# Soy consumption reduces the risk of non-small-cell lung cancers with *epidermal growth factor receptor* mutations among Japanese

**Keitaro Matsuo,[1,6,7] Akio Hiraki,[1] Hidemi Ito,[1] Takayuki Kosaka,[2] Takeshi Suzuki,[1] Kaoru Hirose,[3] Kenji Wakai,[4] Yasushi Yatabe,[5] Tetsuya Mitsudomi[2] and Kazuo Tajima[1,6]**

[1]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute; Departments of [2]Thoracic Surgery and [5]Pathology and Molecular Diagnosis, Aichi Cancer Center Central Hospital, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681; [3]Department of Planning and Information, Aichi Prefectural Institute of Public Health, 7-6 Azanagare, Tsujimachi, Kita-ku, Nagoya 462-8576; Departments of [4]Preventive Medicine/Biostatistics and Medical Decision Making and [6]Epidemiology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan

(Received December 12, 2007/Revised January 19, 2008; February 17, 2008/Accepted February 18, 2008/Online publication April 21, 2008)

***Epidermal growth factor receptor (EGFR)*** **mutations play substantial roles in genesis and proliferation of non-small-cell lung cancers (NSCLCs). We recently found that reproductive factors have a substantial impact on risk of development of NSCLCs featuring such *EGFR* mutations. Therefore, we explored the influence of dietary habits on NSCLC risk with reference to the *EGFR* mutational status. We conducted a case-control study using 353 patients with NSCLCs (122 *EGFR* mutated and 231 *EGFR* wild-type) and 1765 age-sex matched non-cancer control subjects. Dietary exposure was based on a semiquantitative food frequency questionnaire and impact of major food items, like meats, seafoods, vegetables and soybean products was assessed by multivariate logistic regression. Soybean products demonstrated a protective association with *EGFR* mutated, but not *EGFR* wild-type NSCLCs, with multivariate-adjusted odds ratios and 95% confidence intervals for the 2nd and 3rd tertile of soybean product consumption of 0.79 (0.50–1.27) and 0.56 (0.34–0.93) relative to those in the lowest tertile (trend *P* = 0.023). In conclusion, soy consumption may exert a protective association against the development of NSCLCs with *EGFR* mutations, providing possible insights into mechanisms of their genesis. (*Cancer Sci* 2008; 99: 1202–1208)**

The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase that activates several signaling pathways resulting in cell proliferation, escape from apoptosis, and increased likelihood of invasion or metastasis, all of which are associated with the cancer phenotype.[1] Elevated levels of EGFR are frequently seen in a variety of epithelial tumors[2] including non-small-cell lung cancers (NSCLCs)[3,4] and activating mutations of *EGFR* such as deletion in Exon 19 or point mutation at codon 858 have been reported in a subset of NSCLCs which are usually highly sensitive to kinase inhibitors such as gefitinib and erlotinib.[5–7] Moreover, striking correlations between the *EGFR* mutation status and characteristics like a non-smoking background, adenocarcinoma-type, female sex, and East Asian ethnicity have been repeatedly demonstrated[5–9] albeit without clarification of the reasons as yet. Previously, we demonstrated a significance of female sex and reproductive factors for risk of *EGFR* mutated ($EGFR^{mut}$) NSCLCs and suggested possible involvement of estrogenic factors in their genesis.[10] Considering several pathological studies also demonstrated expression of hormonal receptors in NSCLCs, our former results motivated us to examine the impact of soy consumption, a primary source of the isoflavone phytestrogens well known to be associated with the hormone-related breast cancer.[11–14]

Here, to assess the significance of dietary soy consumption with reference to the *EGFR* status of NSCLCs, in parallel with other food items, we conducted a case-control study in 353 patients with NSCLCs and 1765 matched controls in a Japanese population.

## Materials and Methods

**Study subjects.** Subjects for this study were part of those in the previous report.[10] A total of 353 patients diagnosed as having primary NSCLCs at the Department of Thoracic Surgery, Aichi Cancer Center Hospital (ACCH) and treated between January 2001 and February 2005 and 1765 age- and sex-matched (1:5 case-control ratio) non-cancer patients who visited ACCH during the same period were examined. The reason for exclusion of the other 82 cases was lack of diet information by semiquantitative food-frequency-questionnaire (SQFFQ). Correspondingly, 410 matched-controls were excluded. All the cases underwent potentially curative pulmonary resection at the Department of Thoracic Surgery, ACCH. The distribution of histological subtypes for NSCLCs among the 353 patients was 280 adenocarcinomas (79.3%), 18 large cell carcinomas (5.1%), 46 squamous cell carcinomas (13.0%) and 9 (2.6%) miscellaneous-type lesions. Control subjects were eligible if they had no past history or current diagnosis of cancer.

All subjects were enrolled in the Hospital-based Epidemiological Research Program at ACCH (HERPACC) at the time of their first visit to our hospital. The study frame-work of this program has been detailed elsewhere.[15,16] Briefly, all first-visit outpatients at ACCH aged 18–79 years are asked to complete a self-administered questionnaire on lifestyle factors which was then checked by trained interviewers. The questionnaire includes items on demographic characteristics, medical history, smoking and drinking habits, regular physical exercise, as well as menstrual and reproductive history and dietary habits before the development of current symptoms. To date, approximately 95% of eligible subjects have completed the questionnaire. All data are loaded into a HERPACC database and periodically linked with the hospital cancer registry system to update the data for cancer incidence. Like most general hospitals in Japan, the ACCH accepts new outpatients who visit of their own volition, with or without a doctor's referral. Thus, even though the ACCH is called a cancer hospital, only 19% of all new outpatients have cancer.[17] Among non-cancer outpatients, 45% present with no abnormal findings on clinical examination and 35% with benign non-specific diseases.[18] We previously showed that the lifestyle patterns of first-visit outpatients accord with those of the general population randomly selected from residents of Nagoya City.[19] Therefore, non-cancer outpatients at ACCH can thus be regarded as appropriate controls for epidemiological studies. The present study was approved by the Institutional Ethics Review Board of ACCH, and all participants provided written informed consent.

