BMC
Cancer

Allelotypes of lung adenocarcinomas featuring *ALK* fusion demonstrate fewer onco- and suppressor gene changes

Hironori Ninomiya[1], Motohiro Kato[4,5], Masashi Sanada[4], Kengo Takeuchi[1,2], Kentaro Inamura[1], Noriko Motoi[1], Hiroko Nagano[1], Kimie Nomura[1], Yukinori Sakao[3], Sakae Okumura[3], Hiroyuki Mano[6,7], Seishi Ogawa[4] and Yuichi Ishikawa[1*]

Abstract

Background: A subset of lung adenocarcinomas harboring an *EML4-ALK* fusion gene resulting in dominant oncogenic activity has emerged as a target for specific therapy. *EML4-ALK* fusion confers a characteristic histology and is detected more frequently in never or light smokers and younger patients.

Methods: To gain insights into etiology and carcinogenic mechanisms we conducted analyses to compare allelotypes of 35 *ALK* fusion-positive and 95 -negative tumours using single nucleotide polymorphism (SNP) arrays and especially designed software which enabled precise global genomic profiling.

Results: Overall aberration numbers (gains + losses) of chromosomal alterations were 8.42 and 9.56 in tumours with and without *ALK* fusion, respectively, the difference not being statistically significant, although patterns of gain and loss were distinct. Interestingly, among selected genomic regions, oncogene-related examples such as 1p34.3 (*MYCL1*), 7q11.2(*EGFR*), 7p21.1, 8q24.21(*MYC*), 16p13.3, 17q12(*ERBB2*) and 17q25.1 showed significantly less gain. Also, changes in tumour suppressor gene-related regions, such as 9p21.3 (*CDKN2A*) 9p23-24.1 (*PTPRD*), 13q14.2 (*RB1*), were significantly fewer in tumours with *ALK* fusion.

Conclusion: Global genomic comparison with SNP arrays showed tumours with *ALK* fusion to have fewer alterations in oncogenes and suppressor genes despite a similar overall aberration frequency, suggesting very strong oncogenic potency of *ALK* activation by gene fusion.

Keywords: Lung adenocarcinoma, *ALK* fusion, SNP array, Allelotype, Copy number

Background

The adenocarcinoma is the most common form of lung cancer worldwide, different subsets having specific genetic backgrounds of great importance for molecular-targeted therapy. For example, somatic mutations of the epidermal growth factor receptor (*EGFR*) are especially prevalent in adenocarcinomas among never smokers, females, and those with Asian ethnicity [1]. On the other hand, *KRAS* mutations are associated with the smoking habit [2] and the two tend to be mutually exclusive. Recently, Soda et al. found a novel fusion gene, *EML4-ALK*, arising from an

inversion on the short arm of chromosome 2 in non-small cell lung carcinomas [3]. *ALK* fusion is a unique example of tyrosine kinase activation by structural chromosome re-arrangement [4].

EML4-ALK fusion is a powerful driving molecular event by itself. The chimeric protein permits ligand-independent dimerization and constitutive activation of *ALK*, resulting in dominant oncogenic activity. Multiple fusion variants of *EML4-ALK* and notable clinicopathological characteristics of fusion positive tumours have been revealed [5-9]. Since the tyrosine kinase is involved and activated by gene fusion, this type of malignancy has emerged as a target for anti-tyrosine kinase therapy [4,10-12].

We have revealed that *ALK* fusion-positive tumours constituted a particular subset in lung adenocarcinomas in

* Correspondence: ishikawa@jfcr.or.jp
[1]Division of Pathology, The Cancer Institute, Ariake 3-8-31, Koutou-ku, Tokyo 135-8550, Japan
Full list of author information is available at the end of the article

BioMed Central

terms of clinical characteristics, histology and etiology, as well as molecular changes [7,8]. It is of great interest to assess global genomic alterations to provide deep insight into their genesis, especially considering these tumours arise in non- or light smokers. Single nucleotide polymorphism (SNP) microarray analysis enables precise high-throughput detection of genomic copy number alterations, gains and losses in the genome contributing to carcinogenesis [13] with gene expression varying consistently with DNA copy number changes [14,15]. We therefore conducted of the present genomic profiling of lung adenocarcinomas with and without *ALK* fusion.

Methods

Patient population and specimens

A series of 130 cases of lung adenocarcinomas, 35 with *EML4-ALK* or *KIF5B-ALK* fusion and 95 cases without, were enrolled in this study. From 1998 to 2008, 1,086 primary lung adenocarcinomas were surgically resected at Thoracic Surgery Division, the Cancer Institute Hospital, Japanese Foundation for Cancer Research (JFCR), Tokyo. All cases were screened as to ALK expression by immunohistochemistry using the iAEP method [6] and for positive cases subsequent RT-PCR and FISH analysis were performed, as previously described [5,6,16]. Among them, sufficient amounts and quality of fresh tumour material were available for 35 cases. Fusion gene variants are listed in Addtional file 1: Table S1. V3 constituted the largest proportion, 31% (11/35), having a breakpoint at exon 20 of *EML4*. A rare variant, *KIF5B-ALK* fusion, was detected in two cases. There was no correlation with fusion variant and pathological subtypes (data not shown). The 95 cases without *ALK* fusion were randomly selected from 730 surgically resected adenocarcinomas from 1995 to 2003 at the same hospital. Tissue specimens were snap-frozen in liquid nitrogen, typically within 20 minutes after resection, and stored at −80°C until use. Genomic DNA was extracted by standard proteinase K digestion and the phenol-chloroform method. To confirm if specimens used for analysis in this study contained a significant amount of tumour cells, typically 50% or more, a neighboring surface was examined histologically with frozen sections. This study was approved by the institutional review board of the JFCR.

