

from the tumors immediately after the PDT using SAFE-3000. We confirmed that all the NPe6 in the tumor had been excited and photobleached by the laser irradiation (664 nm) and that no additional laser irradiation was needed for curative treatment.

**Conclusions:** This novel PDD system using SAFE-3000 and NPe6 improved the quality and efficacy of PDT and avoided misjudgement of the dose of the photosensitizer or laser irradiation in PDT. PDT using NPe6 will become a standard option of treatments for centrally located early lung cancer.

© 2007 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Photodynamic therapy (PDT), used as a treatment modality for many cancers, uses a tumor-specific photosensitizer and laser irradiation to induce the production of reactive oxygen species in the cancer cells [1,2]. Since the report of the first modern clinical trial of PDT by Dougherty et al. [3], PDT using the photosensitizer, photofrin, has been applied for the treatment of many cancers, and is approved by the United States Food and Drug Administration (FDA) for the treatment of early stage lung cancer as well as advanced esophageal and lung cancers [1,4–6]. In Japan, PDT is recommended as a treatment option for centrally located early-stage lung cancers in the therapeutic guidelines for lung cancer established by the Japanese Ministry of Health, Labor and Welfare using the principles of evidence-based medicine [7–9]. Recently, the second-generation photosensitizer, mono-L-aspartyl chlorine e6 (taraporfin sodium, laserphyrin, NPe6), which has a major absorption band at 664 nm, was approved for use in the diagnosis/treatment of centrally located early lung cancer by the Japanese government [7–11]. A phase II clinical study using NPe6 and a diode laser for early-stage lung cancer demonstrated excellent antitumor effects and safety, including a significantly lower skin photosensitivity as compared to that observed with photofrin [12]. The Japanese government approved the use of NPe6 for PDT in 2003, and the product has been available in the Japanese market from June 2004.

Furukawa et al. reported that it was important to accurately define the tumor extent and depth before PDT, based on their analysis of the histopathological features, in particular, the depth of invasion of the bronchial wall by the cancer, of the recurrent lesions after CR in 114 cases of centrally located early lung cancers who underwent PDT [13]. It has been reported that PDT is associated with a high success rate only when the tumor is confined to the mucosa or submucosa, and that an important correlation exists between the response rate and the tumor size [14–16].

Sutedja et al. reported that autofluorescence bronchoscopy improved staging of radiographically occult lung cancer and had an impact on therapeutic strategy [17]. We have reported the usefulness of a photodynamic diagnosis (PDD) system based on autofluorescence bronchoscopy and the photosensitizer photofrin for the diagnosis of centrally located early-stage lung cancer [7,8,10,11]; red fluorescence emitted from the tumor following excitation of the photosensitizer by laser irradiation, e.g., using the Lung Imaging Fluorescence Endoscopy (LIFE) system, was detected in the cases with cancer, whereas green fluorescence was detected from the normal mucosa. Recently, an autofluorescence diagnosis system integrated into a

videoendoscope (SAFE-3000, Pentax, Tokyo) was established [18]. SAFE-3000 consists of a color CCD videoendoscopy-based autofluorescence system with two light sources, namely, a xenon lamp for white light and a diode laser (408 nm) as an autofluorescence mode excitation light source. Ikeda et al. reported the feasibility of use of SAFE-3000 as a very useful laser system for the early diagnosis of centrally located early cancers [18].

In this study, we examined the usefulness of this new PDD system using SAFE-3000 and NPe6 to accurately define the margin of the tumor prior to PDT and improve the efficacy of PDT for lung cancer. In cases with large tumors or anatomical problems that make laser irradiation difficult, it is often difficult to decide whether or not additional laser irradiation might be needed. Therefore, we compared the red fluorescence emitted from the tumor excited by SAFE-3000 with a diode laser (408 nm) just before and after the PDT, and conducted a dosimetric analysis to determine the appropriate dose of the laser irradiation and of the photosensitizer.

## 2. Materials and methods

### 2.1. Patient selection

Between June 2004 and March 2006, 29 patients with 38 centrally located early lung cancer lesions who were provided written informed consent, have been treated by PDT using NPe6 and a diode laser (664 nm), and also been examined immediately before and after the PDT by PDD using SAFE-3000. This study was approved by the Ethical Research Committee of Tokyo Medical University.

### 2.2. Criteria for the diagnosis of centrally located early lung cancer

Lung cancers located no distal to the segmental bronchi and determined histologically to be carcinoma in situ or carcinoma showing only limited invasion with no evidence of invasion beyond the bronchial cartilage and classified as squamous cell carcinoma are defined as centrally located early lung cancers [7,8,19]. We always determined the tumor depth by EBUS, and tumors did not invade the bronchial wall beyond the cartilage and were confined to the basal membrane of the mucosa, submucosa or intracartilaginous layers of the bronchial wall. In 2003, the Japan Photodynamic Association and Japanese Society of Laser Surgery and Medicine established the following therapeutic criteria for PDT in cases with centrally located early lung cancer [7,8]: patients with (1) endoscopically assessable early lung cancer, (2) normal chest X-ray and CT imaging,

(3) no metastasis to lymph nodes or no distant metastasis as revealed by routine clinical diagnostic methods including fluorodeoxyglucose position emission tomography (FDG-PET) for staging.

### 2.3. Autofluorescence bronchoscopy (SAFE-3000)

The newly developed videoendoscopy-based autofluorescence bronchoscopy (AF) system is referred to as SAFE-3000 [18]. In this system, normal bronchial tissue emits intense green autofluorescence when excited by blue light from a diode laser (408 nm), whereas abnormal tissue lacks the green autofluorescence due to the differences in the tissue structure, metabolic state, and blood flow.

### 2.4. Procedures of PDD and PDT and follow-up

PDT was performed using NPe6 and an aluminum gallium indium phosphorus (AlGaInP) diode laser system (PD laser, Panalase6405, Matsushita Electric Industrial Co., Ltd., Osaka, Japan) [7,8,12]. Laser irradiation (664 nm) for the PDT was performed via a quartz fiber inserted through the biopsy channel of the endoscope, 4 h after the administration of the photosensitizer talaporfin sodium (40 mg/m<sup>2</sup>). The total energy of the laser irradiation was 100 J/cm<sup>2</sup>, 150 mW/cm<sup>2</sup>. Just before the PDT, we performed PDD using SAFE-3000 with a diode laser (408 nm) to define the tumor margin, based on the red fluorescence emitted from the tumor. Immediately after the NPe6-PDT, we again performed PDD using SAFE-3000 to determine the change in the intensity of the red fluorescence emitted from the tumor as compared with that observed just before the PDT.

Cytologic and histologic examinations via fiberoptic bronchoscopy were performed at 1, 2 and 3 months and thereafter at 3-month intervals in the first year and 6-month intervals after the second years after PDT. The antitumor effect of the initial treatment was rated based on endoscopic measurement of the tumor size using forceps and morphologic appearance, and the findings on histopathologic examination of biopsy specimens, according to the general rules of the Japan Lung Cancer Society and the Japan Society of Clinical Oncology [7,8]. The antitumor effect was again evaluated at 3 months after the PDT. The tumors were

**Table 1** Characteristics of cases

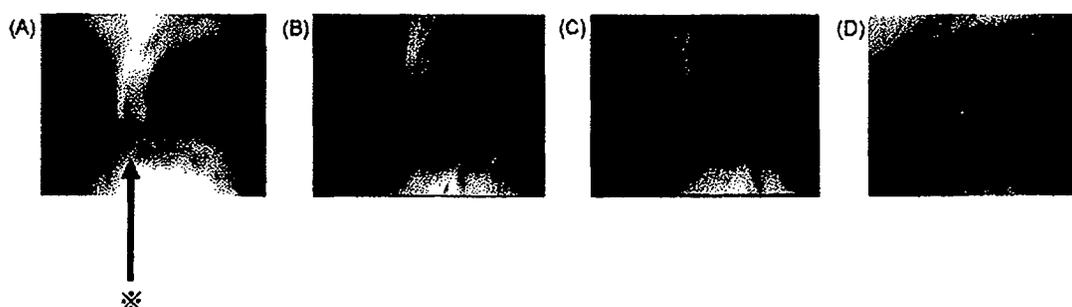
Total number of patients (lesions)	29 patients (38 lesions)
Age (years)	Mean (72.3)
Gender	Male: 29 patients
Histology	Squamous cell carcinoma: 38 lesions
Endoscopic findings	
Thickened type	35 lesions (C-stage 0: 34 lesions) (C-stage I: 1 lesion)
Polypoid type	2 lesions (C-stage 0: 2 lesions)
Nodular type	1 lesion (C-stage 0: 1 lesion)

then classified as showing complete response (CR) (no microscopically demonstrable tumor in brushing and/or biopsy specimens over period of 4 weeks)

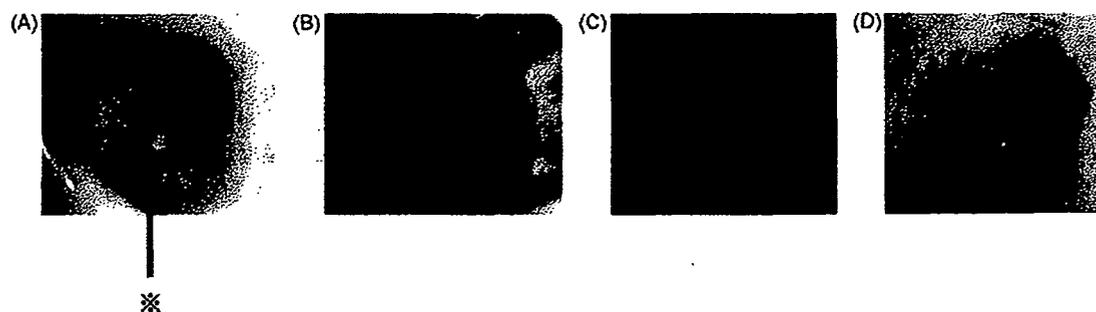
## 3. Results

### 3.1. PDD using SAFE-3000 and NPe6

Between July 2004 to March 2006, 29 patients (38 lesions) had centrally located early lung cancer and underwent NPe6-PDT and PDD based on the therapeutic criteria. All were male with a median age of 73 years. The histological type of the cancer was squamous cell carcinoma and the clinical stage 0 (including carcinoma in situ) in 37 lesions, C-stage IA (T1N0M0) in 1 lesion (Table 1). In accordance with the therapeutic criteria for PDT of lung cancer, we determined the tumor depth and tumor extent by endobronchial ultrasonography (EBUS) [16,20,21], optical coherence tomography (OCT) [22,23] and SAFE-3000 [18,24], to evaluate the suitability of the patients as candidates for PDT. The Japanese Lung Cancer Society classified centrally located early lung cancers, based on the endoscopic findings, as the thickened type (Fig. 1), polypoid type (Fig. 2) or the nodular type (Fig. 3) [18]. The thickened type, which is characterized by superficial lesions showing subtle mucosal changes of the



**Fig. 1** A 63-year-old man with centrally located early stage lung cancer. (A) Fiberoptic bronchoscopy revealed a superficial lesion. A thickened-type squamous cell carcinoma can be visualized at the carina (the bifurcation of right and left main bronchus). (B) Photodynamic diagnosis using SAFE-3000 and NPe6 was conducted before the PDT. The red fluorescence excited by the diode laser (408 nm) from the SAFE-3000 system revealed the cancerous lesion. (C) Loss of the red fluorescence from the tumor was confirmed by PDD using SAFE-3000 immediately after the NPe6-PDT. (D) CR was achieved 3 months after PDT.



**Fig. 2** A 83-year-old man with centrally located early stage lung cancer. (A) Fiberoptic bronchoscopy revealed a polypoid lesion. A polypoid-type squamous cell carcinoma can be visualized at left B<sup>3</sup>. (B) Photodynamic diagnosis using SAFE-3000 and NPe6 was conducted before the PDT. The red fluorescence excited by the diode laser (408 nm) from the SAFE-3000 system revealed the cancerous lesion. (C) Loss of the red fluorescence from the tumor was confirmed by PDD using SAFE-3000 immediately after the NPe6-PDT. (D) CR was achieved 3 months after PDT.



