Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access

Author manuscript

Cancer Res. Author manuscript; available in PMC 2019 May 07.

Published in final edited form as:

Cancer Res. 2013 July 01; 73(13): 3821–3832. doi:10.1158/0008-5472.CAN-13-0031.

Combination of Protein Coding and Noncoding Gene Expression as a Robust Prognostic Classifier in Stage I Lung Adenocarcinoma

Ichiro Akagi[#1,2], **Hirokazu Okayama**[#1,6], **Aaron J. Schetter**[1], **Ana I. Robles**[1], **Takashi Kohno**[4], **Elise D. Bowman**[1], **Dickran Kazandjian**[1], **Judith A. Welsh**[1], **Naohide Oue**[7], **Motonobu Saito**[6], **Masao Miyashita**[2], **Eiji Uchida**[2], **Toshihiro Takizawa**[3], **Seiichi Takenoshita**[6], **Vidar Skaug**[8], **Steen Mollerup**[8], **Aage Haugen**[8], **Jun Yokota**[5], and **Curtis C. Harris**[1]

[1]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland [2]Division of Surgery for Organ Function and Biological Regulation, Tokyo [3]Division of Molecular Medicine and Anatomy, Graduate School of Medicine, Nippon Medical School, Tokyo [4]Division of Genome Biology, National Cancer Center Research Institute, Tokyo [5]Division of Multistep Carcinogenesis, National Cancer Center Research Institute, Tokyo [6]Department of Organ Regulatory Surgery, Fukushima Medical University School of Medicine, Fukushima [7]Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan [8]Section for Toxicology, Department of Chemical and Biological Working Environment, National Institute of Occupational Health, Oslo, Norway

[#] These authors contributed equally to this work.

Abstract

Prognostic tests for patients with early-stage lung cancer may provide needed guidance on postoperative surveillance and therapeutic decisions. We used a novel strategy to develop and validate a prognostic classifier for early-stage lung cancer. Specifically, we focused on 42 genes with roles in lung cancer or cancer prognosis. Expression of these biologically relevant genes and their association with relapse-free survival (RFS) were evaluated using microarray data from 148 patients with stage I lung adenocarcinoma. Seven genes associated with RFS were further

---

Corresponding Author: Curtis C. Harris, National Cancer Institute, 37 Convent Drive, MSC 4258, Building 37, Room 3068A, Bethesda, MD 20892. Phone: 301-496-2048; Fax: 301-496-0497; curtis_harris@nih.gov.

Authors' Contributions

Conception and design: I. Akagi, H. Okayama, A.J. Schetter, D. Kazandjian, M. Saito, J. Yokota, C.C. Harris

Development of methodology: I. Akagi, H. Okayama, A.J. Schetter, D. Kazandjian, C.C. Harris

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): I. Akagi, H. Okayama, T. Kohno, E.D. Bowman, J.A. Welsh, N. Oue, V. Skaug, S. Mollerup, A. Haugen, C.C. Harris

Analysis and interpretation of data (e.g.,statistical analysis, biostatistics, computational analysis): I. Akagi, H. Okayama, A.J. Schetter, A.I. Robles, T. Takizawa, V. Skaug, J. Yokota, C.C. Harris

Writing, review, and/or revision of the manuscript: I. Akagi, H. Okayama, A.J. Schetter, A.I. Robles, E.D. Bowman, M. Miyashita, T. Takizawa, V. Skaug, S. Mollerup, A. Haugen, C.C. Harris

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): I. Akagi, E.D. Bowman, D. Kazandjian, C.C. Harris

Study supervision: I. Akagi, M. Miyashita, E. Uchida, S. Takenoshita, J. Yokota, C.C. Harris

Disclosure of Potential Conflict of Interest

No potential conflicts of interest were disclosed.

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

examined by quantitative reverse transcription PCR in 291 lung adenocarcinoma tissues from Japan, the United States, and Norway. Only BRCA1, HIF1A, DLC1, and XPO1 were each significantly associated with prognosis in the Japan and US/Norway cohorts. A Cox regression-based classifier was developed using these four genes on the Japan cohort and validated in stage I lung adenocarcinoma from the US/Norway cohort and three publicly available lung adenocarcinoma expression profiling datasets. The results suggest that the classifier is robust across ethnically and geographically diverse populations regardless of the technology used to measure gene expression. We evaluated the combination of the four-gene classifier with miRNA miR-21 (MIR21) expression and found that the combination improved associations with prognosis, which were significant in stratified analyses on stage IA and stage IB patients. Thus, the four coding gene classifier, alone or with miR-21 expression, may provide a clinically useful tool to identify high-risk patients and guide recommendations regarding adjuvant therapy and postoperative surveillance of patients with stage I lung adenocarcinoma.

Introduction

Surgery with curative intent is the standard of care for stage I non–small cell lung cancer (NSCLC; National Comprehensive Cancer Network, NCCN, Guidelines; ref. 1). However, even after successful surgery and with histologically negative lymph nodes, 20% to 30% of patients with stage I NSCLC will recur (2). Although adjuvant chemotherapy can improve survival in patients with stage II or IIIA disease, its benefit in stage I patients is controversial (3). Therefore, it is critical to develop biomarkers that can identify patients with stage I NSCLC at high risk of recurrence who may benefit from adjuvant therapy.

Protein coding and noncoding gene expression have been used to develop prognostic classifiers for patients with various types of cancer (4–9) including stage I lung cancer (10–17). In many examples, the associations reported in single cohorts have failed to provide clinically useful information in additional patient populations (18). On the basis of recommendations outlined by Subramanian and Simon (18), we sought to develop a clinically useful, prognostic classifier in early-stage lung cancer to improve decisions about therapy and postoperative surveillance. We focused our analysis on 42 genes with a known mechanistic role in lung cancer and/or an association with cancer prognosis to maximize the potential of developing a biologically relevant classifier. We evaluated 291 primary tumors from 3 geographically and ethnically diverse populations by quantitative reverse transcription PCR (qRT-PCR) to identify genes with robust associations with prognosis. Our sample sizes were of sufficient power to achieve this task. A Cox regression-based classifier was then produced using linear gene expression values of the 4 protein-coding genes and we have made all data, methodologies, and scripts publicly available to allow readers to reproduce the results. We conducted stratified analyses of tumor–node–metastasis (TNM) stage IA and stage IB to identify high-risk patients who would benefit from adjuvant chemotherapy. We further tested the robustness of our prognostic classifier by evaluating 3 large, publicly available lung adenocarcinoma microarray datasets. All statistical models were evaluated with both univariate and multivariate models adjusting for clinically relevant risk factors such as age, smoking, and stage. We present results and coefficients of our final model in sufficient detail to allow readers to easily test the prognostic classifier in additional

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

patient populations. Finally, we combined this coding gene classifier with the expression of miR-21, a miRNA that we have shown to be associated with RFS and cancer-specific mortality in early-stage lung cancer (17), to determine whether this combination improved associations with prognosis in stage I lung adenocarcinoma.

Materials and Methods

Patients and tissue samples

We analyzed 291 tumor samples from 3 cohorts of patients with lung adenocarcinoma from National Cancer Center Hospital (Tokyo, Japan; Japan cohort, $n = 199$), the Metropolitan Baltimore area of the United States (U.S. cohort, $n = 67$), and the Haukeland University Hospital (Bergen, Norway; Norway cohort, $n = 25$). The Japan cohort was recruited from National Cancer Center Hospital between 1998 and 2008. The U.S. cohort was recruited between 1987 and 2009. The Norway cohort ($n = 25$) was recruited between 1988 and 2003. Further information about these cohorts has been described elsewhere (17).

