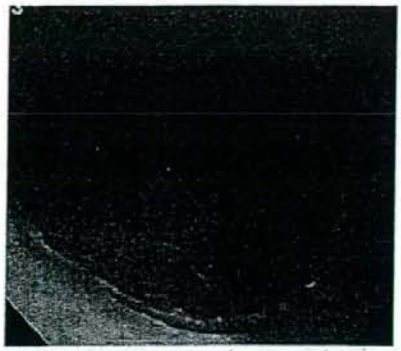

Figure 3. An en bloc resection was achieved without complication in 100 minutes.

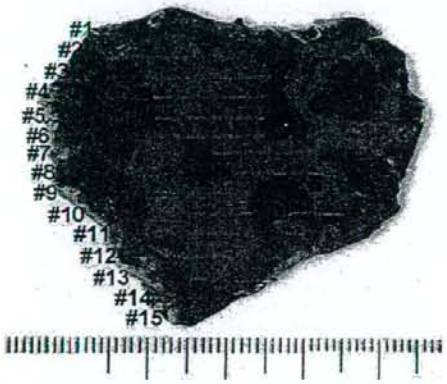

Figure 4. The resected specimen was 50 mm in diameter, with both the lateral and vertical margins negative by endoscopy. Both the lateral and vertical margins were negative on histopathologic examination, and the depth of invasion was m2. A curative local resection was achieved in this case.

RESULTS

During the study period, 4 patients were treated with ESD. All 4 lesions were eligible for outcome analysis. Clinical characteristics of the patients are presented in Table 1. Each of the 4 ESD cases was successfully resected en bloc, with no complications. The mean (SD) ESD time was 58 ± 42 minutes (range 15-100 minutes), and the mean (SD) size of the resected specimens was 35 ± 15 mm (range 15-50 mm). On histopathologic examination, the lateral and vertical margins were negative in 3 of 4 ESD cases, but the depth of invasion was sm1 in 2 of those cases, so additional CRT was performed on those patients. A curative local resection was achieved in the other case. None of the patients developed local recurrence or distant metastasis in the follow-up period. There were no immediate or late complications related to ESD procedures reported. The median (SD) follow-up time was 3 ± 2 months (range 0-6 months) for the ESD group.

TABLE 1. Clinical characteristics of patients

No. lesions	4
Stage before CRT (stage I/II)	3/1
Days after CRT (median)	749 (range 217-1377)
Residual/recurrent	2/2
Tumor depth (m/sm)	2/2
Tumor size (mean [SD]) (mm)	35 ± 15 (range 15-50)
Procedure time (mean [SD]) (min)	58 ± 42
En bloc resection rate	100% (4/4)
Tumor-free lateral margin rate	75% (3/4)
Local recurrence rate	0% (0/4)
Complication (perforation)	0 (0%)

DISCUSSION

The ESD technique, by using a B-knife[21-24] and an IT-knife,[17,23,24] enhanced the en bloc resection rate, thereby increasing the likelihood of curative results for local residual or recurrent tumors. In fact, ESDs with a B-knife and an IT-knife are performed to treat superficial neoplastic lesions, such as gastric and colonic neoplasms, at the NCCH.[17,22-24] ESD has enabled us to treat recurrent gastric cancers after EMR. As indicated in our previous reports,[26] about 5% of such cases involved perforations, although virtually all of the perforation cases were successfully treated by means of endoscopic clipping, without the need for additional surgery.

The esophagus is located in the mediastinum, so the risks of ESD are further enhanced, and perforations must be avoided. The newly developed B-knife results in a safer ESD, because the electric current is localized at the needle tip.[21] The IT-knife [17,23,24] also decreases the risk of perforation as a result of the insulated tip attached to the end of the needle. A B-knife was mainly used for the dissection of fibrosis caused by CRT. The combined use of these two instruments has enabled us to safely perform ESDs even for local recurrence of residual tumors after CRT with successful results similar to our experience in the colorectum.[23,24] Although the number of patients who underwent ESD in our series was limited and the follow-up periods were short, there were no cases of recurrence after ESD during any of the follow-up periods. Further follow-up data are required, however, for meaningful recurrence and survival analyses.

For comparison, 17 local recurrent or residual tumors (10 recurrent tumors, 7 residual tumors) in 14 patients treated at the NCCH between January 2005 and December

2005 by conventional strip biopsy (EMR) were included as historical controls. Ten of the EMR lesions were stage I treated by CRT, and the other 7 were stage II lesions. The 17 EMRs were performed from 134 days to 636 days after the initial CRT.

Analysis showed a significant difference between the 2 treatment groups in terms of en bloc resection rates, with 100% (4/4) in the ESD group compared with 47% (8/17) in the EMR group ($P = .05$), despite the tumor size being significantly larger in the ESD group. Further analysis showed a difference between the 2 groups in terms of resection margin involvement, with 25% (1/4) in the ESD group and 65% (11/17) in the EMR group (not significant). The higher en bloc resection rate and lower incidence of margin involvement in the ESD group compared with the EMR group resulted in a higher curability rate.

It is recognized that ESD for local recurrent or residual tumors is difficult because of fibrosis, which results after CRT. Although it is still not technically feasible to perform either EMR or ESD for an invasive SCC deeper than sm2 (close to the muscle layer), ESD enables us to resect invasive SCC for both sm1 and sm2. Surgical treatment for esophageal SCC is difficult, with a poor quality-of-life reported for patients after surgery, whereas a higher recurrence rate has been reported after CRT treatment. ESD or EMR should be performed initially, therefore, followed by CRT to treat possible lymph node metastasis when EUS or magnifying endoscopy examinations reveal no evidence of deeper invasion to the muscle layer as previously reported.[27]

In conclusion, ESD for recurrent or residual superficial esophageal tumors after CRT with a B-knife or an IT-knife achieves the goal of an en bloc resection with a low rate of incomplete treatment without greater risk than the EMR technique. ESD should be the reference procedure, therefore, for treating such lesions.

ACKNOWLEDGMENTS

We thank Christopher Dix for editing this manuscript and Paul Fockens, MD, for critically reading the manuscript.

DISCLOSURES

The authors report that there are no disclosures relevant to this publication.

REFERENCES

1. Muto M, Nakane M, Katada C, et al. Squamous cell carcinoma in situ at oropharyngeal and hypopharyngeal mucosal sites. Cancer 2004;101:1375-81.
2. Gono K, Obi T, Yamaguchi M, et al. Appearance of enhanced tissue features in narrow-band endoscopic imaging. J Biomed Opt 2004;9:568-77.
3. McCulloch P, Ward J, Tekkis PP, et al. Mortality and morbidity in gastro-oesophageal cancer surgery: initial results of ASCOT multicentre prospective cohort study. BMJ 2003;327:1192-7.
4. Karl RC, Schreiber R, Boulware D, et al. Factors affecting morbidity, mortality, and survival in patients undergoing Ivor Lewis esophagogastrectomy. Ann Surg 2000;231:635-43.
5. Muro K, Arai T, Hamanaka H. Chemoradiotherapy for superficial (sm2/sm3) esophageal cancer: chemoradiotherapy for clinical stage I esophageal cancer [Japanese with English abstract]. Stomach Intestine 2002;37:1305-14.
6. Ohtsu A, Boku N, Muro K, et al. Definitive chemoradiotherapy for T4 and/or M1 lymph node squamous cell carcinoma of the esophagus. J Clin Oncol 1999;17:2915-21.
7. Fiorica F, Di Bona D, Schepis F, et al. Preoperative chemoradiotherapy for oesophageal cancer: a systematic review and meta-analysis. Gut 2004;53:925-30.
8. Balart J, Balmaña J, Rius X, et al. Treatment of oesophageal cancer with preoperative chemoradiotherapy may increase operative mortality. Eur J Surg Oncol 2003;29:884-9.
9. Hattori S, Muto M, Ohtsu A, et al. EMR as salvage treatment for patients with locoregional failure of definitive chemoradiotherapy for esophageal cancer. Gastrointest Endosc 2003;58:65-70.
10. Noguchi H, Naomoto Y, Kondo H, et al. Evaluation of endoscopic mucosal resection for superficial esophageal carcinoma. Surg Laparosc Endosc Percutan Tech 2000;10:343-50.
11. Narahara H, Iishi H, Tatsuta M, et al. Effectiveness of endoscopic mucosal resection with submucosal saline injection technique for superficial squamous carcinomas of the esophagus. Gastrointest Endosc 2000;52:730-4.
12. Inoue H, Kawano T, Tani M, et al. Endoscopic mucosal resection using a cap: techniques for use and preventing perforation. Can J Gastroenterol 1999;13:477-80.
13. Inoue H, Endo M, Takeshita K, et al. A new simplified technique of endoscopic esophageal mucosal resection using a cap-fitted panendoscope (EMRC). Surg Endosc 1992;6:264-5.
14. Gotoda T, Yanagisawa A, Sasako M, et al. Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers. Gastric Cancer 2000;3:219-25.
15. Gotoda T, Sasako M, Ono H, et al. Evaluation of the necessity for gastrectomy with lymph node dissection for patients with submucosal invasive gastric cancer. Br J Surg 2001;88:444-9.
16. Soetikno R, Kaltenbach T, Yeh R, et al. Endoscopic mucosal resection for early cancers of the upper gastrointestinal tract. J Clin Oncol 2005;23:4490-8.
17. Ono H, Kondo H, Gotoda T, et al. Endoscopic mucosal resection for treatment of early gastric cancer. Gut 2001;48:225-9.
18. Fujishiro M, Yahagi N, Kakushima N, et al. En bloc resection of a large semicircular esophageal cancer by endoscopic submucosal dissection. Surg Laparosc Endosc Percutan Tech 2006;16:237-41.
19. Fujishiro M, Yahagi N, Kakushima N, et al. Endoscopic submucosal dissection of esophageal squamous cell neoplasms. Clin Gastroenterol Hepatol 2006;4:688-94.
20. Fujishiro M, Yahagi N, Nakamura M, et al. Successful outcomes of a novel endoscopic treatment for GI tumors: endoscopic submucosal dissection with a mixture of high-molecular-weight hyaluronic acid, glycerin, and sugar. Gastrointest Endosc 2006;63:243-9.
21. Sano Y, Fu KI, Saito Y, et al. A newly developed bipolar-current needle-knife for endoscopic submucosal dissection of large colorectal tumors. Endoscopy 2006;38(Suppl 5):E95.
22. Saito Y, Emura F, Matsuda T, et al. A new sinker-assisted endoscopic submucosal dissection for colorectal tumors. Gastrointest Endosc 2005;62:297-301.
23. Saito Y, Uraoka T, Matsuda T, et al. A pilot study to assess safety and efficacy of carbon dioxide insufflation during colorectal endoscopic submucosal dissection under conscious sedation. Gastrointest Endosc 2007;65:537-42.
24. Saito Y, Uraoka T, Matsuda T, et al. Endoscopic treatment of large superficial colorectal tumors: a cases series of 200 endoscopic submucosal dissections (with video). Gastrointest Endosc 2007;66:966-73.

25. Uraoka T, Fujii T, Saito Y, et al. Effectiveness of glycerol as a submucosal injection for EMR. Gastrointest Endosc 2005;61:736-40.
26. Minami S, Gotoda T, Ono H, et al. Complete endoscopic closure of gastric perforation induced by endoscopic resection of early gastric cancer using endoclips can prevent surgery (with video). Gastrointest Endosc 2006;63:596-601.
27. Shimizu Y, Kato M, Yamamoto J, et al. EMR combined with chemoradiotherapy: a novel treatment for superficial esophageal squamous-cell carcinoma. Gastrointest Endosc 2004;59:199-204.

Received July 13, 2007. Accepted October 2, 2007.

Current affiliations: Division of Endoscopy (Y.S., H.T., H.S., K.T., C.Y., S.N., T.M.), Division of Pathology (Y.N.), Division of Gastrointestinal Oncology (K.K.), National Cancer Center Hospital, Tokyo, Japan.

