A Highly Specific and Sensitive Massive Parallel Sequencer-based Test for Somatic Mutations in Non-Small Cell Lung Cancer.

Yohiaki Inoue, MD[1, 2], Jun Shiihara, MD, PhD[3], Hitoshi Miyazawa, MS[4], Hiromitsu Ohta[4], MD, PhD, Yoshiaki Nagai, MD[4], PhD, Kunihiko Kobayashi, MD, PhD[5], Mitsuo Nakayama, MD, PhD[2], and Koichi Hagiwara, MD, PhD[4, 5*]

[1]Graduate School, Saitama Medical University, 38 Morohongo, Moroyama, Saitama 350-0495, Japan; [2]Depertment of General Thoracic Surgery, Saitama Medical Center, 1981 Kamoda, Kawagoe, Saitama 350-8550, Japan; [3]Depertment of Respiratory Medicine, Graduate School of Medicine, Tohoku University, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-0872, Japan; [4]Depertment of Respiratory Medicine, Saitama Medical University, 38 Morohongo, Moroyama, Saitama 350-0495, Japan; Department of Respiratory Medicine, Saitama International Medical Center, 1397-1 Yamane, Hidaka, Saitama 350-1298, Japan; [5]Clinical Laboratories for Cardiovascular Diseases, Jichi Medical University, 1-847 Amanuma-cho, Oomiya-ku, Saitama-shi, Saitama 330-8503, Japan.

Correspondence are addressed to:
Koichi Hagiwara
Clinical Laboratories for Cardiovascular Diseases
Saitama Medical Center, Jichi Medical University
1-847 Amanuma-cho, Oomiya-ku, Saitama-shi, Saitama 330-8503, Japan
Phone: +81-48-647-2111 ext 6701
FAX: +81-48-644-6262
E-mail: hagiwark@me.com

Word count: 3534
Number of Figures: 1
Number of Tables: 4
Quantity of Supplemental Data: 6
Doc S1, Doc S2, Doc S3, Table S1, Fig. S1, Fig. S2.

Summary

Cancer treatment is increasingly being determined based on mutations for molecular targeting therapy. Therefore, it is critical to establish accurate, sensitive, and specific tests that can scan for mutational hot spots in patients to determine the most effective course of treatment. We here report the development of the Mutation Investigator using Next-era Sequencer (MINtS) system, which was designed to screen for mutation hot spots related to non-small cell lung cancer (mutations in the *EGFR*, *KRAS*, and *BRAF* genes). We focused on reducing the false-mutant read rate, which is of particular relevance to patients, as the resulting false-positive for a mutation would lead to a wrong target and an ineffectual treatment course. The MINtS system shows high specificity and sensitivity (≥0.99) even applied to samples with a cancer cell content of 1%, and is relatively inexpensive involving a straightforward 5-step protocol. The MINtS was capable of correctly detecting 6 different mutations in the admixtures simulating samples with a cancer cell content of 1%. Out of 96 clinical specimens of non-small cell lung cancer including both tissue and cytological samples, the MINtS detected *EGFR* mutation in 27 (28%), *KRAS* mutation in 18 (19%), and *BRAF* mutation in 2 (2%). The statistical algorithm and protocol that we present herein is expected to be a resource for clinical practice to help determining targeted and individualized treatment strategies.

Keywords

BRAF Kinases, erbB-1, Genes, K-ras Genes, Mutation, Oncogenes,

Introduction

Recent progress in the molecular targeting therapy for lung cancer has consolidated the importance of mutation testing for choosing appropriate treatment regimens [1]. Accordingly, epidermal growth factor receptor (*EGFR*) mutation test has been introduced in clinical practice in many countries. Several drugs targeting different cancer driver genes [2] are already on the market, and many similar drugs are expected to follow suit in the near future. There is an increasing need for multiple-gene mutation tests that are clinically applicable, and can simultaneously screen all somatic mutations that might predict the response to drugs currently available or expected to be available. The massive parallel sequencer is a particularly attractive instrument for the development of such a multiple-gene mutation test.

