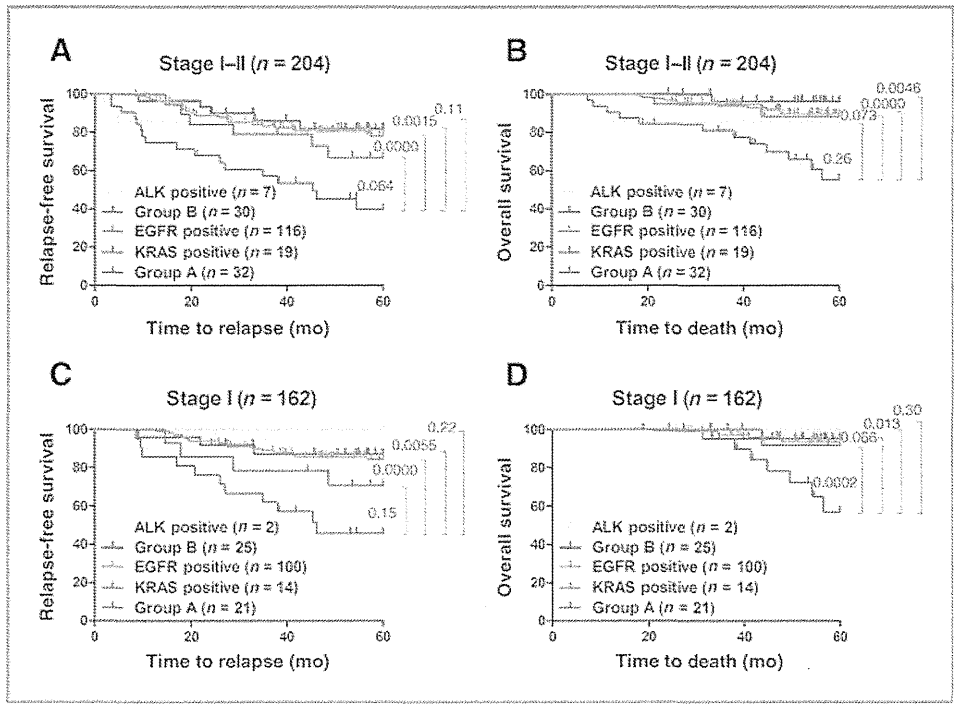

Figure 2. Kaplan–Meier survival curves for RFS and OS of 204 lung adenocarcinoma cases according to *EGFR*-positive, *KRAS*-positive, *ALK*-positive, group A, and group B. RFS and OS of stage I–II (A, B) and stage I (C, D) cases are shown.

of *KRAS*-positive and *ALK*-positive cases. Accordingly, multivariate analyses of 162 stage I patients further showed significant differences in RFS and OS between group A cases and *EGFR*-positive cases, and also between group A cases and group B cases. Because numbers of *KRAS*-positive cases and *ALK*-positive cases were small, we next compared RFS and OS between group A patients and patients in all 4 other groups combined ("Others" in Table 3). Differences in RFS as well as those in OS were highly significant and independent of staging. These results strongly indicated that group A patients comprise a distinct subclass of *EGFR/KRAS/ALK*-negative lung adenocarcinomas, and the prognoses of group A patients were the worst among the 5 groups of patients.

Clustering of lung adenocarcinomas with poor prognosis by gene expression profiling

We next carried out unsupervised hierarchical clustering of all the 226 adenocarcinoma cases, including 127 *EGFR*-positive cases and 20 *KRAS*-positive cases, to investigate whether expression profiling with a set of 174 genes with 190 probes could extract group A cases as a unique subset among all adenocarcinomas, and whether the profiling could be useful for prognosis prediction of patients with any genotypes of adenocarcinomas in general. As shown in Supplementary Fig. S4, clustering patterns of all the 226 patients were very similar to those of the 79 patients consisting of 11 *ALK*-positive cases and 68 triple-negative cases. In particular, the 11 *ALK*-positive cases comprised a small cluster in the right side of Cluster 1 (Cluster 1b), supporting that *ALK*-positive adenocarcinomas show unique expression profiles among all adenocarcinomas. Group A and group B cases also have a tendency to accumulate in Clusters 1a and Cluster 2, respectively. However, group A cases often comprise clusters with the *KRAS*-positive cases, whereas group B cases were distributed with the *EGFR*-positive cases. Therefore, group A and group B triple-negative adenocarcinomas were not exclusive with the *EGFR*-positive and *KRAS*-positive adenocarcinomas by expression profiling of these 174 genes. Therefore, expression profiling with a set of the 174 genes was concluded to be useful to distinguish *ALK*-positive adenocarcinomas among all lung adenocarcinomas.

However, RFS of 119 patients in Cluster 1 was significantly worse than RFS of 85 patients in Cluster 2 (HR = 3.73, P = 0.00016). When Cluster 1 was further divided into 2 subclasses 1a and 1b of the right and left sides, respectively, Cluster 1a containing most of group A patients showed the worst prognosis among the 3 subclasses (Supplementary Fig. S4). Therefore, the expression signature of these 174 genes was indicated to be useful for prognostic prediction of adenocarcinoma patients, in particular of triple-negative adenocarcinoma patients.

