Multiplex Diagnosis of Oncogenic Fusion and MET Exon Skipping by Molecular Counting Using Formalin-fixed Paraffin Embedded Lung Adenocarcinoma Tissues

Abstract

Background: Fusions of the $\textit{ALK}$, $\textit{RET}$, $\textit{ROS1}$, $\textit{BRAF}$, and $\textit{NRG1}$ genes and intronic $\textit{MET}$ mutations are druggable oncogene alterations in lung adenocarcinoma (LADC) that cause expression of aberrant transcripts. Because these aberrant transcripts are both infrequent (less than 5%) and mutually exclusive, multiplex assays are required to detect them in tumor samples.

Samples and methods: Aberrant transcripts of the six oncogenes listed above (36 transcripts in total) were examined in a molecular counting (MC) assay, which counts RNA molecules via the simultaneous hybridization of several probes. Forty-one surgical resected LADC tissue samples found to express one of these aberrant oncogenic transcripts upon whole transcriptome sequencing (test cohort: n = 22) or RT-PCR (validation cohort: n = 19) analyses were subjected to MC, followed by biopsied tumor tissues.

Results: Threshold values for the diagnosis of each of the 36 transcripts were determined in frozen and formalin-fixed paraffin embedded (FFPE) samples from the test cohort. Based on these threshold values, the MC assay diagnosed expression of oncogenic transcripts in the validation cohort samples with 100% accuracy. The assay also accurately detected oncogenic fusions in bronchial lavage fluid and transbronchial biopsies.

Conclusions: The MC assay enables multiplex detection of oncogenic fusion and exon-skipped transcripts in tumor samples, including FFPE samples, obtained in the clinic. (207 words)

Key words: personalized medicine, molecular counting, oncogene fusion, MET exon skipping, genetic diagnosis
INTRODUCTION

Analyses of lung adenocarcinoma (LADC) genomes have identified several oncogene alterations that cause expression of aberrant transcripts in tumor cells. A representative example is the fusion of oncogenes to other partner genes, which causes constitutive activation of protein kinases encoded by these oncogenes.\(^1,^2\) These alterations are a mutually exclusive driver mutation in lung carcinogenesis and are a target for therapies based on protein tyrosine kinase inhibitors (TKIs). A well-known example is the oncogenic \(EGFR\) mutation, which is a target for therapy using EGFR-TKIs. Indeed, LADCs harboring the \(ALK\) oncogene fusion (3–5% of all LADC cases) respond very well to ALK inhibitors such as crizotinib,\(^3\) ceritinib,\(^4\) and alectinib.\(^5\) \(RET\) and \(ROS1\) oncogene fusions are also rare (1–3% of LADC\(^6\)-\(^10\) in each case), but they are a promising target for therapy using TKIs. Several clinical trials focusing on \(RET\) and \(ROS1\) fusion-positive LADC have been undertaken, and the results of these studies were recently reported.\(^11,^12\) In addition, there is a report of single case of \(RET\) fusion-positive LADC who responded to \(RET\) TKI.\(^13\) These reports indicate the therapeutic effects of \(RET\) and \(ROS1\) TKIs. Recently, we and others identified \(BRAF\) and \(NRG1\) oncogene fusions as another therapeutic target in ~10% of patients with invasive mucinous adenocarcinoma of the lung.\(^14,^15\) Recent studies also identified another type of oncogene alteration that generates aberrant oncogenic transcripts in approximately 3% of LADCs: skipping of the coding exon 14 in \(MET\) oncogene transcripts due to aberrant splicing, which is mainly associated with intronic mutations.\(^16,^17\) The skipped transcript produces a constitutively active MET protein lacking the CBL binding domain that negatively regulates MET kinase. Indeed, in recent reports, a MET TKI showed antitumor activity in patients with LADCs expressing exon 14-skipped \(MET\) transcripts,\(^18,^19\) suggesting that this alteration is also a promising therapeutic target.

