the informed consent of patients undergoing surgery between May 1997 and June 2009 at the Cancer Institute Hospital (Tokyo, Japan). This study was approved by the Institutional Review Board of the Japanese Foundation for Cancer Research. The RCC cases used in the tumor tissue array analysis were consecutive, and histologic diagnosis was performed according to the World Health Organization classification.[7] For the tumor tissue arrays, 1 or 2 spots of the most representative tumor area were selected and cored (2 mm in diameter) using a tissue-arraying instrument (KIN-1, Azumaya, Tokyo, Japan).

Karyotyping

Karyotype analyses with Q-banding were performed as described previously[10] with informed, written consent for cytogenetic studies obtained from the patients.

IHC

A representative tissue block from each lesion (cases 1 to 3) and the tissue array blocks were sectioned at 4-μm thickness, deparaffinized in xylene, and rehydrated through different grades of ethanol. IHC including TFEB (V-17, Santa Cruz, CA; diluted 1:100) and TFE3 (P-16, Santa Cruz; 1:500) was performed using the Histofine SAB-PO Goat kit (Nichirei, Tokyo, Japan) in a Dako Autostainer (Dako, Carpinteria, CA). IHC including HMB45 (HMB45, Dako; diluted 1:50), Melan-A (A103, Dako; 1:50), Cathepsin-K (182-12G5, Daiichi Fine Chemicals, Toyama, Japan; 1:500), E-cadherin (NCH-38, Dako; 1:100), α-methylacyl-CoA racemase (AMACR, 13H4, Abcam, Cambridge, UK; 1:50), CD10 (56C6, Novocastra, Burlingame, CA; 1:100), AE1/AE3 (AE1/AE3, Dako; 1:100), epithelial membrane antigen (EMA, E29, Dako; 1:500), CK7 (OVTL 12/30, Progen, Heidelberg, Germany; 1:50), and Vimentin (V9, Dako; 1:1000) was performed using the EnVision+ dextran polymer kit (Dako) in a Dako Autostainer.

5′-RACE

Total RNAs were isolated from the lesions, which had been grossly dissected and snap-frozen in liquid nitrogen within 20 minutes of removal and stored at −80°C. Total RNAs were purified using an RNeasy Mini Kit (QIAGEN, Valencea, CA). RNAs were treated with an RNase-Free DNase Set (QIAGEN). RNA quality and absence of genomic DNA contamination were checked by formaldehyde agarose gel electrophoresis. 5′-RACE (rapid amplification of cDNA ends) was performed using the SMART RACE cDNA Amplification kit (Clontech, Palo Alto, CA) with a UPM (universal primer mix) and the TFEB-563R primer (5′-CTGGACAGGGGTAG CGTGTTGGGC-3′) according to the manufacturer's instructions. The amplified fragments were subjected to a direct sequence analysis.

Polymerase Chain Reaction

Genomic DNAs were isolated from the frozen lesions using a QIAamp DNA Mini Kit (QIAGEN). Polymerase chain reaction (PCR) experiments were carried out with Alpha-F (5′-GCAGGCGGAGCTT GAGGAAAC-3′) and TFEB-392R (5′-AGCACGTG TCCAGCTCGCA-3′) primers. The PCR analysis was performed using TaKaRa LA Taq (Takara Bio, Otsu, Japan). The PCR conditions were as follows: 94°C for 1 minute, followed by 32 cycles of 10 seconds at 98°C, and for 1 minute 30 seconds at 68°C. The amplified fragments were subjected to a direct sequence analysis using the same primers.

Reverse Transcription Polymerase Chain Reaction

Total RNAs were reverse transcribed with random primers using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA). Reverse transcription polymerase chain reaction (RT-PCR) experiments were

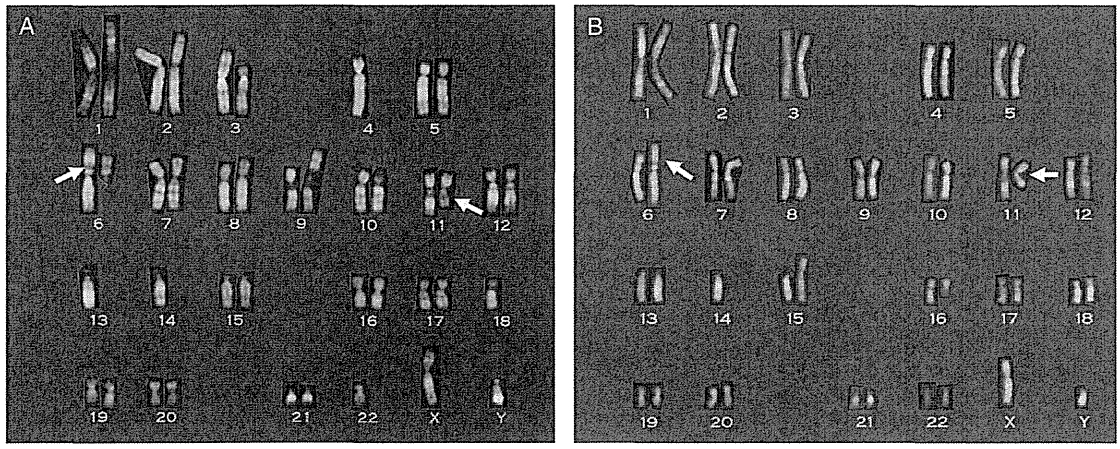


FIGURE 1. Q-banding karyotypes show a t(6;11)(p21.1;q12∼13) chromosomal rearrangement (arrows) in case 2 (A) and case 3 (B). A, 41,XY,i(1)(q10),del(3)(p12),-4,del(6)(q12),t(6;11)(p21.1;q12∼13),i(9)(q10),add(10)(p11),-13,-14,-18,add(19)(q12),-22. B, 45,XY,t(6;11)(p21.1;q12∼13),-14,t(15;16)(p10;p10).

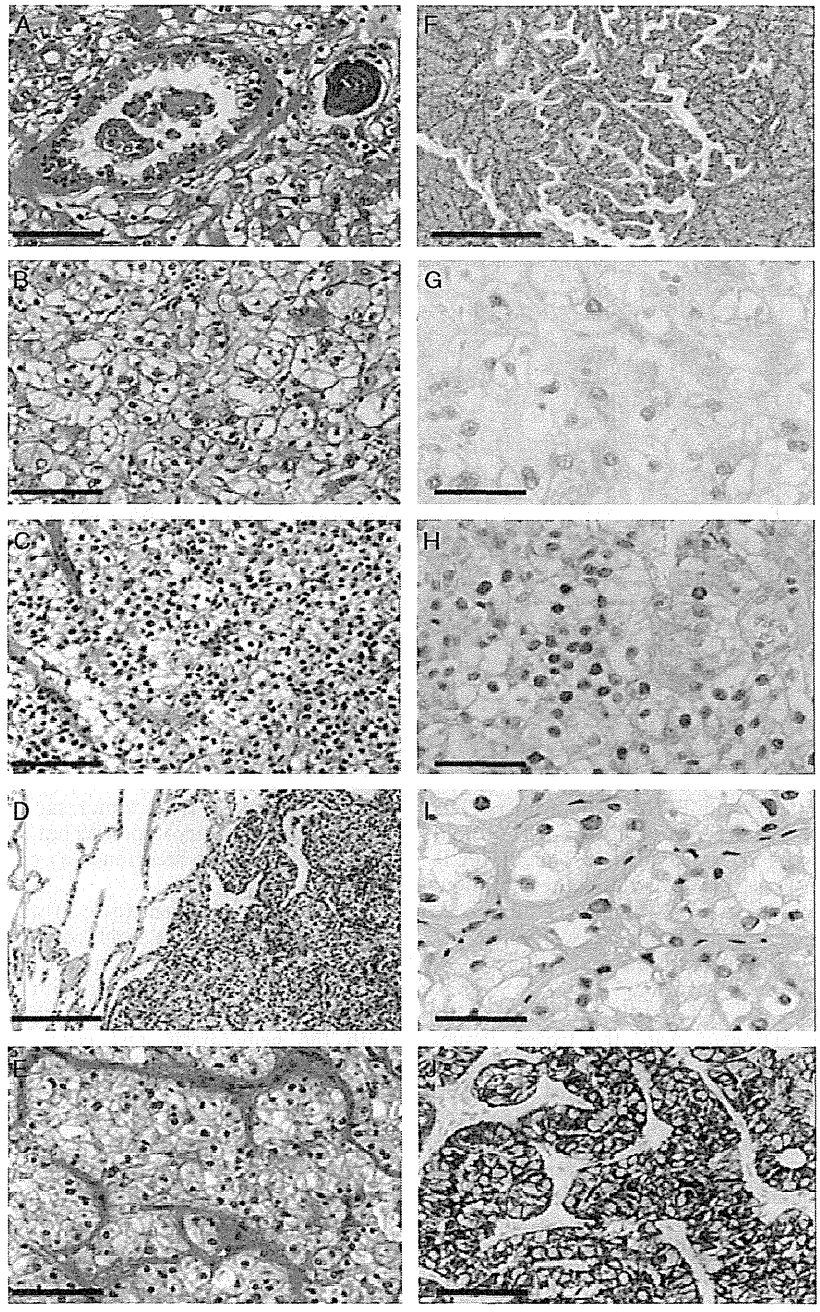


FIGURE 2. A, Pseudopapillary structures with hyaline cores and scattering psammoma bodies were observed in case 1. Bar, 100 μm. B, Voluminous clear and eosinophilic cells with occasional high-grade nuclear atypia were also observed in case 1. Bar, 100 μm. The case 2 tumor, in both the lymph node metastasis (C: Bar, 100 μm) and lung metastasis (D: Bar, 200 μm), consisted of clear and eosinophilic cells with thin-walled staghorn-shaped vasculature. E, The case 3 tumor contained a mixture of eosinophilic and clear cells with voluminous cytoplasm, perinuclear clearance, and well-defined cell borders, showing a nested pattern, separated by relatively thick fibrovascular walls. Bar, 100 μm. F, The papillary architecture was very limited in case 3. Bar, 500 μm. The TFEB IHC analyses showed nuclear staining in all 3 cases (G, case 1; H, case 2; I, case 3). Bar, 50 μm. J, CK7 was diffusely positive in case 3. Bar, 100 μm.

TABLE 1. Summary of Clinical Histories and IHC Features of TFEB Translocation RCCs With IHC Results of TFE3 Translocation RCCs and Common Subtypes of RCCs

TFEB RCC Case No.	Age (y)	Sex	Size (cm)	Status	F/U (mo)	TFEB	TFE3	HMB45	Melan-A	Cath-K	E-cad	AMACR	AE1/AE3	EMA
1	18	M	7	NED	18	+	–	+	+	NA	NA	NA	–	–
2	10	M	12.2	NED	26	+	–	+	+	+	NA	NA	–	–
3	14	F	4.5	NA	NA	+	NA	+	+	+	NA	NA	–	NA
4	42	F	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5	17	F	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
6	18	F	2.8	NED	18	+	NA	+	+	NA	NA	NA	–	–
7	20	F	9.5	NED	30	+	NA	+	+	+	NA	NA	–	NA
8	9	F	2.0	NA	NA	+	NA	+	+	+	NA	NA	NA	NA
9	33	M	6.0	NA	NA	+	NA	+	+	+	NA	NA	NA	–
10	42	F	NA	Meta	36	+	–	+	+	NA	NA	NA	–	NA
11	54	F	7	NED	36	+	–	+	+	+	NA	NA	–	NA
12	6	F	5	NED	3	+	NA	+	+	+	NA	NA	+	NA
13	34	F	15	NED	50	+	–	+	+	NA	+	+	+	+
14	36	M	20	DOD	3	+	–	+	+	NA	–	+	–	+
15	9	F	NA	NED	42	+	–	NA	NA	NA	NA	NA	NA	NA
16	22	M	4.0	NED	42	NA	–	+	+	NA	NA	NA	NA	NA
17	24	F	13.6	NED	20	NA	–	+	+	NA	NA	NA	NA	NA
18	39	F	1.0	NED	17	NA	–	+	+	NA	NA	NA	NA	NA
19	22	M	10	NED	NA	+	–	+	+	NA	+	+	+	–
20	26	M	4.3	NED	6	+	–	+	+	NA	NA	NA	+	–
Case 1	57	M	10.2	NED	8	+	–	+	+	+	+	+	+	–
Case 2	37	M	NA	DOD	120	+	–	+	+	+	+	+	+	–
Case 3	47	M	3.2	NED	12	+	–	–	+	–	+	–	+	+
Total	23(median) (6 to 57)	M:F = 10:13	6.5(median) (1.0-20)	Recurrence = 17% (3/18)		100% (18/18)	0% (0/15)	95% (19/20)	100% (20/20)	90% (9/10)	83% (5/6)	83% (5/6)	47% (7/15)	27% (3/11)

Cath-K indicates Cathepsin-K; DOD, died of disease; E-cad, E-cadherin; F/U, follow-up; Meta, metastasis; NA, not available; NED, no evidence of disease; papRCC, References: TFEB RC case no.1[1,2,6], no.2[1,2,13], no.3[2,11,13], no.4[11], no.5[11], no.6[2,6], no.7[2,13], no.8[2,13], no.9[2,13], no.10[12], no.11[12,13,15], no.12[3,13], no.13[5], no.14[5], no.15[8],

carried out with Alpha-F and TFEB-563R primers. The RT-PCR analysis was performed using AmpliTaq Gold (Applied Biosystems, Foster City, CA). The RT-PCR conditions were as follows: 94°C for 10 minutes, followed by 35 cycles of 1 minute at 94°C, and for 1 minute 10 seconds at 68°C. The amplified fragments were subjected to a direct sequence analysis using the same primers.