[7]To whom correspondence should be addressed. E-mail: kmatsuo@aichi-cc.jp



doi: 10.1111/j.1349-7006.2008.00812.x

***EGFR* mutation analysis.** Assays for *EGFR* mutations were described elsewhere.[10] The majority of tumor samples were analyzed by reverse transcriptase-polymerase chain reaction (RT-PCR) coupled with direct sequencing. When frozen tissue for RT-PCR was not available, the cases were examined with a DNA-based assay using formalin-fixed, paraffin-embedded tissue. Both methods have been described in detail elsewhere.[8,20,21] Briefly, total RNA was isolated using an RNAeasy kit (Qiagen, Valencia, CA, USA), followed by RT-PCR with a QIAGEN OneStep RT-PCR kit (Qiagen). The purified PCR products were directly sequenced with an ABI PRISM 3100 instrument (Applied Biosystems, Foster City, CA, USA). For DNA extracted from paraffin-embedded tissues, we used the cycleave technique with the Smart Cycler system (SC-100, Cepheid, Sunnyvale, CA, USA) and fragment analysis with the ABI PRISM 3100 for detection of point mutations of codon 858 and gene deletions in exon 19, which together account for approximately 90% of mutations in the EGFR gene.[7] Comparability of two methods was 99% for condon 858 point mutation and deletion in exon 19 and was 95% for overall.[21]

**Measurement of lifestyle exposure.** All exposures were assessed from the self-administered questionnaire with checks by trained interviewers, as completed at the first visit to ACCH before the diagnostic procedure was conducted. Subjects were specifically questioned about their lifestyle before the onset of the symptoms which occasioned their visit to ACCH.

Smoking status was divided into three categories: never, former and current. Former smokers were defined as those who quit smoking at least 1 year before the time of the survey. Cumulative doses of smoking were evaluated as pack-years (PYs), the product of the numbers of packs consumed in one day and the number of years of smoking. Age at starting smoking and years since quitting were asked of current smokers and former smokers, respectively. Alcohol consumption of each type of beverage (Japanese sake, beer, shochu, whiskey and wine) was determined by the average number of drinks per day, which was then converted into a Japanese sake (rice wine) equivalent, the 'go' (180 mL) which contains 23 g of ethanol. Total alcohol consumption was estimated as the sum of pure alcohol consumption in grams.

Consumption of food items was based upon the SQFFQ, for which validity and reproducibility have been described elsewhere.[22–24] Our questionnaire covers 47 single food items with frequencies in the eight categories of never or seldom, 1–3 times/month, 1–2 times/week, 3–4 times/week, 5–6 times/week, once/day, twice/day, and three or more times/day. Values for the consumption of selected food groups (soybean, meat, fish, seafood other than fish, green-yellow vegetables, vegetables other than green-yellow vegetables, and fruits) were calculated as the sums of consumption of contributing single food items as estimated from the food frequency and portion size.

**Statistical analysis.** We used STATA version 8 (STATA Corp., College Station, TX, USA) for all analyzes. The impact of each lifestyle factor was evaluated with odds ratios (ORs) in unconditional logistic regression models adjusted for age, sex and smoking. Conditional logistic regression models were not applied to avoid sparsity of data in stratified analysis. Differences in continuous variables between groups were evaluated by the Mann–Whitney test when appropriate. Food items were categorized into tertiles (T1 to T3) according to the distribution among controls. Trend tests were accomplished with scores which were defined as median value of each category in unconditional logistic regression models adjusted for age, sex, energy intake and smoking. We defined two-sided *P*-values of less than 0.05 as statistically significant.

## Results

Baseline characteristics for the controls and cases are shown in Table 1. $EGFR^{mut}$ cases constituted approximately 35% among cases. As surgical cases tended to be checked for *EGFR* status, adenocarcinoma with histology was dominant. Smoking status significantly differed between the controls and the $EGFR^{wt}$ cases. Current smokers and heavy smokers (PY more than 40) were more common than in controls in $EGFR^{wt}$ but not $EGFR^{mut}$ cases. Alcohol drinking did not show any association, independent of the *EGFR* mutation status. Adenocarcinomas predominated among $EGFR^{mut}$ cases and even in the $EGFR^{wt}$ cases accounted for approximately two-thirds.

Table 2 shows the impact of food groups on $EGFR^{mut}$ and $EGFR^{wt}$ NSCLCs. Consumption of soybean products was associated with a statistically significant decrease in the risk of $EGFR^{mut}$ NSCLCs. ORs for the second and third tertiles in comparison with the lowest tertile were thus 0.79 (0.50–1.27) and 0.56 (0.34–0.93), the trend being statistically significant ($P = 0.023$). Consumption of other food products including meat, fish and seafood showed no significant influence on either type of NSCLC. Intake of vegetables exhibited a significant protective association only for $EGFR^{wt}$ NSCLCs and similarly protective trend was observed with fruits and vegetables other than green-yellow vegetables.