Precision medicine for LADC patients based on the TKIs described above requires the
diagnosis of multiplex oncogene aberrations in tumor cells. Since most patients subjected to TKI therapy are inoperable, diagnosis must be performed using small amounts of biopsied tissues or archived surgical specimens that have been formalin-fixed and paraffin embedded (FFPE) for histopathological examination. In fact, techniques such as fluorescence in situ hybridization (FISH), which detects gross chromosome rearrangements, are often used to diagnose gene fusions in these clinical specimens. However, the recent discovery of a number of uncommon but druggable oncogene fusions in LADC, along with the finding that they are mutually exclusive, means that multiplex assays that enable the simultaneous identification of several alterations are required. Furthermore, in contrast to hot spot mutations such as EGFR, KRAS, and BRAF, which activate oncogenes in tumor cells, the presence of highly diverse genomic fusion points located within intron sequences hampers the detection of gene fusions by multiplex genome PCR-based assays using next-generation sequencers. In addition, identifying the intronic mutations responsible for MET exon skipping using genomic DNA is difficult due to their highly diverse locations and the occurrence of passenger mutations. Multiplex reverse transcription (RT)-PCR-based assays using tumor RNA are often used to detect aberrant transcripts caused by gene fusions and exon skipping. This is because there are few aberrant transcript patterns, despite the diversity with respect to the location of the responsible mutations. However, the poor quality of the RNA obtained from FFPE tumor tissues, particularly from archived tissues, makes it difficult to develop reliable RT-PCR-based multiplex assays.

Molecular counting (MC) using the nCounter system (NanoString Technologies, Seattle, WA) is a method of quantifying RNA that is not based on RT and PCR; this method directly counts RNA molecules via the simultaneous hybridization of multiple probes. Indeed, previous studies demonstrate the feasibility of the method for quantifying RNA expression and detecting gene fusions. Here, we used this method to detect 36 oncogenic transcripts, 35 of which were derived
from five oncogenes (ALK, RET, ROS1, NRG1, and BRAF) fused to partner genes and the MET exon-skipped transcript above. To examine the feasibility of this method, RNA samples obtained from 62 LADC tissues samples from 41 patients (39 FFPE and 23 snap-frozen tissues in total; see Fig. 1B) were subjected to MC. These samples were selected from 608 consecutive LADC patients because whole RNA sequencing or RT-PCR analyses of snap-frozen tissues revealed that they either did or did not express aberrant oncogenic transcripts.

MATERIALS AND METHODS

Subjects
Consecutive LADC cases who underwent surgical resection between 1997 and 2008 at the National Cancer Center Hospital, Tokyo (n = 608) were screened for EGFR, KRAS, BRAF, and HER2 hot spot mutations using the high resolution melting (HRM) method using genomic DNA obtained from snap-frozen tumor tissues.24 The EML4- and KIF5B-ALK, KIF5B- and CCDC6-RET, and CD74-, EZR-, and SLC34A2-ROS1 fusions were examined by RT-PCR of RNA obtained from snap-frozen tumor tissues as previously described.8,10,25 Histological diagnosis was based on the latest edition of the World Health Organization classification of lung tumors.26 RNA sequencing was performed on the 160 cases that yielded sufficient RNA to examine whether they expressed oncogene fusions and exon-skipped MET transcripts. The results were confirmed by RT-PCR as previously described.21

Sixty-four LADC cases of ALK, RET, ROS1, NRG1, and BRAF gene fusions or MET exon skipping (Fig. 1A) were identified. After excluding 23 cases due to insufficient sample, the remaining 41 were divided into two cohorts and subjected to MC analysis. The test cohort included 22 cases in which whole RNA sequencing using next-generation sequencing had been performed on RNA obtained from snap-frozen tumor tissues. Of these, 13 frozen tissue RNA samples and 21 FFPE tissue RNA samples were available for study (Fig. 1B). The validation cohort comprised 19 cases in
which RT-PCR analysis had been performed to examine the presence of ALK, RET, ROS1, NRG1, and BRAF fusions and MET exon skipping. Of these, 10 frozen tissue RNA samples and 18 FFPE tissue RNA samples were available for study (Fig. 1B). In addition, two FFPE tumor tissues obtained by transbronchial biopsy and diagnosed as positive for the RET and ROS1 fusions, respectively, and 18 bronchial lavage samples (including two harboring the EML4-ALK fusion) that were obtained by endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA), were also tested in the MC assay. This study was approved by the institutional review board of the National Cancer Center and all patients gave written informed consent.

