

## **Supplemental Appendix 1**

### **MATERIALS AND METHODS**

#### **Sample collection and handling**

All tumour tissue samples were formalin-fixed, paraffin-embedded (FFPE) diagnostic samples, archival or recently collected for recruitment to this study. All tissue samples went through pathological evaluation to confirm the diagnosis of non-small cell lung cancer (NSCLC) and to estimate the percentage of tumour cells. Owing to the time-consuming nature of macro-dissection, the limited availability of micro-dissections of tumours, and a readily available, sensitive method for epidermal growth factor receptor (EGFR) mutation testing, we did not include macro- or micro-dissection in tumour sample processing. Instead, we defined a cut-off of 2% tumour cell content as a sample quality check according to the minimum requirement of amplification refractory mutation system (ARMS) technology (about 1% analytical sensitivity). Samples below this threshold were rejected. Four to eight sections (5 µm thickness) of each qualifying tumour tissue sample were mounted on slides and used for DNA extraction.

For malignant pleural effusion (MPE) samples, 250–500 mL was collected from each patient, centrifuged at 1000 *g* for 10 min at room temperature within 1 h of sample collection.

Ten mL (10 mL) of each supernatant was transferred into a new collection tube and stored at -80°C until ready to proceed to DNA extraction. The cell pellets were fixed using standard 10% formalin and embedded in paraffin to form the MPE cell block. The MPE cell blocks were pathologically evaluated in the same way as the tumour tissue samples to confirm the diagnosis and the percentage of tumour cells. The tumour cell content had to be  $\geq 2\%$  for the sample to be included in the mutation analysis step. Ten to 15 sections (4 µm thickness) from each qualifying MPE cell block sample were used for DNA extraction.

Plasma was prepared from 10 mL of ethylene diamine tetraacetic acid-anticoagulated peripheral whole blood by centrifugation for 10 min at 2500 g at 4°C within 1 h after collection, and then stored at -80°C until DNA extraction.

For tumour tissue and MPE cell block samples, an additional three sections (4 µm thickness) were prepared on slides for EGFR mutation detection using the mutant-specific immunohistochemistry (IHC) assay.

### **DNA extraction and quality check**

The QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) was used for DNA extraction from tumour tissue samples and MPE cell block samples, as advised in the user manual. The QIAamp Circulating Nucleic Acid Kit (Qiagen) was used for DNA extraction from MPE supernatant and plasma, as advised in the user manual. Extracted DNA samples were quantified by real-time quantitative polymerase chain reaction (PCR) method using a commercial Taqman assay for the RNase P gene (Life Technologies, New York, USA). The concentration of each DNA sample was normalised to 0.4 ng/µL if the original concentration was >0.4 ng/µL.

EGFR mutation detection by ARMS method ADx-ARMS kit (Amoy Diagnostics, Xiamen, China), which has been approved by the Chinese State Food and Drug Administration for in vitro diagnostics use and has been used for scientific research,[1] was used for the EGFR mutation detection in this study. This kit can be used with multiple real-time PCR platforms, including Stratagene 3005P, with which our laboratory is equipped. Conversely, the Qiagen Scorpion ARMS TheraScreen kit can only be used with the RGQ real-time PCR platform, with which few laboratories in China are equipped. Before the application of this kit for the study, we carried out an evaluation study, using 115 cases of lung adenocarcinoma tumour FFPE samples (tumour content: 2–80%), to compare the ADx-ARMS kit with the Qiagen

Scorpion ARMS kit (a previous version that could be used with Stratagene 3005P and was used for EGFR mutation detection in IPASS and other studies), achieving a concordance rate of 96.5% (data not shown). This kit detects the 29 most common EGFR mutations so far described in lung cancer. All experiments were performed following the manufacturer's instructions. Briefly, 5  $\mu$ L DNA was added to 45  $\mu$ L of the PCR master mix for each assay, which contained PCR primers, fluorescent probes, PCR buffer and DNA polymerase. The PCR cycling conditions were: 5-min incubation at 95°C, followed by 15 cycles of 95°C for 25 s, 64°C for 20 s, 72°C for 20 s and then 31 cycles of 93°C for 25 s, 60°C for 35 s, 72°C for 20 s. Fluorescent signal was collected from FAM and HEX channels. Genotypes were determined according to threshold count (Ct) and/or change in Ct value, as indicated in the manufacturer's instructions.

