Table 2 Patient background and clinicopathological data

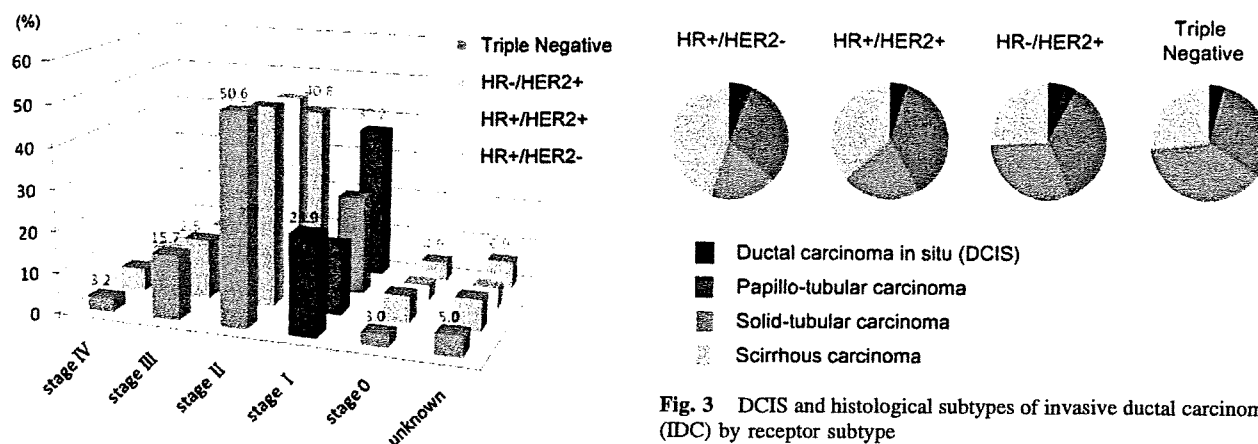|	Receptor subtype			
	HR+/HER2−	HR+/HER2+	HER2	TN
Number of patients (%)	8,039 (68.7)	892 (7.6)	977 (8.3)	1,797 (15.4)
Age median (range)	56 (NR–100)	54 (23–93)	56 (22–95)	57.5 (NR–94)
Ratio of bilateral breast cancer (%)	6.6[b]	5.9[a]	4.8[a]	6.2[b]
Incidence of breast cancer family history (%)	8.6	8.4	24.1	28.1
Ratio of premenopausal patients (%)	37.1[c]	38.8[c]	24.1[d]	28.1[d]
Height (cm) mean ± SD	154.3 ± 6.3	154.9 ± 6.1	154.0 ± 6.2	153.8 ± 6.3
Weight (kg) mean ± SD	54.7 ± 9.0	54.5 ± 8.8	53.9 ± 8.6	54.2 ± 9.0
BMI mean ± SD	23.0 ± 3.7	22.7 ± 3.5	22.7 ± 3.3	22.9 ± 3.5
Tumor size (cm) mean ± SD	2.6 ± 2.1[e]	3.2 ± 2.2	3.5 ± 2.6[f]	3.4 ± 2.7
Incidence of positive lymph node involvement (%)	20.6	34.9	38.5	32.2
Incidence of distant metastasis (%)	2.5	5.6	5.6	3.2
Incidence of breast-conserving surgery (%)	53.9	41.3	35.1	45.0

NR no record

[a] HER2 positive versus [b]HER2 negative according to ratio of bilateral breast cancer; $P = 0.040$, Fisher's exact probability test

[c] Hormone receptor-positive group versus [d]hormone receptor-negative group according to the ratio of premenopausal patients; $P < 0.0001$, Fisher's exact probability test

[e] HR+/HER2− subtype versus [f]TN subtype according to tumor size; $P < 0.00001$, standard t test of mean and standard deviation (SD)

Fig. 2 Stage at diagnosis by receptor subtype

Fig. 3 DCIS and histological subtypes of invasive ductal carcinoma (IDC) by receptor subtype
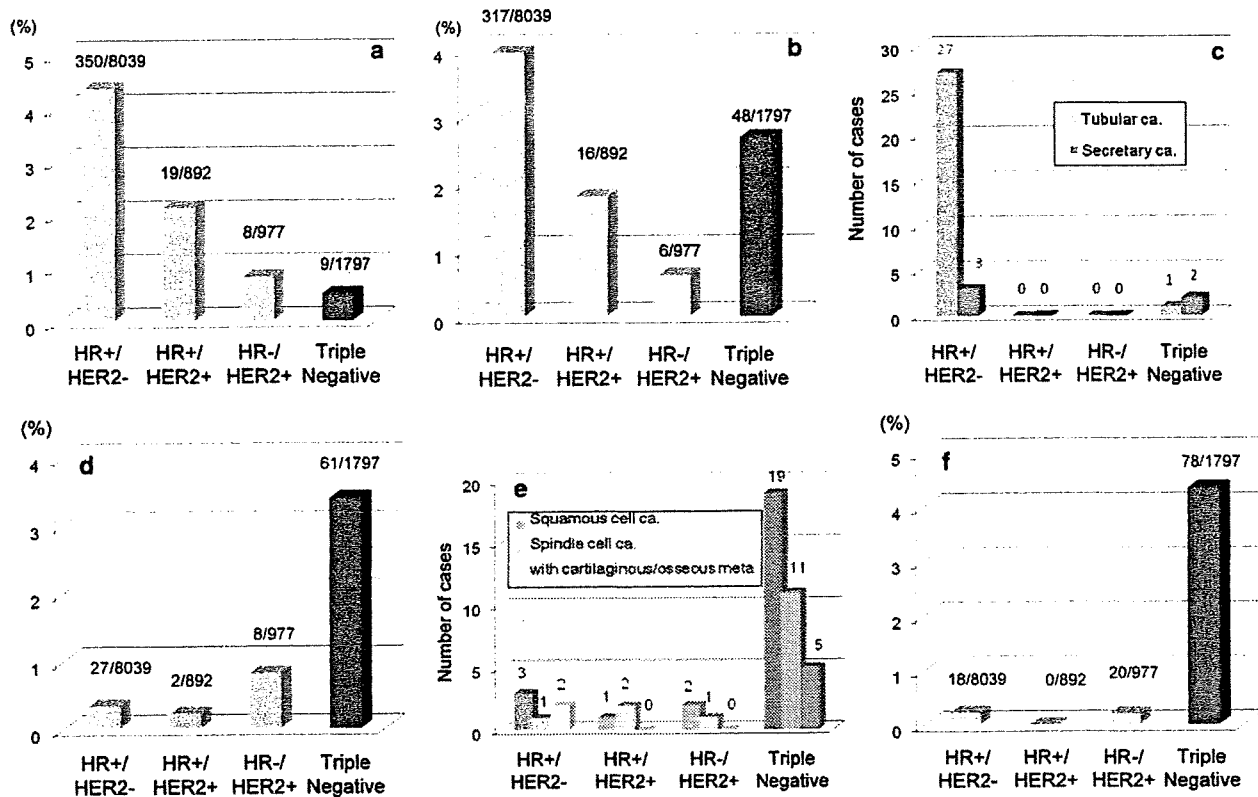
Pathological findings

From a viewpoint of morphologic classification, whereas scirrhous carcinoma was most frequently found in ER+/HER2− subtype, solid-tubular carcinoma prevailed in TN breast cancer (Fig. 3). As to breast cancers of special types, mucinous carcinoma occurred rarely in the TN group, but was quite frequent among ER+/HER2− subtype patients (Fig. 4a). Invasive lobular carcinoma was found reasonably frequently in the HR+/HER2− type (Fig. 4b). Tubular and secretary carcinomas were mostly found in patients with a HR+/HER2− subtype (Fig. 4c).

Medullary carcinomas were observed frequently in the TN group (Fig. 4d). Squamous cell carcinoma, spindle cell carcinoma, or metaplastic carcinoma with bone/cartilage metaplasia was likewise very common in patients in the TN subtype (Fig. 4e). This group also included the highest percentages of apocrine carcinomas (Fig. 4f).

Selection of chemotherapeutic regimens

Two main chemotherapeutic regimens were administered: (1) anthracycline-containing regimens (ACR), which included: doxorubicin plus cyclophosphamide (AC), epirubicin plus C (EC), C plus A plus 5-fluorouracil (CAF) and CEF, and (2) taxane (paclitaxel or docetaxel)-containing

(%)

a

5 · 350/8039

4 ··

3 –

2 ·

19/892

1 ·

8/977

9/1797

0 ··

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

(%) 317/8039

b

4 ·

3 ·

16/892

48/1797

2 ·

6/977

1 ·

0

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

30 · 27

c

Number of cases

25 ·

20 ·

Tubular ca.

Secretary ca.

15 ·

10 ·

5 ·

3

0 0   0 0   1 2

0

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

(%)

4 · d

61/1797

3 ·

2 ·

8/977

1 · 27/8039 2/892

0 ·

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

e

19

20 ·

Number of cases

15 ·

Squamous cell ca.

Spindle cell ca.

with cartilaginous/osseous meta

11

10 ·

5 ·

3

1   1   0   1   0

5

0 ·

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

(%)

5 · f

78/1797

4 ·

3 ·

2 ·

1 · 18/8039 0/892 20/977

0 ·

HR+/ HER2– HR+/ HER2+ HR–/ HER2+ Triple Negative

Fig. 4 a Incidence of mucinous carcinoma by receptor subtype. b Incidence of invasive lobular carcinoma by receptor subtype. c Incidence of tubular or secretary carcinomas by receptor subtype. d Incidence of medullary carcinoma by receptor subtype. e Incidence of metastatic carcinoma by receptor subtype. f Incidence of apocrine carcinoma by receptor subtype

Table 3 Chemotherapy according to receptor subtype

	Subtype			
	HR+/HER2–	HR+/HER2+	HR–/HER2+	TN
Number of cases	8,039	892	977	1,797
Node positive cases (%)	20.6	34.9	38.5	32.2
Number of patients treated by chemotherapy	3,913	696	940	1,563
Incidence of patients treated by ACR (%)	28.3	45.0	53.7	49.5
Incidence of patients treated by taxanes (%)	16.9	29.9	36.7	29.2
Incidence of neoadjuvant chemotherapy (%)	25.0	33.9	27.2	25.0

ACR anthracycline-containing regimen

regimens. In one patient there was a possibility that these two main regimens might have been administered concomitantly. ACR was administered to 28.3% of HR+/HER2– subtype tumors and taxane-based regimens to 16.9%. However, the incidence of axillary lymph nodes metastases was 20.6% in this subtype, which was smaller compared to other subtypes. In the other three subtype groups, as shown in Table 3, patients invariably received ACR or taxane regimens. Although the incidence of neoadjuvant chemotherapy was almost identical in each subtype group, the HR+/HER2+ group tended to be treated with this type of chemotherapy a little more frequently (Table 3).

Discussion

Of 14,749 breast cancer cases in Japan in 2004, 11,705 (79.4%) were examined for their ER, PgR, and HER2 status. TN tumors, defined as negative for all three receptors, accounted for 15.5% (1,819 cases). This was the largest collection of data on the prevalence rate of TN

tumors in Japan and was gathered from a large patient sample. Although there were no restrictions placed on the various centers for the methods they used for determining the presence of the receptors, or the criteria employed for their definition, we assumed that most cases had been examined using immunohistochemistry, since this technique for the detection of the receptors was in widespread use in 2004. It is possible that the hormone receptor status of some cases in this current study were incorrectly determined, because the definition criteria had not been established at that time in 2004. Most Japanese institutions regarded 0 or 2 on the Allred score as negative; others used a cutoff value of 10% for the determination of ER/PgR. Currently, the criteria for HER2 positivity are: 3+ with the IHC method, or 2+ with the IHC method and positive with the FISH method. However, in 2004, a tumor was defined as positive for HER2 if the IHC method resulted in 3+ alone, or in 2+ 3+.

Hence, in the present analyses, there were no strict criteria in place for the determination of ER, PgR, or HER2, leaving each institution to apply its own criteria. Now, however, it is considered that standardized analytical methods and definition criteria have been adopted nationwide, so that future analyses will be more reliable. Nevertheless, despite this limitation, we consider that the results of this population study are clinically very significant because of the large number of cases (over 11,000) and participating institutions (over 350).

The basal-like subtype accounts for 15–20% of breast cancers, irrespective of the method of analysis or ethnic group [6]. However, premenopausal African–American patients have a significantly higher incidence of this subtype compared to other patients [7, 11]. It is well known that the pathological and biological characteristics of breast cancer are significantly worse in young African–American patients and that they show a clinically poor prognosis. Therefore, the high incidence of basal-like subtype in young African–American patients correlates with the high histological grade of the tumors and the poor prognosis of the disease in this specific patient subgroup. On the other hand, it has been reported that this basal-like subtype is comparatively rare in Japanese women, with an incidence of only 8% documented in a recent study of 793 breast cancer cases in Japan [8]. We were interested in a possible familial nature of TN tumors, because of suggestions of a link to BRCA1 mutation. In the present study, however, we could not find any evidence of a family history of breast cancer in this subtype or that it affected younger women than other subtypes. This may have been due to the fact that various subtypes of breast cancer are included in the definition of 'TN,' although the basal-like subtype is thought to comprise 40–80% of cases [7, 12].

On the other hand, HR+/HER2− subtype tumors tended to be smaller and the patients freer of lymph node metastases at the time of diagnosis, while the HER2 subtypes were often positive for regional or distant lymph node metastases. This finding indicates the HR+/HER2− subtype tends to be detected at an earlier stage than the HR−/HER2+ subtype, which was characteristically diagnosed at an advanced stage. However, even if these tumors were both detected at an early stage, the difference in outcome would not be affected since HER2-positive tumors progress more rapidly.

Regarding histological subtypes, scirrhous carcinoma and solid-tubular carcinoma tended to be found more frequently in the HR+/HER2− and the TN subtypes, respectively. Although invasive lobular carcinoma was also found in the TN type, its true incidence is unclear as it is rarely difficult to distinguish from scirrhous carcinoma. This TN subgroup also included many cases of medullary and metaplastic carcinomas. Spindle cell and squamous cell carcinomas of the TN tumors showed metaplasia derived from invasive ductal carcinoma and exhibited characteristics of basal-like tumors [12]. However, medullary and apocrine carcinomas, which were included in the metaplastic carcinomas, have a better prognosis than the common type and, among TN breast cancers, should be regarded as different from the more common basal-like breast cancer.