Primary lung tumors and adjacent noncancerous tissues were procured from patients undergoing surgical resections without preoperative chemotherapy or radiation treatment. Tissues were snap-frozen immediately after surgery and stored at −80°C. Histopathology was classified according to the World Health Organization Classification of Tumor system. Only patients with the diagnosis of pure adenocarcinoma or adenocarcinoma with a bronchioloalveolar carcinoma (BAC) component were used, whereas those of adenocarcinoma in situ (formerly pure BAC) were excluded.

Patient demographics are listed in Table 1. Cases were originally staged on the basis of American Joint Committee on Cancer (AJCC) sixth edition and were restaged to AJCC seventh edition where possible. The U.S. and Norway cohorts showed similar 5-year survival rates, TNM staging, gender, and age at diagnosis. Thus, to increase the statistical power for all further analyses, they were combined. All patients consented to tissue specimen collection. This study was carried out under the approval of the Institutional Review Board at the NIH, Regional Committees for Medical and Health Research Ethics in Norway, and the Internal Review Board for National Cancer Center at Japan.

RNA isolation and mRNA qRT-PCR

RNA was extracted from frozen tissue samples using TRIzol (Invitrogen), and was assessed via the Bioanalyzer 2100 system (Agilent Technologies). Data collection was completed while blinded to clinical outcomes. TaqMan Gene expression assays (Applied Biosystems) were loaded into 96.96 dynamic arrays (Fluidigm Corporation) in duplicate and quantitative real-time PCR (qRT-PCR) reactions were carried out using BioMark Real-Time PCR System according to manufacturer's instructions (Fluidigm Corporation). TaqMan assays included *DNMT1* (Assay ID Hs00154749_m1), *BRCA1* (ID Hs00173233_m1), *HIF1A* (ID Hs00936371_m1), *CA9* (ID Hs00154208_m1), *CCT3* (ID Hs00195623_m1), *DLC1* (ID Hs00183436_m1), and *XPO1* (ID Hs00418963_m1). 18S (ID Hs03003631_m1) was used as a normalization control. Undetectable signals were treated as missing data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gene expression arrays

Publicly available gene expression datasets.—Microarray data generated using the Japanese cohort (19) are available at Gene Expression Omnibus (accession number GSE31210). Additional publicly available microarray data, including the Bhattacharjee cohort (20) and National Cancer Institute Director's Challenge cohort (21), were used for validation and obtained through ONCOMINE 2.0 (Compendia Bioscience). The Tomida cohort (15) was obtained from Gene Expression Omnibus (accession number GSE13213). Selection criteria for all publicly available datasets required each dataset to include survival information for more than 50 TNM stage I patients and have expression data for *BRCA1*, *HIF1A*, *DLC1*, and *XPO1*. The normalized expression values were obtained from each dataset and were not processed further. To build the gene signature, we averaged the expression values for 2 probes corresponding to *BRCA1* in the Oncomine 2.0 cohorts. There were 3 probes (A_23_P252721, A_24_P940115, and A_23_P112016) for *DLC1* in the Tomida cohort. A_23_P112016 was excluded because of missing values and the other 2 were averaged.

Statistical analysis and gene classifier development

Patients were dichotomized on the basis of the median expression value for each gene to evaluate the association between gene expression and survival by the Kaplan–Meier log-rank test using Graphpad Prism v5.0 (Graphpad Software Inc). Cox regression was carried out using Stata 11.2 (Staga-Corp LP). Coefficients from multivariate Cox regression models on continuous expression values for *BRCA1*, *HIF1A*, *DLC1*, and *XPO1* from the Japan cohort were used to build the 4 coding gene classifier scores for all cohorts. The association between the 4-coding gene classifier and survival was assessed for significance by P_{trend} and by the log-rank test where appropriate. For Cox regression analysis, age was treated as a continuous variable and smoking status was dichotomized into more than 20 pack-years and less than 20 pack-years. Gene expression data, clinical information, and stata coding to generate the 4-coding gene classifier are publicly available for download (22).

miRNA measurements

Global miRNA expression patterns were measured with NanoString Human microRNA assays using 100 ng of total RNA, according to manufacturer's instructions (NanoString Technologies. miR-21 expression values were normalized on the basis of the average expression of the 5 most highly expressed miRs that do not include miR-21(miR-720, miR-26a, miR-126, miR-16, and miR-29a). Using the expression of the 5 highest miRNAs was thought to be more precise than using lower expressed miRNAs as normalization controls.

Absolute quantification of miR-21 copies per cell by qRT-PCR

To calculate the copies of miR-21 per tumor cell, we first estimated the total RNA content per cell using a series of total RNA extraction from 2 lung adenocarcinoma cell lines, A549 and NCI-H23. Briefly, trypsinized cells were counted and a series of cell suspension (100K, 330K, 1.0M, 3.3M, 10M cells in triplicate) was pelleted, washed, and then subjected to total

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA extraction by TRIzol. The total RNA quantity was determined by NanoDrop and this data was used to generate a standard curve to estimate amount of RNA per cell.

Copy numbers of miR-21 were calculated on the basis of comparing levels of miR-21 in lung tumors with a standard curve of serial diluted, synthetic miR-21 (Integrated DNA Technologies, Inc). Synthetic *C. elegans* miR-54 was added to all samples as a quality control of both reverse transcription and PCR. For 49 tumors from the 3 independent cohorts, 40 ng of total RNA was used for reverse transcription. qRT-PCR was conducted in triplicate (miR-21) or duplicate (cel-miR-54). qRT-PCR was conducted using standard TaqMan PCR protocol as described previously.(17) Absolute copy number of miR-21 was determined by generating a standard curve of synthetic miR-21.

Results

XPO1, BRCA1, HIF1A, CA9, DLC1, and *CCT3* expression are associated with relapse-free survival of stage I–II lung adenocarcinoma in the Japan cohort

Our strategy for developing the coding gene classifier is found in Supplementary Fig. S1. Forty-two genes were selected on the basis of literature support for a role in lung cancer (Supplementary Table S2). We analyzed microarray data on TNM stage I (AJCC 6th edition) patients with lung cancer from the Japan cohort (n = 148) and examined associations of those genes with RFS. Seven genes (*DNMT1, XPO1, BRCA1, HIF1A, CA9, DLC1*, and *CCT3*) were significantly associated with RFS (P < 0.01) and selected for further analysis (Supplementary Table S2). qRT-PCR measurements significantly correlated with the microarray data (P < 0.001) for 6 of the 7 genes (Supplementary Fig. S2). *DNMT1* expression by qRT-PCR did not correlate with microarray data and was omitted from further analysis.

qRT-PCR expression for each gene was dichotomized as based on median expression for the Japan cohort (n = 199). *BRCA1* [HR, 2.05; 95% confidence interval (CI), 1.17–3.58; P = 0.012], *HIF1A* (HR, 1.79; 95% CI, 1.03–3.11; P = 0.038), *CA9* (HR, 3.25; 95% CI, 1.79–5.90; P = 0.001), *CCT3* (HR, 2.14; 95% CI, 1.22– 3.74; P = 0.008), *DLC1* (HR, 0.44; 95% CI, 0.25–0.77; P = 0.004), and *XPO1* (HR, 2.02; 95% CI, 1.15–3.53; P = 0.014) were each significantly associated with RFS (Supplementary Table S3) further validating our microarray results.