Minimum set of genes characterizing triple-negative lung adenocarcinomas with poor prognosis

The above results implied that triple-negative adenocarcinomas can be classified into 2 distinct subgroups by expression profiling and prognoses of these 2 groups are significantly different from each other. Accordingly, expression of several genes among the 174 genes was expected to be independently associated with prognosis of triple-negative adenocarcinoma patients. Therefore, we next selected genes whose expression was associated with prognosis from the 174 genes evaluated by the 190 probes. To evaluate the prognostic value of each probe and to make a comparative study for association of gene expression with prognosis in other cohorts possible, we took a minimum P value approach for grouping the patients for survival analysis because of the following reason. A database

Table 3. Hazard ratios for relapse-free and overall survivals in lung adenocarcinomas

			Univariate		Multivariate	
Survival	**Case (*n*)**	**Variable**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Relapse free	Stage I–II (204)	Age	1.03 (0.99–1.07)	0.12	1.04 (0.99–1.08)	0.092
		Sex (male/female)	1.39 (0.82–2.38)	0.22	1.00 (0.49–2.05)	0.99
		Smoking habit (ever/never)	1.43 (0.84–2.44)	0.19	1.10 (0.54–2.24)	0.80
		pStage (II/I)	1.86 (1.41–2.45)	1.3E-05	3.50 (1.93–6.34)	3.6E-05
		Subgroup				
		Group A/ALK (+)	4.78 (0.63–35.99)	0.13	6.01 (0.76–47.82)	0.09
		Group A/KRAS (+)	2.43 (0.96–6.17)	0.062	2.85 (1.10–7.35)	0.031
		Group A/EGFR (+)	3.58 (1.93–6.64)	5.3E-05	2.76 (1.44–5.29)	0.0022
		Group A/Group B	4.58 (1.69–12.42)	0.0028	4.10 (1.50–11.24)	0.0061
		Group A/Others	3.56 (2.00–6.34)	1.6E-05	3.04 (1.68–5.53)	2.5E-04
	Stage I (162)	Age	1.01 (0.96–1.06)	0.69	1.00 (0.95–1.05)	0.97
		Sex (male/female)	0.99 (0.50–1.96)	0.98	0.83 (0.33–2.07)	0.69
		Smoking habit (ever/never)	1.06 (0.54–2.08)	0.87	0.97 (0.39–2.45)	0.95
		Subgroup				
		Group A/ALK (+)	—	—	—	—
		Group A/KRAS (+)	2.31 (0.73–7.28)	0.15	2.36 (0.73–7.62)	0.15
		Group A/EGFR (+)	4.33 (2.00–9.35)	2.0E-04	4.51 (2.05–9.91)	1.7E-04
		Group A/Group B	5.36 (1.49–19.24)	0.010	5.52 (1.50–20.37)	0.010
		Group A/Others	4.18 (2.03–8.60)	1.0E-04	4.32 (2.06–9.09)	1.1E-04
Overall	Stage I–II (204)	Age	1.03 (0.98–1.08)	0.33	1.03 (0.98–1.09)	0.21
		Sex (male/female)	1.69 (0.82–3.48)	0.16	0.89 (0.33–2.41)	0.82
		Smoking habit (ever/never)	1.91 (0.92–3.97)	0.084	1.46 (0.54–3.92)	0.45
		pStage (II/I)	2.07 (1.45–2.97)	7.2E-05	3.93 (1.83–8.44)	4.6E-04
		Subgroup				
		Group A/ALK (+)	2.95 (0.38–22.78)	0.30	3.50 (0.41–29.85)	0.25
		Group A/KRAS (+)	3.12 (0.88–11.09)	0.079	3.31 (0.91–12.03)	0.069
		Group A/EGFR (+)	4.59 (2.06–10.23)	2.0E-04	3.35 (1.44–7.81)	0.005
		Group A/Group B	6.83 (1.53–30.54)	0.012	5.68 (1.24–25.95)	0.025
		Group A/Others	4.50 (2.17–9.36)	5.7E-05	3.61 (1.68–7.78)	0.0010
	Stage I (162)	Age	0.99 (0.93–1.06)	0.73	0.98 (0.91–1.04)	0.45
		Sex (male/female)	1.15 (0.43–3.08)	0.79	0.77 (0.20–3.00)	0.70
		Smoking habit (ever/never)	1.47 (0.55–3.91)	0.45	1.26 (0.32–4.89)	0.74
		Subgroup				
		Group A/ALK (+)	—	—	—	—
		Group A/KRAS (+)	5.79 (0.71–47.3)	0.10	5.61 (0.67–46.84)	0.11
		Group A/EGFR (+)	5.83 (2.04–16.71)	0.0010	6.06 (2.08–17.71)	9.8E-04
		Group A/Group B	9.13 (1.12–74.34)	0.039	9.32 (1.10–78.61)	0.040
		Group A/Others	6.30 (2.34–16.99)	2.8E-04	6.47 (2.33–17.98)	3.4E-04

named PrognoScan was recently developed by coauthors of this study (26). In the PrognoScan database, minimum *P* values for the association of gene expression with prognosis of all probes in a platform are available for a number of cohorts that have been published. Therefore, it was possible to validate the present findings using data from various other cohorts by the same criteria. According to the method described previously (26), corrected minimum *P* values were calculated for each probe to control the error rate for the evaluation of the association with RFS and OS. Expression of 11 genes evaluated with 12 probes (2 probes for the *DEPDC1* gene) showed significant associations with both RFS and OS in 62 triple-negative adenocarcinomas and also in 46 stage I triple-negative adenocarcinomas (Table 4). Among the 11 genes, expression of 10 genes was positively correlated with poor prognosis, whereas that of the remaining 1 gene, *KIF19*, expression was negatively correlated with poor prognosis.