There was no apparent correlation between the choice of chemotherapeutic regimen and tumor receptor subtype, both an anthracycline-containing regimen (ACR) and taxanes were used depending on the degree of progression. Neoadjuvant therapy was used in 27.2 and 25% of HR−/HER2+ and TN tumors, respectively, indicating that in 2004 this therapy was being used in large resectable tumors.

In conclusion, we analyzed data from a large number of breast cancer cases registered by the Japanese Breast Cancer Society in order to characterize and advance our understanding of the TN subtype of breast cancer. The present study demonstrated that it was important to establish standard analytical methods and criteria for detection of ER, PgR, and HER2 and that in particular it was necessary to define TN breast cancer more carefully. In the future, we need to follow up the prognosis and response to chemotherapy in these TN breast cancer cases in an attempt to characterize the subtype in more detail [2]. TN breast cancer simulates basal-like tumor, which has been classified from gene profiles. The basal-like type of TN breast cancer is diagnosed by IHC methods based on the expression of EGFR and cytokeratin 5/6. We are looking into further analyzing the cases from the 2004 registry from the perspectives of immunohistochemistry, prognosis, and use of adjuvant chemotherapy.

References

1. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer. 2007;109:1721–8.
2. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13:2329–34.
3. Nishimura R, Arima N. Is triple negative a prognostic factor in breast cancer? Breast Cancer. 2008;15:303–8.
4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA. 2001;98:10869–74.
5. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA. 2003;100:8418–23.
6. Kobayashi S. Basal-like subtype of breast cancer: a review of its unique characteristics and their clinical significance. Breast Cancer. 2008;15:153–8.
7. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006;295:2492–502.
8. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. Breast. 2007;16(Suppl 2):S72–7.
9. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol. 1999;17:1474–81.
10. Umemura S, Kurosumi M, Moriya T, Oyama T, Arihiro K, Yamashita H, et al. Immunohistochemical Evaluation for hormone receptors in breast cancer: a practically useful evaluation system and handling protocol. Breast cancer. 2006;13:232–5.
11. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008;109:123–39.
12. Reis-Filho JS, Milanezi F, Steele D, Savage K, Simpson PT, Nesland JM, et al. Metaplastic breast carcinomas are basal-like tumours. Histopathology. 2006;49:10–21.

Association between frequent CpG island methylation and HER2 amplification in human breast cancers

Kotoe Terada[1,2], Eriko Okochi-Takada[1], Sadako Akashi-Tanaka[3], Kazuaki Miyamoto[4], Kiyomi Taniyama[4], Hitoshi Tsuda[5], Kiyoshi Asada[1], Michio Kaminishi[2] and Toshikazu Ushijima[1,*]

[1]Carcinogenesis Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, [2]Department of Surgical Metabolic Care and Endocrine Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan, [3]Department of Surgery, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, [4]Institute for Clinical Research, National Hospital Organization Kure Medical Center/Chugoku Cancer Center, Hiroshima 737-0023, Japan and [5]Pathology Laboratory, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

*To whom correspondence should be addressed. Tel: +81 3 3547 5240;
Fax: +81 3 5565 1753;
Email: tushijim@ncc.go.jp

The presence of frequent methylation of CpG islands (CGIs), designated as the CpG island methylator phenotype in some cancers, is associated with distinct clinicopathological characteristics, including gene amplification, in individual tumor types. Amplification of HER2 in human breast cancers is an important prognostic and therapeutic target, but an association between HER2 amplification and frequent CGI methylation is unknown. To clarify the association, we here quantified methylation levels of promoter CGIs of 11 genes, which are unlikely to confer growth advantage to cells, in 63 human breast cancers. The number of methylated genes in a cancer did not obey a bimodal distribution, and the 63 cancers were classified into those with frequent methylation ($n = 16$), moderate methylation ($n = 26$) and no methylation ($n = 21$). The incidence of HER2 amplification was significantly higher in the cancers with frequent methylation (11 of 16) than in those with no methylation (2 of 21, $P = 0.001$). Also, the number of methylated genes correlated with the degree of HER2 amplification ($r = 0.411$, $P = 0.002$). Correlation analysis with clinicopathological characteristics and methylation of CDKN2A, BRCA1 and CDH1 revealed that frequent methylation had significant correlation with higher nuclear grades ($P = 0.001$). These showed that frequent methylation had a strong association with HER2 amplification in breast cancers and suggested that frequent methylation can be a determinant of various characteristics in a fraction of human breast cancers.

Introduction

Aberrant DNA methylation is deeply involved in the development and progression of human cancers (1–4). Methylation of CpG islands (CGIs) in promoter regions is a major mechanism for inactivation of tumor suppressor genes. At the same time, maintenance of appropriate DNA methylation levels is known to be important for maintenance of genome integrity. DNA hypomethylation can lead to genomic instability and increased tumor incidence in mice (5,6) and is associated with loss of heterozygosity in human cancers (7,8). On the other hand, aberrant DNA methylation precedes loss of heterozygosity in human liver cancers (9).

The presence of frequent methylation of CGIs in a cancer was first described in colorectal cancers and designated as the CpG island methylator phenotype (CIMP) (10). Depending upon tumor tissue

types, the presence of frequent CGIs methylation, or CIMP, can be clearly observed and is associated with distinct clinicopathological features. For example, by careful selection of marker genes and their quantitative methylation analysis, CIMP in colorectal cancers was shown to be strongly associated with BRAF mutations (11). In neuroblastomas, both in Japan and Germany, CIMP was observed as a distinct entity associated with poor prognosis and MYCN amplification (12,13). Remarkably, all the cases with MYCN amplification had frequent methylation, with only one exception. Cases with CIMP but without MYCN amplification had a better prognosis than those that had both and a worse prognosis than those that had neither. This complete containment of tumors with MYCN amplification within CIMP-positive tumors suggested that CIMP could precede gene amplification or that at least the presence of frequent aberrant DNA methylation was associated with gene amplification.

Gene amplification of HER2, which is a member of the epidermal growth factor receptor family (14), is very important in human breast cancers. Initially, HER2 amplification was found to be present in 15–30% of newly diagnosed breast cancer cases and to be associated with increased metastatic potential and decreased overall survival (15). Suppression of HER2 activity was shown to have antitumor activity, and antibodies against HER2 were developed as a therapeutic agent against breast cancers. Now, it is well known that a humanized antibody against HER2, such as trastuzumab, is very effective against breast cancers with HER2 amplification (16,17). Nevertheless, inducers of HER2 amplification remain unknown.

In this study, we aimed to clarify whether or not the presence of frequent CGI methylation was associated with HER2 amplification in human breast cancers. For this end, from the genes silenced in human cancers (18,19), we selected genes whose silencing is unlikely to confer growth advantage and avoided selection bias of cells with methylation. Also, we performed quantitative methylation analysis of their putative nucleosome-depleted regions (20), which are most resistant to DNA methylation (21). Association between frequent CGI methylation and clinicopathological characteristics, including silencing of three tumor-suppressor genes (CDKN2A, BRCA1 and CDH1), was also analyzed.

Materials and methods

Patients and tissue samples

Sixty-three breast cancer tissue specimens were obtained from patients who underwent mastectomy or breast-conserving surgery (stage I 22 cases; stage II 26 cases; stage III 15 cases and stage IV 0 case). Informed consent was obtained from all the patients, and analysis was approved by the institutional review boards. Cancer tissues were frozen after resection and stored at −80°C until extraction of genomic DNA. High-molecular weight DNA was extracted by the phenol–chloroform method. Histological types were evaluated according to the criteria of the Japanese Breast Cancer Society (22).

Bisulfite modification and quantitative methylation-specific polymerase chain reaction

Completely methylated DNA and completely unmethylated DNA were prepared by methylating genomic DNA with SssI methylase (New England Biolabs, Beverly, MA) and amplifying genomic DNA with the GenomiPhi amplification system (GE Healthcare, Buckinghamshire, UK), respectively. Bisulfite modification was performed using 1 μg of BamHI-digested genomic DNA as described previously (23). The modified DNA was suspended in 40 μl of Tris-EDTA buffer, and an aliquot of 1 μl was used for polymerase chain reaction (PCR) with a primer set specific to methylated or unmethylated sequences (supplementary Table 1 is available at Carcinogenesis Online). Using the completely methylated DNA and completely unmethylated DNA, an annealing temperature specific for each primer set was determined. Real-time PCR was performed using SYBR® Green I (BioWhittaker Molecular Applications, Rockland, ME) and an iCycler Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). The number of DNA molecules with methylated sequences and
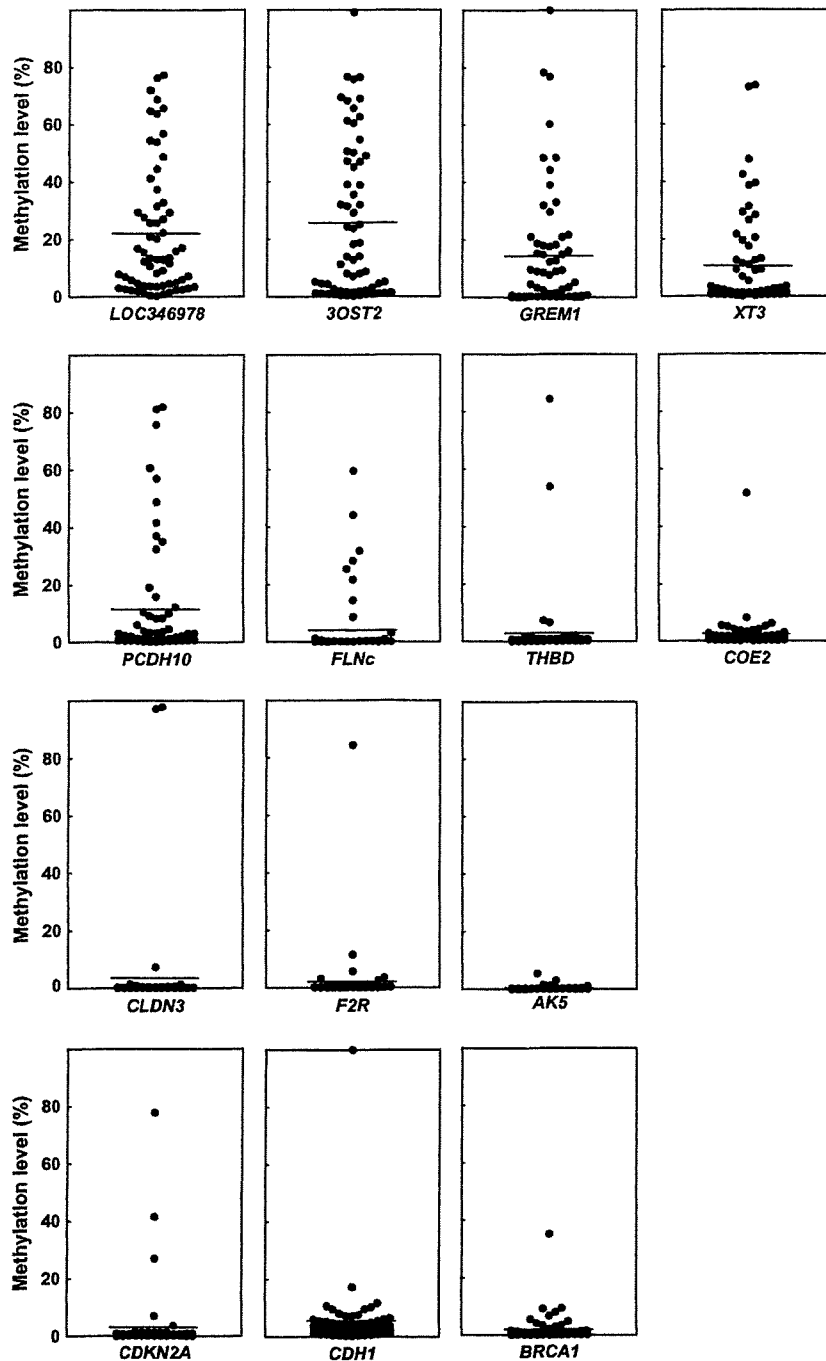
Fig. 1. Methylation levels in the 63 breast cancer samples. The numbers of DNA molecules methylated and unmethylated in promoter CGIs were obtained by quantitative methylation-specific PCR, and a methylation level was calculated as a fraction of methylated DNA molecules among the total DNA molecules. Some cancers had no methylation and the others had various levels of methylation. The methylation level in cancers was considered to represent the fraction of cancer cells in a sample and less occasionally the fraction of cells with methylation among cancer cells. We confirmed that we detected dense methylation of promoter CGIs by sequencing the quantitative methylation-specific PCR products obtained using primers specific to methylated DNA molecules (suppplementary Figure 3 is available at *Carcinogenesis* Online).

that with unmethylated sequences in a test sample were measured by comparing its amplification with those of standard samples that contained $10–10^6$ DNA molecules. The standard samples were prepared by cloning PCR products of methylated and unmethylated sequences into the pGEM-T Easy vector (Promega, Madison, WI) or by purifying their PCR products using the Wizard SV Gel and PCR clean-up system (Promega). The 'methylation level' was

calculated as the fraction of methylated DNA molecules among the total DNA molecules.

Fluorescence in situ hybridization analysis of the HER2 amplification

Fluorescence *in situ* hybridization was performed using a PathVysion kit (Abbot Molecular, Des Plaines, IL) with our modification (24). The *HER2* locus

467

and centromere of chromosome 17 (CEP17) were labeled by SpectrumOrange and SpectrumGreen fluorescence, respectively, and nuclei were counterstained with 4′, 6-diamidino-2-phenylindole. *HER2* and CEP17 signals were counted in 60 nuclei under a fluorescence microscope. Cancers with *HER2*:CEP17 ratio ≥2 were determined as *HER2* amplification positive.

Analysis of *3OST2* expression on cell growth

MCF7 Tet-Off cell line was purchased from Clontech Laboratories (Mountain View, CA). Full-length *3OST2* complementary DNA, cloned from human mammary epithelial cells, was inserted into the multiple cloning site of pTRE2hyg vector (Clontech Laboratories). The MCF7 Tet-Off cell line was transfected with the vector, and a stable clone was obtained by selection using hygromycin. Growth curves were analyzed by counting the cell numbers for the parental cell line, stable clones transfected with *3OST2*-expressing vector and with empty vector (without doxycycline). Overexpression of *3OST2* complementary DNA was confirmed by real-time reverse transcription–PCR analysis.