**Fig. 3** A 65-year-old man with centrally located early stage lung cancer. (A) Fiberoptic bronchoscopy revealed a nodular lesion. A nodular-type squamous cell carcinoma can be visualized at right upper bronchus. (B) Photodynamic diagnosis using SAFE-3000 and NPe6 was conducted before the PDT. The red fluorescence excited by the diode laser (408 nm) from the SAFE-3000 system revealed the cancerous lesion. (C) Loss of the red fluorescence from the tumor was confirmed by PDD using SAFE-3000 immediately after the NPe6-PDT. (D) CR was achieved 3 months after PDT.

bronchial surface, is the predominant type. In this study, among the 38 lesions, the thickened type accounted for 35 lesions, the polypoid type accounted for 2 lesions, and the nodular type accounted for one lesion (Table 1). In thickened type of lesions, in particular, the malignant lesion showed only slight redness and loss of clarity, which was too subtle to recognize. It was difficult to clearly confirm the tumor extent and to irradiate accurately to the tumor by diode-laser (664 nm) (Fig. 1A). However, just before the NPe6-PDT, we were able to clearly detect red fluorescence from the tumor by the PDD system using SAFE-3000, and therefore, to accurately determine the tumor extent (Figs. 1B–3B). We performed cytological examination from the boarder area, which we observed by SAFE-3000, and we confirmed tumor extent cytologically. Therefore, we were able to irradiate to the tumor thoroughly. No differences in the fluorescence pattern from the tumors were noted among three types (thickened, polypoid and nodular) of tumors mentioned above. In addition, immediately after the PDT, we were able

to confirm the loss of the red fluorescence from the tumor lesions (Figs. 1C–3C). The data obtained after the PDT confirmed that all the NPe6, in the tumor had been excited by the laser irradiation (664 nm) and that the red fluorescence of the tumor could, therefore, no longer be observed. Thus, further irradiation by PD laser (664 nm) after the PDT session was not required and the dosimetric assessment in relation of the doses of NPe6 and laser irradiation required for PDT was considered to be appropriate for the cancer treatment.

### 3.2. The efficacy of NPe6-PDT

The evaluation of the efficacy of PDT is shown in Table 2. The complete response rate of the centrally located early lung cancer lesions to NPe6-PDT was 92.1% (35/38 lesions). Of these, 33 lesions were  $\leq 10$  mm and 5 were  $>10$  mm in diameter prior to the PDT, with CR rates of 93.9% and 80.0%, respectively. Two lesions  $\leq 10$  mm in diameter in which the

**Table 2** Results of PDT for early stage lung cancer

Tumor size (cm)	No. of lesions	CR (rate, %)	PR	Recurrence after CR
$\leq 1.0$	33	31 (93.9)	2	3 (9.7%)
$1.0 <$	5	4 (80.0)	1	4 (100%)
Total	38	35 (92.1)	3	7

**Table 3** Recurrent cases after CR

Case no.	Age	Tumor size (mm)	Interval of CR (months)	Additional treatment
1	63	8	21	PDT (CR)
2	72	8	10	PDT, radiation
3	63	8	10	PDT, chemotherapy
4	63	15	9	PDT, radiation
5	80	15	15	PDT (CR)
6	78	15	9	PDT, chemotherapy
7	59	25	8	Operation

peripheral margin could not be visualized showed PR. There was no difference in response rates based on depth of cancer nor type of lesions (thickened, polypoid, nodular). We performed chemotherapy and then achieved CR for three cases, which had not obtained CR by PDT. We performed cytological and/or pathological examinations at PDT sites and we confirmed recurrences at PDT sites. Recurrent tumors were local sites, neither lymph nodes metastasis nor distant metastasis.

However, in seven patients (seven lesions) recurrences were encountered after CR was achieved by NPe6-PDT (Table 3). In all the seven cases, the recurrent lesions were thickened types. One patient with a lesion >20 mm in diameter underwent surgery and six patients (six lesions) underwent NPe6-PDT again after reassessing their suitability as candidates for PDT by EBUS, OCT and SAFE-3000. Of these six lesions, CR after the repeat PDT was achieved in only two cases, not achieved in four cases. We performed radiotherapy in two cases, and chemotherapy in two cases. All patients in this study were disease-free and still alive (Table 3).

### 3.3. Complications of NPe6-PDT

In this study, patients undergoing NPe6-PDT were required to avoid sunlight for 2 weeks, even though the skin photosensitivity was less strong as compared to that with photofrin-PDT. NPe6-PDT did not cause extensive necrotic changes in the tumors and clean-up bronchoscopy was rarely needed. The period of hospitalization after the laser-irradiation for NPe6-PDT was 7–10 days. These results suggest that NPe6-PDT may be cost-effective and improve patient's quality of life. In two lesions, we observed bronchial stenosis at the site of the tumor, which had spread around the bronchus, necessitating cylindrical laser irradiation.

## 4. Discussion

In this study, we were able to observe the red fluorescence emitted by the tumor cells because of laser excitation (408 nm) of the NPe6 contained in them, using SAFE-3000 just prior to PDT; this enabled us to clearly determine the tumor margin of the centrally located early lung cancer lesions, independent of the endoscopic findings (Figs. 1B–3B). Thus, this novel PDD technique using the new autofluorescence bronchoscopy system, SAFE-3000, and NPe6, allows more accurate assessment of the tumor

margin, and therefore, of the quality and efficacy of PDT. Since tumor lesions are characterized by hypervascularity and increased redness, emission of red fluorescence can be observed from the tumor cells by excitation of a tumor-selective photosensitizer administered prior to the irradiation by blue light (408 nm) from diode laser using SAFE-3000.

We have been examining the feasibility of using photodynamic diagnostic (PDD) systems since 1980 [10,11,25,26]. The potential for the application of hematoporphyrin derivative (HpD) fluorescence using a krypton ion laser for localization of early lung cancer was demonstrated in an animal model by Hayata and Dougherty. The development of PDD systems are fraught with problems including skin photosensitivity and poor selectivity, e.g., of the photosensitizer photofrin, and also autofluorescence interference. In order to overcome the problems, we developed the excimer laser fluorescence image analyzer system [10,11]. The red fluorescence images of the excited photosensitizer and the autofluorescence are amplified separately and then processed by a data analyzer in this system. However, in the R/G fluorescence bronchoscopy system, the rate of false-positive results was found to be very high, i.e., 30%. We also developed a PDD system based on laser-induced fluorescence (originally for lung cancer) endoscopy after administration of a photosensitizer [10,11,27]. The sensitivity of this LIFE-PDD system was more than 90%, however, the false-positive rate remained high.

In this study, we examined a new PDD system developed by us, using SAFE-3000. The advantages of this PDD systems are as follows: increased sensitivity, ability to clearly define the tumor margin, a color CCD videoendoscopy-based AF system equipped with a diode laser and a hand switch to easily switch between the white light and AF modes. There are several kinds of AF systems available for the early detection of cancer lesions [28–32]. However, most are not equipped with a diode laser system to excite the photosensitizer in the tumor. As shown in Figs. 1B–3B, the red fluorescence from the tumor and the border between the normal mucosa and cancer lesion can be clearly visualized with SAFE-3000.

In Figs. 1C–3C, the loss of the red fluorescence from the tumor in the thickened, polypoid as well as the nodular types of cancer immediately after the PDT can be clearly visualized. Invasion of the bronchial cartilage has been reported in 18% and 27% of lesions, respectively, with the nodular and polypoid types of early lung cancer [33], therefore, additional laser-irradiation was often required for these types of early lung cancer. However, our data from the present study suggest that all of the NPe6 in the tumor had been completely excited by the laser irradiation, and

no additional laser irradiation was necessary. The results indicated that the doses of the NPe6 as well as the laser irradiation were appropriate for the treatment of early lung cancer in-dependent upon bronchoscopical findings (thickened, nodular, polypoid types). Therefore, there was no difference in response rates based on type of lesions.

As Furuse et al. reported, the longitudinal extent of the tumor was the only independent predictive factor for CR after PDT [15]; in our study, the tumor size was also an important factor for CR after NPe6-PDT. In two lesions less than 10 mm in diameter, CR was not achieved by NPe6-PDT, even though the peripheral extent of the tumors could be clearly visualized. For these two lesions, we selected cylindrical fiber irradiation, because the tumor extended around the bronchial wall. The reason for the inability to achieve CR in these cases with lesions less than 10 mm could be explained by possible inappropriate estimation of the tumor margin and insufficient laser irradiation.

In this study, the CR rate was high and all patients were alive. However, Furukawa et al. reported that the majority of patients who underwent PDT, were at an advanced age, with poor cardiopulmonary function and died from other disease except for lung cancer, and 5-year survival rate was relatively low. Therefore, we have to follow up all patients and have to analyze the relationship between CR rate and 5-year survival rate after PDT using talaporfin sodium in the future. Moreover, in order to improve the efficacy of PDD and PDT, we should develop the fibers for irradiation of the tumors that could facilitate irradiation of areas that are normally difficult to irradiate.

### Conflict of interest statement

There was no financial support for the authors nor does any author have a financial relationship with a commercial entity that has an interest in this manuscript.

### Acknowledgement

This work was supported in part by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT). KAKENHI 18791002 (to J.U.).

### References

- [1] Dougherty TJ, Gomer CJ, Henderson BW, Jori G, Kessel D, Korbek M, et al. Photodynamic therapy. *J Natl Cancer Inst* 1998;90:889–905.
- [2] Oleinick NL, Morris RL, Belichenko I. The role of apoptosis in response to photodynamic therapy: what, where, why, and how. *Photochem Photobiol Sci* 2001;1:1–21.
- [3] Dougherty TJ, Lawrence G, Kaufman JH, Boyle D, Weishaupt KR, Goldfarb A. Photoradiation in the treatment of recurrent breast carcinoma. *J Natl Cancer Inst* 1979;62:231–7.
- [4] Cheng M, Pennathur A, Luketich JD. Role of photodynamic therapy in unresectable esophageal and lung cancer. *Lasers Surg Med* 2006;38:396–402.
- [5] Cortese DA, Edell ES, Kinsey JH. Photodynamic therapy for early stage squamous cell carcinoma of the lung. *Mayo Clin Proc* 1997;72:595–602.
- [6] Boxem AJM, Westerga J, Venmans BJW, Postmus PE, Sutedja T. Photodynamic therapy, Nd-YAG laser and electrocautery for treating early-stage intraluminal cancer: which to choose? *Lung Cancer* 2001;31:31–6.
- [7] Usuda J, Kato H, Okunaka T, Furukawa K, Tsutsui H, Yamada K, et al. Photodynamic therapy (PDT) for lung cancers. *J Thorac Oncol* 2006;5:489–93.
- [8] Kato H, Usuda J, Okunaka T, Furukawa K, Honda H, Sakaniwa N, et al. Basic and clinical research on photodynamic therapy at Tokyo Medical University Hospital. *Lasers Surg Med* 2006;38:371–5.
- [9] Mathur PN, Edell E, Sutedja T, Vergnon JM. Treatment of early non-small cell lung cancer. *Chest* 2003;123:176–80.
- [10] Kato H, Usuda J, Okunaka T, Furukawa K. The history of the study of photodynamic therapy (PDT) and photodynamic diagnosis (PDD) in the department of Surgery, Tokyo Medical University. *Photodiagn Photodyn Ther* 2004;1:107–10.
- [11] Kato H, Harada M, Ichinose S, Usuda J, Tsuchida A, Okunaka T. Photodynamic therapy (PDT) of lung cancer: experience of the Tokyo Medical University. *Photodiagn Photodyn Ther* 2004;1:49–55.
- [12] Kato H, Furukawa K, Sato M, Okunaka T, Kusunoki Y, Kawahara M, et al. Phase II clinical study of photodynamic therapy using mono-L-aspartyl chlorine e6 and diode laser for early superficial squamous cell carcinoma of the lung. *Lung Cancer* 2003;42:103–11.
- [13] Furukawa K, Kato H, Konaka C, Okunaka T, Usuda J, Ebihara Y. Locally recurrent central-type early stage lung cancer <1.0 cm in diameter after complete remission by photodynamic therapy. *Chest* 2005;128:3269–75.
- [14] Furuse K, Fukuoka M, Kato H, Horai T, Kubota K, Kodama N, et al. A prospective phase II study on photodynamic therapy with photofrin II for centrally located early-stage lung cancer. *J Clin Oncol* 1993;11:1852–7.
- [15] Miyazu Y, Miyazawa T, Kurimoto N, Iwamoto Y, Kanoh K, Kohno N. Endobronchial ultrasonography in the assessment of centrally located early-stage lung cancer before photodynamic therapy. *Am J Respir Crit Care Med* 2002;165:832–7.
- [16] Usuda K, Saito Y, Nagamoto N, Sato M, Sagawa M, Kanma K, et al. Relation between bronchoscopic findings and tumor size of roentgenographically occult bronchogenic squamous cell carcinoma. *J Thoracvasc Surg* 1993;106:1098–103.
- [17] Sutedja TG, Codrington H, Rissc EK, Breuer RH, Van Mourik JC, Golding RP, et al. Autofluorescence bronchoscopy improves staging of radiographically occult lung cancer and has an impact on therapeutic strategy. *Chest* 2001;120:1327–32.
- [18] Ikeda N, Honda H, Hayashi A, Usuda J, Kato Y, Tsuboi M, et al. Early detection of bronchial lesions using newly developed videoendoscopy-based autofluorescence bronchoscopy. *Lung Cancer* 2006;52:21–7.
- [19] Ikeda S. Atlas of early cancer of major bronchi. Tokyo: Igakushoin Publisher; 1975.
- [20] Kurimoto N, Murayama M, Yoshioka S, Nishisaka T, Inai K, Dohi K. Assessment of usefulness of endobronchial ultrasonography in determination of depth of tracheobronchial tumor invasion. *Chest* 1999;115:1500–6.
- [21] Larsen SS, Vilmann P, Krasnik M, Dirksen A, Clementsen P, Mattbaek N, et al. Endoscopic ultrasound guided biopsy performed routinely in lung cancer staging spares futile thoracotomies: preliminary results from a randomized clinical trial. *Lung Cancer* 2005;49:377–85.
- [22] Tsuboi M, Hayashi A, Ikeda N, Honda H, Kato Y, Ichinose S, et al. Optical coherence tomography in the diagnosis of bronchial lesions. *Lung Cancer* 2005;49:387–94.
- [23] Whiteman SC, Yang Y, Pittius DG, Stephens M, Parmer J, Spiteri MA. Optical coherence tomography: real-time imaging bronchial airways microstructure and detection of inflam-