The results of stratified analyzes by sex, age, smoking and histology are shown in Table 3. Soybean product intake conferred a significantly decreased risk for $EGFR^{mut}$ but not $EGFR^{wt}$ NSCLCs in males. A similar trend was observed in females treated separately, although without significance. Regarding age, protective association of soy consumption for $EGFR^{mut}$ NSCLCs were clearly seen in the younger age groups (<55 and 55–64 years), while being less clear at older ages. At all levels of smoking, soy consumption showed risk reduction for $EGFR^{mut}$ NSCLCs, though not significant. With $EGFR^{wt}$ NSCLCs, the middle level exposure group (1–40 PYs) showed a decreased risk; increased risk was observed for never smokers; and no change was observed for the high level exposure group. Taking into account the histopathology, protective association were clear for $EGFR^{mut}$ but not $EGFR^{wt}$ adenocarcinomas. Due to the small number of cases $EGFR^{mut}$ nonadenocarcinomas were difficult to evaluate. Stratified analysis for menopausal status among women showed a protective association for $EGFR^{wt}$ NSCLCs among premenopausal women, although very small number of cases might limit interpretability. $EGFR^{wt}$ NSCLCs did not show any significant protective association of soy consumption.

We further examined the single food items contributing to total soy consumption, miso soup, natto (fermented soy) and tofu as shown in Table 4. Miso soup consumption conferred a statistically significantly decreased risk of $EGFR^{mut}$ NSCLC overall and in males. A non-significant similar trend was observed for females. High consumption of natto also showed protective association for $EGFR^{mut}$ but not for $EGFR^{wt}$ tumors. Associations between risk and tofu consumption were not obvious overall as well as in each sex.

## Discussion

In this study of the impact of food consumption on the risk NSCLCs divided by *EGFR* mutation status, soy consumption was found to be significantly associated with a lowered risk of $EGFR^{mut}$ but not $EGFR^{wt}$ NSCLCs. To our knowledge, this is the first indication that soy consumption may influence molecular events contributing to the subset of NSCLCs featuring *EGFR* mutations, as the hormonal receptor status of breast cancers is known to be affected by reproductive factors.[25]

Soy products contain isoflavones, phytestrogens such as genistein, daizein and glycitein[12] and there is strong epidemiologic evidence of protective effects of soy consumption for breast cancer.[11–14] In the present study, a relatively higher consumption of soybean products decreased the risk of NSCLCs with *EGFR* mutations only. In addition, we and others have previously

**Table 1. Characteristics of controls and NSCLC cases seperated by *EGFR* mutation status**

| | Male | | | | | Female | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Control | Case | | | | Control | Case | | | |
| | | $EGFR^{mut}$ | Comparison with controls (*P*-values) | $EGFR^{wt}$ | Comparison with controls (*P*-values) | Comparison by mutation status (*P*-values) | | $EGFR^{mut}$ | Comparison with controls (*P*-values) | $EGFR^{wt}$ | Comparison with controls (*P*-values) | Comparison by mutation status (*P*-values)* |
| *n* | 1080 | 46 | | 170 | | | 685 | 76 | | 61 | | |
| Age (SD)† | 62.6 ± 9.4 | 62.8 ± 9.7 | | 62.5 ± 9.4 | | | 60.7 ± 11.0 | 63.5 ± 8.8 | | 57.2 ± 12.6 | | |
| Smoking | | | | | | | | | | | | |
| Never smoker | 254 (23.5%) | 14 (30.4%) | 0.702 | 7 (4.1%) | <0.001 | <0.001 | 594 (86.7%) | 69 (90.8%) | 0.24 | 43 (70.5%) | 0.003 | 0.004 |
| Former smoker | 415 (38.4%) | 15 (32.6%) | | 57 (33.5%) | | | 26 (3.8%) | 4 (5.3%) | | 5 (8.2%) | | |
| Current smoker | 408 (37.8%) | 17 (37.0%) | | 106 (62.4%) | | | 65 (9.5%) | 3 (4.0%) | | 13 (21.3%) | | |
| Unknown | 3 (0.3%) | 0 (0.0%) | | 0 (0.0%) | | | 0 (0.0%) | 0 (0.0%) | | 0 (0.0%) | | |
| 0 pack-years | 255 (23.6%) | 14 (30.4%) | 0.563 | 7 (4.1%) | <0.001 | <0.001 | 594 (86.7%) | 69 (90.8%) | 0.018 | 43 (70.5%) | <0.001 | 0.009 |
| 1–40 pack-years | 448 (41.5%) | 18 (39.1%) | | 43 (25.3%) | | | 86 (12.6%) | 4 (5.3%) | | 13 (21.3%) | | |
| >40 pack-years | 352 (32.6%) | 14 (30.4%) | | 115 (67.7%) | | | 3 (0.4%) | 2 (2.6%) | | 5 (8.2%) | | |
| Unknown | 25 (2.3%) | 0 (0.0%) | | 5 (2.9%) | | | 2 (0.3%) | 1 (1.3%) | | 0 (0.0%) | | |
| Drinking | | | | | | | | | | | | |
| Never drinker | 300 (27.8%) | 12 (26.1%) | 0.611 | 48 (28.2%) | 0.091 | 0.939 | 473 (69.1%) | 50 (65.8%) | 0.239 | 44 (72.1%) | 0.837 | 0.342 |
| Former drinker | 135 (12.5%) | 3 (6.5%) | | 12 (7.1%) | | | 24 (3.5%) | 0 (0.0%) | | 1 (1.6%) | | |
| Current drinker | 644 (59.6%) | 31 (67.4%) | | 109 (64.1%) | | | 186 (27.2%) | 26 (34.2%) | | 16 (26.2%) | | |
| Unknown | 1 (0.1%) | 0 (0.0%) | | 1 (0.6%) | | | 2 (0.3%) | 0 (0.0%) | | 0 (0.0%) | | |
| Histopathology | | | | | | | | | | | | |
| Adenocarcinoma | | 43 (93.5%) | | 110 (64.7%) | | 0.002 | | 75 (98.7%) | | 52 (85.3%) | | 0.014 |
| Squamous cell carcinoma | | 2 (4.4%) | | 40 (23.5%) | | | | 0 (0.0%) | | 4 (6.6%) | | |
| Large cell carcinoma | | 1 (2.2%) | | 13 (7.7%) | | | | 0 (0.0%) | | 4 (6.6%) | | |
| Others | | 0 (0.0%) | | 7 (4.1%) | | | | 1 (1.3%) | | 1 (1.6%) | | |