BRCA1, HIF1A, DLC1, and *XPO1* are associated with cancer-specific mortality in the combined US/Norway cohort

All 6 genes were measured by qRT-PCR in the combined US/ Norway cohort (stage I–II, n = 92). The expression of *BRCA1* (HR, 3.21, 95% CI, 1.70–6.07; P < 0.001), *HIF1A* (HR, 2.01, 95% CI, 1.07–3.57; P = 0.029), *DLC1* (HR, 0.45; 95%CI, 0.25–0.85; P = 0.013), and *XPO1* (HR, 2.06; 95%CI, 1.12–3.76; P = 0.019) were each significantly associated with cancer-specific mortality in the combined US/Norway cohort by Cox regression (Supplementary Table S3).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A 4-coding gene classifier is associated with prognosis in 5 independent cohorts

We showed that *BRCA1*, *HIF1A*, *DLC1*, and *XPO1* are associated with prognosis in multiple cohorts from different regions of the world providing strong evidence that these can be useful prognostic biomarkers. In an attempt to make a robust prognostic classifier for lung cancer, we developed a Cox regression model using the expression of these 4-coding genes. Guidelines for prognostic factor studies in NSCLC recommends including the results in stage II patients with low risk of recurrence as well stage I patients (18). Therefore, we built a gene classifier on all of the stage I and II patients in the Japan cohort ($n = 199$) using multivariate Cox regression on linear expression values of each of the 4 genes. The resulting model was classifier score = (0.104 × *BRCA1*) + (0.133 × *HIF1A*) + (−0.246 ×*DLC1*) + (0.378 × *XPO1*). This model was applied to the Japan and US/Norway cohorts using qRT-PCR expression data and to 3 publicly available datasets (Director's cohort, $n = 378$; Bhattacharjee cohort, $n = 100$; Tomida cohort, $n = 92$) using microarray expression data. Characteristics of these cohorts are found in Supplementary Table S1.

The resulting classifier score was categorized as low, medium, or high based on tertiles. The 4-coding gene classifier was significantly associated with prognosis in stage I–II patients in all 5 cohorts: Japan ($P < 0.001$), US/Norway ($P = 0.001$), Director's ($P = 0.002$), Bhattacharjee ($P = 0.019$), and Tomida ($P = 0.014$) cohorts (Supplementary Fig. S3). These results provide strong evidence that the 4-coding gene classifier is robust and will lead to reproducible predictions in ethnically and geographically diverse populations.

A 4-gene classifier is associated with prognosis in stage I lung cancer in 5 independent cohorts

The goal of our study was to develop a prognostic gene classifier for early-stage lung adenocarcinoma. Therefore, we focused on stage I patients. The 4-coding gene classifier was significantly associated with prognosis in stage I lung adenocarcinoma for all 5 cohorts including the Japan ($P < 0.001$, $n = 149$), US/Norway ($P < 0.001$, $n = 67$), Director's ($P < 0.001$, $n = 276$), Bhattacharjee ($P = 0.036$, $n = 76$), and Tomida ($P = 0.008$, $n = 79$) cohorts (Fig. 1). In univariate Cox regression models, high-risk group was associated with prognosis in the Japan (HR, 3.84; 95% CI, 1.53–9.64; $P = 0.004$), US/Norway (HR, 8.03; 95% CI, 2.54–25.28; $P < 0.0005$) cohorts, Director's (HR, 2.68; 95% CI, 1.50–4.79; $P = 0.001$), Bhattacharjee (HR, 2.61; 95% CI, 1.04–6.56; $P = 0.042$), and Tomida (HR, 4.73; 95% CI, 1.32–16.96; $P = 0.017$) cohorts. Multivariate Cox regression showed that these associations were independent of other clinical characteristics (Table 2). These data suggest that the 4-coding gene classifier has potential to be used with other clinical characteristics to help identify stage I patients at high risk of cancer relapse.

Subgroup analysis was carried out on stage IB patients (Fig. 1). The 4-gene classifier was significantly associated with prognosis stage IB patients in the Japan ($P = 0.029$, $n = 49$), US/Norway ($P = 0.013$, $n = 38$), Director's ($P = 0.003$, $n = 162$), and Bhattacharjee ($P = 0.020$, $n = 40$) cohorts further showing the potential of this protein-coding gene classifier as a prognostic biomarker for lung cancer.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The patients in this study were staged on the basis of AJCC sixth edition at the time of diagnosis. The 4-gene classifier was developed and validated on the basis of AJCC sixth edition staging information. In 2009, the AJCC seventh edition TNM staging was developed and published. To determine how our classifier performs with AJCC seventh edition staging, we restaged patients to AJCC seventh edition for cases with available data (Table 1) and found that the 4-gene classifier was significantly associated in AJCC seventh edition TNM stage I lung cancer patients in both the Japan ($P < 0.001$, Fig. 2) and the US/Norway cohorts ($P = 0.003$, Fig. 3).

The 4-coding gene classifier and noncoding miR-21 are independently associated with prognosis in stage I lung adenocarcinoma

We previously reported that high miR-21 expression in tumors was associated with poor prognosis in stage I, lung adenocarcinoma (17). That study used the same Japan and US/Norway cohorts as the current study and provides an opportunity to determine whether the combination of miR-21 and 4-coding gene classifier improves prognostic use. In our previous study, we used qRT-PCR to measure miR-21 in lung tumors. We sought to determine whether another method of measuring miR-21 provided the same results in a way that may be easier to translate to the clinic. For this, we used nCounter Human miRNA assays, which provide a method for digital detection of hundreds of miRNAs with minimal sample preparation and no amplification. miR-720, miR-26a, miR-16, miR-126, and miR-29 were the highly expressed miRNAs (excluding miR-21) and none of these miRNAs were associated with prognosis (Supplementary Fig. S4). Therefore, we decided to normalize miR-21 expression to the geometric mean of these 5 miRNAs. We observed similar results when comparing the nCounter Human miRNA assay measurement of miR-21 with our previous publication using qRT-PCR. Using the nCounter assays, higher than median expression of miR-21 was significantly associated with worse prognosis in stage I patients in both the Japan and US/Norway cohorts. Interestingly, associations of miR-21 with prognosis were stronger when using nCounter assays to measure miR-21 compared with our previously reported qRT-PCR measurements of miR-21. These data were analyzed on the basis of both AJCC seventh edition staging (Figs. 2 and 3) and AJCC sixth edition staging (Supplementary Figs. S5 and S6).We evaluated whether the combination of miR-21 and the 4-gene classifier was superior to either classifier alone. Kaplan–Meier analysis (Figs. 2 and 3) shows that patients with a low 4-gene classifier score and low miR-21 (categorized as low risk) had the best prognosis. In general, patients categorized as high risk by only one of these markers had an intermediate prognosis and patients with high 4-gene classifier/high miR-21 (categorized as high risk) had the worst prognosis, regardless of TNM stage groups. Multivariate analysis showed that both the 4-gene classifier and miR-21 were statistically independent of one another in the Japan and the US/Norway cohort (Table 3). These results suggest that the 4 coding gene classifier and miR-21 expression can be used together as a prognostic biomarker for stage I lung adenocarcinoma.

Although it is clear that increased miR-21 expression is associated with poor survival, it is unclear what this expression level is in terms of copies per cell. We next estimated that lung tumor cells has approximately 50,000 copies of miR-21 per cell on average. This was calculated using a standard curve of serially diluted, synthetic miR-21 (Supplementary Fig.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

S7) and known amounts of tumor RNA. The total RNA per cell for the lung cancer cell lines A549 and NCI-H23 was estimated to be 19.4 pg/cell and 20.1 pg/cell, respectively (Supplementary Fig. S8). Therefore, we used 20 pg of RNA per tumor cell calculate copies of miR-21 per tumor cell. Our copy number estimates are similar to other published estimates for lung tissue (23).

Discussion

Our objective was to build a prognostic gene classifier for early-stage lung adenocarcinoma to help guide clinical decisions. We identified and validated a prognostic gene classifier in 5 independent patient cohorts. The associations of the 4 coding gene classifier with prognosis were significant in stage I patients across ethnically and geographically diverse populations, suggesting that this classifier has potential to identify high risk, patients with early-stage lung cancer who may benefit from adjuvant chemotherapy.