Sequencing analysis of quantitative methylation-specific PCR products

Quantitative methylation-specific PCR products of seven genes, *3OST2*, *FLNc*, *GREM1*, *THBD*, *PCDH10*, *XT3* and *LOC346978*, were cloned into pGEM-T Easy Vector (Promega). For each sample, ~10 clones were cycle sequenced using T7 primer, 5′-TAATACGACTCACTATAGGG-3′ and an Applied Biosystems 310 sequencer (Applied Biosystems, Foster City, CA).

Statistical analysis

Increasing or decreasing trends in no methylation, moderate methylation and then frequent methylation cancers were analyzed by the Mantel–Haenszel chi-square test. Differences between the frequent methylation *HER2*-positive can-

cers and moderate methylation (or no methylation) *HER2*-positive cancers were analyzed by the chi-square test. Correlation between the degree of *HER2* amplification and the number of methylated genes was analyzed using Pearson's correlation coefficient. All the analyses were performed using SPSS (SPSS, Chicago, IL).

Results

Quantitative methylation analysis of breast cancers

From 20 and 14 genes that were methylated in human breast and gastric cancers, respectively (18,19), we selected 11 genes (*LOC346978*, *3OST2*, *GREM1*, *XT3*, *PCDH10*, *FLNc*, *THBD*, *COE2*, *CLDN3*, *F2R* and *AK5*) and quantified their methylation levels in 63 breast cancers. These genes, except for *3OST2* and *CLDN3*, were not expressed in normal human mammary epithelial cells (18,19,25,26), and their silencing was unlikely to confer growth advantage to cells. Also, introduction of *3OST2* complementary DNA into MCF7 cells did not cause growth suppression (supplementary Figure 1 is available at *Carcinogenesis* Online), and its silencing was unlikely to confer growth advantage. Therefore, the majority of the 11 genes were considered to be suitable to detect the presence of a cellular environment that tends to induce methylation of promoter CGIs. We also analyzed methylation of three tumor suppressor genes (*CDKN2A*, *BRCA1* and *CDH1*) for clinicopathological analysis.

Quantitative methylation analyses of the 14 genes showed that some cancers had no methylation and the others had various levels
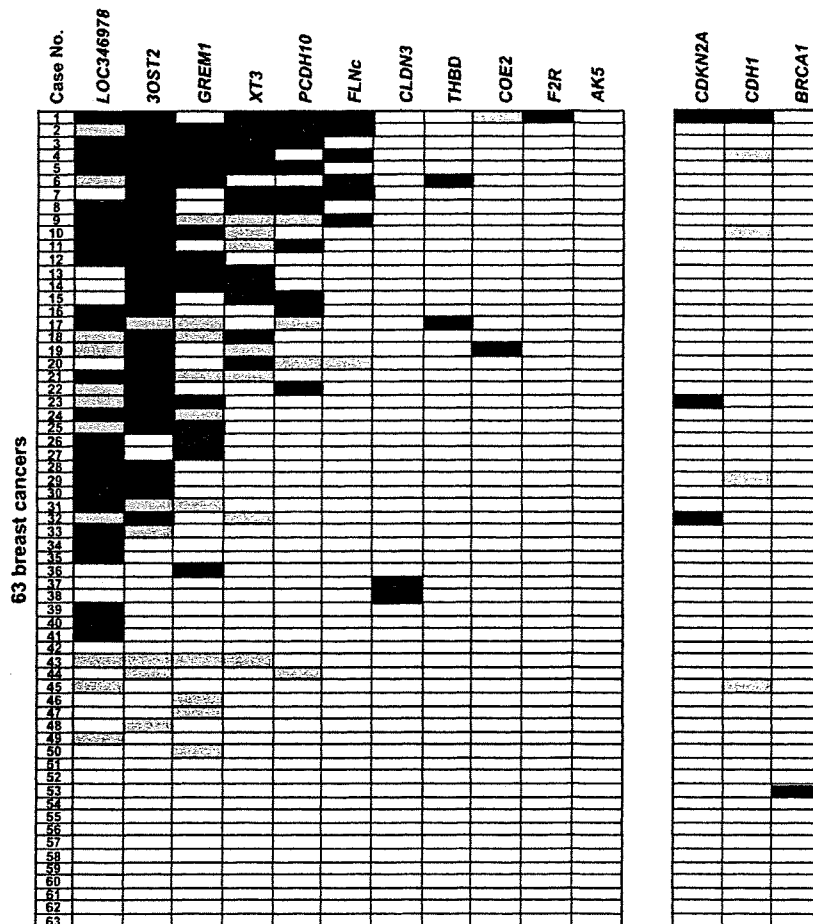
Fig. 2. Methylation profile of the 11 marker genes and three tumor suppressor genes in 63 breast cancers. Methylation in each sample was scored as positive or negative using two different cutoff values, and the 63 samples were aligned by the number of methylated CGIs. Methylation-positive samples using 10 and 20% as cutoff values are shown by gray and black boxes, respectively.

468

of methylation (Figure 1). Such distribution of methylation levels was typically observed for *FLNc*, *THBD*, *CLDN3*, *F2R* and *CDKN2A*. The presence of such distribution confirmed previous findings that cancer samples could essentially be classified into two groups: cancers with methylation of a specific gene and those without (11,19,27). Counting cancer cells in the tissue section samples showed that two samples with least cancer cells contained cancer cells with fractions of 19.8 ± 5.2% and 22.9 ± 0.3% (mean ± SD). Based on these data, we adopted two cutoff values 10 and 20% to score each cancer sample as positive or negative. When overall distribution of methylation was examined, similar patterns of cancers with methylation were observed using the two cutoff values (Figure 2). Using either value, the number of methylated genes in a cancer did not obey bimodal distribution and looked quite similar (Figure 3). Therefore, we adopted a cutoff value of 20% to score individual cancers as positive or negative for methylation.

Then, the 63 cancers were classified by the frequency of CGI methylation. To avoid biases due to a cutoff number of methylated genes, we classified the cancers into three groups, those with no methylation, moderate methylation and frequent methylation, using two different cutoff numbers for frequent methylation. Using a cutoff number of three methylated genes or more, 16, 26 and 21 cases were classified into cancers with frequent methylation, moderate methylation and no methylation, respectively. Using a cutoff number of four methylated genes or more, 8, 34 and 21 cases were classified into those with frequent methylation, moderate methylation and no methylation, respectively.

Association between frequent CGI methylation and the HER2 amplification

The presence of *HER2* amplification was analyzed by fluorescence *in situ* hybridization, and 24 of 63 (38%) cancers had *HER2* amplification (supplementary Figure 2 is available at *Carcinogenesis* Online). The extent of amplification ranged from 2.0- to 16.8-fold. Using a cutoff number of three for frequent methylation, the fractions

of cancers with *HER2* amplification were 11/16, 11/26 and 2/21 in cancers with frequent methylation, moderate methylation and no methylation, respectively (Figure 4A). Using a cutoff number of four, it was 6/8, 16/34 and 2/21, respectively (Figure 4B).

When correlation between the degree of CGI methylation and fraction of cancers with *HER2* amplification was examined by trend analysis, a highly significant increasing trend was observed from cancers with no methylation, to those with moderate methylation and then to those with frequent methylation ($P < 0.001$ for both of cutoff numbers). When cancers with frequent methylation and those with no methylation were compared, the former had a significantly higher fraction ($P = 0.003$ and 0.001 for cutoff numbers of four and three, respectively). Also, the degree of *HER2* amplification showed a correlation with the number of methylated genes (correlation coefficient = 0.411, $P = 0.002$) (Figure 5 and supplementary Table 2 is available at *Carcinogenesis* Online). This demonstrated that frequent CGI methylation had an association with *HER2* amplification.
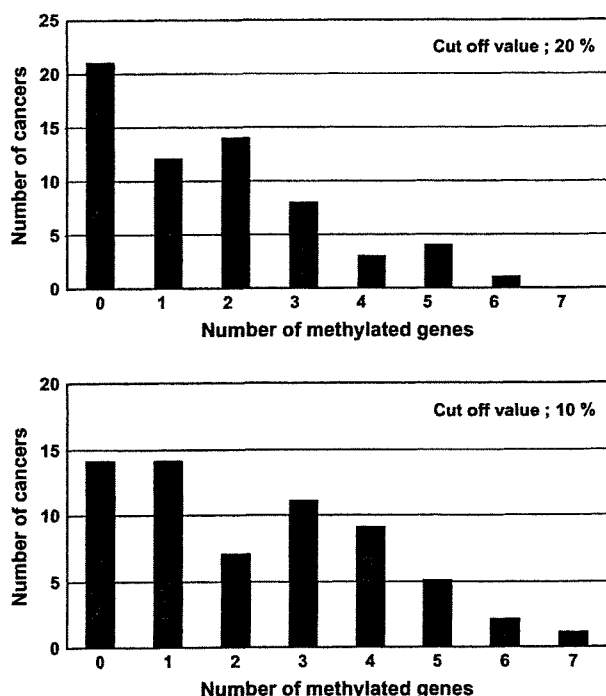

Fig. 3. Distribution of the number of methylated genes in a cancer. Similar distribution patterns were observed using two different cutoff values. The distribution was not bimodal, and involvement of multiple mechanisms in frequent methylation in breast cancers was suggested.
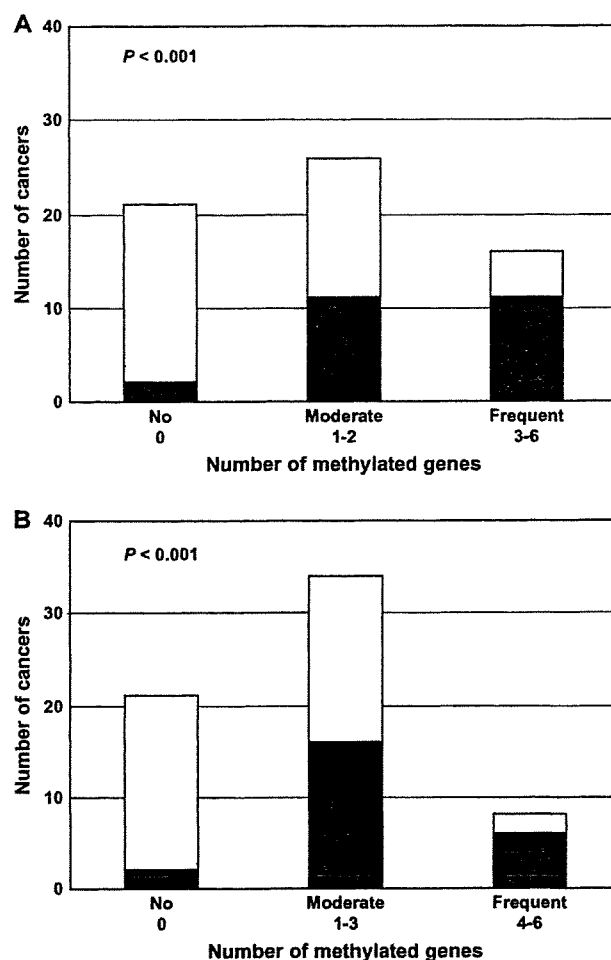

Fig. 4. The correlation between the degree of frequent methylation and *HER2* amplification. Two different cutoff numbers were used to define frequent methylation. These analyses adopted a cutoff value of 20% for methylation-positive. (A) Frequent methylation was defined as cancers with methylation of three or more genes. (B) Frequent methylation was defined as cancers with methylation of four or more genes. Whichever cutoff number was used, a clear increasing trend of *HER2*-positive cancers in no methylation, moderate methylation and then frequent methylation groups was observed ($P < 0.001$ for both of the two cutoff numbers). Closed and open boxes represent cancers with and without *HER2* amplification, respectively.

469

Association between frequent CGI methylation and other clinicopathological features, including methylation of tumor suppressor genes

The correlation between frequent CGI methylation and methylation of three tumor suppressor genes, *CDKN2A*, *CDH1* and *BRCA1*, was analyzed (Table I). However, none of the three genes showed any correlation ($P = 0.557$, 0.157 and 0.232, respectively). Regarding other clinicopathological characteristics, the degree of frequent CGI methylation correlated with higher nuclear grades ($P = 0.001$). The degree of frequent CGI methylation tended to show correlations with advanced pathological stage ($P = 0.068$) and post-menopausal status ($P = 0.044$). However, no association was observed with lymph node metastasis and negative expression of estrogen receptor (ESR) or progesterone receptor (PGR).

Discussion

The present study demonstrated for the first time that frequent CGI methylation in breast cancers had a highly significant association with *HER2* amplification. Regarding DNA methylation and HER2 over-
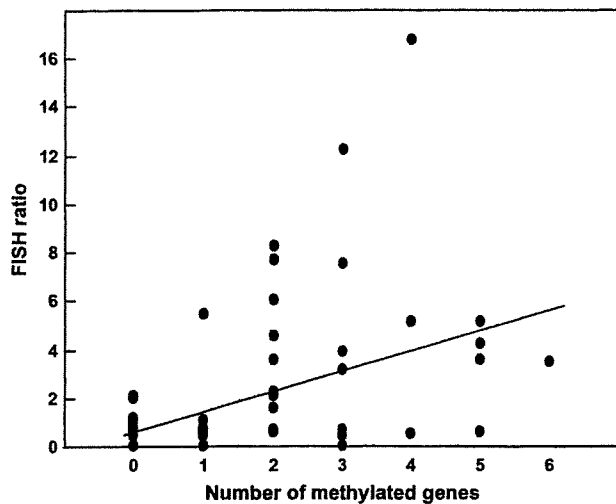

Fig. 5. The correlation between the number of methylated genes and degree of *HER2* amplification. The degree of *HER2* amplification showed a strong correlation with the number of methylated genes (correlation coefficient = 0.411, $P = 0.002$).

