**Table 5.** Distribution of epithelial tumors by sex and age group (years)

| | 0–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90–99 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Male | 3[a] | 8 | 38 | 98 | 225 | 298 | 288 | 79 | 14 | 1051 |
| Female | 0 | 12 | 25 | 44 | 98 | 173 | 229 | 119 | 26 | 726 |
| Total | 3 | 20 | 63 | 142 | 323 | 471 | 517 | 198 | 40 | 1777 |

[a]Numbers show numbers of patients

**Table 6.** Distribution of nonepithelial tumors by sex and age group (years)

| | 0–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90–99 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Male | 0[a] | 0 | 1 | 1 | 7 | 4 | 4 | 2 | 1 | 20 |
| Female | 0 | 2 | 3 | 1 | 0 | 3 | 3 | 0 | 0 | 12 |
| Total | 0 | 2 | 4 | 2 | 7 | 7 | 7 | 2 | 1 | 32 |

[a]Numbers show numbers of patients

**Table 7.** Distribution of tumor location by sex

| | Lip | Tongue | Gingiva | Oral floor | Palate | Buccal mucosa | Maxillary sinus | Jawbone | Oral cavity | Major salivary gland | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Male | 13/0 | 468/1[a] | 281/10 | 121/0 | 19/3 | 99/0 | 37/3 | 4/3 | 2/0 | 10/0 | 1054/20 |
| Female | 5/0 | 261/0 | 295/8 | 42/1 | 23/2 | 85/0 | 10/0 | 1/0 | 1/1 | 7/0 | 730/12 |
| Total | 18/0 | 729/1 | 576/18 | 163/1 | 42/5 | 184/0 | 47/3 | 5/3 | 3/1 | 17/0 | 1784/32 |

[a]Epithelial tumors / Nonepithelial tumors

**Table 8.** Average age at the first consultation, by tumor location

| Location | Lip | Tongue | Gingiva | Oral floor | Palate | Buccal mucosa | Maxillarysinus | Jawbone | Oral cavity | Major salivary gland |
|---|---|---|---|---|---|---|---|---|---|---|
| Average age (years) | 69.6 | 62.0 | 68.1 | 64.6 | 69.3 | 69.1 | 64.1 | 43.8 | 62.5 | 57.9 |
| Number of tumors | 18 | 730 | 594 | 164 | 47 | 184 | 50 | 8 | 4 | 17 |

**Table 9.** Number of epithelial tumors by location

| Location | Lip (18) | | | Tongue (729) | | | | Gingiva (576) | | Oral floor (163) | | | Palate (42) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Upper lip | Lower lip | Labial commisure | Dorsal tongue | Lateral tongue | Ventral tongue | Anterior 2/3 | Upper gingiva | Lower gingiva | Anterior | Lateral | Not specified | Hard palate |
| | 2 | 9 | 7 | 17 | 579 | 101 | 32 | 214 | 362 | 93 | 52 | 18 | 42 |

| Buccal mucosa (184) | | | | | Others (4) | Jawbone and sinus (51) | | | | Major salivary gland (17) | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Buccal mucosa | Upper lip | Lower lip | Gingiv-buccal sulcus | Retro molar | Oral cavity | Maxillary sinus | Maxilla | Intra osseous (mandible) | Mandible | Parotid | Subman-dibular | Sublingual | |
| 128 | 7 | 4 | 23 | 22 | 4 | 46 | 1 | 3 | 1 | 9 | 6 | 2 | 1784 |

Age, sex, and primary tumor location for epithelial and nonepithelial tumor cases

*Epithelial tumor cases*

Epithelial tumors were observed in 1777 patients (98.2%), 1051 males (59.1%) and 726 females (40.1%; male: female ratio 1.45:1; Table 5). The average age at the time of the first consultation was 65.3 ± 13.9 years (range, 12-98 years). The age of female patients was significantly higher (67.8 ± 14.3 years) than that of male patients (63.6 ± 13.1 years; $P < 0.05$). Primary tumor locations in these patients showed a tendency similar to that in all of the patients investigated, as tumor incidence was higher in order of the tongue ($n$ = 729; 40.9%), gingiva ($n$ = 576; 32.4%), buccal mucosa ($n$ = 184; 10.3%), and oral floor ($n$ = 163; 9.1%; Table 9).

As for the histopathological diagnosis, squamous cell carcinoma Not otherwise specified (NOS) (8070/3) accounted for the largest proportion ($n$ = 1282; 71.9%) of the 1784 epithelial tumors studied. However, when 329 tumors

**Table 10.** Histopathological diagnosis of epithelial tumors

| Histopathological diagnosis | No. of tumors |
|---|---|
| 8000/3 (Neoplasm, malignant) | 13 |
| 8010/3 (Carcinoma, NOS) | 5 |
| 8020/3 (Carcinoma, undifferentiated, NOS) | 2 |
| 8051/3 (Verrucous carcinoma, NOS) | 52 |
| 8052/3 (Papillary squamous cell carcinoma) | 4 |
| 8070/2 (Squamous cell carcinoma in situ, NOS) | 27 |
| 8070/3 (Squamous cell carcinoma, NOS) | 1282 |
| 8071/3 (Squamous cell carcinoma, keratinizing, NOS) | 273 |
| 8072/3( Squamous cell carcinoma, large cell, nonkeratinizing) | 11 |
| 8073/3 (Squamous cell carcinoma, small cell, nonkeratinizing) | 7 |
| 8074/3 (Squamous cell carcinoma, spindle cell) | 5 |
| 8076/2 (Squamous cell carcinoma in situ with questionable stromal invasion) | 1 |
| 8076/3 (Squamous cell carcinoma, microinvasive) | 1 |
| 8140/3 (Adenocarcinoma, NOS) | 19 |
| 8200/3 (Adenoid cystic carcinoma) | 39 |
| 8430/3 (Mucoepidermoid carcinoma) | 31 |
| 8470/3 (Mucinous cystadenocarcinoma, NOS) | 2 |
| 8550/3 (Acinar cell carcinoma) | 3 |
| 8560/3 (Adenosquamous carcinoma) | 2 |
| 8940/3 (Mixed tumor, malignant, NOS) | 2 |
| 8980/3 (Carcinosarcoma, NOS) | 1 |
| 8982/3 (Myoepithelial carcinoma, NOS) | 1 |
| 9270/3 (Odontogenic tumor, malignant) | 1 |

**Table 11.** TN Classification of epithelial tumors

| | N0 | N1 | N2a | N2b | N2c | N3 | NX | Total |
|---|---|---|---|---|---|---|---|---|
| T0 | 1[a] | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Tis | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| T1 | 381 | 16 | 2 | 3 | 4 | 0 | 4 | 410 |
| T2 | 572 | 136 | 8 | 29 | 15 | 0 | 5 | 765 |
| T3 | 126 | 43 | 6 | 41 | 13 | 1 | 0 | 230 |
| T4 | 142 | 79 | 18 | 72 | 44 | 9 | 7 | 371 |
| TX | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| Total | 1228 | 274 | 34 | 145 | 76 | 10 | 17 | 1784 |

[a] Numbers show numbers of tumors

**Table 12.** TN classification of M1 cases

| | T2 | T3 | T4 | Total |
|---|---|---|---|---|
| N0 | 2[a] | 1 | 2 | 5 |
| N1 | 1 | 2 | 1 | 4 |
| N2a–c | 1 | 1 | 2 | 4 |
| N3 | 0 | 0 | 4 | 4 |
| Total | 4 | 4 | 9 | 17 |

[a] Numbers show numbers of tumors

**Table 13.** Distribution of T classification of epithelial tumors by location

| | T0 | Tis | T1 | T2 | T3 | T4 | TX | Total |
|---|---|---|---|---|---|---|---|---|
| Lip | 0[a] | 0 | 11 | 6 | 1 | 0 | 0 | 18 |
| Tongue | 0 | 2 | 245 | 339 | 93 | 50 | 0 | 729 |
| Gingiva | 0 | 1 | 73 | 228 | 66 | 208 | 0 | 576 |
| Oral floor | 0 | 0 | 35 | 82 | 16 | 30 | 0 | 163 |
| Palate | 0 | 0 | 12 | 15 | 1 | 13 | 1 | 42 |
| Buccal mucosa | 1 | 1 | 33 | 82 | 30 | 37 | 0 | 184 |
| Maxillary sinus | 0 | 0 | 1 | 6 | 17 | 22 | 0 | 46 |
| Jawbone | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 5 |
| Oral cavity | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 4 |
| Major salivary gland | 0 | 0 | 0 | 6 | 4 | 7 | 0 | 17 |
| Total | 1 | 4 | 410 | 765 | 230 | 371 | 3 | 1784 |

[a] Numbers show numbers of tumors

of other types (8052/3, 8070/2,8071/3, 8072/3, 8073/3, 8074/3, 8076/2, and 8076/3) were included in the analysis, 1611 tumors were squamous cell carcinoma (90.3%). Among salivary gland tumors, adenoid cystic carcinoma (8200/3) had the highest incidence ($n$ = 39; 39.4%) followed by mucoepidermoid carcinoma (8430/3) ($n$ = 31; 31.3%) and adenocarcinoma (8140/3) ($n$ = 19; 19.2%; Table 10).

As for T classification, T2 was predominant ($n$ = 765; 42.9%), followed by T1 ($n$ = 410; 23.0%), T4 ($n$ = 371; 20.8%), and T3 ($n$ = 230; 12.9%). For N classification, N0 was predominant ($n$ = 1228; 68.8% 69.1), followed by N1 ($n$ = 274; 15.4%), N2 ($n$ = 255; 14.3%), and N3 ($n$ = 10; 0.6%). For TN classification, T2N0 was the highest ($n$ = 572; 32.1%), followed by T1N0 ($n$ = 381; 21.4%), T4N0 ($n$ = 142; 8.0%), and T2N1 ($n$ = 136; 7.6%; Table 11). For M classification, M1 was observed in 17 cases (1.0%), which were classified by TN classification as follows: 4 cases of T2, 4 cases of T3, and 9 cases of T4. N classification does not contribute to M classification (Table 12). As for histopathological diagnosis of the M1 cases, squamous cell carcinoma NOS (8070/3) was predominant in 11 cases, followed by squamous cell carcinoma keratinizing NOS (8017/3) in 2 cases, and verrucous carcinoma NOS (8051/3), adenocarcinoma NOS (8140/3), adenoid cystic carcinoma (8200/3), and adenosquamous carcinoma (8560/3) in 1 case each.

In a comparison of T classification by location, the initial stages T0, Tis, T1, and T2 accounted for the majority of the tumors in the lip, tongue, palate, and buccal mucosa, while they accounted for half of those in the gingiva. In contrast, advanced stage cases accounted for the majority of tumors located in the maxillary sinus, jawbone, and major salivary gland (Table 13).

*Nonepithelial tumors* Thirty-two nonepithelial tumors were observed in 32 patients (1.8%), none of whom had multiple tumors. Twenty of these patients (62.5%) were men and 12 (37.5%) were women (Table 4). Their average age at the first consultation was 59.2 ± 16.5 years (range, 27–92 years). Further, the average age of male patients was 63.0 ± 14.0 years and that of females was 52.9 ± 18.9 years.

The location of the primary site was in the maxillary gingiva for 10 tumors (31.3%), while the maxilla had the majority ($n$ = 19; 59.4%) when 4 tumors (12.5%) found in the maxillary sinus and 5 (15.6%) in the hard palate were added. The mandibular gingiva followed, with 8 tumors (25.0%; Table 14). As for the histopathological diagnosis, malignant melanoma (8720/3) accounted for the largest proportion ($n$ = 13; 40.6%), followed by malignant lymphoma (9590/3, 9670/3, 9680/3) ($n$ = 8; 25.0%; Table 15). The size of the tumors ranged from 1 to 7 cm, and averaged 3.7 cm (median, 4 cm). None of the nonepithelial tumors was described by M classification as M1.

**Table 14.** Number of nonepithelial tumors by location

| Location | Tongue (1) | Gingival (18) | | Oral floor (1) | Palate (5) | Jawbone and maxillary sinus (7) | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | Lateral tongue | Maxillary gingiva | Mandibular gingiva | Anterior | Hard palate | Maxillary sinus | Intraosseous | Mandible | |
| | 1 | 10 | 8 | 1 | 5 | 4 | 2 | 1 | 32 |

**Table 15.** Histopathological diagnosis of nonepithelial tumors

| MMe | FS | MFH | LS | AS | OS | MCS | ML | MMy | PCl | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 1 | 2 | 2 | 1 | 1 | 1 | 8 | 2 | 1 | 32 |

Mme, malignant melanoma; FS, fibrosarcoma; MFH, malignant fibrous histiocytoma; LS, leiomyosarcoma; AS, angiosarcoma; OS, osteosarcoma; MCS, mesenchymal chondrosarcoma; ML, malignant lymphoma; MMy, multiple myeloma; PC, plasmacytoma

## Discussion

The incidence of cancer in the oral and maxillofacial region varies greatly depending on country, region, and ethnic group.[3] Even within the same country and region, incidence by location of primary tumors varies over time.[4] In Japan, a nationwide survey[1] (hereinafter referred to as the 1986 National Survey) was made of 1508 patients treated at 109 oral surgery institutions in fiscal 1986, 16 years prior to the present survey. In addition, the registration of malignant head and neck tumors (hereinafter referred to as the Nara Prefecture Survey) was conduced as a regional cancer registration in Nara Prefecture.[5] In the 1986 National Survey, patient occupation, route of referral, family history of cancer, and comorbidity were investigated.