Presented at 15th United European Gastroenterology Week, October 27-31, 2007, Paris, France (Gastroenterology 2007;39[Suppl]:A224).

Reprint requests: Yutaka Saito, MD, Division of Endoscopy, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan.

Endoscopic submucosal dissection for cancers of the remnant stomach after distal gastrectomy

Ryuta Takenaka, MD, Yoshiro Kawahara, MD, Hiroyuki Okada, MD, Takao Tsuzuki, MD, Satoru Yagi, MD, Jun Kato, MD, Nobuya Ohara, MD, Tadashi Yoshino, MD, Atsushi Imagawa, MD, Shigeatsu Fujiki, MD, Rie Takata, MD, Masahiro Nakagawa, MD, Motowo Mizuno, MD, Tomoki Inaba, MD, Tatsuya Toyokawa, MD, Kohsaku Sakaguchi, MD

Okayama, Tsuyama, Hiroshima, Takamatsu, and Kan-onji, Japan

Background: Endoscopic submucosal dissection (ESD) of early gastric cancer is less invasive than surgical resection, and if technically feasible, it may result in less long-term morbidity than does incisional surgery. However, ESD is technically difficult in patients who have had a previous distal gastrectomy.

Objective: Our purpose was to retrospectively assess the results of ESD of early gastric cancer in the remnant stomach.

Design: Case series.

Setting and Patients: A total of 31 lesions in 30 patients with early remnant gastric cancer were treated with ESD at Okayama University Hospital, Tsuyama Central Hospital, Hiroshima City Hospital, Kagawa Prefectural Central Hospital, and Mitoyo General Hospital from March 2001 to January 2007.

Intervention: ESD.

Main Outcome Measurements: En bloc resection rate, complete resection rate, operation time, and complications.

Results: En bloc resection and complete resection were achieved in 30 (97%) and in 23 (74%) lesions, respectively. The median operation time required for ESD in the remnant stomach was 113 minutes (range 45-450 minutes). Perforation occurred in 4 (13%). The incidence of delayed bleeding requiring blood transfusion was 0%.

Limitation: Short duration of follow-up.

Conclusions: ESD is feasible in the remnant stomach but has a relatively high complication rate and should only be performed by experienced endoscopists.

Abbreviations: EMR, endoscopic mucosal resection; ESD, endoscopic submucosal dissection.

0016-5107/$32.00
doi:10.1016/j.gie.2007.10.021

Distal gastrectomy for benign disease appears to lead to an increased risk for the development of gastric cancer.[1,2] Although remnant gastric cancers are usually detected at an advanced stage and surgical resection of the total remnant stomach has been accepted for a long time,

The Suppression of *Aurora-A/STK15/BTAK* Expression Enhances Chemosensitivity to Docetaxel in Human Esophageal Squamous Cell Carcinoma

Eiji Tanaka,[1] Yosuke Hashimoto,[1] Tetsuo Ito,[1] Kan Kondo,[1] Motoshige Higashiyama,[1] Shigeru Tsunoda,[1] Cristian Ortiz,[1] Yoshiharu Sakai,[1] Johji Inazawa,[2] and Yutaka Shimada[1]

Abstract

Purpose: We previously reported that the expression of Aurora-A was frequently up-regulated in human esophageal squamous cell carcinoma (ESCC) tissues as well as cell lines and the up-regulation contributed to a poor prognosis. In this study, we assessed the possibility of Aurora-A suppression as a therapeutic target for ESCC using ESCC cell lines.

Experimental Design: We established subclones using vector-based short hairpin RNA (shRNA). Then, we investigated the effect of Aurora-A suppression on proliferation and cell cycle changes *in vitro*. Next, chemosensitivity against docetaxel was investigated by tetrazolium salt–based proliferation assay (WST assay) and cell number determinations, and furthermore, the type of cell death induced by docetaxel was analyzed by flow cytometry. Finally, to examine the effect of Aurora-A shRNA on proliferation and chemosensitivity against docetaxel *in vivo*, a s.c. tumor formation assay in nude mice was done.

Results: We established two genetically different stable cell lines (510 A and 1440 A) in which levels of Aurora-A were reduced. Cell growth was inhibited by 38.7% in 510 A and by 24.3% in 1440 A *in vitro* compared with empty vector–transfected controls (510 m and 1440 m), and this growth inhibition was mediated through G_2-M arrest as confirmed by flow cytometry. Next, in a WST assay, the IC_{50} for Aurora-A shRNA-transfected cells was lower than that of empty vector–transfected cells (510 A, 2.7×10^{-7} mol/L; 510 m, 4.8×10^{-7} mol/L; 1440 A, 2.6×10^{-7} mol/L; 1440 m, 4.9×10^{-7} mol/L). In addition, 0.3 nmol/L docetaxel induced a notable level of apoptosis in Aurora-A shRNA-transfected cells compared with empty vector–transfected cells. In the assay of s.c. tumors in nude mice, tumor growth in 510 A was inhibited by 36.1% compared with that in 510 m, and in tumors treated with docetaxel, the suppression of Aurora-A resulted in 44.0% tumor growth suppression *in vivo*.

Conclusions: These results indicated that Aurora-A might play an important role in chemosensitivity to docetaxel, and the suppression of its expression might be a potential therapeutic target for ESCC.

Several proteins strictly regulate the process of cellular division. Defects in chromatid segregation cause genetic instability, a condition associated with tumorigenesis. During the proliferation of normal cells, the centrosome ensures the equal segregation of chromosomes to the postmitotic daughter cells by organizing the bipolar mitotic spindle. In contrast, in cancer cells, multipolar mitotic spindles and various centrosomal anomalies, such as supernumerary centrosomes, centrosomes of abnormal size and shape, aberrantly phosphorylated centrosomic proteins, and prematurely split centrosomes, are frequently observed (1–7).

Aurora-A, a member of the Aurora/Ipl1p family of cell cycle–regulating serine/threonine kinases, is expressed at interphase mitotic centrosomes and the spindle poles in the nucleus where it regulates segregation of chromosomes and cytokinesis. Recent studies have shown that the ectopic expression of Aurora-A in mouse NIH/3T3 cells and Rat 1 fibroblasts causes centrosome amplification and transformation *in vitro* as well as tumorigenesis *in vivo* (8, 9). Furthermore, the up-regulation of Aurora-A expression in diploid human breast epithelial cells leads to abnormal numbers of centrosomes and the induction of aneuploidy (8). A correlation between the up-regulated expression of Aurora-A and clinical aggressiveness has also been reported for several cancers (10–14). Moreover, recent reports showed that the up-regulation of Aurora-A resulted in resistance

Authors' Affiliations: [1]Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan and [2]Department of Molecular Cytogenetics, Graduate School of Biomedical Science, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Received 5/16/06; revised 10/26/06; accepted 11/6/06.

Grant support: Japanese Ministry of Education, Culture, Sports, Science, and Technology Grant-in-Aid 14370385.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Requests for reprints: Yutaka Shimada, Department of Surgery, Graduate School of Medicine, Kyoto University, Kawaracho 54, Shogoin Sakyo-ku, Kyoto 606-8507, Japan. Phone: 81-75-751-3626; Fax: 81-75-751-4390; E-mail: shimada@kuhp.kyoto-u.ac.jp.

doi:10.1158/1078-0432.CCR-06-1192

to apoptosis induced by taxanes in a human cancer cell line (15, 16) and inhibition of the expression of Aurora-A resulted in potent antitumor activity and chemosensitizing activity to taxanes in pancreatic cancer (17). These findings suggest that *Aurora-A* is a critical kinase-encoding gene and a potential chemotherapeutic target.

Esophageal squamous cell carcinoma (ESCC) is one of the most lethal malignancies in the world, including Japan, despite the recent refinement of various therapeutic strategies, including surgery, chemotherapy, radiotherapy, and combined therapy (18, 19). Docetaxel is used as monotherapy or in combination with other agents to treat ESCC, but its activity is far from satisfactory (20, 21). Therefore, identifying and targeting genes conducive to the treatment of ESCC, such as enhancement of conventional chemotherapy, is necessary to improve the survival of patients with this type of refractory cancer.

Using comparative genomic hybridization, we previously investigated changes in the copy number of chromosomes in 29 ESCC cell lines and found that a chromosome gain of the proximal part of 20q, where the *Aurora-A* gene is located, is one of the most common sites of aberrations (19 of 29, 65.5%; ref. 22). We also reported that the expression of Aurora-A was frequently up-regulated in ESCC tissues as well as cell lines, and this contributed to a poor prognosis (14). In the present study, to further elucidate the possibility of using *Aurora-A* in the treatment of human ESCC, we analyzed the phenotypic changes of cultured ESCC cells induced by suppression of Aurora-A expression using a plasmid vector-mediated short hairpin RNA (shRNA) expression system, especially synergistic enhancement of the cytotoxicity of docetaxel.

Materials and Methods

Cell culture. All tested ESCC cell lines of the KYSE series were established in our laboratory and maintained in RPMI 1640 (Life Technologies, Gaithersburg, MD) and Ham's F12 (Nissui Pharmaceutical, Tokyo, Japan) mixed (1:1) medium containing 2% fetal bovine serum (23). HeLa cells were purchased from the American Type Culture Collection (Rockville, MD), cultured in DMEM (Life Technologies) with 10% FCS, and used as a positive control (24, 25).

Western blot analysis. Cells were lysed in lysis buffer [50 mmol/L Tris-HCl (pH 7.5), 150 mmol/L NaCl, 5 mmol/L EDTA, 1% Triton X-100] containing Complete Mini protease inhibitor (Roche Diagnostics, Mannheim, Germany). Cell lysates were sonicated and the protein concentration was estimated by the Bradford method using bicinchoninic acid protein assay reagent (Pierce, Rockford, IL). Cell lysates (15 μg) were electrophoresed on 2% to 15% gradient polyacrylamide gel (Daiichi Pure Chemicals, Tokyo, Japan) and transferred to polyvinylidene difluoride membranes (Millipore, Bedford, MA) using a semidry transfer blot system (Bio-Rad, Hercules, CA). After blocking with TBS containing 1% Tween 20 and 5% skim milk for 1 h, the membranes were incubated at 4°C overnight with anti-human Aurora-A polyclonal antibody (diluted 1:100; TransGenic, Inc., Kumamoto, Japan) or anti-human β-actin monoclonal antibody (1:2,000; Sigma, Inc., St. Louis, MO). The membranes were subsequently incubated at room temperature for 1 h with secondary antibody and analyzed using enhanced chemiluminescence plus reagent (Amersham, Buckinghamshire, United Kingdom). Quantitative analysis was done on a Macintosh computer using the public domain NIH Image program version 1.61 (developed at the NIH and available on the Internet).[3]

Immunofluorescent staining. 510 A, 510 m, 1440 A, and 1440 m cells were cultured onto collagen-coated glass coverslips (BD Biosciences, Bedford, MA). Then, cells were washed twice with PBS and fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. The fixed cells were incubated with 0.3% Triton X-100 solution for 15 min at room temperature, and the cells were blocked for 1 h with 2% fetal bovine serum in PBS at room temperature. Subsequently, the cells were incubated with anti-human Aurora-A polyclonal antibody (diluted 1:50; TransGenic) for 1 h at room temperature. After washing twice with PBS, the cells were incubated with goat anti-rabbit FITC-conjugated secondary antibody (Invitrogen Co., Carlsbad, CA) for 1 h at room temperature. The cells were washed and mounted in glycerol and viewed under a laser scanning microscope (Axioskop 2 plus, Carl Zeiss Co. Ltd., Jena, Germany).

