studies, this difference did not hold true for patients with stage II disease. They also demonstrated that their 3-gene signature was better at predicting survival in their training cohort stage I patients than stage, histology, or smoking status.

Skrzypski et al. examined the expression pattern of 29 genes selected by cDNA studies to test their clinical prognostic value in early-stage squamous cell carcinoma (SCC) of the lung (49). From 2000 to 2004, freshly frozen primary tumor specimens were obtained at the time of the surgery from 66 SCC patients and gene expression of the 29 genes was assessed by quantitative RT-PCR using low-density arrays. Expression values were dichotomized using the median value as the cutoff. The univariate analysis showed 10 genes with prognosis value: PH4 (P=0.01); macrophage-colony stimulating factor (CSF1), which attracts macrophages and induce them to express EGF (P= 0.002); EGFR (P=0.05); KIAA0974 (P=0.02); ANLN (P=0.02); carbonic anhydrase IX (CA IX), which is regulated by hypoxia and plays a role in chemoresistance (P=0.007); VEGFC (P=0.03); neurotrophic tyrosine receptor kinase 1 (P=0.04); fibronectin (P=0.002); insulin receptor (P= 0.03). In the multivariate analysis of survival, CSF1, EGFR and CA IX, and tumor size emerged as significant variables (P=0.005, 0.02, <0.0001, 0.02, respectively).

Roepman et al. aimed to develop a gene expression profile for stage I and stage II NSCLC, allowing identification of patients with a high risk of disease recurrence within 2 to 3 years after initial diagnosis.