The present investigation was conducted using the registration form for Head and Neck Cancer (for initial registration) prepared by the Japan Society for Head and Neck Cancer. In this survey, social background and previous diseases were not investigated. As in the 1986 National Survey, the present subjects were patients who had a checkup at a member institution during the 1-year period from January to December 2002, and the survey did not include those who consulted other institutions. Further, 319 registration forms were excluded due to incomplete descriptions. Accordingly, the results of this survey did not cover all oral cancer cases in Japan for the year of study. However, we consider that the findings provide a description of patients who were diagnosed and treated by oral surgeons.

### Participating institutions

In comparison to the 1986 National Survey, the number of institutions that participated in the present survey was increased, as registration forms were returned by 148 institutions. Further, participation by departments of dentistry and oral surgery in general hospitals increased, though the number of patients per institution varied widely, from 1 to 59. Characteristically, dental colleges and medical colleges reported nearly the same number of patients, while general hospitals had fewer. We consider that this might be because of the recent increases in numbers of supervisory and specialist oral surgeons, and increases in the number of institutions, as society members, that deal with oral cancers.

### Age and sex

The incidence of cancer in the oral and maxillofacial region was more frequent in older patients. It has been frequently reported that the average age at the first consultation was in the 60s for both men and women,[4,6–8] though other reports have noted that cancer cases were more frequently observed in younger patients, in their 40s and 50s.[9,10] However, in a study that compared age over time, it was generally concluded that the number of older patients was increasing.[4,9]

In the 1986 National Survey, patients in their 50s and 60s accounted for the largest proportions, at 25% each, followed by those in their 70s. In the present survey, patients in their 70s were predominant, accounting for 29.0%, indicating an increasing tendency for elderly patients. Especially in women, the number of patients in their 80s was larger than that of those in their 50s, showing an apparent aging trend. In Japan, in patients with cancer in all locations, the ratio of elderly patients (over 70 years old) is tending to increase; it was 39% in 1985 and had increased to 47% in 2000, while it was expected that this proportion would increase to 52% in 2005.[11] As for oral cancer, the proportion of patients over 70 years old was 29.3% in the 1986 National Survey and 42.3% in the present investigation. Thus, it was confirmed that oral cancer is increasing among elderly individuals, similarly to other types of cancer. However, a variety of diseases often coexist in elderly cancer patients, which makes standard treatment difficult and necessitates individualized therapy choices.

Shiboski et al.[12] reported that the proportion of younger patients aged 20–44 years with cancer of the tongue and palatine tonsil was on the increase in the United States. In Scandinavian countries, it was also noted that the number of younger patients with cancer of the tongue showed an increasing tendency.[13] In Japan, patients aged 49 years and younger accounted for 18.1% of the total in the 1986 National Survey, while this age group accounted for 13.0% in the present survey. Further, the number of patients aged 45 years or younger in the present survey amounted to 142 (7.8%), which was a slight decrease from the previous survey. Llewellyn et al.[14] reported that the risk factors for oral cancer in patients aged 45 years or younger were smoking that started at a young age (16 years or younger) and excessive alcohol intake, especially for males, while the

risk factors for females were reported to be hereditary predisposition, virus infection, and sexual activity.[15] In the present survey, no investigation was made regarding habits such as smoking and drinking, though these factors should be examined in a future survey.

Tarvainen et al.[4] reported that age at the first consultation was slightly higher in patients with lip cancer than in those with cancer at other sites of the oral cavity, while Hogan et al.[8] found that age was higher, in the descending order, for cancer of the buccal mucosa, lip, and gingiva. Further, Howell et al.[16] noted that the proportion of patients aged 70 years or older was highest for cancer of the gingiva, followed by lip cancer. In the present survey, the proportion of patients aged 70 years or older was higher, in descending order, for cancer of the lip, buccal mucosa, and palate, confirming that, for patients with lip cancer, the average age at the first consultation was high. However, it was impossible to determine the cause of the high age at first consultation in patients with lip cancer on the basis of the present results. A previous report[4] suggested that there were relationships between the incidence of lip cancer and exposure to ultraviolet light and smoking rate, features which should be examined in a future study.

In the 1986 National Survey, the ratio of males to females among patients with cancer in the oral and maxillofacial region was 1.67:1. However, this ratio varies widely; Iamaroon et al.[6] reported it to be 1.3:1, and Rawashdeh et al.,[7] Ravi et al.,[10] and Hogan et al.[8] reported ratios of 1.77:1, 3.27:1, and 1.73:1, respectively. Shimizu[17] compared survey results for the period from 1930 to 1960 with those of the 1986 National Survey and noted that the ratio of males tended to be higher in the latter survey. Meanwhile, Howell et al.[16] found that the ratio of female patients was increasing over time. In the present survey, the ratio of males to females was 1.45:1, which indicated a slightly higher ratio of female patients as compared to that in the 1986 National Survey. The cause of this increase is not known, though we consider that the following factors may be involved. The age distribution in the present survey, with elderly patients aged 70 years or older accounting for 40% of the subjects, was slightly higher than in the 1986 National Survey. In addition, the proportion of female patients was also higher, and it has been shown, in a national nutrition survey[18] conducted by the Ministry of Health, Welfare and Labor of Japan, that the smoking rate among females has increased slightly, while that of males has decreased. It is considered necessary to continue this type of survey over time.

For the ratio of males to females by tumor location, there have been reports that the male ratio was higher among patients with lip cancer,[4] whereas that of females tended to be higher among those with cancer of the gingiva and salivary gland as compared to patients with cancer at other sites.[16] However, in Japan, cancer of the gingiva, especially the lower gingiva, is frequently seen in male patients.[19] In the present survey, cancer of the lip tended to occur more frequently in males, while that of the gingiva occurred more frequently in females. The factors that affect the ratio of males to females by tumor location are considered to be wide-ranging, and the present results could not elucidate those details.

## Primary location

The incidence of oral cancer by primary location varies depending on the geographic region. Some reports have noted the tongue[6,8,10] as the primary site, while others note the lip[4,16] and buccal mucosa.[9,20] Tarvainen et al.[4] reported that the incidence of lip cancer in males had declined recently, and cited decreases in the smoking rate and in the number of outdoor workers as contributing factors. It was also reported that in regions where the incidence of cancer in the buccal mucosa was frequent, such as Karachi, Pakistan, the disease might be caused by the consumption of local foods such as betel quid, and that the incidence showed an increasing tendency in both males and females, especially among younger individuals.[9] In Japan, cancer of the tongue was reported to be the most frequent type of oral cancer in both the 1986 National Survey and the Nara Prefecture Survey. In the present survey, cancer of the tongue had the highest incidence ($n = 730$:40.2%).

Kawabe et al.[21] reported that 29 of 200 cases of oral cancer developed into multiple cancer, while Kawakami et al.[22] found that of 164 patients with cancer in the oral and maxillofacial region, 38 (23.2%) had multiple primary cancers and 12 had multiple cancers. Further, Saikawa and Ebihara[23] reported that, of 984 patients with oral cancer, 5 demonstrated simultaneous double cancers, with both the first and second tumors located in the oral cavity. In the 1986 National Survey, it was shown that there were 13 cases (0.9%) of multiple cancers in more than two locations. In the present survey, multiple cancers were found in 6 patients (0.3%), similar to the low incidence found in the 1986 National Survey. However, it is likely that metachronous multiple cancer was not included, as that survey period was only 1 year. In the present survey, of the patients with multiple cancers, 1 patient was aged 60 years and the others were older than 75. We consider it necessary to diagnose and treat elderly patients by taking multiple cancers into consideration.

## Histopathological diagnosis

In the histopathological diagnosis of cancer in the oral and maxillofacial region, it has been frequently reported that squamous cell carcinoma was predominant, followed by salivary gland carcinomas and malignant lymphoma consisting mainly of adenoid cystic carcinomas.[1,5,7,8,16,24,25] In the present survey, squamous cell carcinoma was the major type, at 90.3%. For other types of cancer, salivary gland tumors originating in the minor salivary gland were frequently observed, though the incidence of salivary gland tumors was low in comparison to that in other reports. As for the reason for this low incidence, we considered that, in some of the cases with tumor location not noted in the registration form, the tumors were actually located in the salivary gland.

### TNM classification

The number of cases by TNM classification and stage classification varies widely depending on the country, region, and institution. In Japan, cases classified as T2 and stage II are the most frequent.[1,5,26] However, the majority of patients in developing countries and regions are reported to be those with advanced cancer.[6,9,25,27] Carvalho et al.[28] compared the clinical features of patients with cancer in the oral and maxillofacial region between patients in developing and developed countries, and reported that the advanced type of cancer was more frequent in the former. Scott et al.[29] compared the incidence of early-stage and advanced-stage cancer in a United States population, and reported that early-stage cancer was significantly more frequent among females and married individuals, whereas advanced-stage cancer was more frequent in the nonwhite population. In the present study, no investigation was made regarding social factors such as occupation, home town, and marital status. In future investigations, to aid in earlier detection, we intend to take a wider variety of factors into consideration.

As for TN classification in the present survey, T2N0 cases were predominant, followed by T1N0 cases; thus, so-called early cancer, such as stage I and stage II, accounted for the majority. In comparison to the 1986 National Survey results, the proportion of T1 and T2 cases had increased slightly, with the proportion of T2 cases the greatest, while that of T3 and T4 cases had decreased. Similar to findings for tongue cancer cases reported in the Nara Prefecture Survey, the number of cases that were diagnosed at a relatively early stage tended to increase. However, the incidence of T3 and T4 cases together was greater than 30%, and when N and M factors were taken into account, advanced cancer cases accounted for nearly half of the cases in the present survey. Regarding N factor, in the present survey, the proportion of N0 and N2 cases had increased, whereas that of N1 and N3 had decreased as compared to the 1986 National Survey. We speculated that the proportion of patients who visit dentistry and oral surgery institutions directly or on a referral basis at an early stage might be increasing. As for the reason for the increase in the number of N2 cases, it is likely that cases that would have previously been diagnosed as N1 might now be correctly diagnosed as N2, according to progress in diagnostic imaging techniques such as computed tomography (CT) scanning.

According to T classification by tumor location, early-stage cases were predominant among patients with tongue cancer in the present survey, a finding which was also seen in the Nara Prefecture Survey. Further, early-stage cancer tended to be observed frequently in the lip and oral floor, whereas advanced-stage cancer tended to be more frequent in patients with cancer of the maxillary sinus and jawbone. It was considered that this higher frequency of advanced-stage cancer might have resulted from difficulty in detecting subjective and objective signs in these locations.

Calhoun et al.[30] studied distant metastasis of squamous cell carcinoma in the head and neck in 727 patients; 83 patients (11.4%) had distant metastasis during their clinical course. In these cases, 18.1% demonstrated distant metastasis at the first consultation. In the 1986 National Survey results, M1 accounted for 22 cases (1.6%), while M1 cases were slightly fewer in the present survey, at 17 cases (1.0%). In the present survey, not all of the distant metastases could be found at the time of the first consultation. We considered that distant metastasis would be observed in a greater number of patients during the treatment course. As the presence of distant metastasis strongly affects treatment and prognosis, it is considered necessary to improve diagnostic accuracy by the use of positron emission tomography (PET).

### Nonepithelial tumors

In the Nara Prefecture Survey, malignant nonepithelial tumors accounted for about 10% of head and neck cancers. Malignant lymphoma was the predominant nonepithelial tumor, with the cervical lymph node and Waldeyer's ring being the most frequent locations. Although it has been reported that the incidence of malignant head and neck lymphoma is gradually increasing,[31] the incidence was extremely low in the present survey, at 8 cases (0.4%), in comparison to 61 cases (4.0%) reported in the 1986 National Survey. It was considered that malignant head and neck lymphoma was likely to have been excluded in the present survey, as malignant tumors were not itemized in the ICD-O code of the registration form (for initial registration). Thus, it is necessary to give extra consideration to the terms used for registration.

Howell et al.[16] reported the incidence of malignant melanoma as 0.1%. In the 1986 National Survey, 16 cases (1.1%) were reported, and Okamoto et al.[24] reported malignant tumors of the head and neck in 6 cases (0.3%). In the present study, there were 13 cases (incidence of 0.7%). The prognosis of malignant melanoma is extremely poor and its progress is rapid.[32] Although the incidence in the present survey was low, malignant melanoma characteristics must be taken into consideration during daily clinical practice at departments of dentistry and oral surgery.

As for so-called sarcomas, Yamaguchi et al.[33] studied 32 cases and reported that the incidence of osteosarcoma was highest, at 9 cases. In the present survey, osteosarcoma was reported in only a single case. Nonepithelial tumors vary in their histologic features and very few cases have been reported as compared to reports of carcinoma. We consider it necessary to continue our nationwide survey, and also to investigate the diagnosis, treatment, and prognosis of sarcomas.

**Acknowledgments** We express our heartfelt gratitude to the representatives of each of the member institutions of the Japanese Society of Oral and Maxillofacial Surgeons for their cooperation in the survey.