### **EGFR mutation detection by Sanger sequencing**

PCR was performed in a 25  $\mu$ L reaction mix containing 1 $\times$  AmpliTaq Gold<sup>®</sup> 360 Master Mix (Life Technologies), 200  $\mu$ M of each primer and 5  $\mu$ L of genomic DNA. Exons 18–21 of the EGFR gene were amplified using the following primers: EGFR exon 18 forward 5'GCTGAGGTGACCCTTGTCTCTGTGT3', EGFR exon 18 reverse 5'ATACAGCTTGCAAGGACTCTGGGCT3'; EGFR exon 19 forward 5'CAGCATGTGGCACCATCTCACAAT3', EGFR exon 19 reverse 5'AGACATGAGAAAAGGTGGGCCTGAG3'; EGFR exon 20 forward 5'GAAGCCACACTGACGTGCCTCTC3', EGFR exon 20 reverse 5'GCTCCTTATCTCCCCTCCCCGTAT3'; EGFR exon 21 forward 5'ATCTGTCCCTCACAGCAGGGTCTTC3', EGFR exon 21 reverse 5'GCAGCCTGGTCCCTGGTGTC3'. The PCR cycling conditions were: 10-min incubation at 95°C, followed by 40 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 60 s and then a final

incubation at 72°C for 10 min. The resulting PCR products were digested with ExoSAP-IT reagent (Affymetrix, Cleveland, Ohio, USA), and then sequenced in forward and reverse directions with BigDye Terminator Kit (Life Technologies) and an ABI 3730XL DNA analyser (Life Technologies) following the manufacturer's instructions. The sequencing data were analysed for mutations after assembly and quality calling with SeqScape sequence analysis software (version 2.5; Life Technologies).

### **EGFR mutation detection by IHC**

Specificity of EGFR L858R mutant-specific (clone 43B2, CST3197; Cell Signaling Technology, Inc., Beverly, MA, USA) and EGFR E746-A750del specific (clone 6B6, CST2085; Cell Signaling Technology, Inc.) antibodies was confirmed by cross IHC staining on NSCLC cell lines with known EGFR mutations (L858R or E746-A750del). FFPE tumour tissue or MPE cell block samples were sectioned (4 µm thickness) and mounted onto coated slides. The slides were dewaxed and rehydrated before antigen retrieval was performed in PT Link (Dako, Carpinteria, California, USA) at 95°C for 15 min in high pH retrieval buffer (K8004; Dako). Endogenous peroxidase activity was blocked with hydrogen peroxide (S2023; Dako) for 5 min at room temperature. Sections were then incubated separately with total EGFR (CST4267; Cell Signaling Technology, Inc.), EGFR (L858R specific, clone 43B2, CST3197; Cell Signaling Technology, Inc.) and EGFR (E746-A750del specific, clone 6B6, CST2085; Cell Signaling Technology, Inc.) antibodies for 1 h at room temperature. Immunocomplexes were detected by incubation with anti-rabbit (K4003; Dako) horseradish peroxidase-labelled polymer and were detected by treatment with diaminobenzidine (K3468; Dako) for 10 min followed by rinsing in tap water. The sections were then counter-stained, dehydrated and cleared in Leica XL autostainer (Leica Biosystems, Illinois, USA), and finally sealed in the ClearVue automated coverslipper (Thermo Scientific, Boston, USA).

Normal immunoglobulin G (X9003; Dako) from the same species of primary antibody was diluted to match the concentration of the primary antibody and used as the negative control.

The IHC intensity was scored following the criteria of Tsai et al:[2] 0 if tumour cells had a complete absence of staining or faint staining intensity of <10%; 1+ if >10% of the tumour cells had faint staining; 2+ if the tumour cells had moderate staining; and 3+ if the tumour cells had strong staining. An expression score of 1+, 2+ and 3+ was considered positive for immunoreactivity. Cases with no EGFR expression (total EGFR negative) were excluded from data analysis in the study.

### **Quality assurance**

Two well-trained senior biologists with molecular genetics expertise executed ARMS and Sanger sequencing experiments together and analysed the data independently. A third senior molecular genetics expert combined the data interpretation together and addressed any discrepancy between the two analysers through three persons' meeting and discussion. A well-trained technician executed the mutant-specific IHC experiments. Two senior pathologists performed staining interpretation independently and any discrepancy was addressed through discussion between the two pathologists. Different samples and methods were executed blindly before final combination by a senior researcher of the study.