Western Blotting

Proteins were extracted from the frozen tissues using RIPA Buffer (Nacalai Tesque, Kyoto, Japan). Equal amounts of protein were separated by SDS-polyaclylamidegel electrophoresis (12%) and transferred to polyvinylidene fluoride membranes (Hybond-P, GE Healthcare, Little Chalfont, UK). For immunoblotting, the following antibodies were used: anti-TFEB rabbit polyclonal antibody (#ab96834, Abcam, Cambridge, UK; 1:1000) and anti-β-Actin rabbit polyclonal antibody (#4967, Cell Signaling Technology, Beverly, MA; 1:1000). Detection was carried out using an ECL detection system (GE Healthcare).

RESULTS

Clinicopathologic and Cytogenetic Features

Our first patient (case 1) was a 57-year-old man who had a 10.2-cm-sized right renal mass and who had undergone a total nephrectomy at our hospital in 2008. During surgery, a fresh tumor sample (ref #29278) was collected, but its karyotype was not obtained. He was still doing well without recurrence as of 8 months after the surgery.

The second patient (case 2) was a 37-year-old man who presented with a right renal mass and who had undergone a total nephrectomy in 1989. Eight years after the nephrectomy, he presented with lung and mediastinal lymph node metastases and was referred to our hospital with the hematoxylin and eosin-stained slides of the previously removed renal mass. We diagnosed the mass as clear cell-type RCC (ccRCC) at the time. Subsequently, he underwent lymph node dissections and partial resection of the lung for the metastatic tumor (4.5 cm in size). At that time, a fresh tumor sample (ref #1479) was collected, and karyotyping revealed a t(6;11)(p21.1;q12~13) chromosomal rearrangement, characterizing the tumor as an RCC with a t(6;11) translocation (Fig. 1A). Thirty months after the surgery for the lung and lymph node metastases, the patient died of multiple metastases to the lung and bone.

The third patient (case 3) was a 47-year-old man who had a 3.2-cm-sized left renal mass and who had undergone a partial nephrectomy at our hospital in 2009. During surgery, a fresh tumor sample (ref #32328) was collected, and karyotyping revealed a t(6;11)(p21.1;q12~13) chromosomal rearrangement (Fig. 1B). As of 12 months after the surgery, he is healthy without recurrence.

TABLE 1. *(continued)*

CK7	Vim	CD10	Sub type	TFEB	TFE3	HMB45	Melan-A	Cath-K	E-cad	AMACR	AE1/AE3	EMA	CK7	Vim	CD10
NA	NA	+	TFE3 RCC	0% (0/39)	95% (37/39)	42% (11/26)	88% (23/26)	60% (6/10)	69% (18/26)	100% (28/28)	23% (6/26)	28% (8/29)	14% (4/28)	63% (17/27)	100% (26/26)
NA	NA	–	cc RCC	0% (0/165)	0% (0/165)	0.6% (1/165)	0% (0/165)	0% (0/165)	26% (43/165)	12% (20/165)	85% (141/165)	70% (116/165)	12% (19/165)	81% (134/165)	93% (153/165)
NA	NA	–	chro RCC	0% (0/23)	0% (0/23)	0% (0/23)	13% (3/23)	4% (1/23)	91% (21/23)	26% (6/23)	100% (23/23)	100% (23/23)	91% (21/23)	4% (1/23)	43% (10/23)
NA	NA	NA	pap RCC	0% (0/12)	0% (0/12)	8% (1/12)	0% (0/12)	0% (0/12)	50% (6/12)	58% (7/12)	100% (12/12)	75% (9/12)	83% (10/12)	100% (12/12)	75% (9/12)
NA	NA	NA													
NA	NA	+													
NA	+	+													
NA	NA	NA													
–	NA	–													
–	+	+													
–	+	–													
NA	NA	NA													
+	+	+													
–	+	+													
NA	NA	NA													
NA	NA	NA													
NA	NA	NA													
NA	NA	NA													
–	+	+													
–	NA	+													
–	+	–													
–	+	–													
+	–	+													
20% (2/10)	89% (8/9)	60% (9/15)													

papillary-type renal cell carcinoma; TFE3 RCC, *TFE3* translocation renal cell carcinoma; Vim, Vimentin.
no.16 - 18[9], no.19[16], no.20[17]. Immunohistochemical results of TFE3 RCCs were drawn from the studies reported by Camparo et al[5] and Martignoni et al.[13]

All of our 3 patients were Japanese adult male office workers without any particular exposure to occupational carcinogens. None of them had a history of other tumors, nor had they received chemotherapy previously. Although the mother of the second patient had suffered from gastric cancer, one of the most common tumors in Japan, the others had no familial history of cancer. The first and second patients were smokers with a cumulative smoking history of 41 and 11 pack-years, respectively. The third patient had no smoking history.

Pathologic and Immunohistochemical Analysis

Histologically, the case 1 tumor had typical features of translocation RCC, such as a papillary architecture including pseudopapillary structures with hyaline cores, scattering psammoma bodies, and voluminous clear and eosinophilic cells with occasional high-grade nuclear atypia (Fig. 2A, B). The case 2 tumor, in both the lymph node metastasis (Fig. 2C) and the lung metastasis (Fig. 2D), consisted of clear and eosinophilic cells with a thin-walled staghorn-shaped vasculature. The metastatic tumor closely resembled ccRCC histologically. The case 3 tumor contained a mixture of eosinophilic and clear cells with a voluminous cytoplasm, occasional perinuclear clearance, and well-defined cell borders, showing a predominant nested pattern, separated by relatively thick fibrovascular walls (Fig. 2E). The sinusoidal vascular network, usually prominent in ccRCC, was indistinct, and the papillary architecture was very limited in this tumor (Fig. 2F). In addition, Hale's colloidal iron staining, useful for the diagnosis of chromophobe-type RCC (chroRCC), was weakly and focally positive in the cytoplasm of these tumor cells. Our initial diagnosis of the case 3 tumor was chroRCC, although its histology was not typical. Of these 3 TFEB RCCs, only case 1 showed the typical histologic features of translocation RCC, established by more abundant cases of *TFE3* translocation RCC (TFE3 RCC), and there were few of the previously reported histologic features of translocation RCC present in cases 2 and 3.

Table 1 summarizes the IHC features and clinical histories of our 3 cases and the previously reported 20 cases with TFEB RCC. For comparison, the IHC results of TFE3 RCCs, reported by Camparo et al[5] and Martignoni et al,[13] and common subtypes of 200 RCCs generated by our tissue array analysis are also attached. All 3 of our present TFEB RCC tumors were immunoreactive to a TFEB antibody, as shown in Figures 2G to I for cases 1 to 3, respectively, and none of the ordinary RCCs were positive. Almost all TFEB RCCs (95%, with the only exception being the present case 3) expressed HMB45, and all TFEB RCCs (100%) expressed Melan-A.

Cathepsin-K was frequently expressed in TFEB RCCs (90%) and TFE3 RCCs (60%), whereas only 1 of the 200 specimens of common subtypes of RCCs (0.5%) was found to express Cathepsin-K. E-cadherin was often positive in TFEB RCCs (83%) and TFE3 RCCs (69%), whereas the positive rates for E-cadherin were 91% in chroRCCs and 50% in papillary-type RCCs. The positive rates of AMACR were high in TFEB RCCs (83%) and TFE3 RCCs (100%); however, papillary-type RCCs often expressed AMACR (58%) as well. Some TFEB RCCs expressed AE1/AE3 (47%), EMA (27%), and CK7 (20%). Interestingly, case 3 had a borderline immunophenotype between TFEB RCC and chroRCC, whereas cases 1 and 2 had the same immunophenotype, which was largely in accordance with the previously reported TFEB RCCs. In fact, case 3 was positive for TFEB and Melan-A, which was characteristic for TFEB RCC, but shared the phenotype of chroRCC, including diffuse positive staining for CK7 (Fig. 2J) and negative staining for HMB45, Cathepsin-K, and AMACR.

Molecular Analysis

We performed 5′-RACE to detect the *Alpha-TFEB* fusion. The TFEB-563R primer, used in 5′-RACE and RT-PCR, was designed to span an exon 5/6 junction and not to yield genomic products. Although the 5′-RACE analyses yielded 1 fusion product for case 1 and case 3, two fusion products were generated for case 2 (Fig. 3A). To confirm the presence of an *Alpha-TFEB* fusion in each tumor, we directly amplified the fusion point of the *Alpha-TFEB* DNA by genomic PCR (Fig. 3B) and the *Alpha-TFEB* complemantary DNA by RT-PCR (Fig. 3C). The TFEB-392R primer, used in genomic PCR experiments, was designed to anneal to the sequence corresponding to exon 4. In case 1 and case 3, both genomic PCR and RT-PCR yielded 1 fusion product. In case 2, RT-PCR yielded 2 independent products (case 2-v1 and case 2-v2), although genomic PCR generated 1 fusion product. Nucleotide sequencing of each genomic PCR and RT-PCR product revealed a break point each (Fig. 3D). Figure 3E shows schematic representations of each fusion pattern. In case 1, the DNA and messenger ribonucleic acid (mRNA) had different fusion points. In case 2, the tumor had 2 variants of mRNA, one of which directly fused to exon 3 of *TFEB* and was identical to the fusion point of the case 1 mRNA. In case 3, the fusion point of *TFEB* was nt 79 of exon 4, downstream of the wild-type ATG in exon 3. A Western blot analysis using an anti-TFEB antibody showed a TFEB protein of almost the same size in all the 3 cases (Fig. 3F). In case 3, although the fusion point of *TFEB* was downstream of the wild-type ATG in exon 3, the protein product size was almost equal to that of the wild-type TFEB protein (approximately 53 kDa) observed in cases 1 and 2.

DISCUSSION

TFEB RCC is a recently recognized tumor entity and may be easily misdiagnosed because of its infrequency and lack of established characteristics. We herein examined 3 new cases of TFEB RCC and performed precise clinicopathologic, cytogenetic, and molecular analyses to elucidate the features of this tumor entity.

The most compelling IHC evidence for TFEB RCC was the detection of nuclear staining for the TFEB protein, although the staining may be technically difficult to perform. Positive staining for TFEB was detected in all 3 of our TFEB RCC cases, and none of the other 200 RCCs expressed TFEB. The prominent sensitivity and specificity for TFEB immunostaining as a characteristic of TFEB RCC will contribute to the definite diagnosis of TFEB RCC. Both TFEB and TFE3 belong to the microphthalmia transcription factor/transcription factor E (MiTF/TFE) subfamily and activate melanocyte-specific promoters. Because of these activities, HMB45 was positive in almost all TFEB RCCs, and all TFEB RCCs expressed Melan-A. As TFEB, TFE3, and MiTF activate the expression of Cathepsin-K, it was not surprising that it was frequently expressed in TFEB RCCs and TFE3 RCCs. As only 1 of the 200 common subtypes of RCCs expressed Cathepsin-K, its immunoexpression may also be useful in distinguishing translocation RCCs from the common subtypes of RCCs, as suggested by Martignoni et al.[13] In contrast, E-cadherin and AMACR were often expressed in translocation RCCs; however, some ordinary RCCs also expressed them, making them nonspecific for a diagnosis. Some previous reports suggested that translocation RCCs are nonimmunoreactive for cytokeratins. However, it should be kept in mind that some translocation RCCs rarely expressed cytokeratins, similar to our case 3 tumor, which diffusely expressed AE1/AE3, EMA, and CK7. With respect to proteins under the control of the MiTF/TFE family, the positive rates for all proteins tested (HMB45, Melan-A, Cathepsin-K, and E-cadherin) were higher in TFEB RCCs than in TFE3 RCCs. The lower positive rate for Cathepsin-K in TFE3 RCCs than in TFEB RCCs can be explained by a recently reported study by Martignoni et al,[14] indicating differential expressions of Cathepsin-K in TFE3 RCC, depending on the specific fusion gene involved. All 8 *ASPSCR1-TFE3* translocation RCCs were negative for Cathepsin-K, whereas 12 of 14 *PRCC-TFE3* translocation RCCs expressed Cathepsin-K. Meanwhile, Argani et al[4] have reported a comprehensive IHC analysis of TFE3 RCC. They have extended the known IHC profile of TFE3 RCC using a panel of new biomarkers. For example, TFE3 RCC frequently expressed renal transcription factors PAX8 and PAX2, and the elevated expression of phosphorylated S6 in TFE3 RCCs suggested the mammalian target of rapamycin pathway as an attractive potential therapeutic target for these neoplasms. Also for TFEB RCC, similar comprehensive IHC analyses with additional new biomarkers are required.