## References

1. Uchida Y (1988) Statistical and epidemic studies of oral cancers diagnosed and treated by oral surgeons in Japan. Jpn J Dent Sci 7:16–26 (in Japanese)

2. Japan Society for Head and Neck Cancer (2001) General rules for clinical studies on head and neck cancer, 3rd edn. Kanehara, Tokyo, pp 22–28 (in Japanese)
3. Franceschi S, Bidoli E, Herrero R, et al. (2000) Comparison of cancers of the oral cavity and pharynx worldwide: etiological clues. Oral Oncol 36:106–115
4. Tarvainen L, Suuronen R, Lindqvist C, et al. (2004) Is the incidence of oral and pharyngeal cancer increasing in Finland? An epidemiological study of 17 383 cases in 1953–1999. Oral Dis 10:167–172
5. Miyahara H, Kitamura H, Sugimura M, et al. (2001) Malignant head and neck tumors in Nara prefecture. Jpn J Cancer Clin 47:333–339 (in Japanese)
6. Iamaroon A, Pattanaporn K, Pongsiriwet S, et al. (2004) Analysis of 587 cases of oral squamous cell carcinoma in northern Thailand with a focus on young people. Int J Oral Maxillofac Surg 33:84–88
7. Rawashdeh MA, Matalka I (2004) Malignant oral tumors in Jordanians, 1991–2001. A descriptive epidemiological study. Int J Oral Maxillofac Surg 33:183–188
8. Hogan LC, Hall GM, Chambers IG (2005) Oral cancer at a Tasmanian territory referral clinic, The Royal Hobart Hospital 1996–2002. Aust Dent J 50:31–36
9. Bhurgri Y (2005) Cancer of the oral cavity. Trends in Karachi South (1995–2002). Asian Pac J Cancer Prev 6:22–26
10. Ravi M, Mamta S, Kumar D, et al. (2003) Age specific incidence rate and pathological spectrum of oral cancer in Allahabad. Indian J Med Sci 57:400–404
11. Oshima A, Kuroish IT, Tajima K (eds) (2004) Cancer statistics in Japan. 2004. Shinohara, Tokyo, pp 202–217 (in Japanese)
12. Shiboski CH, Schmidt BL, Jordan RCK (2005) Tongue and tonsil carcinoma; increasing trends in the U.S. population ages 20–44 years. Cancer 103:1843–1849
13. Annertz K, Anderson H, Biorklund A, et al. (2002) Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. Int J Cancer 101:95–99
14. Llewellyn CD, Johnson NW, Warnakulasuriya KAAS (2004) Risk factors for oral cancer in newly diagnosed patients aged 45 years and younger: a case-controlled study in Southern England. J Oral Pathol Med 33:525–532
15. Scully C (2002) Oral squamous cell carcinoma: from an hypothesis about a virus, to concern about possible sexual transmission. Oral Oncol 38:227–234
16. Howell RE, Wright BE, Dewar R (2003) Trends in the incidence of oral cancer in Nova Scotia from 1983 to 1997. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 95:205–212
17. Shimizu M (1989) Incidence and classification of oral cancer. In: Shimizu M, Kohama G (eds) Mundhohlen Karzinom Diagnose und Behandlung. Dental Diamond, Tokyo, pp 66–91 (in Japanese)
18. Health and Walfare Statics Association (2003) Journal of Health and Walfare Statics suppl 50(9). Health and Walfare Statics Association, Tokyo, pp 81–83 (in Japanese)
19. Kohama G (1989) Surgical treatment of carcinoma of the mandibular gingiva (including alveolar mucosa and retromolar region). In: Shimizu M, Kohama G (eds) Mundhohlen Karzinom Diagnose und Behandlung. Dental Diamond, Tokyo, pp 212–231 (in Japanese)
20. Wahid A, Ahmad S, Sajjad M (2005) Pattern of carcinoma of oral cavity reporting at dental department of Ayub Medical College. J Ayub Med Coll Abbottabad 17:65–66
21. Kawabe R, Omura S, Saito T, et al. (1999) Risk factors for multiple primary squamous cell carcinomas of the oral cavity. Jpn J Oral Maxillofac Surg 45:421–426 (in Japanese)
22. Kawakami M, Ikemura K (2004) Multiple primary malignant tumors in patients with squamous cell carcinoma of the oral cavity. Influence on the 5-year survival rate. J Jpn Stomatol Soc 53:9–13 (in Japanese)
23. Saikawa M, Ebihara S (2000) Multiple primary neoplasms in patients with squamous cell carcinoma of the oral cavity, mesopharynx, and hypopharynx. Stomato-pharyngol 12:349–360 (in Japanese)
24. Okamoto M, Takahashi H, Yao K, et al. (1999) Clinical trend of head and neck malignancy: a statistical analysis of 26-year experience at the Kitazato University. Oto-Rhino-Laryngology, Tokyo 42:474–482 (in Japanese)
25. Otoh EC, Johnson NW, Danfillo IS, et al. (2004) Primary head and neck cancers in North Eastern Nigeria. West Afr J Med 23: 305–313
26. Aoki S, Kawabe R, Hirota M, et al. (2003) Ten-year clinicostatistical study of oral squamous cell carcinoma. Yokohama Med J. 54:539–546 (in Japanese)
27. Busquets JM, Garcia HA, Pinedo JT, et al. (2003) Clinicopathological characteristics of head and neck squamous cell carcinoma in Puerto Ricans. P R Health Sci J 22:259–264
28. Carvalho AL, Singh B, Spiro RH, et al. (2004) Cancer of the oral cavity: a comparison between institutions in a developing and a developed nation. Head Neck 26:31–38
29. Scott SE, Grunfeld EA, McGurk M (2005) The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma. Oral Oncol 41:396–403
30. Calhoun KH, Fulmer P, Weiss R, et al. (1994) Distant metastases from head and neck squamous cell carcinomas. Laryngoscope 104:1199–1205
31. Yoshino K (2004) Trends of head and neck cancer – is it on the increase or decrease? 1. Epidemiology of head and neck cancer. Jibi Innkoka, Tokeibu Geka (Otolaryngology – Head and Neck Surgery) 76:345–350 (in Japanese)
32. Tanaka N, Mimura M, Ogi K, et al. (2004) Primary malignant melanoma of the oral cavity: assessment of outcome from the clinical records of 35 patients. Int J Oral Maxillofac Surg 33: 761–765
33. Yamaguchi S, Nagasawa H, Suzuki T, et al. (2004) Sarcomas of the oral and maxillofacial region: a review of 32 cases in 25 years. Clin Oral Investig 8:52–55

*Int. J. Cancer:* **124,** 2106–2115 (2009)


# Transglutaminase 3 as a prognostic biomarker in esophageal cancer revealed by proteomics

**Norihisa Uemura[1,2,3], Yukihiro Nakanishi[4], Hoichi Kato[2], Shigeru Saito[1,5], Masato Nagino[3], Setsuo Hirohashi[1] and Tadashi Kondo[1]***

[1]*Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan*
[2]*Division of Esophageal Surgery, National Cancer Center Hospital, Tokyo, Japan*
[3]*Division of Surgical Oncology, Department of Surgery, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[4]*Pathology Division, National Cancer Center Research Institute, Tokyo, Japan*
[5]*Chem and Bio Informatics Department, Infocom Corporation, Tokyo, Japan*

**To develop a prognostic biomarker for esophageal squamous cell carcinoma (ESCC), we examined the proteomic profile of ESCC using two-dimensional difference gel electrophoresis (2D-DIGE), and identified proteins associated with prognosis by mass spectrometry. The prognostic performance of the identified proteins was examined by immunohistochemistry in additional cases. We identified 22 protein spots whose intensity was statistically different between ESCC cases with good ($N = 9$; survived more than 5 years without evidence of recurrence) and poor ($N = 24$; died within 2 years postsurgery) prognosis, within the patient group that had two or more lymph node metastases. Mass spectrometric protein identification resulted in 18 distinct gene products from the 22 protein spots. Transglutaminase 3 (TGM3) was inversely correlated with shorter patient survival. The prognostic performance of TGM3 was further examined by immunohistochemistry in 76 ESCC cases. The 5-year disease-specific survival rate was 64.5% and 32.1% for patients with TGM3-positive and TGM3-negative tumors, respectively ($p = 0.0033$). Univariate and multivariate analyses revealed that TGM3 expression was an independent prognostic factor among the clinicopathologic variables examined. It is noteworthy that the prognostic value of TGM3 was shown to be higher than those of the lymph node metastasis, intramural metastasis and vascular invasion status. These results establish TGM3 as a novel prognostic biomarker for ESCC for the first time. Examination of TGM3 expression may provide novel therapeutic strategies to prevent recurrence of ESCC.**


**Key words:** esophageal cancer; prognosis; proteomics; TGM3 protein; two-dimensional difference gel electrophoresis

Esophageal cancer is the 8th most common cancer[1] and the 6th leading cause of cancer death worldwide.[2] Despite the use of modern surgical techniques in combination with radio- and chemotherapy, early recurrence is common and the overall 5-year survival rate remains below 40%.[3–5] Although the use of adjuvant and neoadjuvant chemotherapies has improved the survival times of esophageal cancer patients,[6] these treatment modalities cause serious side effects in a large number of patients and only benefit a limited number of patients in terms of overall survival times. On the other hand, 45–52% of patients with resectable esophageal cancer treated with surgery alone survive for more than 5 years.[7,8] The patients who can be completely cured by surgery alone receive unnecessary and harmful combination therapy. The response to treatment such as surgery or chemo-radiotherapy is variable, even when the patients are at the same clinical stage, and is not predicted by the existing diagnostic modalities. Accurate risk stratification is therefore of paramount importance to either avoid potential morbidity due to over-treatment or prevent further progression of disease.

Global mRNA expression studies have identified the gene clusters associated with the progression of esophageal cancer,[9–11] suggesting that multiple gene and protein alterations are implicated.[12] These alterations can be considered as potential biomarkers for detecting cancer, determining prognosis, and monitoring disease progression or therapeutic response. However, none of them has been proven to be clinically useful, and the response to treatment such as surgery or chemo-radiotherapy is not predicted by the existing diagnostic modalities. Practical biomarkers to predict response to treatment have long been desired to optimize therapeutic strategies and improve clinical outcomes.

The proteome is a functional translation of the genome. The genomic aberrations of cancer cells are transcribed to the transcriptome, translated to the proteome, then determining cancer phenotypes. In this sense, the proteome is a functional translation of the genome, directly regulating tumor behavior. It is obvious that proteomic features more directly reflect the tumor characters than genomic contents do. Proteomic studies can generate unique data about the final products of genome information. Many lines of evidence demonstrated discordance between mRNA and protein expression.[13–15] In addition, examining DNA sequences and measuring mRNA expression do not accurately predict the status of post-translational modifications such as phosphorylation and glycosylation, which play a key role in regulating the malignant behavior of cancer cells. Taken together, the proteome can be a rich source for biomarker identification.

In this study, we performed a proteomic study to identify biomarkers to predict the clinical outcome of esophageal squamous cell carcinoma (ESCC) patients. We used laser microdissection to recover tumor cells and neighboring normal epithelial cells from surgical specimens of esophageal cancer cases, and subjected the recovered cells to proteomic analysis using two-dimensional difference gel electrophoresis (2D-DIGE). We took particular note of postoperative prognosis in advanced esophageal cancer treated with surgery alone, and discovered prognostic biomarker candidates to optimize the existing surgical treatment strategy. As a result, transglutaminase 3 (TGM3) was identified as a prognostic biomarker candidate. The prognostic performance of TGM3 was successfully validated by immunohistochemistry in 76 additional ESCC cases. This is the first report concerning the prognostic value of TGM3 expression in ESCC. By measuring TGM3 expression in primary tumors, we will be able to refine the prognostic protocol and optimize current therapeutic strategies.

## Material and methods

### *Patients and clinical information*

We examined primary tumor tissues from 82 ESCC patients who underwent surgery at the National Cancer Center Hospital consecutively from 1998 to 2002. All patients underwent curative resection, and were not treated with chemo- or radiotherapy. The

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

Grant sponsors: Ministry of Health, Labor and Welfare, National Cancer Institute of Biomedical Innovation of Japan (Promotion of Fundamental Studies), Foundation for Promotion of Cancer Research (Japan).

***Correspondence to:** Proteome Bioinformatics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Fax: +81-33-547-5298. E-mail: takondo@ncc.go.jp

Received 28 August 2008; Accepted after revision 15 October 2008
DOI 10.1002/ijc.24194
Published online 3 December 2008 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – CLINICOPATHOLOGICAL DATA OF THE 58 ESOPHAGEAL CANCER CASES EXAMINED

| | All cases | Good prognosis group | Poor prognosis group | *p* |
|---|---|---|---|---|
| Number of cases | 58 | 34 | 24 | |
| Age (mean ± SD) (yr) | 62.3 ± 8.4 | 59.9 ± 9.5 | 63.0 ± 6.3 | 0.447 |
| Gender | | | | 0.015[1] |
| Male | 50 | 26 | 24 | |
| Female | 8 | 8 | 0 | |
| Histologic differentiation | | | | 0.416 |
| Well differentiated | 21 | 15 | 6 | |
| Moderately differentiated | 25 | 14 | 11 | |
| Poorly differentiated | 12 | 5 | 7 | |
| Tumor location | | | | 0.316 |
| Upper | 5 | 2 | 4 | |
| Middle | 25 | 13 | 12 | |
| Lower | 28 | 20 | 8 | |
| Macroscopic classification | | | | 0.974 |
| 1 | 1 | 1 | 0 | |
| 2 | 34 | 20 | 14 | |
| 3 | 23 | 13 | 10 | |
| Tumor size (mean ± SD) (cm) | 6.2 ± 2.2 | 6.0 ± 1.9 | 6.4 ± 2.4 | 0.673 |
| Number of LN metastases | | | | <0.001[1] |
| 0 | 13 | 13 | 0 | |
| 1 | 12 | 12 | 0 | |
| 2 | 5 | 1 | 4 | |
| 3 | 7 | 2 | 5 | |
| ≥4 | 21 | 6 | 15 | |
| SCC (mean ± SD) (ng/ml) | 1.5 ± 1.5 | 1.3 ± 1.2 | 1.7 ± 1.7 | 0.303 |
| CEA (mean ± SD) (ng/ml) | 3.0 ± 1.4 | 2.8 ± 1.1 | 3.3 ± 1.8 | 0.200 |
| Lymphatic invasion | | | | 0.025[1] |
| Negative | 22 | 16 | 6 | |
| Positive | 36 | 18 | 18 | |
| Vascular invasion | | | | 0.024[1] |
| Negative | 20 | 16 | 4 | |
| Positive | 38 | 18 | 20 | |
| Intramural metastasis | | | | 0.007[1] |
| Absent | 50 | 33 | 17 | |
| Present | 8 | 1 | 7 | |
| Prognosis | | | | |
| Good prognosis group[2] | 34 | | | |
| Poor prognosis group[3] | 24 | | | |

[1]Considered to be significant ($p < 0.05$).–[2]Survived more than 5 years without evidence of recurrence.–[3]Died within 2 years postsurgery.

patients were newly diagnosed with thoracic ESCC and were followed up for at least 5 years after surgery. The overall clinicopathological data of the cases are summarized in Table I, while information on the individual cases is available in Supplemental Table S1. Two or three tissue fragments, less than 10 $mm^3$ in volume, were grossly obtained from primary tumors. Matched normal mucosal tissues located at least 5 cm away from the tumor margins were also included in this study. The resected tissues were snap-frozen in liquid nitrogen and stored at −80°C until use. The recovered specimens were histologically examined and the clinicopathological stage was determined according to the International Union against Cancer tumor-node-metastasis (TNM) classification.[16] All cases were classified as T3N0-1M0. This study was approved by the ethics committee of the National Cancer Center and written informed consent was obtained from the patients.