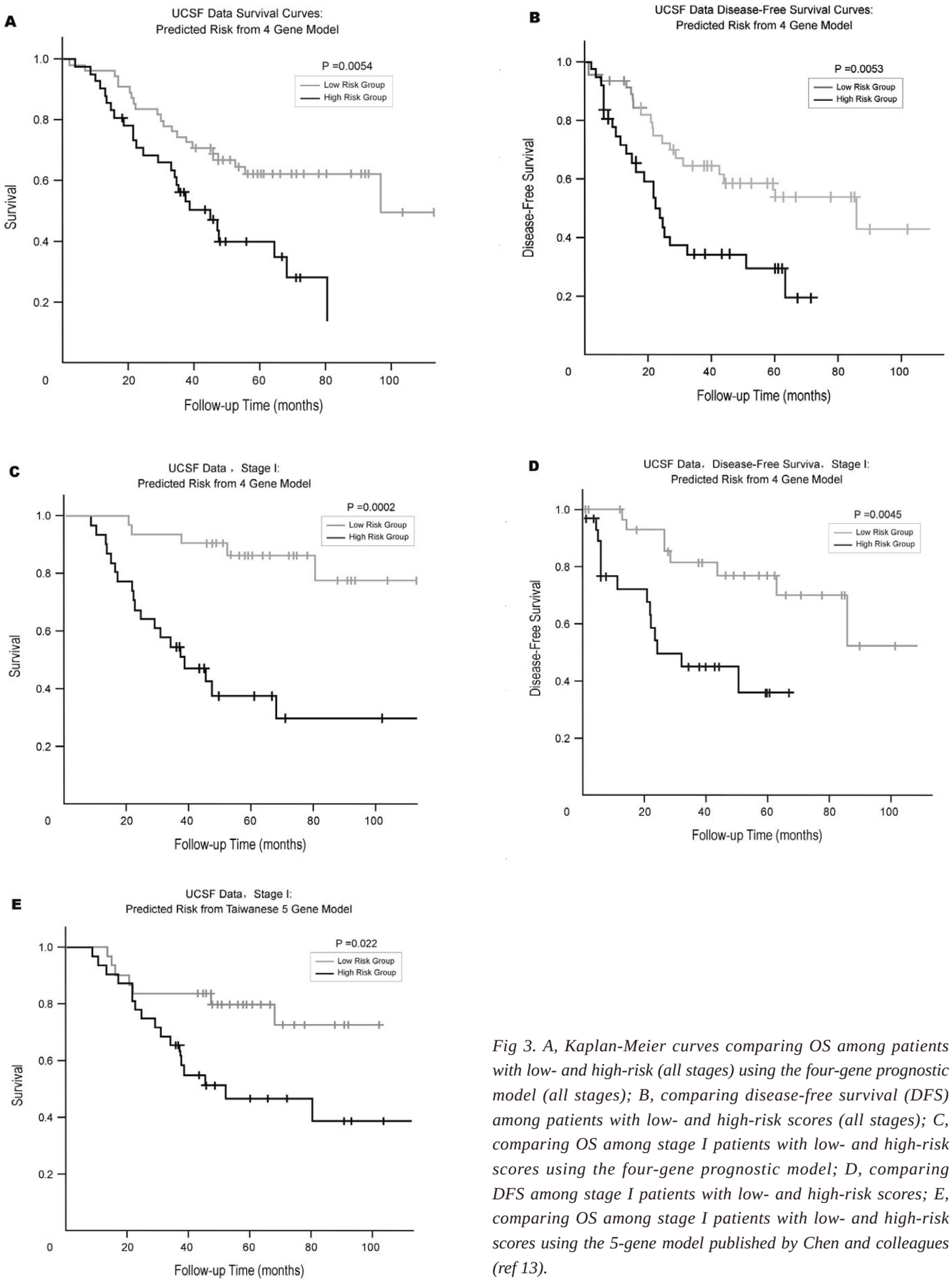

Fig 3. A, Kaplan-Meier curves comparing OS among patients with low- and high-risk (all stages) using the four-gene prognostic model (all stages); B, comparing disease-free survival (DFS) among patients with low- and high-risk scores (all stages); C, comparing OS among stage I patients with low- and high-risk scores using the four-gene prognostic model; D, comparing DFS among stage I patients with low- and high-risk scores; E, comparing OS among stage I patients with low- and high-risk scores using the 5-gene model published by Chen and colleagues (ref 13).

Whole-genome gene expression microarrays were used to analyze frozen tumor samples from 172 NSCLC patients (pT1-2, N0-1, M0) from five European institutions, who had undergone complete surgical resection. A 72-gene expression prognostic NSCLC classifier was developed. Based on the classifier score, patients were classified as either high or low risk of disease recurrence. Patients classified as low risk showed a significantly better recurrence-free survival both in the training set ($P < 0.001$; $n = 103$) and in the independent validation set ($P < 0.05$; $n = 69$). It was found that the 72-gene signature was closely associated with recurrence-free and overall survival in early-stage NSCLC patients and may become a tool for patient selection for adjuvant therapy (62).

Reed et al. proposed a qRT-PCR–based 2-gene signature for adenocarcinoma (48). Pooling microarray analysis of NSCLC cell lines in conjunction with correlation mapping of genes highly expressed in other tumors produced 14 candidate genes. These genes were tested by qRT-PCR on 20 adenocarcinoma samples yielding a 2-gene signature (CK19 and EpCAM2). This 2-gene signature revealed survival differences in high- and low-risk patients in their training cohort (HR, 6.2) and in a separate validation cohort (HR, 4.5) by Kaplan-Meier analysis. Raz et al. proposed a qRT-PCR–based 4-gene signature for adenocarcinoma (50). Seventy-six cancer-related candidate genes were selected from 217 genes demonstrated to have prognostic significance in previously published studies by content experts and literature review. Sixty-one of these genes for which reliable qRT-PCR data could be produced were assayed using qRT-PCR in a cohort of 120 adenocarcinoma samples. Cross-validation using Cox proportional hazards regression supported a 4-gene signature (WNT3A, RND3, LCK, and ERBB3). When applied to a cross-validated cohort of 70 patients with stage I adenocarcinoma, statistically significant differences in OS (87% vs. 38%) and disease-free survival (77% vs. 35%) were shown for high- and low-risk patients. This compared favorably with the 5-gene signature of Chen et al. (13) (Fig 3). When applied to the Raz et al cross-validated cohort, the signature developed by Chen et al demonstrated 5-year OS of 80% and 47%, respectively, for high- and low-risk patients. Notably, 2 of the genes (ERBB3 and LCK) overlapped between the Raz et al and Chen et al signatures.

Identified prognostic classifiers for early-stage NSCLC indicated large differences in sample numbers, microarray platform, and classifier design. Although a great variety of statistical models have been used, the performance of the different classifiers is similar with overall accuracies between 70% and 80% and a hazard ratio of 3 to 4. The overlap in profile genes, however, is limited to only 5 of a total of 327 genes **(Fig. 4)** even though it includes