So far, only 9 previous cases of TFEB RCC have been genetically analyzed. The reported *Alpha* fusion points have been scattered in the 1.2-kb bcr (break points: nt 1070,[11] 1419,[11] 1580,[2] 1631,[15] 1672,[2] 1760,[2] 1795,[2] 1810,[17] and 2274.[3]) The reported *TFEB* fusion points have been located in the 289-bp bcr, numbered from the 5′ end

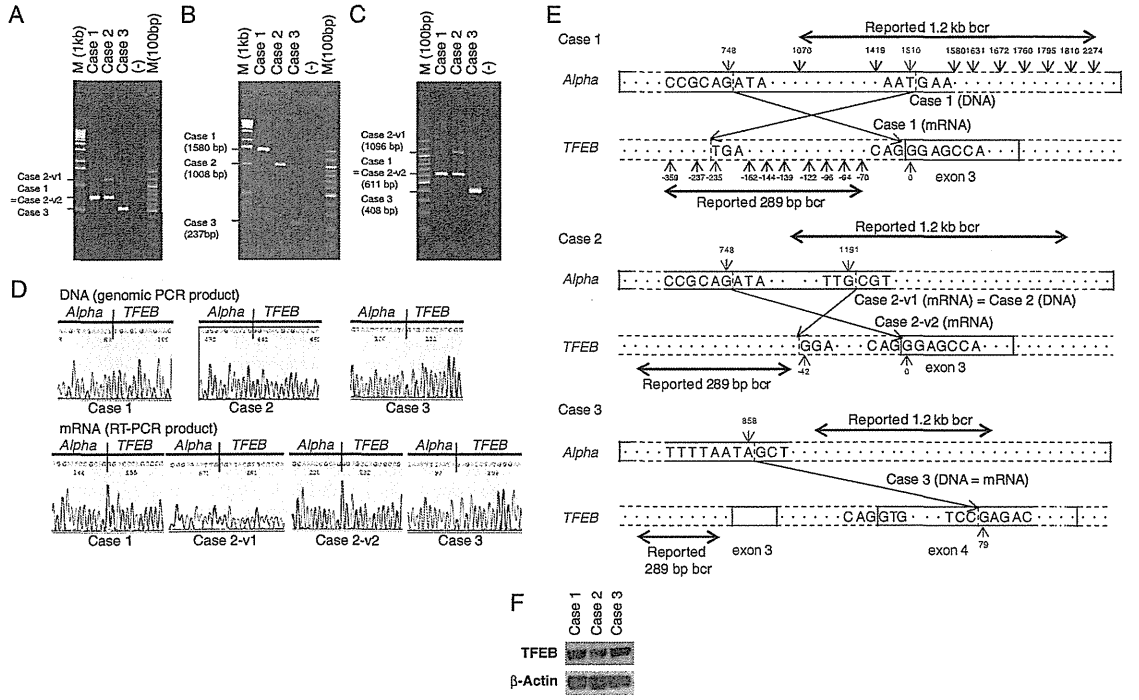


FIGURE 3. A, 5′-RACE amplification of the fusion point of the *Alpha-TFEB* fusion gene. M (1 kb), size markers (1 kb ladder); M (100 bp), size markers (100 bp ladder); Lane (–), no-template control. B, Genomic PCR amplification of the fusion point of the *Alpha-TFEB* fusion gene. C, RT-PCR amplification of the fusion point of the *Alpha-TFEB* mRNA. D, Nucleotide sequencing of each genomic PCR and RT-PCR product around each break point. E, Schematic representations of each fusion pattern. F, Western blot analysis showed a TFEB protein of almost the same size in the 3 cases.

of exon 3 of *TFEB* (break points: nt −359,[15] −237,[11] −162,[2] −144,[2] −139,[2] −122,[2] −96,[11] −94,[17] and −70[2]). All the break points of *TFEB* were upstream of *TFEB* exon 3, and the RT-PCR products were reported to yield only a single transcript. Both genomic PCR and RT-PCR analyses had been performed in only 3 previous cases,[2,11] all of which yielded equal fusion points at the DNA and mRNA levels. Although it was revealed that genomic PCR and RT-PCR products can be different, both of the product sizes were similar. It is easier to perform genomic PCR than RT-PCR for clinical samples because of the more labile nature of mRNA compared with DNA; therefore, the ability to diagnose TFEB RCC by genomic PCR provides a significant benefit with regard to the ease of clinical diagnosis and evaluation.

By examining our 3 new TFEB RCC cases genetically, we revealed that there are more diverse fusion patterns than previously thought. Case 1 had different fusion points between the genomic and transcript coordinates. Case 2 had 2 variants of mRNA. Case 3 had a fusion point of *TFEB* in exon 4, downstream of the wild-type ATG in exon 3; however, nuclear expression of the TFEB protein was detected, and the Western blotting analysis identified a protein similar in size to the wild-type TFEB protein. The diverse fusion patterns of TFEB RCCs must therefore be considered when designing PCR or RT-PCR primers to avoid false-negative results. As *Alpha* is an intronless gene that lacks splicing sites, transcripts of *Alpha-TFEB* may be modified after transcription of the gene, resulting in shorter nucleotides, as observed in case 1 and the case 2 variant 2.

Histologically, TFEB RCCs can present only a few characteristic features, which are difficult to recognize. Two of our 3 TFEB RCCs (cases 2 and 3) had scarce histologic features for translocation RCC and were initially misdiagnosed as other subtypes of RCC. TFEB RCC can also present a heterogeneous immunophenotype, similar to our case 3 tumor, which showed borderline features between TFEB RCC and chroRCC. Although the previously reported TFEB RCCs were predominantly seen in younger patients, all 3 of our patients with TFEB RCC were adults and included the oldest patient diagnosed so far. Therefore, this type of tumor should be included in differential diagnoses even in cases of adult RCC. Case 2 represents the third TFEB RCC case with an aggressive course, confirming the potential of TFEB RCC for a fatal outcome.

Owing to the possibility of TFEB RCCs with scarce characteristic histologic features, TFEB IHC, a powerful diagnostic tool revealed by Argani et al,[2] should be

performed for a more accurate diagnosis. This may be even more powerful when combined with the other IHC analyses referred to in Table 1, even if TFEB RCC is not strongly suspected. Further, the diverse fusion patterns of TFEB RCCs must be kept in mind when designing PCR or RT-PCR primers. More thorough examinations with our results in mind are likely to increase the number of TFEB RCC cases.

As TFEB RCC is exceedingly rare, the genetic and clinicopathologic characteristics have not been fully elucidated. We have herein examined 3 new TFEB RCC cases and revealed that TFEB RCCs have heterogeneous clinicopathologic features and more diverse fusion patterns than previously thought. Our study will contribute to the correct diagnosis of TFEB RCC and will likely increase the number of TFEB RCC cases. Further genetic and clinicopathologic investigations with additional new cases are required to clarify and establish the features of this tumor entity.

ACKNOWLEDGMENT

The authors thank Mr Motoyoshi Iwakoshi, Ms Miyuki Kogure, Mr Hiroshi Yokoo, and Ms Tomoyo Kakita for their valuable technical assistance. They also thank Ms Yuki Takano, Ms Yumiko Toriyama, and Ms Harumi Hodozuka for their secretarial work.

REFERENCES

1. Argani P, Hawkins A, Griffin CA, et al. A distinctive pediatric renal neoplasm characterized by epithelioid morphology, basement membrane production, focal HMB45 immunoreactivity, and t(6;11)(p21.1;q12) chromosome translocation. *Am J Pathol.* 2001;158:2089–2096.
2. Argani P, Lae M, Hutchinson B, et al. Renal carcinomas with the t(6;11)(p21;q12): clinicopathologic features and demonstration of the specific alpha-TFEB gene fusion by immunohistochemistry, RT-PCR, and DNA PCR. *Am J Surg Pathol.* 2005;29:230–240.
3. Argani P, Lae M, Ballard ET, et al. Translocation carcinomas of the kidney after chemotherapy in childhood. *J Clin Oncol.* 2006;24:1529–1534.
4. Argani P, Hicks J, De Marzo AM, et al. Xp11 translocation renal cell carcinoma (RCC): extended immunohistochemical profile emphasizing novel RCC markers. *Am J Surg Pathol.* 2010;34:1295–1303.
5. Camparo P, Vasiliu V, Molinie V, et al. Renal translocation carcinomas: clinicopathologic, immunohistochemical, and gene expression profiling analysis of 31 cases with a review of the literature. *Am J Surg Pathol.* 2008;32:656–670.
6. Davis IJ, Hsi BL, Arroyo JD, et al. Cloning of an Alpha-TFEB fusion in renal tumors harboring the t(6;11)(p21;q13) chromosome translocation. *Proc Natl Acad Sci U S A.* 2003;100:6051–6056.
7. Eble JN, Sauter G, Epstein JI, et al, ed. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs.* Lyon: IARC press; 2004.
8. Geller JI, Argani P, Adeniran A, et al. Translocation renal cell carcinoma: lack of negative impact due to lymph node spread. *Cancer.* 2008;112:1607–1616.
9. Hora M, Hes O, Urge T, et al. A distinctive translocation carcinoma of the kidney ["rosette-like forming," t(6;11), HMB45-positive renal tumor]. *Int Urol Nephrol.* 2009;41:553–557.
10. Komai Y, Fujiwara M, Fujii Y, et al. Adult Xp11 translocation renal cell carcinoma diagnosed by cytogenetics and immunohistochemistry. *Clin Cancer Res.* 2009;15:1170–1176.
11. Kuiper RP, Schepens M, Thijssen J, et al. Upregulation of the transcription factor TFEB in t(6;11)(p21;q13)-positive renal cell carcinomas due to promoter substitution. *Hum Mol Genet.* 2003;12:1661–1669.
12. Martignoni G, Tardanico R, Pea M, et al. t(6;11) renal cell tumor: a clinicopathological study of two cases in adults (abstr 715). *Mod Pathol.* 2005;18:155A.
13. Martignoni G, Pea M, Gobbo S, et al. Cathepsin-K immunoreactivity distinguishes MiTF/TFE family renal translocation carcinomas from other renal carcinomas. *Mod Pathol.* 2009;22:1016–1022.
14. Martignoni G, Gobbo S, Camparo P, et al. Differential expression of cathepsin K in neoplasms harboring TFE3 gene fusions. *Mod Pathol.* 2011. DOI: 10.1038/modpathol.2011.93.
15. Pecciarini L, Cangi MG, Lo Cunsolo C, et al. Characterization of t(6;11)(p21;q12) in a renal-cell carcinoma of an adult patient. *Genes Chromosomes Cancer.* 2007;46:419–426.
16. Suarez-Vilela D, Izquierdo-Garcia F, Mendez-Alvarez JR, et al. Renal Translocation Carcinoma With Expression of TFEB: Presentation of a Case With Distinctive Histological and Immunohistochemical Features. *Int J Surg Pathol.* 2009. DOI:10.1177/1066896909340531.
17. Zhan HQ, Wang CF, Zhu XZ, et al. Renal cell carcinoma with t(6;11) translocation: a patient case with a novel Alpha-TFEB fusion point. *J Clin Oncol.* 2010;28:e709–713.

Lung Cancer 75 (2012) 119–125

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

ASCL1-coexpression profiling but not single gene expression profiling defines lung adenocarcinomas of neuroendocrine nature with poor prognosis

Takeshi Fujiwara[a], Miyako Hiramatsu[b], Takayuki Isagawa[a], Hironori Ninomiya[b], Kentaro Inamura[b], Shumpei Ishikawa[a], Masaru Ushijima[c], Masaaki Matsuura[c], Michael H. Jones[a,d], Miyuki Shimane[a,d], Hitoshi Nomura[a,d], Yuichi Ishikawa[b,*], Hiroyuki Aburatani[a]

[a] *Genome Science Division, Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Tokyo, Japan*
[b] *Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan*
[c] *Bioinformatics Group, The Genome Center, Japanese Foundation for Cancer Research, Tokyo, Japan*
[d] *Chugai Pharmaceutical Co., Ltd., Kamakura Research Laboratories, Kamakura, Japan*

ARTICLE INFO

Article history:
Received 24 October 2010
Received in revised form 20 May 2011
Accepted 23 May 2011

Key words:
Lung
Adenocarcinoma
Neuroendocrine
Gene expression
Subtype
Prognosis

ABSTRACT

Background: Lung adenocarcinoma is heterogeneous regarding histology, etiology and prognosis. Although there have been several attempts to find a subgroup with poor prognosis, it is unclear whether or not adenocarcinoma with neuroendocrine (NE) nature has unfavorable prognosis.
Materials and methods: To elucidate whether a subtype of adenocarcinoma with NE nature has poor prognosis, we performed gene expression profiling by cDNA microarray for 262 Japanese lung cancer and 30 normal lung samples, including 171 adenocarcinomas, 56 squamous cell carcinomas and 35 NE tumors. A co-expression gene set with *ASCL1*, an NE master gene, was utilized to classify tumors by non-negative matrix factorization, followed by validation using an *ASCL1* knock-down gene set in DMS79 cells as well as an independent cohort ($n = 139$) derived from public microarray databases as a test set.
Results: The co-expression gene set classified the adenocarcinomas into alveolar cell (AL), squamoid, and NE subtypes. The NE subtype, which clustered together almost all the NE tumors, had significantly poorer prognosis than the AL subtype that clustered with normal lung samples ($p = 0.0075$). The knock-down gene set also classified the 171 adenocarcinomas into three subtypes and this NE subtype also had the poorest prognosis. The co-expression gene set classified the independent database-derived American cohort into two subtypes, with the NE subtype having poorer prognosis. None of the single NE gene expression was found to be linked to survival difference.
Conclusion: Co-expression gene set with *ASCL1*, rather than single NE gene expression, successfully identifies an NE subtype of lung adenocarcinoma with poor prognosis.

1. Introduction

Lung cancer is divided into small-cell and non-small cell carcinomas. Non-small cell carcinoma is further divided into three histological types: adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Adenocarcinoma is the most common histological type of lung cancer in most countries [1]. However, adenocarcinoma is very heterogeneous in terms of histology, etiology and prognosis. Some adenocarcinomas are related to smoking and have poor prognosis but others are not. There have been many attempts to further classify lung adenocarcinomas and to find a subtype with poor prognosis. Performing expression profiling by cDNA microarray, it was reported that a subtype of lung adenocarcinoma with high relative expression of neuroendocrine genes had poorer prognosis than others [2]. However, that finding was not verified by protein expression studies using immunohistochemistry [3], so it is still unclear that a subset of adenocarcinomas with neuroendocrine nature which can be defined have prognostic value.