- matory/neoplastic morphologic changes. *Clin Cancer Res* 2006;12:813–8.
- [24] Ikeda N, Hayashi A, Iwasaki K, Honda H, Tsuboi M, Usuda J, et al. Comprehensive diagnostic bronchoscopy of central type early lung cancer. *Lung Cancer* 2007;56:295–302.
- [25] Hayata Y, Kato H, Ono J, Iimura I, Matsushima Y, Hayashi N, et al. Fluorescence fiberoptic bronchoscopy in the diagnosis of early lung cancer. *Recent Results Cancer Res* 1982;82:121–30.
- [26] Hayata Y, Kato H, Konaka C, Ono J, Matsushima Y, Yoneyama K, et al. Fiberoptic bronchoscopic laser photoradiation for tumor localization in lung cancer. *Chest* 1982;82:10–4.
- [27] Kusunoki Y, Imamura F, Uda H, Mano M, Horai T. Early detection of lung cancer with laser-induced fluorescence endoscopy and spectrofluorometry. *Chest* 2000;118:1776–82.
- [28] Lam S, MacAulay C, Hung J, et al. Detection of dysplasia and carcinoma in situ with a lung imaging fluorescence endoscope device. *J Thorac Cardiovasc Surg* 1993;105:1035–40.
- [29] Pasic A, Vonk-Noordegraaf A, Risse EKJ, Postmus PE, Sutedja TG. Multiple suspicious lesions detected by autofluorescence bronchoscopy predict malignant development in the bronchial mucosa in high risk patients. *Lung Cancer* 2003;41:295–301.
- [30] Pierard P, Faber J, Hutsebaut J, Martin B, Plat G, Sculier JP, et al. Synchronous lesions detected by autofluorescence bronchoscopy in patients with high-grade preinvasive lesions and occult invasive squamous cell carcinoma of proximal airways. *Lung Cancer* 2004;46:341–7.
- [31] Baas P, Triesscheij M, Burgers S, Pel RV, Stewart F, Aalders M. Fluorescence detection of pleural malignancies using 5-aminolaevulinic acid. *Chest* 2006;129:718–24.
- [32] Chhajed PN, Shibuya K, Hoshino H, Chiyo M, Yasufuku K, Hiroshima K, et al. A comparison of video and autofluorescence bronchoscopy in patients at high risk of lung cancer. *Eur Respir J* 2005;25:951–5.
- [33] Konaka C, Hirano T, Kato H, Furuse K, Takada M, Saito Y, et al. Comparison of endoscopic features of early-stage squamous cell lung cancer and histological findings. *Br J Cancer* 1999;80:1435–9.

## Letter to the Editor

**Pleural sarcomatoid malignant mesothelioma consisting of histiocytoid cells***To the Editor:*

Pleural sarcomatoid malignant mesothelioma (PSMM) consists of spindle cells arranged in fascicles or having a haphazard fashion. The pattern often resembles fibrosarcoma, but marked anaplasia and bizarre multinucleated tumor cells may result in a picture closely mimicking that of malignant fibrous histiocytoma.<sup>1</sup> We herein report an extremely rare case of PSMM consisting of histiocytoid cells without any inflammatory infiltrate. We discuss its differential diagnosis and the key to establishing an accurate diagnosis.

A 76-year-old Japanese woman, a non-smoker, consulted Osaka Prefectural Medical Center for Respiratory and Allergic Diseases complaining of left chest pain that had persisted for the previous 2 weeks. She had worked as a spinning-mill worker for more than 30 years and occupational exposure to asbestos was highly suspected. She had a past history of pleuritis and pericarditis of unknown origin. Chest X-ray demonstrated two fist-sized subpleural nodules in the left mediastinal and parietal costal pleura without any pulmonary lesions. There was a large amount of bloody pleural effusion on the left side. Pleural effusion cytology did not demonstrate any neoplastic findings. The course of the disease was rapidly progressive. A tentative clinical diagnosis of pleuritis carcinomatosa was made and percutaneous needle biopsies were obtained from the subpleural lesions. Histologically, most tumor cells infiltrated into a myxoid stroma and grew as a diffuse pattern (Fig. 1a). Sheets and nests of cells were absent. Inflammatory infiltrate was slight. The tumor cells were mainly composed of histiocytoid cells (Fig. 1b). These histiocytoid cells were ovoid or polyhedral in shape, with pale eosinophilic or foamy cytoplasm. Their nuclei were medium to large, ovoid or angulated, contained fine chromatin, and inconspicuous nucleoli (Fig. 1c). Multinucleated giant tumor cells were rarely recognized. Mitoses were 2/10 high-power fields (HPF). Necrosis and hemorrhagic areas were absent. Reactive histiocytes were sometimes present among these neoplastic cells. Immunohistochemical studies showed that these tumor cells were positive for AE1/AE3 (Fig. 2a), D2-40 (Fig. 2b), HBME-1, vimentin and WT-1 (Fig. 2c), but negative for BerEP4, CA125, calretinin, CAM5.2, CD3, CD15, CD30, CD34, CD45RB, CD79 $\alpha$ , CEA, cytokeratin 5/6, cytokeratin 7, desmin, epithelial membrane antigen, napsin A, PE-10, PG-M1, S-100, Smooth-muscle actin, thrombomodulin and thyroid transcription factor-1. MIB-1 was positive in 20% of tumor cells. PAS reaction was positive in some

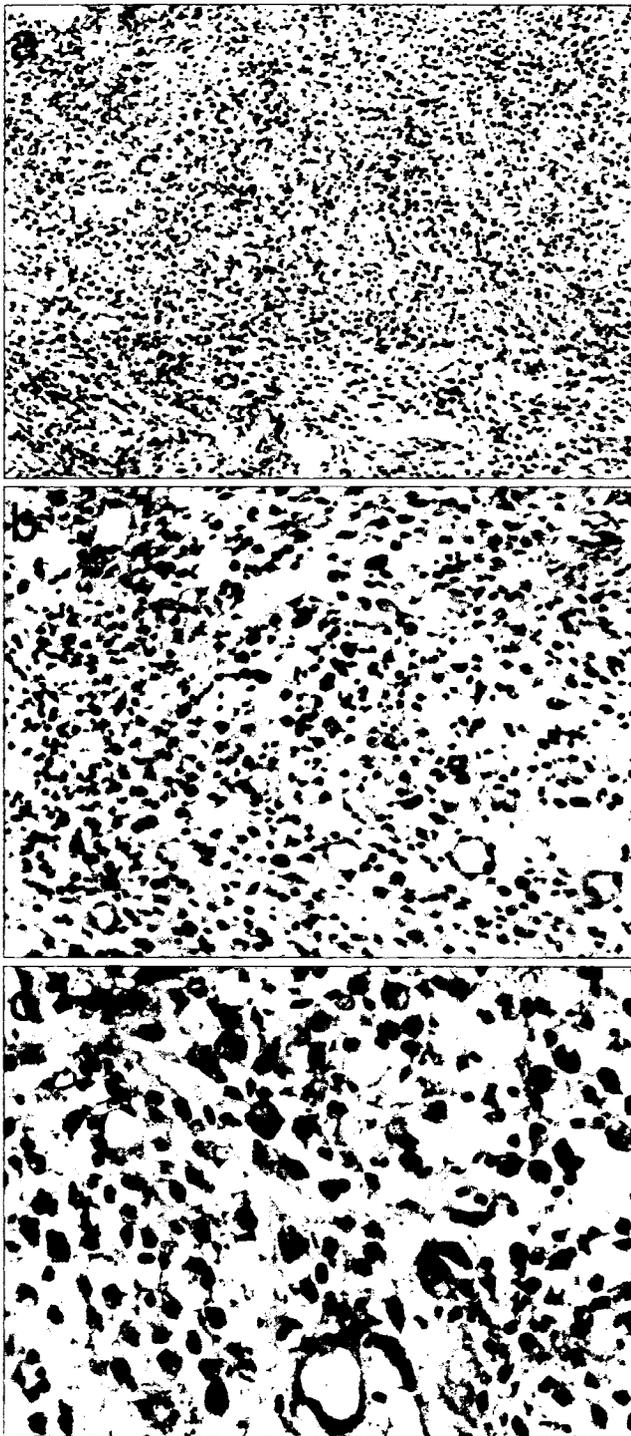
tumor cells and was partly digested with diastase. Alcian blue stain was positive in stroma digested with hyaluronidase and negative in the cytoplasm. Colloidal iron stain was positive in the stroma and digested with hyaluronidase. Taken together, we made a diagnosis of pleural sarcomatoid malignant mesothelioma consisting of histiocytoid cells.

Recognition of this rare variant of malignant mesothelioma is important because of the possibility of confusing it with lymphohistiocytic variant of anaplastic large cell lymphoma (LHALCL), sarcomatoid carcinoma (SC) of the lung or thymus, inflammatory malignant fibrous histiocytoma (IMFH), inflammatory myofibroblastic tumor (IMT) or lymphohistiocytoid mesothelioma (LHM). In LHALCL, large-sized atypical cells with immunohistochemical CD30 positivity are admixed with histiocytes and plasma cells.<sup>2</sup> However, in this tumor, immunohistochemical studies showed that the atypical cells were negative for CD30. As for SC, lesions of the lung and anterior mediastinum were absent, and spindle or giant atypical cells were not apparent histologically in the present case. IMFH is composed of xanthogranulomatous inflammation with scattered atypical large cells with prominent nucleoli. The present tumor lacked xanthogranulomatous inflammation, and histiocytoid cells were not so large and nucleoli were inconspicuous. IMT is composed of a variable mixture of collagen, inflammatory cells and cytologically bland spindle cells having myofibroblastic differentiation. However, in the present case, collagen and these bland spindle cells were absent. LHM is a variant of sarcomatoid malignant mesothelioma, characterized by diffuse proliferation of large, ovoid histiocyte-like and spindle cells, uniformly intermixed with a prominent lymphocytic or lymphoplasmacytic infiltrate.<sup>3,4</sup> But in the present case the lymphocytic or lymphoplasmacytic infiltrate was subtle.