**P*-values were calculated for homogeneity across three groups.
†Mann–Whitney ranksum test was applied.
EGFR, epidermal growth factor receptor; SD, standard deviation.

doi: 10.1111/j.1349-7006.2008.00812.x

**Table 2. Impact of selected food group intake on NSCLC with or without EGFR mutation**

| Item (median, range) | Total | | | | |
|---|---|---|---|---|---|
| | Control[†] | Case | | | |
| | | $EGFR^{mut}$ | | $EGFR^{wt}$ | |
| | | *N* | OR (95% CI)[‡] | *N* | OR (95% CI)[‡] |
| **Soybean products** | | | | | |
| T1 (19.8, 0.0–32 g/day) | 586 | 46 | 1.00 (Reference) | 84 | 1.00 (Reference) |
| T2 (44.1, 33–54.6 g/day) | 586 | 43 | 0.79 (0.50–1.27) | 87 | 1.19 (0.85–1.68) |
| T3 (81.8, 54.7–300 g/day) | 585 | 33 | 0.56 (0.34–0.93) | 60 | 0.89 (0.61–1.29) |
| | | | *P* for trend = 0.023 | | *P* for trend = 0.499 |
| **Meat** | | | | | |
| T1 (12.5 0.0–21.3 g/day) | 586 | 41 | 1.00 (Reference) | 82 | 1.00 (Reference) |
| T2 (26.6, 21.4–32.6 g/day) | 586 | 37 | 0.90 (0.55–1.48) | 84 | 1.11 (0.79–1.57) |
| T3 (54.1, 32.7–230 g/day) | 585 | 44 | 1.15 (0.71–1.86) | 65 | 0.79 (0.55–1.14) |
| | | | *P* for trend = 0.576 | | *P* for trend = 0.259 |
| **Fish** | | | | | |
| T1 (14.4 0.0–20.1 g/day) | 586 | 39 | 1.00 (Reference) | 85 | 1.00 (Reference) |
| T2 (35.4, 20.2–43.1 g/day) | 586 | 43 | 1.00 (0.62–1.63) | 83 | 1.07 (0.76–1.50) |
| T3 (64.9, 43.2–270 g/day) | 585 | 40 | 0.98 (0.59–1.62) | 63 | 0.80 (0.55–1.16) |
| | | | *P* for trend = 0.939 | | *P* for trend = 0.273 |
| **Other seafood** | | | | | |
| T1 (8.39, 0.0–12.0 g/day) | 586 | 41 | 1.00 (Reference) | 80 | 1.00 (Reference) |
| T2 (15.7, 12.0–18.7 g/day) | 586 | 35 | 0.87 (0.53–1.42) | 77 | 1.01 (0.71–1.44) |
| T3 (31.7, 18.8–150 g/day) | 585 | 46 | 1.19 (0.74–1.91) | 74 | 0.92 (0.65–1.32) |
| | | | *P* for trend = 0.480 | | *P* for trend = 0.668 |
| **Green-yellow vegetables** | | | | | |
| T1 (25.3, 0.0–39.8 g/day) | 586 | 37 | 1.00 (Reference) | 103 | 1.00 (Reference) |
| T2 (57.8, 39.9–76.8 g/day) | 586 | 45 | 1.06 (0.64–1.74) | 75 | 0.78 (0.56–1.10) |
| T3 (123.4, 76.9–500 g/day) | 585 | 40 | 0.76 (0.45–1.29) | 53 | 0.69 (0.47–1.00) |
| | | | *P* for trend = 0.282 | | *P* for trend = 0.044 |
| **Other vegetables** | | | | | |
| T1 (24.7, 0.0–41.6 g/day) | 586 | 42 | 1.00 (Reference) | 99 | 1.00 (Reference) |
| T2 (55.5, 41.7–70.2 g/day) | 586 | 36 | 0.81 (0.49–1.33) | 69 | 0.72 (0.51–1.02) |
| T3 (107.5, 70.3–410 g/day) | 585 | 44 | 0.84 (0.51–1.37) | 63 | 0.78 (0.53–1.11) |
| | | | *P* for trend = 0.525 | | *P* for trend = 0.158 |
| **Fruit** | | | | | |
| T1 (21.9, 0.0–33.9 g/day) | 586 | 35 | 1.00 (Reference) | 98 | 1.00 (Reference) |
| T2 (70.5, 34.0–84.9 g/day) | 586 | 37 | 0.93 (0.55–1.56) | 78 | 0.91 (0.65–1.27) |
| T3 (151.2, 85.0–510 g/day) | 585 | 50 | 1.10 (0.66–1.85) | 55 | 0.72 (0.49–1.06) |
| | | | *P* for trend = 0.602 | | *P* for trend = 0.096 |

[†]Eight controls were excluded from analysis due to incomplete SQFFQ data.
[‡]ORs were adjusted for age, sex, energy intake, and smoking in unconditional logistic model.
EGFR, epidermal growth factor receptor; CI, confidence interval; NSCLC, non-small-cell lung cancers; OR, odds ratio.

reported that soy product consumption is associated with a decreased risk of all NSCLCs.[26–29] These findings may provide insights into both the genesis and differentiation of NSCLCs. Soybean isoflavones are heterocyclic phenols closely related in structure to estrogen and their very weak estrogenic activity despite higher affinity for estrogen receptors than estrogens themselves endows them with antiestrogenic effects.[30] Several epidemiologic studies have shown associations between reproductive factors and NSCLC risk, although the direction of the impact was not consistent.[31–40] Moreover, we previously reported elevated risk with a longer fertile life supporting a positive association between estrogenic exposure in NSCLCs and *EGFR* mutation.[10] Taken the data together, one may hypothesize that hormones play an important role in *EGFR* mutagenesis in the lung or outcome in lesions bearing *EGFR* mutations.