Achaete-scute complex homolog 1 (Drosophila) (*ASCL1*) is a transcription factor belonging to the basic helix loop helix (BHLH) family [4,5]. This gene is a master gene for neuroendocrine differentiation and is expressed in fetal and adult neuroendocrine tissues as well as in neuroendocrine tumors such as thyroid medullary carcinomas and small cell lung carcinomas (SCLCs) [6,7]. Here we focused on this gene to characterize lung adenocarcinomas with neuroendocrine nature. We set out to identify a subtype of adenocarcinoma with neuroendocrine gene expression and to examine if such a subtype had poorer prognosis. We examined the possibility that in addition to expression of any single gene such as chromogranin A (*CHGA*) and synaptophysin (*SYP*), a co-expression gene set

* Corresponding author.
E-mail address: ishikawa@jfcr.or.jp (Y. Ishikawa).

0169-5002/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.lungcan.2011.05.028

was important for prognosis. A co-expression gene set was defined as genes that had similar expression patterns as a defined gene [8]. It has been shown in a number of studies that co-expression is correlated with functional relationships [9,10], suggesting that, since *ASCL1* is a master gene for neuroendocrine differentiation, an *ASCL1* co-expression gene set may classify lung adenocarcinomas into functional subtypes.

In this study, we attempted to classify 171 lung adenocarcinomas by single gene expression and by a co-expression gene set with regard to *CHGA*, *SYP* and *ASCL1*. For validation, we used two datasets: a gene set by *ASCL1* knock-down in DMS79 cells, a SCLC cell line highly expressing *ASCL1*, and an independent dataset of 139 adenocarcinoma patients derived from a public database. As a result, we successfully identified three subtypes of lung adenocarcinoma with survival difference by an *ASCL1* co-expression gene set. We also showed the co-expression gene set is robust for the independent cohort.

2. Materials and methods

2.1. Tumor samples and clinicopathological analysis

Clinical samples consisted of 262 Japanese lung cancers and 30 normal lungs, and were consecutively collected from tumors resected surgically at the Cancer Institute Hospital of Japanese Foundation for Cancer Research (JFCR) between 1994 and 2000 with informed consent. Histological diagnosis was made according to the WHO classification [1]. The lung cancers included 171 adenocarcinomas (81 females and 90 males, average age 62.7), 56 squamous cell carcinomas, 8 large cell neuroendocrine carcinomas (LCNEC), 15 SCLC, and 12 carcinoids. All the adenocarcinomas were staged pathologically based on the International Association of the Study of Lung Cancer manual, graded by differentiation degrees, and characterized by cell typing according to our previous analysis (i.e. columnar, hobnail and polygonal cell types) [11]. Cumulative smoking was measured by a smoking index, a product of cigarettes per day and duration in years. Survival was based on lung cancer-specific mortality, and was analyzed by the Kaplan–Meier method. Survival difference was analyzed by log-rank test. This study was approved by the institutional review boards of JFCR and RCAST.

2.2. Microarray

Total RNA was extracted from the samples. Expression levels were measured using a CHUGAI 41K (NCBI GEO GPL 962) microarray, with which we have published several studies on lung cancer [12–14]. cDNA derived from 2 μg cRNA was aminoallyl labeled with Cy5 (sample) or Cy3 (reference), hybridized overnight at 42 °C in a buffer containing 50% formamide, 5× SSC, 0.1% sodium dodecyl sulphate, 0.25 g/L human cot1 DNA, and 0.125 g/L poly dT and washed to a final stringency of 1 × SSC at 42 °C. After washing, slides were immediately scanned with an Axon GenePix 4000B scanner and quantified by Axon GenePix Pro software (Axon, Union City, CA, USA). All samples were subjected to LOcally WEighted Scatter plot Smoother (LOWESS); smoothing rate 20% [15]. Control signals were cut off at value 10. Datasets were normalized by their median. (GSE1037, GSE2088, GSE20853)

2.3. Grouping by single neuroendocrine gene expression

Based on expression of *CHGA*, *SYP* and *ASCL1*, the 171 adenocarcinomas were divided into the high- and low-expression groups, each having 86 and 85 cases, respectively. Survival was compared between the two groups for each gene.

2.4. Selection of neuroendocrine co-expression gene sets

A single probe, each for *CHGA*, *SYP* and *ASCL1* with the largest range of expression was selected. The top one hundred genes with the highest correlation to each gene were extracted from the expression data by Pearson correlation, and the *CHGA*-, *SYP*- and *ASCL1* co-expression gene sets were created.

2.5. Clustering by non-negative matrix factorization and consensus matrix

We employed non-negative matrix factorization (NMF) for clustering. NMF has several advantages, as compared with other clustering methods such as hierarchical clustering and self organizing maps. For example, they are more stable than self organizing map, they are independent of gene numbers, and consequently more robust than hierarchical clustering, and finally numbers of clusters are given by cophenetic correlation coefficients, so they are more objective [16]. Previously, using NMF, we successfully subdivided lung squamous cell carcinoma into two groups with statistically significant differences in survival rates [14]. To maximize NMF performance, the dataset was preprocessed as reported previously [17]. Briefly, the dataset was converted into log10 and row values were separated threshold zero. Negative values are transformed into absolute values. If all row values were null, we deleted the row because NMF could not be applied to non-negative matrix. NMF is performed on the MATLAB platform (The Math Works, Inc) [17]. We set iteration 30–50 for NMF and tested model-rank between 2 and 10.

2.6. Statistical analysis

Datasets were processed by using GeneSpring software (Agilent Technologies Inc). Statistical analysis was performed using JMP software (SAS Institute Inc.). Gene ontology was analyzed using EASE Software [18]. Significance level was set to be $p < 0.05$. For univariate and multivariate analyses for prognosis, we employed the Cox's proportional hazard model.

2.7. ASCL1 siRNA experiment

DMS79 cells, expressing highly *ASCL1*, were cultured in 5% CO_2 at 37.1 °C in RPMI-1640 medium supplemented with 10% (v/v) heat-inactivated fetal calf serum. Two siRNAs for targeting *ASCL1* (National Center for Biotechnology Information (NCBI) accession number: NM-004316) were designed and synthesized by Invitrogen Inc. DMS79 cells were transfected with 20 nM siRNA using Lipofectamine2000 (Invitrogen Inc.), according to manufacturer's instructions. After cells were cultured for 72 h, total RNA was extracted with the TRIzol (Invitrogen Inc.) according to the manufacturer's instruction. Gene expression profile was measured using a Human Genome U133 Plus 2.0 Array (Affymetrix, Inc.) (GSE21216).

2.8. Validation by other datasets

An independent web-derived dataset from a public database with a cohort of 139 American lung adenocarcinomas was employed for validation [2]. Probe sets for the *ASCL1* co-expression gene set were mapped to the GeneChip U95 array, then filtered by means of signals, values of which for at least one probe set were over 100 with SD > 20.

3. Results

3.1. Survival of subgroups by a single neuroendocrine gene expression

We first examined if there was any survival difference between two groups with different expression levels of a single NE gene. For *CHGA*, *SYP* and *ASCL1*, no difference was observed between two groups with high- and low-expression of each of the genes (data not shown).

3.2. Classification by neuroendocrine co-expression gene sets

The appropriate number of subtypes (k) was examined by neuroendocrine co-expression gene sets, using the NMF consensus matrix [16]. Briefly, co-expression gene sets were extracted from entire datasets. Classification of the dataset by the *ASCL1* co-expression gene set (Supplementary Table 1) was evaluated by the consensus matrix. The largest value of cophenetic correlation coefficient between $k=2$ and 10 was at $k=3$ (Figs. 1 and 2), implying that the $k=3$ subclassification model was most robust. Then survival rates of each adenocarcinoma subtype were examined and a significant difference was observed among subtypes. Meanwhile, the *CHGA*- and *SYP*-co-expression gene sets did not provide any significant difference among subtypes (data not shown).

3.3. Clinicopathological features of the three subtypes of lung adenocarcinoma

Since three-subtyping models were determined by the *ASCL1* co-expression gene set, clinicopathological features of each subtype were investigated. The first subtype clustered together with 30 normal lung samples, the second subtype clustered together with 42 squamous cell carcinomas samples, and the third subtype clustered with neuroendocrine tumors including 13 SCLC, 8 LCNEC and 12 carcinoid samples (33/35 NE tumors, 94%). Therefore, we named each subtype as alveolar cell (AL), squamoid (SQ) and neuroendocrine (NE) subtypes, respectively (Table 1).

We analyzed how each subtype was associated with clinical variables. The NE and SQ subtypes had significantly poorer prognosis than the AL subtype ($p=0.0075$) (Fig. 3). As shown in Table 2, the subtypes were associated with differentiation grading, cumulative smoking and sex. The proportion of poorly-differentiated tumors was highest in the NE subtype. The SQ subtype was characterized by high smoking index, as expected. Therefore, AL, SQ and NE subtypes had features predominantly similar to normal lungs, squamous cell carcinoma and NE tumors (including SCLC and LCNEC) respectively.

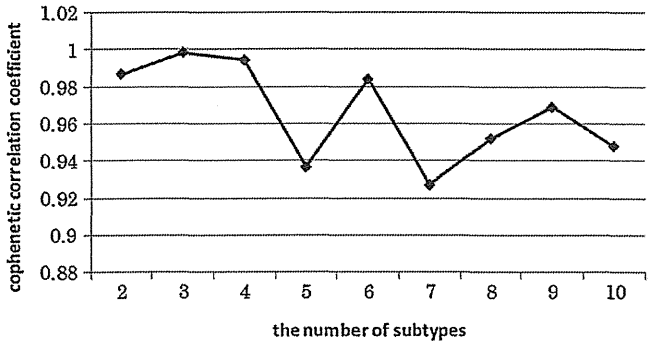


Fig. 1. Cophenetic correlation coefficients of the *ASCL1* co-expression gene set by non-negative matrix factorization (NMF). The data set was analyzed which number of subtypes was most robust from 2 to 10 subtypes by the NMF method. As shown by the figure, a three-subtype model was most robust. Vertical axis: the cophenetic correlation coefficient, horizontal axis: the subtype number.

Table 1
Subtypes of lung adenocarcinoma based on profiling by *ASCL1* co-expression gene set.

	Alveolar cell subtype (AL)	Squamoid subtype (SQ)	Neuroendocrine subtype (NE)
Adenocarcinoma	77	64	30
SCLC	0	2	13
LCNEC	0	0	8
Carcinoid	0	0	12
SqCC	1	42	13
Normal lung	30	0	0

SCLC, small cell lung carcinoma; LCNEC, large cell neuroendocrine carcinoma; SqCC, squamous cell carcinoma.

Regarding pTNM classification, the three subtypes had no statistically significant correlation with the TNM staging as shown in Table 2.

Univariate and multivariate analyses for prognosis were performed for subtypes, differentiation grades, smoking, stages, cell types, and TNM grades (Table 3). In the univariate analysis, the NE and SQ subtypes had significantly poor prognosis as compared with the AL subtype ($p=0.0047$ and 0.015, respectively). Also, moderate and poor differentiation was associated with prognosis as compared with well differentiated tumors ($p<0.002$). Using factors shown to be significant in the univariate analysis, we performed multivariate analyses. Although the first round multivariate analysis using many variables didn't demonstrate significant results for NE and SQ subtypes, the second round multivariate analysis using two subtype factors (NE and SQ as compared with AL subtype) and the pN factor, conferred the significant association of the NE and SQ subtypes with prognosis ($p=0.027$ and 0.014).

3.4. Functional analysis of the ASCL1 gene set

We analyzed gene ontology to identify any functional relevance of the *ASCL1* co-expression gene set. Unique gene symbols were extracted and enrichments of gene ontology (Biological Process) were statistically analyzed (EASE Score was 0.1 or less). As a result (Table 4), 14 ontological categories were defined, encompassing 10 cell cycle categories and 4 neuronal development categories. In contrast, categories of the *ASCL1* knock-down gene set contained pathways related to neuropeptide signaling, G-protein coupled receptor protein signaling and regulation of synapse. Taken together, the *ASCL1* co-expression gene set was characterized by the cell cycle ontological term.

3.5. Validation by siRNA experiment

siRNA against *ASCL1* was transfected into DMS79 cells. After confirmation of down regulation of *ASCL1* expression, a gene set including 91 probes down-regulated compared to control cells, was selected (Supplementary Table 2). The knock-down gene set may include genes directly regulated by *ASCL1* while the co-expression gene set may include genes with cellular functions related to *ASCL1*. The 171 adenocarcinomas were classified by the *ASCL1* knock-down gene set using NMF. The maximum cophenetic correlation coefficient was also 3. Subsequently, survival analysis was carried out for the three subtypes. As expected, they had a significant survival difference ($p=0.0007$). Each subtype was similar to the AL, SQ or NE subtype by the *ASCL1* co-expression gene set (Data not shown).

3.6. Validation by other datasets

To confirm whether our method could be effectively utilized against datasets from other institutions, the *ASCL1* co-expression gene set was applied to a web-derived cohort of 139 American

Table 2
Clinicopathological characteristics of each subtype identified by NMF using the *ASCL1* co-expression gene set.

		Subtype			*p*-Value
		AL subtype	SQ subtype	NE subtype	
Stage	1	47	31	12	0.1345
	2	8	6	1	
	3	16	18	14	
	4	5	6	1	
	NA	1	3	2	
pT category	pT1	36	23	13	0.1384
	pT2	28	24	6	
	pT3	0	4	2	
	pT4	12	10	8	
	NA	1	3	1	
pN category	pN0	47	40	13	0.1826
	pN1	14	11	3	
	pN2	13	9	10	
	pN3	1	0	1	
	NA	2	4	3	
pM category	pM0	70	55	26	0.9208
	pM1	6	5	3	
	NA	1	4	1	
Differentiation	well	44	19	8	0.0034
	moderate	22	34	15	
	poor	10	9	7	
	NA	1	1	0	
Smoking index	(median)	0	566	290	0.0003
	(average)	275	636	546	0.0011
	NA	0	0	0	
Cell type	Columnar	8	10	7	0.0881
	Hobnail-Clara	52	34	13	
	Polygonal	7	12	6	
	NA	10	8	4	

NA: not available

Table 3
Univariate and multivariable analyses of subtypes, differentiation grades and smoking as prognostic factors.