The patients that survived more than 5 years without evidence of recurrence were categorized in the good prognosis group ($N = 39$) while the patients that died within 2 years post surgery were categorized in the poor prognosis group ($N = 28$). The proteomic profiles of these two sample groups were compared.

We performed immunohistochemistry on 76 cases, which included 14 cases that were not categorized in either group. The clinicopathological data of the 76 esophageal cancer cases are demonstrated in Table II.

### *Laser microdissection*

Specific cell populations were recovered by laser microdissection according to our previous reports[17,18] (Fig. 1*a*). A 1 $mm^2$ of microdissected area, recorded during microdissection, was recovered from hematoxylin-stained tissues for each 2D-DIGE gel. As tumor tissues could not be recovered in 9 cases (Supplemental Table S1) because of poor preservation, we finally examined 58 tumor tissues and 53 normal epithelium tissues.

### *2D-DIGE and image analysis*

2D-DIGE was performed as previously described.[17,18] In brief, a common internal control sample was created by mixing a small portion of all protein samples used in this study, which was labeled with Cy3 fluorescent dye (CyDye DIGE Fluor saturation dye, GE Healthcare Biosciences, Uppsala, Sweden). Individual samples were labeled with Cy5 fluorescent dye (CyDye DIGE Fluor saturation dye, GE Healthcare Biosciences). These differently labeled protein samples were mixed together and separated by two-dimensional gel electrophoresis (2D-PAGE) according to their isoelectric point and molecular weight. The first dimension separation was achieved using a 24 cm-length immobiline gel (IPG, pI 4–7, GE Healthcare Biosciences) and Multiphor II (GE Healthcare Biosciences), while the second dimension separation using a home-made gradient gel with GiantGelRunner (Biocraft, Tokyo, Japan), with a separation distance of 36 cm. The gels were scanned using a laser scanner (Typhoon Trio, GE Healthcare Biosciences) at the appropriate wavelength for Cy3 or Cy5. For all protein spots, the Cy5 intensity was normalized with the Cy3 intensity in the same gel using the Progenesis SameSpots software version 3 (Nonlinear Dynamics, Newcastle, UK), so that gel-to-gel variations were canceled out (Fig. 1*b*). We monitored the system reproducibility by running the same sample twice (case 15; Supplemental Table S1). The scatter plot showed that the intensity

TABLE II – UNIVARIATE AND MULTIVARIATE ANALYSIS OF PROGNOSTIC FACTORS AND RELATIONSHIP BETWEEN CLINICOPATHOLOGIC VARIABLES AND TGM3 EXPRESSION

| Variable | Number of cases | Disease-specific survival | | Multivariate analysis of tumor-specific survival by Cox regression | | | TGM3 positive (no. cases) | TGM3 negative (no. cases) | Correlation with TGM3 expression *p* value[2] |
|---|---|---|---|---|---|---|---|---|---|
| | | 5 yr (%) | Log-rank (P) | *p* | Relative risk | 95% CI | | | |
| All cases | 76 | 52.6 | | | | | 48 | 28 | |
| Age (yr) | | | 0.4599 | | | | | | 0.345 |
| <65 | 42 | 57.1 | | | | | 29 | 13 | |
| ≥65 | 34 | 47.1 | | | | | 19 | 15 | |
| Gender | | | 0.0597 | | | | | | 0.548 |
| Male | 64 | 48.4 | | | | | 9 | 3 | |
| Female | 12 | 75.0 | | | | | 39 | 25 | |
| Histologic differentiation | | | 0.1551 | | | | | | 0.884 |
| Well | 25 | 64.0 | | | | | 15 | 10 | |
| Nonwell | 51 | 47.0 | | | | | 33 | 18 | |
| Tumor location | | | 0.5276 | | | | | | 0.198 |
| Upper | 5 | 40.0 | | | | | 5 | 0 | |
| Middle, lower | 71 | 53.5 | | | | | 43 | 28 | |
| Macroscopic classification | | | 0.5805 | | | | | | 0.079 |
| 1, 2 | 43 | 55.8 | | | | | 23 | 20 | |
| 3 | 33 | 48.3 | | | | | 25 | 8 | |
| Tumor size (cm) | | | 0.5669 | | | | | | 0.311 |
| <6.0 | 37 | 56.6 | | | | | 26 | 11 | |
| ≥6.0 | 39 | 48.7 | | | | | 22 | 17 | |
| Pathologic N status | | | 0.0013[1] | 0.025[1] | 4.101 | 1.197–14.044 | | | 0.527 |
| pN0 | 18 | 88.9 | | | | | 13 | 5 | |
| pN1 | 58 | 41.3 | | | | | 35 | 23 | |
| SCC (ng/ml) | | | 0.3090 | | | | | | 0.987 |
| ≤1.5 | 61 | 55.7 | | | | | 38 | 23 | |
| >1.5 | 15 | 40.0 | | | | | 10 | 5 | |
| CEA (ng/ml) | | | 0.1402 | | | | | | 0.548 |
| ≤5 | 67 | 55.2 | | | | | 41 | 26 | |
| >5 | 9 | 29.6 | | | | | 7 | 1 | |
| Lymphatic invasion | | | 0.0084[1] | 0.351 | 1.472 | 0.654–3.313 | | | 0.450 |
| Negative | 30 | 70.0 | | | | | 21 | 9 | |
| Positive | 46 | 41.2 | | | | | 27 | 19 | |
| Vascular invasion | | | 0.0082[1] | 0.228 | 1.608 | 0.743–3.478 | | | 0.158 |
| Negative | 31 | 74.2 | | | | | 23 | 8 | |
| Positive | 45 | 37.8 | | | | | 25 | 20 | |
| Intramural metastasis | | | 0.0072[1] | 0.778 | 1.142 | 0.452–2.884 | | | 0.229 |
| Absent | 68 | 57.3 | | | | | 45 | 23 | |
| Present | 8 | 12.5 | | | | | 3 | 5 | |
| TGM3 | | | 0.0033[1] | 0.015[1] | 0.430 | 0.218–0.848 | | | |
| Negative | 28 | 32.1 | | | | | | | |
| Positive | 48 | 64.5 | | | | | | | |

*Abbreviation:* 95% CI, 95% confidence interval.
[1]Considered to be significant ($p < 0.05$).–[2]Fisher's exact test for categorical variables and Mann-Whitney *U* test for continuous variables.

value of 95% of protein spots was scattered within a 2-fold value difference, and that the correlation coefficient was 0.8352, demonstrating the high reproducibility of our profiling method (Fig. 1*c*). The spot intensity data were exported from the Progenesis SameSpots software as Excel files, amenable to data analysis.

*Data analysis*

As a preprocess of data analysis, raw intensity data for each experiment were log2 transformed and then Z score transformation was applied to standardize the distribution of the intensity data.[19] Hierarchical clustering was performed with the Euclidean distance and unweighted pair group methods, using the arithmetic average (UPGMA) method on the standardized data to reveal the global features of the proteomic profiles acquired. To identify the spots that had different intensity between the 2 groups, the z-test was used for each spot. As the obtained *p*-value list possibly included false positive results due to multiple tests, we estimated the false discovery rate (FDR) following the Benjamini-Hochberg procedure according to the previous report.[20] We subsequently selected the spots so that the FDR is less than 0.05. For the comparison of normal tissues with tumors, we chose the spots the intensity ratio of group means was at least 4 times above the aforementioned FDR criteria.

*Mass spectrometric protein identification*

The proteins corresponding to the protein spots detected were identified by mass spectrometry according to our previous report.[21] Cy5-labeled proteins separated by 2D-PAGE were recovered in gel plugs and digested with modified trypsin (Promega, Madison, WI). The trypsin digests were subjected to liquid chromatography coupled with tandem mass spectrometry, a Finnigan LTQ linear ion trap mass spectrometer (Thermo Electron, San Jose, CA) equipped with a nano-electrospray ion source (AMR, Tokyo, Japan). The Mascot software (version 2.1, Matrix science, London, UK) was used to search for the mass of the peptide ion peaks against the SWISS-PROT database (*Homo sapiens*, 16,529 sequence in Sprot_52.5 fasta file). Proteins with a Mascot score of 34 or more were subjected to protein identification. When multiple proteins were identified in a single spot, the proteins with the highest number of peptides were considered as those corresponding to the spot.

*Pathway analysis of expression data*

Pathway analysis of the protein expression pattern was performed using the MetaCore software (GeneGo, St. Joseph, MI). MetaCore identifies networks based on a manually curated

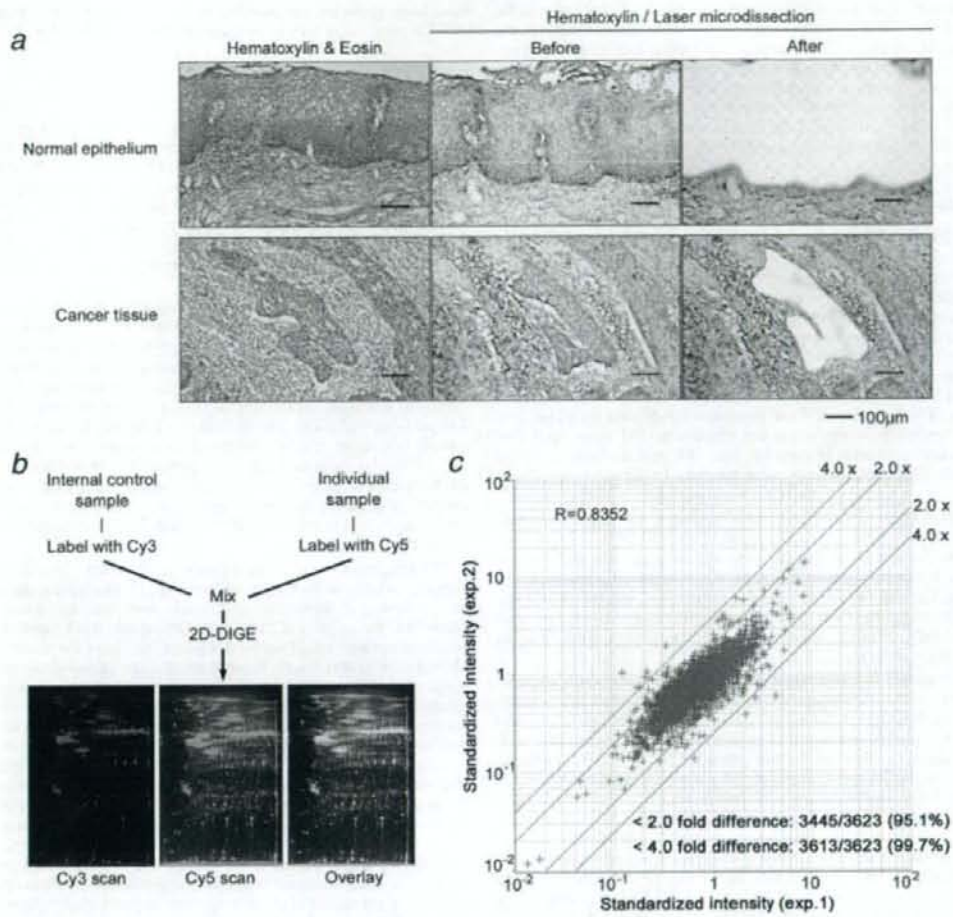


FIGURE 1 – (*a*) Laser microdissection of ESCC. Tissue sections were stained with hematoxylin and eosin for histological observation, while neighboring serial sections were stained with hematoxylin alone for the proteomic study. (*b*) The extracted proteins were labeled with fluorescent dyes and separated by 2D-DIGE. (*c*) Scattergram of 2 independent experiments demonstrating the high reproducibility of 2D-DIGE.

database containing known molecular interactions, functions, and disease interrelationships using proteome data sets. The pathways are identified by the probability that a random set of proteins the same size as the input list would give rise to a particular mapping by chance.

*Immunohistochemistry and tissue microarray*

Immunohistochemical staining for TGM3 was performed on methanol-fixed, paraffin-embedded tissue sections from 76 cases (Supplemental Table S1) using the Dako REAL EnVision Detection System (DAKO, Glostrup, Denmark) following the manufacturer's instructions. The sections were deparaffinized, dehydrated and blocked by 3 mL/L $H_2O_2$ in methanol for 30 min to remove endogenous peroxidase activity. The sections were autoclaved in 10 mM citrate buffer (pH 6.0) at 121°C for 10 min. The primary antibody used was a rabbit polyclonal, mono-specific antibody against TGM3 (HPA004728; Atlas antibodies, Stockholm, Sweden) at a dilution of 1:100. One pathologist (Y. N.) and one medical doctor (N. U.) reviewed the sections stained with anti-TGM3 antibody in a blinded fashion regarding clinical data. The normal esophageal epithelium served as an internal positive control. Cases in which more than 10% tumor cells were positively stained with anti-TGM3 antibody were considered as TGM3 positive, while cases with less than 10% TGM3 positive tumor cells were considered as a TGM3 negative. Staining was evaluated at the dominant differentiation area of the tumor if considerable tumor heterogeneity was present.