In the current study, we established a multiple-gene mutation test employing a massive parallel sequencer. We then used this test to investigate cancer driver genes whose mutations are relevant to non-small cell lung cancer (NSCLC) and are detectable in genomic DNA. The clinical setting for the mutation testing varies among countries. The specifications of our test were determined based on the Japanese consensus established for the *EGFR* mutation test [3]. However, we expect that slight modification of the test will make it possible to apply our system to clinical settings in many countries. In particular, the Japanese consensuses adopted for our system are as follows. First, the test is applicable to a sample for which the cancer cell content is as low as 1%, because the cancer cell content has been reported to be $\geq 1\%$ in most "pathologically cancer-positive" samples [4]. Second, the test is applicable to both cytological and tissue samples, because 1/3 of the samples submitted to the mutation test are cytological samples, and 2/3 are tissue samples in Japan [3]. Third, the preferred sample type is

frozen samples or samples that were immediately stored in a preservative (e.g., RNAlater^{R}). Because nucleic acids in formalin-fixed, paraffin-embedded (FFPE) samples are often severely degraded or chemically modified [5-8], the use of FFPE samples are discouraged. There has been no consensus established for the specificity and sensitivity of the *EGFR* mutation test, and we thus established a criterion of our own: sensitivity and specificity ≥ 0.99 (i.e., false-positive rate and false-negative rate ≤ 0.01). Errors in the mutation test result in assignment of suboptimal treatment to the patients, thus significantly affecting the patients' outcomes [1]. Therefore, from the patients' perspective, a false-positive rate or a false-negative rate > 0.01 is considered unacceptable.

Determining the sensitivity and specificity of multiple-gene mutation tests requires special consideration. Because the test investigates multiple mutation hot spots simultaneously, errors occurring in each spot can accumulate and inflate the error rate of the whole test. Therefore, the false-positive rate and false-negative rates for each mutation hot spot should be much smaller than 0.01.

Here, we report the construction of the Mutation Investigator using Next-era Sequencer (MINtS) system, which investigates the mutations in *EGFR* [9 10 11], Kirsten rat sarcoma viral oncogene homolog gene (*KRAS*) [12], and v–raf murine sarcoma viral oncogene homolog B1 gene (*BRAF*) [13 14] from a sample with a cancer cell content of 1%, with specificity and sensitivity ≥ 0.99. The reactions and statistical algorithm of the MINtS will be a prototype on which we are able to construct a more comprehensive multiple-gene mutation test.

Materials and Methods

Definition of Terms

Read: A "read" refers to a stretch of nucleotide sequence determined by the sequencer. A single read corresponds to the nucleotide sequence of a single DNA molecule. In the current study, polymerase chain reaction (PCR) amplicons were sequenced. Therefore, a single read corresponds to a nucleotide sequence of a single molecule of the PCR-amplified fragments.

Informative read: In the MINtS, a PCR amplified fragment is sequenced in both strands producing a sense-strand read and an anti-sense strand read. When both reads have a high quality score (i.e., Phred score ≥ 30) and have a matched sequence, the sense-strand read is an informative read. An informative read scarcely have errors due to the sequencer, although it still has errors occurring during PCR.

False-mutant read: A "false-mutant read" is an informative read that has a mutant sequence, but nevertheless they originated from the normal genomic sequence. A false-mutant read can occur from a DNA replication error during PCR. The false-mutant reads cause false positive results in the mutation tests.

True-mutant read: A "true-mutant read" is an informative read that has a mutant sequence, and they actually originated from a mutant genomic sequence.

Mutant reads: "Mutant-reads" refer to all false-mutant reads plus the true-mutant reads. It is impossible to discriminate the former from the latter.