Mutation analysis of EGFR, KRAS and TP53

For *EGFR* mutation analysis, exons 18 to 21 were amplified by PCR with specific oligo-primers. For point mutations in exon 18, PCR products were directly sequenced. Fragment analysis was performed for exons 19 and 20 deletions and insertion mutations. The presence of one point mutation in exon 21 was detected by genotyping analysis. To examine *TP53* mutations, direct sequencing from exons 5 to 10 was carried out. For *KRAS* mutation analysis, codons 12, 13 and 61 were examined by direct sequencing. Primers and detailed procedures were as described previously [17].

Histological diagnosis and clinical staging

Histological diagnosis was made on the basis of World Health Organization (WHO) classification [18] by experienced pathologists (N.M. and Y.I.). Pathological staging was based on the AJCC/UICC staging manual of lung cancer [19]. Differentiation grading of adenocarcinoma was determined essentially according to the Japan Lung Cancer Society criteria as illustrated previously [20]. Briefly, well-differentiated (w/d) tumors are composed chiefly of glands lined by, or of papilla covered by, one-layered tumor cells. Also, Adenocarcinoma in situ (AIS) is included in this category. Moderately differentiated (m/d) lesions comprise glands showing a cribriform pattern, fused with one another, or glands lined by, or papillae covered by, tumor cells demonstrating obvious piling-up. Poorly differentiated (p/d) carcinomas show mainly solid growth and only occasionally glandular/papillary patterns and/or mucus production. Blood vessel and lymphatic invasion was also explored microscopically, with hematoxylin-eosin and elastic-fiber stained sections of maximum tumour diameter made from paraffin-embedded specimens.

SNP array analysis and comparisons of allelic imbalance at the chromosome arm level and in selected cancer-related regions

Extracted DNA was subjected to Affymetrix GeneChip Mapping 250K arrays. Allelic imbalance was analyzed using software termed the Copy Number Analyzer for Affymetrix Gene Chip Mapping (CNAG Ver. 2.0) [21]. After appropriate normalization of mean array intensities, signal ratios between tumours and anonymous normal references were calculated in an allele-specific manner, and allele-specific copy numbers were inferred from the observed signal ratios based on the hidden Markov model using the CNAG/AsCNAR software [21-23]. With this procedure, genomic profiles of *ALK* fusion-positive and -negative tumours were obtained. Datas have been deposited at NCBI's Gene Expression Omnibus data repository under GEO series accession number GSE41536.

Comparison was at two levels; a chromosome arm level and a smaller, specific gene locus level. To do this, first we compared average numbers of chromosome arms altered between the two groups [24]. We called gain or loss of each chromosomal arm when copy number change stretched more than 80% of entire length. Secondly, we compared recurrent copy number aberrations at twenty-one cancer-related loci with gains and five with losses. These specific regions were selected based on previous studies of the lung cancer genome

[25,26] and through our global mapping with CNAG. The selected regions with relevant genes were as follows: for gains, 1p34.3 (*MYCL1*), 1q21.2 (*S100 family*), 3q29 (*MUC4*), 5p15.33 (*TERT*), 6p21.1 (*VEGF*), 7p11.2 (*EGFR*), 7p21.1, 7q31.2 (*MET*), 8q24.21 (*MYC*), 10q11.22, 12p12.1 (*KRAS*), 12q14.1 (*CDK4*), 12q15 (*MDM2*), 14q13.3 (*TTF1*), 16p13.3, 17q12 (*ERBB2*), 17q25.1, 19q12 (*CCNE1*), 20q13.2, 20q13.32, 20q13.33 (*TNFSF6B*); and for losses, 9p21.3 (*CDKN2A*), 9p23-p24.1 (*PTPRD*), 10q23.31 (*PTEN*), 13q14 (*RB1*), 17p13.1 (*TP53*).

Statistical analysis

Clinicopathological parameters of cases with or without *ALK* fusion and the frequencies of chromosome arms changed and copy numbers of targeted loci were compared by the chi-square test or the Fisher's exact test as appropriate. The average number of chromosome arms altered with or without *ALK* fusion was compared with Students' *t*-test. Statistical significance was defined as $P=0.05$ or less.

Results

Comparisons of clinicopathological profiles of tumours with or without *ALK* fusion

Clinicopathological profiles of patients are summarized in Table 1. *ALK* fusion-positive cases were significantly younger and featured significantly more never-smokers ($P=0.05$, $P=0.004$, respectively). *ALK* fusion-positive tumours were histologically adenocarcinomas with notable characteristics such as poor differentiation as well as an acinar type structure and mucin production, as reported previously [7-9]. In this study, distribution of histological subtypes differed between two groups, namely, "acinar" subtype accounted for nearly forty percent in *ALK* fusion positive group (Table 1). The frequencies of vascular invasion, both of blood and lymph vessels, did not significantly differ between the two groups ($P=0.738$, $P=0.273$, respectively). In addition, the distribution of pathological stages did not vary ($P=0.532$).

Mutational status of TP53, EGFR and KRAS

Data for the mutational status of *TP53, EGFR and KRAS* in the two groups are summarized in Table 1. Twenty-one cases had *TP53* mutations. Only one case with *ALK* fusion (Case 9: 1/35, 3%) harbored a mutation, a G/A transition at codon 273, as compared to 20 cases without *ALK* fusion (20/95, 21%), the mutation rates being significantly different ($P=0.014$) (Table 1, Additional file 1: Table S2). Twelve (12/21, 57%) of the *TP53*-mutated cases had a smoking history.