Construction of an Aurora-A shRNA expression vector. To construct a vector for Aurora-A shRNA, the pSUPERIOR-puro (OligoEngine, Seattle, WA) was digested with *Bgl*II and *Hin*dIII (TaKaRa Bio, Shiga, Japan). A chemically synthesized oligonucleotide encoding an Aurora-A short hairpin small interfering RNA, including a loop motif, was inserted downstream of the H1 promoter of the plasmid using a DNA ligation kit (TaKaRa Bio) and cloned. The sequence of the oligonucleotide targeted at Aurora-A is 5′-ATGCCCTGTCTTACTGTCA-3′ for KYSE 510, corresponding to positions 853 to 871 within the Aurora-A mRNA sequence (17). To confirm the result, we designed another sequence for KYSE 1440, 5′-GCCGGTTCAGAATCAGAAG-3′, corresponding to positions 335 to 353 within the Aurora-A mRNA. For the negative control vector, an empty pSUPERIOR-puro was used. We checked the internal stabilities of each sequences using Oligo 4.0-s software (National Biosciences, Inc., Plymouth, MN) and found that the 3′-end in the sense strand was less stable than the 5′-end, which is known to cause less off-target effects according to literature (26). Furthermore, we checked each sequence using BLAST Web site software[4] and found that each sequence was specific to *Aurora-A*. Moreover, we also stably transfected each sequence to HeLa cell, and preliminary experiments were undertaken to further confirm their effects (Supplementary Data 1).

Transfections. The ESCC cell lines KYSE 510 and KYSE 1440 were stably transfected with the Aurora-A shRNA expression vector or the empty pSUPERIOR-puro vector using Lipofectin reagent (Invitrogen) as suggested by the manufacturer's instruction. Briefly, 2 μg of each plasmid DNA and 20 μL of Lipofectin reagent together with Opti-MEM I medium (Invitrogen) were used with serum-free medium for 8 h. Cells were incubated for another 48 h with normal growth medium, and subsequently, the cell clones were selected against 1.0 μg/mL puromycin (Nacalai Tesque, Kyoto, Japan) for 3 weeks, and we picked up single colonies originated from single cells and expanded to obtain stably transfected cell lines.

WST assay for sensitivity to doxetaxel. Cytotoxic activity against docetaxel was measured by the tetrazolium salt–based proliferation assay (WST-8 assay; Wako Chemicals, Osaka, Japan) following the manufacturer's instructions. Briefly, cells were cultured in 96-well microtiter plates in 90 μL of growth medium (4,000 cells per well) and incubated for 24 h for sufficient cell growth. Then, 10 μL of a graded concentration of docetaxel (10^{-5} to 10^{-10} mol/L) were added into each well and cultured for 48 h. Control cultures received normal growth medium only. After 48 h, 10 μL of WST-8 solution were added to each well and the plates were incubated at 37°C for another 3 h. Absorbance at 450 and 640 nm was measured using the Delta Soft ELISA analysis program, and cell viability was measured and compared with that of control cells. Each experiment was carried out independently and repeated at least thrice. The IC_{50} value was defined as the concentration that reduced the absorbance in each test by 50%.

Cell proliferation assay. Cells were cultured in 6-cm dishes (2×10^4 per dish) and incubated for 24 h for sufficient cell growth and then treated with medium containing 0.3, 0.6, or 1.0 nmol/L of docetaxel for

[3] http://rsb.info.nih.gov/nih-image/

[4] http://www.ddbj.nig.ac.jp/search/blast-j.html

48 h. Then, they were cultured for another 48 h with normal growth medium. Control cultures received normal growth medium only. Cells were harvested with trypsin/EDTA every 48 h for 4 days and enumerated using a cell counter (Coulter Z1, Beckman Coulter, Fullerton, CA). A comparison was made with the control culture to examine the effect of suppressing the expression of Aurora-A on cell proliferation and to investigate the effect of the suppression on chemosensitivity to docetaxel. The experiment was repeated at least thrice.

Flow cytometry for analyzing the cell cycle. A flow cytometric analysis of DNA content was done to assess the cell cycle phase distribution. Cells were harvested at the 70% confluent stage and fixed in 70% ethanol at −20°C. After being washed with PBS, the cells were treated with PBS containing RNase A (100 mg/mL) at 37°C for 30 min. After centrifugation, the cells were resuspended in PBS containing propidium iodide (50 μg/mL) and stained at room temperature for 30 min. DNA content was evaluated using a FACScan flow cytometer (Becton Dickinson, Mountain View, CA) and CellQuest software (Becton Dickinson). The experiment was repeated thrice.

Flow cytometry for detection of apoptotic cells. KYSE cells transfected with empty vector or shRNA against Aurora-A were cultured in 6-cm dishes (2×10^4 per dish) and treated as already described for the proliferation experiment. Briefly, cells were incubated for 24 h and treated with medium containing 0, 0.3, 0.6, or 1.0 nmol/L of docetaxel for 48 h. Then, they were cultured for another 48 h with normal growth medium. Subsequently, floating cells in the medium and adherent cells were collected. Using an Annexin V-FITC Apoptosis Detection kit (Medical & Biological Laboratories Co. Ltd., Woburn, MA), cells were stained with Annexin V-FITC and propidium iodide according to the manufacturer's instructions. Annexin V binds to cells that express phosphatidylserine on the outer layer of the cell membrane, and propidium iodide stains the cellular DNA of cells with a compromised cell membrane. This allows for the discrimination of live cells (unstained) from early apoptotic cells (stained only with Annexin V) and late apoptotic or necrotic cells (stained with both Annexin V and propidium iodide). Untreated cells and cells treated with 3% formaldehyde for 30 min served, respectively, as negative and positive controls for double staining. Cells were analyzed immediately after staining by using a FACScan flow cytometer and the CellQuest software. For each measurement, >10,000 cells were counted.

Tumor formation assay in nude mice. Suspensions of 1.0×10^6 KYSE 510-derived cells (Aurora-A shRNA-transfected cells, 510 A; empty vector-transfected cells, 510 m) in PBS (50 μL) were injected s.c. into the left flanks of 5-week-old male BALB/c *nu/nu* mice (Japan SLC, Shizuoka, Japan) at day 0. The inoculation was conducted in five mice, and tumor growth was estimated from the average volume of tumors. Tumor volume was calculated by the formula $1/2 \times L^2 \times W$ (L = length and W = width of the tumor). At 46 days after inoculation, all mice were sacrificed, and s.c. tumors were resected and fixed in 10% formaldehyde/PBS. The tumors were paraffin embedded and stained with H&E and for Aurora-A. Immunohistochemical staining for Aurora-A was done as reported previously (14). All the animal experiments were done in accordance with institutional guidelines of the Kyoto University.

Treatment with docetaxel in nude mice. 510 m and 510 A tumors were generated as above. Briefly, suspensions of 1.0×10^6 cells in PBS (50 μL) were injected s.c. into the left flanks of 5-week-old male BALB/c *nu/nu* mice at day 0. The inoculation was conducted in five mice, and mice were treated with docetaxel (7.0 mg/kg body weight; i.p., thrice weekly) or with 0.1 mL PBS (pH 7.4; i.p., thrice weekly) as described previously with modifications (27). Tumor growth was estimated from the average volume of tumors, and tumor volume was calculated as already described. At 39 days after inoculation, all mice were sacrificed, and s.c. tumors were resected, and tumor size was compared. All the animal experiments were done in accordance with institutional guidelines of the Kyoto University.

Statistical analysis. All experiments were done in duplicate or triplicate. The Bonferroni multiple comparison test and t test were used for the statistical analysis of comparative data using StatView version 5 (SAS Institute, Cary, NC). Values of $P < 0.05$ were considered significant and are indicated by asterisks in the figures.

Results

Expression of Aurora-A and sensitivity to docetaxel in ESCC cell lines. Levels of Aurora-A protein expression in three cancer cell lines were determined by Western blotting (Fig. 1A). KYSE 510 and KYSE 1440 had higher levels of expression than KYSE 110. Next, the sensitivity of the cell lines to docetaxel was checked by WST assay. As shown in Fig. 1B, IC_{50} of KYSE 510 and KYSE 1440 ($4.9 \times 10^{-7} \pm 0.20 \times 10^{-7}$ mol/L and $4.6 \times 10^{-7} \pm 0.46 \times 10^{-7}$ mol/L, respectively) was higher than that of KYSE 110 ($2.4 \times 10^{-8} \pm 0.21 \times 10^{-8}$ mol/L; Fig. 1B). Thus, the cell lines with the higher levels of Aurora-A protein (KYSE 510 and KYSE 1440) were more resistant to docetaxel than the cell line with the lower level (KYSE 110).

Vector-based Aurora-A shRNA decreased proliferation in ESCC cell lines. To assess the role of the overexpression of Aurora-A in ESCC cells, we first established subclones via the transfection of vector-based shRNA for Aurora-A in KYSE 510 and KYSE

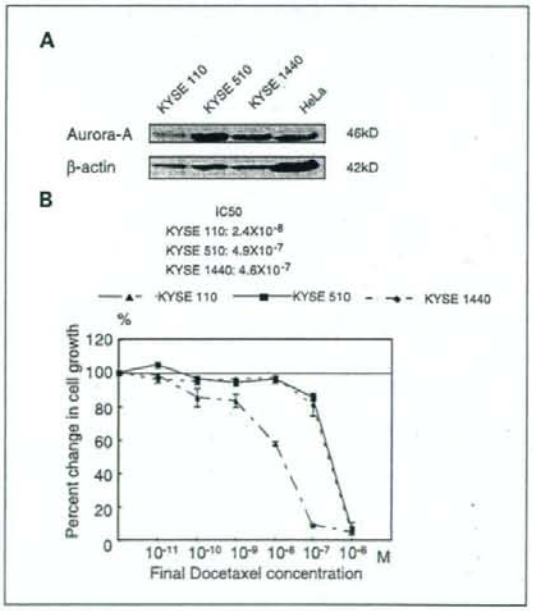

Fig. 1. Aurora-A expression and chemosensitivity to docetaxel in ESCC cell lines. *A*, expression of Aurora-A protein in ESCC cell lines. Immunoblots were probed with the anti-Aurora-A antibody and the anti-β-actin antibody. The HeLa cell line was included as a positive control of Aurora-A expression. *B*, cytotoxic activities of docetaxel in ESCC cell lines were measured by WST-8 assay. Graded concentrations of docetaxel (10^{-6} to 10^{-10} mol/L) were added to the wells, and the cells were cultured for 48 h. Control cultures received normal growth medium only. Cell viability was measured as absorbance at 450 nm, and values were compared with the control. The IC_{50} value was defined as the concentration that reduced the absorbance in each test by 50%. X axis, final docetaxel concentration (10^{-6} to 10^{-11} mol/L). The experiments were repeated thrice.

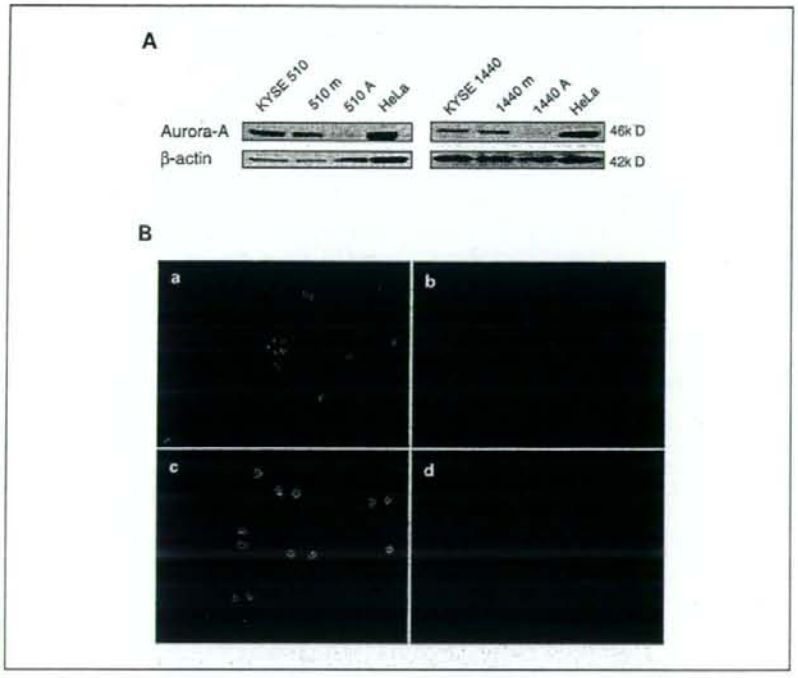

Fig. 2. Vector-based Aurora-A shRNA decreased proliferation of ESCC cell lines. *A*, suppression of Aurora-A protein expression with the Aurora-A shRNA vector. The KYSE 510 and KYSE 1440 cells were stably transfected with an empty vector (510 m or 1440 m) or an Aurora-A shRNA vector (510 A or 1440 A). The expression of Aurora-A protein in these clones was examined by Western blotting. Bottom, protein expression of β-actin. *B*, representative immunofluorescent staining of Aurora-A in an Aurora-A shRNA-transfected cells (510 A or 1440 A) or an empty vector–transfected cells (510 m or 1440 m). a, immunofluorescent staining of Aurora-A in empty vector–transfected cells (510 m). Magnification, ×200. Strong nuclear staining was observed. b, immunofluorescent staining of Aurora-A shRNA-transfected cells (510 A). Magnification, ×200. Weak nuclear staining was observed. c, immunofluorescent staining of Aurora-A in empty vector–transfected cells (1440 m). Magnification, ×200. Strong nuclear staining was observed. d, immunofluorescent staining of Aurora-A in Aurora-A shRNA-transfected cells (1440 A). Magnification, ×200. Weak nuclear staining was observed.