We first selected 174 genes as being preferentially expressed in either *ALK*-positive adenocarcinomas or triple-negative adenocarcinomas by the criteria of "probes whose expression levels in any adenocarcinomas with *EGFR* or *KRAS* mutations were lower than the mean expression

Table 4. List of genes whose expression is associated with relapse free survival and overall survival of patients with lung adenocarcinoma

		NCC														CAN/DF		HLM		MSK		UM		Nagoya		Duke		Seoul	
		TN, Stage I–II				TN, Stage I				All Stage I–II				All Stage I		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III	
Dataset		RFS $n = 62$		OS $n = 62$		RFS $n = 46$		OS $n = 46$		RFS $n = 204$		OS $n = 204$		RFS $n = 162$		OS $n = 162$		OS $n = 82$		OS $n = 79$		OS $n = 104$		OS $n = 178$		OS $n = 117$		RFS $n = 111$		RFS $n = 138$	
Gene symbol	Probe ID (for NCC)	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR
DEPDC1	222958_s_at	0.00	2.3	0.00	3.0	0.00	3.0	0.02	2.7	0.00	2.1	0.00	1.8	0.00	2.0	0.00	2.1	—	—	—	—	0.03	1.1	—	—	0.00	1.6	0.04	1.0	0.01	0.9
DEPDC1	235545_at	0.00	1.8	0.01	2.3	0.00	2.4	0.04	2.6	0.00	1.4	0.00	1.6	0.00	1.3	0.00	2.2														
FOSL2	218881_s_at	0.01	1.7	0.03	1.7	0.02	1.8	0.00	3.3	0.00	1.2	0.00	1.7	0.00	1.4	0.00	2.4	—	—	0.00	1.7	—	—	0.02	0.7	—	—	—	—	0.01	1.0
MCM4	222037_at	0.00	1.8	0.00	3.0	0.01	2.0	0.04	2.6	0.00	1.4	0.00	1.8	0.00	1.5	0.00	2.1	—	—	—	—	—	—	—	—	0.00	1.7	—	—	—	—
UBE2S	202779_s_at	0.00	3.0	0.02	16.0	0.01	2.8	0.02	16.6	0.00	1.6	0.02	1.4	0.00	1.6	0.00	2.1	—	—	—	—	0.05	1.0	—	—	—	—	—	—	—	—
CD300A	217078_s_at	0.01	1.7	0.00	2.1	0.04	1.7	0.00	2.8	0.00	1.1	0.00	1.5	0.01	1.2	0.01	1.7	—	—	—	—	0.00	1.5	—	—	—	—	—	—	—	—
SLITRK4	232636_at	0.02	1.7	0.03	1.7	0.00	2.9	0.00	2.5	0.01	1.1	0.00	1.4	0.04	1.1	0.00	2.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—
KRT16	209800_at	0.00	2.0	0.00	2.5	0.00	2.4	0.00	2.7	0.00	1.2	0.00	1.4	0.01	1.2	0.01	1.7	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SIGLEC9	210569_s_at	0.00	1.9	0.01	2.0	0.00	2.1	0.04	2.1	0.00	1.6	0.00	1.4	0.00	1.7	0.00	2.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
DIAPH3	232596_at	0.02	1.5	0.00	3.0	0.05	1.6	0.03	2.6	0.00	1.2	0.00	2.1	0.00	1.5	0.00	2.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—
LOC152225	1562048_at	0.01	1.5	0.00	2.3	0.02	1.7	0.00	2.7	0.00	1.3	0.00	1.9	—	—	0.00	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
KIF19	1553314_a_at	0.01	-1.5	0.05	-1.6	0.00	-3.0	0.00	−2.5	—	—	0.03	−1.4	—	—	—	—	—	—	—	—	—	—	—	—	0.00	−1.4	—	—	—	—

Abbreviations: NCC, National Cancer Center; TN, Triple-negative.

HRs (log2 ratio) with corrected *P* value < 0.05 are shown.