EGFR and *KRAS* mutations were not detected among *ALK* fusion-positive tumours. This fact that *ALK* rearrangement was mutually exclusive with *EGFR* and *KRAS* mutations ($P<0.0001$, $P=0.189$, respectively) is in

Table 1 Comparison of clinicopathological parameters in cases with or without ALK fusion

	with fusion	without fusion	P
n	35	95	
Age (years)	58.5	62.8	0.050
gender			
male	14	47	0.337
female	21	48	
smoking			
never	25	41	0.004
ever	10	54	
pStage			
I	20	60	0.532
II-IV	15	35	
differentiation grade			
wel	4	44	<0.001
mod+por	31	51	
Predominant subtype			
papillary	21	77	0.019
Acinar	13	13	
Bronchioloalveolar	0	4	
solid with mucin	0	1	
signet	1	0	
lymphatic invasion			
-	24	68	0.738
+	11	27	
Vessel invasion			
-	15	51	0.273
+	20	44	
TP53 mutation			
-	34	75	0.014
+	1	20	
EGFR mutation			
-	35	40	<0.0001
+	0	55	
KRAS mutation			
-	35	88	0.189
+	0	7	

line with our previous studies [8]. The *EGFR* mutation rate was 58% (55/95) in *ALK* fusion-negative cases and decreased with the smoking burden: 70.7% (29/41) in never smokers, 62.5% (15/24) in light smokers (0<pack-years<20) and 36.7% in heavy smokers (more than 20 pack-years) (11/30) (Additional file 2: Figure S1). *KRAS* mutations were identified in 7.4% (7/95) of *ALK* fusion-negative cases, and detected only among smokers. Though *KRAS* mutations were examined through

codons 12, 13 and 61, they were found only in codon 12. The *KRAS* mutation rate increased along with the elevation of smoking index (Additional file 2: Figure S1). These findings for *EGFR* and *KRAS* mutations are consistent with previous reports from Japan, the prevalence being quite different from that in the United States [27-29].

DNA copy number alterations of chromosome arms

We compared the allelotypes of each chromosome arm between the two groups. Global views of chromosome aberrations are shown in Figure 1. Note that in *ALK* fusion-positive tumours, genomic copy number changes were more evenly distributed over the chromosome arms and high copy number gains (dark-red) in short genomic segments were less frequently encountered than with *ALK*-fusion negative examples. Significantly different patterns of respective chromosomal arm gain and loss were noted between the two groups. In fact, 5q, 8p, 9q, 11p and 11q were significantly more amplified, and 6q was more deleted in *ALK* fusion-positive tumours, whereas, in *ALK* fusion-negative tumours, 17q was more amplified, and 8p and 9p were more deleted (Figure 1, Table 2a, Additional file 1: Table S3-S5). *P*-values for comparisons of the aberration frequency in each chromosome arms are shown in Additional file 1:

Table S5. When comparing global chromosome instability levels between the two groups, average numbers of chromosome arms with copy number gain or loss were 8.42 ±7.46 and 9.56 ±7.90 for tumours with and without *ALK* fusion, respectively, as detailed in Table 3, the difference not being statistically significant.

Chromosomal number alterations with advancement of pathological stage

Chromosome aberration might be expected to increase as tumours progress in stages and, if so, numbers of chromosome arms with gain and/or loss might be larger in advanced tumours. In fact however, when we compared the number of chromosome arm altered between pathological stage I and II-IV, total number did not increase in pathological stage II-IV, though only *ALK* fusion-negative tumours showing significant elevation of chromosomal gain (Figure 2).

Comparison of gain and loss frequency of selected loci

We selected twenty-one loci with recurrent copy number gain and five loci with loss to compare small-scale genomic aberrations. All the loci examined and *P*-values are summarized in Additional file 1: Table S6, S7. In Figure 3, stacked bar charts are shown indicating the
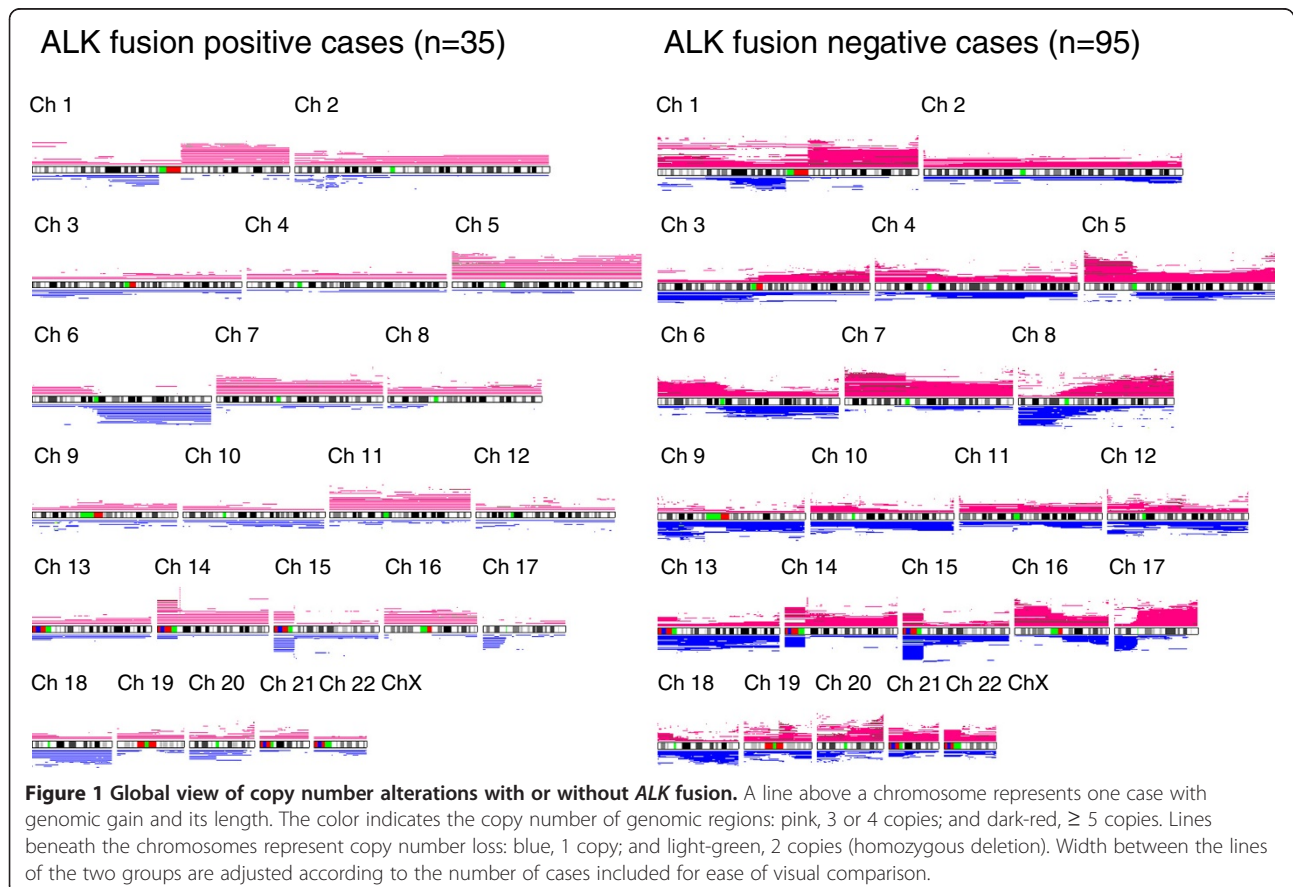