1440 cells. The levels of Aurora-A expression were efficiently reduced by 87% in the stable subclone 510 A and by 90% in 1440 A but were not reduced in empty vector–transfected clones (510 m and 1440 m; Fig. 2A). We also confirmed suppression of Aurora-A expression by immunofluorescent staining and found that Aurora-A immunoreactivity in established stable subclone 510 A and 1440 A was homogeneously depleted (Fig. 2B). We investigated the effect of suppressing the expression of Aurora-A on proliferation in ESCC cell lines. With suppression of Aurora-A expression, cell growth was inhibited by 38.7 ± 10.5% in 510 A and by 24.3 ± 10.0% in 1440 A compared with empty vector–transfected cells (510 m and 1440 m; $P < 0.01$ and 0.01, respectively; Fig. 2C). To investigate the growth suppression caused by inhibition of Aurora-A expression, cell cycle changes in 510 A and 1440 A as well as in empty vector–transfected cells were examined by flow cytometry. The population of cells in G_2-M phase was significantly larger in 510 A (41.13 ± 6.94%) and 1440 A (38.35 ± 1.06%) than that in the empty vector–transfected cells (510 m: 24.12 ± 2.33%, $P < 0.01$; 1440 m: 22.55 ± 2.25%, $P < 0.01$; Fig. 2D).

The effect of suppressing Aurora-A on chemosensitivity to docetaxel **in vitro.** Chemosensitivity to docetaxel was investigated by WST assay and cell number determinations. In the WST assay, we found that IC_{50} of stable transfectants of shRNA for Aurora-A was lower than that of the empty vector transfectant (510 A, $2.7 \times 10^{-7} \pm 0.4 \times 10^{-7}$ mol/L; 510 m, $4.8 \times 10^{-7} \pm 0.7 \times 10^{-7}$ mol/L; 1440 A, $2.6 \times 10^{-7} \pm 0.8 \times 10^{-7}$ mol/L; 1440 m, $4.9 \times 10^{-7} \pm 0.4 \times 10^{-7}$ mol/L; Fig. 3A and B). We also confirmed the effect of suppressing Aurora-A expression on chemosensitivity to docetaxel by cell number determinations. In the cell number determinations, the suppression of the expression of Aurora-A allowed even 0.3 nmol/L docetaxel to be effective (Fig. 3C and D). Discrepancy in effective concentration of docetaxel between WST assay and cell number determinations could be explained by the difference in the initial numbers of cells, cell concentration at the beginning of each experiment, as well as the difference in size of culture plates and observation period. We repeated the cell number determinations with the same cell concentration and observation period as WST assay and confirmed that growth-inhibitory effect of docetaxel in these two different experiments was similar (Supplementary Data 2). Importantly, in respective experiments, we found that the suppression of Aurora-A expression enhanced chemosensitivity to docetaxel. To investigate the effect of chemosensitivity to docetaxel by the suppression of Aurora-A expression, apoptotic cells in the 510 A and 1440 A clones as well as empty vector–transfected clones were examined by flow cytometry and the type of cell death induced by docetaxel was assessed. In the Aurora-A shRNA-transfected cells, 0.3 nmol/L docetaxel induced a notable level of apoptosis compared with the empty vector–transfected cells (510 A, 71.3%; 510 m, 36.12%; 1440 A, 78.1%; 1440 m, 26.28%; Fig. 3E and F). These findings suggested that the suppression of Aurora-A expression augmented the apoptosis induced by docetaxel. To confirm the validity of the experiment, cells treated with 3% formalin for 30 min and stained with both Annexin V and propidium iodide served as a positive control for double staining (Supplementary Data 3).

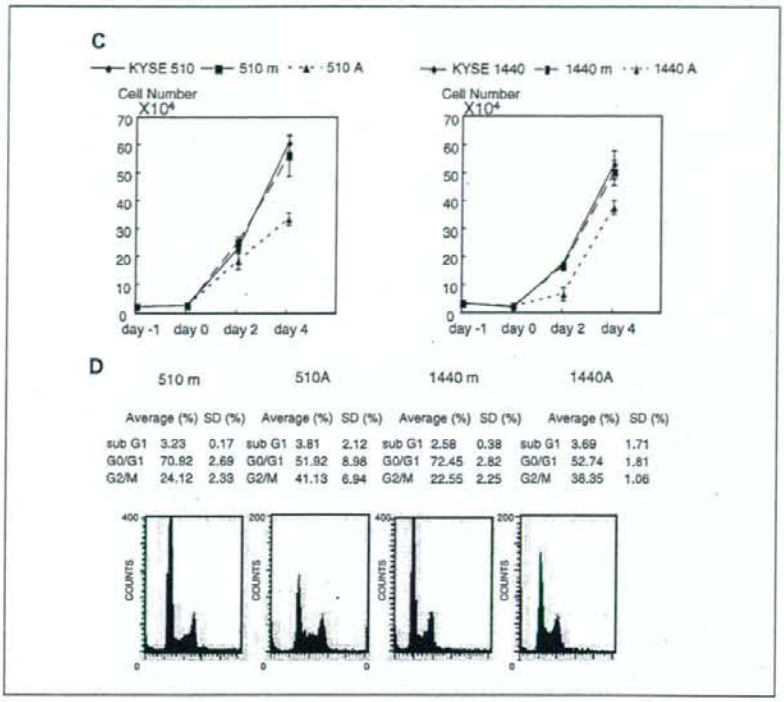

Fig. 2 *Continued. C, in vitro* growth assay in cells transfected with the Aurora-A shRNA vector (510 A or 1440 A) or an empty vector (510 m or 1440 m). The number of cells was counted every 2 d. The experiments were repeated thrice. *D*, cell cycle profiles obtained by flow cytometry. Ratios of cell populations in G_1 and G_2-M were the average and SD. These experiments were done thrice, and representative results are shown. The experiments were repeated thrice, and the representative results are shown.

The effect of Aurora-A suppression on tumor growth **in vivo.** To examine the possible effect of Aurora-A shRNA on tumor growth *in vivo*, s.c. tumor formation assay in nude mice was done. As shown in Fig. 4A, the tumors formed from Aurora-A shRNA-transfected cells (510 A) were clearly smaller than that formed from empty vector–transfected cells (510 m). Then, we examined tumor volume and weight at 46 days after inoculation. The average tumor volume of the mice at day 46 was reduced by 36.1% in 510 A compared with 510 m ($P = 0.03$; Fig. 4C), and the average tumor weight was also decreased by 31.5% in 510 A compared with 510 m ($P = 0.03$; Fig. 4C). All of the tumors were stained with H&E and for Aurora-A (Fig. 4D), and Aurora-A immunoreactivity was confirmed to be reduced in tumors grown from Aurora-A shRNA-transfected cells but not in tumors grown from empty vector–transfected cells.

The effect of Aurora-A suppression on chemosensitivity to docetaxel **in vivo.** To examine the possible effect of Aurora-A shRNA on chemosensitivity to docetaxel *in vivo*, s.c. tumors were generated in nude mice followed by treatment with docetaxel or PBS. The tumors formed from Aurora-A shRNA-transfected cells (510 A) were apparently smaller than those formed from empty vector–transfected cells (510 m) after the treatment with docetaxel at day 39 (Fig. 5A, *b* and *d*). To confirm this, we measured tumor volume at 39 days after inoculation. As shown in Fig. 5C, following the treatment with docetaxel, the average tumor volume of 510 A and 510 m was 175.3 ± 70.4 mm^3 and 312.8 ± 28.0 mm^3, respectively, and thus, the suppression of Aurora-A expression resulted in a 44.0% inhibition of tumor growth ($P = 0.03$).

Discussion

We previously reported that the expression of Aurora-A was frequently up-regulated in ESCC tissues and cell lines and contributed to a poor prognosis (14). In an attempt to determine the potential of Aurora-A as a therapeutic target, we used a vector-based shRNA technique to knock out its expression and analyzed its phenotypes.

In the current study, we were able to suppress the expression of Aurora-A using vector-based shRNA for two different target sequences in two different ESCC cell lines and obtained very similar results at each examination *in vitro*. That is, a reduction of Aurora-A protein expression was clearly related to cell growth inhibition and increased sensitivity to docetaxel.

Recently, a relationship between Aurora-A activity and G_2-M transition was reported in cancer cells (28). In the current study, we showed that the suppression of Aurora-A expression caused an accumulation of the cells in the G_2-M phase *in vitro*, resulting in the inhibition of proliferation of ESCC cell lines. Moreover, in recent report, Hata et al. (17) showed that the suppression of Aurora-A expression had an antitumorigenic effect *in vivo* in pancreatic cancer, and we similarly showed that the suppression of Aurora-A expression in ESCC cell lines inhibited tumor growth *in vivo* using our vector-mediated shRNA strategy. Our results were consistent with Hata et al. in