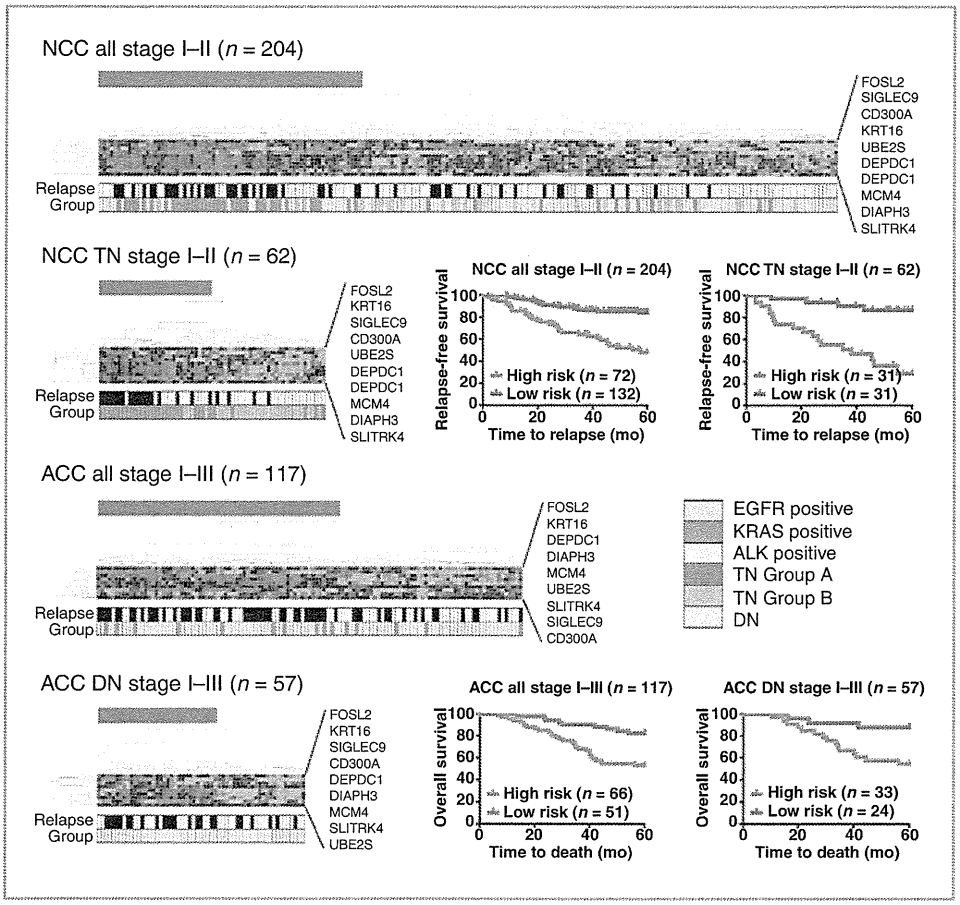

Figure 3. Unsupervised hierarchical clustering based on the expression of a set of 9 genes. All 204 stage I–II adenocarcinomas and 62 triple-negative (TN) stage I–II adenocarcinomas of the National Cancer Center (NCC) data set subjected to survival analysis were analyzed, and a cluster with higher expression of these genes than the other cluster was recognized as a high-risk group (red bar). Results of 117 adenocarcinomas, including 57 double-negative (DN) adenocarcinomas, of the Aichi Cancer Center (ACC) data set are shown below.

level of a total of 54,675 probes." Then, 11 of the 174 genes were further selected as being associated with prognosis of patients with triple-negative adenocarcinomas. Therefore, higher expression of several genes among the 11 genes was predicted to be associated with poorer prognosis, even when all adenocarcinoma cases, including *EGFR*-positive, *KRAS*-positive, and *ALK*-positive adenocarcinomas were analyzed together. Furthermore, triple-negative adenocarcinomas with poor prognosis would be separated into a high-risk group classified with this procedure. For this reason, we next analyzed all 204 adenocarcinoma cases. Among the 11 genes with 12 probes, 9 genes with 10 probes showed significant associations with both RFS and OS in all 204 adenocarcinoma cases and also in 162 stage I adenocarcinoma cases. *LOC152225* and *KIF19* were excluded because of no significant associations in stage I adenocarcinoma cases. As predicted, higher expression of the 9 genes was correlated with poorer prognosis in the analysis of RFS and OS among 204 stages I and II cases and also among 162 stage I cases.

The result strongly indicated that unsupervised hierarchical clustering using this 10 probe set (9 genes) would separate the patients into high-risk and low-risk groups for prognosis and that all group A triple-negative adenocarcinoma patients with poor prognosis would be classified into the high-risk group (Fig. 3 and Supplementary Table S3). As expected, expression profiling of these 9 genes successfully separated the 204 patients into high-risk and low-risk groups with significantly different RFS (HR = 3.79, 95% CI = 2.19–6.55, P = 1.9E-06) as well as OS (HR = 5.72, 95% CI = 2.53–12.87, P = 2.5E-05). Furthermore, if 62 triple-negative cases only were separated with these 9 genes, HRs for both RFS and OS were much higher than those with separation of all the 204 cases. All the relapsed cases in group A were separated into the high-risk group in the analyses of both cases (all the 204 cases and the 62 triple-negative cases only), supporting that triple-negative adenocarcinomas cases with poor prognosis can be selected as a high-risk group from all the adenocarcinoma cases by expression profiling of these 9 genes (Fig. 3). This profiling further separated 162 stage I cases as well as 46 stage I triple-negative adenocarcinoma cases into high-risk and low-risk groups with significantly different RFS as well as OS (Supplementary Fig. S5 and Supplementary Table S3). Again, HRs for both RFS and OS were much higher in triple-negative adenocarcinoma cases than in all adenocarcinoma cases. Accordingly, high levels of expression in these 9 genes were concluded to be distinct characteristics of triple-negative adenocarcinomas with poor prognosis.

Validation of associations using independent expression profiling data

To validate the present findings using the data of other cohorts, we searched for expression profiling data with

mutation data of the *EGFR*, *KRAS*, and *ALK* genes in various databases. However, there has been no cohort in which expression profiles specifically in triple-negative adenocarcinomas were analyzed. Therefore, unsupervised hierarchical clustering using these 9 genes was done on a cohort of 117 Japanese lung adenocarcinoma cases because expression profile data as well as *EGFR/KRAS* mutation data were available only in this cohort (32). This study included 57 adenocarcinoma cases without *EGFR* and *KRAS* mutations. Although a different array platform was used, the data for all the 9 genes were available for clustering. These cases were separated into 2 groups of 33 cases and 24 cases (Fig. 3). OS of the 33 cases was significantly shorter than that of the 24 cases (HR = 3.17, 95% CI = 1.17–8.63, $P = 2.4E-02$; Supplementary Table S3). As with our cohort, the high-risk group showed a significantly higher HR of 2.73, even when all the 117 cases were analyzed together. Although *ALK* mutation data were not available for this cohort, the results strongly supported that expression profiling of the 9 genes would be highly informative for prediction of prognosis of lung adenocarcinoma patients, in particular patients with *EGFR*- and *KRAS*-negative adenocarcinomas.

