Profiling Gene Expression Ratios of Paired Cancerous and Normal Tissue Predicts Relapse of Esophageal Squamous Cell Carcinoma

Yoshio Ishibashi, Nobuyoshi Hanyu, Koji Nakada, Yutaka Suzuki, Takashi Yamamoto, Katsuhiro Tanaka, Kiyoshi Ohkawa, Noriko Hashimoto, Toshiharu Nakajima, Hirohisa Saito, Masato Matsushima, and Mitsuyoshi Urashima

Department of Surgery [Y. I., N. H., K. N., Y. S., T. Y., K. Y.], Department of Biochemistry I [K. O.], and Division of Clinical Research and Development [M. M., M. U.], The Jikei University School of Medicine, Minato-ku, Tokyo 105-8461, and National Center for Child Health and Development [N. Has., T. N., H. S.], Setagaya-ku, Tokyo 154-8567, Japan

ABSTRACT

Esophageal squamous cell carcinoma has heterogeneous clinical outcomes that cannot be predicted well using any existing clinical or molecular prognostic factors. Gene expression profiling may enable more precise prediction of the clinical outcome of these patients. We developed a new approach using gene expression ratios of paired cancerous and normal tissue specimens from the same patient to reduce the effects of variation among individuals. Using oligonucleotide microarrays, we analyzed total RNA expression levels corresponding to 12,600 transcript sequences in 24 paired cancerous and normal tissue operative specimens from 12 patients with esophageal squamous cell carcinoma. Hierarchical clustering analysis using gene expression ratios (cancer:normal) divided the 12 patients into two groups: all 7 patients in the first cluster survived without relapse (median follow-up, 483 days), whereas all 5 patients in the second cluster relapsed (median relapse-free survival time, 280 days; log-rank test, \( P = 0.006 \)). In contrast, clustering either with cancerous tissue alone or with normal tissue alone did not show significant differences in the outcomes. The expressions of a variety of genes related with cell cycle, gene-repair, apoptosis and chemoradiotherapy resistance were up-regulated in the poor prognostic cluster. These results suggest that ratios of paired gene expression profiles may more efficiently predict relapse-free survival of esophageal squamous cell carcinoma than existing prognostic factors or than gene expression profiling with cancerous tissue alone.

INTRODUCTION

In general, squamous cell cancer in the esophagus is an aggressive tumor with poor prognosis, with or without chemoradiotherapy (1). However, the clinical outcomes of esophageal cancer are heterogeneous, and some patients not only with early stage but also with advanced stage cancer may survive longer than clinically expected, and vice versa. Yet, only the stage based on TNM classification is widely accepted as a prognostic factor at present (2), which is still far from an accurate predictor. Molecular analyses of esophageal cancer have largely focused on individual candidate genes, with particular emphasis on \( p16^\text{INK4A} \) and \( p53 \) (3), which are no better prognostic factors than existing clinical staging and TNM classification.

The recent development of microarray analysis provides the opportunity to take a genome-wide approach to predict clinical outcomes in a variety of cancers based on the molecular classification of a similar pathological group of cancers with different prognoses (4). However, unlike classical clinical studies, data on thousands of gene expression patterns from a much smaller number of samples must be examined. In addition, whether differential gene expression originates from variation among individuals or from clinical heterogeneity of a given cancer remains to be clarified.

We hypothesized that using the gene expression ratio of paired cancerous and normal tissue samples from the same patient for a given gene could reduce the effects of individuality and, consequently, increase the accuracy of predicting clinical outcomes compared with analysis of cancerous tissue alone.

MATERIALS AND METHODS

Patients. Paired esophageal cancer and normal tissue samples (total 24) were obtained from 12 patients with esophageal cancer after provision of written informed consent, and a protocol approved by the Ethics Committee for Biomedical Research of the Jikei Institutional Review Board, Jikei University School of Medicine, Japan. Eligibility requirements for this study included squamous cell carcinoma of the esophagus that was confirmed by more than two board pathologists and that was free of cancer invasion in the tumor margin and was clinically limited to the locoregional area (no distant metastasis in TNM classification; stage I to III). Tumor stage and grade were classified according to the fifth edition of the TNM classification of the International Union against Cancer (2). In the present study, primary tumors \( T_1 \) and \( T_2/3 \) in TNM classification were defined as early (E) and advanced (A), respectively. All patients had adequate general health including hepatic, renal, and bone marrow reserve, and could tolerate the planned surgical procedure. Patients were ineligible if they had Barrett’s esophageal adenocarcinoma, stage IV based on TNM classification, presence of other primary cancers, inability of total resection, or prior chemotherapy and/or radiotherapy. On an outpatient basis, the 12 patients were periodically (1–4 months) examined to exclude relapse of the disease using standard examinations including blood examination of tumor markers, chest X-ray, endoscopy, and computed tomography of the chest and abdomen.