In snap-frozen tissue samples, the percentage of tumor cells was deduced to range from 10% to 90% (mean=22%) using the Global Parameter Hidden Markov Model method based on somatic mutation and allelic imbalance data as previously described. The results were consistent with those of FFPE tissue samples estimated by hematoxylin and eosin staining. Thus, the percentage of tumor cells in samples in the present study was indicated to be >10%.

**RNA Preparation**

RNA was extracted from FFPE tissue sections (five sections, each 5 µm thick) using the RNeasy FFPE kit (QIAGEN, Hilden, Germany) or from lavage fluid using the RNeasy Mini kit (QIAGEN). Total RNA was extracted from grossly dissected, snap-frozen tissue samples using TRIzol (Invitrogen, Carlsbad, CA), according to the manufacturer’s instructions. RNA was quantified using the Nanodrop 8000 (Thermo Scientific, Wilmington, DE).

**MC Analysis**

The nCounter system uses a set of probes to detect transcripts: a single 5’ reporter probe (~50-mer) linked to a fluorescence barcode tag (6-mer) and a biotinylated 3’ capture probe (~50-mer). Seventy
sets of probes (Table S1) were designed to simultaneously detect 36 aberrant transcripts, i.e., 35 fusion transcripts for 5 oncogenes and 1 exon 14-skipped MET transcript (Table 1). The probes were synthesized by NanoString Technologies (Seattle, WA). Thirty-eight of the seventy sets were designed to hybridize with chimeric portions of the fusion transcripts, while the other 24 sets were designed to hybridize with wild-type ALK, RET, and NRG1 transcripts to examine gene fusion-induced imbalances in the expression of the 5’ and 3’ regions of the transcripts. Eight sets of probes were designed to hybridize with internal control genes: GAPDH, OAZ1, POLR2A, and GUSB.

The nCounter assay was performed according to the manufacturer’s instructions. Briefly, 150 ng and 400 ng of RNA from fresh frozen and FFPE tissues, respectively, were used in the assay to ensure that the counts for an internal control gene, GAPDH, were >5000 per sample. RNA was hybridized with the probe sets for 16 hours at 67°C. Samples were processed using an automated nCounter Sample Prep Station (NanoString Technologies, Inc.). Cartridges containing immobilized and aligned reporter complexes were subsequently imaged on an nCounter Digital Analyzer (NanoString Technologies, Inc.) set at a data resolution of 555 fields of view. Reporter counts were collected and normalized using nSolver analysis software version 2.0 (NanoString Technologies, Inc.).

**Determination of Gene Fusion and Exon Skipping**

Data were normalized in two steps as previously described. First, six positive internal controls were used to remove potential systemic differences between individual hybridization experiments. The sum of the signal intensity ($S_i$) from the six positive control probes was calculated for sample $i$ and then scaled using the normalization factor, $S_{mean}/S_i$. Second, the scaled signal intensity of sample $i$ was further normalized using four housekeeping genes to remove any effects that might be attributable to differences in the amount of input RNA. For example, if $H_i$ represents the genes in
sample \( i \), the second normalization factor would be defined as \( H_{\text{mean}}/H_i \). 