The current standard of care for stage I NSCLC is lobectomy and mediastinal lymph node dissection, without adjuvant chemotherapy. There is a need for biomarkers to identify stage IA patients who might benefit from adjuvant therapy and stage IB patients who could be spared from adjuvant chemotherapy. We propose that this 4-coding gene classifier comprising *HIF1A*, *DLC1*, *BRCA1*, and *XPO1* can be used to guide therapeutic decisions for stage I patients. Stage I patients defined as high risk may be suitable for earlier or more aggressive intervention. Some studies suggest that TNM stage IB patients should be given postoperative adjuvant chemotherapy (24, 25) while others do not agree (26–28). NCCN guidelines indicate that recurrent NSCLC or metastases should be evaluated for the presence of EGF receptor (EGFR) mutations or EML4–ALK fusions to help determine appropriate therapies. Future studies should address whether the 4-coding gene classifier presented here can be used alone or with EGFR and ALK status to help provide guidance on which therapies should be given to high risk, early-stage patients.

HIF1A, *DLC1*, *XPO1*, and *BRCA1* have all been implicated in cancer biology and may be causally associated with aggressive disease. Therefore, the altered expression of any of these genes may alter tumor biology to create more aggressive tumors that are either more likely to metastasize or will rapidly develop resistance chemotherapies. *HIF1A* overexpression is a common event in multiple types of carcinomas and has been associated with aggressive tumor behavior and overall poor prognosis (29–32). *HIF1A* was part of a lung cancer prognostic classifier reported by Lau and colleagues (33) *XPO1* can modulate both nuclear processing and nuclear-cytosolic transport of miRNAs (34, 35), BRCA1 (36), and TP53 (37, 38), and *XPO1* was also part of a lung cancer prognostic classifier reported by Wan and colleagues(16). *DLC1* is a tumor-suppressor gene frequently deleted or silenced in many tumor types, including lung (39, 40). In particular, *DLC1* methylation was significantly associated with the presence of lung metastatic disease (41). Germline mutation of *BRCA1* is most notably associated with familial susceptibility to breast and ovarian cancers (42). However, *BRCA1* overexpression leads to resistance to chemotherapeutic drugs, owing to its role in DNA repair and antiapoptotic cellular pathways (43). However, a recent study showed that high expression of *BRCA1* mRNA was an indicator of poor prognosis in patients with lung cancer that did not receive adjuvant chemotherapy (44). The Japan cohort

Author Manuscript

in our study is composed primarily of patients who did not receive adjuvant chemotherapy. Thus, the prosurvival role of *BRCA1* may extend beyond enhanced chemotherapeutic resistance to encompass resistance to endogenous oxidative damage (45). In addition to the lung cancer cohorts presented in this study, increased *BRCA1* expression is associated with worse prognosis of other types of human cancer (Supplementary Fig. S9). BRCA1 has multiple functions including DNA repair and DNA recombination (46). BRCA1 may enhance DNA repair of the endogenous DNA double-strand breaks that are found at higher levels in tumors (47). Therefore, elevated BRCA1 may increase cancer cell survival and contribute to the poor prognosis of lung cancer cases and further studies are warranted. Several clinical studies are currently recruiting patients with stage II–IV NSCLC with the purpose of studying *BRCA1* mRNA levels in association with chemotherapy (NCT00478699, NCT00617656, and NCT00705549 at the ClinicalTrials.gov registry). On the basis of our findings, we would argue that there is sufficient evidence for a clinical study of *BRCA1* mRNA levels in stage I NSCLC.

In our study, the combination of the coding gene classifier and miR-21 proved superior at predicting prognosis than either alone. Overexpression of miR-21 has been described across solid tumors, including lung cancer (17, 48). To our knowledge, this is the first report estimating copy numbers per cell for miR-21 in lung tumors. We also find that measurement of miR-21 by NanoString Human microRNA assays may be a more robust prognostic classifier than measuring miR-21 by qRT-PCR. A possible reason for this is that the NanoString assays used 5 highly expressed miRNAs as normalization controls and this may be more stable than using RNU66 as a normalization control, as we did in our previous study (17).

miR-21 has an oncogenic role in lung cancer. OncomiR addiction to miR-21 has been shown in an animal model (49). In a mouse model of NSCLC, miR-21 overexpression enhanced tumorigenesis and its deletion reduced it, providing a direct link between miR-21 and lung carcinogenesis (50). miR-21 targets many genes (51) involved in the cancer cell phenotypes associated with the Hallmarks of Cancer (52). In addition, miR-21 decreases SOD3 (53) and increases resistance to the induction apoptosis in lung cancer cells (54). These and other studies identify miR-21 as a potential therapeutic target for lung cancer (55).

A strength of our study is that we imposed a requirement for each of the genes included in the final model were associated with survival in 2 independent and ethnically diverse sample cohorts. As with any analysis on historical cohorts, a limitation of this study is that the associations are observed in patients that were diagnosed with cancer several years ago and as clinical treatments of lung cancer change, there is no guarantee that these associations would still be relevant. This further justifies a need for prospective examination of this gene signature. Our analysis was conducted using fresh frozen tissues. It will be interesting if these associations can be found when using formalin-fixed paraffin-embedded tissue, which is more readily available for clinical research of both prospective and retrospective studies. In conclusion, our study provides supporting evidence for the use of coding and noncoding gene expression analysis within a clinical setting to help guide therapeutic decisions in lung adenocarcinoma, particularly, stage I.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Izumi Horikawa for his advice.

Grant Support

This research was supported by the Intramural Research Program of the National Cancer Institute, NIH, Department of Defense Congressionally Directed Medical Research Program Grant PR093793, the Norwegian Cancer Society, a Grant-in-Aid from the Ministry of Health, Labor and Welfare for the Third-term Comprehensive 10-year Strategy for Cancer Control, Japan, and a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan for Scientific Research on Innovative Areas (22131006).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. http://www.nccn.org.
2. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet 2000; 355:479–85. [PubMed: 10841143]
3. Tsuboi M, Ohira T, Saji H, Miyajima K, Kajiwara N, Uchida O, et al. The present status of postoperative adjuvant chemotherapy for completely resected non-small cell lung cancer. Ann Thorac Cardiovasc Surg 2007;13:73–7. [PubMed: 17505412]
4. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49–54. [PubMed: 12469122]
5. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845–56. [PubMed: 16239904]
6. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004;350: 1828–37. [PubMed: 15115829]
7. 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. [PubMed: 12118244]
8. Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, et al. Erlotinib in lung cancer - molecular and clinical predictors of outcome. N Engl J Med 2005;353:133–44. [PubMed: 16014883]
9. 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. [PubMed: 14990636]
10. 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. [PubMed: 17194181]
11. 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. [PubMed: 17948124]
12. Lee ES, Son DS, Kim SH, Lee J, Jo J, Han J, et al. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. Clin Cancer Res 2008;14:7397–404. [PubMed: 19010856]
13. Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JM, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 2006;66:7466–72. [PubMed: 16885343]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