Reference frequency of the falsely-mutant reads: The reference frequency of the false-mutant reads was determined as the greatest frequency of the false-mutant reads in 34 samples for each mutation hot spot, and was used in the statistical analysis.

Allowable false positive rate: The allowable false-positive rate is the false-positive rate

allowed for each mutation hot spot to maintain the false-positive rate of the whole test ≤ 0.01.

Allowable false negative rate: The allowable false-negative rate is the false-negative rate allowed for each mutation hot spot to maintain the false-negative rate of the whole test ≤ 0.01.

Samples

Normal human genomic DNA isolated from the immortalized B lymphocyte cell lines established from Japanese volunteers was purchased from the Japanese Collection of Research Bioresources (Osaka, Japan). Clinical samples were collected at either the Saitama Medical University Hospital or the Saitama International Medical Center. Samples containing cancer cells were identified as described previously [4 3]. Briefly, cytological samples were suspended in saline at the site of isolation and were divided into 2 parts. One part was submitted to the pathological department. When the pathologist confirmed the presence of cancer cells, the other part, which had been stock frozen or had stocked in a preservative, was subjected to the mutation test. Tissue samples were serially thin sectioned. When pathologist confirmed the presence of cancer cells in one section, the others were used for the mutation test. Plasmids containing G719S, DelE746-A750, T790M, or L858R mutations of *EGFR*, G12S, or Q61K mutations of *KRAS*, and G469A, or V600E mutations of *BRAF* were artificially synthesized. A 1% admixture sample was prepared to simulate a clinical sample with a cancer cell content of 1% for each mutation by mixing each plasmid with human placental DNA (Promega, Madison, WI, USA) at a molar ratio of 1:200 [3].

List of the mutation hot spots

Table 1 lists the mutation hot spots investigated by the MINtS. Each mutation is known to be relevant to the response to the molecular-targeting drugs.

PCR amplification and sequencing

DNA was subjected to a multiplex PCR to amplify each mutation hot spot (**Doc. S1** and **Doc. S2**). Adaptors and indexes were added by 2 serial PCRs, named add-adaptor PCR and add-index-adaptor PCR, respectively. Here, adaptors are short stretches of nucleotides 24- to 34-bp long, and utilized by the MiSeq during sequencing reaction. Indexes are short stretches of nucleotides 8-bp long, and utilized for the discrimination of multiple patients because different indexes were attached to the amplicons from different patients (**Doc. S2**). The amplified products of the add-index-adaptor PCR from multiple patients were mixed, and sequenced on the MiSeq (Illumina; San Diego, CA, USA) in a single run (**Fig. 1**).

For multiplex PCR, 10 ng of genomic DNA was amplified in a 25-μL solution containing multiplex PCR primer mix (250 nM each), 1 x KOD Plus buffer, 200 nM dNTPs, 1 mM $MgSO_4$, and 0.5 units of KOD polymerase (Toyobo; Osaka, Japan). The PCR was performed with a 94°C hold for 120 s followed by 28 cycles of 94°C for 15 s, 62°C for 30 s and 68°C for 30 s. For add-adaptor PCR, a 1-μL sample of the multiplex PCR was add to a 24-μL solution containing add-adaptor PCR primers (50 nM each), 1× KOD Plus buffer, 200 nM dNTPs, 1 mM $MgSO_4$, and 0.5 units of KOD polymerase. The PCR was performed with a 94°C hold for 120 s followed by 6 cycles of 94°C for 15 s, 62°C for 30 s and 68°C for 30 s. For the add-adaptor PCR reaction from different patient was amplified using different pair of add-adaptor PCR primers, where index