Univariate	*n*	Hazard	95% Confidence interval		
			Low	High	*p*-Value
ASCL1 subtype NE/AL	105	2.94	1.41	6.09	0.0047
ASCL1 subtype SQ/AL	138	2.15	1.16	4.10	0.015
Differentiation moderate/well	141	2.77	1.47	5.52	0.0014
Differentiation poor/well	125	4.12	1.88	8.99	0.0005
Smoking index > 400	167	1.00	0.58	1.69	0.99
Stage 2/1	105	4.68	1.74	11.68	0.0034
Stage 3/1	138	7.39	3.82	15.19	<0.0001
Stage 4/1	101	10.10	3.72	25.38	<0.0001
Polygonal/columnar	50	2.29	1.00	5.52	0.050
Hobnail clara/columnar	123	0.72	0.35	1.63	0.41
≥pT3	165	4.36	2.49	7.51	< 0.0001
≥pN1	161	5.30	2.97	9.86	< 0.0001
≥pM1	164	1.92	0.79	3.97	0.14

Multivariate	*n* = 142	Hazard	95% Confidence interval		
			Low	High	*p*-Value
ASCL1 subtype NE/AL		1.77	0.73	4.26	0.20
ASCL1 subtype SQ/AL		1.77	0.85	3.77	0.13
Differentiation moderate/well		0.98	0.44	2.26	0.96
Differentiation poor/well		0.79	0.22	2.83	0.72
Stage 2/1		0.87	0.18	3.69	0.86
Stage 3/1		1.22	0.30	4.45	0.78
Stage 4/1		1.67	0.31	7.15	0.53
Polygonal/columnar		3.26	1.16	9.45	0.025
Hobnail clara/columnar		0.84	0.37	2.06	0.69
≥pT3		2.53	1.12	6.25	0.025
≥pN1		3.25	1.21	10.26	0.018

Multivariate analysis	*n* = 161	Hazard	95% Confidence interval		
			Low	High	*p*-Value
ASCL1 subtype NE/AL		2.40	1.11	5.07	0.027
ASCL1 subtype SQ/AL		2.21	1.17	4.26	0.014
≥pN1		5.37	3.01	10.03	< 0.0001

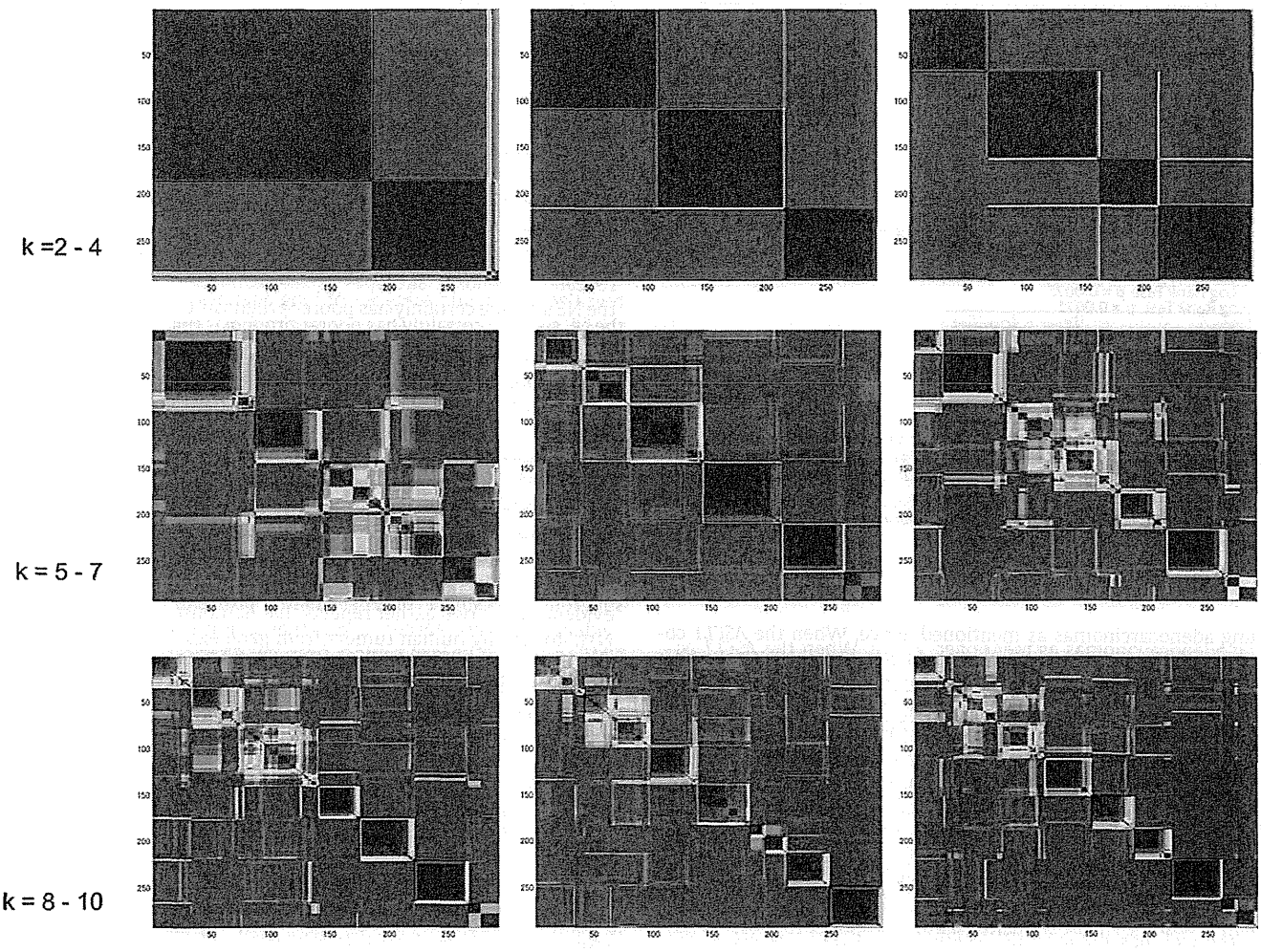
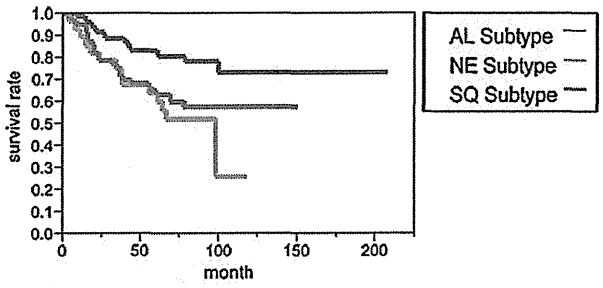


Fig. 2. The consensus matrices of the *ASCL1* co-expression gene set for 2–10 subtypes shown at the time of classifying a data set according to the NMF method. Vertical and horizontal axes are numbers of samples. The colors demonstrate a level of consensus; from blue = 0 (a sample does not go into the same cluster at all by all the trial of NMF) to red = 1 (a sample is classified into the same cluster according to all the trial of NMF) Hierarchical clustering is carried out in both directions.

Table 4
ASCL1 gene set gene ontology enrichment by EASE.

Gene category	List hits	List total	Population hits	Population total	EASE score	Gene symbol
Mitotc cell cycle	7	46	266	8173	0.00	CDK5R1; CDKN2D; CENPF; DDX11; GTSE1; LOC81691; TOP2A
Neurogenesis	7	46	327	8173	0.01	ASCL1; CACNAIA; CDK5R1; FOXG1B; LHX2; MSI1; PCP4
G2/M transition of mitotic cell cycle	3	46	40	8173	0.02	CDK5R1; CDKN2D; DDX11
Morphogenesis	10	46	774	8173	0.02	ASCL1; CACNA1A; CDK5R1; EVC; FOXG1B; HOXB5; LHX2; MSI1
Cell cycle	8	46	550	8173	0.03	CDK5R1; CDKN2C; CDKN2D; CENPF; DDX11; GTSE1; LOC81691; TOP2A
Regulation of cell cycle	6	46	322	8173	0.03	CDK5R1; CDKN2C; CDKN2D; CENPF; DDX11; GTSE1
Development	13	46	1281	8173	0.04	ASCL1; CACNA1A; CDK5R1; CRLF3; EVC; FOXG1B; HOXB5; HOXC5;
DNA replication	4	46	143	8173	0.04	CENPF; DDX11; LOC81691; TOP2A
Cell proliferation	9	46	801	8173	0.07	CDK5RI; CDKN2C; CDKN2D; CENPF; DDX11; GTSE1; LOC81691; PSIP1
Nucleobase, nucleoli	17	46	2058	8173	0.08	ASCL1; CACNA1A; DDX11DDX12; ELAVL4; FOXG1B; FOXN4; HMGN2; HOXB5
Central nervous system	3	46	86	8173	0.08	CDK5R1; FOXG1B; PCP4
Organogenesis	8	46	691	8173	0.08	ASCL1; CACNA1A; CDK5R1; EVC; FOXG1B; LHX2; MSI1; PCP4
Regulation of transcription	12	46	1313	8173	0.10	ASCL1; CACNA1A; FOXG1B; FOXN4; HMGN2; HOXB5; HOXC5
Nucleotide excision repair	2	46	19	8173	0.10	DDX11; DDX12



Log Rank Test p = 0.0075

Subtype	Dead	Censored
Alveolar Cell Subtype	16	60
Squamoid Subtype	25	37
Neuroendocrine Subtype	14	15

Fig. 3. Survival curves of lung adenocarcinoma subtypes based on the *ASCL1* co-expression gene set model. The adenocarcinoma data set was classified according to the NMF method by the *ASCL1* co-expression gene set, and performed Kaplan–Meier survival analysis for each subtype. Note that the neuroendocrine (NE) and squamoid (SQ) subtypes had significantly unfavorable prognosis as compared with the alveolar cell (AL) subtype.

lung adenocarcinomas as mentioned above. When the *ASCL1* co-expression gene set was mapped to the data set of UNIGENE, a set of 71 probes was obtained (Supplementary Table 1). Four genes are overlapped in two gene sets (Supplementary Fig. 1). Next, clustering was carried out using NMF with this probe set. The maximum cophenetic correlation coefficient was two, implying the most robust number of subtypes was two. When survival analysis was performed according to the two subtypes model using 125 cases for which survival data were available, the first group ($n = 11$) had significantly poorer prognosis than the second ($n = 114$) ($p = 0.0098$). This group with poor prognosis included all the cases belonging to the C2 group with neuroendocrine nature and poor prognosis ($n = 9$) in the web-based cohort [2]. Therefore, our *ASCL1* co-expression gene set was proven robust.

4. Discussion

In this study, we sought to classify lung adenocarcinomas to prognostic subtypes by NE genes and their co-expression genes. First, we tried to classify them with individual gene expression of *ASCL1*, *CHGA* or *SYP*. However, expression of each NE gene did not correlate with survival. Next, we used a co-expression gene set of *ASCL1*, *CHGA* or *SYP* and only *ASCL1* co-expression gene set delineated AL, SQ and NE subtypes with different prognosis. We demonstrated that the NE subtype had poorer prognosis than the AL subtype, based on co-expression gene sets, but not when based on single genes. By using the *ASCL1* co-expression (i.e. neuroendocrine-related) gene set we obtained not only the NE subtype but the AL and SQ subtypes. Although the reason why such subtypes unrelated to neuroendocrine nature were obtained is unclear, it is interesting that we got only two subtypes of adenocarcinoma when we applied the gene set to the web-based American cohort ($n = 125$), as described above. Adenocarcinoma subtypes may depend upon causation including ethnicity. For validation, we created an *ASCL1* knock-down gene set in DMS79 cells and showed that the knock-down gene set also produced three subgroups from the 171 adenocarcinoma and that there were significant survival differences. For additional validation, we tried to classify an independent cohort ($n = 139$) of adenocarcinoma derived from a public database and confirmed the *ASCL1* co-expression gene set was robust because the gene set successfully produced a NE subgroup with poorer prognosis. Although a multi-site blinded validation study was recently performed [19], we believe that further studies should be carried out for more detailed stratification and outcome prediction of lung cancer patients using a wide range of populations such as ours, which encompassed the largest number of cases from a single institution.

Whether lung adenocarcinoma with NE differentiation has poorer outcome has been contentious. Previous detailed pathological studies based on immunohistochemistry for NE markers such as chromogranin A, synaptophysin and CD56 did not demonstrate any significant survival difference [3]. However, our study clearly revealed prognostic subtypes of lung adenocarcinoma, indicating the NE subtype certainly has poorer prognosis than the AL subtype. To detect subtypes with poor prognosis, we should bear in mind the two issues: the insufficiency of focusing on a single NE gene, and the importance of *ASCL1*, the NE master gene. Further studies on protein expression are needed, which should include an ASLC1 antibody as well as choose antibodies for multiple NE genes.

For clustering analysis we employed an NMF, a statistical method which was initially developed to recognize image patterns, for example, in parts of a face [20]. Compared with other clustering methods, NMF has several advantages as mentioned above. In particular, it is possible to determine the objective number of subgroups by using a consensus matrix [16]. Our findings support evidence from two earlier reports that demonstrated the utility of NMF to classify human tumors from gene expression profile data [16,21]. In fact, we also succeeded in classifying squamous cell carcinoma of the lung to subgroups with different clinical outcome [14]. Significantly, by using NMF again, we were able to classify the adenocarcinomas to subgroups with statistically significant differences in survival.