14. 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. [PubMed: 17202451]
15. Tomida S, Takeuchi T, Shimada Y, Arima C, Matsuo K, Mitsudomi T, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. J Clin Oncol 2009;27: 2793–9. [PubMed: 19414676]
16. Wan YW, Sabbagh E, Raese R, Qian Y, Luo D, Denvir J, et al. Hybrid models identified a 12-gene signature for lung cancer prognosis and chemoresponse prediction. PLoS ONE 2010;5:e12222. [PubMed: 20808922]
17. Saito M, Schetter AJ, Mollerup S, Kohno T, Skaug V, Bowman ED, et al. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: a retrospective analysis of three cohorts. Clin Cancer Res 2011;17:1875–82. [PubMed: 21350005]
18. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? J Natl Cancer Inst 2010; 102:464–74. [PubMed: 20233996]
19. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes up-regulated in ALK-positive and EGFR/KRAS/ ALK-negative lung adenocarcinomas. Cancer Res 2012;72:100–11. [PubMed: 22080568]
20. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 2001;98:13790–5. [PubMed: 11707567]
21. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 2008;14: 822–7. [PubMed: 18641660]
22. http://www3.cancer.gov/intra/lhc/Supplemental_Dat3a_and_coding_CR.zip.
23. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD. Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. RNA 2008;14:35–42. [PubMed: 18025253]
24. Kato H, Ichinose Y, Ohta M, Hata E, Tsubota N, Tada H, et al. A randomized trial of adjuvant chemotherapy with uracil-tegafur for adenocarcinoma of the lung. N Engl J Med 2004;350:1713–21. [PubMed: 15102997]
25. Roselli M, Mariotti S, Ferroni P, Laudisi A, Mineo D, Pompeo E, et al. Postsurgical chemotherapy in stage IB non small cell lung cancer: long-term survival in a randomized study. Int J Cancer 2006;119: 955–60. [PubMed: 16550600]
26. 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. [PubMed: 15972865]
27. Douillard JY, Rosell R, De Lena M, Carpagnano F, Ramlau R, Gonzales-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. Lancet Oncol 2006;7: 719–27. [PubMed: 16945766]
28. Wakelee H, Dubey S, Gandara D. Optimal adjuvant therapy for non-small cell lung cancer–how to handle stage I disease. Oncologist 2007;12:331–7. [PubMed: 17405898]
29. Giatromanolaki A, Koukourakis MI, Sivridis E, Turley H, Talks K, Pezzella F, et al. Relation of hypoxia inducible factor 1 alpha and 2 alpha in operable non-small cell lung cancer to angiogenic/molecular profile of tumours and survival. Br J Cancer 2001;85:881–90. [PubMed: 11556841]
30. Birner P, Schindl M, Obermair A, Plank C, Breitenecker G, Oberhuber G. Overexpression of hypoxia-inducible factor 1alpha is a marker for an unfavorable prognosis in early-stage invasive cervical cancer. Cancer Res 2000;60:4693–6. [PubMed: 10987269]
31. Zhong H, De Marzo AM, Laughner E, Lim M, Hilton DA, Zagzag D, et al. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. Cancer Res 1999;59: 5830–5. [PubMed: 10582706]
32. Aebersold DM, Burri P, Beer KT, Laissue J, Djonov V, Greiner RH, et al. Expression of hypoxia-inducible factor-1alpha: a novel predictive and prognostic parameter in the radiotherapy of oropharyngeal cancer. Cancer Res 2001;61:2911–6. [PubMed: 11306467]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

33. 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. [PubMed: 18065728]
34. Bussing I, Yang JS, Lai EC, Grosshans H. The nuclear export receptor XPO-1 supports primary miRNA processing in *C. elegans* and *Drosophila*. EMBO J 2010;29:1830–9. [PubMed: 20436454]
35. Castanotto D, Lingeman R, Riggs AD, Rossi JJ. CRM1 mediates nuclear-cytoplasmic shuttling of mature microRNAs. Proc Natl Acad Sci U S A 2009;106:21655–9. [PubMed: 19955415]
36. Brodie KM, Henderson BR. Characterization of BRCA1 protein targeting, dynamics, and function at the centrosome: a role for the nuclear export signal, CRM1, and Aurora A kinase. J Biol Chem 2012;287: 7701–16. [PubMed: 22262852]
37. Cai X, Liu X. Inhibition of Thr-55 phosphorylation restores p53 nuclear localization and sensitizes cancer cells to DNA damage. Proc Natl Acad Sci U S A 2008;105:16958–63. [PubMed: 18952844]
38. Freedman DA, Levine AJ. Nuclear export is required for degradation of endogenous p53 by MDM2 and human papillomavirus E6. Mol Cell Biol 1998;18:7288–93. [PubMed: 9819415]
39. Yuan BZ, Jefferson AM, Baldwin KT, Thorgeirsson SS, Popescu NC, Reynolds SH. DLC-1 operates as a tumor suppressor gene in human non-small cell lung carcinomas. Oncogene 2004;23:1405–11. [PubMed: 14661059]
40. Durkin ME, Yuan BZ, Zhou X, Zimonjic DB, Lowy DR, Thorgeirsson SS, et al. DLC-1: a Rho GTPase-activating protein and tumour suppressor. J Cell Mol Med 2007;11:1185–207. [PubMed: 17979893]
41. Castro M, Grau L, Puerta P, Gimenez L, Venditti J, Quadrelli S, et al. Multiplexed methylation profiles of tumor suppressor genes and clinical outcome in lung cancer. J Transl Med 2010;8:86. [PubMed: 20849603]
42. Black DM, Solomon E.The search for the familial breast/ovarian cancer gene. Trends Genet 1993;9:22–6. [PubMed: 8434413]
43. Kennedy RD, Quinn JE, Mullan PB, Johnston PG, Harkin DP. The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst 2004;96:1659–68. [PubMed: 15547178]
44. Rosell R, Skrzypski M, Jassem E, Taron M, Bartolucci R, Sanchez JJ, et al. BRCA1: a novel prognostic factor in resected non-small-cell lung cancer. PLoS ONE 2007;2:e1129. [PubMed: 17987116]
45. Saha T, Rih JK, Roy R, Ballal R, Rosen EM. Transcriptional regulation of the base excision repair pathway by BRCA1. J Biol Chem 2010;285: 19092–105. [PubMed: 20185827]
46. Silver DP, Livingston DM. Mechanisms of BRCA1 tumor suppression. Cancer Discov 2012;2:679–84. [PubMed: 22843421]
47. Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. Science 2008;319: 1352–5. [PubMed: 18323444]
48. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257–61. [PubMed: 16461460]
49. Medina PP, Nolde M, Slack FJ. OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. Nature 2010;467:86–90. [PubMed: 20693987]
50. Hatley ME, Patrick DM, Garcia MR, Richardson JA, Bassel-Duby R, van Rooij E, et al. Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21. Cancer Cell 2010;18:282–93. [PubMed: 20832755]
51. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J 2012;18:244–52. [PubMed: 22647361]
52. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74. [PubMed: 21376230]
53. Zhang X, Ng WL, Wang P, Tian L, Werner E, Wang H, et al. MicroRNA-21 modulates the levels of reactive oxygen species by targeting SOD3 and TNFalpha. Cancer Res 2012;72:4707–13. [PubMed: 22836756]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

54. Seike M, Goto A, Okano T, Bowman ED, Schetter AJ, Horikawa I, et al. MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. Proc Natl Acad Sci U S A 2009;106:12085–90. [PubMed: 19597153]