We examined TGM3 expression using our home-made tissue microarray containing 59 normal tissues and 323 tumor tissues (Supplemental Table S2).

*Statistical analysis*

The correlation between TGM3 expression and clinicopathological features was evaluated using the Fisher's exact test for categorical variables and the Mann-Whitney *U* test for continuous variables. The disease-specific survival time was calculated from the

first resection of the primary tumor to death of disease-specific causes. All time-to-event end points were computed by the Kaplan-Meier method.[22] Potential prognostic factors were identified by univariate analysis using the log-rank test. Independent prognostic factors were evaluated using the Cox's proportional hazards regression model. $p$ value differences of $<0.05$ were considered to be significant. Statistical analyses were performed using the SPSS 11.0 statistical package (SPSS, Chicago, IL).

## Results

2D-DIGE generated quantitative expression profiles that included 3,623 protein spots per sample. Based on the overall similarity of the acquired protein expression profiles, the samples were divided into 2 groups: tumor tissues and normal epithelial tissues (Supplemental Fig. S1); that is, the proteomic profiles reflected the tissue origin of the sample. Considerable differences were observed between the proteomic profile of tumors and normal tissues; we found 200 protein spots that matched the criteria of an FDR $< 0.001$ and a fold difference $>4$ between the tissue groups. The intensity of 33 of these spots indicated increased protein expression levels while the remaining 167 spots indicated decreased expression in tumor tissues. All proteins corresponding to these 200 protein spots were identified (Supplemental Table S3).

The samples were not grouped according to the prognosis group to which they belonged based on their overall protein expression features. Similarly, no protein spots with significantly different intensity between these two groups were observed. However, the gender, the number of lymph node metastases, the lymphatic and vascular invasion status, and the intramural metastasis status were significantly different between the patients groups with different prognosis (Table I).

The number of lymph node metastases is one of the major prognostic factors in esophageal cancer.[23] We classified the patients based on their lymph node metastasis status into the good and bad prognosis group, and found 22 protein spots with significantly different intensity between the two groups (FDR $< 0.05$). The localization of the 22 spots on the two-dimensional image is shown in Figure 2*a* (enlarged image in Supplemental Fig. S2). Mass spectrometric protein identification revealed that the 22 protein spots corresponded to 18 distinct gene products (Table III, Fig. 2*b* and Supplemental Table S4). Pathway analysis using a MetaCore software analysis tool showed that 17 of the 18 identified proteins were part of a network (Fig. 2*c*) in which STAT1, p53 and HNF4 seemed to be key proteins. TGM3 was connected to STAT1 through Sp1, which directly regulates TGM3 expression[24] and is an intermediary of p53, which is known to be a prognostic factor of several malignancies including esophageal cancer.[25] TGM3 spots seemed 3 times in the list of the 22 protein spots with consistently lower intensity in the poor prognosis group.

To further validate the prognostic value of TGM3 expression in ESCC, we examined the expression of TGM3 in 76 ESCC cases using immunohistochemistry. Both cytoplasmic and nuclear TGM3 staining were observed, depending on the case (Fig. 3*a*, enlarged image in Supplemental Fig. S3), although only the former has been reported previously[26,27] and was considered as indicating positive staining in this study.

The 5-year disease-specific survival rate was significantly higher in the 48 TGM3-positive compared with the 28 TGM3-negative cases (64.5 *versus* 32.1%; $p = 0.0033$; Fig. 3*b*, Table II). Multivariate analysis revealed that TGM3 expression was an independent predictor of disease-specific survival (Table II). The immunohistochemical expression of TGM3 did not correlate with any other clinicopathological variables (Table II).

In tissue microarray analysis, TGM3 was shown to be expressed in all normal squamous epithelia and squamous cell carcinomas examined, including those arising in the skin, lung, oral cavity and uterus. In addition, TGM3 was expressed in a few cases of nonsquamous epithelia and nonsquamous cell carcinomas, including those arising in the breast, prostate and thyroid gland (Fig. 3*c*).

## Discussion

A variety of treatments is currently available for esophageal cancer.[6] The choice of treatment is crucial, as the response is diverse between the patients, even when they are diagnosed at the same clinical stage.[6] Treatment-related complications may easily lead to serious and occasionally fatal adverse reactions, such as myocardial infarction, heart failure, and pneumonia.[28,29] Therefore, by predicting response to treatment and optimizing individualized therapy, we will be able to improve the clinical outcome of esophageal cancer patients.

Novel prognostic modalities have long been desired to improve the management of ESCC. Global genomic and transcriptomic expression studies have been conducted to detect prognostic molecular biomarkers for ESCC.[30,31] However, these studies did not result in the identification of novel practical biomarkers, because they used too many molecules to predict clinical outcome, and did not perform sufficient verification experiments in clinical-scale sample sets using practical methods such as immunohistochemistry. Although proteomics has much potential to reveal the molecular background of esophageal cancer, this is the first report to employ proteomics to identify prognostic biomarkers in esophageal cancer, and to successfully establish TGM3 as a single prognostic biomarker.

The prognosis of esophageal cancer patients may be affected by various factors, including those used in TNM classification and the presence of intramural metastasis and vascular invasion.[29] Therefore, the molecular background of tumors from patients with similar prognosis could vary even when they have the same T and M stage, as in this study. Indeed, we did not identify any protein spots with different intensity between the good and poor prognosis groups. We assumed that this was probably due to the high heterogeneity of the molecular background of the tumors. With this notion, we subsequently focused our analysis on comparing the proteomic profiles of patients with different survival periods within the patient group that had two or more lymph node metastases at the time of pathological diagnosis. This analysis is clinically significant because this patient group generally has poor prognosis and is in more need of the development of suitable prognostic biomarkers. As a consequence, we successfully identified 22 protein spots that had different intensity between the aforementioned patient subgroups. These observations suggest that following a strategy that is based on the use of such clinically relevant parameters is an effective way to identify the proteins that correlate with the malignant potential of tumors.

We identified 18 proteins of prognostic value in the primary tumors. Network analysis revealed that 17 of them were linked through the STAT1, p53 and HNF4 transcription factors, all of which are aberrantly regulated in esophageal cancer. Suppression of the EGF-STAT1 pathway leads to progression of esophageal cancer.[32] p53 immunoreactivity has been detected in 34–67% of ESCC cases,[33–36] and is significantly correlated with cancer-specific death.[25] HNF4alpha expression significantly correlates with MUC4 expression,[37] which is a mediator of tumor growth and metastasis by acting as a ligand for the ErbB2 tyrosine kinase receptor.[38–40] These observations suggest that a limited number of transcription factors may affect a large number of genes, resulting in poor prognosis for esophageal cancer.

We considered that TGM3 was a strong prognostic biomarker candidate, because its immunohistochemical expression clearly correlated with clinical outcome in our study and it has also been shown to be potentially relevant to ESCC.[27] TGM3 expression has been previously correlated with certain malignant phenotypes in ESCC. Liu *et al.* reported TGM3 expression and histological grade are inversely correlated in ESCC,[27] while it should be noted that Mendez *et al.* reported that TGM3 expression is inversely

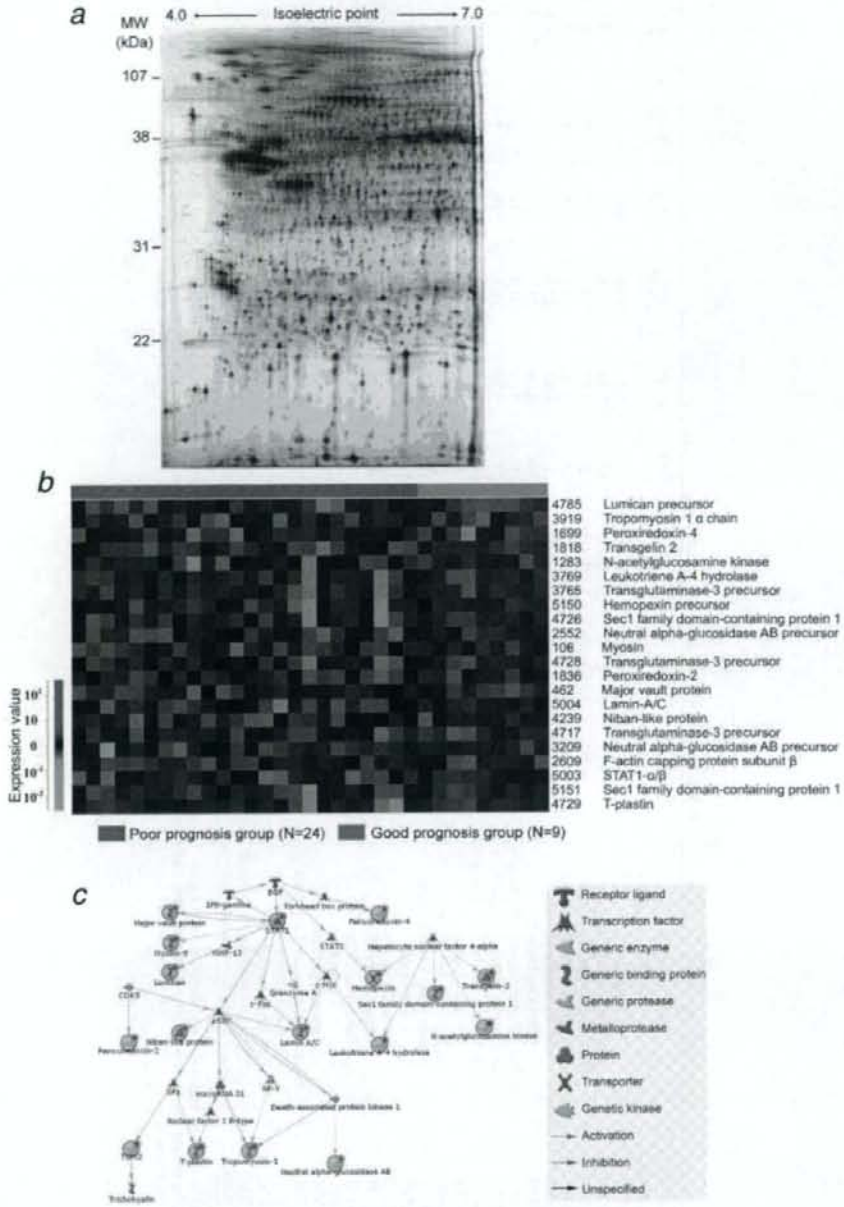


FIGURE 2 – Identification of proteins differentially expressed in ESCC. (*a*) A representative two-dimensional gel image showing the localization of proteins of ESCC tissues. The 22 spots identified in this study are circled and numbered. The spot numbers correspond to those in Figure 2*b*, Table III, and Supplemental Table S2. The image is shown enlarged in Supplemental Figure S2. (*b*) Hierarchical clustering of the 33 ESCC cases based on the intensity of the 22 protein spots. Purple, poor prognosis group; brown, good prognosis group. Right, spot numbers and protein names. (*c*) Pathway analysis of the identified proteins. Seventeen of eighteen of the identified proteins were shown to be part of a network by pathway analysis using the MetaCore software analysis tool.

TABLE III – LIST OF THE IDENTIFIED PROTEINS

| Spot no.[1] | Accession no.[2] | Identified protein | z-value | *p*-value | FDR | Overall rank | pI (cal)[3] | MW (cal; kD)[3] | Protein score[4] | Peptide matches | Sequence coverage (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 106 | P35579 | Myosin-9 | −3.99 | 6.70E-05 | 2.85E-02 | 8 | 5.5 | 227.6 | 420 | 8 | 3.8 |
| 462 | Q14764 | Major vault protein | −3.85 | 1.18E-04 | 3.33E-02 | 12 | 5.34 | 99.6 | 88 | 2 | 4.8 |
| 1283 | Q9UJ70 | *N*-acetylglucosamine kinase | −4.94 | 7.85E-07 | 1.42E-03 | 2 | 5.81 | 37.7 | 97 | 2 | 6.1 |
| 1699 | Q13162 | Peroxiredoxin-4 | 3.69 | 2.21E-04 | 3.80E-02 | 21 | 5.86 | 30.7 | 186 | 4 | 13.7 |
| 1818 | P37802 | Transgelin-2 | −6.51 | 7.47E-11 | 2.71E-07 | 1 | 8.41 | 22.5 | 228 | 3 | 18.6 |
| 1836 | P32119 | Peroxiredoxin-2 | −3.95 | 7.86E-05 | 2.85E-02 | 10 | 5.66 | 22 | 211 | 4 | 24.2 |
| 2552 | Q14697 | Neutral alpha-glucosidase AB precursor | −4.03 | 5.51E-05 | 2.85E-02 | 7 | 5.74 | 107.3 | 137 | 2 | 3.4 |
| 2609 | P47756 | F-actin capping protein subunit beta | −3.73 | 1.89E-04 | 3.80E-02 | 17 | 5.36 | 31.6 | 224 | 3 | 13.4 |
| 3209 | Q14697 | Neutral alpha-glucosidase AB precursor | −3.75 | 1.74E-04 | 3.80E-02 | 16 | 5.74 | 107.3 | 419 | 9 | 12.1 |
| 3765 | Q08188 | Transglutaminase-3 | −4.36 | 1.29E-05 | 1.17E-02 | 4 | 5.62 | 76.9 | 888 | 16 | 24 |
| 3769 | P09960 | Leukotriene A-4 hydrolase | −4.64 | 3.52E-06 | 4.26E-03 | 3 | 5.8 | 69.9 | 702 | 12 | 20.5 |
| 3919 | P09493 | Tropomyosin-1 alpha chain | 3.73 | 1.90E-04 | 3.80E-02 | 18 | 4.69 | 32.7 | 104 | 2 | 6.7 |
| 4239 | P02545 | Lamin-A/C | −3.83 | 1.31E-04 | 3.38E-02 | 14 | 6.57 | 74.4 | 282 | 4 | 7.7 |
| 4717 | Q08188 | Transglutaminase-3 | −3.76 | 1.67E-04 | 3.80E-02 | 15 | 5.62 | 76.9 | 133 | 2 | 2.7 |
| 4726 | P02790 | Hemopexin precursor | −4.09 | 4.39E-05 | 2.65E-02 | 6 | 6.55 | 52.4 | 159 | 4 | 6.7 |
| 4728 | Q08188 | Transglutaminase-3 | −3.95 | 7.77E-05 | 2.85E-02 | 9 | 5.62 | 76.9 | 535 | 10 | 14.6 |
| 4729 | P13797 | T-plastin | −3.61 | 3.02E-04 | 4.98E-02 | 22 | 5.52 | 70.9 | 122 | 3 | 5.3 |
| 4785 | P51884 | Lumican precursor | 3.89 | 9.91E-05 | 3.27E-02 | 11 | 6.16 | 38.7 | 159 | 3 | 8.6 |
| 5003 | P42224 | Signal transducer and activator of transcription 1-alpha/beta | −3.71 | 2.07E-04 | 3.80E-02 | 19 | 5.74 | 87.9 | 114 | 3 | 4.3 |
| 5004 | Q96TA1 | Niban-like protein 1 | −3.85 | 1.20E-04 | 3.33E-02 | 13 | 5.81 | 83.1 | 292 | 5 | 10.8 |
| 5150 | Q8WVM8 | Sec1 family domain-containing protein 1 | −4.10 | 4.10E-05 | 2.65E-02 | 5 | 5.89 | 72.7 | 115 | 2 | 3.7 |
| 5151 | Q8WVM8 | Sec1 family domain-containing protein 1 | −3.70 | 2.15E-04 | 3.80E-02 | 20 | 5.89 | 72.7 | 94 | 3 | 8.7 |