Figure 1 Global view of copy number alterations with or without *ALK* fusion. A line above a chromosome represents one case with genomic gain and its length. The color indicates the copy number of genomic regions: pink, 3 or 4 copies; and dark-red, ≥ 5 copies. Lines beneath the chromosomes represent copy number loss: blue, 1 copy; and light-green, 2 copies (homozygous deletion). Width between the lines of the two groups are adjusted according to the number of cases included for ease of visual comparison.

Table 2 Comparisons of significantly altered chromosomal arms between adenocarcinomas with and without *ALK* fusion

Category	Gain	Loss
More frequent *with* ALK fusion	5q, 8p, 9q, 11p, 11q	6q
More frequent *without* ALK fusion	17q	8p, 9q

percentage gain or loss of the selected loci. Interestingly, copy numbers (and related genes) at 1p34.3 (*MYCL1*), 7p11.2 (*EGFR*), 7p21.1, 8q24.21 (*MYC*), 16p13.3, 17q12 (*ERBB2*) and 17q25.1 were significantly less gained, and those at 9p21.3 (*CDKN2A*), 9p23-p24.1 (*PTPRD*), 13q14.2 (*RB1*) were significantly less deleted in *ALK* fusion-positive tumours than fusion-negative ones, with loci related to both oncogenes and tumour suppressor genes having fewer changes in tumours with *ALK* fusion. There were no oncogene-related loci with more gains and no suppressor gene-related loci with more losses in tumours with *ALK* fusion.

Homozygous deletions were found only at 9p21, at frequencies similar between the two groups, although the summed frequencies of heterozygous and homozygous deletions at 9p21.3 did significantly differ. In the group without *ALK* fusion, all the cases with the homozygous deletion harbored *EGFR* mutations.

MYCL1, *EGFR*, *MYC* and *ERBB2* are well-known oncogenes and *CDKN2A* and *RB1* are tumour suppressor genes related to lung carcinogenesis. *PTPRD* has been suggested to function as a tumour suppressor in several tumours, including lung cancers [30] and brain tumours [31]. Notably, 5p15.33, including *TERT* (telomerase reverse transcriptase), had the highest rate of gain in both groups regardless of *ALK* fusion (Additional file 3: Figure S2 and Additional file 1: Table S7).

Taken together, *ALK* fusion-positive tumours showed similar levels of overall chromosome instability, but when focusing on particular cancer-related regions, significantly fewer copy number gains at oncogene-related loci and significantly fewer deletions at suppressor gene-related loci.

Discussion

Recurrent chromosome translocation has been accepted to play an important role in the pathogenesis of hematological malignancies, but not of solid tumours. Recently, however,

Table 3 Comparisons of numbers of chromosome arms with aberrations between adenocarcinomas with or without *ALK* fusion

	with *ALK* fusion (n=35)	without *ALK* fusion (n=95)	*P*
Gains	5.97±6.75	6.21±6.95	0.859
Losses	2.46±3.06	3.35±4.34	0.196
Total	8.42±7.46	9.56±7.90	0.454

Note that significant differences are not detected.

chromosome rearrangements in solid tumours such as prostate cancer and non-small cell lung cancer have been reported [32]. *ALK* fusion was originally described in anaplastic large-cell lymphoma as a chimeric protein *NPM-ALK* resulting from a translocation. More recently, evidence has accumulated that the *EML4-ALK* fusion gene defines a novel subclass of lung adenocarcinomas with distinct clinicopathological features [7-9], so that it has emerged as a target for therapy. We focused here for the first time on allelic imbalance of tumours with *ALK* fusion with a novel technique which has already shown the involvement of loss of A20 function in the pathogenesis of a subset of B-cell lymphomas [33] and gain of function of C-CBL tumour suppressor in myeloid neoplasms [34]. Applying this methodology, we demonstrated that lung adenocarcnomas with *ALK* fusion feature less amplification of loci with oncogenes and fewer deletions of loci related to tumour suppressor genes, although global chromosome aberrations were similar between tumours with and without *ALK* fusion. suggesting that the fusion gene is a driver mutation, not just a passenger mutation.