55. Croce CM. miRNAs in the spotlight: understanding cancer gene dependency. Nat Med 2011;17:935–6. [PubMed: 21818092]

Author Manuscript

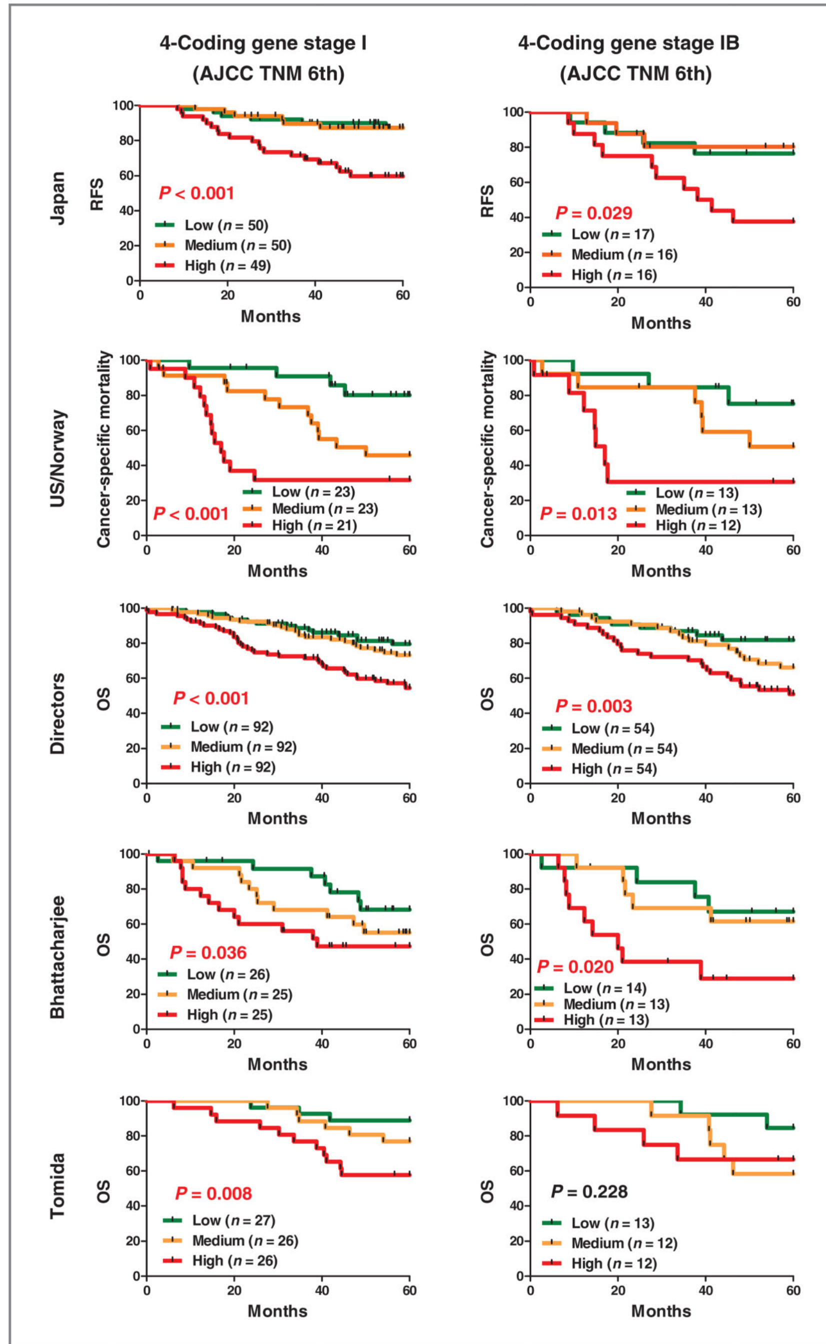

Figure 1.

Kaplan-Meier analysis of the 4-coding gene classifier in stage I lung adenocarcinoma (using AJCC sixth edition staging) from 5 independent cohorts, including the Japan ($n = 149$, 5-year RFS), US/Norway ($n = 67$, 5-year cancer-specific mortality), Director's [$n = 276$, 5-year overall survival (OS)], Bhattachajee ($n = 76$,5-yearOS), and Tomida ($n = 79$, 5-year OS) cohorts. Subgroup analysis on stage IB tumors were shown in right panels. *P* values were calculated by log-ranktest for trend.

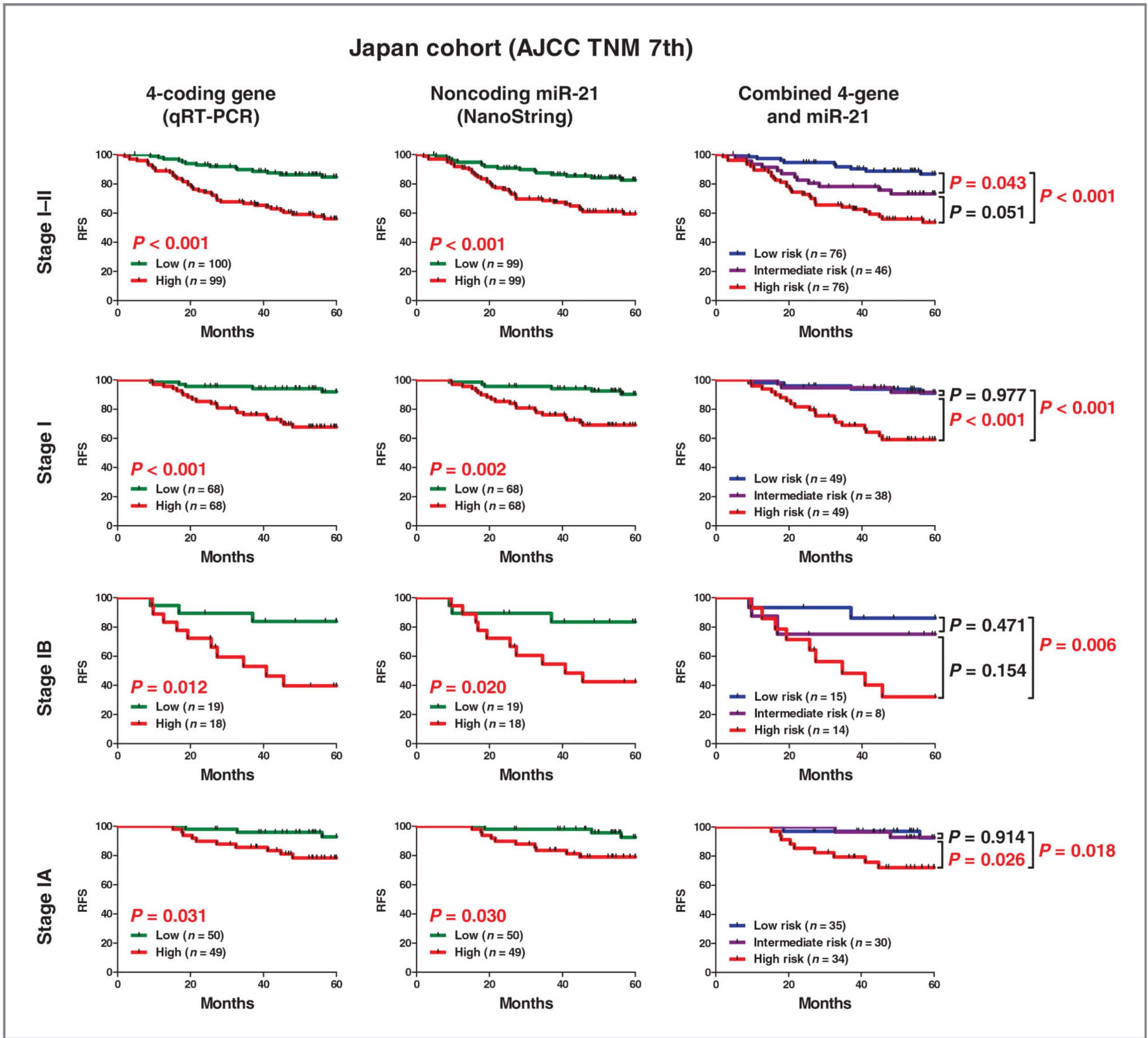

Figure 2.