In our analysis of gene ontology, we found that *ASCL1* co-expression gene set is enriched in ontology related to "cell cycle". This fact is consistent with the former study, where unsupervised clustering of adenocarcinomas produced three subtypes (bronchoid, squamoid and magnoid) and proliferation markers were predominant in the magnoid subtype [22]. The gene that appeared most frequently in the ontology related to "cell cycle" is cyclin-dependent kinase 5 regulatory subunit 1 (p35) (*CDK5R1*). *CDK5R1* is active in the nervous system, essential in neuronal development, and may participate in the onset of a neurodegenerative disease [23]. The gene, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11(CHL1-like helicase homolog, S. cerevisiae) (*DDX11*) appeared in 8 gene ontology sets. DDX11 is a RNA helicase family protein with the conserved motif Asp-Glu-Ala-Asp (DEAD) and is required for sister chromatid adhesion or cell division [24]. Forkhead box G1 (*FOXG1B*), present in all four gene ontology sets, is a transcription factor with a forkhead domain and relevant to nerve system development [25].

Stratification of the lung adenocarcinoma patients into tumor subtypes with different prognosis should be crucial in selecting the best treatment in future. Furthermore, such subtype information needs to be integrated in ongoing large-scale cancer genome sequencing projects, because each subtype may have an unique repertoire of somatically mutated cancer genes [26].

5. Conclusion

We evaluated whether neuroendocrine co-expression gene sets classify lung adenocarcinoma into significant subtypes. By using NMF, we identified three subtypes of lung adenocarcinoma; AL-, SQ- and NE-subtypes, which have alveolar cell, squamoid, and neuroendocrine cell features respectively. Analysis of survival for patients showed the NE subtype had significantly poorer prognosis than the AL subtype.

Conflict of interest

Nothing declared.

Acknowledgements

The authors wish to acknowledge Dr. Sakae Okumura for providing clinical samples and information and Dr. Yujin Hoshida for his kind advice for analysis of microarray data. This study was supported by Grants-in-Aids for Scientific Research, includilng (S) 20221009, from the Ministry of Education, Culture, Sports, Science and Technology, and by grants from the Ministry of Health, Labour and Welfare; the Smoking Research Foundation; the National Institute of Biomedical Innovation; the Princess Takamatsu Cancer Research Fund; and Foundation for Promotion of Cancer Research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.lungcan.2011.05.028.

References

[1] Travis WD, Brambilla E, Müller-Hermelink HK, et al. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart. Lyon: IARC; 2004.
[2] 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 USA 2001;98:13790–5.
[3] Ionescu DN, Treaba D, Gilks CB, Leung S, Renouf D, Laskin J, et al. Nonsmall cell lung carcinoma with neuroendocrine differentiation—an entity of no clinical or prognostic significance. Am J Surg Pathol 2007;31:26–32.
[4] Jiang SX, Kameya T, Asamura H, Umezawa A, Sato Y, Shinada J, et al. hASH1 expression is closely correlated with endocrine phenotype and differentiation extent in pulmonary neuroendocrine tumors. Mod Pathol 2004;17:222–9.
[5] Jiang T, Collins BJ, Jin N, Watkins DN, Brock MV, Matsui W, et al. Achaete-scute complex homologue 1 regulates tumor-initiating capacity in human small cell lung cancer. Cancer Res 2009;69:845–54.
[6] Sippel RS, Carpenter JE, Kunnimalaiyaan M, Chen H. The role of human achaete-scute homolog-1 in medullary thyroid cancer cells. Surgery 2003;134:866–71.
[7] Borges M, Linnoila RI, van de Velde HJ, Chen H, Nelkin BD, Mabry M, et al. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature 1997;386:852–5.
[8] Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. Genome Res 2004;14:1085–94.
[9] 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.
[10] Ge H, Liu Z, Church GM, Vidal M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet 2001;29:482–6.
[11] Hashimoto T, Tokuchi Y, Hayashi M, Kobayashi Y, Nishida K, Hayashi S, et al. Different subtypes of human lung adenocarcinoma caused by different etiological factors. Am J Pathol 2000;157:2133–41.
[12] Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, Miyoshi M, et al. Integrated classification of lung tumors and cell lines by expression profiling. Proc Natl Acad Sci USA 2002;99:12357–62.
[13] Jones MH, Virtanen C, Honjoh D, Miyoshi T, Satoh Y, Okumura S, et al. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. Lancet 2004;363:775–81.
[14] Inamura K, Fujiwara T, Hoshida Y, Isagawa T, Jones MH, Virtanen C, et al. Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. Oncogene 2005;24:7105–13.
[15] Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.
[16] Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USA 2004;101:4164–9.
[17] Kim PM, Tidor B, Keating AE. Subsystem identification through dimensionality reduction of large-scale gene expression data. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. Genome Res 2003;13:1706–18.
[18] Hosack DA, Dennis Jr G, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol 2003;4:R70.
[19] Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 2008;14(8):822–7.
[20] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401:788–91.
[21] Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69(18):7385–92.
[22] Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol 2006;24(31):5079–90.
[23] Shelton SB, Johnson GV. Cyclin-dependent kinase-5 in neurodegeneration. J Neurochem 2004;88:1313–26.
[24] Parish JL, Rosa J, Wang X, Lahti JM, Doxsey SJ, Androphy EJ. The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. J Cell Sci 2006;119:4857–65.
[25] Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, et al. FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet 2008;83:89–93.
[26] International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International Cancer Genome Consortium International network of cancer genome projects. Nature 2010;464(7291):993–8.

nature medicine

RET, ROS1 and ALK fusions in lung cancer

Kengo Takeuchi[1,2], Manabu Soda[3], Yuki Togashi[1,2], Ritsuro Suzuki[4], Seiji Sakata[1], Satoko Hatano[1], Reimi Asaka[1,2], Wakako Hamanaka[2], Hironori Ninomiya[2], Hirofumi Uehara[5], Young Lim Choi[6], Yukitoshi Satoh[5,7], Sakae Okumura[5], Ken Nakagawa[5], Hiroyuki Mano[3,6] & Yuichi Ishikawa[2]

Through an integrated molecular- and histopathology-based screening system, we performed a screening for fusions of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1, receptor tyrosine kinase (ROS1) in 1,529 lung cancers and identified 44 ALK-fusion–positive and 13 ROS1-fusion–positive adenocarcinomas, including for unidentified fusion partners for ROS1. In addition, we discovered previously unidentified kinase fusions that may be promising for molecular-targeted therapy, kinesin family member 5B (KIF5B)-ret proto-oncogene (RET) and coiled-coil domain containing 6 (CCDC6)-RET, in 14 adenocarcinomas. A multivariate analysis of 1,116 adenocarcinomas containing these 71 kinase-fusion–positive adenocarcinomas identified four independent factors that are indicators of poor prognosis: age ≥50 years, male sex, high pathological stage and negative kinase-fusion status.

Echinoderm microtubule associated protein like 4 (EML4)-ALK was the first targetable fusion oncokinase to be identified in non–small cell lung cancer (NSCLC)[1]. This fusion is found in approximately 4–6% of lung adenocarcinomas[2,3]. ROS1 is another receptor tyrosine kinase that forms fusions in NSCLC[4]. Solute carrier family 34 (sodium phosphate), member 2 (SLC34A2)-ROS1 and CD74 molecule, major histocompatibility complex, class II invariant chain (CD74)-ROS1 were identified in 1 out of 41 NSCLC cell lines and 1 out of 150 lung cancer samples, respectively[4]. However, the oncogenic ability of these ROS1 fusion proteins and the incidence of ROS1 fusions in lung cancers are still unclear.

We screened for known and unknown kinase fusions in lung cancers using a histopathology-based system with tissue microarrays of 1,528 surgically removed tissues (**Supplementary Methods** and **Supplementary Appendix**). Immunohistochemistry of antibodies to ALK using the intercalated antibody-enhanced polymer method[2,3,5–7] detected 45 tumors with ALK kinase domain expression (**Supplementary Fig. 1**). In 44 adenocarcinomas, multiplex RT-PCR[2,3] identified 41 *EML4-ALK*–positive and 3 *KIF5B-ALK*–positive adenocarcinomas, including a previously unidentified KIF5B-ALK fusion variant, K17;A20 (**Supplementary Table 1**). Further, we used fluorescence *in situ* hybridization (FISH) for split and fusion assays to confirm the presence of ALK fusions[2,3,8]. The FISH results for the *ALK* split assay, the *EML4-ALK* fusion assay and the *KIF5B-ALK* fusion assay in the 44 adenocarcinomas were all consistent with the presence of the corresponding fusion gene (**Supplementary Figs. 2** and **3**). The remaining tumor that was positive for antibodies to ALK as determined by immunohistochemistry (a large-cell neuroendocrine carcinoma) was negative in the FISH assays and expressed wild-type ALK. ALK fusions existed in 3.0% (44 out of 1,485) of the NSCLCs and 3.9% (44 out of 1,121) of the adenocarcinomas. We included 20 previously reported ALK-fusion–positive and 304 ALK-fusion–negative tumors, all of which were screened with multiplex RT-PCR. Because specimens of these 324 patients were collected consecutively during the period of tissue collection, they served as positive and negative controls, respectively[1–3,8,9]. The immunohistochemistry results using the intercalated antibody-enhanced polymer method were complete matches in the 20 fusion-positive and the 304 fusion-negative tumors.

We used split FISH assays for the screening for *ROS1* gene rearrangement (**Fig. 1**). In 11 of the 13 *ROS1* split FISH-positive tumors (**Fig. 1a**), 5′ rapid amplification of complementary DNA ends (5′ RACE) identified two known and three unknown fusion partners for *ROS1*: *TPM3*, *SDC4*, *SLC34A2*, *CD74* and *EZR* (**Fig. 1b** and **Supplementary Table 1**); RT-PCR confirmed this finding (**Fig. 1c**). In a 5′-RACE–negative tumor (ROS#12) (again, where split FISH is used to detect candidate fusion genes of interest by the presence of rearrangements and RACE is used for the identification of fusion partners), each fusion-specific RT-PCR (using a common reverse primer) amplified the same band, which contained an *LRIG3* sequence. This tumor was proven fusion-positive in RT-PCR specific to *LRIG3-ROS1*, an unidentified fusion. Fusion FISH results confirmed that all 12 cases harbored the corresponding fusion (**Fig. 1a**). All fusion FISH assays for these six ROS1 fusions were negative for the tumor ROS#13 (the frozen material had been consumed), indicating an unknown fusion partner for *ROS1*. *ROS1* split FISH screening failed for nine NSCLCs, including five adenocarcinomas. We identified ROS1 fusions in 0.9% (13 out of 1,476) of the NSCLCs and 1.2% (13 out of 1,116) of the adenocarcinomas.

We performed *KIF5B* split FISH to discover new fusion kinases, as we previously identified KIF5B-ALK fusions in lung cancer[3]. As such, we hypothesized that KIF5B might be rearranged in lung cancer. In 24 *KIF5B* split FISH-positive tumors, 3′ RACE identified an in-frame fusion between *KIF5B* exon 23 and *RET* exon 12

[1]Pathology Project for Molecular Targets, the Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan. [2]Division of Pathology, the Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan. [3]Division of Functional Genomics, Jichi Medical University, Tochigi, Japan. [4]Department of Hematopoietic Stem Cell Transplantation Data Management and Biostatistics, Nagoya University Graduate School of Medicine, Nagoya, Japan. [5]Department of Thoracic Surgical Oncology, Thoracic Center, the Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan. [6]Department of Medical Genomics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [7]Present address: Department of Thoracic Surgery, Kitasato University School of Medicine, Kanagawa, Japan. Correspondence should be addressed to K.T. (kentakeuchi-tky@umin.net).

Received 28 September 2011; accepted 3 January 2012; published online 12 February 2012; doi:10.1038/nm.2658

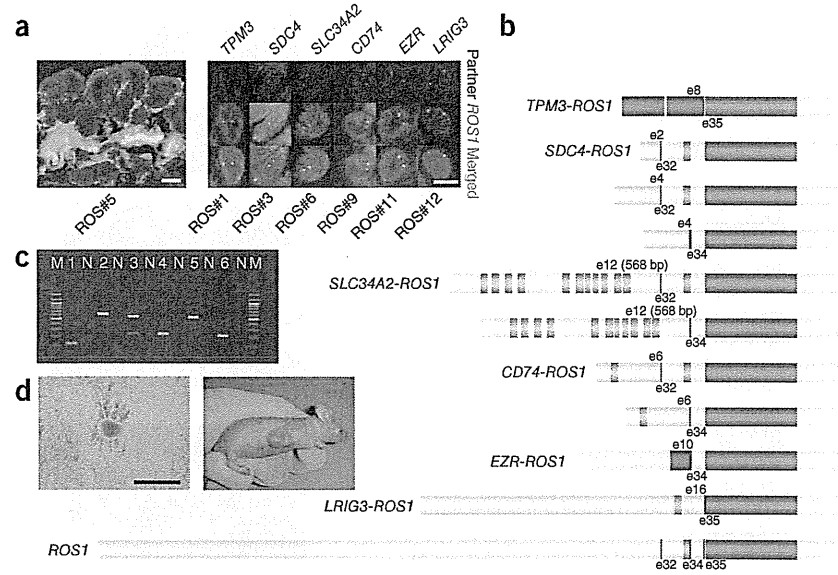


Figure 1 Identification of ROS1 fusions. (**a**) *ROS1* split (left) and fusion (right) FISH assay data (scale bars, 20 μm). In the split assay, multiple tumor cells harbored individual 3′ side signals (green), indicating the presence of a *ROS1* rearrangement. In the fusion assay, a fusion signal (yellow) was observed in the representative tumor cell of each subject, which is consistent with the presence of t(1;6)(q21.2;q22) for *TPM3-ROS1*, t(6;20)(q22;q12) for *SDC4-ROS1*, t(4;6)(q15.2;q22) for *SLC34A2-ROS1*, t(5;6)(q32;q22) for *CD74-ROS1*, inv(6)(q22q25.3) for *EZR-ROS1* or t(6;12)(q22;q14.1) for *LRIG3-ROS1*. (**b**) The break points of ROS1 are exons 32, 34 and 35. All of the break points allow the resulting fusion to harbor the kinase domain of ROS1 (red), and the exon 32 break point allows the resulting fusion to harbor the transmembrane domain of ROS1 (orange). In the fusion partners, dark blue and orange represent coiled-coil and transmembrane domains, respectively. Coiled-coil domains may contribute to homodimelization, but only TPM3 and EZR contained these domains. In contrast to ALK and RET fusions, the role of the fusion partner's coiled-coil domain is unknown in ROS1 fusions. e, exon. (**c**) Results for fusion-specific RT-PCR for tumors ROS#1 (lane 1, TPM3-ROS1, T8;R35, predicted product size of 119 bp), ROS#3 (lane 2, SDC4-ROS1, S2;R32, 596 bp), ROS#6 (lane 3, SLC34A2-ROS1, S13del2046;R32 and S13del2046;R34, 544 bp and 235 bp, respectively), ROS#8 (lane 4, CD74-ROS1, C6;R34, 230 bp), ROS#10 (lane 5, EZR-ROS1, E10;R34, 527 bp), and ROS#12 (lane 6, LRIG3-ROS1, L16;R35, 218 bp). M and N represent the size marker (100-bp ladder) and the non-template control, respectively. (**d**) The transforming potential of the ROS1 fusion. Mouse 3T3 fibroblasts infected with a retrovirus encoding SDC4-ROS1 derived from tumor ROS#4 formed multiple foci (scale bar, 1 mm). All of the four nude mice injected with the corresponding 3T3 cells developed a subcutaneous tumor (right).