Previous studies suggest that aberrations in the six oncogenes examined herein are mutually exclusive.\textsuperscript{28,29} This was supported by the whole transcriptome sequencing results for the test cohort samples examined; therefore, samples expressing one particular aberrant oncogene transcript were used as negative controls for the other aberrant oncogene transcripts. Threshold values used to determine the presence/absence of a particular aberrant transcript were defined as the mean plus five standard deviations (SDs) above the value for tumor samples harboring aberrations in other transcripts. Based on these threshold values, the false-positive detection rate was predicted to be less than 0.0002%. Threshold values were set for both fresh frozen and FFPE tissues. To determine imbalances in the expression of the 5' and 3' regions of gene transcripts, the 3' overexpression ratio was calculated as follows: 3'-imbalanced probe counts/5'-imbalanced probe counts. Threshold values for 3' overexpression for a particular gene were defined as the mean value plus five SDs above the value for tumor samples with aberrations in other genes.

**Immunostaining for the MET Protein**

The Bench-Mark XT automated slide processing system (Ventana, Tucson, AZ) was used to stain for c-MET, as previously described.\textsuperscript{30} In brief, after tissue sections were deparaffinized using EZ Prep (Ventana), heat-induced epitope retrieval with CC1 (Ventana) was performed, and the slides were incubated with primary antibodies against c-MET (CONFIRM Anti-Total c-MET, clone SP44; Ventana). Immunoreactivity was detected using the UltraView DAB Universal Detection Kit, followed by counter-staining with Hematoxylin II (Ventana) and Bluing Reagent (Ventana). Immunopositive cases were defined as those exhibiting cytoplasmic and/or membranous staining in more than 10% of cells.
RESULTS

Test and Validation Cohorts

Forty-one LADCs, each expressing one aberrant oncogenic transcript, were selected from 608 consecutive surgical cases and divided into two cohorts (22 in the test cohort and 19 in the validation cohort). Samples were then examined in the MC assay (Fig. 1A). RNAs from snap-frozen tissues from the 22 test cohort samples had already been subjected to whole transcriptome sequencing; therefore, each expressed 1 of the 36 aberrant transcripts (Table 1). RNA from snap-frozen tissues from the 19 validation cohort samples had been subjected to RT-PCR; therefore, each expressed 1 of the 36 aberrant transcripts.

MC Assay to Determine the Expression of Aberrant Oncogene Transcripts

The expression of 23 fusion transcripts and 1 skipped MET transcript was ascertained by counting the signals generated by the fusion probes alone. However, the fusion probes were not able to discriminate nine fusion transcripts that shared 5’-partner gene exons. In these cases, oncogene fusion was determined from the results of fusion probe assays in conjunction with the results from the 3'/5' imbalance assays (Fig. 1C).

The 22 test cohort cases expressed 13 aberrant transcripts: 12 fusion transcripts and 1 exon 14-skipped MET transcript (Table 1). For 13 of the 22 cases, RNAs from snap-frozen tissues were available and were subjected to MC. The background counts were calculated from data obtained for negative samples, and the threshold count for each aberrant transcript was set as the mean plus five SDs (as described in the Materials and Methods section). Indeed, using these criteria, all 13 samples were verified as \( \log_2 \left[ \frac{\text{fusion probe counts}}{\text{threshold counts}} \right] > 0 \), and the results of the imbalance assays for the RET and NRG1 genes enable us to judge KIF5B-RET (K15:R12) and CD74-ROS1 (C6:R34), respectively (Supplementary Fig. S1A). The threshold values for 39 FFPE
samples was re-calculated in the same way; the results clearly distinguished between cases that were positive or negative for gene fusion or \( MET \) skipping (Supplementary Fig. S1B). The expression values determined by MC showed a good correlation with those determined by RNA sequencing, indicating the quantitative nature of the MC assay (Supplementary Fig. S2).