The combined 4-coding gene classifier with noncoding miR-21 classifier is associated with RFS and this association is significantly better than either classifier alone in the Japan cohort. Kaplan–Meier analysis for the 4-gene classifier (left), noncoding miR-21 (middle), and the combination (right) in stage I–II, stage I, stage IB, and IA, respectively (using AJCC seventh edition staging).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

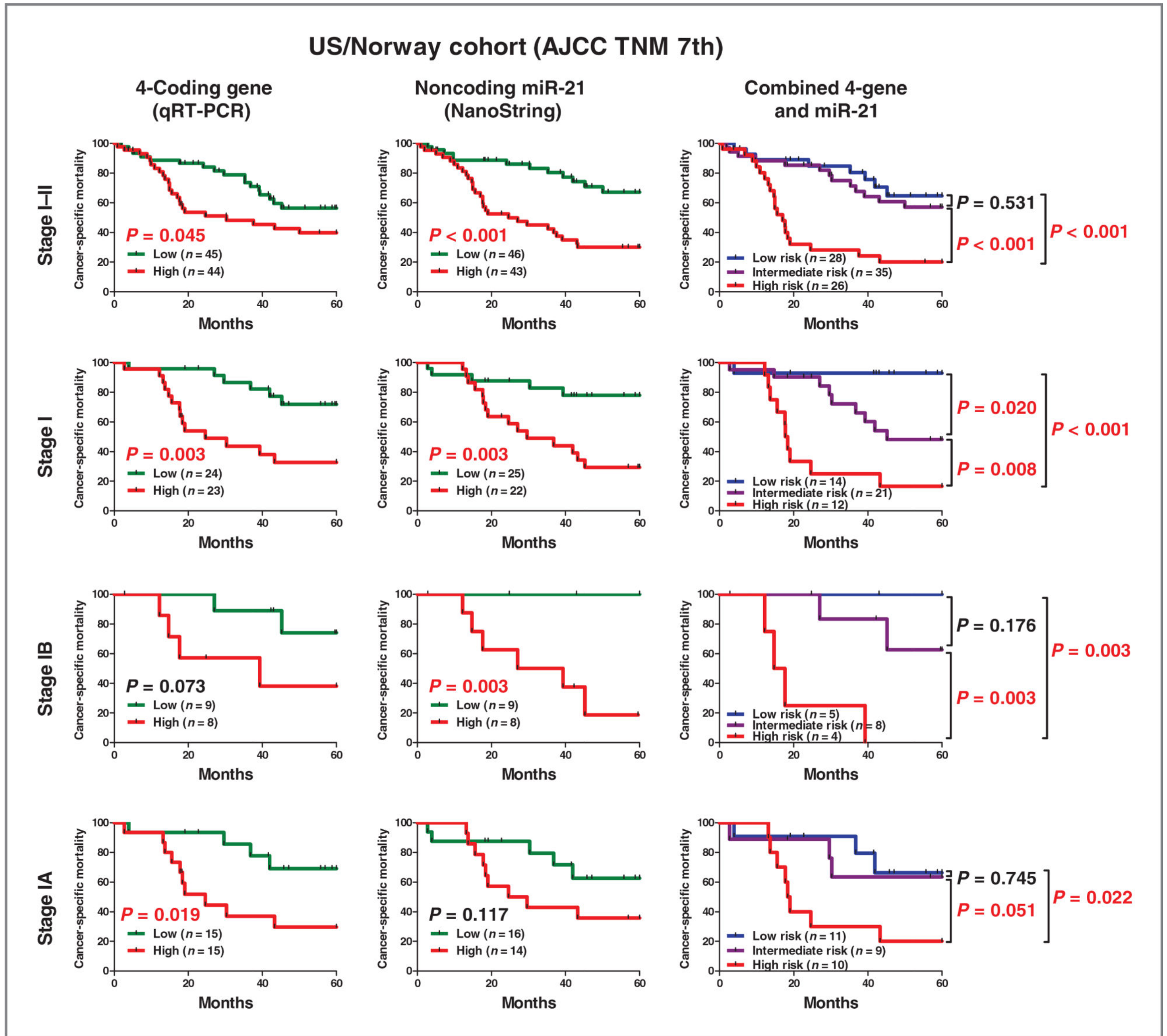

Figure 3.

The combined 4-coding gene classifier with noncoding miR-21 classifier is associated with cancer-specific mortality and this association is significantly better than either classifier alone in the US/Norway cohort. Kaplan–Meier analysis for the 4-gene classifier (left), noncoding miR-21 (middle), and the combination (right) in stage I–II, stage I, stage IB, and IA, respectively (using AJCC seventh edition staging).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Characteristics of study populations of patients in the Japan, Norway, and U.S. cohorts

	Japan cohort (n = 199)	Norway cohort (n = 25)	U.S. cohort (n = 67)
Age, y			
Mean (SD)	59.4 (7.7)	64.0 (11.8)	64.9 (10.0)
Range	30–76	37–82	40–90
Gender (%)			
Male	97 (48.7)	15 (60.0)	37 (55.2)
Female	102 (51.3)	10 (40.0)	30 (44.8)
Race (%)			
Caucasian	0 (0.0)	25 (100.0)	43 (64.2)
African-American	0 (0.0)	0 (0.0)	24 (35.8)
Asian	199 (100.0)	0 (0.0)	0 (0.0)
Histology (%)			
Adenocarcinoma	199 (100.0)	25 (100.0)	67 (100.0)
Tumor size (cm)			
Mean (SD)	3.0 (1.6)	3.8 (1.7)	3.7 (2.1)
Range	0.9–14.0	2.0–6.5	0.9–10.5
Unknown	0	13	2
AJCC TNM 6th stage (%)			
IA	100 (50.3)	6 (24.0)	23 (34.3)
IB	49 (24.6)	14 (56.0)	24 (35.8)
II	50 (25.1)	5 (20.0)	20 (29.9)
AJCC TNM 7th stage (%)[a]			
IA	99 (49.7)	6 (24.0)	24 (36.4)
IB	37 (18.6)	5 (20.0)	12 (18.2)
II	63 (31.7)	5 (20.0)	30 (45.5)
IB or II	0 (0.0)	7 (28.0)	0 (0.0)
Unknown	0 (0.0)	2 (8.0)	0 (0.0)
Smoking history (%)			
Never	98 (49.2)	1 (4.0)	4 (6.0)

	Japan cohort (n = 199)	**Norway cohort (n = 25)**	**U.S. cohort (n = 67)**
<20 pack years	32 (16.1)	11 (44.0)	8 (11.9)
20 pack years	69 (34.7)	12 (48.0)	54 (80.6)
Unknown	0 (0.0)	1 (4.0)	1 (1.5)
Adjuvant therapy (%)			
Adjuvant chemotherapy	8 (4.0)		0 (0.0)
None	191 (96.0)		59 (88.1)
Unknown	0 (0.0)	25 (100.0)	8 (11.9)[b]

[a]Cases were restaged to AJCC 7th edition based on tumor size and/or pathology reports where possible.

[b]No information on the timing of therapies (possibly either after surgery or after recurrence).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Univariate and Multivariate Cox regression of the 4-gene classifier in 5 independent cohorts (AJCC TNM sixth edition, stage I patients)