[1]Spot numbers refer to those in Figure 1A and Supplemental Figure 1.–[2]Accession numbers of proteins were derived from Swiss-Prot and NCBI nonredundant databases.–[3]Theoretical isoelectric point and molecular weight obtained from Swiss-Prot and the ExPASy database. (http://au.expasy.org).–[4]Mascot score for the identified proteins based on the peptide ions score ($p < 0.05$) (http://www.matrixscience.com).

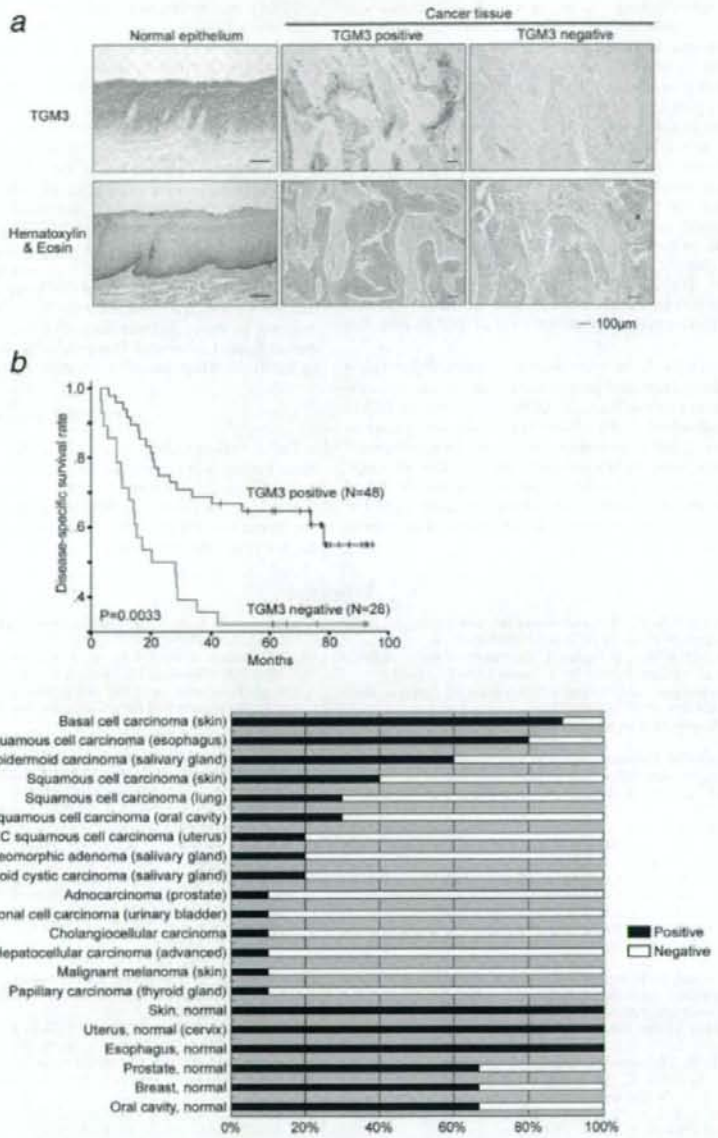


FIGURE 3 – Validation of the differential expression of TGM3 in relation to prognosis. (*a*) Immunohistochemistry; TGM3 was expressed in all normal esophageal epithelia and 63% cancer tissues. The image is shown enlarged in Supplemental Figure S3. (*b*) Kaplan-Meier survival curves for ESCC patients in relation to TGM3 expression. Patients with positive TGM3 expression had significantly better ($p = 0.0033$) prognosis than patients with negative TGM3 expression. (*c*) Tissue microarray data. TGM3 was expressed in normal squamous epithelia and squamous cell carcinomas arising in a range of tissues (a detailed list is shown in Supplemental Table 2), and some adenocarcinomas and their normal counterparts.

correlated with lymph node metastasis of oral squamous cell carcinoma.[41] Although Liu *et al.* reported that TGM3 expression correlated with histological grade in ESCC,[27] we did not observe any correlation between TGM3 expression and the clinico-pathological parameters examined, including the tumor stage. This discordance is probably due to the different antibody used, the different surgical procedure followed,[42] and the different clinical background of the patients included in the studies. A multi-institutional

validation study that will take into account such differences will be required to firmly establish TGM3 as a practical prognostic biomarker. In oligomicroarray analysis in ESCC, the TGM3 gene was reported to be suppressed and to correlate with lymph node metastasis.[31] Although these reports suggested that TGM3 may be involved in cancer progression, they have not shown the prognostic or other practical value of TGM3 expression examination in ESCC. Genome and transcriptome studies have listed many biomarker candidates included TGM3,[31] without, however, detecting or proposing single biomarkers of potential practical use. In this study, we detected the proteins associated with the survival of ESCC patients using a proteomic approach, and subsequently selected TGM3 as an individual prognostic biomarker candidate. We then showed that the positive or negative immunohistochemical expression of TGM3 in our study corresponds to TGM3 expression as assessed by the proteomics tools we employed, and thus can be used to assess the expression level of TGM3 (Fig. 3*a*) in practice.

TGM3 plays a key role in epidermal terminal differentiation through cross-linking structural proteins such as involucrin, loricrin and small proline-rich proteins.[43] Although the role of TGM3 has been well established in the differentiation of skin keratinocytes,[44] little information is available concerning its involvement in esophageal epithelium. TGM3 stabilizes the cornified envelope of cells, a process that precedes the transition of keratinocytes to corneocytes by apoptosis. Therefore, down-regulation of TGM3 in ESCC may interrupt the potentially critical initiation of apoptosis, thereby favoring tumor cell survival.

TGM3 expression was decreased in ESCC tissues compared with normal tissues; all normal tissues strongly expressed TGM3 compared to only 63% of the ESCC tissues. This observation is consistent with previous microarray studies[45,46] that indicated that TGM3 is down-regulated in many types of malignancies compared with the corresponding normal tissues, suggesting that the reduced expression of TGM3 may play a common role in the carcinogenesis of not only ESCC but also other carcinomas.

In conclusion, we performed the first esophageal cancer proteomics study that uses a large-scale clinical sample set that includes prognostic information, and identified TGM3 expression as a novel prognostic indicator in ESCC. The use of a common internal control sample in 2D-DIGE, and the use of laser microdissection contributed to accurate protein expression profiling. The immunohistochemical examination of TGM3 expression may help identify patients with high risk for recurrence, and may improve the clinical outcome of these patients through closer postoperative follow-up and additional treatment. Our results therefore provide the possibility for the development of novel strategies for ESCC management.

## Acknowledgements

The excellent technical support of Ms. Yukiko Kobori and Ms. Mina Fujishiro in electrophoresis and Ms. Satoko Kouda in the immunohistochemical study are greatly appreciated. Norihisa Uemura is a recipient of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research (Japan) within the framework of the 3rd Term Comprehensive 10-Year Strategy for Cancer Control.

## References

1. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. Int J Cancer 1999;80:827–41.
2. Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. Int J Cancer 1999;83:18–29.
3. Ilson DH. Oesophageal cancer: new developments in systemic therapy. Cancer Treat Rev 2003;29:525–32.
4. Enzinger PC, Mayer RJ. Esophageal cancer. N Engl J Med 2003;349: 2241–52.
5. Mariette C, Balon JM, Piessen G, Fabre S, Van Seuningen I, Triboulet JP. Pattern of recurrence following complete resection of esophageal carcinoma and factors predictive of recurrent disease. Cancer 2003;97:1616–23.
6. Shinoda M, Hatooka S, Mori S, Mitsudomi T. Clinical aspects of multimodality therapy for resectable locoregional esophageal cancer. Ann Thorac Cardiovasc Surg 2006;12:234–41.
7. Ando N, Iizuka T, Kakegawa T, Isono K, Watanabe H, Ide H, Tanaka O, Shinoda M, Takiyama W, Arimori M, Ishida K, Tsugane S. A randomized trial of surgery with and without chemotherapy for localized squamous carcinoma of the thoracic esophagus: the Japan Clinical Oncology Group Study. J Thorac Cardiovasc Surg 1997;114:205–9.
8. Ando N, Iizuka T, Ide H, Ishida K, Shinoda M, Nishimaki T, Takiyama W, Watanabe H, Isono K, Aoyama N, Makuuchi H, Tanaka O, et al. Surgery plus chemotherapy compared with surgery alone for localized squamous cell carcinoma of the thoracic esophagus: a Japan Clinical Oncology Group Study–JCOG9204. J Clin Oncol 2003;21: 4592–6.
9. Tamoto E, Tada M, Murakawa K, Takada M, Shindo G, Teramoto K, Matsunaga A, Komuro K, Kanai M, Kawakami A, Fujiwara Y, Kobayashi N, et al. Gene-expression profile changes correlated with tumor progression and lymph node metastasis in esophageal cancer. Clin Cancer Res 2004;10:3629–38.
10. Kan T, Shimada Y, Sato F, Ito T, Kondo K, Watanabe G, Maeda M, Yamasaki S, Meltzer SJ, Imamura M. Prediction of lymph node metastasis with use of artificial neural networks based on gene expression profiles in esophageal squamous cell carcinoma. Ann Surg Oncol 2004;11:1070–8.
11. Gomes LI, Esteves GH, Carvalho AF, Cristo EB, Hirata R, Jr, Martins WK, Marques SM, Camargo LP, Brentani H, Pelosof A, Zitron C, Sallum RA, et al. Expression profile of malignant and nonmalignant lesions of esophagus and stomach: differential activity of functional modules related to inflammation and lipid metabolism. Cancer Res 2005;65:7127–36.
12. Vallbohmer D, Lenz HJ. Predictive and prognostic molecular markers in outcome of esophageal cancer. Dis Esophagus 2006;19:425–32.
13. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardia SL, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM, Beer DG. Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics 2002;1:304–13.
14. Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell 2005;8:393–406.
15. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–30.
16. Sobin L, Wittekind C. UICC TNM Classification of malignant tumors, 6th edn. New York: Wiley-Liss, 2002.
17. Kondo T, Seike M, Mori Y, Fujii K, Yamada T, Hirohashi S. Application of sensitive fluorescent dyes in linkage of laser microdissection and two-dimensional gel electrophoresis as a cancer proteomic study tool. Proteomics 2003;3:1758–66.
18. Kondo T, Hirohashi S. Application of highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dyes) to laser microdissection and two-dimensional difference gel electrophoresis (2D-DIGE) for cancer proteomics. Nat Protoc 2006;1:2940–56.
19. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. J Mol Diagn 2003;5:73–81.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
21. Okano T, Kondo T, Kakisaka T, Fujii K, Yamada M, Kato H, Nishimura T, Gemma A, Kudoh S, Hirohashi S. Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis. Proteomics 2006;6: 3938–48.
22. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
23. Tachibana M, Kinugasa S, Yoshimura H, Shibakita M, Tonomoto Y, Dhar DK, Nagasue N. Clinical outcomes of extended esophagectomy with three-field lymph node dissection for esophageal squamous cell carcinoma. Am J Surg 2005;189:98–109.
24. Lee JH, Jang SI, Yang JM, Markova NG, Steinert PM. The proximal promoter of the human transglutaminase 3 gene. Stratified squamous epithelial-specific expression in cultured cells is mediated by binding of. Sp1 and ets transcription factors to a proximal promoter element. J Biol Chem 1996;271:4561–8.
25. Aloia TA, Harpole DH, Jr, Reed CE, Allegra C, Moore MB, Herndon JE, II, D'Amico TA. Tumor marker expression is predictive of survival in patients with esophageal cancer. Ann Thorac Surg 2001;72: 859–66.
26. Hitomi K, Presland RB, Nakayama T, Fleckman P, Dale BA, Maki M. Analysis of epidermal-type transglutaminase (transglutaminase 3)