The weaker association in females is one of the limitation of this interpretation, suggesting other possibility for soy's impact on *EGFR* mutation status. Soy products may exert various non-estrogenic effects, reviewed in detail by Messina *et al*.[30] The inhibitory effect of genistein on tyrosine kinase[41,42] might be relevant to the question of the putative involvement of soy products in the EGFR-mediated phosphorylation cascade in $EGFR^{mut}$ NSCLCs. Further, Peterson *et al*. reported that genistein inhibits EGFR-mediated signal transduction distal to EGFR tyrosine autophosphorylation.[43] The epidemiologic association as well as biological significance of soy products thus requires further evaluation.

In our study, consumption of fruits and vegetables showed protective association only for $EGFR^{wt}$ NSCLC. Fruit and vegetable consumption is regarded as convincing protective factors for smoking related cancers occurring, including lung, and possible mechanisms behind this association include antioxidative effect by micronutrients such as beta-carotene, vitamins C and E.[44] Our current finding are accordant with our previous finding that $EGFR^{wt}$ NSCLC were smoking related NSCLC.[10]

Several methodological points with this study should be addressed. First, most remarkable advantage of this study is an availability of *EGFR* mutational status for hundreds of cases along with lifestyle data. This situation is rarely available with population-based study setting. Next, all lifestyle information

**Table 3. Impact of soybean product intake on NSCLC with or without EGFR mutation**

| | Control† | | | Cases | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | EGFR$^{mut}$ | | | | EGFR$^{wt}$ | | | |
| | Soy consumption (*n*) | | | Soy consumption [*n*, OR‡ (95%CI)] | | | | Soy consumption [*n*, OR‡ (95% CI)] | | | |
| | T1 | T2 | T3 | T1 | T2 | T3 | Trend-*P* | T1 | T2 | T3 | Trend-*P* |
| **Sex** | | | | | | | | | | | |
| Male | 379 | 344 | 353 | 20 | 17 | 9 | | 64 | 67 | 39 | |
| | | | | 1.00 (Reference) | 0.94 (0.45–1.96) | 0.39 (0.16–0.94) | 0.035 | 1.00 (Reference) | 1.21 (0.81–1.79) | 0.74 (0.47–1.16) | 0.243 |
| Female | 207 | 242 | 232 | 26 | 26 | 24 | | 20 | 20 | 21 | |
| | | | | 1.00 (Reference) | 0.73 (0.39–1.35) | 0.67 (0.35–1.25) | 0.238 | 1.00 (Reference) | 1.18 (0.59–2.37) | 1.39 (0.68–2.83) | 0.365 |
| **Age** | | | | | | | | | | | |
| <55 | 175 | 124 | 108 | 12 | 8 | 5 | | 32 | 17 | 8 | |
| | | | | 1.00 (Reference) | 0.66 (0.24–1.85) | 0.49 (015–1.59) | 0.221 | 1.00 (Reference) | 0.91 (0.47–1.76) | 0.56 (0.24–1.34) | 0.227 |
| 55–64 | 174 | 206 | 182 | 19 | 12 | 7 | | 23 | 29 | 23 | |
| | | | | 1.00 (Reference) | 0.47 (0.21–1.05) | 0.36 (0.14–0.95) | 0.026 | 1.00 (Reference) | 1.16 (0.62–2.15) | 0.99 (0.51–1.92) | 0.979 |
| 65–79 | 237 | 256 | 295 | 15 | 23 | 21 | | 29 | 41 | 29 | |
| | | | | 1.00 (Reference) | 1.38 (0.66–2.87) | 0.85 (0.41–1.79) | 0.547 | 1.00 (Reference) | 1.60 (0.92–2.78) | 1.06 (0.59–1.90) | 0.861 |
| **Smoking§** | | | | | | | | | | | |
| 0 pack-years | 256 | 281 | 306 | 30 | 27 | 26 | | 13 | 16 | 21 | |
| | | | | 1.00 (Reference) | 0.69 (0.38–1.25) | 0.59 (0.32–1.07) | 0.094 | 1.00 (Reference) | 1.40 (0.63–3.14) | 2.11 (0.96–4.63) | 0.058 |
| 1–40 pack-years | 191 | 181 | 161 | 8 | 10 | 4 | | 27 | 19 | 10 | |
| | | | | 1.00 (Reference) | 1.49 (0.50–4.39) | 0.55 (0.15–2.04) | 0.358 | 1.00 (Reference) | 0.80 (0.42–1.52) | 0.47 (0.21–1.03) | 0.059 |
| >40 pack-years | 131 | 116 | 108 | 8 | 5 | 3 | | 43 | 49 | 28 | |
| | | | | 1.00 (Reference) | 0.70 (0.19–2.63) | 0.42 (0.09–2.07) | 0.282 | 1.00 (Reference) | 1.33 (0.80–2.21) | 0.83 (0.47–1.48) | 0.515 |
| **Histopathology¶** | | | | | | | | | | | |
| Adenocarcinoma | 470 | 464 | 460 | 44 | 41 | 33 | | 54 | 64 | 44 | |
| | | | | 1.00 (Reference) | 0.79 (0.49–1.28) | 0.59 (0.35–0.98) | 0.042 | 1.00 (Reference) | 1.36 (0.92–2.03) | 1.01 (0.65–1.57) | 0.957 |
| non-Adenocarcinoma | 116 | 122 | 125 | 2 | 2 | 0 | | 30 | 23 | 16 | |
| | | | | 1.00 (Reference) | 0.90 (0.12–6.55) | NE†† | 0.187 | 1.00 (Reference) | 0.78 (0.41–1.48) | 0.58 (0.29–1.16) | 0.124 |
| **Menopausal status‡‡** | | | | | | | | | | | |
| Premenopausal | 50 | 40 | 29 | 6 | 2 | 1 | | 7 | 5 | 2 | |
| | | | | 1.00 (Reference) | 0.20 (0.03–1.26) | 0.11 (0.01–1.12) | 0.037 | 1.00 (Reference) | 1.15 (0.32–4.10) | 0.77 (0.13–4.38) | 0.858 |
| Postmenopausal | 156 | 201 | 202 | 20 | 17 | 9 | | 64 | 68 | 39 | |
| | | | | 1.00 (Reference) | 0.90 (0.46–1.75) | 0.84 (0.43–1.66) | 0.623 | 1.00 (Reference) | 1.18 (0.50–2.81) | 1.60 (0.70–3.67) | 0.253 |