Table I. Association between frequent CGI methylation and clinicopathological features, including methylation of tumor suppressor genes

	No methylation	Moderate methylation	Frequent methylation	P value
CDKN2A methylation (+/−)	2/19	0/26	1/15	0.557
BRCA1 methylation (+/−)	1/20	0/26	0/16	0.232
CDH1 methylation (+/−)	0/21	0/26	1/15	0.157
Menopausal (pre/post)	12/9	9/17	4/12	0.044
Stage (I/II/III)	9/11/1	9/8/9	4/7/5	0.068
Lymph node metastasis (positive/negative)	6/15	13/13	7/9	0.308
ESR (positive/negative)	15/6	14/12	11/5	0.779
PGR (positive/negative)	17/4	14/12	9/7	0.100
Nuclear grade (1/2/3)	4/10/7	2/7/17	0/2/14	0.001

Frequent methylation was defined as breast cancers with methylation of three or more genes. Increasing or decreasing trends were tested by Mantel–Haenszel chi-square.

expression, Fiegl *et al.* (28) previously found that methylation levels of four genes (*CDH13*, *PGR*, *HSD17B4* and *MYOD1*) and one gene (*BRCA1*), which were selected from 35 genes, correlated with HER2 expression positively and inversely, respectively ($P = 0.01$–0.04). Methylation levels of individual genes in cancers are affected by the content of cancer cells, and, also, the correlation observed in the study was considered to be due to interaction between a function of an individual gene and HER2 overexpression. In contrast, here, we focused on the abnormality in epigenetic regulation in cancers. To estimate its degree, we used marker genes that were unlikely to confer growth advantages even if methylated, scored their methylation as positive or negative and integrated the information from the 11 marker genes into the frequency of methylation in a cancer sample. The cancers were classified into three groups, namely those with frequent methylation, moderate methylation and no methylation. As a result, a very strong association between frequent methylation and *HER2* amplification ($P < 0.001$) was demonstrated. Also, the degree of frequent methylation showed a clear correlation with the degree of *HER2* amplification. *BRCA1* methylation did not correlate with the degree of frequent methylation or *HER2* amplification ($P = 0.806$).

The association between frequent methylation and *HER2* amplification has clinical implications. It is known that *HER2* amplification status can show a discrepancy between primary and metastatic sites in a small fraction of patients (29). There is a possibility that HER2-negative breast cancers at initial diagnosis change into HER2 positive at their recurrence and that the presence of frequent methylation at the initial diagnosis can be used to predict such cases. Since accurate detection of HER2-positive cancers is very important to implement appropriate treatment, including trastuzumab (17), future studies to predict the *HER2* amplification status using frequent methylation and to clarify the mechanism of the association are warranted. Also, the effect of frequent methylation on long-term survival is important. So far, only 5 of 63 cases suffered from recurrence (one frequent methylation, three moderate methylation and one no methylation cases), and the effect cannot be statistically analyzed. Since the association between *HER2* amplification and poor survival (without trastuzumab) is well established, the effect of frequent methylation on long-term survival seems worth being analyzed in the future.

Some breast cancers with *HER2* amplification belonged to the moderate methylation or no methylation groups although the majority of cancers with *HER2* amplification belonged to the frequent methylation group. This was in contrast with the case of neuroblastomas, where all the neuroblastomas with *MYCN* amplification had frequent methylation, CIMP, with only one exception (12,13). Therefore, the relationship between frequent methylation and *HER2* amplification in breast cancers seems more complex than the relationship between CIMP and *MYCN* amplification in neuroblastomas. Not only frequent methylation could lead to *HER2* amplification through chromosomal instability (9), which was our initial expectation, but also *HER2* amplification could lead to frequent methylation or they might have common inducers.

The degree of frequent methylation also correlated with higher nuclear grades. It also tended to show association with advanced stages and post-menopausal status. It has been reported that *CDH1* methylation was associated with negative ESR and PGR expressions ($P = 0.06$ and 0.09, respectively) and that frequent methylation of seven tumor suppressor genes was associated with poor differentiation (30). It has also been reported that PGR expression was negatively associated with *ESR1*, *TGFBR2*, *PPTGS2* and *CDH13* methylation ($P = 0.01$–0.04) (31) and that ESR and PGR expressions were positively and negatively associated with *HIN-1/RASSF1A* and *RIL/CDH13* methylation, respectively (32). Taken together, the frequent methylation in breast cancers was weakly associated with advanced stages, negative PGR and ESR expressions and poor differentiation (higher nuclear grades). Nevertheless, the correlation between frequent methylation and *HER2* amplification was much stronger than these associations in our study. It was considered that quantitative analysis of marker genes was advantageous to clarify the strong association.

470

The cutoff value of methylation levels to score cancer samples as positive or negative for methylation was determined based upon the fraction of cancer cells in two samples with their smallest contents (20%). To count methylation in a fraction of cancer cells, we also tested a cutoff value, 10%, but quite similar results were obtained (Figure 3). Regarding the cutoff number for frequent methylation, we tried three and four but observed a highly significant association using both numbers (Figure 4). This excluded a possibility that a false-positive association between frequent methylation and *HER2* amplification was observed due to arbitrary cutoff values or numbers. Also, we confirmed that we detected dense methylation of promoter CGIs by our quantitative methylation-specific PCR analysis by sequencing the PCR products. Almost all the CpG sites in the products were densely methylated (supplementary Figure 3 is available at *Carcinogenesis* Online). Finally, we confirmed that the methylation detected in cancer tissues originated from cancer cells. Methylation levels of nine genes that showed high methylation levels ($\geq 10\%$) in some cancer samples were measured in 11 pairs of non-cancerous breast and cancer tissues (supplementary Figure 4 is available at *Carcinogenesis* Online). The methylation levels of all the genes were elevated only in cancer tissues, and the methylation we detected was considered to originate from cancer cells.

In summary, frequent methylation in breast cancers had a strong association with *HER2* amplification.

Supplementary material

Supplementary Tables 1 and 2 and Figures 1–4 can be found at http://carcin.oxfordjournals.org/

Funding

Acknowledgements

References

1. Herman,J.G. *et al.* (2003) Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.*, 349, 2042–2054.
2. Baylin,S.B. *et al.* (2006) Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer*, 6, 107–116.
3. Esteller,M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, 8, 286–298.
4. Jones,P.A. *et al.* (2007) The epigenomics of cancer. *Cell*, 128, 683–692.
5. Chen,R.Z. *et al.* (1998) DNA hypomethylation leads to elevated mutation rates. *Nature*, 395, 89–93.
6. Gaudet,F. *et al.* (2003) Induction of tumors in mice by genomic hypomethylation. *Science*, 300, 489–492.
7. Nakagawa,T. *et al.* (2005) DNA hypomethylation on pericentromeric satellite regions significantly correlates with loss of heterozygosity on chromosome 9 in urothelial carcinomas. *J. Urol.*, 173, 243–246.
8. Rodriguez,J. *et al.* (2006) Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res.*, 66, 8462–9468.
9. Kondo,Y. *et al.* (2000) Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis—a comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. *Hepatology*, 32, 970–979.
10. Toyota,M. *et al.* (1999) CpG island methylator phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA*, 96, 8681–8686.
11. Weisenberger,D.J. *et al.* (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat. Genet.*, 38, 787–793.
12. Abe,M. *et al.* (2005) CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas. *Cancer Res.*, 65, 828–834.
13. Abe,M. *et al.* (2007) Marked and independent prognostic significance of the CpG island methylator phenotype in neuroblastomas. *Cancer Lett.*, 247, 253–258.
14. King,C.R. *et al.* (1985) Amplification of a novel v-*erb*B-related gene in a human mammary carcinoma. *Science*, 229, 974–976.
15. Slamon,D.J. *et al.* (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science*, 235, 177–182.
16. Arteaga,C.L. (2003) Trastuzumab, an appropriate first-line single-agent therapy for HER2-overexpressing metastatic breast cancer. *Breast Cancer Res.*, 5, 96–100.
17. Hudis,C.A. (2007) Trastuzumab—mechanism of action and use in clinical practice. *N. Engl. J. Med.*, 357, 39–51.
18. Miyamoto,K. *et al.* (2005) Identification of 20 genes aberrantly methylated in human breast cancers. *Int. J. Cancer*, 116, 407–414.
19. Enomoto,S. *et al.* (2007) Lack of association between CpG island methylator phenotype in human gastric cancers and methylation in their background non-cancerous gastric mucosae. *Cancer Sci.*, 98, 1853–1861.
20. Li,B. *et al.* (2007) The role of chromatin during transcription. *Cell*, 128, 707–719.
21. Ushijima,T. (2005) Detection and interpretation of altered methylation patterns in cancer cells. *Nat. Rev. Cancer*, 5, 223–231.
22. Sakamoto,G. *et al.* (2005) General rules for clinical and pathological recording of breast cancer 2005. *Breast Cancer*, 12 (suppl.), S1–S27.
23. Kaneda,A. *et al.* (2004) Decreased expression of the seven *ARP2/3 complex* genes in human gastric cancers. *Cancer Lett.*, 212, 203–210.
24. Taniyama,K. *et al.* (2008) Tyrosine1248-phosphorylated HER2 expression and HER2 gene amplification in female invasive ductal carcinomas. *Breast Cancer*, 15, 231–240.
25. Van Rompay,A.R. *et al.* (1999) Identification of a novel human adenylate kinase. cDNA cloning, expression analysis, chromosome localization and characterization of the recombinant protein. *Eur. J. Biochem.*, 261, 509–517.
26. Reference database for gene Expression Analysis (RefExA). http://www.Isbm.org/database/index.html.
27. Ogino,S. *et al.* (2006) CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut*, 55, 1000–1006.
28. Fiegl,H. *et al.* (2006) Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. *Cancer Res.*, 66, 29–33.
29. Simon,R. *et al.* (2001) Patterns of her-2/neu amplification and overexpression in primary and metastatic breast cancer. *J. Natl Cancer Inst.*, 93, 1141–1146.
30. Li,S. *et al.* (2006) DNA hypermethylation in breast cancer and its association with clinicopathological features. *Cancer Lett.*, 237, 272–280.
31. Widschwendter,M. *et al.* (2004) Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res.*, 64, 3807–3813.
32. Feng,W. *et al.* (2007) Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast Cancer Res.*, 9, R57.

Improved Method for Analysis of RNA Present in Long-Term Preserved Thyroid Cancer Tissue of Atomic Bomb Survivors

Kiyohiro Hamatani,[1] Hidetaka Eguchi,[1,2] Mayumi Mukai,[1] Kazuaki Koyama,[1] Masataka Taga,[1] Reiko Ito,[1] Yuzo Hayashi,[3] and Kei Nakachi[1]

Background: Since many thyroid cancer tissue samples from atomic bomb (A-bomb) survivors have been preserved for several decades as unbuffered formalin-fixed, paraffin-embedded specimens, molecular onco-logical analysis of such archival specimens is indispensable for clarifying the mechanisms of thyroid carcino-genesis in A-bomb survivors. Although *RET* gene rearrangements are the most important targets, it is a difficult task to examine all of the 13 known types of *RET* gene rearrangements with the use of the limited quantity of RNA that has been extracted from invaluable paraffin-embedded tissue specimens of A-bomb survivors. In this study, we established an improved 5' rapid amplification of cDNA ends (RACE) method using a small amount of RNA extracted from archival thyroid cancer tissue specimens.

Methods: Three archival thyroid cancer tissue specimens from three different patients were used as in-house controls to determine the conditions for an improved switching mechanism at 5' end of RNA transcript (SMART™) RACE method; one tissue specimen with *RET/PTC1* rearrangement and one with *RET/PTC3* re-arrangement were used as positive samples. One other specimen, used as a negative sample, revealed no detectable expression of the *RET* gene tyrosine kinase domain.

Results: We established a 5' RACE method using an amount of RNA as small as 10 ng extracted from long-term preserved, unbuffered formalin-fixed, paraffin-embedded thyroid cancer tissue by application of SMART technology. This improved SMART RACE method not only identified common *RET* gene rearrangements, but also isolated a clone containing a 93-bp insert of rare *RTE/PTC8* in RNA extracted from formalin-fixed, paraffin-embedded thyroid cancer specimens from one A-bomb survivor who had been exposed to a high radiation dose. In addition, in the papillary thyroid cancer of another high-dose A-bomb survivor, this method detected one novel type of *RET* gene rearrangement whose partner gene is acyl coenzyme A binding domain 5, located on chromosome 10p.

Conclusion: We conclude that our improved SMART RACE method is expected to prove useful in molecular analyses using archival formalin-fixed, paraffin-embedded tissue samples of limited quantity.

Background

MOLECULAR ONCOLOGICAL ANALYSES of archival tissue specimens are indispensable, especially for studies of rare cancers or cancers associated with uncommon past events, such as radiation exposures in Thorotrast treatment, nuclear power plant accidents, and atomic bombings. In re-cent years, application of new molecular techniques, includ-ing polymerase chain reaction (PCR) in the use of archival tissue samples, is anticipated to advance the understanding of the molecular mechanisms of these cancers (1).

Since many thyroid cancer tissue samples from atomic bomb (A-bomb) survivors have been preserved for a long time as unbuffered formalin-fixed, paraffin-embedded specimens, analysis of such archival specimens is indispensable for clari-fying the characteristics of thyroid carcinogenesis in A-bomb survivors. *RET* gene rearrangements are thought to play an initial and critical role in papillary thyroid cancer develop-ment, although wide variations in the prevalence of such re-arrangements, ranging from 3% to 70%, have been observed in different geographic areas (2–4). To date, at least 13 rear-ranged forms of the *RET* gene (*RET/PTC 1–9, PCM1-RET,*

[1]Department of Radiobiology/Molecular Epidemiology, Radiation Effects Research Foundation, Hiroshima, Japan.
[2]Translational Research Center, Saitama Medical University, International Medical Center, Saitama, Hidaka, Japan.
[3]Geriatric Health Service Facility Hidamari, Hiroshima, Japan.

43

ELKS-RET, ΔRFP-RET, and *HOOK3-RET*) have been identified, of which *RET/PTC1* and *RET/PTC3* are by far the most frequent, from papillary thyroid cancers in patients both with and without radiation-exposure history (3,5–7). Analysis of *RET* gene rearrangements is one of the most important issues in molecular thyroid cancer research among A-bomb survivors. However, it is a difficult task to examine all types of *RET* rearrangements because of the limited quantity of paraffin-embedded tissue specimens from A-bomb survivors; namely, it is almost impossible to investigate all 13 rearranged forms of the *RET* gene using such limited amounts of RNA.