**MC Assay of the Validation Cohort Samples**

The 19 cases in the validation cohort were next tested in the MC assay using the same threshold values and the criteria set from the test cohort. These cases expressed 11 aberrant transcripts: ten fusion transcripts and one exon 14-skipped \( MET \) transcript (Table 1). The 19 cases yielded 10 RNA samples from snap-frozen tissues and 18 from FFPE tissues. As expected, all 19 cases were accurately judged as expressing aberrant transcripts in both snap-frozen and FFPE samples (Fig. 3A, B). In one \( NRG1 \) fusion-positive case (NR-002), \( CD74 \_ex6 \) and \( CD74-NRG1 \) (C8:N6) were the major transcripts detected. These results were consistent with those from the RT-PCR, which showed that \( CD74-NRG1 \) (C6:N6) was the minor transcript co-expressed with \( CD74-NRG1 \) (C8:N6) as the major transcript in this case.

**MC Assay of Non-surgical Samples**

The MC assay was next used to detect aberrant transcripts in non-surgical specimens obtained from the daily lung cancer clinic and used for pathological diagnosis. Two FFPE transbronchial biopsy specimens, which were diagnosed as \( CCDC6-RET \) and \( CD74-ROS1 \) fusion by RT-PCR and FISH, respectively, were subjected to MC. The results clearly showed that these FFPE tissues were positive for both of these fusions (Fig. 3A). In addition, 18 bronchial lavage samples were also subjected to MC analysis. The results showed that two \( EML4-ALK \) fusion-positive samples were diagnosed correctly (Fig. 3B).
DISCUSSION

Here, we describe an MC assay that enables the multiplex diagnosis of druggable aberrations for six oncogenes: \textit{ALK}, \textit{RET}, \textit{ROS1}, \textit{BRAF}, and \textit{NRG1} fusions, and \textit{MET} exon 14 skipping, which are not detectable by NGS-based hot spot sequencing of genomic DNA due to the highly diverse locations of the fusion and mutation points. Samples that had already been identified as either positive or negative were used to set the threshold values for the MC assay. The assay accurately diagnosed aberrant expression of transcripts for the six druggable oncogenes in both frozen and FFPE samples. These results demonstrate the utility of MC assay for detecting gene fusions\textsuperscript{22,23} by increasing the number of oncogenes examined to date. In addition, we showed that the MC assay can be used to analyze non-surgical specimens (e.g., transbronchially biopsied and bronchoalveolar lavage samples) that are obtained in the daily lung cancer clinics and used for pathological diagnosis. Notably, the aberrations in the six oncogenes examined in the present study were mutually exclusive; they were also mutually exclusive with hot spot-activating mutations in the \textit{EGFR}, \textit{HER2}, and \textit{BRAF} oncogenes in our LADC cohort (Supplementary Fig. 3). The high accuracy and robustness of the MC assay, even when used for FFPE tumor tissues, indicate that it will complement other diagnostic assays used to detect hotspot \textit{EGFR}, \textit{HER2}, and \textit{BRAF} oncogene mutations, and thereby improve the diagnosis of LADC. Other novel druggable (but infrequent) oncogene fusions have been identified in lung cancer.\textsuperscript{31-34} The fact that the probe set used for the MC assay can be easily tailored to the genes being examined is a great advantage. On the other hand, the MC assay has a limitation in that it required a relatively high amount of RNA (150 ng for frozen tissues and 400 ng for FFPE tissues), indicating that it is not applicable for samples from which an insufficient amount of RNA is obtained.
The present study also focused on MET exon 14 skipping, a novel druggable aberration being examined in clinical trials as a therapeutic target.\textsuperscript{18,19} Among seven such cases, we found that two were negative after immunostaining FFPE tumor tissues for the MET protein (Supplementary Table 2; Supplementary Fig. 4); therefore, immunostaining is not a suitable method for detecting MET activation caused by exon skipping. Aberrant RNA splicing associated with intronic mutations may be a mechanism that drives carcinogenesis by activating oncogenes and inactivating tumor suppressor genes. Thus, the MC assay will be suitable for detecting aberrantly spliced transcripts as well as gene fusions.