Associations of *DEPDC1* expression with prognosis of NSCLC patients

Associations of gene expression with prognosis in various cancers are available from the PrognoScan database (22). Therefore, associations of expression of these 9 genes with prognosis of NSCLC patients were examined in 7 other cohorts (Table 4). Notably, *DEPDC1* expression was positively associated with poor prognosis in 4 of the 7 cohorts; MSK, Nagoya, Duke, and Seoul. The results strongly indicated that *DEPDC1* expression can be a novel prognostic marker for patients with NSCLC. Representative data showing the association of *DEPDC1* expression with prognosis in 204 adenocarcinoma patients obtained from the minimum *P* value approach are shown in Supplementary Fig. S6. Associations of *DEPDC1* expression with RFS and OS were validated by quantitative RT-PCR analysis of 204 stages I and II cases and also of 162 stage I cases (Supplementary Fig. S3).

FOSL2 expression was associated with prognosis in 3 of the 7 cohorts, whereas *MCM4*, *CD300A*, and *UBE2S* expression was associated in 1 cohort, respectively (Table 4).

Discussion

In this study, we attempted to characterize *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas by genome-wide expression profiling. For this purpose, we selected a set of genes that are not transcriptionally activated in any *EGFR*-positive and *KRAS*-positive adenocarcinomas, and obtained 2 pieces of unique evidence. One is that *ALK*-positive adenocarcinomas show unique expression profiles in comparison with any other types of adenocarcinomas. The other is that there is a group of patients with extremely poor prognosis among triple-negative adenocarcinomas. This group, herein designated as group A, of patients showed much worse prognoses than patients with *EGFR*, *KRAS*, or *ALK* mutations and also than the other group, group B, of patients with triple-negative adenocarcinomas.

ALK-positive adenocarcinomas are sensitive to ALK TKIs with an overall response rate of 55% (8). Therefore, for the clinical application of ALK-targeted therapy, it is indispensable to develop a simple and reliable method for detection of *ALK* rearrangements in lung adenocarcinomas. Here, we showed that *ALK* expression is exclusively high only in *ALK*-positive adenocarcinomas and that several other genes, including *GRIN2A*, are overexpressed together with *ALK* specifically in *ALK*-positive adenocarcinomas. Therefore, GRIN2A can be a biomarker for detection of *ALK*-positive adenocarcinomas. *GRIN2A* encodes an *N*-methyl-D-aspartate (NMDA) receptor, which is a neurotransmitter-gated ion channel involved in regulation of synaptic function in the central nervous system (33). It was noted that the *GRIN2A* gene was recently reported to be frequently mutated in melanoma (34). Therefore, although the biological significance of *GRIN2A* upregulation in *ALK*-positive adenocarcinomas remains unclear, *GRIN2A* expression may play some important role in the phenotype unique to *ALK*-positive adenocarcinomas. Expression profiles unique to *ALK*-positive adenocarcinomas, shown here, will be also informative to improve clinical detection of *ALK* rearrangements.

Group A cases were discriminated by expression profiling of 9 genes among stage I–II cases who received complete surgical resection of tumors. Therefore, this gene set will be applicable as biomarkers to select lung adenocarcinoma patients who will benefit from adjuvant therapy after surgery, in particular to select them among patients with triple-negative adenocarcinomas. For this reason, combined analyses of this expression profiling with mutational analyses of the *EGFR*, *KRAS*, and *ALK* genes will be appropriate to pick out triple-negative adenocarcinoma patients with poor prognosis from all the adenocarcinoma patients. Molecular targeting drugs against triple-negative adenocarcinomas are not available at present; therefore, genes upregulated in group A cases will also be applicable as targets for therapy. *DEPDC1* was previously identified as being upregulated in bladder cancer and breast cancer (35–37). Because *DEPDC1* expression was hardly detectable in any normal tissues except testis, it has been considered as a cancer/testis antigen and also as a promising target of therapeutic drugs (35, 36). This study showed that *DEPDC1* is preferentially expressed in triple-negative adenocarcinomas with poor prognosis. In the PrognoScan database, *DEPDC1* expression is shown to be positively associated with poor prognosis in bladder cancer, multiple myeloma, breast cancer, glioma, and melanoma. Therefore, *DEPDC1* could be a novel target for diagnosis as well as therapy in various cancers, including lung adenocarcinoma.

Identification of genetic alterations that occur specifically in group A cases will be also of great importance for the development of target therapy for stages I and II lung adenocarcinoma patients with poor outcomes. Group A cases include males and ever-smokers as a majority (Table 1); therefore, group A cases were likely to carry several genetic alterations induced by tobacco carcinogens leading to poor outcomes. Identification of genetic alterations in

group A adenocarcinomas will further facilitate the development of targeted therapies for lung adenocarcinomas with poor prognosis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Dr. Teruhiko Yoshida and Ms. Sachiyo Mimaki for their efforts in expression profiling.

Grant Support

This work was supported in part by grants-in-aid from the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control and from the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO: 10–41). K. Shiraishi was an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

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.