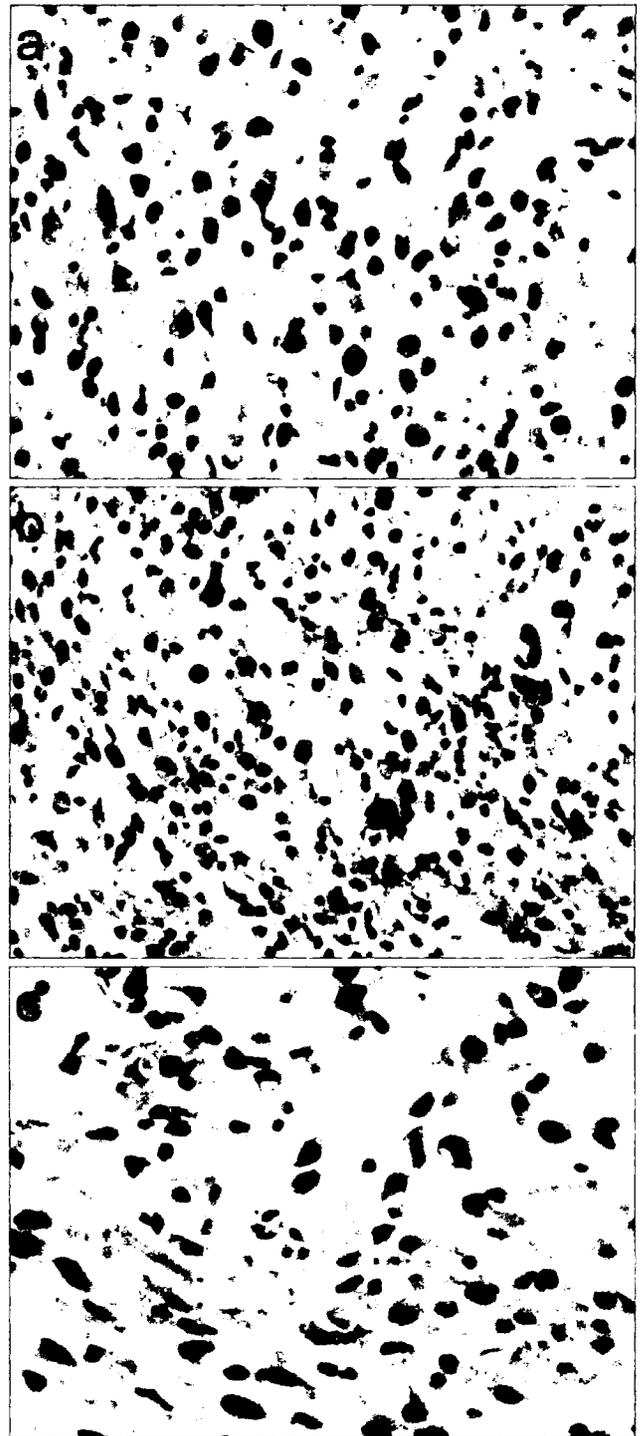
We made a diagnosis of PSMM consisting of histiocytoid cells based on percutaneous needle biopsy specimens. The keys to accurate diagnosis were as follows: (i) recognition of serosal tumor; (ii) immunohistochemical positivity for AE1/AE3, vimentin and mesothelial markers including D2-40,<sup>5</sup> WT1 and HBME-1; and (iii) being aware of the varied histopathological manifestations of PSMM.

Kunimitsu Kawahara,<sup>1</sup> Teruaki Nagano,<sup>1</sup> Kaoru Matsui,<sup>2</sup>  
Toshiaki Kawai,<sup>3</sup> and Teruo Iwasaki<sup>4</sup>

Departments of <sup>1</sup>Pathology, <sup>2</sup>Thoracic Malignancy and  
<sup>4</sup>Thoracic Surgery, Osaka Prefectural Medical Center for  
Respiratory and Allergic Diseases, Habikino and  
<sup>3</sup>Department of Pathology, National Defense Medical  
College, Tokorozawa, Japan



**Figure 1** Microscopic view of the tumor. (a) Tumor cells infiltrating diffusely in a myxoid stroma. (b) Tumor cells mainly composed of histiocytoid cells. (c) The tumor cells were ovoid or polyhedral in shape with pale eosinophilic or foamy cytoplasm. Stain: HE.



**Figure 2** Immunohistochemical staining of the tumor. The histiocytoid cells had positive staining for (a) AE1/AE3, (b) D2-40 and (c) WT1.

#### REFERENCES

1 Churg A, Roggli V, Galateau-Salle F *et al.* Mesothelioma. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, eds.

*World Health Organization Classification of Tumors. Pathology and Genetics of Tumors of the Lung, Pleura, Thymus and Heart.* Lyon: IARC Press, 2004; 128–36.  
2 Pileri S, Falini B, Delsol G *et al.* Lymphohistiocytic T-cell lymphoma (anaplastic large cell lymphoma CD30+/Ki-1+ with a high

- content of reactive histiocytes). *Histopathology* 1990; **16**: 383–91.
- 3 Henderson DW, Attwood HD, Constance TJ, Shilkin KB, Steele RH. Lymphohistiocytoid mesothelioma: A rare lymphomatoid variant of predominantly sarcomatoid mesothelioma. *Ultrastruct Pathol* 1988; **12**: 367–84.
  - 4 Khalidi HS, Medeiros LJ, Battifora H. Lymphohistiocytoid mesothelioma. An often misdiagnosed variant of sarcomatoid malignant mesothelioma. *Am J Clin Pathol* 2000; **113**: 649–54.
  - 5 Chu AY, Litzky LA, Pasha TL, Acs G, Zhang PJ. Utility of D2-40, a novel mesothelial marker, in the diagnosis of malignant mesothelioma. *Mod Pathol* 2005; **18**: 105–10.

# Polymorphisms, Mutations, and Amplification of the *EGFR* Gene in Non-Small Cell Lung Cancers

Masaharu Nomura<sup>1</sup>, Hisayuki Shigematsu<sup>1</sup>, Lin Li<sup>2</sup>, Makoto Suzuki<sup>1</sup>, Takao Takahashi<sup>1</sup>, Pila Estess<sup>3</sup>, Mark Siegelman<sup>3</sup>, Ziding Feng<sup>2</sup>, Harubumi Kato<sup>4</sup>, Antonio Marchetti<sup>5</sup>, Jerry W. Shay<sup>6</sup>, Margaret R. Spitz<sup>7</sup>, Ignacio I. Wistuba<sup>8</sup>, John D. Minna<sup>1,9,10</sup>, Adi F. Gazdar<sup>1,3\*</sup>

**1** Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **2** Cancer Prevention Research, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **3** Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **4** First Department of Surgery, Tokyo Medical University, Tokyo, Japan, **5** Pathology Unit, Clinical Research Center, Center of Excellence on Aging, University Foundation, Chieti, Italy, **6** Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **7** Department of Epidemiology, University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America, **8** Department of Pathology, University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America, **9** Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **10** Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America

**Funding:** This research was supported by grants from the Specialized Program of Research Excellence in Lung Cancer (PS0CA70907) and the Early Detection Research Network (SU01CA8497102), National Cancer Institute, Bethesda, Maryland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Academic Editor:** William Pao, Memorial Sloan-Kettering Cancer Center, United States of America

**Citation:** Nomura M, Shigematsu H, Li L, Suzuki M, Takahashi T, et al. (2007) Polymorphisms, mutations, and amplification of the *EGFR* gene in non-small cell lung cancers. *PLoS Med* 4(4): e125. doi:10.1371/journal.pmed.0040125

**Received:** March 2, 2006  
**Accepted:** February 9, 2007  
**Published:** April 24, 2007

**Copyright:** © 2007 Nomura et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** AI, allelic imbalance; CA-SSR1, CA simple sequence repeat 1; *EGFR*, epidermal growth factor receptor; FISH, fluorescence in situ hybridization; HBEC, human bronchial epithelial cell; LAD, longer allele dominant; NSCLC, non-small cell lung cancer; PBMC, peripheral blood mononuclear cell; SAD, short allele dominant; SNP, single nucleotide polymorphism; TK, tyrosine kinase; TKI, tyrosine kinase inhibitor; WT, wild-type

\* To whom correspondence should be addressed. E-mail: adi.gazdar@utsouthwestern.edu

## ABSTRACT

### Background

The *epidermal growth factor receptor (EGFR)* gene is the prototype member of the type I receptor tyrosine kinase (TK) family and plays a pivotal role in cell proliferation and differentiation. There are three well described polymorphisms that are associated with increased protein production in experimental systems: a polymorphic dinucleotide repeat (*CA simple sequence repeat 1 [CA-SSR1]*) in intron one (lower number of repeats) and two single nucleotide polymorphisms (SNPs) in the promoter region, −216 (G/T or T/T) and −191 (C/A or A/A). The objective of this study was to examine distributions of these three polymorphisms and their relationships to each other and to *EGFR* gene mutations and allelic imbalance (AI) in non-small cell lung cancers.

### Methods and Findings

We examined the frequencies of the three polymorphisms of *EGFR* in 556 resected lung cancers and corresponding non-malignant lung tissues from 336 East Asians, 213 individuals of Northern European descent, and seven of other ethnicities. We also studied the *EGFR* gene in 93 corresponding non-malignant lung tissue samples from European-descent patients from Italy and in peripheral blood mononuclear cells from 250 normal healthy US individuals enrolled in epidemiological studies including individuals of European descent, African-Americans, and Mexican-Americans. We sequenced the four exons (18–21) of the TK domain known to harbor activating mutations in tumors and examined the status of the *CA-SSR1* alleles (presence of heterozygosity, repeat number of the alleles, and relative amplification of one allele) and allele-specific amplification of mutant tumors as determined by a standardized semiautomated method of microsatellite analysis. Variant forms of SNP −216 (G/T or T/T) and SNP −191 (C/A or A/A) (associated with higher protein production in experimental systems) were less frequent in East Asians than in individuals of other ethnicities ( $p < 0.001$ ). Both alleles of *CA-SSR1* were significantly longer in East Asians than in individuals of other ethnicities ( $p < 0.001$ ). Expression studies using bronchial epithelial cultures demonstrated a trend towards increased mRNA expression in cultures having the variant SNP −216 G/T or T/T genotypes. Monoallelic amplification of the *CA-SSR1* locus was present in 30.6% of the informative cases and occurred more often in individuals of East Asian ethnicity. AI was present in 44.4% (95% confidence interval: 34.1%–54.7%) of mutant tumors compared with 25.9% (20.6%–31.2%) of wild-type tumors ( $p = 0.002$ ). The shorter allele in tumors with AI in East Asian individuals was selectively amplified (shorter allele dominant) more often in mutant tumors (75.0%, 61.6%–88.4%) than in wild-type tumors (43.5%, 31.8%–55.2%,  $p = 0.003$ ). In addition, there was a strong positive association between AI ratios of *CA-SSR1* alleles and AI of mutant alleles.

### Conclusions

The three polymorphisms associated with increased *EGFR* protein production (shorter *CA-SSR1* length and variant forms of SNPs −216 and −191) were found to be rare in East Asians as compared to other ethnicities, suggesting that the cells of East Asians may make relatively less intrinsic *EGFR* protein. Interestingly, especially in tumors from patients of East Asian ethnicity, *EGFR* mutations were found to favor the shorter allele of *CA-SSR1*, and selective amplification of the shorter allele of *CA-SSR1* occurred frequently in tumors harboring a mutation. These distinct molecular events targeting the same allele would both be predicted to result in greater *EGFR* protein production and/or activity. Our findings may help explain to some of the ethnic differences observed in mutational frequencies and responses to TK inhibitors.

The Editors' Summary of this article follows the references.

## Introduction

*Epidermal growth factor receptor* (*EGFR*, also known as *ERBB1*) belongs to the *ERBB* gene family of receptor tyrosine kinases (TKs), and is a major regulator of several distinct and diverse signaling pathways [1–3]. It is frequently overexpressed in many malignancies including non-small cell lung cancer (NSCLC), and overexpression may be associated with a negative prognosis [4,5]. A recent finding that mutations of the gene in lung cancers predict, somewhat imprecisely, response to TK inhibitors (TKIs) has generated much interest [6–10]. Mutations are limited to the first four exons of the TK domain, and occur more often in individuals with adenocarcinoma histology, East Asian origin, female gender, and never smoker status. However, exceptions exist to the correlation between mutation status and response to TKIs, suggesting that other factors may play a role. Recently, *EGFR* amplification has been identified as a further factor that may predict response to therapy [11,12]. Experimental evidence indicates that polymorphisms of the gene may also regulate protein expression.

*CA simple sequence repeat 1* (*CA-SSR1*) is a highly polymorphic locus containing 14–21 CA dinucleotide repeats and is located at the 5' end of the long intron one of the *EGFR* gene, lying upstream and in close proximity to a second enhancer [13,14]. The allele size distribution of *CA-SSR1* demonstrates ethnic differences, with East Asians having longer repeats than individuals of European descent or African-Americans [15]. By interacting with the second or downstream enhancer, a lower *CA-SSR1* repeat number was found to modulate *EGFR* transcription in vivo and in vitro, and to be correlated with increased transcription and protein expression [13,14].

The relationship between *CA-SSR1* repeat length and *EGFR* overexpression has been extensively studied in breast cancers [16,17]. Localized amplification of the *CA-SSR1* repeat, usually limited to the shorter allele, occurs frequently in breast cancers, is related to *EGFR* expression, and demonstrates a field effect, indicating that it is an early event during multistage pathogenesis [18]. In head and neck cancer, patients with a lower number of *CA-SSR1* repeats (total of both alleles < 35 repeats) had a statistically significantly increased likelihood of responding to erlotinib [19].

In addition to *CA-SSR1*, two kinds of single nucleotide polymorphisms (SNPs) in the promoter region may correlate with increased promoter activity and expression of *EGFR* mRNA. One of the SNPs is located –216 bp upstream from the initiator ATG (adenine as +1), and the change of nucleoside is guanine to thymine. This is an important binding site for the transcription factor SP1 that is necessary for activation of *EGFR* promoter activity [20]. The variant forms, –216 G/T or T/T, are more frequent in individuals of European descent and African-Americans than in Asians [21]. The other SNP, –191 C/C, is located in the *EGFR* promoter region near one of four transcription regions (–214 to –200) [22]. This SNP may also be associated with increased protein expression, and the minor forms, –191 C/A or A/A are also rare among Asians [21].