Genetic instability was here categorized into two groups for simplicity, at the chromosomal level and at the nucleotide level. We earlier found the former to play a more important role in lung carcinogenesis, the frequency of LOH (loss of heterozygosity) being higher in less-differentiated tumours [35]. *ALK* fusion positive tumours are more common among non-smokers and the younger population, similar to those with *EGFR* mutations. We had expected fewer chromosome aberrations in *ALK* fusion-positive tumours because tumours arising in such people usually harbor less LOH and a lower *TP53* mutation rate than smokers [36-38]. Contrary to our expectation, the global copy number changes at the chromosomal arm level did not differ between the two groups, although significant differences of alteration frequency at the individual chromosomal arms were seen. In addition, only *ALK* fusion-negative tumours showed an increase of the frequency of chromosome arm gain with the advancement of disease stage. Furthermore, at the smaller-genomic scale level, *ALK* fusion-positive tumours were less amplified at the loci containing *EGFR* family genes, 7p11.2 (*EGFR*), 17q12 (*ERBB2*) and other loci, 1p34.3 (*MYCL*), 7p21.1, 8q24.21 (*MYC*), 16p13.3 and 17q25.1. *EGFR* and *ERBB2* play important roles by dimerizing when their ligands binds to produce downward growth signals to the tumour cells. Mutations and activation of these genes may drive carcinogenesis [39], and increased expression is associated with a poor prognosis in NSCLCs [40-43]. *ALK* fusion positive tumours are speculated to be less dependent on the actions of oncogenes and tumour-suppressor genes induced by copy number changes. Our results may also indicate that there

	p-Stage	with *ALK* fusion	P	without *ALK* fusion	P
Gain	I	5.05±6.07	n.s.	4.97±6.41	0.017
	II-IV	7.20±7.59		8.53±7.43	
Loss	I	1.90±3.11	n.s.	3.60±4.66	n.s.
	II-IV	3.20±2.93		2.88±3.80	
Gain +Loss	I	6.95±6.78	n.s.	8.57±7.71	n.s.
	II-IV	10.4±8.09		11.41±8.11	

Figure 2 Comparisons of numbers of chromosome arms altered with or without *ALK* fusion in different pathological stages. Note that, whereas tumours in higher stages show more gains than stage I tumours when the tumours have no *ALK* fusion, *ALK* fusion positive tumours exhibit no such difference. p-Stage; pathological stage, n.s.; not statistically significant.

is common and frequent chromosome abnormality in lung adenocarcinomas independent of *ALK* fusion, such as the 5p15.33 region, including *TERT*.

As for genomic loss, 9p21.3 (*CDKN2A*), 9p23-p24.1 (*PTPRD*) and 13q14.2 (*RB1*) were significantly less frequently deleted in *ALK* fusion-positive tumours. Homozygous deletion was seen only at 9p21.3 including *CDKN2A* and limited to *EGFR*-mutated tumours among *ALK* fusion-negative neoplasms as reported in the literature [44] and also seen in *ALK*-fusion positive ones. That deletion of 9p23-24.1 and 13q14.2 including tumour suppressor genes was rare in *ALK* fusion-positive tumours suggests that they can grow even if the functions of these suppressor genes are retained.

Of all the selected loci, 5p15.33 containing *TERT* (telomerase reverse transcriptase isoform 2) showed the highest frequency of recurring gain regardless of *ALK* fusion. The enzyme is important for telomere regeneration and maintenance resulting in a growth advantage and Zhang et al. reported that the locus is a frequent target of amplification during tumourigenesis [45]. Copy number gain of this locus significantly correlates with telomerase activity [46] and is one of the most consistent alterations in the early stages of non-small cell lung cancer [47]. In addition, increased susceptibility to lung cancer development associated with a SNP polymorphism of this locus has been reported [48,49]. The fact that most human tumour cells have telomerase activity indicates that its acquisition is vital for carcinogenesis and cell immortalization, and it might explain the reason why lung adenocarcinomas with or without *ALK* fusion shows similar frequency of copy number gain of this locus.

Our results have some therapeutic relevance. The fact that there are less involvement of other oncogenes and tumor suppressor genes may be related to dramatic responses to targeted drugs because of intact cellular processes including apoptosis pathways. In this regard, there is an interesting paper by Camidge et al. [50], demonstrating the inverse relationship between fused and isolated red copy number on FISH might suggest

Figure 3 Significant differences in copy number change detected at seven loci for gain and three loci for loss among twenty-six selected loci. Colors of the stacked bars represent copy number: orange, 3,4 copies; red, ≥ 5 copies; blue, 1 copy loss (heterozygous deletion), light green: 2 copy losses (homozygous deletion). 17q12, 17q25.1 show remarkable differences in copy number gain between *ALK* fusion-positive and -negative tumours.

the *ALK* fusion positive tumor was a "near-diploid" subtype of non-small cell lung cancer. Comparing closely, however, between their and our results, our study clearly revealed the overall frequency of chromosome aberrations are similar between *ALK* fusion positive and negative tumors, suggesting not "near-diploid". But, certainly, we need more investigations on genomic instability of *ALK* fusion positive tumors.

It is well known that smoking causes genomic changes with allelic imbalance [20]. As shown in Table 1, smokers dominate never smokers in the group without fusion whereas the fusion-positive group has more never smokers than smokers. Since the tumors without *ALK* fusion include *EGFR*-mutated tumors, most of which are from never smokers, the *ALK* fusion-negative group is certainly heterogeneous. In due course, a study that describes comparisons of allelotypes of non-smoker's tumors between with *ALK* fusion and with *EGFR* mutation should be warranted.