Received April 23, 2011; revised October 19, 2011; accepted October 21, 2011; published OnlineFirst November 11, 2011.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
3. Colvy T, Noguchi M, Henschke C, Vazquez M, Geisinger K, Yokose T, et al. Adenocarcinoma. In:Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 35–44.
4. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol 2011;12:175–80.
5. Herbst RS, Heymach JV, Lippman SM: Lung cancer. N Engl J Med 2008;359:1367–80.
6. Janku F, Stewart DJ, Kurzrock R. Targeted therapy in non-small-cell lung cancer–is it becoming a reality? Nat Rev Clin Oncol 2010;7: 401–14.
7. Bronte G, Rizzo S, La Paglia L, Adamo V, Siragusa S, Ficorella C, et al. 2288; Driver mutations and differential sensitivity to targeted therapies: a new approach to the treatment of lung adenocarcinoma. Cancer Treat Rev 2010;36 Suppl 3:S21–9.
8. Gerber DE, Minna JD. ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. Cancer Cell 2010;18:548–51.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. Nat Rev Cancer 2007;7:778–90.
10. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 2009;361:958–67.
11. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al.: Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947–57.
12. Konstantinopoulos PA, Karamouzis MV, Papavassiliou AG. Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. Nat Rev Drug Discov 2007;6:541–55.
13. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
14. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 2006;24:1679–88.
15. Angulo B, Suarez-Gauthier A, Lopez-Rios F, Medina PP, Conde E, Tang M, et al. Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. J Pathol 2008;214:347–56.
16. Motoi N, Szoke J, Riely GJ, Seshan VE, Kris MG, Rusch VW, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. Am J Surg Pathol 2008;32:810–27.
17. Meyerson M, Carbone D. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. J Clin Oncol 2005;23:3219–26.
18. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 1–344.
19. 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.
20. Robins JM, Gail MH, Lubin JH. More on "Biased selection of controls for case-control analyses of cohort studies". Biometrics 1986;42: 293–9.
21. Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med 2004;61:e59.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8. Available from: http://rana.lbl.gov/eisen/.
23. Takano T, Ohe Y, Sakamoto H, Tsuta K, Matsuno Y, Tateishi U, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. J Clin Oncol 2005;23: 6829–37.
24. Takano T, Ohe Y, Tsuta K, Fukui T, Sakamoto H, Yoshida T, et al. Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. Clin Cancer Res 2007;13:5385–90.
25. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, et al. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. Clin Cancer Res 2008;14:6618–24.
26. Mizuno H, Kitada K, Nakai K, Sarai A. 2288; PrognoScan: a new database for meta-analysis of the prognostic value of genes. 2288; BMC Med Genomics 2009;2:18. Available from: http://gibk21.bio.kyutech.ac.jp/PrognoScan/index.html.
27. Mitsudomi T. Advances in target therapy for lung cancer. Jpn J Clin Oncol 2010;40:101–6.
28. Suda K, Tomizawa K, Mitsudomi T. Biological and clinical significance of KRAS mutations in lung cancer: an oncogenic driver that contrasts with EGFR mutation. Cancer Metastasis Rev 2010;29: 49–60.
29. Zhang X, Zhang S, Yang X, Yang J, Zhou Q, Yin L, et al. Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. Mol Cancer 2010;9:188.
30. Boland JM, Erdogan S, Vasmatzis G, Yang P, Tillmans LS, Johnson MR, et al. Anaplastic lymphoma kinase immunoreactivity correlates with ALK gene rearrangement and transcriptional up-regulation in non-small cell lung carcinomas. Hum Pathol 2009; 40:1152–8.
31. Mino-Kenudson M, Chirieac LR, Law K, Hornick JL, Lindeman N, Mark EJ, et al. A novel, highly sensitive antibody allows for the routine detection of ALK-rearranged lung adenocarcinomas by standard immunohistochemistry. Clin Cancer Res 2010;16:1561–71.

32. 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.

33. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet 2010;42:1021–6.

34. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. Nat Genet 2011;43:442–6.

35. Kanehira M, Harada Y, Takata R, Shuin T, Miki T, Fujioka T, et al. Involvement of upregulation of DEPDC1 (DEP domain containing 1) in bladder carcinogenesis. Oncogene 2007;26:6448–55.

36. Harada Y, Kanehira M, Fujisawa Y, Takata R, Shuin T, Miki T, et al. Cell-permeable peptide DEPDC1-ZNF224 interferes with transcriptional repression and oncogenicity in bladder cancer cells. Cancer Res 2010; 70:5829–39.

37. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W. Identification of early molecular markers for breast cancer. Mol Cancer 2011;10:15.