Clontech's (Mountain View, CA) switching mechanism at 5′ end of RNA transcript (SMART™) technology is based on the terminal transferase activity of reverse transcriptase, which adds a few additional nucleotides, primarily deoxycytidine, to the 3′ end of cDNA, creating an extended template at the 5′ end in combination with a SMART oligonucleotide (8). Therefore, PCR amplification of cDNA using a set of adaptor-specific primer and gene-specific primer will produce the cDNA that contains the complete 5′ end of mRNA (9).

We improved and established the 5′ rapid amplification of cDNA ends (RACE) method (10) using a very small amount of RNA extracted from unbuffered paraffin-embedded thyroid cancer tissue by application of SMART technology. In addition to the analysis of common *RET* gene rearrangements, this method succeeded in detecting not only rare *RET* gene rearrangements but also one *RET* gene rearrangement that has not yet been reported.

Materials and Methods

Tissue

All cancer tissue specimens used for this study were prepared from archival unbuffered formalin-fixed, paraffin-embedded blocks. Three thyroid cancer tissue specimens from three different patients were used as in-house controls to determine the conditions for an improved SMART RACE method; one tissue specimen with *RET/PTC1* rearrangement and one with *RET/PTC3* rearrangement were used as positive samples. One other specimen, used as a negative sample, revealed no detectable expression of the *RET* gene tyrosine kinase (TK) domain. The control samples had been preserved at room temperature for 19–21 years. Papillary thyroid cancer specimens from A-bomb survivors, preserved at room temperature for 20–50 years, were collected after approval from ·the Human Investigation Committee and the Ethics Committee for Genome Research at the Radiation Effects Research Foundation (RERF). After deparaffinization of 5 μm sections by Hemo-De (Fujisawa Yakuhin Kogyo, Osaka, Japan) and staining with methyl green (Sigma-Aldrich, St. Louis, MO), cancerous regions were isolated using disposable scalpels. All cancerous regions from two to four successive tissue sections were combined for RNA extraction.

RNA extraction

RNA was isolated from dissected tissue using the High Pure RNA Paraffin kit according to the manufacture's instructions (Roche Diagnostics, Manheim, Germany), with some modifications. Briefly, dissected tissue was digested with proteinase K at 55°C overnight, followed by DNase I treatment. After the lysate was purified by High Pure filter, RNA was eluted twice

with 100 μL of RNase-free water. RNA was then precipitated by ethanol in the presence of 2 μL of ethachinmate (Nippon Gene, Tokyo, Japan) as a carrier and resuspended in 30 μL of RNase-free water.

cDNA synthesis

To enhance template activity of RNA extracted from archival unbuffered formalin-fixed, paraffin-embedded tissue specimens, total RNA was first heated in 10 mM of citrate buffer (pH 4.0) at 70°C for 45 minutes and then precipitated by ethanol (11). Then, 100 ng of total RNA and 50 pmol/μL of random primers (9 mer) were heated in 11 μL of RNase-free water at 65°C for 10 minutes and chilled in ice water. A mixture consisting of 4 μL of 5× reverse transcription (RT) buffer, 2 μL of 20 mM DTT, 1 μL of 10 mM dNTPs, and 1 μL of RNase Inhibitor (20 U/μL; TaKaRa, Tokyo, Japan) was added to the RNA solution and incubated at room temperature for 5 minutes. After addition of 1 μL of Rever Tra Ace (100 U/μL; Toyobo, Osaka, Japan), the reaction mixture was incubated at 42°C for 45 minutes.

Improved SMART RACE

After 1 μL of various concentrations of SMART adaptor (SMART II A oligonucleotide; Clontech) was added to 20 μL of cDNA solution, the mixture was further incubated at 42°C for 60 minutes and heated at 70°C for 15 minutes to inactivate reverse transcriptase. SMART RACE-PCR was performed in a 25 μL volume containing 1×PCR buffer; 200 μM each of dATP, dCTP, dGTP, and dTTP; 3 mM of $MgCl_2$; 0.4 μM of each primer (SMART adaptor-specific primer, S-RACE 1, 5′-AAGCAGTGGTAACAACGCAGAGTA-3′; exon 12 of *RET* gene-specific primer, *RET*-Ex12PR9, 5′-TCCGAGGGAATTC CCACTTT-3′); 0.5 U of FastStart Taq DNA polymerase containing a thermostable proof-reading protein (Roche Diagnostics); and 2 μL of SMART adaptor-treated cDNA. PCR was carried out on a DNA Engine using the following cycle conditions: at 95°C for 3 minutes; 45 cycles at 95°C for 30 seconds, at 60°C for 30 seconds, and at 72°C for 45 seconds; and a final extension at 72°C for 5 minutes. The first PCR products were diluted at 1000- to 10,000-fold in water, and 1 μL of diluted samples was used as a template for semi-nested SMART RACE-PCR with SMART adaptor and nested primer (*RET* gene-specific primer, *RET*-Ex12A4; Table 1). The second PCR conditions were as described in the first PCR, with the exception of annealing temperature (66°C instead of 60°C) and number of cycles (25 cycles instead of 45 cycles).

Cloning and sequencing of cDNA fragments

Five microliters of second SMART RACE-PCR products was electrophoresed on an 8% acrylamide gel and viewed with ethidium bromide. Target candidate bands were determined by *Bam*HI (TaKaRa) digestion of aliquots of second SMART RACE-PCR products, because the fragments amplified by SMART RACE-PCR included a *Bam*HI recognition site located at the 5′ end of exon 12 of the *RET* gene. Target candidate cDNA fragments derived from 50 μL of second SMART RACE-PCR products were eluted from 8% acrylamide gel and cloned into *Hinc*II-digested and dephosphorylated pUC118 vector using TaKaRa Blunting Kination Ligation kit (TaKaRa). Plasmid DNA containing a longer insert than 70 bp was

TABLE 1. PRIMERS AND OLIGONUCLEOTIDES USED FOR DETECTION OF mRNA EXPRESSION BY REVERSE TRANSCRIPTION-POLYMERASE CHAIN REACTION

	Primer name	Sequence	Annealing temperature
BCR	BCR-S2	5'-GGTGGGCGACCTCTTCCAGA-3' (20 mer)	58°C
	BCR-A2	5'-TCCACGAAGGCCCGGTACAC-3' (20 mer)	
Kinase domain of RET	RET-Ex12S5	5'-GAGCAGGGTACACCACGGTGG-3' (21 mer)	60°C
	RET-Ex13A2	5'-CTCACTCGGGGAGGCGTTCT-3' (20 mer)	
RET/PTC1	PTC1S14	5'-AACCGCGACCTGCGCAAAGC-3' (20 mer)	58°C
	RET-Ex12A	5'-GAGGGAATTCCCACTTTGGA-3' (20 mer)	
RET/PTC3	PTC3/4–12	5'-ACCCAAAAGCAGACCTTGGA-3' (20 mer)	56°C
	RET-Ex12A4	5'-GAGGGAATTCCCACTTTGGA-3' (20 mer)	
RET/PTC8	PTC8S1	5'-GAGTGATCTTTCTAGCAAAACACA-3' (24 mer)	58°C
	RET-Ex12A4	5'-GAGGGAATTCCCACTTTGGA-3' (20 mer)	
	PTC8-oligo77	5'-AAGAGAGGGAGAGTGATCTTTCTAGCAAAA CACAGCTGTTACAGGAGGATCCAAAGT GGGAATTCCCTCAAGCCGAA-3' (77 mer)	
ACBD5-RET	ACBD5S1	5'-ACATTGCAGACTGCTCCTCA-3' (20 mer)	58°C
	RET-Ex12A4	5'-GAGGGAATTCCCACTTTGGA-3' (20 mer)	
	ACBD5-RET-oligo77	5'-TCAACATCAACATTGCAGACTGCTCCTCAG CCCACCTCACAGGAGGATCCAAAGTGGG AATTCCCTCGGAAGAACTT-3' (77 mer)	

sequenced using DNA sequencer CEQ8000 (Beckman Coulter, Fullerton, CA), because the total length of the SMART adaptor and the 5' portion of exon 12 of RET was 55 bp.

Detection of RET/PTC8 expression in papillary thyroid cancer from A-bomb survivors by RT-PCR

Amplification of the BCR gene as an internal control by RT-PCR was at first conducted for examination of availability of RNA extracted from archival tissue samples. The cDNA derived from 10 ng of total RNA was used as a template for RT-PCR. RT-PCR for detection of RET/PTC8 and ACBD5-RET was performed in a 25 μL volume of solution containing 1×PCR buffer; 200 μM each of dATP, dCTP, dGTP, and dTTP; MgSO$_4$ (3.5 mM for RET/PTC8 and 3.0 mM for ACBD5-RET); 0.4 μM of each primer (PTC8S1, RET-Ex12A4 for RET/PTC8; ACBD5S1, RET-Ex12A4 for ACBD5-RET); and 0.5 U of Platinum Taq DNA polymerase High Fidelity (Invitrogen, Carlsbad, CA). PCR conditions consisted of initial denaturation (95°C for 3 minutes), followed by 40 cycles (denaturation at 95°C for 30 seconds, annealing at 58°C for 30 seconds, and extension at 68°C for 45 seconds), and final extension at 68°C for 5 minutes. To detect expression of BCR, the TK domain of the RET gene, RET/PTC1, and RET/PTC3, RT-PCR was conducted using Platinum Taq DNA polymerase (Invitrogen) in the presence of 1×PCR buffer; 200 μM each of dATP, dCTP, dGTP, and dTTP; 2.5 mM MgCl$_2$ (2.0 mM for RET/PTC3); and 0.4 μM of each primer at 40 cycles for BCR, RET/PTC1, and RET/PTC3, and 36 cycles for the TK domain. Reaction conditions were as described in RET/PTC8, except for extension temperature (72°C instead of 68°C). Primer sets and annealing temperatures are summarized in Table 1. RT-PCR products of RET/PTC8 and ACBD5-RET were confirmed to be actual products by sequencing after cloning of corresponding bands into a cloning vector.

RT-PCR products other than RET/PTC8 and ACBD5-RET were confirmed to be actual products by digestion of restriction enzymes, BamHI for RET/PTC1 and RET/PTC3, AluI for BCR (TaKaRa), and HaeIII (New England Biolabs, Beverly, MA) for the RET gene TK domain, which existed within each amplified target fragment. For positive control of RT-PCR on the BCR gene, the RET gene TK domain and RET/PTC1, cDNA derived from a human thyroid cell line (TPC1) with RET/PTC1 rearrangement was used as a template. For positive control of RT-PCR on RET/PTC8 and ACBD5-RET, a mixture of 77 base-synthesized nucleotides (PTC8-oligo77 for RET/PTC8 and ACBD5-RET-oligo77 for ACBD5-RET, respectively) and genomic DNA at a molar ratio of 1:1 was used as a template. For negative control, H$_2$O was used as a template. No amplification of RNA without RT was observed.

Results

Concentration of SMART adaptor

All cDNA syntheses in this study were performed with random primer (9 mer), since RNA extracted from archival formalin-fixed, paraffin-embedded thyroid cancer tissue specimens was invariably degraded to some degree. Using RNA prepared from each of two in-house controls harboring RET/PTC1 and RET/PTC3 rearrangements, first we examined the effects of concentration of SMART adaptor on the amplification efficiency by SMART RACE-PCR with gene-specific and adaptor-specific primers. One microliter of various concentrations of SMART adaptor was added to the reaction mixture immediately after cDNA synthesis had been completed. The mixture was further incubated at 42°C for 60 minutes in the presence of SMART adaptor. Since the total length of the SMART adaptor and the 5' portion of exon 12 of RET is 55 bp, the length of amplified target fragments requires more than 70 bp to identify the counterpart gene. In addition, it is important to conduct cloning of obvious and intense fragments for successful isolation and identification of the counterpart gene. Therefore, among clear-cut fragments of

more than 70 bp, fragments with a BamHI digestion site were determined as target candidates, since the fragments amplified by SMART RACE-PCR included BamHI recognition site at the 5' end of exon 12 of the RET gene. All fragments in Figure 1A–D that could be recognized as a band were confirmed to be digested with BamHI. As shown in lanes 2–4 of Figure 1A and C, PCR products looked like smears without clear and intense bands when adaptor concentration was high (2–10 μM). PCR products with 0.2–1 μM of SMART adaptor revealed an obvious and intense band of more than 70 bp compared with the bands at other higher concentrations (lanes 5–7, Fig. 1A and C, indicated by asterisks), though the sizes of these target candidate fragments differed somewhat among various concentrations of SMART adaptor. Treatment with 0.1 μM of SMART adaptor showed that amplified bands of more than 70 bp in length were observed only infrequently (lane 8, Fig. 1A, C).