Conclusions

Although overall frequencies of aberrations at the chromosome arm level do not appear to significantly differ between *ALK* fusion-positive and -negative tumors, smaller genomic regions including cancer-related genes do show significant variation. Thus tumors with *ALK* fusion feature significantly fewer copy number gains and losses at loci containing oncogenes and tumor-suppressor genes, respectively. This implies that *ALK* fusion itself exerts very strong driving forces for tumorigenesis, in other words, that *ALK* fusion is a driver mutation, not just a passenger mutation.

Additional files

Additional file 1: Table S1. Frequencies of fusion variants of *ALK* rearrangements. **Table S2.** Cases with *TP53* mutations and their smoking status. **Table S3.** Chromosomal arms and number of cases with gain with or without ALK fusion. **Table S4.** Chromosomal arms and number

of cases with loss with or without ALK fusion. **Table S5.** P-values for comparisons of the frequencies of chromosome aberrations in all chromosome arms between tumours with or without *ALK* fusion. **Table S6.** Number of cases with copy number gain or loss at selected loci with or without *ALK* fusion. **Table S7.** Significance of the differences in frequencies of copy number changes (gains and losses) between tumours with or without *ALK* fusion.

Addtional file 2: Figure S1. Mutation rates for EGFR, TP53 and KRAS according to cumulative smoking are shown. EGFR and KRAS mutations were only detected among ALK fusion negative cases, so ALK fusion positive cases were not included in the analysis. Note the gradually decrease in EGFR mutation rate with increase in cumulative smoking. KRAS mutations were detected only among smokers.

Additional file 3: Figure S2. Comparisons of copy number alteration rates at selected loci with or without *ALK* fusion. Note that 5p15.33 including *TERT* shows the highest gain both in *ALK* fusion positive and negative tumours, the frequencies being identical.

Competing interest
The authors have no potential conflicts of interest.

Authors' contributions
HN, MK, SO, HM and YI designed the study. HN, KT, KI, NM, HM and YI performed pathological and/or genomic diagnosis of tumors. HN, MK, MS and SO obtained microarray data and carried out bioinformatics analysis. HN and KN analyzed mutations. YS, SO and YI collected samples and/or provided detailed clinical data of patients. HN and YI drafted the manuscript. All authors read and approved the final manuscript.

Authors' information
HM has found ALK fusion in lung cancer with own developed cDNA library. MK, MS and SO detected genes responsible for hematological disorders through same algorithm with this study, CNAG/AsCAR. KT has created a novel diagnostic method to detect *ALK* fusion positive lung cancer. YI has found characteristic pathological features of *ALK* positive cancer.

Author details
[1]Division of Pathology, The Cancer Institute, Ariake 3-8-31, Koutou-ku, Tokyo 135-8550, Japan. [2]Pathology Project for Molecular Targets, The Cancer Institute, Ariake 3-8-31, Koutou-ku, Tokyo 135-8550, Japan. [3]Thoracic Oncology Center, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Ariake 3-8-31, Koutou-ku, Tokyo 135-8550, Japan. [4]Cancer Genomics Project, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan. [5]Department of Pediatrics, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan. [6]Department of Medical Genomics, Graduate School of Medicine, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan. [7]Division of Functional Genomics, Jichi Medical University, Tochigi 329-0498, Japan.

References
1. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, Herz J, Minna JD, Gazdar AF: **Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers.** *J Natl Canc Inst* 2005, **97**:339–346.
2. Mounawar M, Mukeria A, Le Calvez F, Hung RJ, Renard H, Cortot A, Bollart C, Zaridze D, Brennan P, Boffetta P, Brambilla E, Hainaut P: **Patterns of EGFR, HER2, TP53, and KRAS mutations of p14arf expression in non-small cell lung cancers in relation to smoking history.** *Cancer Res* 2007, **67**:5667–5672.
3. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H: **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.** *Nature* 2007, **448**:561–566.
4. Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, Choi HG, Kim J, Chiang D, Thomas R, Lee J, Richards WG, Sugarbaker DJ, Ducko C, Lindeman N, Marcoux JP, Engelman JA, Gray NS, Lee C, Meyerson M, Jänne PA: **EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer.** *Clin Cancer Res* 2008, **14**:4275–4283.
5. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, Enomoto M, Takada S, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H: **Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts.** *Clin Cancer Res* 2008, **14**:6618–6624.
6. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, Inamura K, Takada S, Ueno T, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H: **KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer.** *Clin Cancer Res* 2009, **15**:3143–3149.
7. Inamura K, Takeuchi K, Togashi Y, Nomura K, Ninomiya H, Okui M, Satoh Y, Okumura S, Nakagawa K, Soda M, Choi YL, Niki T, Mano H, Ishikawa Y: **EML4-ALK fusion is linked to histological characteristics in a subset of lung cancers.** *J Thorac Oncol* 2008, **3**:13–17.
8. Inamura K, Takeuchi K, Togashi Y, Hatano S, Ninomiya H, Motoi N, Mun MY, Sakao Y, Okumura S, Nakagawa K, Soda M, Choi YL, Mano H, Ishikawa Y: **EML4-ALK lung cancers are characterized by rare other mutations, a TTF-1 cell lineage, an acinar histology, and young onset.** *Mod Pathol* 2009, **22**:508–515.
9. Yoshida A, Tsuta K, Nakamura H, Kohno T, Takahashi F, Asamura H, Sekine I, Fukayama M, Shibata T, Furuta K, Tsuda H: **Comprehensive histologic analysis of ALK-rearranged lung carcinomas.** *Am J Surg Pathol* 2011, **5**:1226–1234.
10. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, Shapiro GI, Costa DB, Ou SH, Butaney M, Salgia R, Maki RG, Varella-Garcia M, Doebele RC, Bang YJ, Kulig K, Selaru P, Tang Y, Wilner KD, Kwak EL, Clark JW, Iafrate AJ, Camidge DR: **Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis.** *Lancet Oncol* 2011, **12**:1004–1012.
11. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Jänne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, Fidias P, Stubbs H, Engelman JA, Sequist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Salgia R, Shapiro GI, Clark JW, *et al*: **Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer.** *N Engl J Med* 2010, **363**:1693–1703.
12. Ou SH, Kwak EL, Siwak-Tapp C, Dy J, Bergethon K, Clark JW, Camidge DR, Solomon BJ, Maki RG, Bang YJ, Kim DW, Christensen J, Tan W, Wilner KD, Salgia R, Iafrate AJ: **Activity of crizotinib (PF02341066), a dual mesenchymal-epithelial transition (MET) and anaplastic lymphoma kinase (ALK) inhibitor, in a non-small cell lung cancer patient with de novo MET amplification.** *J Thorac Oncol* 2011, **6**:942–946.
13. Osada H, Takahashi T: **Genetic alterations of multiple tumor suppressors and oncogenes in the carcinogenesis and progression of lung cancer.** *Oncogene* 2002, **21**:7421–7434.
14. Järvinen AK, Autio R, Kilpinen S, Saarela M, Leivo I, Grénman R, Mäkitie AA, Monni O: **High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx.** *Gene Chromosome Canc* 2008, **47**:500–509.
15. Lo KC, Stein LC, Panzarella JA, Cowell JK, Hawthorn L: **Identification of genes involved in squamous cell carcinoma of the lung using synchronized data from DNA copy number and transcript expression profiling analysis.** *Lung Cancer* 2008, **59**:315–331.
16. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H, Ishikawa Y: **RET, ROS1, and ALK Fusions in Lung Cancer.** *Nat Med* 2012, **18**:378–381.
17. Inamura K, Togashi Y, Nomura K, Ninomiya H, Hiramatsu M, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y: **let-7 microRNA expression is reduced in bronchioloalveolar carcinoma, a non-invasive carcinoma, and is not correlated with prognosis.** *Lung Cancer* 2007, **58**:392–396.