in human stratified epithelia and cultured keratinocytes using monoclonal antibodies. J Dermatol Sci 2003;32:95–103.
27. Liu W, Yu ZC, Cao WF, Ding F, Liu ZH. Functional studies of a novel oncogene TGM3 in human esophageal squamous cell carcinoma. World J Gastroenterol 2006;12:3929–32.
28. Ishikura S, Nihei K, Ohtsu A, Boku N, Hironaka S, Mera K, Muto M, Ogino T, Yoshida S. Long-term toxicity after definitive chemoradiotherapy for squamous cell carcinoma of the thoracic esophagus. J Clin Oncol 2003;21:2697–702.
29. Igaki H, Kato H, Tachimori Y, Sato H, Daiko H, Nakanishi Y. Prognostic evaluation for squamous cell carcinomas of the lower thoracic esophagus treated with three-field lymph node dissection. Eur J Cardiothorac Surg 2001;19:887–93.
30. Hirasaki S, Noguchi T, Mimori K, Onuki J, Morita K, Inoue H, Sugihara K, Mori M, Hirano T. BAC clones related to prognosis in patients with esophageal squamous carcinoma: an array comparative genomic hybridization study. Oncologist 2007;12:406–17.
31. Uchikado Y, Inoue H, Haraguchi N, Mimori K, Natsugoe S, Okumura H, Aikou T, Mori M. Gene expression profiling of lymph node metastasis by oligomicroarray analysis using laser microdissection in esophageal squamous cell carcinoma. Int J Oncol 2006;29:1337–47.
32. Watanabe G, Kaganoi J, Imamura M, Shimada Y, Itami A, Uchida S, Sato F, Kitagawa M. Progression of esophageal carcinoma by loss of EGF-STAT1 pathway. Cancer J 2001;7:132–9.
33. Shimaya K, Shiozaki H, Inoue M, Tahara H, Monden T, Shimano T, Mori T. Significance of p53 expression as a prognostic factor in oesophageal squamous cell carcinoma. Virchows Arch A Pathol Anat Histopathol 1993;422:271–6.
34. Sarbia M, Porschen R, Borchard F, Horstmann O, Willers R, Gabbert HE. p53 protein expression and prognosis in squamous cell carcinoma of the esophagus. Cancer 1994;74:2218–23.
35. Furihata M, Ohtsuki Y, Ogoshi S, Takahashi A, Tamiya T, Ogata T. Prognostic significance of human papillomavirus genomes (type-16, -18) and aberrant expression of p53 protein in human esophageal cancer. Int J Cancer 1993;54:226–30.
36. Wang DY, Xiang YY, Tanaka M, Li XR, Li JL, Shen Q, Sugimura H, Kino I. High prevalence of p53 protein overexpression in patients with esophageal cancer in Linxian. China and its relationship to progression and prognosis. Cancer 1994;74:3089–96.
37. Piessen G, Jonckheere N, Vincent A, Hemon B, Ducourouble MP, Copin MC, Mariette C, Van Seuningen I. Regulation of the human mucin MUC4 by taurodeoxycholic and taurochenodeoxycholic bile acids in oesophageal cancer cells is mediated by hepatocyte nuclear factor 1alpha. Biochem J 2007;402:81–91.
38. Carraway KL, Ramsauer VP, Haq B, Carothers Carraway CA. Cell signaling through membrane mucins. Bioessays 2003;25:66–71.
39. Singh AP, Moniaux N, Chauhan SC, Meza JL, Batra SK. Inhibition of MUC4 expression suppresses pancreatic tumor cell growth and metastasis. Cancer Res 2004;64:622–30.
40. Fauquette V, Perrais M, Cerulis S, Jonckheere N, Ducourouble MP, Aubert JP, Pigny P, Van Seuningen I. The antagonistic regulation of human MUC4 and ErbB-2 genes by the Ets protein PEA3 in pancreatic cancer cells: implications for the proliferation/differentiation balance in the cells. Biochem J 2005;386:35–45.
41. Mendez E, Fan W, Choi P, Agoff SN, Whipple M, Farwell DG, Futran ND, Weymuller EA, Jr, Zhao LP, Chen C. Tumor-specific genetic expression profile of metastatic oral squamous cell carcinoma. Head Neck 2007;29:803–14.
42. Pennathur A, Luketich JD. Resection for esophageal cancer: strategies for optimal management. Ann Thorac Surg 2008;85:S751–6.
43. Kalinin AE, Kajava AV, Steinert PM. Epithelial barrier function: assembly and structural features of the cornified cell envelope. Bioessays 2002;24:789–800.
44. Kim IG, Gorman JJ, Park SC, Chung SI, Steinert PM. The deduced sequence of the novel protransglutaminase E (TGase3) of human and mouse. J Biol Chem 1993;268:12682–90.
45. Chen BS, Wang MR, Xu X, Cai Y, Xu ZX, Han YL, Wu M. Transglutaminase-3, an esophageal cancer-related gene. Int J Cancer 2000;88:862–5.
46. Luo A, Kong J, Hu G, Liew CC, Xiong M, Wang X, Ji J, Wang T, Zhi H, Wu M, Liu Z. Discovery of Ca2+-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. Oncogene 2004;23:1291–9.

# Molecular background of α-fetoprotein in liver cancer cells as revealed by global RNA expression analysis

Shigeru Saito,[1,2] Hidenori Ojima,[3] Hitoshi Ichikawa,[4] Setsuo Hirohashi[1] and Tadashi Kondo[1,5]

[1]Proteome Bioinformatics Project, National Cancer Center Research Institute; [2]Chem and Bio Informatics Department, Infocom Corporation; [3]Pathology Division, National Cancer Center Research Institute; [4]Cancer Transcriptome Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan

(Received April 16, 2008/Revised August 1, 2008; August 12, 2008/Accepted August 14, 2008/Online publication November 24, 2008)

**α-Fetoprotein (AFP) is considered to be a diagnostic and prognostic biomarker in hepatocellular carcinoma (HCC). However, the role of AFP in the development of HCC is presently obscure. We hypothesized that a certain set of genes is expressed in a manner coordinate with AFP, and that these genes essentially contribute to the malignant characteristics of AFP-producing HCC. To address this hypothesis, we carried out global mRNA expression analysis of 21 liver cancer cell lines that produce varying levels of AFP. We identified 213 genes whose mRNA expression levels were significantly correlated with that of AFP ($P < 0.0001$). These included liver-specific transcription factors for AFP and other albumin family genes. Eighteen HCC-associated genes and 11 genes associated with malignancies other than HCC showed significant correlations with AFP production levels. Genes involved in lipid catabolism, blood coagulation, iron metabolism, angiogenesis, and the Wnt and mitogen-activated protein kinase pathways were also identified. Text data mining revealed that participation in the transcription factor network could explain the connection between 78 of the identified genes. Glypican 3, which is a component of the Wnt pathway and contributes to HCC development, had the fifth highest correlation coefficient with AFP. Reactivity to specific antibodies confirmed the significant correlation between AFP and glypican 3 expression in HCC tissues. These observations suggest that AFP-producing liver cancer cells may have a unique molecular background consisting of cancer-associated genes. From this genome-wide association study, novel aspects of the molecular background of AFP were revealed, and thus may lead to the identification of novel biomarker candidates. (*Cancer Sci* 2008; 99: 2402–2409)**

Hepatocellular carcinoma (HCC) is one of the most common and aggressive malignancies worldwide and is the third leading cause of cancer death.[1] It is a major health problem with high prevalence in Asia and Africa,[2,3] and recent studies indicated that the incidence of HCC has increased substantially in the USA and UK over the last three decades.[4,5] The prognosis for HCC patients remains dismal at present, and novel diagnostic modalities as well as improvement of the therapeutic strategies currently in use are required to improve the clinical outcome for HCC patients.

Altered α-fetoprotein (AFP) level is a hallmark of HCC development;[6] a considerable proportion of HCC patients have elevated plasma AFP, and diagnostic value of AFP was suggested in the patients with liver cirrhosis.[7] Plasma AFP is a useful prognostic indicator, as the median survival rate of HCC patients with markedly elevated AFP is significantly shorter than that of patients with normal or moderately elevated AFP.[8] Preoperative AFP levels are predictive of HCC recurrence,[9,10] and may therefore be used in deciding therapeutic options for HCC patients after surgery. The lens culinaris agglutinin-reactive fraction of AFP, in particular, has been shown to be significantly associated with portal vein invasion and poor clinical outcomes.[11] AFP has been shown to function as a superoxide dismutase[12] and as an apoptotic factor,[13–15] and to directly promote proliferation in cultured cells.[16–20] Nevertheless, the molecular background of HCC associated with increased AFP levels in HCC patients and the mechanisms underlying the association of AFP with the onset of HCC and poor prognosis are presently unclear.

We hypothesized that a certain set of genes are expressed in a manner coordinate with AFP, and that these genes are responsible for the greater tumor size, portal vein thrombosis, and lack of histological differentiation that are observed in HCC tumors with higher AFP expression.[8] We have previously reported that the expression levels of 11 proteins correlate highly with that of AFP.[21] In the present study, we generated gene expression profiles of 21 liver cancer cell lines using DNA microarrays and investigated the genes whose expression level correlated significantly with AFP mRNA levels. The functional properties of the identified genes and their association with each other at the transcription level were examined using a text data mining program. The correlation or otherwise of AFP expression with the product of each identified gene was validated in HCC tissues using specific antibodies.

## Materials and Methods

**Cell lines.** The following 21 cell lines were used: HuH-7, JHH-7, JHH-5, HepG2, HT17, HuH-1, Hep3B, Li-7, PLC/PRL/5, KIM-1, KYN-2, HLE, HLF, JHH-4, JHH-6, SK-Hep-1, KYN-3, PH5-CH, PH5-T, RBE, and SSP-25. Details of these cell lines are summarized in our previous proteomics report.[21]

**Clinical specimens.** HCC tissues were obtained from 23 HCC patients at the time of surgery, fixed in formalin, and embedded in paraffin. The project was approved by the institute's ethical committee and written informed consent for the use of the tissues for research purposes was obtained from the donors.

**Western blotting.** Cellular proteins were extracted from the cell lines using a urea lysis buffer (6 mol/L urea, 2 mol/L thiourea, 3% CHAPS, 1% Triton X-100), and 30 μg protein was separated by sodium dodecylsulfate–polyacrylamide gel electrophoresis with an e-PAGE system (ATTO, Tokyo, Japan) as described previously.[21] Immunoblot analysis was carried out using primary antibodies against AFP (1:200 dilution, clone ZSA06; Zymed Laboratories, South San Francisco, CA, USA) and β-actin (1:1000 dilution, clone AC-15; Sigma, St Louis, MO, USA), peroxidase-conjugated secondary antibody (1:1000 dilution; GE Healthcare, Uppsala, Sweden), and enhanced chemiluminescence (GE Healthcare). The enhanced chemiluminescence signal was monitored using Fuji LAS-1000 (Fuji Film, Tokyo, Japan) and measured with ImageQuant TL (GE Healthcare).

**Gene expression analysis.** For gene expression analysis, total RNA was prepared from the 21 cell lines using an RNeasy mini

[5]To whom correspondence should be addressed.
E-mail: takondo@gan2.res.ncc.go.jp



doi: 10.1111/j.1349-7006.2008.00973.x

kit (Qiagen, Hilden, Germany). The integrity of the purified RNA was confirmed using 2100 Bioanalyzer and an RNA 6000 nano LabChip kit (Agilent Technologies, Santa Clara, CA, USA). The DNA microarray used was a Human Genome U133 plus 2.0 array (Affymetrix, Santa Clara, CA, USA). Target cRNA was prepared from 1 μg of the purified RNA with a one-cycle cDNA synthesis kit and 3′-amplification reagents for *in vitro* transcription amplification and biotin-labeling (Affymetrix). Hybridization to the microarrays, washing, staining with the antibody amplification procedure, and scanning were carried out according to the manufacturers' instructions. The scanned image data were processed using the GeneChip Operating Software (version 1.4; Affymetrix). The signal expression value of each probe set was calculated and normalized by setting the signal value mean for each experiment to 100 so that minor differences between the experiments were adjusted. Among the 54 675 probes on the DNA microarray, we first selected 12 091 probes based on MicroArray Quality Control Project database analysis,[22] in which the intraplatform and interplatform reproducibility was examined for each gene using individual RNA samples and quantitative reverse transcription–polymerase chain reaction (RT-PCR). The 12 901 probes corresponded to 12 901 unique genes assigned to NCBI Entrez Gene ID numbers. The data were log-transformed (base 2) to produce a closer to normal distribution for statistical analysis.

To measure the similarity of gene expression profiles between AFP and other genes, we used Pearson's correlation coefficient $r_{ij}$ as follows:

$$r_{ij} = \frac{\sum_{k=1}^{m}(p_{ik}-\overline{p}_i)(p_{jk}-\overline{p}_j)}{\sqrt{\sum_{k=1}^{m}(p_{ik}-\overline{p}_i)^2 \sum_{k=1}^{m}(p_{jk}-\overline{p}_j)^2}},$$

where $m$ is the number of observations, $p_i$ is the AFP expression profile, $p_j$ is the expression profile of the gene in question, and $P_i$ is the arithmetic mean of $p_i$ over $m$ observations. We used the $z$-transforms of the observed correlation coefficients, calculated as follows:

$$z_{ij} = \frac{1}{2}\ln\frac{1+|r_{ij}|}{1-|r_{ij}|}.$$

The $Z$-statistic approximately follows the standard normal distribution:[23]

$$Z = \frac{z_{ij}}{\sqrt{1/(m-3)}}$$

We tested the observed correlation coefficients statistically under the null hypothesis that H0: $r_{ij} = 0$, with the significance level set at $\alpha/2$, that is, we rejected the null hypothesis if $Z > Z_{\alpha/2}$. The significance level was set at $\alpha = 0.0001$ in consideration for multiple tests.