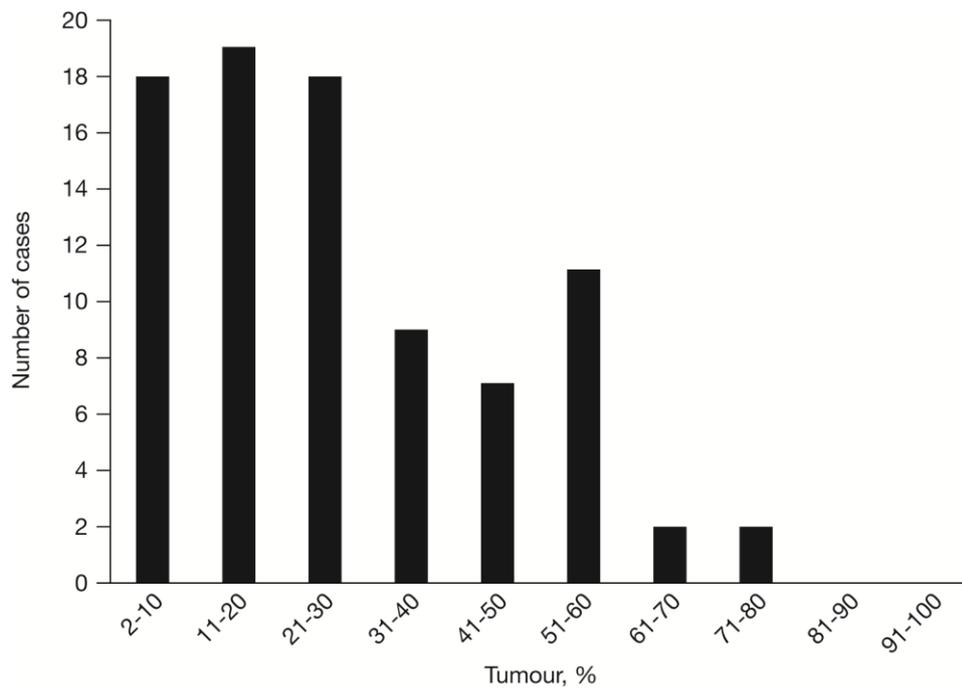
## **RESULTS**

### **Sample characteristics**

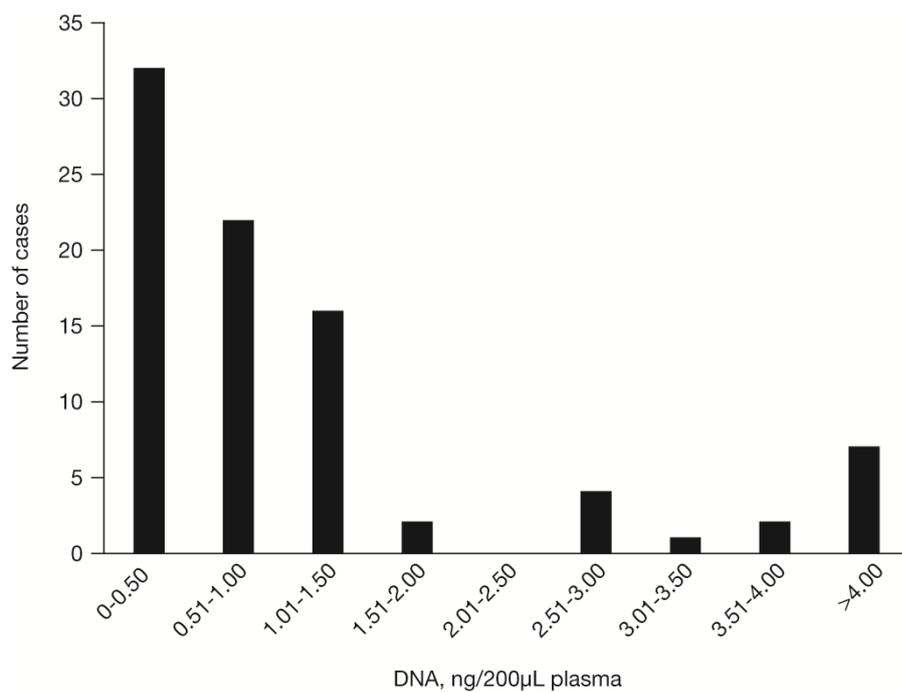
The tumour tissue samples were obtained from surgery of the primary lung tumour (n=4), small biopsy of the primary lung tumour (n=25), biopsy of a metastatic lymph node (n=48) and biopsy of metastatic loci other than a lymph node (n=9). Of the 86 patients who provided adequate tumour tissue samples (tumour content distribution is illustrated in Figure 1), 31

also provided MPE samples. Twenty-one of these MPE cell block samples passed pathology review and were included for EGFR mutation analyses, together with their supernatant counterparts. A further six patients whose MPE cell block samples passed the quality check but who had inadequate tumour tissue samples were included for EGFR mutation analyses for MPE samples only. The concentration of five tumour DNA samples, four MPE supernatant DNA samples and 21 plasma DNA samples was  $<0.4$  ng/ $\mu$ L but  $>0.2$  ng/ $\mu$ L; the concentration of seven tumour DNA samples, two MPE block DNA samples, two MPE supernatant DNA samples and 28 plasma DNA samples was  $<0.2$  ng/ $\mu$ L. The overall yield of amplifiable cfDNA from 86 plasma samples is illustrated in Figure 2.

**Supplemental Appendix Figure 1** Tumour content in the 86 tumour tissue samples.



**Supplemental Appendix Figure 2** Yield of amplifiable cell-free DNA from 86 plasma samples.



## References

1. Liu Y, Liu B, Li XY, *et al.* A comparison of ARMS and direct sequencing for EGFR mutation analysis and tyrosine kinase inhibitors treatment prediction in body fluid samples of non-small-cell lung cancer patients. *J Exp Clin Cancer Res* 2011;**30**:111.
2. Tsai T-H, Wu S-G, Chang Y-L, *et al.* Effusion immunocytochemistry as an alternative approach for the selection of first-line targeted therapy in advanced lung adenocarcinoma. *J Thorac Oncol* 2012;**7**:993-1000.