†Eight controls were excluded from analyzes due to incomplete SQFFQ status.
‡ORs were adjusted for age, sex, energy intake and smoking in unconditional logistic model.
§Twenty-seven controls and six cases were excluded because of unknown status.
¶Controls matched with adenocarcinoma cases were used for adenocarcinoma analysis. Same condition is applied to nonadenocarcinoma analysis.
††NE indicate not estimate due to lack of case in this category.
‡‡Analysis was limited to female with known menopausal status.
EGFR, epidermal growth factor receptor; CI, confidence interval; NSCLC, non-small-cell lung cancers; OR, odds ratio; SQFFQ, semiquantitative food-frequency-questionnaire.

was based on self-reporting, there could be the possibility of information bias; however, regarding bias in soy consumption, it is unlikely having disease change subjects' response because effect of soy on NSCLC have never been accepted in public. Moreover, as the data were collected before diagnosis and patients with NSCLCs had no information regarding their *EGFR* mutation status, bias due to having the disease seems again unlikely. The hospital-based design could be another limitation; however, similarity in lifestyle between our controls and population[19] might warrant appropriateness of our study. Finally, our study was limited to ethnic Japanese and the results therefore cannot necessarily be extrapolated to other population groups. Considering the higher consumption of soy products in Japan than in Western countries, and their potential protective effects against *EGFR*-mutated NSCLCs, evaluations in other population is important. Also, there might be unknown genetic difference in regard to risk of *EGFR* mutation, further evaluation requires taking genetic background in the population into consideration.

In conclusion, the present findings provide evidence that soy product consumption reduces the risk of *EGFR*-mutated NSCLCs. Further investigation of these findings in other ethnic groups is warranted.

## Acknowledgments

This study was supported by a Grant-in Aid for Scientific Research from the Ministry of Education, Science, Sports, Culture and Technology of Japan, a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare, and by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan.
Authors are grateful for the technical support by the staff of Division of Epidemiology and Prevention, Nagoya, Japan.

## References

1 Hunter T, Cooper JA. Epidermal growth factor induces rapid tyrosine phosphorylation of proteins in A431 human tumor cells. *Cell* 1981; **24**: 741–52.
2 Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. *Eur J Cancer* 2001; **37** (Suppl. 4): S9–15.
3 Arteaga CL. Overview of epidermal growth factor receptor biology and its role as a therapeutic target in human neoplasia. *Semin Oncol* 2002; **29** (Suppl. 14): 3–9.