(tumor RET#11). *RET* split FISH on the tissue arrays identified 22 fusion-positive tumors in 1,528 lung cancers (**Fig. 2a**), from which a multiplex RT-PCR system that captures all possible *KIF5B-RET* fusions detected 12 fusion-positive tumors: eight tumors with the fusion of *KIF5B* exon 15 and *RET* exon 12 (K15;R12) and one tumor each with the K16;R12, K22;R12, K23;R12 and K24;R11 fusions (**Fig. 2b** and **Supplementary Table 1**). The *KIF5B-RET* fusion FISH results were consistent with the presence of inv(10)(p11.22q11.2) in all 12 of these tumors (**Fig. 2a**).

In a routine histopathological diagnosis, we encountered an adenocarcinoma that showed a mucinous cribriform pattern (**Fig. 2c**) that was previously reported as a histopathological marker for the presence of EML4-ALK (**Supplementary Fig. 4**)[9–11]. Notably, this adenocarcinoma (tumor RET#14) was negative for ALK fusion and was positive for *CCDC6-RET*, as determined by FISH and inverse RT-PCR; the latter fusion gene was first described in thyroid cancer[12]. RT-PCR identified another tumor positive for the *CCDC6-RET* fusion (RET#13) in the remaining 10 tumors. The 14 RET-positive tumors (out of the total 1,528 tumors tested, with one additional tumor (RET#14) found through routine pathology diagnostic service) were also positive in the revised multiplex RT-PCR that captured *EML4-ALK*, *KIF5B-ALK*, *KIF5B-RET* and *CCDC6-RET* simultaneously (**Fig. 2d**). The RET kinase domain expression using real-time RT-PCR was weak or undetectable for the remaining nine tumors determined to be positive in the *RET* split FISH screening. Perhaps the genomic rearrangement occurred downstream of the *RET* break points. *RET* split FISH screening failed in three NSCLCs, including two adenocarcinomas. RET#14 was the index case found in routine pathology diagnostic service but not in the 1,528 cohort. RET fusions existed in 0.9% (13 out of 1,482) of the NSCLCs and 1.2% (13 out of 1,119) of the adenocarcinomas. The 14 RET fusion-positive subjects did not receive vandetanib.

We concluded that the rearrangements described above are somatic without using any matched normal tissues. Our histopathology-based screening method preserves the samples' histological architecture. This allows observers to confirm that internal non-tumor cells, for example, epithelial cells, inflammatory cells or fibroblasts, are negative in a test of interest.

All 71 kinase-fusion–positive (44 ALK, 13 ROS1 and 14 RET fusions) lung cancers were exclusively adenocarcinomas (6% of all adenocarcinomas in the present study), were positive for antibodies to TTF1, which is regarded as a marker for lung adenocarcinoma, as determined by immunohistochemistry (excluding two ALK-positive tumors) and were negative for *EGFR* and *KRAS* mutations. Thirteen of the 44 ALK-positive tumors (30%) were weakly positive for p63 expression (were weakly positive for a squamous cell carcinoma marker, p63) (**Supplementary Table 1**). Thirty-three tumors showed a mucinous cribriform pattern in at least 5% of their area; 22 tumors had this pattern in >25% of their area (**Fig. 2c**, **Supplementary Table 1** and **Supplementary Fig. 4**). The frequency of mucinous cribriform carcinoma was significantly higher in the kinase-fusion–positive group of tumors than in the 77 fusion-negative adenocarcinomas (22 out of 71 compared to 7 out of 77, respectively; $P = 0.00088$). Notably, we observed this pattern preferentially in EML4-ALK–positive tumors (70%, 29 out of 41); all three CD74-ROS1–positive tumors also showed this pattern. Recognizing this pattern in routine pathology diagnoses led to the identification of the CCDC6-RET fusion (tumor RET#14). In organs other than the lung, secretory breast carcinoma, which is characterized by a cribriform pattern with abundant secretory material, harbors the ets variant 6 (ETV6)-neurotrophic tyrosine kinase, receptor, type 3 (NTRK3) fusion (ref. 13). We identified an ALK-fusion–positive renal cell carcinoma that showed a mucinous cribriform pattern[7]. This pattern may be linked to the presence of particular kinase fusions[10], and this possibility warrants further study.

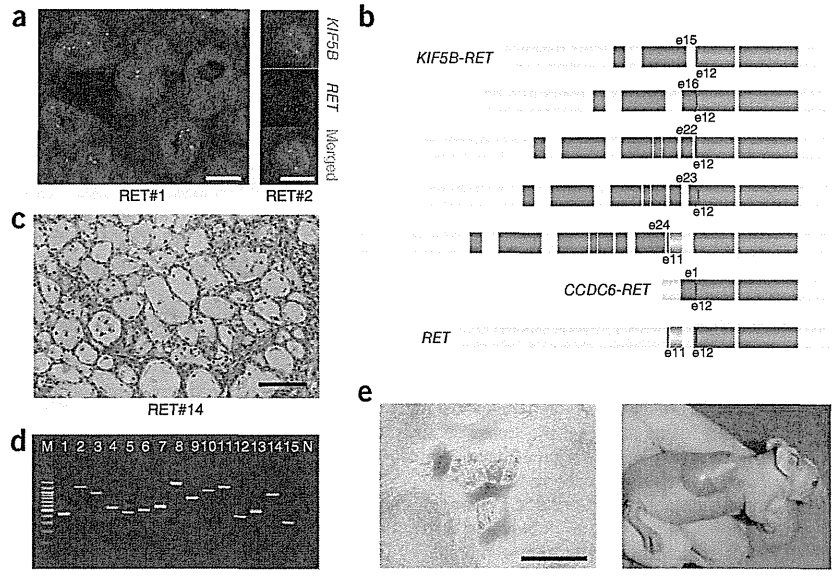


Figure 2 Discovery of RET fusions. (**a**) *RET* split (left) and fusion (right) FISH assay data (scale bars, 20 μm). In the split assay, multiple tumor cells harbored individual 3′ side signals (green), indicating the presence of *RET* rearrangement. In the fusion assay, a fusion signal (yellow) was observed in the representative tumor cell of subject RET#2, which is consistent with the presence of inv(10)(p11.22q11.2). (**b**) The break points of RET are exons 11 and 12. Both of the break points allow the resulting fusion to harbor the kinase domain of RET (red), and the exon 11 break point allows the resulting fusion to harbor the transmembrane domain of RET (orange). In the fusion partners, dark blue represents a coiled-coil domain, which probably contributes to the homodimelization of the fusion. Only the longer isoforms of RET and the RET fusions are shown. (**c**) Subject RET#14 showed the representative histopathology of mucinous cribriform carcinoma (scale bar, 100 μm). (**d**) The results for fusion-specific RT-PCR for subjects ALK#10 (lane 1, EML4-ALK, E13;A20, predicted product size of 432 bp), ALK#16 (lane 2, EML4-ALK, E20;A20, 1185 bp), ALK#26 (lane 3, EML4-ALK, E6;A20, 913 bp), ALK#38 (lane 4, EML4-ALK, E14;ins11del49A20, 546 bp), ALK#39 (lane 5, EML4-ALK, E2;A20, 454 bp), ALK#40 (lane 6, EML4-ALK, E13;ins69A20, 501 bp), ALK#41 (lane 7, EML4-ALK, E14;del14A20, 570 bp), ALK#42 (lane 8, KIF5B-ALK, K17;A20, 1,483 bp), ALK#44 (lane 9, KIF5B-ALK, K24;A20, 814 bp), RET#6 (lane 10, KIF5B-RET, K15;R12, 1,104 bp), RET#9 (lane 11, KIF5B-RET, K16;R12, 1,293 bp), RET#10 (lane 12, KIF5B-RET, K22;R12, 420 bp), RET#11 (lane 13, KIF5B-RET, K23;R12, 525 bp), RET#12 (lane 14, KIF5B-RET, K24;R11, 999 bp) and RET#13 (lane 15, CCDC6-RET, C1;R12, 352 bp). M and N represent the size marker (100-bp ladder) and non-template control, respectively. (**e**) The transforming potential of the KIF5B-RET fusion. Mouse 3T3 fibroblasts infected with a retrovirus encoding K15;R12L derived from tumor RET#7 fomed multiple foci (scale bar, 1 mm). All of the four nude mice injected with the corresponding 3T3 cells developed a subcutaneous tumor (right).

Supplementary Tables 1–4 summarize the clinicopathological features of the subjects. Briefly, young age, low smoking index and small tumor size characterized the kinase-fusion–positive group of subjects (**Supplementary Table 2**). A multivariate analysis of the adenocarcinomas revealed four independent factors that were indicators of poor prognosis: age ≥50 years, male sex, high pathological stage and negative kinase-fusion status (**Supplementary Table 3**). There was no significant difference in overall survival between the kinase-positive and epidermal growth factor receptor (EGFR)-mutant groups (P = 0.32). **Supplementary Table 4** shows the clinicopathological features of the subjects stratified by each fusion.

The transforming ability of CCDC6-RET and all of the ALK fusions, excluding K17;A20, was shown previously[1–3,8,12]. 3T3 cells infected with a virus expressing K17;A20, tropomyosin 3 (TPM3)-ROS1, syndecan 4 (SDC4)-ROS1, SLC34A2-ROS1, CD74-ROS1, ezrin (EZR)-ROS1, leucine-rich repeats and immunoglobulin-like domains 3 (LRIG3) (transcript variant 2)-ROS1 or KIF5B-RET (with both the longer (RET51) and shorter (RET9) RET isoforms) led to multiple transformed foci formation in culture and in subcutaneous tumors in a nude mouse tumorigenicity assay (**Figs. 1d, 2e** and **Supplementary Fig. 5**).

To test whether vandetanib, an inhibitor of vascular endothelial growth factor receptor (VEGFR-2), VEGFR-3, EGFR and RET[14], might be effective for the treatment of RET-fusion–positive tumors, we induced Flag-tagged EML4-ALK (E13;A20) or KIF5B-RET (K15;R12L and K15;R12S) in Ba/F3 cells, which are dependent on interleukin-3 (IL-3) for growth. All transfected cells, including those without any kinase fusion, proliferated in the presence of IL-3, but only cells expressing E13;A20 or K15;R12L grew in the absence of IL-3 (**Supplementary Fig. 6a**). In the absence of IL-3, vandetanib inhibited the proliferation of cells expressing K15;R12L (**Supplementary Fig. 6c**) but not the proliferation of cells expressing E13;A20 (**Supplementary Fig. 6d**). Crizotinib was not effective in inhibiting the proliferation of Ba/F3 cells expressing K15;R12L (**Supplementary Fig. 7**).

In 1985, a 3T3 assay identified *RET* as a rearranged transforming gene[15]. RET fusions have been identified exclusively in papillary thyroid carcinoma and are more frequently observed in radiation-associated thyroid cancers (for example, in survivors of the Chernobyl accident[16], atomic bomb survivors[17] and post-radiation therapy patients[18]). Therefore, a retrospective comparison of RET fusions in individuals with lung cancer with and without a history of radiation exposure warrants further study. If a positive association is found between RET fusion and radiation exposure in these studies, it might be desirable for individuals with internal or therapeutic exposure to irradiation (for example, those individuals involved in the Fukushima accident) to be monitored prospectively for lung cancer as well as thyroid cancer.

In Japan, more than 40% of lung adenocarcinomas in younger individuals harbor EGFR mutations[19]. In this study, 16% (17 out of 107) of younger individuals (≤50 years of age) with adenocarcinoma harbored a kinase fusion. Collectively, as long as molecular target diagnoses are properly performed, >50% of the individuals with lung adenocarcinoma in this generation may benefit from treatment with corresponding kinase inhibitors. Integrated pathology-based screening techniques can also be used for the selection of individuals to receive this treatment[20]. The results of our study will facilitate the development of a molecular classification of lung adenocarcinomas that is closely related to both the pathogenesis and the treatment of disease. This study was approved by the Institutional Review Board of the Cancer Institute Hospital, and all subjects provided informed consent.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemedicine/.