In summary, we have developed an assay that enables the simultaneous diagnosis of multiple druggable oncogene aberrations in FFPE lung cancer tissues with a high degree of accuracy and robustness. The advantage of this assay is that probe sets can be changed easily to accommodate changes in the genes being investigated. The MC assay can be used not only to detect aberrant transcripts but also to examine gene expression levels, as exemplified previously by PAM50, a Food and Drug Administration-approved assay used to classify intrinsic breast cancer subtypes.\textsuperscript{35} Thus, multiplexed assays that simultaneously detect aberrant transcripts and measure the expression levels of other druggable genes are worth developing.

REFERENCES


Figure Legends

FIGURE 1.

(A) Selection of study subjects. Forty-one lung adenocarcinoma cases harboring oncogenic gene fusions or exon skipping were selected from 608 consecutive patients that underwent surgical resection. Patients were classified into two cohorts. Cases in which RNA isolated from snap-frozen tumor samples had been analyzed by whole transcriptome (RNA) sequencing were included in the test cohort (n = 22), while those in which RNA isolated from snap-frozen tumors samples had been analyzed by RT-PCR were included in the validation cohort (n = 19).

(B) Sample characteristics and molecular counting (MC) assay results. RNA was extracted from fresh frozen and/or FFPE tissues. Status was determined by counting the signals generated by fusion probes (F) or those generated by fusion and imbalance probes (FI). Black box: Diagnosis based on counts generated by the fusion probe only. Gray box: Diagnosis based on counts generated by the FI probes.

(C) Strategy for diagnosing gene fusions and MET exon skipping. Left: 26 cases of gene fusion and 1 case of MET exon 14 skipping were diagnosed based on counts generated by the fusion probes alone. Right: Nine cases of gene fusion were diagnosed based on counts generated by both the FI probes.

FIGURE 2: Results for the validation cohort. The results for fresh frozen samples (A) and FFPE samples (B) are shown. Upper: The reporter counts for the fusion probes are expressed as log₂ (fusion probe counts/threshold counts). Values are shown when the counting results were greater than the threshold. Lower: The results of 3’ overexpression are expressed as log₂(3’/5’ probe count ratio/ 3’/5’ probe count threshold)
FIGURE 3: Results for the non-surgical samples. The results from biopsy samples (A) and lavage samples (B) are shown. The reporter counts of the fusion probes are expressed as $\log_2$ (fusion probe counts/threshold counts). Values are shown when the counting results were greater than the threshold.
**Supplementary Fig. S1.** Counts for the test cohort. The results for fresh frozen samples (A) and FFPE samples (B) are shown. Upper: The reporter counts for the fusion probes are expressed as log$_2$ (fusion probe counts/threshold counts). Values obtained when the counting results were greater than the threshold are shown. Lower: The results for 3’ overexpression are expressed as log$_2$ (3’/5’ probe count ratio/3’/5’ probe count threshold).

**Supplementary Fig. S2.** The correlation of the expression levels of \( RET \) fusion transcripts between those estimated by RNA sequencing and those estimated by the MC assay. Four \( RET \) fusion-positive test cohort samples, which had been analyzed by the TopHat-fusion algorithm in our previous study\(^2\), were subjected to the correlation analysis. The levels of fusion transcripts in RNA sequence data were expressed as the number of sequencing reads spanning the partner and \( RET \) genes normalized by the total read numbers among the four samples, while those in the MC assay were expressed as the fusion probe counts normalized by the counts of housekeeping genes. The correlation was estimated by calculating the \( R^2 \) value.

**Supplementary Fig. S3.** Pie chart showing oncogenic aberrations that occur in LADC in a mutually exclusive manner. Surgical specimens from 319 stage I–II lung adenocarcinomas, which were selected from the original cohort of 608 based on the availability of frozen tumor tissues for genome/transcriptome analyses, were subjected to analysis.\(^8\)

**Supplementary Fig. S4.** Representative images of a c-MET-positive (A) and -negative (B) case of LADC showing exon 14 skipping of the \( MET \) gene. In c-MET-positive cases, both the cell membrane and cytoplasm are stained brown (original magnification, ×200).