Addition timing of SMART adaptor

When cDNA was synthesized in the presence of SMART adaptor (2–10 μM), electrophoretic images of PCR products were similar to those in the case of addition of adaptor after completion of the cDNA synthesis (compare lanes 2–4 in Fig. 1B and D with lanes 2–4 in Fig. 1A and C). When cDNA was prepared in the presence of a lower concentration (0.2–1 μM) of SMART adaptor, the electrophoretic patterns differed from those where adaptor was added after completion of cDNA synthesis (compare lanes 5–7 in Fig. 1B and D with lanes 5–7 in Fig. 1A and C, respectively). Namely, weak bands of more than 70 bp were observed (lanes 5–7, Fig. 1B, D). Although treatment with 0.1 μM of SMART adaptor occasionally produced clear and intense bands (lane 8, Fig. 1B, indicated by an arrow), amplified bands of more than 70 bp in length were rarely observed (lane 8, Fig. 1D).
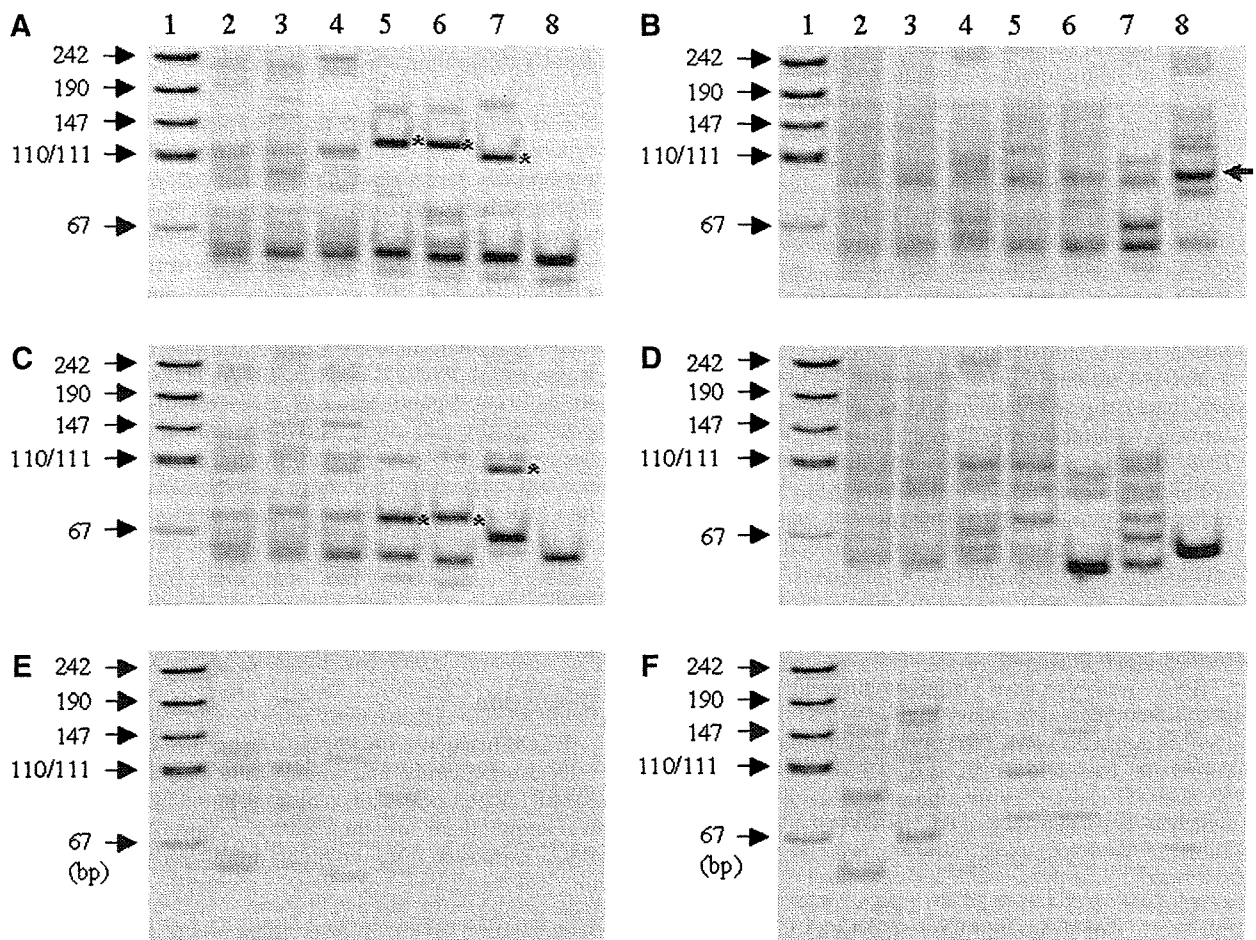

FIG. 1.   Effects of concentration and addition timing of SMART adaptor on improved SMART RACE-PCR amplification. (A, B) SMART RACE-PCR products with total RNA from specimen with RET/PTC1. (C, D) SMART RACE-PCR products with total RNA from specimen with RET/PTC3. (E, F) SMART RACE-PCR products with total RNA that revealed no detectable expression of the RET gene tyrosine kinase domain. (A, C, E) reveal RACE-PCR products when various concentrations of SMART adaptor were added after completion of cDNA synthesis (lane 2, 10 μM; lane 3, 5 μM; lane 4, 2 μM; lane 5, 1 μM; lane 6, 0.4 μM; lane 7, 0.2 μM; lane 8, 0. 1 μM). Lane 1 indicates pUC19-MspI digest for DNA size marker. (B, D, F) indicate RACE-PCR products when cDNA was synthesized in the presence of SMART adaptor. The asterisks and arrow indicate clear and intense candidate fragments measuring more than 70 bp. PCR, polymerase chain reaction; RACE, 5' rapid amplification of cDNA ends; SMART, switching mechanism at 5' end of RNA transcript.

When RNA extracted from archival samples with no detectable expression of *RET* gene TK domain was used, several weak bands were observed regardless of the addition timing of SMART adaptor. However, few bands (more than 70 bp) were found ·in the cases of 0.1, 0.2, or 0.4 μM of SMART adaptor (lanes 6–8, Fig. 1E, F). Not all weak bands in Figure 1E and F could be digested with *BamHI*. No bands were observed when H$_2$O was used as template for RACE-PCR. In addition, SMART RACE-PCR without reverse transcription produced no amplified bands whatsoever.

Incubation time

We examined the effects of incubation time after addition of adaptor on amplification by SMART RACE-PCR using RNA from specimens with *RET/PTC1*. After SMART adaptor was added to cDNA solution to give a final concentration of 0.2 μM, the reaction was incubated for various durations at 42°C. Among target candidate bands of about 90 bp (lanes 2–6, Fig. 2, indicated by asterisks) found in SMART RACE-PCR products until 45 minutes of incubation time, a case treated for 45 minutes revealed the most intense band (lane 6, Fig. 2). Incubation for 60 minutes resulted in a larger intense target candidate band (lane 7, Fig. 2) than did other incubation times. In the case of 90-minute incubation (lane 8, Fig. 2), the larger candidate band disappeared and a weak 90 bp candidate band reappeared (indicated by an asterisk).

We next examined how efficiently cDNA fragments of the target gene can be isolated by this improved SMART RACE method. Target candidate bands of about 100 bp (lane 7, Fig. 1A) and 90 bp (lane 7, Fig. 1C) in the second PCR products were eluted from 8% acrylamide gel and cloned onto a pUC118 plasmid vector. Plasmid DNA with a longer insert than 70 bp was sequenced. As shown in Table 2, all 36 screened colonies proved to be plasmid DNA harboring *RET/PTC1* fragments of 97 bp in length. *RET/PTC3* fragments with a length of 89 bp were also detected in 35 out of 36 colonies, as shown in Table 2. This suggests that almost all candidate fragments amplified by the improved SMART RACE method are actually target cDNA fragments.

RET/PTC8 in thyroid cancer tissue from A-bomb survivors

RNA was extracted from unbuffered formalin-fixed, paraffin-embedded papillary thyroid cancer tissue specimens from 64 subjects exposed to A-bomb radiation, and *BCR* gene expression was detected in 52 of the 64 cases. Among these 52 cases, 11 showed expression of *RET* gene TK domain, of which 9 cases had *RET/PTC1* or *RET/PTC3*, as shown in Table 3. Two papillary thyroid cancer cases showing no
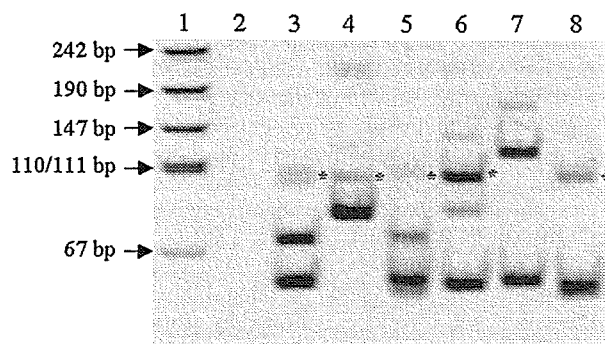

FIG. 2. Effect of incubation time on improved SMART RACE-PCR amplification. The reaction was incubated for various times at 37°C after SMART adaptor was added to cDNA solution to give a final concentration of 0.05 μM (lane 2, 0 minute; lane 3, 10 minutes; lane 4, 20 minutes; lane 5, 30 minutes; lane 6, 45 minutes; lane 7, 60 minutes; lane 8, 90 minutes). Lane 1 indicates pUC19-*MspI* digest for DNA size marker. The asterisks indicate fragments of about 90 bp with *BamHI* recognition site.

RET/PTC1 or *RET/PTC3* rearrangement were examined for type of rearrangement using this improved SMART RACE method. The target band in the second PCR products (Fig. 3A, indicated by an arrow) was eluted and cloned into a cloning vector. Since 23 out of 24 white colonies contained plasmid DNA with *BamHI* site and a longer insert than about 70 bp, plasmid DNA in the 23 colonies was sequenced. The sequencing of the portion upstream of *RET* exon 12 found in all clones was identical to the 5′ part of the kinectin 1 gene, whose insert length was 93 bp (Fig. 3B). As shown in Figure 3C, the band corresponding to *RET/PTC8* was also detected in the papillary thyroid cancer specimen from the same A-bomb survivor by RT-PCR. Further, the improved SMART RACE method enabled us to identify a novel rearrangement of the *RET* gene in papillary thyroid cancer of another A-bomb survivor. This new type of *RET/PTC*, whose partner gene is acyl coenzyme A binding domain 5 (*ACBD5*) located on chromosome 10p, is now being analyzed for its tumorigenicity. Expression of the *ACBD5-RET* fusion gene was confirmed in RNA from the cancer specimens by RT-PCR (Fig. 3C).

Discussion

RACE reaction with RNA extracted from fresh tissues or cells can be easily and successfully performed using the SMART RACE kit (Clontech). On the other hand, as far as we know, there are thus far no reports on RACE with RNA extracted from archival unbuffered formalin-fixed,

TABLE 2. ACCURACY OF TARGET cDNA FRAGMENT AMPLIFIED BY IMPROVED SWITCHING MECHANISM AT 5′ END OF RNA TRANSCRIPT 5′ RAPID AMPLIFICATION OF cDNA ENDS METHOD

	No. of screened white color colonies	No. of BamHI-positive colonies	No. of confirmation by sequencing	Efficiency (%)
No 1[a]	36	36	36	100
No 2[b]	36	36	35[c]	97

[a]Thyroid cancer with *RET/PTC1*.
[b]Thyroid cancer with *RET/PTC3*.
[c]One clone failed to perform sequencing.

TABLE 3. NUMBER OF CASES EXPOSED TO ATOMIC BOMB RADIATION IN THIS STUDY

Total tested	BCR positive	RET-TK positive	RET/PTC1 positive	RET/PTC8 positive	ACBD5-RET positive
64	52	11	9[a]	1	1

[a]One case carried both RET/PTC1 and RET/PTC3.
TK, tyrosine\kinase.

paraffin-embedded tissue samples. We succeeded in identification and isolation of a partner gene of rearranged RET using RNA extracted from archival unbuffered formalin-fixed and paraffin-embedded thyroid cancer tissue specimens by improved SMART RACE method. A series of experiments from cDNA synthesis to PCR was conducted at least three times. Similar electrophoresis patterns could be reproduced, though the sizes of the target bands differed somewhat. Size differences of amplified fragments between experiments may largely have been due to degradation of RNA extracted from archival samples, since such a phenomenon was not observed when cDNA was prepared with the random primer (9 mer) using intact RNA from fresh samples (data not shown).

We found that the SMART RACE method enabled isolation of 5' upstream fragments in target cDNA with RNA extracted from archival formalin-fixed, paraffin-embedded tissue specimens by adjusting the concentration and the timing of the addition of SMART adaptor. The optimal concentration of SMART adaptor (ultimately around 0.2 μM) was lower than that in conventional SMART RACE with oligo (dT) using intact RNA (9). This finding was contrary to our expectation that a higher concentration of adaptor would be required because of degradation of the RNA extracted from archival formalin-fixed and paraffin-embedded tissue. Since a large amount of SMART adaptor brought in by cDNA used as a PCR template has a GGG sequence at the 3' end, this adaptor
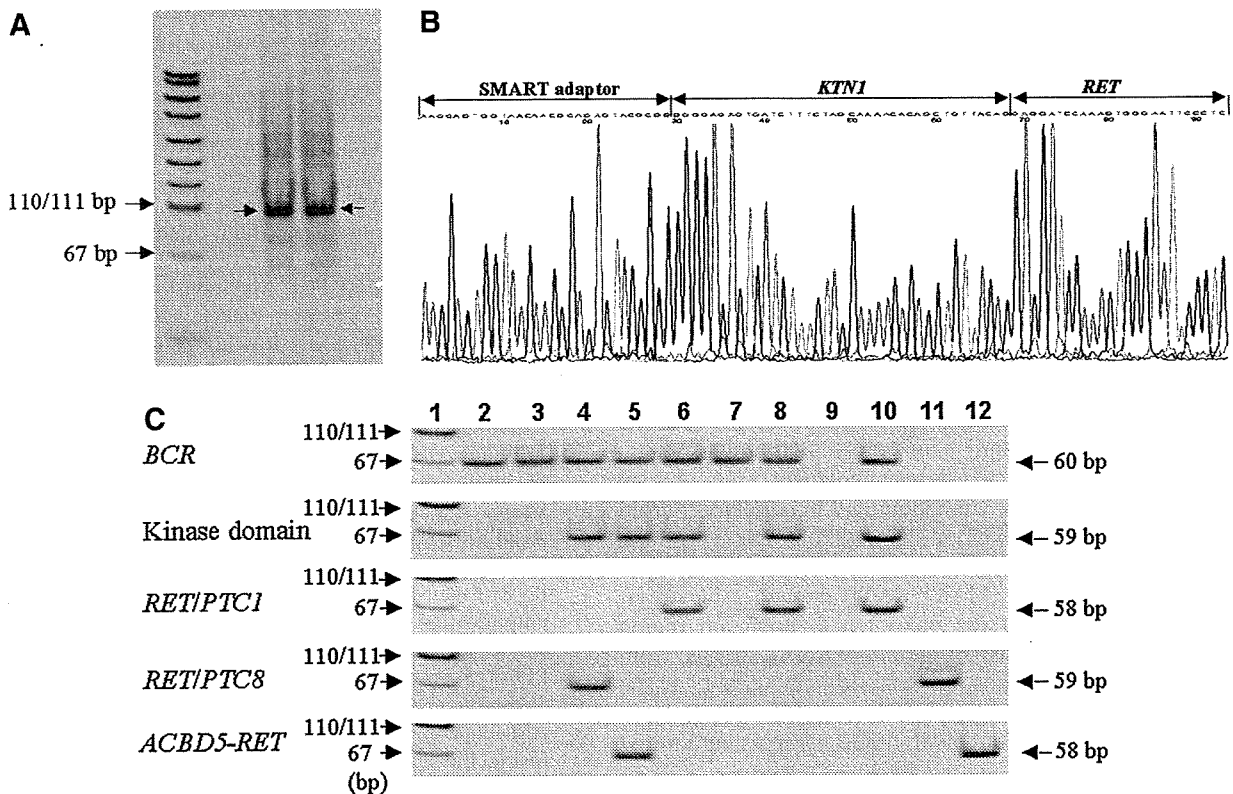