**Quantitative RT-PCR.** cDNA was generated from mRNA using the SuperScript III kit (Invitrogen, Carlsbad, CA, USA). Quantitative amplification was carried out using the 7500 Real-time PCR system (Applied Biosystems) and was monitored with TaqMan Gene Expression Assays using premade primers, human glyceraldehydes-3-phosphate dehydrogenase (GAPDH), and TaqMan Universal PCR Master Mix according to the manufacturer's instructions (Applied Biosystems). All experiments were carried out in triplicate. The following 16 cancer-associated genes were examined: *AFP*, glypican 3 (*GPC3*), thrombopoetin (*THPO*), S100 calcium binding protein P (*S100P*), meprin Aα (*MEP1A*), prospero-related homeobox 1 (*PROX1*), frequently rearranged in advanced T-cell lymphomas 2 (*FRAT2*), carcinoembryonic antigen-related cell adhesion molecule 1 (*CEACAM1*), frequently rearranged in advanced T-cell lymphomas (*FRAT1*), v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (*ERBB3*), α2-HS-glucoprotein (*AHSG*), v-raf murine sarcoma viral oncogene homolog B1 (*BRAF*), suppression of tumorigenicity (*ST7*), visinin-like-1 (*VSNL1*), regucalcin (*RGN*), and secretagogin EF-hand calcium binding protein (*SCGN*).

**Text data mining.** To find the transcriptional regulation network of the identified genes, we carried out text data mining using MetaCore (GeneGo, Saint Joseph, MI, USA; http://www.genego.com). Dijkstra's shortest path algorithms were first calculated using a prefilter based on tissue type (fetal or non-fetal liver).(24) Genes involved in transcriptional regulation were then extracted from the networks. Finally, genes not involved in the networks were excluded and a transcriptional regulation network of AFP-related genes was obtained.

**Immunohistochemical study.** Immunohistochemical staining for AFP and GPC3 was carried out using an automated immunohistochemical stainer according to the manufacturer's protocol (Envision; Dako Cytomation, Glostrup, Denmark). Serial sections of formalin-fixed, paraffin-embedded tissues (4 μm thick) were placed on silane-coated slides. Sections with the maximum tumor diameter were selected for immunohistochemical evaluation. A polyclonal antibody against AFP (rabbit, 1:100 dilution; Dako Cytomation) and an antibody against GPC3 (clone 1G12, 1:2000 dilution; BioMosaics, Burlington, VT, USA) were used. All sections were evaluated by H.O. and T.K. without knowledge of any clinical or pathological information; cases for which consensus was not reached were re-evaluated using a dual-headed microscope. AFP and GPC3 expression in the sections was scored as follows: negative, no membranous or cytoplasmic expression in the cancer cells; positive 1+, membranous and/or cytoplasmic expression observed in less than 50% of cancer cells; 2+, membranous and/or cytoplasmic expression in >50% of cancer cells.

## Results

**Expression of AFP in the liver cancer cell lines.** We examined the AFP expression levels in 21 liver cancer cell lines using western blotting, DNA microarrays, and quantitative RT-PCR (Fig. 1a). AFP measured by DNA microarrays and quantitative RT-PCR showed consistent expression at the mRNA level in the 21 cell lines ($r3 = 0.763$). AFP expression at the protein level was concordant with that at the mRNA level as measured by DNA microarrays ($r1 = 0.613$) and quantitative RT-PCR ($r2 = 0.546$) (Fig. 1b). To examine the genes that had expression patterns similar to AFP, we used DNA microarray data because western blotting and quantitative RT-PCR do not generate gene expression data in a genome-wide manner. The presence of a significant correlation between western blotting data, DNA microarray data, and quantitative RT-PCR data suggested that the use of microarray data for measuring gene expression is quite acceptable.

**Genes associated with AFP expression.** From an initial DNA microarray data set consisting of 12 091 genes, we used MicroArray Quality Control criteria to select the 213 genes whose expression level significantly correlated with that of AFP ($P < \alpha = 0.0001$) (Supporting Table S1). The correlation coefficient value of the selected genes was at least 0.724980 (Supporting Table S1). The selected genes included *hepatocyte nuclear factor 4, alpha (HNF4A)*, *transcription factor 1 (TCF1)* and *forkhead box A3 (FOXA3)*, the transcription factors for AFP (Table 1; Supporting Table S2). Expression of the albumin family of genes, which is regulated by the same transcription factors as AFP, such as transthyretin, albumin, and vitamin D binding protein, was also associated with AFP expression (Table 1; Supporting Table 2). We found that the expression of 18 liver cancer-associated genes and 11 genes reported to be associated with malignancies other than liver cancer correlated highly with AFP expression

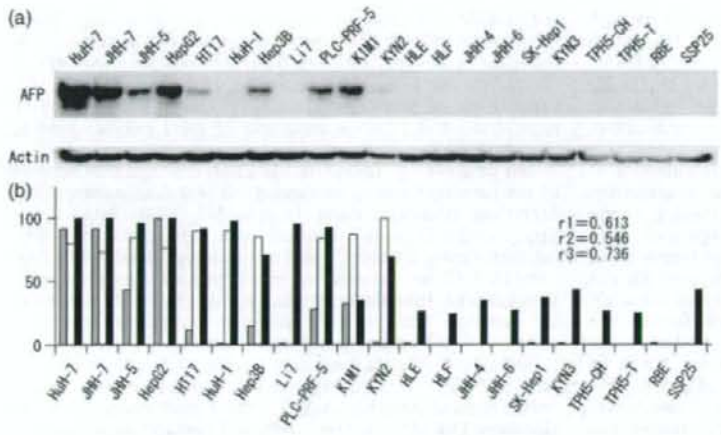


**Fig. 1.** α-Fetoprotein (AFP) expression in the liver cancer cell lines examined. (a) AFP expression using Western blotting. Actin served as positive control. (b) AFP expression at the protein level was concordant with that at the mRNA level as measured by DNA microarrays and quantitative reverse transcription–polymerase chain reaction (RT-PCR). r1, r2, and r3 are correlation coefficiency values; r1, western blotting versus DNA microarray; r2, western blotting versus quantitative RT-PCR; r3, quantitative RT-PCR versus DNA microarray.

**Table 1. List of α-fetoprotein (AFP)-associated genes that were correlated with AFP and malignancies**

| Correlation coefficiency value | Gene symbol | Gene title |
|---|---|---|
| Transcription factors for AFP | | |
| 0.881559 | HNF4A | Hepatocyte nuclear factor 4α |
| 0.800777 | TCF1 | Hepatic nuclear factor 1 (HNF1) |
| 0.748229 | FOXA3 | Forkhead box A3 |
| Albumin family | | |
| 0.901225 | ALB | Albumin |
| 0.738336 | GC | Vitamin D binding protein |
| 0.917934 | TTR | Transthyretin (prealbumin, amyloidosis type I) |
| Hepatocellular carcinoma-related genes | | |
| 0.931226 | GPC3 | Glypican 3 |
| 0.896721 | ASGR2 | Asialoglycoprotein receptor 2 |
| 0.885392 | ASGR1 | Asialoglycoprotein receptor 1 |
| 0.875061 | AHSG | α-2-HS-glycoprotein |
| 0.836128 | THPO | Thrombopoietin |
| 0.814368 | HPN | Hepsin |
| 0.804988 | VTN | Vitronectin |
| 0.799604 | PROX1 | Prospero-related homeobox 1 |
| 0.771268 | CEACAM1 | Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) |
| 0.75632 | ERBB3 | Erythroblastic leukemia viral oncogene homolog 3 (avian) |
| 0.741689 | REL | v-rel reticuloendotheliosis viral oncogene homolog (avian) |
| 0.72498 | FGFR3 | Fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism) |
| 0.725946 | GLUL | Glutamate-ammonia ligase (glutamine synthetase) |
| 0.76969 | GJB1 | Gap junction protein β1, 32 kDa (connexin 32, Charcot-Marie-Tooth neuropathy, X-linked) |
| 0.833362 | PC | Pyruvate carboxylase |
| 0.850565 | VIL1 | Villin 1 |
| 0.734071 | ZG16 | Zymogen granule protein 16 |
| 0.724989 | PRAP1 | Proline-rich acidic protein 1 |
| Genes associated with cancer other than liver cancer | | |
| 0.790659 | EVA1 | Epithelial V-like antigen 1 |
| 0.745726 | SHD | Src homology 2 domain containing transforming protein D |
| 0.80636 | MEP1A | Meprin Aα (PABA peptide hydrolase) |
| 0.83576 | S100P | S100 calcium binding protein P |
| 0.790596 | FRAT2 | Frequently rearranged in advanced T-cell lymphomas 2 |
| 0.766911 | FRAT1 | Frequently rearranged in advanced T-cell lymphomas |
| 0.730513 | CHEK2 | CHK2 checkpoint homolog (*Schizosaccharomyces pombe*) |
| 0.777384 | ST7 | Suppression of tumorigenicity 7 |
| 0.75987 | DMD | Dystrophin (muscular dystrophy, Duchenne and Becker types) |
| 0.810027 | VSNL1 | Visinin-like 1 |
| 0.724989 | PRAP1 | Proline-rich acidic protein 1 |



doi: 10.1111/j.1349-7006.2008.00973.x

(Table 1; Supporting Table S2). As the aim of the present study was to find genes whose expression correlated significantly with AFP expression, and to discuss the possible mechanisms underlying the contribution of aberrant AFP expression to the HCC phenotypes, validation of the detected gene expression levels using other methods was critical. The expression of 16 selected genes, including *AFP*, was examined by quantitative RT-PCR, showing consistent expression levels for all genes except *S100P* and *BRAF* (Supporting Fig. S1).

**Reported biological function of genes associated with AFP expression.** The identified genes were grouped based on their function as reported previously: 11 genes are known to be involved in lipid metabolism, 14 in the blood coagulation pathway, six in iron metabolism, four in angiogenesis, and three encode complement factors (Table 2). Five of the genes identified are reported to be involved in signal transduction, including three in the Wnt signaling pathway and two in the mitogen-activated protein kinase pathway (Table 2).

**Transcription network of the identified genes.** We used a text data mining approach to examine the transcriptional network of the identified genes (Fig. 2; a higher resolution image of the network is shown in Supporting Fig. S2). This literature-based interpretation

**Table 2. List of α-fetoprotein (AFP)-associated genes that were involved in normal regulatory pathways**

| Correlation | Gene symbol | Gene title |
|---|---|---|
| Lipid catabolism[11] | | |
| 0.953605 | APOB | Apolipoprotein B |
| 0.932435 | APOA1 | Apolipoprotein A-I |
| 0.928801 | FABP1 | Fatty acid binding protein 1, liver |
| 0.911320 | APOC2 | Apolipoprotein C-II |
| 0.897877 | APOC3 | Apolipoprotein C-III |
| 0.789613 | DHCR24 | 24-Dehydrocholesterol reductase |
| 0.785963 | APOH | Apolipoprotein H |
| 0.743957 | LIPC | Lipase, hepatic |
| 0.741831 | SCARB1 | Scavenger receptor class B |
| 0.741049 | HMGCS2 | 3-Hydroxy-3-methylglutaryl-Coenzyme A synthase 2 |
| 0.741007 | PCSK9 | Kexin type 9 |
| Blood coagulation[14] | | |
| 0.931987 | SERPIND1 | Heparin cofactor |
| 0.899619 | PROZ | Vitamin K-dependent plasma glycoprotein |
| 0.875732 | F7 | Coagulation factor VII |
| 0.874013 | F10 | Coagulation factor X |
| 0.854635 | SERPINF1 | α-2 antiplasmin |
| 0.840493 | FGL1 | Fibrinogen-like 1 |
| 0.832607 | LOC55908 | Hepatocellular carcinoma-associated gene TD26 |
| 0.811034 | KNG1 | Kininogen 1 |
| 0.793770 | SERPINF2 | α-2 antiplasmin PEDF |
| 0.776527 | PROC | Inactivator of coagulation factors Va and VIIIa |
| 0.773841 | FGG | Fibrinogen gamma chain |
| 0.764485 | SERPINC1 | Antithrombin |
| 0.749890 | F13B | Coagulation factor XIII, B polypeptide |
| 0.746955 | F5 | Coagulation factor V |
| Iron metabolism[6] | | |
| 0.872017 | LEAP-2 | Liver-expressed antimicrobial peptide 2 |
| 0.860938 | TF | Transferrin |
| 0.825157 | HPX | Hemopexin |
| 0.783582 | TFR2 | Transferrin receptor 2 |
| 0.751094 | HAMP | Hepcidin antimicrobial peptide |
| 0.727492 | SLC11A2 | Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2 |
| Angiogenesis[4] | | |
| 0.763843 | ANGPTL3 | Angiopoietin-like 3 |
| 0.739551 | HRG | Histidine-rich glycoprotein |
| 0.724980 | FGFR3 | Fibroblast growth factor receptor 3 |
| 0.811034 | KNG1 | Kininogen 1 |
| Complement[3] | | |
| 0.803829 | C8A | Complement component 8, α polypeptide |
| 0.766028 | C2 | Complement component 2 |
| 0.775684 | C8B | Complement component 8, β polypeptide |
| Wnt pathway[3] | | |
| 0.931226 | GPC3 | Glypican 3 |
| 0.790596 | FRAT2 | Frequently rearranged in advanced T-cell lymphomas 2 |
| 0.766911 | FRAT1 | Frequently rearranged in advanced T-cell lymphomas |
| Mitogen-activated protein kinase pathway[2] | | |
| 0.832252 | MAP3K13 | Mitogen-activated protein kinase kinase kinase 13 |
| 0.797606 | MAPK6 | Mitogen-activated protein kinase 6 |