doi: 10.1111/j.1349-7006.2008.00812.x

**Table 4. Impact of soy and soy-containing food items on NSCLC with or without EGFR mutation**

| Item | Total | | | | | Male | | | | | Female | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Control | Case | | | | Control | Case | | | | Control | Case | | | |
| | | $EGFR^{mut}$ | $OR^{\dagger}$ (95% CI) | $EGFR^{wt}$ | $OR^{\dagger}$ (95% CI) | | $EGFR^{mut}$ | $OR^{\dagger}$ (95% CI) | $EGFR^{wt}$ | $OR^{\dagger}$ (95% CI) | | $EGFR^{mut}$ | $OR^{\dagger}$ (95% CI) | $EGFR^{wt}$ | $OR^{\dagger}$ (95% CI) |
| **Miso soup** | | | | | | | | | | | | | | | |
| ≤3–4 times/week | 758 | 62 | 1.00 (Reference) | 99 | 1.00 (Reference) | 433 | 25 | 1.00 (Reference) | 69 | 1.00 (Reference) | 325 | 37 | 1.00 (Reference) | 30 | 1.00 (Reference) |
| ≤once a day | 811 | 56 | 0.69 (0.46–1.03) | 103 | 1.08 (0.79–1.48) | 512 | 20 | 0.46 (0.24–0.91) | 81 | 1.12 (0.78–1.61) | 299 | 36 | 0.87 (0.52–1.45) | 22 | 1.00 (0.54–1.82) |
| twice a day | 152 | 3 | 0.23 (0.07–0.77) | 22 | 1.17 (0.70–1.96) | 110 | 1 | 0.12 (0.02–0.92) | 15 | 0.90 (0.49–1.68) | 42 | 2 | 0.35 (0.08–1.57) | 7 | 2.33 (0.89–6.15) |
| Unknown | 44 | 1 | | 7 | | 25 | 0 | | 5 | | 19 | 1 | | 2 | |
| | | | *P* for trend = 0.005 | | *P* for trend = 0.499 | | | *P* for trend = 0.003 | | *P* for trend = 0.946 | | | *P* for trend = 0.216 | | *P* for trend = 0.252 |
| **Natto** | | | | | | | | | | | | | | | |
| ≤1–3 times/month | 550 | 42 | 1.00 (Reference) | 91 | 1.00 (Reference) | 387 | 21 | 1.00 (Reference) | 70 | 1.00 (Reference) | 163 | 21 | 1.00 (Reference) | 21 | 1.00 (Reference) |
| ≥3–4 times/week | 864 | 57 | 0.63 (0.40–0.99) | 98 | 0.85 (0.61–1.16) | 490 | 17 | 0.57 (0.28–1.17) | 72 | 0.89 (0.61–1.29) | 374 | 40 | 0.69 (0.38–1.22) | 26 | 0.71 (0.38–1.35) |
| ≥once a day | 301 | 21 | 0.59 (0.33–1.06) | 32 | 0.91 (0.58–1.43) | 174 | 8 | 0.56 (0.22–1.43) | 20 | 0.84 (0.48–1.47) | 127 | 13 | 0.62 (0.29–1.34) | 12 | 1.01 (0.45–2.29) |
| Unknown | 50 | 2 | | 10 | | 29 | 0 | | 8 | | 21 | 2 | | 2 | |
| | | | *P* for trend = 0.052 | | *P* for trend = 0.482 | | | *P* for trend = 0.150 | | *P* for trend = 0.467 | | | *P* for trend = 0.199 | | *P* for trend = 0.839 |
| **Tofu** | | | | | | | | | | | | | | | |
| ≤1–3 times/month | 570 | 35 | 1.00 (Reference) | 73 | 1.00 (Reference) | 387 | 22 | 1.00 (Reference) | 56 | 1.00 (Reference) | 179 | 13 | 1.00 (Reference) | 17 | 1.00 (Reference) |
| 1–2 times/week | 617 | 50 | 1.07 (0.66–1.74) | 88 | 1.30 (0.92–1.84) | 490 | 13 | 0.54 (0.25–1.15) | 70 | 1.33 (0.90–1.97) | 238 | 37 | 1.85 (0.93–3.69) | 18 | 1.06 (0.51–2.23) |
| ≥3–4 times/week | 509 | 34 | 0.80 (0.48–1.36) | 62 | 1.14 (0.78–1.66) | 174 | 10 | 0.61 (0.27–1.39) | 37 | 0.97 (0.61–1.53) | 239 | 24 | 1.15 (0.55–2.38) | 25 | 1.46 (0.73–2.94) |
| Unknown | 69 | 3 | | 8 | | 29 | 1 | | 7 | | 29 | 2 | | 1 | |
| | | | *P* for trend = 0.398 | | *P* for trend = 0.450 | | | *P* for trend = 0.170 | | *P* for trend = 0.946 | | | *P* for trend = 0.988 | | *P* for trend = 0.269 |

$^{\dagger}$ORs were adjusted for age, sex and smoking in unconditional logistic model. Unknown subjects were excluded from trend analysis.
EGFR, epidermal growth factor receptor; CI, confidence interval; NSCLC, non-small-cell lung cancers; OR, odds ratio.