Samples. Each cancer tissue sample was divided into two specimens: one for pathological confirmation that the sample was composed of more than 80% cancer cells and the other for RNA extraction. Paired normal tissue was obtained from an area that was at least 4 cm away from any cancerous tissue. Normal tissues were also divided into two specimens: one for pathological confirmation that the sample did not contain either cancer cells or premalignant tissue morphologically. Total RNA was extracted using the RNeasy kit (Qiagen, Chuoh-ku, Tokyo, Japan), and biotinylated cRNAs were then generated. Samples were hybridized onto Affymetrix U95A Array oligonucleotide arrays (Affymetrix, Santa Clara, CA). Arrays were subsequently developed with phycoerythrin-conjugated streptavidin and were scanned to obtain quantitative gene expression levels. Paired normal and cancerous tissue specimens from each patient were handled simultaneously during RNA extraction and hybridization.

Statistics. Genes with expression levels \( \leq 0 \) units in at least one sample were deleted, which resulted in a reduction in the number of genes from 12,626 to 5,843. Gene expression ratios of paired cancerous and normal tissues were simply calculated as unit levels of cancerous tissue divided by unit levels of cancerous tissue divided by unit levels of cancerous tissue \( (> 0 \) units), which were then transformed with \( \log_2 \log(\text{cancer/normal}) = \log(\text{cancer}) - \log(\text{normal}) \). Next, these were normalized with the \( z \) value. Hierarchical clustering analysis was then performed using Spotfire software version 7.0 (Spotfire, Somerville, MA). Associations between categorical parameters were examined using Fisher’s exact and \( \chi^2 \) tests. The Kaplan-Meier product-limit method was used to estimate relapse-free survival. For comparison purposes, the Mantel log-rank test was per-
formed. Differences between groups were considered significant if the \( P \) was < 0.05.

RESULTS

Patient characteristics are shown in Table 1. Age ranged from 52 to 82 (mean, 66 ± 10) years, and the majority were males. Patients 8 and 9 were brothers. All of the patients underwent surgery between November 2, 2000, and January 29, 2002, and were followed for 121–644 days.

The summary of raw data were as follows; mean of normal, 1016; quartile of normal, 140 (25/312(50)/777(75)); mean of cancer, 1102; quartile of cancer, 156(25)/369(50)/937(75). Correlation-of-law data between normal and cancer were 0.86 in 70,116 observations. SDs among 12 normal samples and 12 cancer samples in each gene were compared with show distributions of heterogeneity in normal and cancer expression profiles (Fig. 1). Before attempting to predict the clinical outcome by gene expression profiling, we examined whether hierarchical clustering can distinguish cancerous tissue from normal counterparts. Twenty-four samples containing 12 normal and 12 cancerous tissue samples were analyzed using hierarchical clustering, resulting in two major clusters: \((a)\) 12 normal samples and 1 cancer and \((b)\) 11 cancer samples. The profiles of the most prominent genes expressed at high levels in cancerous and normal tissue samples are shown in Fig. 2A and B, respectively. Of the 29 kinds of genes expressed at a higher level in cancerous than in normal tissue (similarity at 2.74 among the gene cluster), 19 (66%) kinds had been previously reported to associate with cellular proliferation and gene repair (Fig. 2A). In contrast, one-half of the 38 kinds of genes expressed at higher levels in normal than in cancer tissue (similarity at 2.41 among the gene cluster) have been associated with enzymatic proteins, and only two genes associated with cellular proliferation and gene repair were present (Fig. 2B).

Hierarchical clustering analyses were then performed for the 12 paired cancer:normal gene expression ratios, 12 cancerous tissue samples alone, and 12 normal tissue samples alone (Fig. 3), and compared with clinical information. Two distinct clusters were formed when the cancer:normal expression ratio was used (Fig. 3A); all of the patients in the first cluster (cluster I) survived without relapse (median follow-up, 483 days), whereas all patients included in the second cluster (cluster II) relapsed (median relapse-free survival time, 280 days; \( P = 0.0013 \)). In contrast, hierarchical clustering using either cancerous tissue alone (Fig. 3B) or normal tissue alone (Fig. 3C) from the 12 patients was not associated with clinical stages, TNM classification, or relapse. Other pathologic prognostic factors (pathologic grade of differentiation and lymphovascular invasion) did not show significant association with any clusters obtained by analysis of cancer:normal ratio, cancerous tissue alone, or normal tissue alone, using the two-sided Fisher’s exact test.