sequences were used to discriminate samples from different patients. A 1-μL sample of the add-adaptor PCR was added to a 24-μL solution containing add-index-adaptor PCR primers (200 nM each), 1× KOD Plus buffer, 200 nM dNTPs, 1 mM $MgSO_4$, and 0.5 units of KOD polymerase. The PCR was performed by a 94°C hold for 120 s followed by 6 cycles of 94°C for 15 s, 62°C for 30 s and 68°C for 30 s. The residual activity of DNA polymerase was quenched by adding 2.5 μL of 10× stop solution (50 mM EDTA, 1% SDS). The add-index-adaptor PCR products for multiple samples were mixed, primers were removed by Sephacryl S-200 spin column (GE Healthcare Life Sciences; Piscataway, NJ, USA) and the final DNA concentration was adjusted to 4 ng/μL. A 10-μL sample of the mixture and a 10-μL sodium hydrate solution (0.2 M) were mixed and kept at room temperature for 5 min. The mixture was neutralized by adding 980 μL of hybridization buffer (HT1: Illumina), 600 μL of which was used to determine the pair-end nucleotide sequences using the MiSeq Reagent kit V3 (Illumina).

Selection of the informative reads

Because the size of each amplicon is small, pair-end reads would tend to overlap at the center of the amplicon where the mutation hot spot is located. Therefore, the hot spot was sequenced in both strands. When the sequencing results of both strands matched at the hot spot and its flanking 10-bp regions and the Phred score [15] was high (≥ 30, i.e., error rate ≤ 0.001 per nucleotide), the read was considered an informative read (**Fig. 1**).

Statistical analysis

The numbers of informative reads that matched the normal sequence or

mutant sequence were counted. A diagnosis of either "negative for mutation" or "positive for mutation" was made when the statistical power was sufficient to determine the mutation in a sample with a cancer cell content of 1%, within the allowable false-positive and -negative rates. Otherwise, the result was undetermined. When the result was "positive for mutation", the cancer cell content was calculated by $\frac{Number\ of\ mutant\ reads}{Number\ of\ total\ reads} \times 2 \times 100$ (percent), assuming that cancer cells are diploid.

A detailed description of the algorithm for the statistical analysis is described in the **Doc. S3**, which describes statistical analysis.

Ethical considerations

The research plan for the current study was approved by the institutional review board of the Saitama Medical University and the participating institutions, and it conforms to the provisions of the Declaration of Helsinki. Patients' samples were collected after written informed consent was obtained.

Results

Strategy overview

Errors caused by the DNA polymerase during PCR and by the sequencer are the main obstacles for achieving a highly specific and sensitive detection of the mutations. The former type of error was reduced by employing a high-fidelity DNA polymerase, KOD. The latter type of error was reduced by sequencing both strands and by selecting the reads with a high Phred score. The five-step strategy is illustrated in **Fig. 1**.

Determination of false-mutant reads frequency

Even when sequencing normal DNA, mutant reads nonetheless appear (i.e., false-mutant reads), which cause false-positive results. To determine the frequency of the false-mutant reads for each mutation hot spot, we performed Steps 1 to 3 (**Fig. 1**) on 34 normal DNA samples (**Table 2**). The maximum frequency of the false-mutant reads in these 34 samples was defined as the reference frequency of the false-mutant reads (**Table 2**), which was used in the statistical analysis.

In a sample with mutation, reads that originated from the mutant genome and have a sequence identical to the normal sequence may occur due to the same mechanism as the false-mutant reads. However, these types of false-normal reads were considered to be very rare and to have little impact on the test results; these reads were therefore ignored.

Allowable false-positive rate and false-negative rates for each mutation hot spot

In a multiple-gene mutation test, the sum of the false positive rates in each

mutation hot spot forms the false-positive rate of the whole test (see **Doc. S3**, which describes statistical analysis). We thus assigned a fraction of the false-positive rate to each mutation spot (the allowable false-positive rate) based on the reference frequency of the false-mutant reads (**Table 2**).