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

ACKNOWLEDGMENTS

We thank M. Iwakoshi, K. Shiozawa, T. Kakita, H. Nagano and K. Nomura for their technical assistance and S. Sengoku for providing administrative assistance. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, as well as by grants from the Japan Society for the Promotion of Science; the Ministry of Health, Labor and Welfare of Japan; the Vehicle Racing Commemorative Foundation of Japan; the Princess Takamatsu Cancer Research Fund; and the Uehara Memorial Foundation.

AUTHOR CONTRIBUTIONS

K.T. conceived of and led the entire project, designed the FISH probes, screened samples using FISH and immunohistochemistry, performed histopathological analyses, generated figures and tables and wrote the manuscript. M.S. performed functional analyses and generated the figures. Y.T. performed inverse RT-PCR and RACE experiments and their corresponding analyses. R.S. conducted statistical analyses. S.S. performed FISH and histopathological analyses. S.H. processed and analyzed the tissue microarrays and FISH screening and generated figures. R.A. processed the FISH probe library. W.H. made and analyzed the database and processed tissue microarrays. H.N., H.U., Y.S., S.O. and K.N. collected specimens and clinical information and were involved in planning the project. Y.L.C. conducted functional analyses. H.M. supervised the functional analyses and planned the project. Y.I. performed histopathological analyses and collected specimens. All authors participated in the discussion and interpretation of the data and the results.

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. Soda, M. *et al. Nature* **448**, 561–566 (2007).
2. Takeuchi, K. *et al. Clin. Cancer Res.* **14**, 6618–6624 (2008).
3. Takeuchi, K. *et al. Clin. Cancer Res.* **15**, 3143–3149 (2009).
4. Rikova, K. *et al. Cell* **131**, 1190–1203 (2007).
5. Takeuchi, K. *et al. Haematologica* **96**, 464–467 (2011).
6. Takeuchi, K. *et al. Clin. Cancer Res.* **17**, 3341–3348 (2011).
7. Sugawara, E. *et al. Cancer* published online, doi:10.1002/cncr.27391 (17 January 2012).
8. Choi, Y.L. *et al. Cancer Res.* **68**, 4971–4976 (2008).
9. Inamura, K. *et al. J. Thorac. Oncol.* **3**, 13–17 (2008).
10. Takeuchi, K. *Pathol. and Clin. Med.* **28**, 139–144 (2010).
11. Jokoji, R. *et al. J. Clin. Pathol.* **63**, 1066–1070 (2010).
12. Grieco, M. *et al. Cell* **60**, 557–563 (1990).
13. Tognon, C. *et al. Cancer Cell* **2**, 367–376 (2002).
14. Flanigan, J., Deshpande, H. & Gettinger, S. *Biologics* **4**, 237–243 (2010).
15. Takahashi, M., Ritz, J. & Cooper, G.M. *Cell* **42**, 581–588 (1985).
16. Ito, T. *et al. Lancet* **344**, 259 (1994).
17. Hamatani, K. *et al. Cancer Res.* **68**, 7176–7182 (2008).
18. Bounacer, A. *et al. Oncogene* **15**, 1263–1273 (1997).
19. Kosaka, T. *et al. Cancer Res.* **64**, 8919–8923 (2004).
20. Han, B. *et al. Cancer Res.* **68**, 7629–7637 (2008).

Original Article

Identification of Anaplastic Lymphoma Kinase Fusions in Renal Cancer

Large-Scale Immunohistochemical Screening by the Intercalated Antibody-Enhanced Polymer Method

Emiko Sugawara, MD[1,2]; Yuki Togashi, MS[1,3]; Naoto Kuroda, MD[4]; Seiji Sakata, MD, PhD[1]; Satoko Hatano, BS[1,3]; Reimi Asaka, BS[1,3]; Takeshi Yuasa, MD, PhD[5]; Junji Yonese, MD, PhD[5]; Masanobu Kitagawa, MD, PhD[2]; Hiroyuki Mano, MD, PhD[6,7]; Yuichi Ishikawa, MD, PhD[3]; and Kengo Takeuchi, MD, PhD[1,3]

BACKGROUND: Several promising molecular-targeted drugs are used for advanced renal cancers. However, complete remission is rarely achieved, because none of the drugs targets a key molecule that is specific to the cancer, or is associated with "oncogene addiction" (dependence on one or a few oncogenes for cell survival) of renal cancer. Recently, an anaplastic lymphoma kinase (ALK) fusion, vinculin-ALK, has been reported in pediatric renal cell carcinoma (RCC) cases who have a history of sickle cell trait. In this context, ALK inhibitor therapy would constitute a therapeutic advance, as has previously been demonstrated with lung cancer, inflammatory myofibroblastic tumors, and anaplastic large cell lymphomas. **METHODS:** Anti-ALK immunohistochemistry was used to screen 355 tumor tissues, using the intercalated antibody-enhanced polymer (iAEP) method. The cohort consisted of 255 clear cell RCCs, 32 papillary RCCs, 34 chromophobe RCCs, 6 collecting duct carcinomas, 10 unclassified RCCs, 6 sarcomatoid RCCs, and 12 other tumors. **RESULTS:** Two patients (36- and 53-year-old females) were positive for ALK as determined by iAEP immunohistochemistry. Using 5′-rapid amplification of complementary DNA ends, we detected *TPM3-ALK* and *EML4-ALK* in these tumors. The results of this study were confirmed by fluorescence in situ hybridization assays. The 2 ALK-positive RCCs were unclassified (mixed features of papillary, mucinous cribriform, and solid patterns with rhabdoid cells) and papillary subtype. They comprised 2.3% of non-clear cell RCCs (2 of 88) and 3.7% of non-clear cell and nonchromophobe RCCs (2 of 54). **CONCLUSIONS:** The results of this study indicate that ALK fusions also exist in adult RCC cases without uncommon backgrounds. These findings confirm the potential of ALK inhibitor therapy for selected cases of RCC. ***Cancer* 2012;118:4427-36.** © *2012 American Cancer Society.*

KEYWORDS: anaplastic lymphoma kinase, molecular-targeted therapy, renal cell carcinoma, immunohistochemistry, intercalated antibody-enhanced polymer.

INTRODUCTION

Renal cancer is one of the major cancers. The incidence and mortality of cases are estimated at 273,518 and 116,368 in the world; 14,963 and 6957 in Japan; and 56,678 and 13,711 in the United States.[1] The 5-year survival rate of patients with localized disease is relatively good: 65% to 93% and 47% to 77% for stages 1 and 2, respectively.[2] For advanced renal cancers (34%-80% and 2%-20% 5-year survival rates in stages 3 and 4, respectively),[2] several molecular-targeted drugs have been recently approved by the US Food and Drug Administration. These drugs, which include sunitinib, sorafenib, temsirolimus, everolimus, bevacizumab, pazopanib, and axitinib, are promising. However, none of them targets a key molecule that is specific to the cancer, or is associated with "oncogene addiction" of renal cancer, namely, the dependence on one or a few oncogenes for maintenance of the malignant phenotype and cell survival.

Anaplastic lymphoma kinase (ALK) fusion is a potential vulnerability, an "Achilles' heel", of many types of human cancer, including lymphoma,[3,4] sarcoma,[5] and carcinoma.[6,7] Experimentally, lung adenocarcinomas developed in EML4-ALK (fusion of ALK with echinoderm microtubule–associated protein like 4) transgenic mice were successfully treated with an ALK inhibitor.[8] The ALK inhibitor crizotinib has recently been used in patients with lung cancer, inflammatory myofibroblastic tumors (IMTs), or anaplastic large cell lymphomas (ALCLs), which harbor various ALK fusions. The compound showed an 81% response rate in ALK-positive lung cancers defined by at least 2 diagnostic methods,[9,10] and a

Corresponding author: Kengo Takeuchi, MD, PhD, Pathology Project for Molecular Targets, The Cancer Institute, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto, Tokyo 135-8550, Japan; Fax: (011) 81 3-3570-0230; kentakeuchi-tky@umin.net

[1]Pathology Project for Molecular Targets, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan; [2]Department of Comprehensive Pathology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan; [3]Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan; [4]Department of Diagnostic Pathology, Kochi Red Cross Hospital, Kochi City, Kochi, Japan; [5]Department of Urology, The Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan; [6]Division of Functional Genomics, Jichi Medical University, Tochigi, Japan; [7]Department of Medical Genomics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

We thank Tomoyo Kakita, Keiko Shiozawa, and Motoyoshi Iwakoshi for their technical assistance, and Sayuri Sengoku for providing administrative assistance.

DOI: 10.1002/cncr.27391, **Received:** October 1, 2011; **Revised:** October 31, 2011; **Accepted:** November 10, 2011, **Published online** January 17, 2012 in Wiley Online Library (wileyonlinelibrary.com)

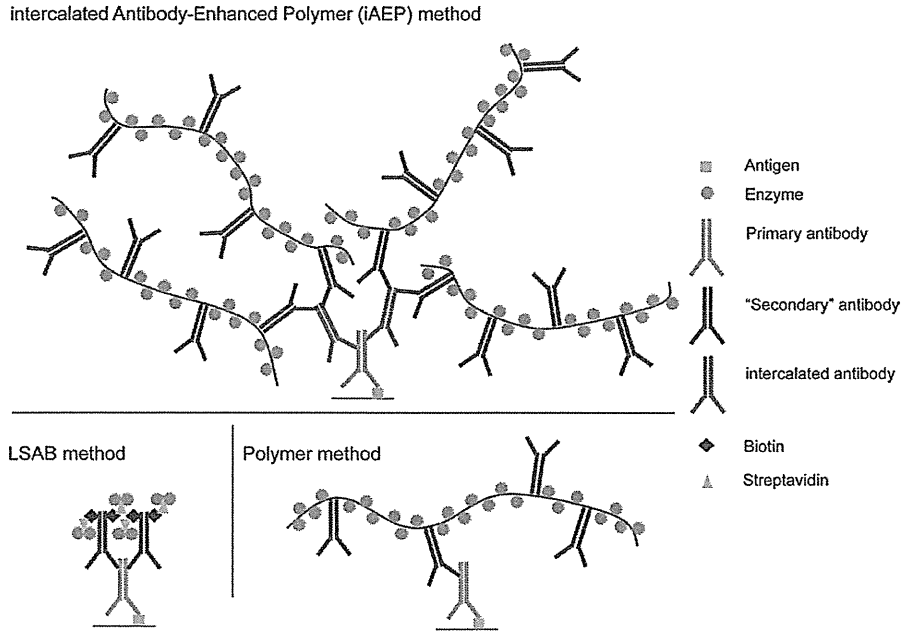


Figure 1. Schematic of intercalated antibody-enhanced polymer (iAEP) method is shown. The labeled streptavidin biotin (LSAB) and polymer methods are common conventional immunohistochemistry methods. In the iAEP method, a step of "intercalated antibody" is added between those of the primary antibody and polymer reagent. Thus, the iAEP method has an additional step compared with the polymer method, but the same number of steps as the LSAB method. There are generally 2 ways to raise the sensitivity of immunohistochemistry. The first is to raise the sensitivity of the antigen-antibody reaction, by increasing the concentration of the primary antibody, using a more sensitive antibody, antigen-retrieval technique, and so forth. The second is to raise the sensitivity of the detection system for the antigen-antibody immune complex. These 2 techniques may appear to generate the same result; however, in principle, they are totally different. The staining results are more likely to differ, especially when the antigen density is very low, such as for EML4-ALK (fusion of echinoderm microtubule-associated protein like 4 with anaplastic lymphoma kinase) or PPFIBP1-ALK (fusion of PTPRF interacting protein binding protein 1 with ALK).[13,24] In such a setting, the latter technique is more advantageous. The staining intensity depends on the density of enzyme in the antigen site. However sensitive a primary antibody is, the antigen-antibody complex cannot exceed the number of antigens. In contrast, it is easy to increase the enzyme density per antigen-antibody complex with use of the latter technique, which includes the iAEP method.

strong response in IMT for several months.[11] Two patients with ALCL who were receiving crizotinib achieved complete remission.[12] These findings indicate that ALK fusion addiction is one of the most promising targets in cancer therapy.

To ensure that such molecular-targeted therapy is effective and less toxic, accurate screening methods to detect ALK fusions are crucial. However, although immunohistochemistry has been a gold standard for the detection of ALK fusions in ALCL and IMT,[13,14] conventional anti-ALK immunohistochemistry is not sensitive enough to detect EML4-ALK, which was first described in lung cancer in 2007.[6,7] To overcome this, we developed a sensitive immunohistochemical tool, the intercalated antibody-enhanced polymer (iAEP) method (Fig. 1).[13] Combined with a conventional anti-ALK mouse monoclonal antibody 5A4, the iAEP method efficiently and consistently detected EML4-ALK in paraffin-embedded sections. In various studies on ALK-positive lung cancer, anti-ALK immunohistochemistry by iAEP or essentially equivalent methods was used to examine surgically resected specimens,[13,15-19] transbronchial lung biopsy specimens,[20] and endobronchial ultrasound-guided transbronchial needle aspiration specimens.[17,21,22] More importantly, some of the patients screened by anti-ALK iAEP immunohistochemical analysis received crizotinib therapy and showed a good response.[16,17,22] Novel ALK fusions, including v6 and v7 of EML4-ALK,[13] kinesin family member 5B (KIF5B)-ALK,[13] sequestosome 1 (SQSTM1)-ALK,[23] and PTPRF interacting protein, binding protein 1 (PPFIBP1)-ALK[24] have been identified using anti-ALK iAEP immunohistochemical analysis. Thus, anti-ALK iAEP immunohistochemistry constitutes a powerful tool for clinical and also research purposes.

The development of anti-ALK antibodies has facilitated the investigation of many types and cases of cancer, including lung cancer.[25-27] Since 1994, ALK-positive tumors have been identified exclusively in lymphoma