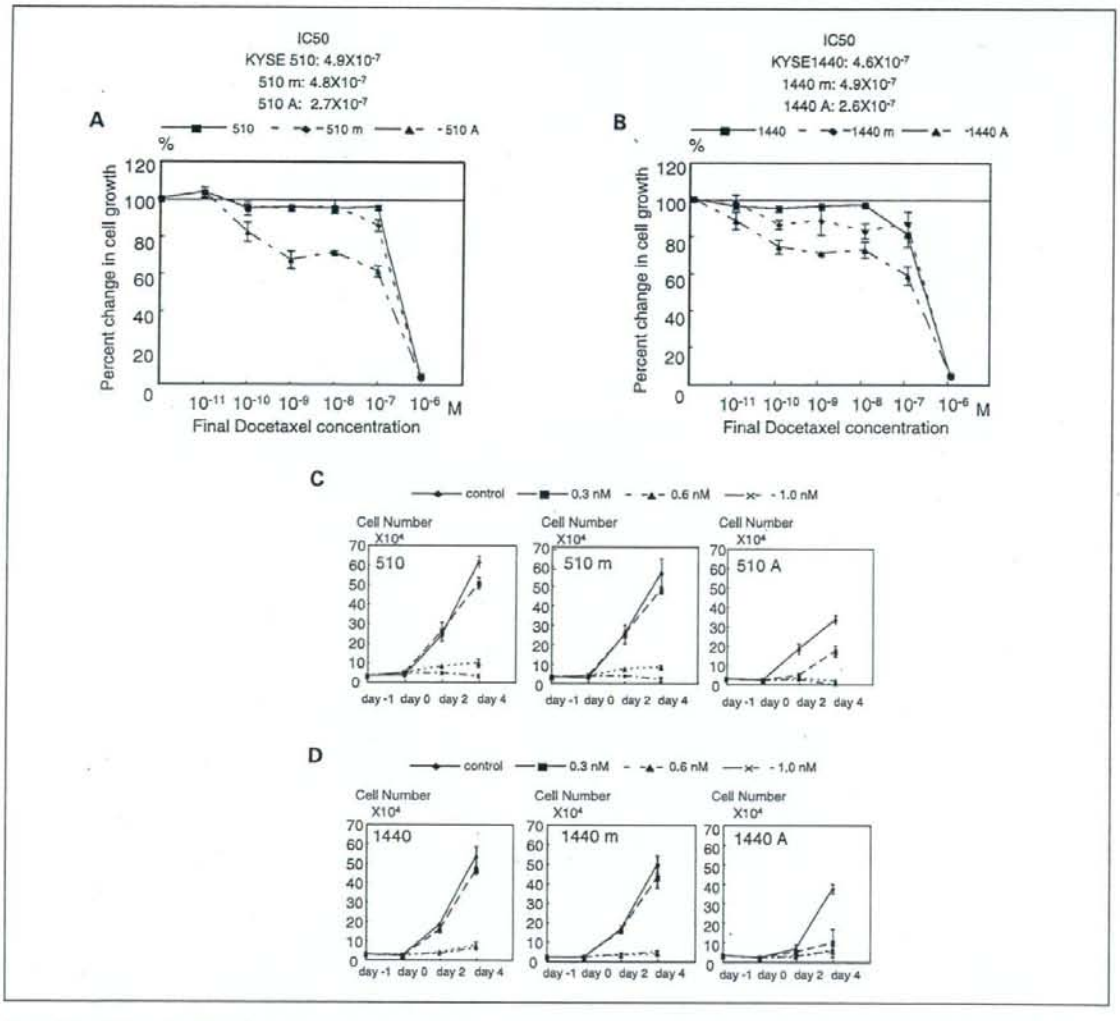

Fig. 3. Effect of the suppression of Aurora-A expression on sensitivity to docetaxel. *A* and *B*, cytotoxic activities of docetaxel were measured by WST-8 assay. Graded concentrations of docetaxel (10^{-5} to 10^{-10} mol/L) were added to the wells, and the cells were cultured for 48 h. Control cultures received normal growth medium only. Cell viability was measured as absorbance at 450 nm, and values obtained were compared with that of the control. The IC_{50} value was defined as the concentration that reduced the absorbance in each test by 50%. X axis, final docetaxel concentration (10^{-6} to 10^{-11} mol/L). The experiments were repeated thrice. *C* and *D*, cytotoxic activities of docetaxel were confirmed by cell counting. Graded concentrations of docetaxel (0, 0.3, 0.6, and 1.0 nmol/L) were added for the first 2 d, and normal growth medium was added for the next 2 d. The number of cells was counted every 2 d. These experiments were repeated thrice.

terms of antitumor activity that the suppression of Aurora-A expression caused in ESCC. Consequently, Aurora-A seemed to be a critical factor for the proliferation of cancer cells, and therefore, it should be a good therapeutic target for halting proliferation of ESCC.

Furthermore, we found that the suppression of Aurora-A expression enhanced the sensitivity to docetaxel-induced apoptosis both *in vitro* and *in vivo*. Taxanes bind to free tubulin and promote the assembly of tubulin into stable microtubules. Hence, they stop cell cycle progression, causing cells in the M phase to accumulate at the metaphase-anaphase transition and subsequently leading them to apoptosis, which is consistent with our findings that the suppression of Aurora-A expression resulted in the accumulation of cells in the G_2-M phase. As previously discussed, Aurora-A is essential for the proper arrangement of centrosomes and microtubules, and Hata et al. showed that a combination of the suppression of Aurora-A expression and use of taxanes resulted in an augmented induction of apoptosis in pancreas cancer *in vitro* (17). Additionally, Anand et al. (15) showed that the overexpression of Aurora-A induced increased resistance to taxanes via a decrease in spindle checkpoint activity *in vitro*. Our results

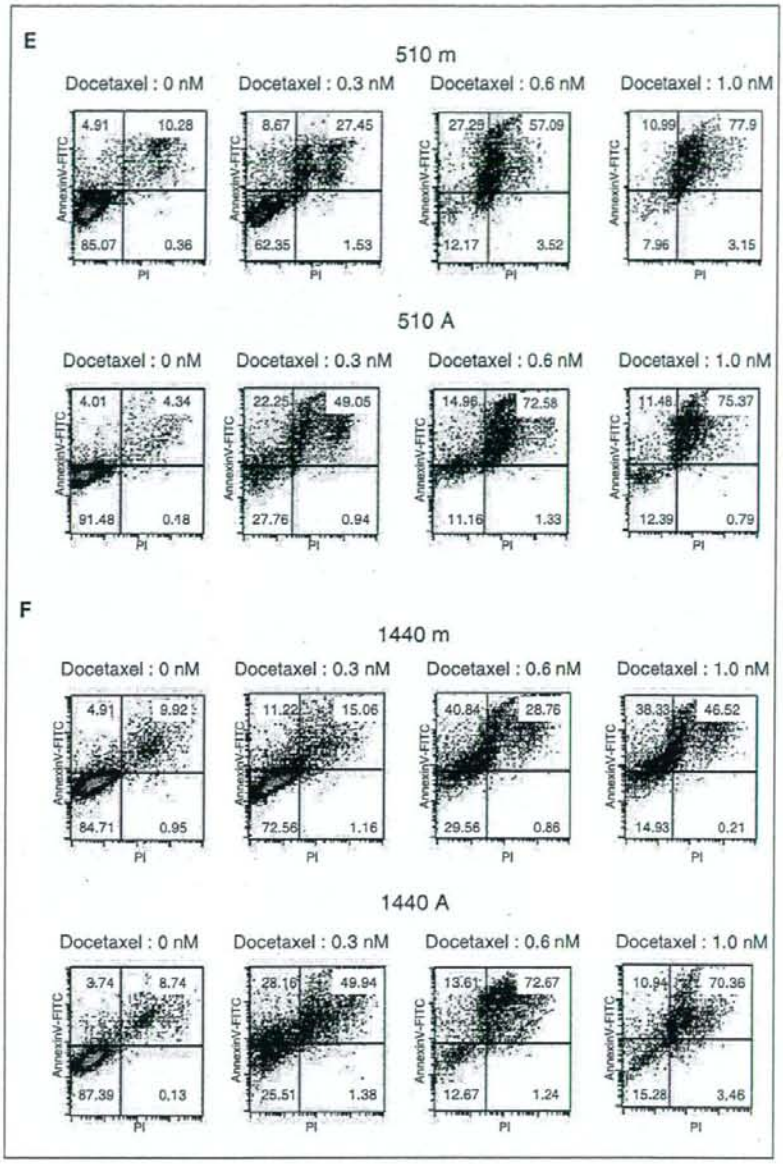

Fig. 3 *Continued.* *E* and *F*, to assess the type of cell death induced by docetaxel, flow cytometry was done. Cells in bottom left quadrant (unstained), top left quadrant (stained only with Annexin V), and top right quadrant [stained with both Annexin V and propidium iodide (*PI*)] represent viable cells, cells in early apoptosis, and cells in late apoptosis, respectively. The experiments were repeated thrice, and the representative results are shown.

are consistent with this study and were able to show that the suppression of Aurora-A expression enhanced chemosensitivity to docetaxel not only *in vitro* but also *in vivo*. These results suggest that, by suppressing Aurora-A expression, spindle checkpoint activity might have recovered and, thus, increased the sensitivity to taxanes. In the meantime, the mechanism that triggers apoptosis after inhibition of Aurora-A expression, as well as the complicated biological activity of Aurora-A, remains to be clarified.

RNA interference has become conventional applications for *in vivo* cancer therapy (29, 30), and an efficient way of delivering small interfering RNA into solid tumors has been developed (31). In the current study, we explored the possibility that the RNA interference–mediated suppression of Aurora-A could be used as a specific gene-targeting therapy to suppress the progression of ESCC. Moreover, the function of Aurora kinase inhibitors (including the patent literature) has been studied recently, revealing potentially promising anticancer

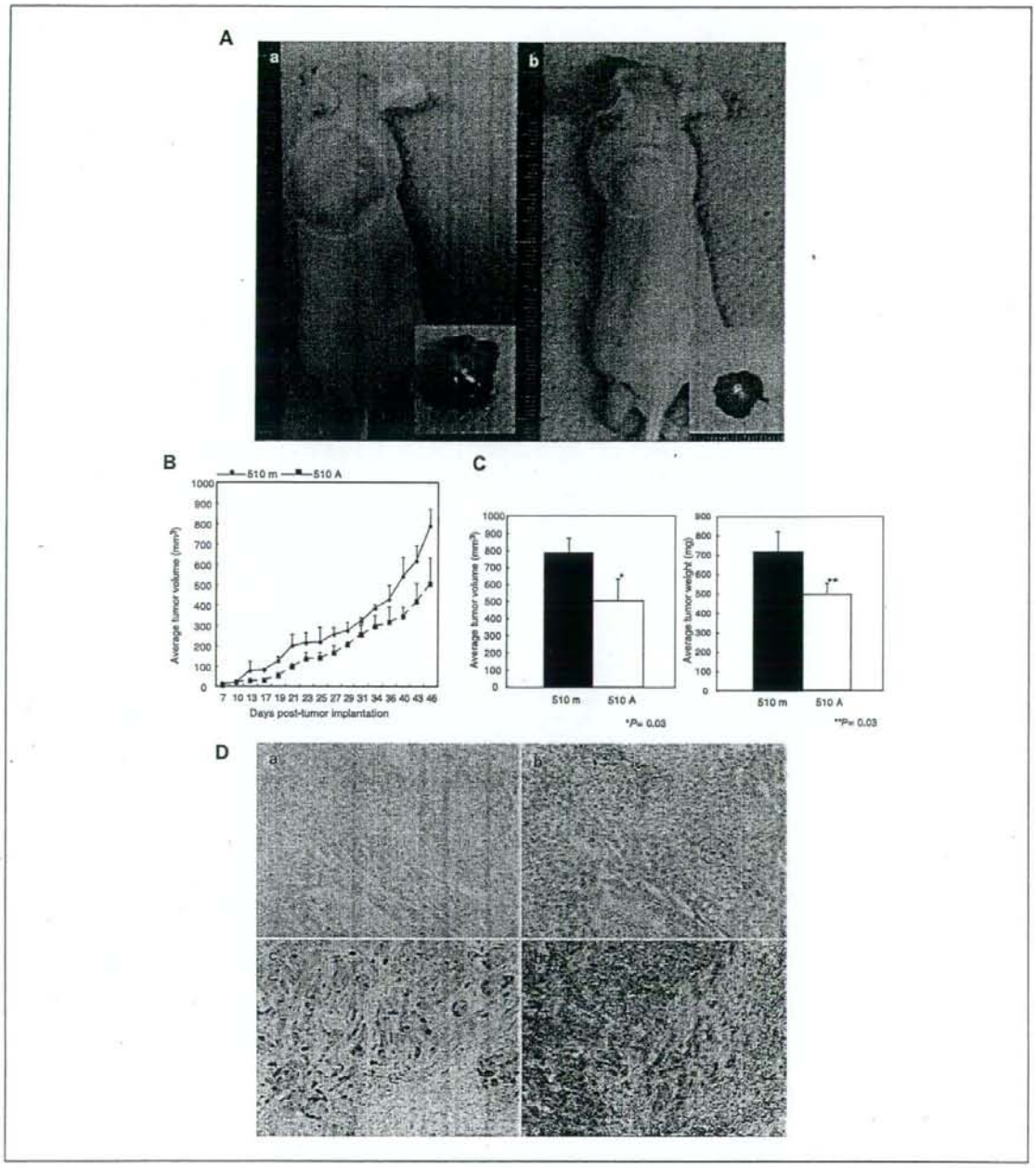

Fig. 4. Effect of the down-regulation of Aurora-A expression on the formation of tumors *in vivo*. *A*, representative features of tumors in a mouse 46 d after the inoculation. a, tumors formed from empty vector – transfected cells (510 m) in the left flank; b, tumors formed from Aurora-A shRNA-transfected cells (510 A) in the left flank. *B*, growth of tumors in the mice injected with Aurora-A shRNA-transfected cells (510 A) or empty vector – transfected cells (510 m). The inoculation was done in five mice. *C*, tumor volume and weight at day 46 after inoculation. Left, black column, average tumor volume at day 46 after the inoculation of empty vector – transfected cells (510 m); white column, average tumor volume at day 46 after the inoculation of Aurora-A shRNA-transfected cells (510 A; $n = 5$). *, $P = 0.03$. Right, black column, average tumor weight at day 46 after the inoculation of empty vector – transfected cells (510 m); white column, average tumor weight at day 46 after the inoculation of Aurora-A shRNA-transfected cells (510 A; $n = 5$). **, $P = 0.03$. *D*, a, immunohistochemical staining of Aurora-A in s.c. tumors at day 46 after the inoculation of Aurora-A shRNA-transfected cells (510 A). Magnification, ×200. b, H&E staining of s.c. tumors at day 46 after the inoculation of Aurora-A shRNA-transfected cells (510 A). Magnification, ×200. c, immunohistochemical staining of Aurora-A in s.c. tumors at day 46 after the inoculation of empty vector – transfected cells (510 m). Magnification, ×200. d, H&E staining of s.c. tumors at day 46 after the inoculation of empty vector – transfected cells (510 m). Magnification, ×200.