The false-negative rate requires a different consideration. Most of the mutations listed in **Table 1** have been shown to occur in a mutually exclusive manner, and 2 mutations at most (*EGFR* exon 20 mutation and one of the others) may occur simultaneously. Accordingly, the allowable false-negative rate of 0.005 for each mutation hot spot was determined to be sufficient to maintain the false-negative rate of the whole test at the desired level of ≤ 0.01.

At this point, all values required for the statistical analysis was ready. Using these values, the power of statistical analysis was calculated: If a mutation hot spot has > 3500 informative reads, the mutation status is determined with specificity and sensitivity $\geq$ allowable false positive and negative rates. With $500 - 3500$ informative reads, the mutation status may sometimes be undetermined. With < 500 informative reads, the DNA quality is considered poor and the analysis is terminated and the result was not determined (see **Doc. S3**, which describes statistical analysis).

Mutation testing using admixture samples

Admixture samples simulating samples with a cancer cell content is of 1% were prepared for 8 mutations: *EGFR* G719S, *EGFR* E746-A750 del type 1, *EGFR* T790M, *EGFR* L858R, *KRAS* G12S, *KRAS* Q61K, *BRAF* G469A, and *BRAF* V600E. Five samples for each admixture (a total of 40) and 45 normal DNA samples were investigated in a single sequencing run. All mutations in the admixture were correctly

detected, while all 45 normal DNA samples were negative for mutation (**Table 3**). The average number of informative reads for one mutation hot spot was 9,970.

Mutation testing using 96 clinical samples sample

We next investigated 96 clinical samples randomly selected from our archives. All samples were isolated in the daily clinical practice and had been tested for the *EGFR* mutation using the PNA-LNA PCR clamp method [16] [17] [4]. All samples were mixed and investigated in a single sequencing run. Ninety-five samples gave definite results, while 1 sample failed due to bad quality of DNA (**Table 4**). The average number of informative reads for one mutation hot spot was 13,700. The results for the *EGFR* mutation matched with those had been determined by the PNA-LNA PCR clamp method, except the one with the lowest calculated cancer cell content of 0.8%. Using the number of the total number of informative reads and the total number of mutant reads, the false-positive rate for this result was calculated to be less than $< 10^{-10}$. We considered that the discrepancy is due to a false-negative result in the PNA-LNA PCR clamp method.

Discussion

In the current study, we constructed a highly sensitive and specific mutation test that can simultaneously investigate multiple somatic mutations using a massive parallel sequencer, MiSeq.

A major challenge in the construction of a mutation test for multiple somatic mutations is the control of the false-positive rate. In a mutation test that investigates cancer-driver genes, false-positive results have a larger clinical impact than false-negative results. When the former occur, the patients will be given a molecular targeting drug for the wrong target, which will provide no effect. When the latter occur, the patients will be given conventional chemotherapy, which may be effective to some extent. Therefore, false-positive results should be more carefully avoided than false-negative results. When simultaneously assessing multiple mutations, the false positive-rate for each mutation hot spot accumulates and increases the false-positive rate of the whole test. Therefore, the number of the genes tested should be kept to a requisite minimum, and the false-positive rate for each mutation hot spot should be maintained as small as possible.

To reduce the false-positive rate, we used a high-fidelity DNA polymerase for PCR. Furthermore, errors originating from the sequencer were reduced by sequencing DNA in both strands. We first selected only reads with a Phred score ≥ 30, so that the probability of error was in the order of $\leq 10^{-3}$, which is generally having good quality for massively parallel sequencers. However, this order of error rate is not feasible for attaining high specificity: any of the allowable false-positive rates listed in **Table 2** is smaller than 10^{-3}, indicating that the allowable false-positive rate will never be attained. Therefore, we sequenced both strands, so that the probability of error was in

the order of $\leq (10^{-3})\times(10^{-3}) = 10^{-6}$. Theoretically, this almost completely removes any errors originating from the sequencer, and enables to attain the allowable false-positive rate for all hot spots.