4 Bunn PA Jr, Franklin W. Epidermal growth factor receptor expression, signal pathway, and inhibitors in non-small cell lung cancer. *Semin Oncol* 2002; **29** (Suppl. 14): 38–44.
5 Paez JG, Janne PA, Lee JC *et al*. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; **304**: 1497–500.
6 Lynch TJ, Bell DW, Sordella R *et al*. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; **350**: 2129–39.
7 Pao W, Miller V, Zakowski M *et al*. EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 2004; **101**: 13306–11.
8 Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res* 2004; **64**: 8919–23.
9 Shigematsu H, Lin L, Takahashi T *et al*. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 2005; **97**: 339–46.
10 Matsuo K, Ito H, Yatabe Y *et al*. Risk factors differ for non-small-cell lung cancers with and without EGFR mutation: assessment of smoking and sex by a case-control study in Japanese. *Cancer Sci* 2007; **98**: 96–101.
11 Adlercreutz H, Mazur W. Phyto-oestrogens and Western diseases. *Ann Med* 1997; **29**: 95–120.
12 Adlercreutz H. Phyto-oestrogens and cancer. *Lancet Oncol* 2002; **3**: 364–73.
13 Bingham SA, Atkinson C, Liggins J, Bluck L, Coward A. Phyto-oestrogens: where are we now? *Br J Nutr* 1998; **79**: 393–406.
14 Sarkar FH, Li Y. The role of isoflavones in cancer chemoprevention. *Front Biosci* 2004; **9**: 2714–24.
15 Tajima K, Hirose K, Inoue M, Takezaki T, Hamajima N, Kuroishi T. A Model of practical cancer prevention for out-patients visiting a hospital: the hospital-based epidemiologic research program at aichi cancer center (HERPACC). *Asian Pac J Cancer Prev* 2000; **1**: 35–47.
16 Hamajima N, Matsuo K, Saito T *et al*. Gene–environment interactions and polymorphism studies of cancer risk in the hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac J Cancer Prev* 2001; **2**: 99–107.
17 Inoue M, Tajima K, Takezaki T *et al*. Epidemiology of pancreatic cancer in Japan: a nested case-control study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC). *Int J Epidemiol* 2003; **32**: 257–62.
18 Inoue M, Tajima K, Hirose K *et al*. Subsite-specific risk factors for colorectal cancer: a hospital-based case-control study in Japan. *Cancer Causes Control* 1995; **6**: 14–22.
19 Inoue M, Tajima K, Hirose K *et al*. Epidemiological features of first-visit outpatients in Japan: comparison with general population and variation by sex, age, and season. *J Clin Epidemiol* 1997; **50**: 69–77.
20 Yatabe Y, Kosaka T, Takahashi T, Mitsudomi T. EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. *Am J Surg Pathol* 2005; **29**: 633–9.
21 Yatabe Y, Hida T, Horio Y, Kosaka T, Takahashi T, Mitsudomi T. A rapid, sensitive assay to detect EGFR mutation in small biopsy specimens from lung cancer. *J Mol Diagn* 2006; **24**: 1679–88.
22 Tokudome S, Goto C, Imaeda N, Tokudome Y, Ikeda M, Maki S. Development of a data-based short food frequency questionnaire for assessing nutrient intake by middle-aged Japanese. *Asian Pac J Cancer Prev* 2004; **5**: 40–3.
23 Tokudome S, Ikeda M, Tokudome Y, Imaeda N, Kitagawa I, Fujiwara N. Development of data-based semi-quantitative food frequency questionnaire for dietary studies in middle-aged Japanese. *Jpn J Clin Oncol* 1998; **28**: 679–87.
24 Tokudome Y, Goto C, Imaeda N *et al*. Relative validity of a short food frequency questionnaire for assessing nutrient intake versus three-day weighed diet records in middle-aged Japanese. *J Epidemiol* 2005; **15**: 135–45.
25 Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiol Biomarkers Prev* 2004; **13**: 1558–68.
26 Takezaki T, Hirose K, Inoue M *et al*. Dietary factors and lung cancer risk in Japanese: with special reference to fish consumption and adenocarcinomas. *Br J Cancer* 2001; **84**: 1199–206.
27 Koo LC. Dietary habits and lung cancer risk among Chinese females in Hong Kong who never smoked. *Nutr Cancer* 1988; **11**: 155–72.
28 Swanson CA, Mao BL, Li JY *et al*. Dietary determinants of lung-cancer risk: results from a case-control study in Yunnan Province, China. *Int J Cancer* 1992; **50**: 876–80.
29 Wakai K, Ohno Y, Genka K *et al*. Risk modification in lung cancer by a dietary intake of preserved foods and soyfoods: findings from a case-control study in Okinawa, Japan. *Lung Cancer* 1999; **25**: 147–59.
30 Messina MJ, Persky V, Setchell KD, Barnes S. Soy intake and cancer risk: a review of the in vitro and in vivo data. *Nutr Cancer* 1994; **21**: 113–31.
31 Gao YT, Blot WJ, Zheng W *et al*. Lung cancer among Chinese women. *Int J Cancer* 1987; **40**: 604–9.
32 Wu AH, Yu MC Thomas DC, Pike MC, Henderson BE. Personal and family history of lung disease as risk factors for adenocarcinoma of the lung. *Cancer Res* 1988; **48**: 7279–84.
33 Taioli E, Wynder EL. Re: Endocrine factors and adenocarcinoma of the lung in women. *J Natl Cancer Inst* 1994; **86**: 869–70.
34 Zhou BS, Wang TJ, Guan P, Wu JM. Indoor air pollution and pulmonary adenocarcinoma among females: a case-control study in Shenyang, China. *Oncol Rep* 2000; **7**: 1253–9.
35 Liao ML, Wang JH, Wang HM, Ou AQ, Wang XJ, You WQ. A study of the association between squamous cell carcinoma and adenocarcinoma in the lung, and history of menstruation in Shanghai women, China. *Lung Cancer* 1996; **14** (Suppl. 1): S215–21.
36 Seow A, Poh WT, Teh M *et al*. Diet, reproductive factors and lung cancer risk among Chinese women in Singapore: evidence for a protective effect of soy in nonsmokers. *Int J Cancer* 2002; **97**: 365–71.
37 Kubik AK, Zatloukal P, Tomasek L, Petruzelka L. Lung cancer risk among Czech women: a case-control study. *Prev Med* 2002; **34**: 436–44.
38 Kreuzer M, Gerken M, Heinrich J, Kreienbrock L, Wichmann HE. Hormonal factors and risk of lung cancer among women? *Int J Epidemiol* 2003; **32**: 263–71.
39 Brenner AV, Wang Z, Kleinerman RA *et al*. Menstrual and reproductive factors and risk of lung cancer among Chinese women, Eastern Gansu Province, 1994–98. *J Epidemiol* 2003; **13**: 22–8.
40 Liu Y, Inoue M, Sobue T, Tsugane S. Reproductive factors, hormone use and the risk of lung cancer among middle-aged never-smoking Japanese women: a large-scale population-based cohort study. *Int J Cancer* 2005; **117**: 662–6.
41 Cunningham BD, Threadgill MD, Groundwater PW, Dale IL, Hickman JA. Synthesis and biological evaluation of a series of flavones designed as inhibitors of protein tyrosine kinases. *Anticancer Drug Des* 1992; **7**: 365–84.
42 Kuriu A, Ikeda H, Kanakura Y *et al*. Proliferation of human myeloid leukemia cell line associated with the tyrosine-phosphorylation and activation of the proto-oncogene c-kit product. *Blood* 1991; **78**: 283d–40.
43 Peterson G, Barnes S. Genistein and biochanin A inhibit the growth of human prostate cancer cells but not epidermal growth factor receptor tyrosine autophosphorylation. *Prostate* 1993; **22**: 335–45.
44 No author listed. Vegetables and fruits. In: Cannon G, ed. *Food, Nutrition and the Prevention of Cancer: a Global Perspective: World Cancer Research Fund/American Institute for Cancer Research* 1997. American Institute for Cancer Research: Washington, DC; 436–46.


doi: 10.1111/j.1349-7006.2008.00812.x