For the reasons discussed above, we investigated the distribution of these SNPs in lung cancer patients and healthy individuals of various ethnicities, the length and allelic imbalance (AI) of *CA-SSR1* in lung cancer patients, and

the relationship between AI of *CA-SSR1* and allele-specific amplification in lung cancer patients with mutations of the *EGFR* gene.

## Methods

Because of the multiple, complex studies performed in this report, we summarize the salient investigations and their results in Table 1.

### Human Bronchial Epithelial Cell and Lung Cancer Cell Lines

All cancer cell lines were cultured in RPMI 1640 (Life Technologies, Rockville, Maryland, United States) supplemented with 5% fetal bovine serum and incubated in humidified air and 5% CO<sub>2</sub> at 37 °C. Most cell lines were established by us at one of two locations. The prefix NCI indicates cell lines established at the National Cancer Institute, and the prefix HCC indicates cell lines established at the Hamon Center for Therapeutic Oncology Research of the University of Texas Southwestern Medical Center.

Human bronchial epithelial cells (HBECS) from healthy individuals or those with lung cancer were immortalized and cultured by us as previously described [23,24]. The cells were cultured in K-SFM medium (Life Technologies) and included 5 ng/ml EGF.

### Clinical Samples

A total of 556 samples of primary lung cancers including adenocarcinomas ( $n = 345$ , 62%), squamous cell carcinomas ( $n = 182$ , 33%), adenosquamous carcinomas ( $n = 16$ , 3%), and large cell carcinomas ( $n = 10$ , 2%) were obtained from four countries, the US, Australia, Japan, and Taiwan, and included 336 (60%) tumors from East Asians and 220 (40%) from other ethnicities (97% of whom were of European descent). None of the cases had prior treatment with TKIs. Samples of tumor containing relatively high percentages of tumor (>70%) were selected and analyzed without microdissection.

Corresponding non-malignant lung tissues were available from 450 of the samples. We also obtained 93 DNA samples from non-malignant lung tissue of European-descent patients with lung cancer in Italy and 250 DNA samples of peripheral blood mononuclear cells (PBMCs) from healthy individuals of European descent ( $n = 75$ ), African-Americans ( $n = 75$ ), and Mexican-Americans ( $n = 100$ ) enrolled in ongoing epidemiological studies in the US for investigation of frequencies of the polymorphisms (Table 2). Institutional Review Board permission and informed consent were obtained at each collection site.

### DNA Extraction

Genomic DNA was isolated from cell lines, frozen primary tumors, and non-malignant tissues by digestion with 100 µg/ml proteinase K (Life Technologies) followed by standard phenol-chloroform (1:1) extraction and ethanol precipitation [25].

### EGFR Gene Mutations

Details about *EGFR* mutation types and methodologies for mutation detection have been published elsewhere [9]. Briefly, we sequenced exons 18–21 of the TK domain of *EGFR* in tumor and corresponding non-malignant tissues. The overall frequency of mutation was 20%, and there were

**Table 1.** Summary of Investigations Performed, Results, and Their Implications

Investigation	Finding	Implication
Ethnic differences in <i>EGFR</i> polymorphisms in <i>CA-SSR1</i> length	<i>CA-SSR1</i> was longer in East Asians than in individuals of European descent, both for shorter allele and for combined allele length	For all three polymorphisms (shorter <i>CA-SSR1</i> length and variant forms of SNPs -216 and -191), the forms associated with increased <i>EGFR</i> protein production are rarer in East Asians
Ethnic differences in <i>EGFR</i> polymorphisms in SNP -216	Variant forms G/T and T/T were more common in individuals of European descent	The forms of the polymorphisms associated with increased protein production tend to co-segregate in lung cancer patients
Ethnic differences in <i>EGFR</i> polymorphisms in SNP -191	Variant forms C/A and A/A were more common in individuals of European descent	
Relationship between <i>CA-SSR1</i> and SNP polymorphisms	NSCLC patients with rare forms of SNPs -216 and -191 had shorter combined allele length for <i>CA-SSR1</i>	
Relationship between SNP -216 variants and <i>EGFR</i> mRNA expression	HBECs that have variant forms tended to make more <i>EGFR</i> mRNA	For SNP -216, data are consistent with higher protein production being associated with the minor form
Effect of <i>CA-SSR1</i> allele length on survival in patients with NSCLC	Patients with longer allele lengths had improved survival	The data are consistent with the concept that patients with less intrinsic protein production have improved survival in the absence of TKI therapy
<i>EGFR</i> mutations in NSCLC	Mutations were present in 25% of cases, and more common in East Asians (35.6%) than in individuals of European descent (11.3%)	This finding confirms previous reports that NSCLC tumors in East Asians have a higher incidence of <i>EGFR</i> mutations
Relationship between <i>EGFR</i> mutations and <i>CA-SSR1</i> AI	Mutations were more frequent in tumors with AI, especially those arising in East Asians and those with SAD	Mutations and AI frequently occur together in East Asian NSCLC tumors with SAD
Determination of whether AI targets mutant or WT allele	In NSCLC cases having both AI and mutation, the copy number of the mutant allele was preferentially increased compared to that of the WT allele	AI preferentially targets the mutant allele

doi:10.1371/journal.pmed.0040125.t001

three kinds of mutations, in-frame deletions in exon 19, missense mutations (predominantly mutation L858R in exon 21, but also in exons 18 or 20), and in-frame duplications/insertions of one to three codons in exon 20. The resistance-associated T790M mutation in exon 20 [9] was not detected in any tumor.

#### Analysis of *EGFR* Polymorphic Sites

We sequenced genomic DNA encompassing the SNP sites in the promoter region of *EGFR* -216 and -191 as described previously [21], using a single PCR reaction.

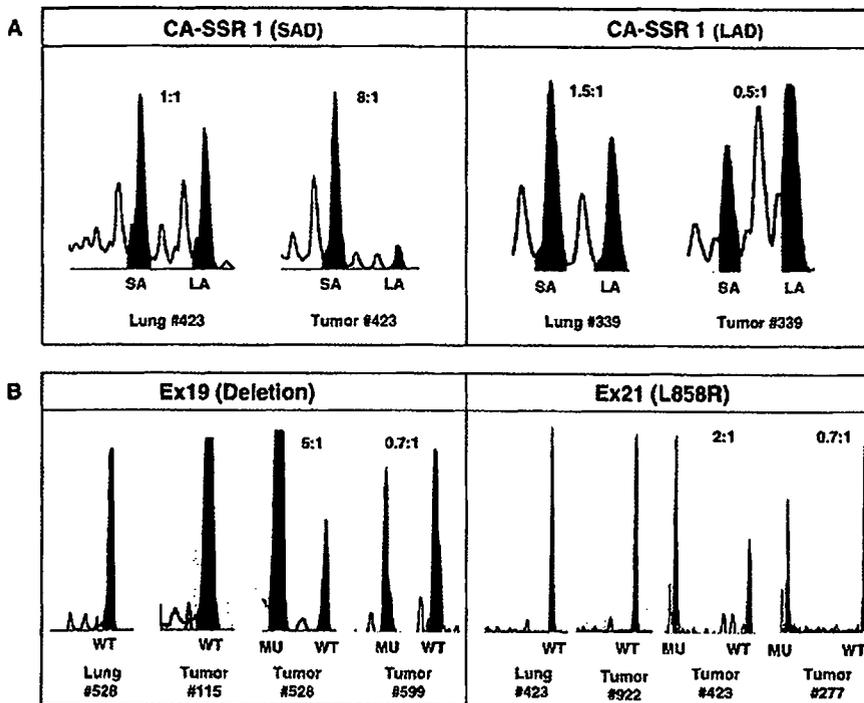
The *CA*-repeat-containing region of intron one was amplified by PCR. The sequences of the primers were 5'-CCA ACC AAA ATA TTA AAC CTG TCT T-3' (forward) and

5'-CTT GAA CCA GGG ACA GCA AT-3' (reverse). For analysis of repeat allele lengths and relative ratios, instrumentation and reagents from Applied Biosystems (Foster City, California, United States) were utilized. The reverse primer was labeled with TAMRA fluorescent dye (6-FAM) at the 5' end. The 25- $\mu$ l PCR reaction mixture contained 100 ng of genomic DNA, 10 $\times$  PCR buffer containing 15 mM MgCl<sub>2</sub>, 2 mM of each dNTP, 10 pmol of each primer, and 1.25 units of HotStart Taq DNA polymerase (Qiagen, Valencia, California, United States). After an initial denaturalization step at 95 °C for 12 min, samples were cycled 35 times as follows: 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s. The final extension was at 72 °C for 20 min. The size of the products (about 80 bp) was

**Table 2.** Summary of Germline (Blood) and Malignant and Non-Malignant Lung Tissues Examined

Sample	Ethnicity	Country					Total
		US	Australia	Japan	Taiwan	Italy	
Healthy individuals without cancer	Individuals of European descent	75					75
	African-Americans	75					75
	Mexican-Americans	100					100
	Total	250					250
Non-malignant lung tissue from NSCLC patients	Individuals of European descent	133	71			93	297
	East Asians	4	1	187	48		240
	Others	7					7
	Total	144	72	187	48	93	544
Malignant lung tissue from NSCLC patients	Individuals of European descent	142	71				213
	East Asians	4	1	251	80		336
	Others	7					7
	Total	153	72	251	80		556

doi:10.1371/journal.pmed.0040125.t002



**Figure 1.** Determination of AI for Heterozygous for *CA-SSR1* and for Tumors Having a Deletion Mutation in Exon 19 or the L858R Mutation in Exon 21. Representative wave patterns are illustrated for (A) the *CA-SSR1* allele and (B) the deletion mutation in exon 19 or L858R mutation in exon 21. Both tumors and corresponding lung tissue were analyzed. Note in (A) the ratio of shorter allele to longer allele is actually 1.3:1, as illustrated for lung #423, due to artifactual preferential amplification of the short allele. Thus, an appropriate correction factor is applied. doi:10.1371/journal.pmed.0040125.g001

confirmed by electrophoresis on 2% agarose gels. After PCR, 1  $\mu$ l of the product plus 0.5  $\mu$ l of Genescan 500 ROX molecular weight standard were denatured in 12  $\mu$ l of Hi-Di Formamide (Applied Biosystems) and separated with a Prism Genetic Analyzer and analyzed by Gene Scan Analysis software 3.1 (Applied Biosystems).

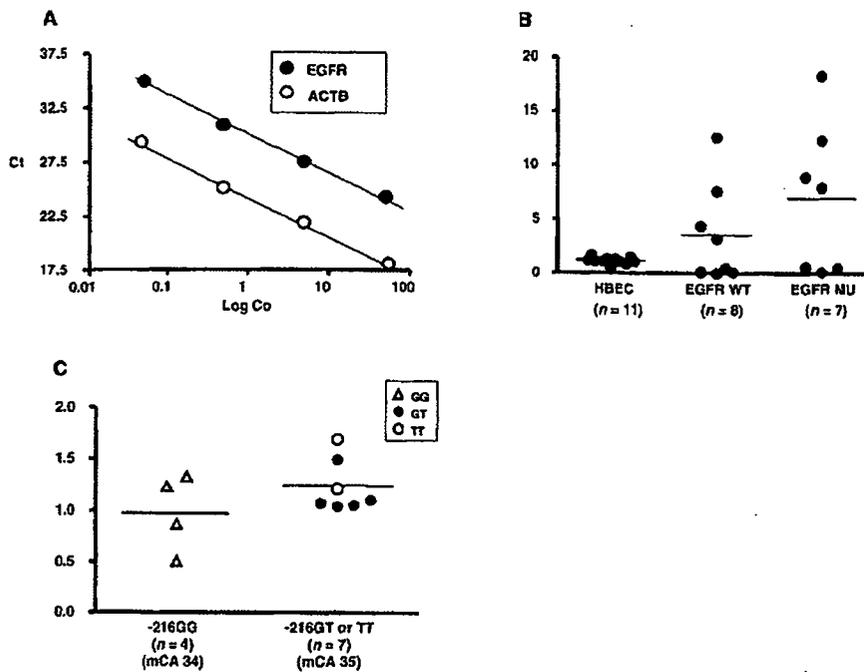
Examination of the resultant traces demonstrated that biallelic (heterozygous) samples showed two sets of waves and two peaks, while the monoallelic (homozygous) samples showed a single set of waves and one peak (Figure 1). The highest peak reflects the repeat number of the *CA-SSR1* allele as determined by the size marker, while the preceding waves (stutter bands) represent PCR-induced artifacts. In samples without AI the shorter peak appears artificially larger as a result of preferential PCR amplification. In non-malignant lung tissue the alleles were presumed to be of equal size, and their ratios were used as a correction factor for this artificial discrepancy.