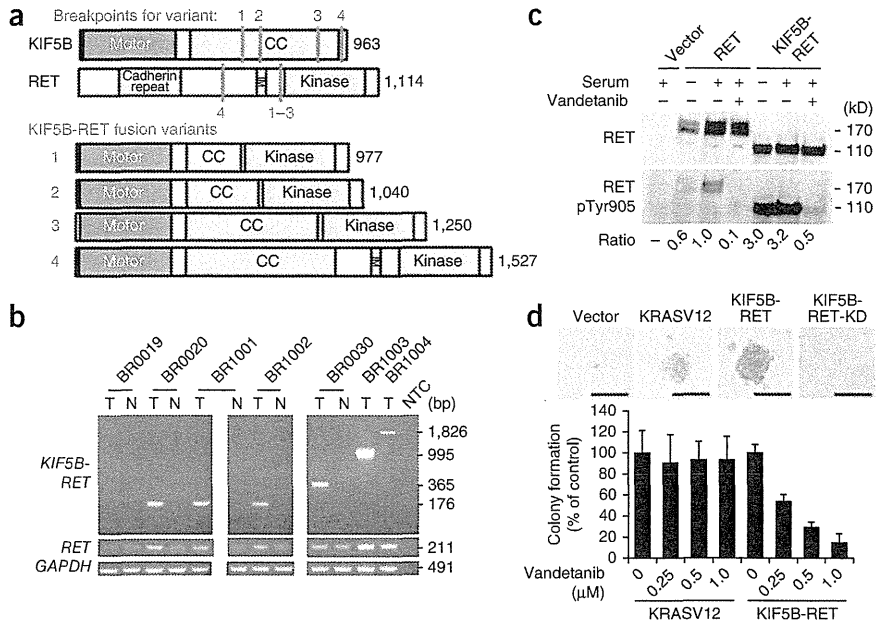

Figure 1 *KIF5B-RET* fusions in LADC. (**a**) Schematic representations of the wild-type KIF5B and RET proteins as well as the four fusion variants identified in this study. The breakpoints for each variant are indicated with red lines. CC, coiled coil; TM, transmembrane. (**b**) Detection of *KIF5B-RET* fusions by RT-PCR. RT-PCR products for the *RET* kinase domain (exons 12 and 13) and *GAPDH* are shown below. Six LADCs positive for *KIF5B-RET* fusions (T) are shown, with four corresponding non-cancerous lung tissues (N), a no-template control (NTC) and one LADC that was negative for the fusion (BR0019). (**c**) Activation of RET kinase activity in the KIF5B-RET protein and the suppression of this activity by vandetanib. H1299 lung cancer cells were transfected with an empty vector, wild-type RET (RET) or KIF5B-RET expression plasmids and treated either with DMSO (serum) or vandetanib, as indicated. The ratios of phosphorylated Tyr905 (pTyr905) RET to total RET signals with respect to wild-type RET after the serum treatment are listed below the gels. (**d**) Anchorage-independent growth of NIH3T3 cells expressing KIF5B-RET protein and the suppression of this growth by vandetanib. Representative pictures of colonies without vandetanib treatment (top). Scale bars, 50 μm. Bar graph showing the percentage (± s.d.) of colonies formed after treatment with the indicated amounts of vandetanib (average results of three independent experiments) with respect to those formed by DMSO-treated cells. The study was approved by the institutional review boards of institutions participating in this study.

United States (an individual of European ancestry) (**Supplementary Fig. 7**), but we detected no fusion transcripts in the 34 subjects from Norway (**Supplementary Table 3**); *KIF5B-RET* fusions occurred in 1–2% of LADCs in both Asians and non-Asians. The individual from the United States with the *RET* fusion was classified as an 'ever smoker', whereas the six individuals from Japan with the *RET* fusion were 'never smokers' (**Table 1**). Therefore, prevalence of LADC with regard to smoking status is unclear. We did not detect the *KIF5B-RET* fusion in other major subtypes of lung cancer, including 234 squamous-cell, 17 large-cell and 20 small-cell lung carcinomas (**Supplementary Table 3**). The fusion was also not present in other types of adenocarcinomas, including those of the ovary ($n = 100$) and colon ($n = 200$) (data not shown), suggesting that it is specific to LADC.

All seven subjects with LADC harboring the *KIF5B-RET* fusion were negative for *EGFR*, *KRAS* and *ALK* mutations or fusions and were negative for mutations in *HER2*, which is an additional driver mutation in LADC[10] (**Table 1** and **Supplementary Table 4**). The mutually exclusive nature of the *RET* fusions and other oncogenic alterations[1,2,11] suggests that the *KIF5B-RET* fusion is a driver mutation. All proteins encoded by the four *KIF5B-RET* fusion variants contained the KIF5B coiled-coil domain, which functions in protein dimerization[12], and retained the full RET kinase domain, similar to other types of oncogenic *RET* fusions observed in thyroid tumors (**Fig. 1a**)[13]. The KIF5B-RET proteins are likely to form a homodimer through the coiled-coil domain of KIF5B, causing an aberrant activation of the kinase function of RET in a manner similar to the *PTC-RET* and *KIF5B-ALK* fusions[7,14]. In fact, the N-terminal portion of the KIF5B coiled-coil region, which is retained in all variants, has been predicted to have the ability to dimerize through two coiled-coil structures[15]. Consistently, when the *KIF5B-RET* variant 1 was exogenously expressed in H1299 human lung cancer cells without wild-type or fusion *RET* expression, Tyr905, which is located in the activation loop of the RET kinase site[15,16], was phosphorylated in the absence of serum stimulation, indicating an aberrant activation of *RET* kinase[16,17] by fusion with *KIF5B* (**Fig. 1c**). This phosphorylation was suppressed by vandetanib, a TKI against RET (as well as other tyrosine kinases, including EGFR and VEGFR)[18] (**Fig. 1c** and **Supplementary Fig. 8**).

Expression of exogenous *KIF5B-RET*, but not *KIF5B-RET*-KD (a kinase-dead mutant corresponding to S765P in wild-type RET[17]), induced morphological transformation (**Supplementary Fig. 9**) and anchorage-independent growth of NIH3T3 fibroblasts in a way that was analogous to the induction caused by mutant *KRAS* (KRASV12) (**Fig. 1d**). Consistently, phosphorylation of Tyr905 was higher in the KIF5B-RET

Table 1 Characteristics of lung adenocarcinomas with the *KIF5B-RET* fusion

Sample	Country	Sex	Age[a]	Smoking	*KIF5B-RET* fusion[b]	Pathological stage	Pathological findings	RET staining	TTF-1 staining	Napsin A staining	Thyroglobulin staining
BR0020	Japan	Male	57	Never	K15; R12 (variant 1)	IIB	Moderately differentiated ADC	+	+	+	–
BR1001	Japan	Female	65	Never	K15; R12 (variant 1)	IB	Well differentiated ADC	+	+	+	–
BR1002	Japan	Female	64	Never	K15; R12 (variant 1)	IB	Well differentiated ADC	+	+	+	–
BR0030	Japan	Male	57	Never	K16; R12 (variant 2)	IA	Well differentiated ADC	+	+	+	–
BR1003	Japan	Male	28	Never	K23; R12 (variant 3)	IA	Well differentiated ADC	+	+	+	–
BR1004	Japan	Female	71	Never	K24; R8 (variant 4)	IA	Moderately differentiated ADC	NT	NT	NT	NT
NCI1580	USA	Male	63	Ever[c]	K15; R12 (variant 1)	II	Moderately differentiated ADC	NT	NT	NT	NT