FIG. 3. Identification of RET/PTC8 in papillary thyroid cancer from one A-bomb survivor. (A) Second PCR products of RNA extracted from an archival formalin-fixed, paraffin-embedded papillary thyroid cancer specimen from an A-bomb survivor (arrows indicate the target band in the second PCR products). (B) Sequence of RET/PTC8 isolated by this improved SMART RACE method. (C) Expression of RET/PTC8 (RT-PCR product) in papillary thyroid cancer in the A-bomb survivor. Lanes 2–8 indicate RT-PCR products from RNA in papillary thyroid cancer specimens of A-bomb survivors. Lanes 2, 3, and 7 show no detectable expression of RET tyrosine kinase domain; lane 4, RET/PTC8; lane 5, ACBD5-RET; lanes 6 and 8, RET/PTC1; lane 9, H₂O for negative control; lane 10, cell line TPC1 harboring RET/PTC1; lane 11, genomic DNA plus 77 base-synthesized nucleotides (PTC8-oligo77) for positive control of RET/PTC8; lane 12, genomic DNA plus 77 base-synthesized nucleotides (ACBD-RET-oligo77) for positive control of ACBD5-RET. Lane 1 indicates pUC19-MspI digest for DNA size marker. A-bomb, atomic bomb; RT, reverse transcription.

may cause nonspecific amplification by the first PCR. As a result, specific and efficient RT-PCR may be hindered. Further, when cDNA in the presence of SMART adaptor was prepared with random primers (9 mer) using RNA extracted from archival tissue specimens, we could not obtain the clear and intense target candidate bands observed in the case of addition of adaptor after completion of cDNA synthesis (Fig. 1A and C vs. 1B and D). However, when cDNA was prepared with random primers using intact total RNA, target candidate bands could be detected regardless of the timing of adding SMART adaptor (data not shown). This suggests that when degraded RNA was used, a larger amount of short cDNA primed with SMART adaptor may be produced, possibly interfering with specific PCR amplification.

In addition to the concentration and the timing of the addition of SMART adaptor, length of incubation time was also an important factor for efficient amplification by our SMART RACE-PCR: optimal incubation time was relatively long (45–60 minutes) despite addition of adaptor after cDNA synthesis completion. Thus, although amplification efficiency by SMART RACE-PCR largely depended on concentration of SMART adaptor, timing of adaptor addition, and incubation time when cDNA was prepared with random primer using RNA extracted from archival formalin-fixed and paraffin-embedded tissue specimens, unlike SMART RACE with intact RNA, we found that the improved SMART RACE method achieved successful isolation of 5' upstream fragments in target cDNA with RNA extracted from archival unbuffered formalin-fixed, paraffin-embedded tissue specimens.

A rare rearrangement of the RET gene, RET/PTC8, previously identified in papillary thyroid cancers in children from areas contaminated by the Chernobyl accident (12), was detected in papillary thyroid cancer of one A-bomb survivor exposed to more than 2 Gy of radiation. Further, a novel rearrangement of the RET gene (ACBD5-RET fusion gene) was identified in papillary thyroid cancer of another A-bomb survivor exposed to more than 1.5 Gy. It is interesting to note that a rare RET/PTC8 rearrangement and a novel ACBD5-RET fusion gene were identified in papillary thyroid cancers in A-bomb survivors exposed to a relatively high radiation dose.

Since this improved SMART RACE method achieves effective amplification of the unknown 5' upstream region of certain genes with even 10 ng of total RNA extracted from archival formalin-fixed and paraffin-embedded tissue specimens, the method is expected to prove useful in molecular analyses using archival tissue samples of limited quantity.

Acknowledgments

Disclosure Statement

The authors declare that they have no commercial associations that might create a conflict of interest in connection with this article.

References

1. Mies C 1994 Molecular biological analysis of paraffin-embedded tissues. Hum Pathol 25:550–560.
2. Suárez HG 1998 Genetic alterations in human epithelial thyroid tumors. Clin Endocrinol 48:531–546.
3. Tallini G, Asa SL 2001 RET oncogene activation in papillary thyroid carcinoma. Adv Anat Pathol 8:345–354.
4. Ciampi R, Nikiforov YE 2007 RET/PTC rearrangements and BRAF mutations in thyroid tumorigenesis. Endocrinology 148:936–941.
5. Saenko V, Rogounovitch T, Shimizu-Yoshida Y, Abrosimov A, Lushnikov E, Roumiantsev P, Matsumoto N, Nakashima M, Meirmanov S, Ohtsuru A, Namba H, Tsyb A, Yamashita S 2003 Novel tumorigenic rearrangement, Δrfp/ret, in a papillary thyroid carcinoma from externally irradiated patient. Mutat Res 527:81–90.
6. Arighi E, Borrello MG, Sariola H 2005 RET tyrosine kinase signaling in development and cancer. Cytokine Growth Factor Rev 16:441–467.
7. Ciampi R, Giordano TJ, Wikenheiser-Brokamp K, Koenig RJ, Nikiforov YE 2007 HOOK3-RET: a novel type of RET/PTC rearrangement in papillary thyroid carcinoma. Endocr-Relat Cancer 14:445–452.
8. Chenchik A, Zhu YY, Diatchenko L, Li R, Hill J, Siebert PD 1998 Generation and use of high quality cDNA from small amounts of total RNA by SMART PCR. In: Siebert PD, Larrick JW (eds) RT-PCR Methods for Gene Cloning and Analysis. Eaton Publishing, Natick, MA, pp 305–319.
9. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PM 2001 Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. BioTechniques 30:892–897.
10. Frohman MA, Dush MK, Martin GR 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85:8998–9002.
11. Hamatani K, Eguchi H, Takahashi K, Koyama K, Mukai M, Ito R, Taga M, Yasui W, Nakachi K 2006 Improved RT-PCR amplification for molecular analyses with long-term preserved formalin-fixed, paraffin-embedded tissue specimens. J Histochem Cytochem 54:773–780.
12. Salassidis K, Bruch J, Zitzelsberger H, Lengfelder E, Kellerer AM, Bauchinger M 2000 Translocation t(10;14)(q11.2;q22.1) fusing the kinectin to the RET gene creates a novel rearranged form (PTC8) of the RET proto-oncogene in radiation-induced childhood papillary thyroid carcinoma. Cancer Res 60:2786–2789.

Address correspondence to:
Kiyohiro Hamatani, Ph.D.
Department of Radiobiology/Molecular Epidemiology
Radiation Effects Research Foundation
5-2 Hijiyama Park, Minami-ku
Hiroshima-shi
Hiroshima 732-0815
Japan

E-mail: hamatani@rerf.or.jp

Lung cancer susceptibility among atomic bomb survivors in relation to CA repeat number polymorphism of *epidermal growth factor receptor* gene and radiation dose

Kengo Yoshida*, Kei Nakachi, Kazue Imai, John B.Cologne[1], Yasuharu Niwa, Yoichiro Kusunoki and Tomonori Hayashi

Department of Radiobiology/Molecular Epidemiology and [1]Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami Ward, Hiroshima City 732-0815, Japan

*To whom correspondence should be addressed. Tel: +81 82 261 3131;
Fax: +81 82 261 3170;
Email: kyoshi@rerf.or.jp

Lung cancer is a leading cause of cancer death worldwide. Prevention could be improved by identifying susceptible individuals as well as improving understanding of interactions between genes and etiological environmental agents, including radiation exposure. The epidermal growth factor receptor (EGFR)-signaling pathway, regulating cellular radiation sensitivity, is an oncogenic cascade involved in lung cancer, especially adenocarcinoma. The cytosine adenine (CA) repeat number polymorphism in the first intron of *EGFR* has been shown to be inversely correlated with EGFR production. It is hypothesized that CA repeat number may modulate individual susceptibility to lung cancer. Thus, we carried out a case–cohort study within the Japanese atomic bomb (A-bomb) survivor cohort to evaluate a possible association of CA repeat polymorphism with lung cancer risk in radiation-exposed or negligibly exposed (<5 mGy) A-bomb survivors. First, by dividing study subjects into *Short* and *Long* genotypes, defined as the summed CA repeat number of two alleles ≤37 and ≥38, respectively, we found that the *Short* genotype was significantly associated with an increased risk of lung cancer, specifically adenocarcinoma, among negligibly exposed subjects. Next, we found that prior radiation exposure significantly enhanced lung cancer risk of survivors with the *Long* genotype, whereas the risk for the *Short* genotype did not show any significant increase with radiation dose, resulting in indistinguishable risks between these genotypes at a high radiation dose. Our findings imply that the EGFR pathway plays a crucial role in assessing individual susceptibility to lung adenocarcinoma in relation to radiation exposure.

Introduction

Despite focused research over decades, lung cancer remains a leading cause of cancer deaths worldwide as well as in Japan (1). Effective prevention of lung cancer could be improved by identifying susceptible individuals as well as seeking a comprehensive understanding of lung carcinogenesis and potentially causal environmental agents. It is conceivable that lung cancer risk is influenced by low-penetrance genetic variation, which may contribute to gene–environment interactions (2,3). Causal environmental factors include cigarette smoking and other factors, including exposure to ionizing radiation (4–6). Epidemiological studies on atomic bomb (A-bomb) survivors conducted by the Radiation Effects Research Foundation (RERF) have revealed a radiation dose-dependent increase in incidence of certain cancers, especially lung cancer (7).

Few molecular epidemiology studies have assessed interindividual variation in radiation sensitivity in relation to lung cancer development among A-bomb survivors. Our previous study showed that large

interindividual variation in somatic mutability in response to prior A-bomb irradiation was associated with subsequent development of all cancers combined, in terms of erythrocyte *glycophorin A*-mutant frequency, which suggests the existence of genetic factors involved in radiation sensitivity and cancer susceptibility (8). Considering the implications of those results, we began searching for gene polymorphisms that might be responsible for individual susceptibility to radiation-associated lung cancer.

The initial event in the development of lung adenocarcinoma is thought to be alteration of the epidermal growth factor receptor (*EGFR*) gene (9). Mutations in *EGFR* are frequently observed in patients with lung adenocarcinoma (10,11). Gene amplification of *EGFR* is also frequently observed in patients with adenocarcinoma as well as squamous cell carcinoma of the lung, resulting in enhanced EGFR levels that lead to poor clinical prognosis and a chemo/radio-resistant phenotype (9,12,13). Therefore, the altered EGFR-signaling pathway, which turns on downstream pathways such as Ras-Raf-Mek and PI3K-Akt, is thought to contribute to the development of non-small-cell lung cancer in general through enhanced cell proliferation, inhibition of apoptosis, invasion and metastasis (4,9).

It is noteworthy that EGFR signaling facilitates cellular resistance to radiation exposure and that EGFR and its signaling cascade are induced by radiation exposure even in the absence of ligand binding (12,14,15). EGFR is thought to be a key molecule in the development of lung cancer in the general population as well as that of radiation-associated lung cancer found in A-bomb survivors. We therefore conducted a case–cohort study within a cohort of A-bomb survivors to assess the relationship between a functional polymorphism of *EGFR* and radiation-associated lung cancer.

The 5'-regulatory sequence of *EGFR* contains two functional gene polymorphisms that are markedly associated with the transcriptional activity of this gene: −216G/T and a highly polymorphic microsatellite sequence consisting of cytosine adenine (CA)-dinucleotide repeats (16,17). In this study, we focused on the CA repeat polymorphism in the first intron of *EGFR* because −216G/T variant alleles are infrequent in Asian populations (16,18). The number of CA repeats also shows substantial variation by ethnicity; larger numbers are found in East Asians than in persons of European descent or African-Americans (18–20). As demonstrated by *in vitro* and *in vivo* studies, CA repeat number is inversely correlated with messenger RNA or protein expression (21–23). Consequently, the association between CA repeat number and risk of various cancers has been examined by several groups, although the results on lung cancer have been inconsistent (24,25). This study aims to address: (i) whether *EGFR* CA repeat number is associated with risk of lung cancer development in the Japanese population who had not been exposed to A-bomb irradiation and (ii) how prior exposure to A-bomb irradiation influences the risk of radiation-associated lung cancer among A-bomb survivors in conjunction with the CA repeat polymorphism.

Subjects and methods

Study subjects

RERF's predecessor research organization, the Atomic Bomb Casualty Commission, established two major cohort studies to assess the health effects of A-bomb radiation exposure (26): the Life Span Study started in 1950 with ~120 000 members in Hiroshima and Nagasaki, including 93 000 A-bomb survivors and the Adult Health Study (AHS) including ~23 000 members who were selected from the Life Span Study and received biennial medical examinations beginning in 1958. The Immunology Study began in the AHS cohort in December 1981 with the aim of investigating radiation effects on the immune systems of A-bomb survivors; blood samples to examine immune-related biomarkers were collected from 9385 AHS participants who visited RERF for medical examinations and who donated blood samples for this study in 1981–2006. Additionally, 7131 persons among the 9385 AHS participants

Abbreviations: A-bomb, atomic bomb; AHS, Adult Health Study; CA, cytosine adenine; EGFR, epidermal growth factor receptor; Gy, Gray; PCR, polymerase chain reaction; RERF, Radiation Effects Research Foundation; RR, relative risk.

donated blood samples (peripheral lymphocytes and/or blood absorbed on paper disks) from which DNA could be extracted.