18. Travis WD, Elisabeth B, Muller-Hermelink HK, Harris CC: *Pathology and Geneticsof Tumours of the Lung, Pleural, Thymus and Heart*. Lyon: IARC press; 2004.
19. AJCC: *Cancer Staging Manual*. 6th edition; 2002:167–177. Chapter 19; Lung - original pages.
20. Ishikawa Y, Furuta R, Miyoshi T, Satoh Y, Okumura S, Nakagawa K, Tsuchiya E: Loss of heterozygosity and the smoking index increase with decrease in differentiation of lung adenocarcinomas: etiologic implications. *Cancer Lett* 2002, **187**:47–51.
21. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S: A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005, **65**:6071–6079.
22. Ogawa S, Nanya Y, Yamamoto G: Genome-wide copy number analysis on GeneChip platform using copy number analyzer for affymetrix GeneChip 2.0 software. *Meth Mol Biol* 2007, **396**:185–206.
23. Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, Kawamata N, Hangaishi A, Kurokawa M, Chiba S, Gilliland DG, Koeffler HP, Ogawa S: Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 2007, **81**:114–126.
24. Danner BC, Gerdes JS, Jung K, Sander B, Enders C, Liersch T, Seipelt R, Gutenberg A, Gunawan B, Schöndube FA, Füzesi L: Comparison of chromosomal aberrations in primary colorectal carcinomas to their pulmonary metastases. *Canc Genet* 2011, **204**:122–128.
25. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, *et al*: Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007, **450**:893–898.
26. Kwei KA, Kim YH, Girard L, Kao J, Pacyna-Gengelbach M, Salari K, Lee J, Choi YL, Sato M, Wang P, Hernandez-Boussard T, Gazdar AF, Petersen I, Minna JD, Pollack JR: Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 2008, **27**:3635–3640.
27. Sasaki H, Hikosaka Y, Kawano O, Moriyama S, Yano M, Fujii Y: Evaluation of Kras gene mutation and copy number gain in non-small cell lung cancer. *J Thorac Oncol* 2011, **6**:15–20.
28. Reinersman JM, Johnson ML, Riely GJ, Chitale DA, Nicastri AD, Soff GA, Schwartz AG, Sima CS, Ayalew G, Lau C, Zakowski MF, Rusch VW, Ladanyi M, Kris MG: Frequency of EGFR and KRAS mutations in lung adenocarcinomas in African Americans. *J Thorac Oncol* 2011, **6**:28–31.
29. Kosaka T, Yatabe Y, Onozato R, Kuwano H, Mitsudomi T: Prognostic implication of EGFR, KRAS, and TP53 gene mutations in a large cohort of Japanese patients with surgically treated lung adenocarcinoma. *J Thorac Oncol* 2009, **4**:22–29.
30. Kohno T, Otsuka A, Girard L, Sato M, Iwakawa R, Ogiwara H, Sanchez-Cespedes M, Minna JD, Yokota J: A catalog of genes homozygously deleted in human lung cancer and the candidacy of PTPRD as a tumor suppressor gene. *Gene Chromosome Canc* 2010, **49**:342–352.
31. Veeriah S, Brennan C, Meng S, Singh B, Fagin JA, Solit DB, Paty PB P, Rohle D, Vivanco I, Chmielecki J, Pao W, Ladanyi M, Gerald WL, Liau L, Cloughesy TC, Mischel PS, Sander C, Taylor B, Schultz N, Major J, Heguy A, Fang F, Mellinghoff IK, Chan TA: The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc Natl Acad Sci USA* 2009, **106**:9435–9440.
32. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, Wang L, Montie JE, Rubin MA, Pienta KJ, Roulston D, Shah RB, Varambally S, Mehra R, Chinnaiyan AM: Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 2007, **448**:595–599.
33. Kato M, Sanada M, Kato I, Sato Y, Takita J, Takeuchi K, Niwa A, Chen Y, Nakazaki K, Nomoto J, Asakura Y, Akatsuka M, Hayashi Y, Mori H, Igarashi T, Kurokawa M, Chiba S, Mori S, Ishikawa Y, Okamoto K, Tobinai K, Nakagama H, Nakahata T, Yoshino T, Kobayashi Y, Ogawa S: Frequent inactivation of A20 in B-cell lymphomas. *Nature* 2009, **459**:712–716.
34. Sanada M, Suzuki T, Shih LY, Otsu M, Kato M, Yamazaki S, Tamura A, Honda H, Sakata-Yanagimoto M, Kumano K, Oda H, Yamagata T, Takita J, Gotoh N, Nakazaki K, Kawamata N, Onodera M, Nobuyoshi M, Hayashi Y, Harada H, Kurokawa M, Chiba S, Mori H, Ozawa K, Omine M, Hirai H, Nakauchi H, Koeffler HP, Ogawa S: Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* 2009, **460**:904–908.