The degree of the amplification of each allele was indicated by the area under the peak as determined by software provided by the instrument's manufacturer. The relative ratios (AI ratios), termed LOH score in previous reports, of the two peaks (shorter peak area under the curve to longer peak area under the curve) in tumor samples were calculated as previously described [26]. The AI ratio was calculated thus: AI ratio =  $(T1 \times N2)/(T2 \times N1)$ , where T indicates tumor, N indicates normal, 1 indicates the area under the peak for the shorter allele, and 2 indicates the area under the peak for the longer allele.

As either peak could be increased in relative size, AI cases were divided into shorter allele dominant (SAD) or longer

allele dominant (LAD) cases. We used the definitions of these two categories as determined previously [26]. SAD cases are defined as cases in which the adjusted AI ratio was greater than 1.27, and LAD cases were those in which the adjusted AI ratio was less than 0.79. For LAD cases, the formula results in ratio values less than unity. Therefore, the ratio was inverted for LAD cases, allowing the AI ratios to reflect the relative size of the longer allele, irrespective of which allele was increased in relative size. We confirmed the previous finding that the ratios of the areas under the curve for the two alleles in constitutional DNA on repeat testing or from different individuals are relatively constant. From an analysis of constitutional DNA from over 500 healthy individuals and cancer patients, we determined that the mean ratio of the two alleles in non-malignant tissues was 1.3, resulting from artifactual preferential amplification of the shorter allele (data not shown). For tumor samples lacking corresponding non-malignant tissue, the AI was determined by the formula AI ratio =  $T1/(T2 \times 1.3)$ .

The primers for investigation of selective amplification of the mutant or wild-type (WT) allele of exon 19 in-frame deletions and the exon 21 point mutation L858R were designed as follows: 5'-TCA CAA TTG CCA GTT AAC GTC T-3' (forward) and 5'-CAG CAA AGC AGA AAC TCA CAT C-3' (reverse) for exon 19, and 5'-ATG AAC TAC TTG GAG GAC CGT C-3' (forward) and 5'-TGC CTC CTT CTG CAT GGT ATT C-3' (reverse) for exon 21. Each forward primer was labeled with TAMRA fluorescent dye (6-FAM) at the 5' end. The conditions for PCR were the same as for *CA-SSR1* except for the annealing temperature (57  $^{\circ}$ C for exon 19 and 61  $^{\circ}$ C for exon 21). The PCR products of exon 21 were cut by



**Figure 2.** Relationship between SNP -216 Variants and EGFR mRNA Expression in HBEC Cultures

(A) Standard curves of EGFR and ACTB. Both slopes of cycle threshold (Ct)/log copies (Log Co) were mostly coincidental.

(B) Comparison of relative ratio of EGFR/ACTB among three groups of cultured cells (HBECs, lung cancer cell lines without EGFR mutations [WT], and lung cancer cell lines with EGFR mutations [MU]).

(C) Comparison of relative ratio of HBECs having SNP -216 G/G versus G/T or T/T. mCA, mean number of CA-SSR1 repeats.

doi:10.1371/journal.pmed.0040125.g002

the restriction enzyme *Sau96I* (New England BioLabs, Ipswich, Massachusetts, United States) and analyzed. The size of each product (about 142 bp for mutant alleles of exon 19, 158 bp for the WT allele of exon 19, 100 bp for mutant the allele of exon 21, and 150 bp for the WT allele of exon 21) was also confirmed by electrophoresis in 2% agarose gels.

The ratio (mutant allele/WT allele) to define amplification of each mutant allele, exon 19 in-frame deletion or the L858R point mutation, was determined by ROC (receiver operating characteristics) curves using the definitive value of AI, 1.27 (data not shown). The definitive ratios for exon 19 and 21 were 0.82 (sensitivity 70%, specificity 68%) and 0.2 (sensitivity 90%, specificity 90%), respectively, and the combined definitive ratio was 0.47 (sensitivity 70%, specificity 61%). We used these ratios as cut-off values to determine whether the mutant allele was amplified. Because of the presence of various amounts of non-malignant cells in the tumor samples,

amplifications of the WT allele could not be determined with certainty.

#### Real-Time PCR for the Expression of EGFR mRNA

cDNA was prepared by reverse transcription of 2  $\mu$ g of RNA from cell lines using SuperScript II reverse transcriptase according to the manufacturer's protocol (Invitrogen, Carlsbad, California, United States). Real-time PCR was performed with the Sybr (SYBR) Green I method using Power SYBR Green PCR Master Mix (Applied Biosystems). ACTB cDNA was used as an internal control. Primer sequences were as follows: 5'-ATA GTC GCC CAA AGT TCC GTG AGT-3' (forward) and 5'-ACC ACG TCG TCC ATG TCT TCT TCA-3' (reverse) for EGFR and 5'-AGT CCT GTG GCA TCC ACG AAA CTA-3' (forward) and 5'-ACT GTG TTG GCG TAC AGG TCT TTT-3' (reverse) for ACTB. Standard curves for EGFR and ACTB were obtained (Figure 2A), and the relative expression ratios of EGFR:ACTB were calculated.

**Table 3.** The Distribution of EGFR Genotypes by Ethnicity for Lung Cancer Patients

NSCLC Patients	SNP -216			SNP -191		
	G/G	G/T or T/T	p-Value <sup>a</sup>	C/C	C/A or A/A	p-Value <sup>a</sup>
Individuals of European descent (n = 306)	39.7%	60.3%	<0.001 <sup>b</sup>	63.0%	37.0%	<0.001 <sup>b</sup>
East Asians (n = 331)	93.4%	6.6%		99.4%	0.6%	

<sup>a</sup>Chi-square test. No significant gender differences were present ( $p = 0.194$ , Fisher's exact test).

<sup>b</sup>General linear regression adjusting for gender, age, smoking, histology, and EGFR mutations.

doi:10.1371/journal.pmed.0040125.t003

**Table 4.** Ethnic Differences in Distribution of the Allele Lengths of CA-SSR1 in Lung Cancer Patients

NSCLC Patients	Shorter Allele Length		Longer Allele Length		Combined Allele Length	
	Mean (SD)	p-Value <sup>a</sup>	Mean (SD)	p-Value <sup>a</sup>	Mean (SD)	p-Value <sup>a</sup>
Individuals of European descent (n = 306)	16.6 (1.4)	<0.001 <sup>b</sup>	18.5 (1.8)	<0.001 <sup>b</sup>	35.1 (2.8)	<0.001 <sup>b</sup>
East Asians (n = 331)	17.9 (2.0)		19.8 (1.2)		37.7 (2.7)	

<sup>a</sup>Two-sample t-test. No significant gender differences were present ( $p = 0.194$ , Fisher's exact test). Therefore, gender is not adjusted for in the comparisons.

<sup>b</sup>General linear regression adjusting for gender, age, smoking, histology, and EGFR mutations.  
SD, standard deviation.

doi:10.1371/journal.pmed.0040125.t004

### Statistical Analyses

We used the Chi-square test (testing the null hypothesis of equal distributions across study groups) to compare the distributions across study groups when outcomes were discrete such as genotypes of the SNP or SAD frequencies. When events were rare, e.g., where the expected cell counts were less than five, Fisher's exact test was used instead for comparisons. We also used Chi-square for an independent test for the assessment of each ethnic group using the Hardy-Weinberg equilibrium model. When outcomes were continuous, such as CA-SSR1 repeat numbers, two-sample t-test and analysis of variance were used. In order to control for potential confounding bias in comparisons of SNP and CA-SSR1 distributions, the multivariate logistic and general linear regression models were used with certain clinicopathological factors such as age, gender, smoking status, and histology as covariates (Tables 3–6). AI ratios of CA-SSR1 plotted against mutant/WT ratios are shown in Figure 3 with the fitted regression lines. The associations between AI ratios and mutant/WT ratios were tested using Pearson's correlation in case of small sample size and extreme values, the nonparametric Wilcoxon rank sum test was used to compare mutant/WT ratios for those with and without SAD. In this paper, all statistical tests and 95% confidence intervals are two-sided. Because of multiple tests,  $p$ -values less than 0.01 were judged to be statistically significant, and  $p$ -values less than 0.05 were judged as moderately significant. Both positive and negative results are reported in the tables and in the text.

### Results

Because of the complex nature of the findings and their interrelationships, a tabular summary of our major findings is presented in Table 1.

### Ethnic Differences in Distribution of Polymorphisms

We examined ethnic differences in the distribution of the minor alleles of the two SNPs –216 and –191 in the promoter region of the EGFR gene and mean CA-SSR1 repeat numbers. A summary of the samples studied from healthy individuals and cancer patients is presented in Table 2. For healthy US individuals, the frequencies of the –216 genotypes showed a borderline statistically significant difference between individuals of European descent, African-Americans, and Mexican-Americans ( $p = 0.08$ ) (Dataset S1). The G/G genotype was present in 46.7% (95% confidence interval: 35.4%–58.0%) of individuals of European descent compared to 60% (48.9%–71.1%) and 63% (53.5%–72.5%) of African-Americans and Mexican-Americans, respectively. The frequencies of the minor forms of the –191 polymorphism were significantly lower ( $p < 0.001$ ) in African-Americans (10.7%, 3.7%–17.7%) than in individuals of European descent (36%, 25.1%–46.9%) and Mexican-Americans (43%, 33.3%–52.7%). Also, the mean CA-SSR1 repeat number was significantly shorter in individuals of European descent (for the shorter, longer, or combined allele lengths) than in African-Americans and Mexican-Americans (combined allele length for individuals of European descent, 35.3, 34.7–35.9, for African-Americans, 36.2, 35.6–36.8, and for Mexican-Americans, 36.8, 36.3–37.3;  $p = 0.001$ ). The differences between African-Americans and Mexican-Americans were relatively modest and only reached significance for the shorter allele length (Dataset S1).

Among US European-descent individuals in this study, there were no significant differences in the frequency of the three polymorphisms between the healthy individuals (DNA from PBMCs) and those with NSCLC (DNA from non-malignant tissue). As shown in Table 3 and Dataset S1, the –216 G/G form was present in 46.7% (35.4%–58.0%) of the healthy individuals and 39.7% (30.8%–47.4%) of the patients with lung cancer ( $p = 0.321$ ), and the –191 C/C genotype was

**Table 5.** The Relationship between Repeat Length of CA-SSR1 and SNPs

CA-SSR1 Allele	SNP –216			SNP –191		
	G/G	G/T or T/T	p-Value <sup>a</sup>	C/C	C/A or A/A	p-Value <sup>a</sup>
Shorter	17.7 (1.9)	16.3 (1.0)	<0.001	17.3 (1.9)	16.8 (1.2)	0.084
Longer	19.6 (1.3)	18.2 (1.8)	<0.001	19.2 (1.7)	18.8 (1.3)	0.011
Combined	37.3 (2.8)	34.5 (2.4)	<0.001	36.6 (3.1)	35.6 (2.3)	0.011

Data are given as mean repeat length (standard deviation).

<sup>a</sup>Two-sample t-test after adjustment for ethnicity.

doi:10.1371/journal.pmed.0040125.t005

**Table 6.** Ethnic Differences in the Relationship between the Length of CA-SSR1 and SNPs -191 and -216

SNP	Genotype	East Asians		Individuals of European Descent	
		Percentage with Shorter Combined CA-SSR1 <sup>a</sup>	p-Value <sup>b</sup>	Percentage with Shorter Combined CA-SSR1 <sup>a</sup>	p-Value <sup>b</sup>
-216	G/G	36.3%	0.001	53.7%	<0.001
	G/T or T/T	72.7%		81.2%	
-191	C/C	38.3%	0.149	76.2%	0.017
	C/A or A/A	100.0%		59.5%	
Both	C/C + G/G	35.8%	<0.001	59.0%	0.119
	Others	75.0%		72.7%	

<sup>a</sup>Having a shorter combined allele is defined as having 36 or fewer combined CA-SSR1 repeats. (This cut-off is based on the overall mean length of the combined allele for both ethnic groups. Using the individual group means for East Asians and individuals of European descent as the cut-offs shows the similar results.)

<sup>b</sup>p-Values are from Fisher's exact tests.

doi:10.1371/journal.pmed.0040125.t006

present in 64% (53.1%–74.9%) of the healthy individuals and 63% (54.8%–71.2%) of the patients with cancer ( $p = 0.941$ ). Also, the mean CA-SSR1 repeat numbers for the short allele, long allele, and combined alleles of healthy European-descent individuals were not significantly different from those of European-descent patients with cancer ( $p = 0.492$ ,  $0.604$ , and  $0.495$ , respectively) (Table 4; Dataset S1). These data permitted us to presume that the polymorphism frequencies in patients with lung cancer follow the pattern of the general population, and we can combine the data from healthy individuals and patients with NSCLC for individuals of European descent, which is the dominant ethnicity of the US, Italy, and Australia populations in this study. Furthermore, no significant differences were observed in this study for the frequencies of all three polymorphisms between individuals of European descent in the US versus in Italy, nor between East Asians in Japan versus in Taiwan (data not shown). Thus, we pooled the data from these two groups and labeled them as “individuals of European descent” and “East Asians,” which were then used for further analyses.