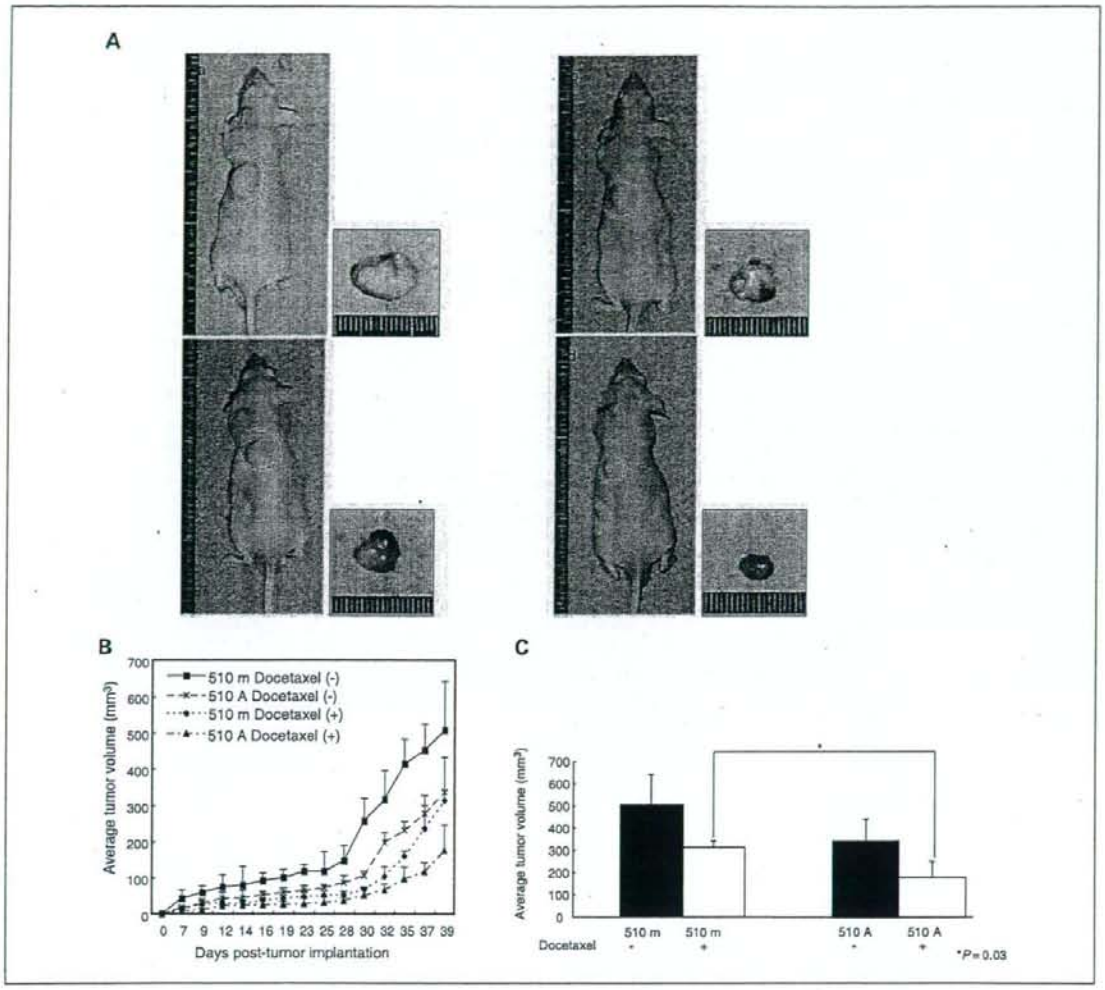

Fig. 5. Effect of the suppression of Aurora-A expression on sensitivity to docetaxel *in vivo*. Mice were treated with docetaxel (7.0 mg/kg body weight; i.p., thrice weekly) or with 0.1 mL PBS (pH 7.4; i.p., thrice weekly). *A*, representative features of tumors 39 d after inoculation. a, empty vector – transfected cells (510 m) were injected s.c. into the left flank and the mice were treated with PBS (i.p., thrice weekly). b, empty vector – transfected cells (510 m) were injected s.c. into the left flank and the mice were treated with docetaxel (7.0 mg/kg body weight; i.p., thrice weekly). c, Aurora-A shRNA-transfected cells (510 A) were injected s.c. into the left flank and the mice were treated with PBS (i.p., thrice weekly). d, Aurora-A shRNA-transfected cells (510 A) were injected s.c. into the left flank and the mice were treated with docetaxel (7.0 mg/kg body weight; i.p., thrice weekly). *B*, growth of tumors in the mice injected with Aurora-A shRNA-transfected cells (510 A) or empty vector – transfected cells (510 m) with or without docetaxel. The inoculation was done in five mice. *C*, tumor volume at day 39 after the inoculation. Left, black column, average tumor volume at day 39 after the inoculation of empty vector – transfected cells (510 m) in mice treated with docetaxel ($n = 5$). Right, black column, average tumor volume at day 39 after the inoculation of Aurora-A shRNA-transfected cells (510 A) in mice treated with PBS; white column, average tumor volume at day 39 after the inoculation of Aurora-A shRNA-transfected cells (510 A) in mice treated with docetaxel ($n = 5$). *, $P = 0.03$.

effects (32, 33). Therefore, our results in combination with these findings suggest that taxane-mediated chemotherapy could be more effective in combination with these anti-Aurora agents in ESCC.

In summary, the suppression of Aurora-A expression is shown to inhibit tumor growth of ESCC and enhanced chemosensitivity to docetaxel both *in vitro* and *in vivo*. Consequently, the therapeutic regimen to suppress the Aurora-A expression is a feasible candidate to become a novel therapeutic strategy for the treatment of ESCC.

Acknowledgments

We thank Sakiko Shimada for assistance with the cell cultures; Takako Murai, Kumi Kodama, Akane Iwase, and Fumie Uemura for technical assistance; and Drs. Atsushi Itami, Go Watanabe, Junichiro Kawamura, and Toshiya Soma for technical advice.

References

1. Ghadimi BM, Sackett DL, Difilippantonio MJ, et al. Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations. Genes Chromosomes Cancer 2000;27:183–90.
2. Kuo KK, Sato N, Mizumoto K, et al. Centrosome abnormalities in human carcinomas of the gallbladder and intrahepatic and extrahepatic bile ducts. Hepatology 2000;31:59–64.
3. Lingle WL, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL. Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. Proc Natl Acad Sci U S A 1998;95:2950–5.
4. Pihan GA, Purohit A, Wallace J, Malhotra R, Liotta L, Doxsey SJ. Centrosome defects can account for cellular and genetic changes that characterize prostate cancer progression. Cancer Res 2001;61:2212–9.
5. Pihan GA, Purohit A, Wallace J, et al. Centrosome defects and genetic instability in malignant tumors. Cancer Res 1998;58:3974–85.
6. Sato N, Mizumoto K, Nakamura M, et al. Correlation between centrosome abnormalities and chromosomal instability in human pancreatic cancer cells. Cancer Genet Cytogenet 2001;126:13–9.
7. Saunders WS, Shuster M, Huang X, et al. Chromosomal instability and cytoskeletal defects in oral cancer cells. Proc Natl Acad Sci U S A 2000;97:303–8.
8. Zhou H, Kuang J, Zhong L, et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy, and transformation. Nat Genet 1998;20:189–93.
9. Bischoff JR, Anderson L, Zhu Y, et al. A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J 1998;17:3052–65.
10. Li D, Zhu J, Firozi PF, et al. Overexpression of oncogenic STK15/BTAK/Aurora A kinase in human pancreatic cancer. Clin Cancer Res 2003;9:991–7.
11. Sen S, Zhou H, Zhang RD, et al. Amplification/overexpression of a mitotic kinase gene in human bladder cancer. J Natl Cancer Inst 2002;94:1320–9.
12. Royce ME, Xia W, Sahin AA, et al. STK15/Aurora-A expression in primary breast tumors is correlated with nuclear grade but not with prognosis. Cancer 2004;100:12–9.
13. Hamada M, Yakushijin Y, Ohtsuka M, Kakimoto M, Yasukawa M, Fujita S. Aurora2/BTAK/STK15 is involved in cell cycle checkpoint and cell survival of aggressive non-Hodgkin's lymphoma. Br J Haematol 2003;121:439–47.
14. Tanaka E, Hashimoto Y, Ito T, et al. The clinical significance of Aurora-A/STK15/BTAK expression in human esophageal squamous cell carcinoma. Clin Cancer Res 2005;11:1827–34.
15. Anand S, Penrhyn-Lowe S, Venkitaraman AR. AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. Cancer Cell 2003;3:51–62.
16. Dutertre S, Prigent C. Aurora-A overexpression leads to override of the microtubule-kinetochore attachment checkpoint. Mol Interv 2003;3:127–30.
17. Hata T, Furukawa T, Sunamura M, et al. RNA interference targeting aurora kinase a suppresses tumor growth and enhances the taxane chemosensitivity in human pancreatic cancer cells. Cancer Res 2005;65:2899–905.
18. Isono K, Sato H, Nakayama K. Results of a nationwide study on the three field lymph node dissection of esophageal cancer. Oncology 1991;48:411–20.
19. 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.
20. Muro K, Hamaguchi T, Ohtsu A, et al. A phase II study of single-agent docetaxel in patients with metastatic esophageal cancer. Ann Oncol 2004;15:955–9.
21. Lordick F, von Schilling C, Bernhard H, Hennig M, Bredenkamp R, Peschel C. Phase II trial of irinotecan plus docetaxel in cisplatin-pretreated relapsed or refractory oesophageal cancer. Br J Cancer 2003;89:630–3.
22. Pimkhaokham A, Shimada Y, Fukuda Y, et al. Nonrandom chromosomal imbalances in esophageal squamous cell carcinoma cell lines: possible involvement of the ATF3 and CENPF genes in the 1q32 amplicon. Jpn J Cancer Res 2000;91:1126–33.
23. Shimada Y, Imamura M, Wagata T, Yamaguchi N, Tobe T. Characterization of 21 newly established esophageal cancer cell lines. Cancer 1992;69:277–84.
24. Cremet JY, Descamps S, Verite F, Martin A, Prigent C. Preparation and characterization of a human aurora-A kinase monoclonal antibody. Mol Cell Biochem 2003;243:123–31.
25. Marumoto T, Honda S, Hara T, et al. Aurora-A kinase maintains the fidelity of early and late mitotic events in HeLa cells. J Biol Chem 2003;278:51786–95.
26. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell 2003;115:209–16.
27. Dandekar DS, Lokeshwar BL. Inhibition of cyclooxygenase (COX)-2 expression by Tet-inducible COX-2 antisense cDNA in hormone-refractory prostate cancer significantly slows tumor growth and improves efficacy of chemotherapeutic drugs. Clin Cancer Res 2004;10:8037–47.
28. Dutertre S, Cazales M, Quaranta M, et al. Phosphorylation of CDC25B by Aurora-A at the centrosome contributes to the G_2-M transition. J Cell Sci 2004;117:2523–31.
29. Brummelkamp TR, Bernards R, Agami R. Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell 2002;2:243–7.
30. Hingorani SR, Jacobetz MA, Robertson GP, Herlyn M, Tuveson DA. Suppression of BRAF(V599E) in human melanoma abrogates transformation. Cancer Res 2003;63:5198–202.
31. Takei Y, Kadomatsu K, Yuzawa Y, Matsuo S, Muramatsu T. A small interfering RNA targeting vascular endothelial growth factor as cancer therapeutics. Cancer Res 2004;64:3365–70.
32. Mortlock AA, Keen NJ, Jung FH, et al. Progress in the development of selective inhibitors of aurora kinases. Curr Top Med Chem 2005;5:807–21.
33. Keen N, Taylor S. Aurora-kinase inhibitors as anticancer agents. Nat Rev Cancer 2004;4:927–36.

BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 6 2007, pages 732–738
doi:10.1093/bioinformatics/btl663

Gene expression

Genomic characterization of multiple clinical phenotypes of cancer using multivariate linear regression models

Shigeyuki Matsui[1,2,*], Masaaki Ito[3], Hiroyuki Nishiyama[3], Hajime Uno[4], Hirokazu Kotani[5], Jun Watanabe[3], Parry Guilford[6], Anthony Reeve[7], Masanori Fukushima[8,2] and Osamu Ogawa[3]

[1]Department of Pharmacoepidemiology, Graduate School of Public Health, Kyoto University, Kyoto, [2]Translational Research Informatics Center, Foundation for Biomedical Research and Innovation, Kobe, [3]Department of Urology, Kyoto University Graduate School of Medicine, Kyoto, [4]Division of Biostatistics, School of Pharmaceutical Sciences, Kitasato University, Tokyo, [5]Department of Pathology, Kyoto University Graduate School of Medicine, Kyoto, Japan, [6]Pacific Edge Biotechnology, University of Otago, Dunedin, [7]Cancer Genetics Laboratory, Department of Biochemistry, University of Otago, Dunedin, New Zealand and [8]Dvision of Clinical Trial Design and Management, Translational Research Center, Kyoto University Hospital, Kyoto, Japan

Received on October 2, 2006; revised on November 27, 2006; accepted on December 25, 2006

Advance Access publication January 18, 2007

Associate Editor: Joaquin Dopazo

ABSTRACT

Motivation: The development of gene expression microarray technology has allowed the identification of differentially expressed genes between different clinical phenotypic classes of cancer from a large pool of candidate genes. Although many class comparisons concerned only a single phenotype, simultaneous assessment of the relationship between gene expression and multiple phenotypes would be warranted to better understand the underlying biological structure.

Results: We develop a method to select genes related to multiple clinical phenotypes based on a set of multivariate linear regression models. For each gene, we perform model selection based on the doubly-adjusted R-square statistic and use the maximum of this statistic for gene selection. The method can substantially improve the power in gene selection, compared with a conventional method that uses a single model exclusively for gene selection. Application to a bladder cancer study to correlate pre-treatment gene expressions with pathological stage and grade is given. The methods would be useful for screening for genes related to multiple clinical phenotypes.

Availability: SAS and MATLAB codes are available from author upon request.

Contact: matsui@pbh.med.kyoto-u.ac.jp

1 INTRODUCTION

The development of comprehensive, gene expression microarray technology has allowed the identification of differentially expressed genes between different clinical phenotypic classes of cancer from a large pool of candidate genes. Although many of such class comparison studies assessed relation of gene expression with a single clinical phenotype, such as stage of cancer, assessment of relationship with multiple clinical phenotypes, e.g. stage and grade of cancer, would be more relevant to better understand the underlying biological structure.

In incorporating multiple phenotypic characteristics of cancer simultaneously, multivariate linear regression models or ANOVA-type models for each gene are a useful analytical tool by introducing phenotypes as covariates for the response variable of gene expression level, although many applications concern modelling of channel-specific intensities to reflect the experimental design using multichannel arrays (e.g., Kerr *et al.*, 2000, 2001; Wolfinger *et al.*, 2001; Dobbin and Simon, 2002) or oligonucleotide arrays (Chu *et al.*, 2002). In correlating gene expression data with multiple phenotypes, several groups of genes with distinct differential patterns may exist. A group of genes may relate to only a single phenotype, e.g. stage or grade, while another group of genes may relate to multiple phenotypes, e.g. both stage and grade, possibly with their interaction.

A commonly used approach for selecting genes related to multiple phenotypes is to assume a full model that incorporates all phenotypes and their possible interactions, e.g. a model with stage, grade and their interaction as covariates, to cover all possible differential patterns for multiple phenotypes. Then, an overall test for the null hypothesis of no effect of all the covariates on gene expression is performed with some control for false positives (Chen *et al.*, 2005; Matsui, 2006). A drawback to this approach is that, although the overall test could detect genes with any differential patterns, the power of this test may be smaller for genes for which reduced models are correct. One may want to detect genes with various differential models equally. Only a single model should not be fitted for all genes exclusively, unless a particular differential pattern is of interest.

*To whom correspondence should be addressed.

One idea to improve the power for various differential models would be to invoke some model selection. Model selection may also be warranted to gain additional insights on selected genes. For example, a group of genes for which the same model is selected have the potential to relate to the same aspect of disease biology. One approach is to perform a structural hypothesis testing. For example, we first assume the full model with stage, grade and their interaction and test if there is an interaction effect between stage and grade for all genes. For those genes where an interaction exists, the full model can be selected. For those where there is not, the reduced model with just stage and grade can be fit and the main effects of stage and grade are tested separately. A critical issue of this approach, however, is how to determine the significance levels for these tests. One possible approach is to determine the significance levels based on a criterion to control an error rate on false positives. However, there could be multiple combinations of significant levels for main and interaction effects for a given level of an overall error rate. Apart from technical issues, one may also argue in the first place that model selection is a different problem of gene selection and hence another criterion should be used for model selection.

In this article, we develop a method to select genes related to multiple clinical phenotypes via model selection based on a set of candidate, multivariate linear regression models. For each gene, all candidate models are fit and a model is selected based on a criterion of prediction error. Specifically we select the model for which the doubly-adjusted R-square proposed by Okuno *et al.* (1977) is maximized. This model selection procedure can be expressed as model selection based on hypothesis testing described earlier, but the significance levels are determined from the perspective of prediction error (see Section 2). Also, this model selection criterion is approximately equal to popular criteria based on Mallow's Cp (Mallows, 1973) and AIC (Akaike, 1973). Meanwhile, the doubly-adjusted R-square is a reasonable statistic for gene selection because of its link to an overall test for the null hypothesis of no effect of all phenotypic variables on gene expression. Procedures to control the multiplicity from examining the large number of genes can be developed. A simulation study is conducted to evaluate the performance of the proposed method. Lastly, we illustrate the methods using a bladder cancer data set to correlate pre-treatment gene expression levels with pathological stage and grade.

2 METHODS

2.1 Multivariate linear regression models

Assume values Y_{ji} for $j=1,\ldots,J$ genes from sample $i=1,\ldots,n$. For cDNA arrays these are typically normalized log-ratios and for oligonucleotide arrays they are normalized log signals. For each gene, the expression data is linked to multiple phenotypes via multivariate linear regression models. We suppose that there are, in total, p variables, $x_1,\ldots,x_p$, for which their association with gene expression is of interest. Note that the set of variables may include those representing interaction effects of different phenotypes, for example, an interaction of two different phenotypes, $x_{k1}x_{k2}$, where x_{k1} and x_{k2} represent the respective phenotypes ($k_1 \neq k_2$). We assume various candidate models using the set of all the p variables and subsets of it as covariates. When k covariates ($1 \leq k \leq p$) are linked to the expression level for gene j, we assume the (candidate) linear model,

$$Y_{ji} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_{ji} \quad (1)$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are regression coefficients. The error term, ε_{ji}, is assumed independently and identically distributed the normal distribution $N(0, \sigma_j^2)$.

2.2 The Doubly-adjusted R-square

In this section, we restrict our attention to analysis of a particular gene and subscript j is dropped for brevity. The doubly-adjusted R-square is related to the prediction sum of squares (PRESS) (Hocking, 1976), defined by $\text{PRESS} = \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2$, where $\hat{y}_{(i)}$ is the predicted value of y_i obtained after omitting sample i from the fitting process ($i=1,\ldots,n$). The PRESS is a statistic related to prediction error. If the model is correctly specified, the expected value of PRESS reduces to $(n+k+1)\sigma^2$. Using the commonly used, unbiased estimator for the parameter σ^2, $RSS/(n-k-1)$, the expected PRESS can be estimated by $(n+k+1)RSS/(n-k-1)$.

The doubly-adjusted R-square is defined by one minus the ratio of estimated expected prediction errors, the expected prediction error under the linear model with k covariates and that under the linear model with no covariate ($k=0$) (Okuno *et al.*, 1977),

$$R^2_{DA} = 1 - \frac{[(n+k+1)/(n-k-1)]RSS}{[(n+1)/(n-1)]TSS},$$

or

$$R^2_{DA} = 1 - \frac{n+k+1}{n+1}\frac{n-1}{n-k-1}(1-R^2) \quad (2)$$

where

$$R^2 = 1 - \frac{RSS}{TSS} \quad (3)$$

denotes the coefficient of multiple determination or the R-square statistic. Here $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$ is the total sum of squares and $RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ is the residual sum of squares, where $\hat{y}_i$ is the fitted value for y_i and $\bar{y} = \sum_{i=1}^{n} y_i/n$ is the average of gene expression across all samples. For R^2_{DA}, the expected PRESS is used, instead of the usual sum of squares such as RSS and TSS, which are used for R^2 in (3).

The statistic R^2_{DA} has a one-to-one correspondence with an F-statistic for the overall test of the null assumption that all of the k regression coefficients are zero. Note that this null assumption represents no effects of all the p variables because no effects of the other $p-k$ variables are assumed in model (1). The F-statistic is given by

$$F = \frac{(TSS - RSS)/k}{RSS/(n-k-1)} \quad (4)$$

which has the relation with R^2,

$$R^2 = \frac{kF}{kF + n - k - 1} \quad (5)$$

As such, R^2_{DA} has correspondence with the F-statistic through (2).

For each gene, we select the model for which R^2_{DA} is maximized over candidate models. Consider comparison of the two models, one with k covariates and the other with $(k+1)$ covariates after the entry of one additional covariate into the model with k covariates. The model with $(k+1)$ covariates will be selected if $R^2_{DA,(k+1)} \geq R^2_{DA,(k)}$.

Another popular criterion for model selection is based on the partial F-statistic, which can be expressed as

$$F_{(k)} = \frac{R^2_{(k+1)} - R^2_{(k)}}{(1 - R^2_{(k+1)})/(n-k-2)}$$

The criterion $R^2_{DA,(k+1)} \geq R^2_{DA,(k)}$ corresponds to

$$F_{(k)} \geq \frac{2n}{n+k+1} \approx 2$$

when $n \gg k$ (Okuno *et al.*, 1977). Another popular criterion is based on Mallow's Cp (Mallows, 1973),

$$Cp = \frac{RSS}{\sigma^2} + 2(k+1) - n$$

for the model with k covariates. The AIC (Akaike, 1973) is also a popular statistic for model selection, although assumption for the error term distribution is needed. For normally distributed error terms with known variance, AIC can be expressed as

$$AIC = \frac{RSS}{\sigma^2} + 2(k+1) + \text{const.}$$

for the model with k covariates, which is identical with Cp. The model with $(k+1)$ covariates may be selected if $Cp_{(k+1)} \leq Cp_{(k)}$ or $AIC_{(k+1)} \leq AIC_{(k)}$. When the unbiased estimate for σ^2 under the model with $(k+1)$ covariates, $RSS_{(k+1)}/(n-k-2)$, is replaced with σ^2, these criteria may correspond to

$$F_{(k)} \geq 2.$$

As such, the criterion based on R^2_{DA} is approximately equal to the criteria based on Cp or AIC.