Current MINtS system has room to investigate additional mutation hot spots. In Step 1 (**Fig. 1**), 8 amplicons were amplified by the multiplex PCR. Our preliminary data suggested that 15 amplicons could safely be amplified in a single tube (data not shown). If more amplicons are tested, Steps 1 and 2 may be performed in separate tubes, and then the products of each reaction can be combined in a single tube, and Steps 3 to 5 are performed as described in the text. A single sequencing run by the MiSeq has a capacity for dozens of amplicons, and thus may not be a limiting factor for the number of amplicons.

Sensitivity and specificity of the MINtS can be increased by adjusting parameters. Even though the parameters are set for samples with a cancer cell content of 1%, they can be set for samples with smaller cancer cell content. However, mutation tests for samples with smaller cancer cell content inevitably requires a larger amount of DNA to suppress sampling errors [18 3]. We think that the current specification of MINtS will be appropriate for the use in the clinical setting.

The cost and possibility of reimbursement from an insurance provider are additional important factors to consider when designing a mutation test for use in clinical medicine. We run several dozens of samples, each of which were tagged by different indexes in Step 2 (**Fig. 1**) for discrimination, in a single run on the MiSeq. Therefore, the consumables required for a single sample would cost approximately 30–50 dollars, which we consider to be acceptable for most clinics. An insurance provider may reimburse the cost of a mutation test for genes that are relevant to the treatment of

cancer, but may not reimburse the costs associated with a mutation test for genes that have no direct relevance to the treatment. Therefore, the list of the mutation hot spots currently investigated by the MINtS (**Table 1**) is considered feasible.

Fusion genes including anaplastic lymphoma receptor tyrosine kinase (*ALK*) fusion genes, ROS proto-oncogene 1, receptor tyrosine kinase (*ROS1*) fusion genes [19 20], and ret proto-oncogene (*RET*) fusion genes [21 22 19] are additionally important cancer-driver genes in NSCLC. They are suitable for examining mRNA isolated from a specimen. Therefore, a mutation test for these genes requires a completely different design, and thus these genes were not included in the current study.

In the current study, we established a mutation test MINtS for investigating multiple somatic mutations with a specificity and sensitivity ≥ 0.99. To achieve this, we controlled the false-positive rate and the false-negative rate for each mutation hot spot. Our approach will be applicable to establish mutation tests for investigating a higher number of mutation hot spots, and thus the current MINtS protocol should serve as a prototype for more comprehensive mutation tests.

Acknowledgments

We sincerely thank Ms. Chieko Ono and Akemi Yokote for technical assistance. This work was supported in part by the Grants-in-Aid for Health and Labor Science [H25-Sohyaku-Ippan-011 to K.H.] from the Ministry of Health, Labor and Welfare, Japan.

Disclosure statement

All authors do not have any conflict of interest to disclose.

References

[1] Kobayashi K, Hagiwara K. Epidermal growth factor receptor (EGFR) mutation and personalized therapy in advanced nonsmall cell lung cancer (NSCLC). Target Oncol. 2013; 8: 27-33.

[2] Pao W, Girard N. New driver mutations in non-small-cell lung cancer. The Lancet Oncol. 2011; 12: 175-80.

[3] Hagiwara K, Kobayashi K. Importance of the cytological samples for the epidermal growth factor receptor gene mutation test for non-small cell lung cancer. Cancer Sci. 2013; 104: 291-7.

[4] Tanaka T, Nagai Y, Miyazawa H, et al. Reliability of the peptide nucleic acid-locked nucleic acid polymerase chain reaction clamp-based test for epidermal growth factor receptor mutations integrated into the clinical practice for non-small cell lung cancers. Cancer Sci. 2007; 98: 246-52.

[5] Goto K, Satouchi M, Ishii G, et al. An evaluation study of EGFR mutation tests utilized for non-small-cell lung cancer in the diagnostic setting. Ann Oncol. 2012; 23: 2914-9.