The two clusters based on gene expression ratios of cancer:normal tissue also showed a significant difference in relapse-free survival time (log-rank test, \( P = 0.006 \); Fig. 4A). There was also a significant difference in relapse-free survival time between presence \((N_1)\) and absence \((N_0)\) of lymph node metastasis \((P = 0.03;\) Fig. 4C), but its difference was smaller than that of hierarchical clustering based on gene expression profiling. Other clinical prognostic factors, including stages based on TNM classification (Fig. 4B), invasiveness of primary tumor (early, \( T_1 \), or advanced, \( T_2 \) and \( T_3 \); Fig. 4D), grade of differentiation, and degree of lymphovascular invasion, did not make significant difference in relapse-free survival time, although the sample size was too small to conclude that these prognostic factors have no impact on predicting clinical outcomes.

The genes in cluster I with high paired cancer:normal expression ratios (Fig. 5A) were classified into several groups based on reported function: mucin and tumor antigen, chromosome \((abiquitous Kruppel-like factor, major centromere autoantigen: CENP-B, histone H1, phosphoribosylpyrophosphate synthetase, CTF5\), signaling to direct cell fate \((Wnt-13, HOX2H, Bcl-2 interacting killer, DRAK2, SIM2, Smad3, RAIG1\), and others. In contrast, those in cluster II were mainly related to ubiquitin-related genes \((abiquitous-conjugating enzyme, ubiquitin carrier protein, SUMO1\), viral oncoregens \((EBV small RNA-associated protein, 100 kDa EBNA coactivator, E6-AP, papillomavirus\); Fig. 5B), cell cycle, and gene repair \([Int-6, dynamin, CBFp300-interacting transactivator (CITED2), c-myc binding protein (MBP-1), serine/threonine protein kinase, SMARCA5, cksb2, ku (p70/p80) subunit, MTPC1, hPMS1, RhAp48, PTTG1IP, CUTLI, replication protein 1\),
Fig. 2. Hierarchical clustering of 24 specimens composed of 12 pairs of cancer (red lines in dendrogram at top) and normal (blue lines) tissue samples. Each column, a single sample; each row, a single gene. Red areas, increased RNA expression; green areas, decreased RNA expression. A, a typical gene profile highly expressed in cancerous tissue (red); B, a typical gene profile highly expressed in normal esophageal tissue (green).

**DISCUSSION**

Profiling gene expression ratios of paired esophageal squamous cancerous and normal tissue specimens allowed us to predict relapse-free survival time more precisely than did any known clinical prognostic factors or the expression profiles of cancerous and normal tissues from each patient. Although the number of analyzed samples in this study was small and validation studies with larger numbers are required, the present findings do have statistically significant relationships (log-rank test in Kaplan-Meier survival curve, \( P = 0.006 \)) that suggest less possibility of a type I error. Importantly, association with clinical outcomes was shown to be significant not by the gene expression profile of either cancerous tissue alone or normal tissue alone, but by the ratio of cancer to normal tissue specimens from the
same patient, suggesting that profiling gene expression ratios may provide more informative findings from a huge number of variables with a small number of samples, and reduce the effects of individuality. Comparisons of gene expression profiling between cancerous tissue and paired normal counterparts have been reported in previous studies (5, 6). However, they did not use the ratio of paired expression for analysis but merely used comparison. As shown in Fig. 2C, the gene expression profiles in normal esophageal tissue were not homogeneous, thus adjustment using normal counterparts is obligatory to produce an accurate comparison among patients. Differences in gene expression profiles may arise not only from the nature of cancer cell biology but also from individuality as well as from organ-specificity. Paired samples from the same patients share the same genomic DNA and same exposure of esophageal mucosa to hazardous environments such as heavy alcohol drinking and smoking in the same duration and amounts, thus the differences in gene expression ratios most likely originate from the biological nature of the esophageal carcinoma, canceling out individuality. Heterogeneity among individuals was high in both normal and cancer gene expression profiles as shown in Fig. 1. By matching original esophageal tissue, organ-specific RNA expression patterns could also be deleted. Therefore, the ratio of cancer:normal expression is thought to predict clinical outcome quite efficiently.