35. Ninomiya H, Nomura K, Satoh Y, Okumura S, Nakagawa K, Fujiwara M, Tsuchiya E, Ishikawa Y: Genetic instability in lung cancer: concurrent analysis of chromosomal, mini- and microsatellite instability and loss of heterozygosity. *Br J Cancer* 2006, **94**:1485–1491.
36. Yoshino I, Fukuyama S, Kameyama T, Shikada Y, Oda S, Maehara Y, Sugimachi K: Detection of loss of heterozygosity by high-resolution fluorescent system in non-small cell lung cancer: association of loss of heterozygosity with smoking and tumor progression. *Chest* 2003, **123**:545–550.
37. Yohena T, Yoshino I, Takenaka T, Ohba T, Kouso H, Osoegawa A, Hamatake M, Oda S, Kuniyoshi Y, Maehara Y: Relationship between the loss of heterozygosity and tobacco smoking in pulmonary adenocarcinoma. *Oncol Res* 2007, **16**:333–339.
38. Le Calvez F, Mukeria A, Hunt JD, Kelm O, Hung RJ, Tanière P, Brennan P, Boffetta P, Zaridze DG, Hainaut P: TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. *Cancer Res* 2005, **65**:5076–5083.
39. Gazdar AF, Shigematsu H, Herz J, Minna JD: Mutations and addiction to EGFR: the Achilles 'heal' of lung cancers? *Trends Mol Med* 2004, **10**:481–486.
40. Woodburn JR: The epidermal growth factor receptor and its inhibition in cancer therapy. *Pharmacol Ther* 1999, **82**:241–250.
41. Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL, Rossi E, Ludovini V, Gregorc V, Toschi L, Franklin WA, Gazdar AF CL, Bunn PA Jr, Hirsch FR: Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol* 2005, **23**:5007–5018.
42. Varella-Garcia M, Mitsudomi T, Yatabe Y: EGFR and HER2 genomic gain in recurrent non-small cell lung cancer after surgery: impact on outcome to treatment with gefitinib and association with EGFR and KRAS mutations in a Japanese cohort. *J Thorac Oncol* 2009, **4**:318–325.
43. Pugh TJ, Bebb G, Barclay L, Sutcliffe M, Fee J, Salski C, O'Connor R, Ho C, Murray N, Melosky B, English J, Vielkind J, Horsman D, Laskin JJ, Marra MA: Correlations of EGFR mutations and increases in EGFR and HER2 copy number to gefitinib response in a retrospective analysis of lung cancer patients. *BMC Canc* 2007, **7**:128.
44. Blons H, Pallier K, Le Corre D, Danel C, Tremblay-Gravel M, Houdayer C, Fabre-Guillevin E, Riquet M, Dessen P, Laurent-Puig P: Genome wide SNP comparative analysis between EGFR and KRAS mutated NSCLC and characterization of two models of oncogenic cooperation in non-small cell lung carcinoma. *BMC Med Genom* 2008, **1**:25.
45. Zhang A, Zheng C, Lindvall C, Hou M, Ekedahl J, Lewensohn R, Yan Z, Yang X, Henriksson M, Blennow E, Nordenskjöld M, Zetterberg A, Björkholm M, Gruber A, Xu D: Frequent amplification of the telomerase reverse transcriptase gene in human tumors. *Cancer Res* 2000, **60**:6230–6235.
46. Saretzki G, Petersen S, Petersen I, Kölble K, von Zglinicki T: hTERT gene dosage correlates with telomerase activity in human lung cancer cell lines. *Cancer Lett* 2002, **176**:81–91.
47. Kang JU, Koo SH, Kwon KC, Park JW, Kim JM: Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Canc Genet Cytogenet* 2008, **182**:1–11.
48. Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, Chang GC, Wu T, Park JY, Hsiao CF, Kim YH, Shen H, Seow A, Yeager M, Tsai YH, Kim YT, Chow WH, Guo H, Wang WC, Sung SW, Hu Z, Chen KY, Kim JH, Chen Y, Huang L, Lee KM, *et al*: The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet* 2010, **6**:e1001051.
49. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njølstad

I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, *et al*: **Lung cancer susceptibility locus at 5p15.33.** *Nat Genet* 2008, **40:**1404–1406.

50. Camidge DR, Theodoro M, Maxson DA, Skokan M, O'Brien T, Lu X, Doebele RC, Barón AE, Varella-Garcia M: **Correlations between the percentage of tumor cells showing an anaplastic lymphoma kinase (ALK) gene rearrangement, ALK signal copy number, and response to crizotinib therapy in ALK fluorescence in situ hybridization-positive nonsmall cell lung cancer.** *Cancer* 2012, **118:**4486–4494.