[6] Greer CE, Wheeler CM, Manos MM. Sample preparation and PCR amplification from paraffin-embedded tissues. PCR Methods Appl. 1994; 3: S113-22.

[7] Liu D, Nakano J, Ueno M, et al. A useful protocol for analyses of mutations of the epidermal growth factor receptor gene. Oncol Rep. 2006; 15: 1503-5.

[8] Marchetti A, Felicioni L, Buttitta F. Assessing EGFR mutations. N Engl J Med. 2006; 354: 526-8; author reply -8.

[9] Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004; 350: 2129-39.

[10] Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304: 1497-500.

[11] Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med. 2010; 362: 2380-8.

[12] Der CJ, Krontiris TG, Cooper GM. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. Proc Natl Acad Sci U S A. 1982; 79: 3637-40.

[13] Peters S, Michielin O, Zimmermann S. Dramatic response induced by vemurafenib in a BRAF V600E-mutated lung adenocarcinoma. J Clin Oncol. 2013; 31: e341-4.

[14] Marchetti A, Felicioni L, Malatesta S, et al. Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. J Clin Oncol. 2011; 29: 3574-9.

[15] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998; 8: 186-94.

[16] Miyazawa H, Tanaka T, Nagai Y, et al. Peptide nucleic acid-locked nucleic acid polymerase chain reaction clamp-based detection test for gefitinib-refractory T790M epidermal growth factor receptor mutation. Cancer Sci. 2008; 99: 595-600.

[17] Nagai Y, Miyazawa H, Huqun, et al. Genetic heterogeneity of the epidermal growth factor receptor in non-small cell lung cancer cell lines revealed by a rapid and

sensitive detection system, the peptide nucleic acid-locked nucleic acid PCR clamp. Cancer Res. 2005; 65: 7276-82.

[18] Pirker R, Herth FJ, Kerr KM, et al. Consensus for EGFR mutation testing in non-small cell lung cancer: results from a European workshop. J Thorac Oncol. 2010; 5: 1706-13.

[19] Takeuchi K, Soda M, Togashi Y, et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med. 2012; 18: 378-81.

[20] Shaw AT, Ou SH, Bang YJ, et al. Crizotinib in ROS1-Rearranged Non-Small-Cell Lung Cancer. N Engl J Med. 2014.

[21] Kohno T, Ichikawa H, Totoki Y, et al. KIF5B-RET fusions in lung adenocarcinoma. Nat Med. 2012; 18: 375-7.

[22] Lipson D, Capelletti M, Yelensky R, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012; 18: 382-4.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Doc. S1. Sequence information of the amplicons.

Doc. S2. PCR primers.

Doc. S3. Statistical analysis.

Table S1. Minimal N_i for obtaining unambiguous results for each mutation hot spot.

Fig. S1. Distribution of error reads.

Fig. S2. Determination of the result

Figure legend

Fig. 1 Overview of the reactions and analysis procedures of MINtS

Step 1: DNA fragments containing each mutation hot spot are amplified by multiplex PCR with 8 amplicons using KOD Plus DNA polymerase. Step 2: Adaptors and indexes are added to both sides of the amplified fragment for loading on the MiSeq sequencer. Adaptors are required for sequencing the fragments on the MiSeq. Indexes are required for discriminating patients whose samples are mixed and investigated together in a single sequencing run. Step 3: Nucleotide sequences of both strands are determined as pair-end reads on the MiSeq. Step 4: Only the reads with a matched sequence and with a high Phred score (≥ 30) (i.e., informative reads) are selected for the statistical analysis. Step 5: The numbers of informative reads that matched with normal or mutant sequences were counted and statistically analyzed. The result is reported as either positive or negative for mutation when the specificity and sensitivity of the whole test of ≥ 0.99 is attained. Otherwise, the result is undetermined. For a detailed description of the statistical analysis, see **Doc. S3**, which describes statistical analysis.