Comparing individuals of European descent and East Asians, the frequency of the minor forms of the -216 polymorphism was significantly higher ( $p < 0.001$ ) in individuals of European descent (60.3%, 54.8%–65.8%) than in East Asians (6.6%, 3.9%–9.3%). This was also true for the minor forms of the -191 polymorphism (individuals of European descent, 37.0%, 31.6%–42.4%; East Asians, 0.6%, 0%–1.4%;  $p < 0.001$ ), as shown in Table 3. In addition, Table 4 shows that both alleles of CA-SSR1 (and the combined allele length) were significantly shorter in individuals of European descent than in East Asians ( $p < 0.001$ ). The comparisons were controlled for potential confounders such as gender, age, and smoking.

#### Relationship between CA-SSR1 Allele Lengths and SNPs

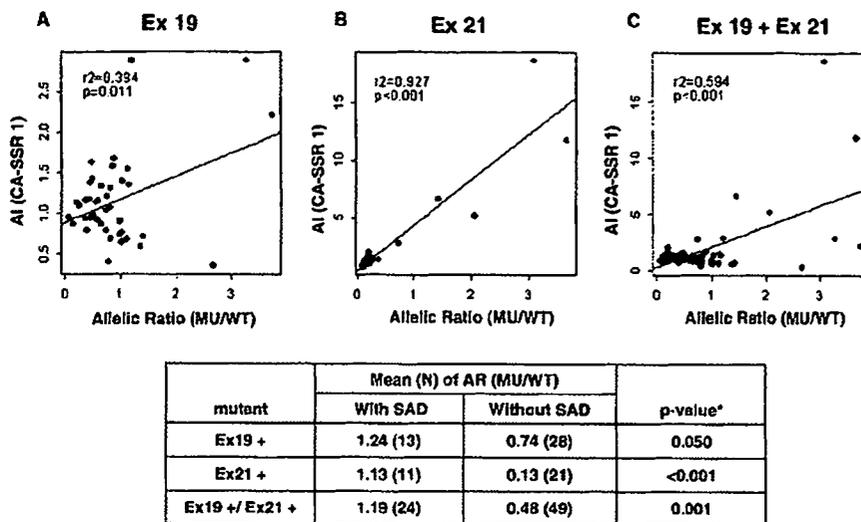
We first examined the concordance of the SNP -216, SNP -191, and CA-SSR repeat polymorphisms. As shown in Table 5, individuals who were homo- or heterozygous for the variant forms of SNP -216 (G/T or T/T) had significantly lower mean CA-SSR repeat numbers in short, long, and combined allele lengths than those who were homozygous for the common form -216 G/G after adjustment for ethnicity. In similar comparisons for the variant forms of SNP -191, there was significant concordance with the longer and combined allele lengths, but not for the shorter allele.

We next investigated the relationship between the combined allele length and the SNPs for different ethnicities. For convenience, since the overall mean CA-SSR1 repeat number for shorter and longer allele combined was 36, we dichotomized the combined allele length as “longer” for those with greater than 36 repeats and as “shorter” for those with 36 repeats or fewer. As shown in Table 6, the frequency of the “shorter” combined allele was significantly higher in individuals with the minor forms of -216 (East Asians, 72.7%, 54.1%–91.3%; individuals of European descent, 81.2%, 76.1%–86.3%) than in those with the common form (East Asians, 36.6%, 30.9%–41.7%; individuals of European descent, 53.7%, 45.8%–61.6%). A similar pattern for SNP -191 was noted in East Asians but not in individuals of European descent. Also, for individuals carrying both variant genotypes of the two SNPs, the frequency of the “shorter” combined allele was observed to be higher than in those with the common forms of the SNPs in both individuals of European descent and East Asians, although the difference was statistically significant only in East Asians (Dataset S2).

#### Relationship between EGFR Expression and the -216 Polymorphism

The polymorphism genotype of the 11 HBEC cultures was determined as previously described. The lines, derived from American individuals of European descent, showed little variation in the repeat length of the shorter CA-SSR1 allele (mean length 16.2, range 16–17). Similarly, for the -191 polymorphism, ten of the cases had the common C/C genotype and only one case demonstrated the C/A genotype. Thus, we were unable to study the effects of these two polymorphisms on gene expression in the HBEC cultures. However, for the -216 polymorphism, four of the cases had the common form, G/G, while the remaining seven cases expressed the variant forms G/T ( $n = 5$ ) or T/T ( $n = 2$ ). Thus, we limited our examination of the relationship of SNPs to EGFR expression to the -216 polymorphism (Figure 2B and 2C).

The standard curves for ACTB and EGFR mRNA expression were straight lines nearly parallel to each other (Figure 2A), permitting us to use the expression ratio of these two genes for comparisons. To further validate our assays, we determined the ratios for the HBECs as well as for eight NSCLC cell lines having the WT form and for seven cell lines having a mutant form of the EGFR gene. As expression in normal



\* Wilcoxon Rank Sum test.

**Figure 3.** The Correlation between AI and Allelic Ratio

The correlation between allelic ratio of *CA-SSR1* (shorter allele/longer allele) and the allelic ratio (AR) of mutant (MU) to WT allele of (A) the exon 19 in-frame deletion ( $r^2 = 0.394$ ,  $p = 0.011$ ), (B) the exon 21 L858R point mutation ( $r^2 = 0.927$ ,  $p < 0.001$ ), or (C) both ( $r^2 = 0.594$ ,  $p < 0.001$ ) in the same mutant cases.

doi:10.1371/journal.pmed.0040125.g003

epithelial cells is low or not detectable in the absence of ligand, the HBECs were cultured in EGF-containing medium (5 ng/ml). Expression in the HBECs was relatively low, with a narrow range (Figure 2B). The lung cancer lines, grown in the absence of added ligand, showed considerable variability of expression. Four WT lines had low expression, while four lines, all having *EGFR* copy number of four or greater, had considerably higher expression levels. Four of the mutant lines, all highly amplified for copy number and lacking the secondary resistance-associated T790M mutation [27,28], had high expression ratios. However three mutant lines had low expression ratios. Two of these lines had the secondary T790M mutation as well as an activating mutation, while the third line had a relatively low copy number.

While the range of expression in the HBECs was modest, we correlated expression with the -216 genotype (Figure 2C). The four lines having the G/G phenotype had a mean expression ratio of 1.0 (range 0.5–1.3). The seven lines having one of the two variant forms had a mean expression ratio of 1.2 (range 1.0–1.7). The two lines homozygous for the variant form T/T were among the three highest expressing lines. While these differences were not significant, they may represent a trend towards higher expression being associated with the variant forms.

The range of relative expression of *EGFR* compared to *ACTB* of lung cancer cell lines was variable. The two high values were observed in the cell lines with *EGFR* mutation. The mean value of cell lines having the common SNP -216 G/G ( $n = 4$ ) was 0.97, compared to 1.24 for the lines with the minor forms SNP -216 G/T or T/T ( $n = 7$ ) (Figure 2C). The range of the number of *CA-SSR1* repeats in the cell lines, all from individuals of European descent, was from 16 to 17 for the shorter allele, 16 to 19 for the longer allele, and 32 to 38 for the combined length. The highest value was observed in the group with the shortest combined number of *CA-SSR1* repeats (32) and one of the minor SNP -216 forms.

#### The Relationship between Polymorphisms and Survival

We also investigated the relationship between the SNP -216, SNP -191, and *CA-SSR* repeat polymorphisms and patient overall survival (Figure S1). We did not observe a relationship between survival and either SNP form or any combination of SNP forms after adjusting for age, gender, ethnicity, smoking, and histology. For the shorter allele of *CA-SSR1* in the tumor cases, the mean length was 17.5. We divided the cases into those having shorter alleles, with mean lengths of 17 or fewer repeats, and those having a mean length of 18 or more repeats. We found that cases having a mean length of 18 or more repeats had improved survival compared to those having shorter allele lengths of 17 or fewer repeats ( $p = 0.017$ ). These findings suggest that patients (in the absence of TKI therapy) whose tumor cells are predicted to make less *EGFR* protein have an improved survival compared to those whose cells are predicted to have higher intrinsic protein production. Similar data have been reported recently from another group [29]. For cases with AI of *CA-SSR1* (see below) or of the mutant allele, no differences in patient survival were noted (data not shown).

#### AI of the *CA-SSR1* Alleles

The degree of amplification of each allele was reflected by the relative area under the peak (Figure 1), and the AI was determined by the ratio of shorter to longer *CA-SSR1* alleles in informative cases where two alleles were of different length. Among 450 tumor cases where the corresponding non-malignant lung tissues were available, there was no difference in the presence of homo- or heterozygosity of allele length or in the repeat length of each allele between tumor and non-malignant tissues (data not shown). These findings permitted us to analyze all 556 cases using the tumor tissues alone. For the *CA-SSR1* alleles, 376 (68%) of 556 cases were informative. The informative rate was similar to that in other previous studies [16,26]. However, in our study the

**Table 7.** Frequencies of AI of Either Allele of *CA-SSRI* by Ethnicity

NSCLC Patients	AI	<i>p</i> -Value <sup>a</sup>	Mutant or WT <i>EGFR</i> Allele <sup>b</sup>	AI	<i>p</i> -Value <sup>a</sup>
All cases ( <i>n</i> = 356)	109/356 (30.6%)		MT ( <i>n</i> = 90)	40/90 (44.4%)	0.002
East Asians ( <i>n</i> = 205)	73/205 (35.6%)	0.019	WT ( <i>n</i> = 266)	69/266 (25.9%)	
Individuals of European descent ( <i>n</i> = 151)	36/151 (23.8%)		MT ( <i>n</i> = 73)	34/73 (46.6%)	0.022
			WT ( <i>n</i> = 132)	39/132 (29.5%)	
			MT ( <i>n</i> = 17)	6/17 (35.3%)	0.24
			WT ( <i>n</i> = 134)	30/134 (22.4%)	

These analyses were limited to informative cases of East Asians from Japan or Taiwan and individuals of European descent from the US and Australia.

<sup>a</sup>Chi-square test with continuity adjustment.

<sup>b</sup>Mutant (MT) *EGFR* alleles are limited to exon 19 deletions and exon 21 L858R.

doi:10.1371/journal.pmed.0040125.t007

informative rate was not consistent across ethnicities: there was an informative rate of 62.8% (211/336) in East Asians and 75.0% (165/220) in other ethnicities. Of the 376 informative cases, we excluded cases with mutations other than deletions in exon 19 or the L858R mutation in exon 21 (*n* = 12) as well as patients of ethnicities other than East Asians and individuals of European descent (*n* = 5) and Asians in the US (*n* = 3). Of the remaining 356 NSCLC cases of East Asian or European descent, 263 had the WT *EGFR* gene and 95 had the mutations in exon 19 or exon 21.

For these 356 cases, we determined the ratios of the *CA-SSRI* alleles as previously described in the Methods section. AI, defined by an allelic ratio greater than 1.27 or less than 0.79, was present in 109 (30.6%) of the cases but was significantly more frequent (*p* = 0.002) in cases with mutant tumors (44.4%, 34.1%–54.7%) than in those with WT tumors (25.9%, 20.6%–31.2%), and in East Asians (35.6%, 29.0%–42.2%) than in individuals of European descent (23.8%, 17.0%–30.6%) (*p* = 0.019) (Table 7; Dataset S3).

The 109 cases with AI were also divided into SAD or LAD. As shown in Table 8 (and Dataset S3), the overall frequency of SAD was 60.3% (49.1%–71.5%) in East Asians and 44.4% (28.2%–60.6%) in individuals of European descent. Also, in East Asians the SAD frequency was significantly higher (*p* = 0.001) in tumors with the exon 19 or exon 21 mutation than in those without mutations (82.4%, 69.6%–95.2%, versus 41.0%, 25.6%–56.4%). This difference, however, was not observed in patients of European descent.