A total of 4764 participants were selected for the Immunogenome cohort study to assess relationships between cancer development and gene polymorphisms among A-bomb survivors, focusing on immune-related genes. The inclusion criteria were as follows: age <80 years at the time of blood collection; radiation dose information available; no prior cancer diagnosis at the time of blood collection and informed consent given for extracting and using DNA for research purposes (living members) or approval of the RERF Ethics Committee for Genome Research (deceased members who died prior to giving informed consent). Among the 4764 cohort members, 1061 incident cancers were identified from the Hiroshima and Nagasaki Tumor Registries, diagnosed between 1981 and 2001. Those included cancers of the stomach ($n = 227$), colon ($n = 165$), rectum ($n = 53$), liver ($n = 115$), lung ($n = 124$), breast ($n = 90$) and thyroid ($n = 47$).

Within the cohort, we defined a sub-cohort for this case–cohort study consisting of 2160 members who were randomly selected, a sampling rate of 0.45. A total of 486 incident cancers, including 62 lung cancer cases, were included in this sub-cohort. Cases were all 124 members who were diagnosed with lung cancer between 1981 and 2001. Their time of entry into the cohort was the year when the blood donation for the Immunology Study was first made. By histological type of lung cancer, the cases consisted of 66 (52.4%) adenocarcinomas, 21 (16.7%) squamous cell carcinomas, 10 (7.9%) small cell carcinomas and 29 (23.0%) other types including large cell carcinomas. In addition, two multiple cancer cases were included; those histological types were counted individually. This study was approved by the RERF Ethics Committee for Genome Research.

Determination of CA repeat number

The number of polymorphic CA repeats in the first intron of *EGFR* was determined by polymerase chain reaction (PCR)-based fragment length analysis using fluorescent primers along with DNA direct sequencing (18,24). First, genomic DNA was extracted from peripheral blood cells with proteinase K digestion and a QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) and subjected to whole genome amplification (GenomiPhi DNA Amplification Kit, GE Healthcare, Little Chalfont, Buckinghamshire, UK). Then, 50–200 ng of the amplified DNA was used for PCR with both carboxyfluorescein fluorochrome-labeled and unlabeled forward primers (5′-GGGCTCACAGCAAACTTCTC-3′) and unlabeled reverse primers (5′-AAGCCAGACTCGCTCATGTT-3′). The 10 µl PCR reaction mixture contained 10× PCR buffer (Sigma–Aldrich, St Louis, MO), 0.5 U Taq DNA polymerase (Sigma–Aldrich), 0.2 mM each of deoxynucleoside triphosphate, 2 mM MgCl₂, 200 nM each unlabeled primer and 20 nM labeled primer. PCR cycle conditions consisted of an initial denaturation step at 94°C for 5 min followed by 40 cycles of 30 s at 94°C, 30 s at 65°C and 60 s at 72°C, with a final elongation step at 72°C for 5 min. After 1 µl of PCR product and 0.5 µl Genescan 500 ROX molecular weight standard (Applied Biosystems, Foster City, CA) were denatured in 10 µl deionized formamide, the number of CA repeats was determined using an ABI 3100 genetic analyzer (Applied Biosystems) and GeneMapper software (version 3.0, Applied Biosystems). The exact number of CA repeats was verified by DNA direct sequencing with 16 and 18 repeat homozygote samples.

Statistical analysis

The data were sampled according to the case–cohort design, which does not require a rare disease assumption (27). The unweighted case–cohort approach was used for analysis (28). Relative risks for cancer incidence were estimated using the Cox proportional hazard model in terms of either follow-up time (years) or age as the underlying time axis. Analyses were performed with SPSS (version 14.0, SPSS, Chicago, IL).

All models included adjustment for age at the time of blood collection, gender, city (Hiroshima versus Nagasaki), smoking status (number of cigarettes per day) and radiation dose. Information on smoking was collected at the time of blood collection. A-bomb radiation dose in Gray (Gy) was estimated using the DS02 dosimetry system (29), based on weighted skin dose computed as the gamma dose plus 10 times the neutron dose. A radiation dose <0.005 Gy was called non-exposed when performing analyses based on dose group.

Results

Table I shows characteristics of cases and sub-cohort members. Cases evidenced higher proportions of males, smokers and persons with the highest radiation doses compared with sub-cohort members.

PCR-based identification of CA repeat number revealed a frequency distribution ranging between 9 and 24, with a bimodal pattern having peaks at 16 (16.6%) and 20 (61.9%) among non-exposed members in

Table I. Characteristics of the study population within the RERF AHS cohort

	Cases	Sub-cohort
Total[a]	124 (100)	2160 (100)
Age at entry[b]	59 (48–74)	56 (41–75)
Gender[a]		
Men	68 (54.8)	779 (36.1)
Women	56 (45.2)	1381 (63.9)
City[a]		
Hiroshima	84 (67.7)	1432 (66.3)
Nagasaki	40 (32.3)	728 (33.7)
Smoking status[a]		
Non-smoker	49 (39.5)	1571 (72.7)
Smoker	64 (51.6)	511 (23.6)
Unknown	11 (8.9)	78 (3.6)
Radiation dose[a]		
<5 mGy	41 (33.1)	906 (41.9)
5–712[c]	37 (29.8)	627 (29.0)
≥712	46 (37.1)	627 (29.0)
Histological type[d]		
Adenocarcinoma	66 (52.4)	
Squamous cell carcinoma	21 (16.7)	
Small cell carcinoma	10 (7.9)	
Other types	29 (23.0)	

[a]Number (%).
[b]Median (5–95% percentiles).
[c]712 mGy: median dose in exposed sub-cohort members.
[d]Two multiple primary cancer cases are included (one is adenocarcinoma and adenosquamous carcinoma, the other is adenocarcinoma and small cell carcinoma). Those histological types are counted individually.

the sub-cohort (Figure 1A). The allele distribution showed close agreement with those from Asian populations as previously reported (19,24). The allele distributions of lung cancer cases and sub-cohort members were compared for the radiation non-exposed population (Figure 1A) or radiation exposed population (Figure 1B). In the non-exposed population, a relative frequency of 16 CA repeats or 20 CA repeats in cases was higher or lower, respectively, than that in sub-cohort members (Figure 1A). On the other hand, there was no substantial difference in relative frequency of 16 CA repeats or 20 CA repeats between exposed cases and exposed sub-cohort members (Figure 1B). In contrast to the differences in cases by radiation exposure status, no distinguishable difference was found between the non-exposed and exposed sub-cohort members.
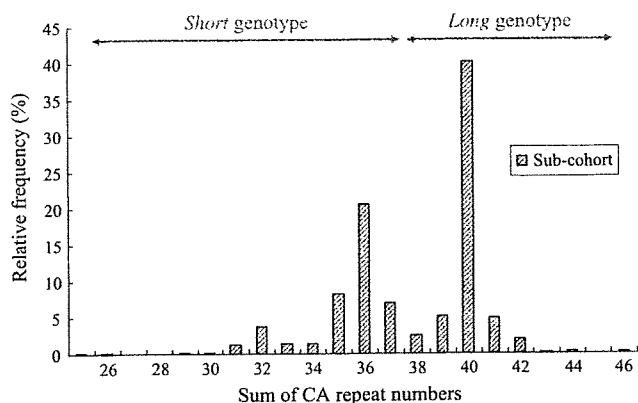
In the analysis to calculate lung cancer risk, we used two methods to genotype the CA repeat polymorphism according to previous studies (25,30): method 1 sums repeat numbers of two alleles in each individual and defines his or her genotype as *Long* (sum ≥38) or *Short* (sum ≤37) (Figure 2); method 2 defines the two alleles separately as *long* (allele repeat number ≥18) or *short* (repeat number ≤17) and combines them for each individual as *long/long, long/short* or *short/short*. We report the results from method 1 genotyping in the text (for results from method 2 genotyping, refer supplementary Tables SI–SIII, available at *Carcinogenesis* Online).

As a preliminary to risk estimation, we confirmed that age, gender, city, smoking status and radiation dose did not influence the distribution of the CA repeat number in the sub-cohort population (data not shown). First, overall relative risk (RR) of lung cancer for the *Short* genotype in all subjects was evaluated by Cox regression analysis with adjustment for age, gender, city, smoking and radiation dose, taking the *Long* genotype as a reference: a significant increase in risk was found for the *Short* genotype (RR: 1.79, 95% CI: 1.14–2.82, Table II).

Next, we stratified the subjects by radiation dose (<5, 5–712 and ≥712 mGy) and evaluated the relative risks for the genotypes using two models: risk ratios for the genotypes within the same dose group (without considering radiation effects on risk) and relative risks for
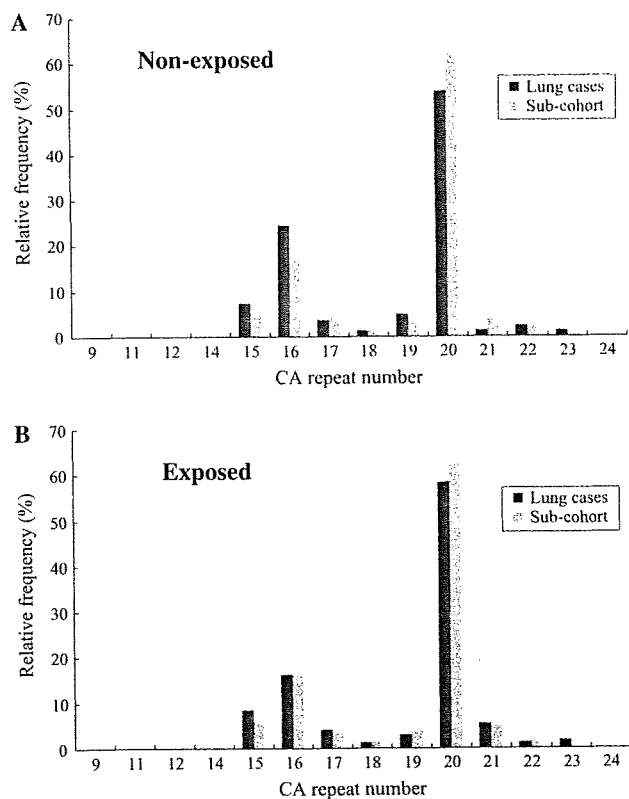
Fig. 1. Relative allele frequencies of the *EGFR* CA repeat polymorphism. (A) Radiation non-exposed (<5 mGy) lung cancer cases (*N* = 82 chromosomes) and sub-cohort members (*N* = 1812). (B) Radiation-exposed (≥5 mGy) lung cancer cases (*N* = 166) and sub-cohort members (*N* = 2 508).

Fig. 2. Relative frequencies of summed CA repeat numbers in each individual in radiation non-exposed (<5m Gy) sub-cohort members; *N* = 906 (individuals). Genotype definition in the genotyping method 1 is indicated at the top of the graph.

Table II. Association between CA repeat polymorphism and lung cancer risk

Genotype	N[a], cases/sub-cohort	RR[b]	95% CI[c]
Long	57/1176	1.0[d]	
Short	56/890	1.79	1.14–2.82

[a]Eleven cases and 78 sub-cohort members were excluded due to lack of information about an adjustment factor.
[b]RR: relative risks adjusted for age, gender, city, smoking and radiation dose.
[c]95% confidence interval.
[d]The genotype serves as reference category.

Table III. CA repeat polymorphism and lung cancer risk among radiation dose-stratified populations

Genotype	N[a], cases/sub-cohort	RR1[b]	95% CI[c]	RR2[d]	95% CI[c]
Non-exposed					
Long	15/488	1.0[e]		1.0[e]	
Short	22/389	2.43	1.23–4.79	2.11	1.09–4.08
5–712 mGy[f]					
Long	18/343	1.0[e]		1.76	0.88–3.53
Short	17/256	1.27	0.65–2.47	2.24	1.11–4.49
≥712 mGy[f]					
Long	24/345	1.0[e]		2.60	1.36–4.96
Short	17/245	0.92	0.49–1.72	2.45	1.22–4.92

[a]Eleven cases and 78 sub-cohort members were excluded due to lack of information about an adjustment factor.
[b]RR1: genetic relative risks, *Short* versus *Long* genotype, adjusted for age, gender, city and smoking.
[c]95% confidence interval
[d]RR2: genetic and radiation-related risks, with reference to non-exposed *Long* genotype.
[e]Reference category.
[f]712 mGy: median dose in exposed sub-cohort members.

95%CI: 0.88–3.53 for 5–712 mGy and RR: 2.60, 95%CI: 1.36–4.96 for ≥712 mGy), whereas that for the *Short* genotype remained unchanged (RR: 2.11, 95%CI: 1.09–4.08 for <5 mGy and RR: 2.45, 95%CI: 1.22–4.92 for ≥712 mGy) independently of radiation exposure status or dose, resulting in almost identical relative risks for both genotypes at the highest dose.

Finally, we extended the analysis to histological type of lung cancer (adenocarcinoma or squamous cell carcinoma). With lung adenocarcinoma, there was a significant increase in risk for the *Short* genotype only in the non-exposed population (RR: 3.33, 95% CI: 1.23–9.02, Table IV) as well as a radiation-associated elevation in risk for the *Long* genotype (RR: 4.37, 95% CI: 1.72–11.1 for ≥712 mGy). With lung squamous cell carcinoma, there were no significant risk differences among these genotypes (data not shown).

Discussion

We investigated the association between CA repeat number polymorphism in the *EGFR* gene and lung cancer risk using a case–cohort study setting, with the aim of elucidating risk of radiation-associated lung cancer in individuals with different *EGFR* genotypes. In the radiation dose-stratified analysis, we found that the *Short* genotype (the smaller sums of CA repeat numbers of two alleles) was significantly associated with an increased risk of lung cancer in survivors who had not been exposed to radiation (Table III). When we further examined the risk of lung cancer by histological types, adenocarcinoma, but not squamous cell carcinoma, showed a significant association between CA repeat genotypes (*Long* and *Short*) and lung cancer risk (Table IV). That is consistent with the clinico-pathological observation that adenocarcinoma is the most frequent histology in lung cancer of non-smokers, with frequent *EGFR* mutations detected

the genotypes as functions of radiation dose (*Long* genotype in the non-exposed population used as a reference). When looking at risk ratios (RR1 in Table III), the *Short* genotype revealed a significantly increased risk of lung cancer in the non-exposed population (RR: 2.43, 95% CI: 1.23–4.79), but this genetic effect on risk disappeared in the exposed population. Relative risk for genotype combined with radiation dose (RR2 in Table III) showed that risk for the *Long* genotype increased with increasing radiation dose (RR: 1.76,