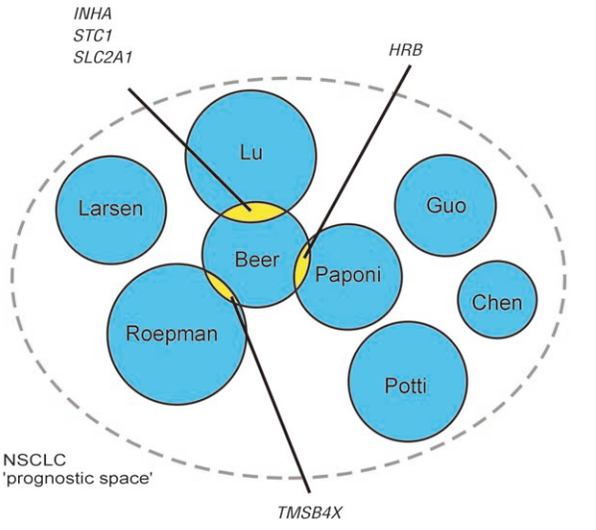

Fig 4. Gene overlap between NSCLC prognostic signatures. Overlap in genes of recent NSCLC survival signatures is limited to 5of a total of 327 genes used. Likely, all identified signatures are subsets from a larger NSCLC prognostic space (ref 62).

two studies (63,64) that reanalyzed existing data (51) but showed respectively no and three genes in overlap (62). Ein-Dor and coworkers (65) demonstrated that biological heterogeneity leads to thousands of samples being required to identify robust and reproducible subsets for most tumor types. These conclusions are supported by the finding that thousands of genes display intratumor heterogeneity, likely caused by the diversity of tumor microenvironments and cell populations (66,67). However, Boutros and coworkers hypothesized that differing statistical methodologies contribute to this lack of overlap (68). To test the hypothesis, they analyzed our previously published quantitative RT-PCR dataset with a semisupervised method. A 6-gene signature was identified and validated in 4 independent public microarray datasets that represent a range of tumor histologies and stages. The result demonstrated that at least 2 prognostic signatures can be derived from this single dataset. They further estimated the total number of prognostic signatures in this dataset with a 10-millionsignature permutation study. Their 6-gene signature was among the top 0.02% of signatures with maximum verifiability, reaffirming its efficacy. Importantly, the analysis identified 1,789 unique signatures, implying that their dataset contains >500,000 verifiable prognostic signatures for NSCLC. The result appears to rationalize the observed lack of overlap among reported NSCLC prognostic signatures (68).

Conclusions

GEP has demonstrated a tremendous potential to drive

personalized medicine in the near future when its use could be applied as a diagnostic tool, molecular staging classification, and more importantly, as prognostic and predictive biomarkers. Even patients with early-stage lung cancer demonstrate significant recurrence rates and lower-than-expected survival rates after surgical resection. The development of genomic prognostic models holds significant promise in our ability to differentiate among those patients who might benefit from additional therapy or lesser surgical procedures. However, we need to improve this technology in order to get results that are reproducible in most instances. Also, the technology must be available and be less cumbersome so it can be easily applied into common clinical practice.

References

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. CA Cancer J Clin 2007;57:43-66.
2. Ponn RB, Lo Cicero J III, Daly BDT. Surgical treatment of non small cell lung cancer. In: Shields TW, Lo Cicero J III, Ponn R, Rusch VW, Editors. General Thoracic Surgery, 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 1548-87.
3. Wright G, Manser RL, Byrnes G, Hart D, Campbell DA. Surgery for early-stage non-small cell lung cancer: systematic review and meta-analysis of randomised controlled trials. Thorax 2006;61:597-603.
4. Strauss GM. Adjuvant chemotherapy of lung cancer: methodologic issues and therapeutic advances. Hematol Oncol Clin North Am 2005;19:263-81.
5. Nesbitt JC, Putnam JB Jr, Walsh GL, Roth JA, Mountain CF. Survival in early stage non-small cell lung cancer. Ann Thorac Surg 1995;60:466-72.
6. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet 2000;355:479-85. [Erratum, Lancet 2000;355:1280.]
7. Hoang T, Xu R, Schiller JH, Bonomi P, Johnson DH. Clinical model to predict survival in chemonaïve patients with advanced non-small-cell lung cancer treated with third-generation chemotherapy regimens based on eastern cooperative oncology group data. J Clin Oncol 2005;23:175-83.
8. Forrest LM, McMillan DC, McArdle CS, Angerson WT, Dagg K, Scott HR. A prospective longitudinal study of performance status, an inflammation-based score (GPS) and survival in patients with inoperable non-small-cell lung cancer. Br J Cancer 2005;92:1834-6.
9. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. Chest 2002;122:1037-57.
10. Meyerson M, Carbone DP. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. J Clin Oncol 2005;23:3219-26.
11. Custodio AB, González-Larriba JL, Bobokova J, Calles A, Alvarez R, Cuadrado E, et al. Prognostic and predictive markers of benefit from adjuvant chemotherapy in early-stage non-small cell lung cancer. J Thorac Oncol 2009;4:891-910.
12. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845-56.
13. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 2007;356:11-20.
14. Endoh H, Tomida S, Yatabe Y, Konishi H, Osada H, Tajima K, et al. Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. J Clin Oncol 2004;22:811-9.
15. Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570-80.
16. Coussens LM, Werb Z. Inflammation and cancer. Nature 2002; 420:860-7.
17. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007; 128:683-92.
18. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature 2008;454:436-44.
19. Chin L, Gray JW. Translating insights from the cancer genome into clinical practice. Nature 2008; 452:553-63.
20. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467-70.
21. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci U S A 1997; 94:13057-62.
22. Ramaswamy S. Translating cancer genomics into clinical oncology. N Engl J Med 2004; 350:1814-6.
23. Paik S, Kim CY, Song YK, Kim WS. Technology insight: Application of molecular techniques to formalin-fixed paraffin-embedded tissues from breast cancer. Nat Clin Pract Oncol 2005;2:246-54.
24. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. J Clin Epidemiol 2007; 60:1205-19.
25. Kratz JR, Jablons DM. Genomic prognostic models in early-stage lung cancer. Clin Lung Cancer 2009;10:151-7.
26. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105-11.
27. Heguy A, O'Connor TP, Luettich K, Worgall S, Cieciuch A, Harvey B-G, et al. Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. J Mol Med 2006;84:318.
28. Borczuk AC, Gorenstein L, Walter KL, Assaad AA, Wang L, Powell CA. Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. Am J Pathol 2003;163:1949-60.
29. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy

failure in patients with multiple types of cancer. J Clin Invest 2005;115:1503-21.
30. Winton T, Livingston R, Johnson D, Rigas J, Johnston M, Butts C, et al. Vinorelbine plus Cisplatin vs. Observation in resected non-small-cell lung cancer. N Engl J Med 2005;352:2589-97.
31. Tsao MS, Zhu C, Ding K, Strumpf D, Pintilie M, Meyerson M, et al. A 15-gene expression signature prognostic for survival and predictive for adjuvant chemotherapy benefit in JBR. 10 patients. J Clin Oncol 2008;26(supp):abstr 7510.
32. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, et al. Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci U S A 2001;98:10787-92.
33. Grever MR, Schepartz SA, Chabner BA. The National Cancer Institute: cancer drug discovery and development program. Semin Oncol 1992; 19:622- 38.
34. Stinson SF, Alley MC, Kopp WC, Fiebig HH, Mullendore LA, Pittman AF, et al. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. Anticancer Res 1992;12:1035-53.
35. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12:1294-300. Epub 2006 Oct 22.
36. Beane J, Spira A, Lenburg ME. Clinical impact of high-throughput gene expression studies in lung cancer. J Thorac Oncol 2009;4:109-18.
37. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497-500.
38. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005; 353:123-32.
39. Rosell R, Taron M, Sanchez JJ, Paz-Ares L. Setting the benchmark for tailoring treatment with EGFR tyrosine kinase inhibitors. Future Oncol 2007; 3:277-83.
40. O'Byrne KJ, Danson S, Dunlop D, Botwood N, Taguchi F, Carbone D, et al. Combination therapy with gefitinib and rofecoxib in patients with platinum-pretreated relapsed non small-cell lung cancer. J Clin Oncol 2007; 25:3266-73.
41. Okano T, Kondo T, Fujii K, Nishimura T, Takano T, Ohe Y, et al. Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor tyrosine kinase inhibitor in lung adenocarcinoma. Clin Cancer Res 2007;13:799-805.
42. Altorki N, Lane ME, Bauer T, Lee PC, Guarino MJ, Pass H, et al. Phase II proof-of-concept study of pazopanib monotherapy in treatment-naive patients with stage I/II resectable non-small-cell lung cancer. J Clin Oncol 2010;28:3131-7. Epub 2010 Jun 1.
43. Nikolinakos PG, Altorki N, Yankelevitz D, Tran HT, Yan S, Rajagopalan D, et al. Plasma cytokine and angiogenic factor profiling identifies markers associated with tumor shrinkage in early-stage non-small cell lung cancer patients treated with pazopanib. Cancer Res 2010;70:2171-9. Epub 2010 Mar 9.
44. Tomida S, Koshikawa K, Yatabe Y, Harano T, Ogura N, Mitsudomi T, et al. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. Oncogene 2004;23:5360-70.
45. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu CQ, Strumpf D, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol 2007;25:5562-9.
46. Bianchi F, Nuciforo P, Vecchi M, Bernard L, Tizzoni L, Marchetti A, et al. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. J Clin Invest 2007;117:3436-44.
47. Seike M, Yanaihara N, Bowman ED, Zanetti KA, Budhu A, Kumamoto K, et al. Use of a cytokine gene expression signature in lung adenocarcinoma and the surrounding tissue as a prognostic classifier. J Natl Cancer Inst 2007;99:1257-69.
48. Reed CE, Graham A, Hoda RS, Khoor A, Garrett-Mayer E, Wallace MB, et al. A simple two-gene prognostic model for adenocarcinoma of the lung. J Thorac Cardiovasc Surg 2008;135:627-34.
49. Skrzypski M, Jassem E, Taron M, Sanchez JJ, Mendez P, Rzyman W, et al. Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung. Clin Cancer Res 2008;14:4794-9.
50. Raz DJ, Ray MR, Kim JY, He B, Taron M, Skrzypski M, et al. A multigene assay is prognostic of survival in patients with early-stage lung adenocarcinoma. Clin Cancer Res 2008;14:5565-70.
51. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-24. Epub 2002 Jul 15.
52. Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570-80.
53. Yu H, Jove R. The STATs of cancer-new molecular targets come on age. Nat Rev Cancer 2004;4:97-105.
54. Kumar A, Commane M, Flickinger TW, Horvath CM, Stark GR. Defective TNF-alpha-induced apoptosis in STAT1-null cells due to low constitutive levels of caspases. Science 1997;278:1630 -2.
55. Rehli M, Krause SW, Schwarzfischer L, Kreutz M, Andreesen R. Molecular cloning of a novel macrophage maturation-associated transcript encoding a protein with several potential transmembrane domains. Biochem Biophys Res Commun 1995;217:661- 7.
56. Chen JJ, Lin YC, Yao PL, Yuan A, Chen HY, Shun CT, et al. Tumor-associated macrophages: the double-edged sword in cancer progression. J Clin Oncol 2005;23:953-64.
57. Furukawa T, Sunamura M, Motoi F, Matsuno S, Horii A. Potential tumor suppressive pathway involving DUSP6/MKP-3 in pancreatic cancer. Am J Pathol 2003;162:1807-15.
58. Müller-Tidow C, Diederichs S, Bulk E, Pohle T, Steffen B, Schwäble J, et al. Identification of metastasis-associated receptor tyrosine kinases in non-small cell lung cancer. Cancer Res 2005;65:1778-82.
59. Zamoyska R, Basson A, Filby A, Legname G, Lovatt M, Seddon B. The influence of the src-family kinases, on T cell differentiation, survival and activation. Immunol Rev 2003;191:107-18.
60. Mahabeleshwar GH, Kundu GC. Tyrosine Kinase p56lck regulates