### AI of Mutant to WT Allele

For cases with the deletions in exon 19 or the L858R mutation in exon 21, the AI of the mutant allele was determined by the mutant/WT allele ratio. A flow chart describing the process of case selection and exclusion is presented in Figure 4. These mutant cases gave us an opportunity to examine the association between AI in amplification of *CA-SSRI* repeats and AI in the ratio of mutant to WT alleles. Specifically, we wished to determine, in cases having both forms of AI, whether the mutant form was selectively amplified in association with selective amplification of the shorter allele of *CA-SSRI*. As described in the Methods section, we devised methods to determine the ratios of mutant to WT alleles for the two most frequent mutations, deletions in exon 19 and the L858R mutation in exon 21, which together account for ~85% of *EGFR* mutations in NSCLC [9]. Of the 109 cases with mutations (in exon 19 or L858R), sufficient DNA was available from 76. Of these 76 samples, 32 (42.1%) tumors had selective imbalance involving the mutant allele. The ratio of *CA-SSRI* alleles was utilized to determine whether AI was present and, if present, which of the two alleles was preferentially overrepresented. Of these 32 samples having AI of the mutant allele, 26 (81.3%) also had AI of *CA-SSRI*. In addition, a positive correlation between AI ratios of *CA-SSRI* and mutant/WT ratios was observed in tumors having either form of mutation (Figure 3). The linear correlation was tested using Pearson's correlation and found to be significant. However, because of the possibility that the observed strong correlation might be driven by extreme

**Table 8.** Frequencies of AI of *CA-SSRI* by Ethnicity

Cases with AI of <i>CA-SSRI</i>	SAD <sup>a</sup>	<i>p</i> -Value <sup>b</sup>	Mutant or WT <i>EGFR</i> Allele <sup>c</sup>	SAD <sup>a</sup>	<i>p</i> -Value <sup>b</sup>
All cases ( <i>n</i> = 109)	60/109 (55.0%)		MT ( <i>n</i> = 40)	30/40 (75.0%)	0.003
East Asians ( <i>n</i> = 73)	44/73 (60.3%)	0.214	WT ( <i>n</i> = 69)	30/69 (43.5%)	
Individuals of European descent ( <i>n</i> = 36)	16/36 (44.4%)		MT ( <i>n</i> = 34)	28/34 (82.4%)	0.001
			WT ( <i>n</i> = 39)	16/39 (41.0%)	
			MT ( <i>n</i> = 6)	2/6 (33.3%)	0.672
			WT ( <i>n</i> = 30)	14/30 (46.7%)	

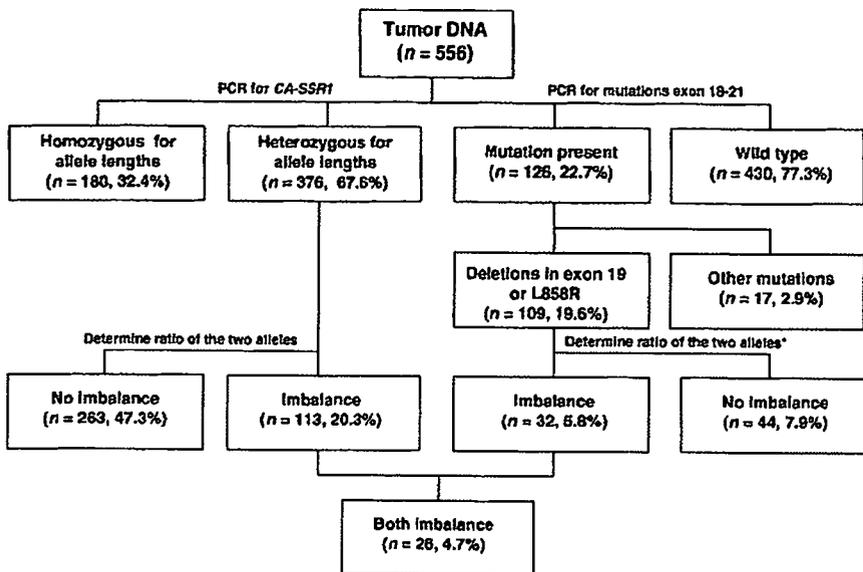
These analyses were limited to informative cases of East Asians from Japan or Taiwan and individuals of European descent from the US and Australia.

<sup>a</sup>Number of SAD cases (analyses limited to cases with AI).

<sup>b</sup>Chi-square test with continuity adjustment.

<sup>c</sup>Mutant (MT) *EGFR* alleles are limited to exon 19 deletions and exon 21 L858R.

doi:10.1371/journal.pmed.0040125.t008



\* 109 cases with mutation in exon 19 or L858R; sufficient DNA was available from 76.

Figure 4. Flow Chart for Examination of the Relationship between AIs of CA-SSRI Length and EGFR Mutations  
doi:10.1371/journal.pmed.0040125.g004

values given the small sample size of the available cases, we used a nonparametric test instead to compare mutant/WT ratios between those with SAD and those without. As expected, for all the mutations under study, the cases with SAD had higher mean mutant/WT ratios than those without SAD. These findings agreed with our hypothesis that in cases demonstrating CA-SSRI imbalance, the mutant allele was more frequently increased in relative copy number compared to the WT allele.

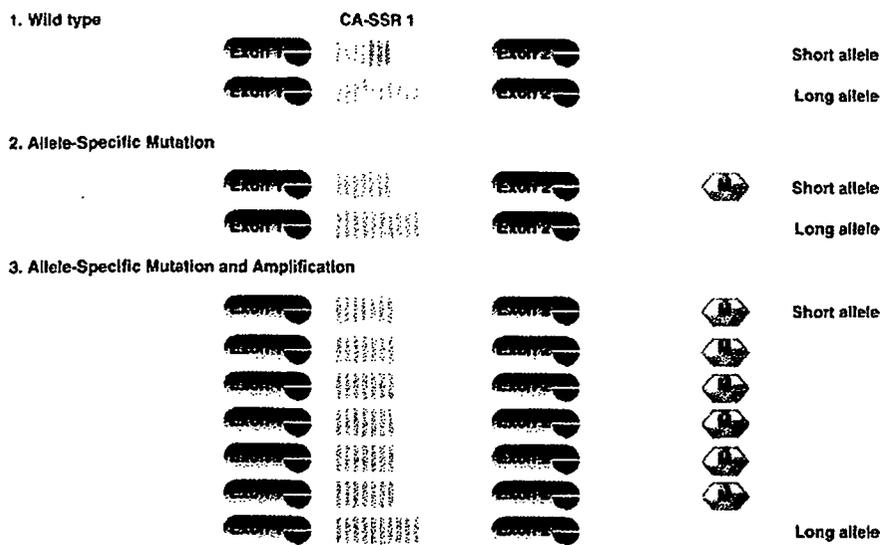
## Discussion

In this report we examined the frequency of three germline polymorphisms in the EGFR gene in healthy individuals of different ethnicities, and in non-malignant and malignant lung tissue from patients with NSCLC. We found ethnic-related differences in polymorphism frequencies consistent with previous reports, indicating that the shorter allele of CA-SSRI and the minor forms of SNPs -191 (C/A or A/A) and -216 (G/T or T/T) are significantly less frequent in East Asians than in individuals of European descent [21]. In addition, we noted a relationship between the presence of the short form of CA-SSRI and the minor forms of the SNPs. The published data [13,19,21,26] and our observations regarding EGFR mRNA expression in HBECs suggest that the shorter CA-SSRI allele lengths and the variant forms of the -191 and -216 polymorphisms are associated with increased intrinsic gene expression. However, most of the data in the literature are from the results of transfection studies or tumor cell lines, and thus may not reflect the state of normal epithelial cells. As sections of non-malignant lung contain only a small minority of epithelial cells, a study of adjacent non-malignant lung tissues from resected cases or peripheral blood cells would not yield meaningful data. In an attempt to overcome these limitations, we studied 11 cultures of immortalized HBECs. These cultures show minimal genetic changes. In the

presence of ligand stimulation, we demonstrated a trend for increased mRNA expression in lines having the SNP -216 G/T or T/T genotypes, consistent with published data. The published reports and our results are consistent with the hypothesis that cells of individuals of East Asian ethnicity express less EGFR protein constitutively than cells of individuals of other ethnicities. However, final experimental proof for this hypothesis is still lacking.

Amplification of the EGFR gene is relatively common in lung and other cancers, and may be associated with mutations of the TK domain in lung cancers [12] or of the extracellular domain in glioblastomas [30]. Two recent reports describe a correlation between copy numbers of the EGFR gene as measured by fluorescence in situ hybridization (FISH) and response to TKIs [11,31]. In this study we used allelic size differences in the CA-SSRI repeat polymorphism to determine AI of the gene. AI was observed in 30.2% of informative cases, a frequency comparable to increased copy number as detected by FISH analyses [32]. AI was significantly more frequent in East Asians and occurred nearly twice as frequently in mutant cases than in WT cases. A relationship between increased copy number by FISH analysis and mutation has also been described previously [12]. While there were no significant differences in the frequencies of either the shorter or longer allele being involved in the imbalance for all of the cases or for all of the mutant cases, in mutant cases arising in East Asians, the shorter allele was twice as likely to be preferentially amplified as the longer allele.

Finally we determined whether the mutant allele was selectively amplified in tumors having both mutation and imbalance. For tumors having deletion mutations in exon 19 or the L858R point mutation in exon 21 (together accounting for 86.5% of all mutations) we devised methods for determining the ratio of mutant to WT alleles. Of 76 cases examined, 42.1% demonstrated imbalance of the mutant allele. This figure is consistent with our finding of an overall



**Figure 5.** Hypothesized Allele-Specific Mutation and Amplification of *EGFR* in Lung Cancers

We hypothesized that CA-SSR1 polymorphism occurs, mutations (M) target the *EGFR* allele with the shorter CA-SSR1 repeat number, and then there is allele-specific amplification. These three events, targeting the same allele, would be predicted to result in greater protein production than random allelic occurrence.

doi:10.1371/journal.pmed.0040125.g005

AI (from analysis of the *CA-SSR1* alleles) percentage of 45.3% in mutant cases, and suggests that in mutation-containing tumors having AI, the mutant allele is the one that is usually amplified. Having found, by separate analyses in mutant cases, that both the shorter *CA-SSR1* allele and the mutant allele were selectively amplified, we performed a correlation of these two forms of imbalance and demonstrated a strong positive association.

Incorporation of our findings and previously published data form the basis of a hypothesis suggesting a close relationship between *CA-SSR1* length, SNP -191 polymorphism, and SNP -216 polymorphism and *EGFR* gene amplification. As mentioned above, all three of these polymorphisms (shorter *CA-SSR1* length and the variant forms of the two SNPs) are reported to be associated with increased *EGFR* production, and they were rarely observed in East Asians. These findings suggest that the cells of most East Asians make less *EGFR* protein than do the cells of individuals of other ethnicities. If a certain critical level of *EGFR* is required to drive the cell toward a malignant phenotype, mutations of the TK domain and autonomous activation of downstream signaling may target East Asians, the subgroup with possibly lower intrinsic protein production. Also, we found in East Asians (but not in individuals of European descent) that mutations target the shorter *CA-SSR1* allele (suggestive of greater protein production) followed by allele-specific amplification of the mutant allele. As illustrated in Figure 5, three events target the same allele: (a) shorter *CA-SSR1* repeat length, (b) activating mutation, and (c) selective amplification of the mutant allele. These interactions favor greater protein production in mutant tumors. A similar observation was made in glioblastomas, which frequently contain a mutation or splicing variant resulting in loss of much of the extracellular domain of *EGFR*. The variant form of the allele frequently demonstrated allele-specific amplification [33]. As previously mentioned, FISH technology has

been used to demonstrate that *EGFR* amplification and mutation often, but not invariably, occur together [12].

## Conclusions

The three polymorphisms associated with increased *EGFR* protein production (shorter *CA-SSR1* length and the variant forms of SNPs -216 and -191) were found to be rare in East Asians as compared to individuals of other ethnicities, suggesting that the cells of East Asians may make relatively less intrinsic *EGFR* protein. Interestingly, especially in tumors from patients of East Asian ethnicity, *EGFR* mutations were found to favor the shorter allele of *CA-SSR1*, and selective amplification of the shorter allele of *CA-SSR1* occurred frequently in tumors harboring a mutation. These distinct molecular events targeting the same allele would both be predicted to result in greater *EGFR* protein production and/or activity. These findings may reveal what underlies some of the ethnic differences observed in mutational frequencies and responses to TKIs.

## Supporting Information

**Alternative Language Abstract S1.** Translation into Japanese by Masaharu Nomura

Found at doi:10.1371/journal.pmed.0040125.sd001 (27 KB DOC).

**Alternative Language Abstract S2.** Translation into French by Masaharu Nomura

Found at doi:10.1371/journal.pmed.0040125.sd002 (31 KB DOC).

**Alternative Language Abstract S3.** Translation into German by Masaharu Nomura

Found at doi:10.1371/journal.pmed.0040125.sd003 (31 KB DOC).

**Alternative Language Abstract S4.** Translation into Spanish by Masaharu Nomura

Found at doi:10.1371/journal.pmed.0040125.sd004 (31 KB DOC).

**Dataset S1.** Ethnic Differences in Polymorphisms

Found at doi:10.1371/journal.pmed.0040125.sd005 (37 KB DOC).