Variable		Univariate analysis		Multivariate analysis[c]	
		HR (95% CI)	*P*	HR (95% CI)	*P*
Japan cohort (*n* = 149)					
4-Gene classifier (qRT-PCR)[a]	Low	Reference	NA	Reference	NA
	Medium	1.04 (0.34–3.23)	0.940	1.30 (0.41–4.11)	0.657
	High	**3.84 (1.53–9.64)**	**0.004**	**3.78 (1.51–9.51)**	**0.005**
		$P_{trend} = 0.002$		**$P_{trend} = 0.003$**	
AJCC 6th Stage	IB/IA	**2.89 (1.43–5.87)**	**0.003**	**2.57 (1.22–5.41)**	**0.013**
Age, y	Continuous	1.00 (0.96–1.05)	0.895	1.00 (0.95–1.05)	0.991
Gender	Male/female	0.95 (0.47–1.94)	0.893	0.80 (0.93–1.65)	0.551
Pack years	20/<20	1.51 (0.72–3.16)	0.271		
US/Norway cohort (*n* = 67)[b]					
4-Gene classifier (qRT-PCR)[a]	Low	Reference	NA		NA
	Medium	**3.48 (1.12–10.79)**	**0.031**	**3.48 (1.11–10.90)**	**0.032**
	High	**8.03 (2.54–25.28)**	**<0.0005**	**8.40 (2.65–26.67)**	**<0.0005**
		$P_{trend} < 0.0005$		**$P_{trend} < 0.0005$**	
AJCC 6th Stage	IB/IA	0.99 (0.46–2.12)	0.971	0.78 (0.36–1.68)	0.527
Age, y	Continuous	1.01 (0.97–1.05)	0.576	1.02 (0.97–1.06)	0.485
Gender	Male/female	0.88 (0.43–1.82)	0.723	0.91 (0.44–1.91)	0.807
Pack years	20/<20	0.97 (0.38–2.46)	0.943		
Director's cohort (*n* = 276)					
4-Gene classifier (microarray)[a]	Low	Reference	NA		NA
	Medium	1.35 (0.71–2.55)	0.362	1.37 (0.72–2.60)	0.332
	High	**2.68 (1.50–4.79)**	**0.001**	**2.68 (1.49–4.80)**	**0.001**
		$P_{trend} < 0.0005$		**$P_{trend} < 0.0005$**	
AJCC 6th stage	IB/IA	1.43 (0.90–2.27)	0.134	1.42 (0.89–2.28)	0.144
Age, y	Continuous	**1.03 (1.01–1.05)**	**0.008**	**1.03 (1.01–1.06)**	**0.006**
Gender	Male/female	1.20 (0.77–1.87)	0.410	0.98 (0.63–1.55)	0.960

Variable		Univariate analysis		Multivariate analysis[c]	
		HR (95% CI)	*P*	HR (95% CI)	*P*
Pack years	20/<20				
Bhattacharjee cohort (*n* = 76)					
4-Gene classifier (microarray)[a]	Low	Reference	NA		NA
	Medium	1.67 (0.65–4.31)	0.290	1.35 (0.51–3.58)	0.541
	High	**2.61 (1.04–6.56)**	**0.042**	**2.69 (1.05–6.94)**	**0.040**
		P_{trend} **= 0.039**		P_{trend} **= 0.036**	
AJCC 6th Stage	IB/IA	1.74 (0.84–3.61')	0.138	2.43 (1.12–5.24)	0.023
Age, y	Continuous	**1.04(1.00–1.08)**	**0.037**	**1.06 (1.01–1.10)**	**0.009**
Gender	Male/female	1.29 (0.64–2.62)	0.475	0.82 (0.38–1.78)	0.624
Pack years	20/<20	1.77 (0.68–4.62)	0.241		
Tomida cohort (*n* = 79)					
4-Gene classifier (microarray)[a]	Low	Reference	NA		NA
	Medium	2.14 (0.53–8.55)	0.283	1.79 (0.44–7.32)	0.418
	High	**4.73 (1.32–16.96)**	**0.017**	**3.92 (1.07–14.36)**	**0.024**
		P_{trend} **= 0.011**		P_{trend} **= 0.024**	
AJCC 6th Stage	IB/IA	1.45 (0.60–3.50)	0.409	1.27 (0.47–3.46)	0.639
Age, y	Continuous	1.02 (0.97–1.06)	0.449	1.01 (0.96–1.06)	0.768
Gender	Male/female	**2.64 (1.02–6.89)**	**0.046**	2.43 (0.93–6.39)	0.071
Pack years	20/<20	1.39 (0.58–3.36)	0.462		

NOTE: Bold, significant values < 0.05.

[a]The 4-coding gene classifier was categorized on the basis of tertiles.

[b]All univariate and multivariate models were adjusted for cohort membership for the US/Norway analyses.

[c]Multivariate models included all variables that were significant in univariate models in at least one cohort.

Abbreviation: NA, not applicable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3.

Cox regression analysis of the 4-gene classifier and miR-21 expression in the Japan and the US/Norway cohorts (AJCC TNM 7th edition)

		Univariate analysis		Multivariate analysis	
Variable		**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Japan cohort (stage I—II, *n* = 199)					
4-Gene classifier (qRT-PCR)[a]	High/low	**3.56 (1.94–6.55)**	**0.000**	**2.39 (1.16–4.90)**	**0.018**
miR-21 (NanoString)[a]	High/low	**2.75 (1.53–4.94)**	**0.001**	1.34 (0.67–2.69)	0.410
AJCC 7th Stage	II/I	**3.19 (1.87–5.45)**	**0.000**	**2.07 (1.15–3.71)**	**0.015**
Age, y	Continuous	1.03 (0.99–1.07)	0.132	1.03 (0.99–1.07)	0.140
Gender	Male/female	1.27 (0.74–2.16)	0.382	1.21 (0.70–2.08)	0.500
Pack years	20/<20	1.62 (0.94–2.79)	0.084		
US/Norway cohort (stage I—II, n = 89))[b]					
4-Gene classifier (qRT-PCR)[a]	High/low	**1.95 (1.04–3.66)**	**0.037**	1.88 (0.96–3.65)	0.064
miR-21 (NanoString)[a]	High/low	**3.38 (1.72–6.65)**	**0.000**	**3.42 (1.66–7.04)**	**0.001**
AJCC 7th stage[c]	II/I	1.60 (0.84–3.03)	0.150	1.45 (0.76–2.75)	0.262
Age (y)	Continuous	1.01 (0.98–1.04)	0.556	1.01 (0.98–1.04)	0.676
Gender	Male/female	1.02 (0.55–1.90)	0.938	0.79 (0.41–1.52)	0.487
Pack years	20/<20	0.86 (0.42–1.78)	0.685		
Japan cohort (stage I, *n* = 136)					
4-Gene classifier (qRT-PCR)[a]	High/low	**4.76 (1.79–12.64)**	**0.002**	**4.14(1.39–12.32)**	**0.011**
miR-21 (NanoString)[a]	High/low	**3.89 (1.56–9.69)**	**0.004**	1.73 (0.60–4.98)	0.312
AJCC 7th Stage	IB/IA	**3.25 (1.50–7.01)**	**0.003**	**3.36 (1.46–7.70)**	**0.004**
Age, y	Continuous	1.00 (0.95–1.06)	0.919	1.00 (0.95–1.05)	0.967
Gender	Male/female	0.98 (0.45–2.14)	0.967	0.83 (0.38–1.84)	0.654
Pack years	20/<20	1.54 (0.69–3.47)	0.294		
US/Norway cohort (stage I, *n* = 47)[b]					
4-Gene classifier (qRT-PCR)[a]	High/low	**4.02 (1.53–10.59)**	**0.005**	**4.68 (1.67–13.15)**	**0.003**
miR-21 (NanoString)[a]	High/low	**4.11 (1.49–11.37)**	**0.006**	**6.55 (1.97–21.78)**	**0.002**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Variable		Univariate analysis		Multivariate analysis	
		HR (95% CI)	***P***	**HR (95% CI)**	***P***
AJCC 7th Stage	IB/IA	0.69 (0.26–1.86)	0.467	1.17 (0.37–3.72)	0.792
Age, y	Continuous	0.99 (0.94–1.04)	0.590	0.98 (0.93–1.03)	0.450
Gender	Male/female	0.91 (0.38–2.19)	0.831	0.40 (0.13–1.24)	0.112
Pack years	20/<20	1.29 (0.41–4.04)	0.666		

NOTE: Bold, significant values < 0.05.

[a] The 4-coding gene classifier and noncoding miR-21 were each categorized on the basis of median.

[b] All univariate and multivariate models were adjusted for cohort membership for the US/Norway analyses.

[c] Upon restaging to AJCC 7th edition, there were 7 cases in the Norway cohort for which it could not be distinguished whether they were TNM stage IB or IIA. These are included in univariate analyses and excluded in multivariate analyses.