Dhanasekaran et al. (7) examined over 700 prostate-cancer samples using a microarray to determine hepsin and pim-1 proteins as clinical prognostic markers. Bigger sample sizes enable us to stratify patients with known prognostic factors; however, the use of numerous arrays raises problems from an economical point of view. When a study population is narrowed depending on an etiologically identical group, sample size can be dramatically reduced. Hedenfalk et al. (8) targeted hereditary breast cancer with mutation at either BRCA1 (n = 7) or BRCA2 (n = 7) and demonstrated different gene expression profiles for the two groups. When investigators tried to abstract the truth from a large dataset obtained from gene expression profiles, they contrived new statistical approaches in addition to hierarchical clustering. Nielsen et al. (9) succeeded in molecular classification of soft-tissue tumors by singular value decomposition using 41 samples. For molecular classification of small round blue-cell tumors (SRBCTs), Khan et al. (10) applied artificial neural networks using 63 samples (23 biopsies and 40 different cell lines). Supervised machine learning has been used for analysis of gene expression profiling in diffuse large B-cell lymphoma and other kinds of carcinomas (11–13). Others applied epidemiological strategy in combination with hierarchical analysis (14, 15). However, to apply gene expression profiling for general clinical use, the predictability of the clinical outcome has to be superior to existing clinical prognostic factors. The present study has proven the efficiency of gene expression profiling for clinical oncology, as Zhang et al. (16) demonstrated with colon cancer using recursive partitioning for the analysis of gene expression profiles.

Many sorts of cell cycle-related genes were highly expressed in our cancerous tissue compared with our normal tissue, which is suspected to be a parallel phenomenon of cell proliferation or a consequence of the tumorigenic process rather than an etiological finding of esophageal carcinoma. Similar phenomena have also been reported in the gene expression profiling of other carcinomas (17, 18). In contrast, profiling gene expression ratios discriminated prognosis of patients efficiently, as described above. These profiles showed different hierarchical clustering from data sets of cancer tissue alone. The enhanced
expression of mucin-related proteins in the better prognostic group (cluster I) may reflect the existence of normal function as esophageal mucosa. Viral oncoproteins related to EBV and HPV E6-Ap were detected in the poor prognostic group (cluster II). No previous reports have demonstrated expression of the EBV gene in esophageal squamous cancer cells, even in a high-prevalence area in China (19). On the contrary, the high frequency of occurrence of HPV in esophageal squamous cell carcinoma implicated HPV as one of the possible etiological factors in this disease (20). However, its prevalence seemed to be lower than first reported, or else the contribution of HPV to the etiology of esophageal cancer differs among geographic areas, being low in European countries and high in Asian counties (21, 22).

Kihara et al. (23) applied microarray analysis to 20 esophageal cancer tissue samples to detect a 52-gene set that predicted sensitivity against chemotherapy. In our study, a variety of genes related to gene-repair, apoptosis, and chemoradiotherapy resistance were up-regulated in the poor prognostic group (cluster II). In particular, higher expression of superoxide dismutase (SOD), MRP, and ubiquitin-conjugating and -carrier protein may be consistent with previous reports (24, 25). MRP is well documented to play a role in the mechanism of chemoresistance, and, recently, VX-710 (biricodar; Incel), which restores drug sensitivity to P-glycoprotein (MDR1)- and MRP1-expressing cells, was shown to have efficacy in patients with advanced chemoresistant breast cancer in a Phase II study (26). Up-regulation of ku70/80 may also prevent DNA damage and apoptosis induced by chemoradiotherapy, as reported previously (27). The higher expression of MTA1 in the poor prognostic group (cluster II) in the present study is consistent with a previous report that an esophageal cancer overexpressing MTA1 showed significantly higher frequencies of adventitial invasion and lymph node metastasis (28).

In conclusion, gene expression profiling using the ratio of paired cancerous and normal tissue samples from the same patients may more efficiently predict the clinical outcome of esophageal squamous cell cancer and delineate molecules that affect the prognosis of the patients.

REFERENCES


Fig. 5. Profiling gene expression ratios of paired cancerous and normal tissues from the same patients. The typical profiles of genes expressed higher in cluster I, in which no relapse cases were observed, are shown in A. The typical profiles of gene expressed higher in cluster II, in which all five cases relapsed, are shown in B, C, D.)
Profiling Gene Expression Ratios of Paired Cancerous and Normal Tissue Predicts Relapse of Esophageal Squamous Cell Carcinoma

Yoshio Ishibashi, Nobuyoshi Hanyu, Koji Nakada, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/63/16/5159

Cited articles
This article cites 26 articles, 9 of which you can access for free at:
http://cancerres.aacrjournals.org/content/63/16/5159.full.html#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
/content/63/16/5159.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.