cell motility and nuclear factor kappa B-mediated secretion of urokinase type plasminogen activator through tyrosine phosphorylation of IkappaBalpha following hipoxia/reoxygenation. J Biol Chem 2003;278:52598-612.

61. Mahabeleshwar GH, Das R, Kundu GC. Tyrosine Kinase, p56lckinduced cell motility, and urokinase-type plasminogen activator secretion involve activation of epidermal growth factor receptor/extracellular regulated kinase pathways. J Biol Chem 2004;279:9733-42.
62. Roepman P, Jassem J, Smit EF, Muley T, Niklinski J, van de Velde T, et al. An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. Clin Cancer Res 2009;15:284-90.
63. Guo L, Ma Y, Ward R, Castranova V, Shi X, Qian Y. Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. Clin Cancer Res 2006;12:3344-54.
64. Lu Y, Lemon W, Liu PY, Yi Y, Morrison C, Yang P, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med 2006;3:e467.
65. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proc Natl Acad Sci USA 2006;103:5923-8.
66. Bachtiary B, Boutros PC, Pintilie M, Shi W, Bastianutto C, Li JH, et al. Gene expression profiling in cervical cancer: An exploration of intratumor heterogeneity. Clin Cancer Res 2006;12:5632-40.
67. Blackhall FH, Pintilie M, Wigle DA, Jurisica I, Liu N, Radulovich N, et al. Stability and heterogeneity of expression profiles in lung cancer specimens harvested following surgical resection. Neoplasia 2004;6:761-7.
68. Boutros PC, Lau SK, Pintilie M, Liu N, Shepherd FA, Der SD, et al. Prognostic gene signatures for non-small-cell lung cancer. Proc Natl Acad Sci U S A 2009;106:2824-8.