[a]Age in years. [b]Fused exon numbers of *KIF5B* (K) and *RET* (R); and variant types (in parentheses) are shown. None of the subjects had oncogenic *EGFR*, *KRAS*, *HER2* or *ALK* mutations or fusions. [c]The number of pack years smoked for this subject is not known. NT, not tested.

protein than in the KIF5B-RET-KD protein. The anchorage-independent growth induced by *KIF5B-RET* was suppressed by treatment with vandetanib (<1 μM), whereas the growth induced by mutant *KRAS* was not (**Fig. 1d**). These results are similar to those observed for *RET* fusions in thyroid cancer[19]. We also detected phosphorylation of the KIF5B-RET protein at Tyr905 in fusion-positive LADC specimens (**Supplementary Fig. 6**). These results suggest that the *RET* fusions are a previously unidentified LADC driver mutation and a potential target for existing TKIs, including vandetanib, which has been recently approved by the US Food and Drug Administration for the treatment of thyroid cancer[18]. Further studies are warranted to promote molecular subtype diagnoses and personalized therapy options for LADC. For this purpose, both the clinical and biological features of this fusion are being investigated. For further information, please see the **Supplementary Note** and **Supplementary Tables 5** and **6**.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

This work was supported in part by the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio), Grants-in-Aid from the Ministry of Health, Labour and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control, the National Cancer Center Research and Development Fund and the Norwegian Cancer Society. National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan. We thank T. Urushidate, S. Ohashi, S. Mitani, K. Yokozawa, S. Wakai, C. Otsubo and H. Isomura of the National Cancer Center and D. Suzuki and K. Nagase of the National Center for Global Health and Medicine for technical assistance. We also thank J.D. Minna and L. Girard of the University of Texas Southwestern Medical Center, K. Kumamoto of Saitama Medical University and A. Okamoto of Jikei University for *RET* fusion screening, N. Morii of the National Institute of Advanced Industrial Science and Technology (AIST) for thermodynamic characterization of the KIF5B protein and M. Maekawa of the GSP laboratory for rapid preparation of the FISH probes.

AUTHOR CONTRIBUTIONS

RNA sequencing: H.I., K.Y., M.H., T.N. and H.S. Sequence data processing: Y.T., S.C. and I.Y. Molecular biological analyses: T.K., Y.S., R.I., H. Ogiwara, T.O., M.E., A.J.S., H. Okayama, A.H., Y.A. and S.O. Clinical and pathological analyses: K.T., K.F., V.S., S.W., I.S. and H.T. Manuscript writing: T.K., H.I. and T.S. Study design: T.K., H.I., C.C.H., T.Y., J.Y. and T.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemedicine/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Janku, F., Stewart, D.J. & Kurzrock, R. *Nat. Rev. Clin. Oncol.* **7**, 401–414 (2010).
2. Gerber, D.E. & Minna, J.D. *Cancer Cell* **18**, 548–551 (2010).
3. Lovly, C.M. & Carbone, D.P. *Nat. Rev. Clin. Oncol.* **8**, 68–70 (2011).
4. Soda, M. *et al. Nature* **448**, 561–566 (2007).
5. Meyerson, M., Gabriel, S. & Getz, G. *Nat. Rev. Genet.* **11**, 685–696 (2010).
6. Mani, R.S. & Chinnaiyan, A.M. *Nat. Rev. Genet.* **11**, 819–829 (2010).
7. Wells, S.A. Jr. & Santoro, M. *Clin. Cancer Res.* **15**, 7119–7123 (2009).
8. Bishop, J.A., Sharma, R. & Illei, P.B. *Hum. Pathol.* **41**, 20–25 (2010).
9. DeLellis, R.A., Shin, S.J. & Treaba, D.O. *Immunohistology of Endocrine Tumors* (Saunders, Philadelphia, Pennsylvania, USA, 2010).
10. Shigematsu, H. *et al. Cancer Res.* **65**, 1642–1646 (2005).
11. Herbst, R.S., Heymach, J.V. & Lippman, S.M. *N. Engl. J. Med.* **359**, 1367–1380 (2008).
12. Hirokawa, N., Noda, Y., Tanaka, Y. & Niwa, S. *Nat. Rev. Mol. Cell Biol.* **10**, 682–696 (2009).
13. Jhiang, S.M. *Oncogene* **19**, 5590–5597 (2000).
14. Takeuchi, K. *et al. Clin. Cancer Res.* **15**, 3143–3149 (2009).
15. Morii, H., Takenawa, T., Arisaka, F. & Shimizu, T. *Biochemistry* **36**, 1933–1942 (1997).
16. Vitagliano, D. *et al. Endocr. Relat. Cancer* **18**, 1–11 (2011).
17. Croyle, M. *et al. Cancer Res.* **68**, 4183–4191 (2008).
18. Commander, H., Whiteside, G. & Perry, C. *Drugs* **71**, 1355–1365 (2011).
19. Carlomagno, F. *et al. Cancer Res.* **62**, 7284–7290 (2002).