2.3 Selection of model and gene

We invoke model selection for each gene before gene selection. For each gene, we perform model selection based on R^2_{DA}, i.e. selecting the model for which is maximized. Then, the maximum R^2_{DA} is used as the statistic for gene selection, and genes with the greatest values of the maximum R^2_{DA} are selected. A rationale for using R^2_{DA} for gene selection is related to its correspondence with the overall F-test of no relations between all the p variables and gene expression as noted in the previous section. For genes for which this null assumption is incorrect, R^2_{DA} for some candidate models or the maximum of R^2_{DA} across candidate models may tend to have larger values, compared with genes for which the null assumption is correct. This tendency may be true for genes with any differential pattern for gene expression as long as this is captured by one of candidate models. Thus, our procedure is to select genes related to multiple phenotypes irrespective of their differential patterns via the optimization in gene selection. Note that other model selection criteria such as Cp and AIC may not have the correspondence with the overall test because of standardization across candidate models for each gene, which would be a drawback for using these statistics for gene selection. For example, if all p variables are used as covariates in estimating σ^2 for each gene, Cp for the model with k covariates may reduce to $2k-p+1$ for all genes.

The multiplicity issue by repeated examinations across genes can be adjusted by controlling the family-wise error rate (FWER) or the false discovery rate (FDR), where the FWER is the probability of making one or more false positives and the FDR is the expected proportion of false positives among the positives declared (Dudoit *et al.*, 2003). Note that the control of these error rates does not relate to individual effects but to the overall test of the null assumption of no effects for all the p variables. Many authors now favour the FDR over the FWER as the appropriate error measure for exploratory microarray studies because the FWER approach can be very conservative. A practical and conceptually simple approach to providing procedures to control the FDR conservatively is to fix the rejection region beforehand and to estimate the FDR (Storey, 2002). A gene is declared to be significant if the maximum R^2_{DA} for the gene is equal to or greater than a cut-off point on this statistic. For a given cut-off point, we can estimate the FDR conservatively by a multivariate permutation procedure with random permutations of the assignment of the set of all p variables to derive the null distribution of the maximum R^2_{DA}, from which the average of (false) positives can be calculated as an estimate of the expected false positives. For controlling FWER, the null distribution of the maximum of the maximum R^2_{DA} across genes is derived from the random permutations, which is referred to the observed value of the maximum R^2_{DA} to obtain an adjusted P-value.

3 RESULTS

3.1 Simulation

We assessed adequacy of the proposed method through a small simulation study. Particularly we compared the power of our method with that of the conventional method that uses a single model with all p variables as covariates exclusively and select genes based on the overall F-test for this model. We considered three distinct phenotypes. The state of each phenotype was dichotomized, and let x_1, x_2 and x_3 be binary variables representing respective phenotypes. Let $x_h = 1$ for a particular state of phenotype and $x_h = 0$ for the other states ($h = 1, 2, 3$). Thus, there are eight combinations in total for the level of the three phenotypes, $(x_1, x_2, x_3) = (0,0,0)$, $(0,0,1)$, $(0,1,0)$, $(0,1,1)$, $(1,0,0)$, $(1,0,1)$, $(1,1,0)$ and $(1,1,1)$. We considered six samples for each combination (the total number of samples was 48).

We assessed the power for selecting a single informative gene which is related to phenotypes. We considered 1000 genes and, of the 1000 genes, 1 gene was informative and the other 999 were non-informative (0.1% of the whole gene is informative). For the informative gene, we assumed the model $\{1\}$, $\{1,2\}$, $\{1,2,1\times2\}$, $\{1,2,3\}$ or $\{1,2,3,1\times2,1\times3,2\times3,1\times2\times3\}$. For example, the model $\{1,2,1\times2\}$ is to a linear model with the three covariates, x_1, x_2 and x_1x_2. The error term ($\varepsilon_{j,i}$ in(1)) is assumed independently and identically distributed the standard normal distribution for all genes. The size of regression coefficients ranged from 0.0 to 2.0. For the informative gene with the model, $\{1,2,1\times2\}$, we assumed the differential patterns where the samples with one of the four combinations for the level of the two binary phenotypic variables, x_1 and x_2, e.g. samples with $(x_1, x_2) = (0,0)$, irrespective of the level of x_3, had higher expressions compared with the other samples. Similarly, for the informative gene with the model, $\{1,2,3,1\times2,1\times3,2\times3,1\times2\times3\}$, we assumed the differential patterns where the samples with one of the eight combinations for the level of the three binary phenotypic variables, x_1, x_2 and x_3, e.g. samples with $(x_1, x_2, x_3) = (0,0,0)$, had higher expressions compared with the other samples.

For each simulation configuration, we performed the proposed method and the conventional method. For the proposed method, we introduced the four variables representing interaction effects, x_1x_2, x_1x_3, x_2x_3 and $x_1x_2x_3$, in addition to the three variables, x_1, x_2 and x_3, as covariates. For the seven covariates, we assumed the following 12 candidate models, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,1\times2\}$,

$\{1, 3, 1\times3\}$, $\{2, 3, 2\times3\}$, $\{1, 2, 3\}$, $\{1, 2, 3, 1\times2, 1\times3, 2\times3\}$ and $\{1, 2, 3, 1\times2, 1\times3, 2\times3, 1\times2\times3\}$. For the conventional method, we employed only the overall F-test for the full model, $\{1, 2, 3, 1\times2, 1\times3, 2\times3, 1\times2\times3\}$, for gene selection. For each procedure, we control the FWER. A null distribution of the maximum of statistic (the maximum R^2_{DA} for the proposed method and the F-statistic for the conventional method) was obtained from 500 permutations, which was referred to the observed value of statistic to obtain an adjusted P-value. For each of a selected set of regression coefficients, Table 1 summarizes empirical power for the informative gene obtained from 500 simulations at the significance level of 10% for the adjusted P-value. We first confirmed that the empirical FWER was almost equal to the nominal level (data not shown). The empirical power of the proposed method can be substantially greater than that of the conventional method when models other than the full model are correct for the informative gene. The conventional method outperformed the proposed method when the full model is correct for the informative gene. Although this could be explained by error in model selection, the power loss was relatively small in our simulation setting.

3.2 The bladder cancer example

Tumour-biopsy tissues were collected at diagnosis from 48 patients with bladder transitional cell carcinomas at the Department of Urology, Kyoto University Graduate School of Medicine between 1990 and 2003. Patients were selected on the basis of availability of tumour-biopsy samples, without regard to clinical information. Microarray experiments were performed for frozen tissues using a cDNA array that contains printings of approximately 30 000 oligonucleotides fabricated by Pacific Edge Biotechnology Limited in New Zealand. In each hybridization, fluorescent cDNA targets were prepared from a tumour mRNA sample and a reference mRNA sample contracted from a pool of cell lines of different cancers. Two replicated arrays made from the same sample of RNA (technical replicates) without reverse labelling or dye swap were averaged to improve precision of the estimate of the expression profile for a given RNA sample. After image analysis and spot filtering, the data were normalized by a locally weighted linear regression method. We excluded genes that had low variance of log ratio expressions across samples, because the observed variability for genes with low variance is more likely to be due to measurement noise than actual biological variability (e.g. Simon *et al.*, 2004). We selected six thousands genes with variance in the top 20th percentile.

As clinical phenotypes, cancer stage and grade from pathological reports were confirmed by a pathologist. We classified stage into two categories with distinct treatment modalities, superficial (≦ pT1) and invasive tumours (>pT1), to understand the underlying biological difference related to therapeutics. Generally, superficial tumours are treated with transurethral surgery or intravesical instillation therapy, while invasive tumours are treated with radical cystectomy or systemic chemotherapy. Of 48 patients, we had 31 superficial and 17 invasive patients. We also classified grade into two categories, low grade (grade 1 or 2) and more malignant, high grade (grade 3). Of the 31 superficial tumours, 25 and 6 were low and high grade, respectively. Of the 17 invasive tumours, 5 and 12 were low and high grade, respectively.

Although several authors have been interested in genes differentially expressed between stage classes (e.g. Sanchez-Carbayo *et al.*, 2003; Modlich *et al.*, 2004), it would seem more likely that multiple factors are related to gene expression simultaneously. We assumed a multivariate linear regression model for each gene. For a single gene, let Y_i be the log-ratio of gene expression measurement for patient i ($i=1,\ldots,48$). Let stage_i be a binary variable that takes 1 if the stage of patient i is invasive and zero otherwise. Let grade_i be a binary variable that takes 1 if the grade of patient i is high and zero otherwise. We also consider an interaction term of stage and grade, $\text{stage}_i\times\text{grade}_i$. For log-transformed intensity ratios, we assume the following four candidate linear models for differential expression,

$$\text{M1: } \mu_i = \beta_0 + \beta_1(\text{stage}_i)$$
$$\text{M2: } \mu_i = \beta_0 + \beta_1(\text{grade}_i)$$
$$\text{M3: } \mu_i = \beta_0 + \beta_1(\text{stage}_i) + \beta_2(\text{grade}_i)$$
$$\text{M4: } \mu_i = \beta_0 + \beta_1(\text{stage}_i) + \beta_2(\text{grade}_i) + \beta_3(\text{stage}_i \times \text{grade}_i)$$

where $\mu_i = \text{E}(Y_i)$ is the expected value or the mean value of the log-ratio for patient i and β_0, β_1, β_2 and β_3 are regression coefficients. The model M1 was to detect genes just related to stage like in previous studies. For simultaneous relation with stage and grade, M3 and M4 were assumed. Note that M4 is a

Table 1. Empirical power (%) and its standard error in parenthesis of the proposed method and the conventional method for each of a selected set of regression coefficients in differential model

Model for informative gene	Regression coefficients (β_1, β_2, β_3, β_{12}, β_{13}, β_{23}, β_{123})	Proposed method	Conventional method
{1}	(2, 0, 0, 0, 0, 0, 0)	95 (1.0)	83 (1.7)
{1, 2}	(1, 1, 0, 0, 0, 0, 0)	35 (2.1)	20 (1.8)
{1, 2, 1×2}	(−2, −2, 0, 2, 0, 0, 0)	79 (1.8)	61 (2.2)
{1, 2, 3}	(1, 1, 1, 0, 0, 0, 0)	69 (2.1)	56 (2.2)
{1, 2, 3, 1×2, 1×3, 2×3, 1×2×3}	(−2, −2, −2, 2, 2, 2, −2)	22 (1.9)	25 (1.9)

The proportion of correct model selection using the criterion based on R^2_{DA} in the proposed method ranged from 60% to 80% for the first four configurations and 46% for the fifth configuration for model {1, 2, 3, 1×2, 1×3, 2×3, 1×2×3}.

saturated model (four regression parameters for four combinations of binary classes, stage and grade). It covers any differential patterns other than additive effects of stage and grade. They include differential expressions between two classes based on the combination of stage and grade. For example, a differential expression between superficial cancer with low grade (the best type of cancer) and the others is obtained by introducing the constraint $\beta_1=\beta_2=-\beta_3$ ($\neq 0$) in M4, which may yield a linear model with a single interaction effect and no main effects. These differential patterns could be identified based on the estimates of parameter coefficients for genes for which M4 is selected as illustrated below.

Table 2 shows an estimate of FDR for each of several cut-off points for the maximum R^2_{DA} obtained by the multivariate permutation procedure with 2000 permutations. We selected the top 100 genes, so that the FDR was around 15%. Table 3 shows classification of 100 selected genes based on selected models and the sign of estimated regression coefficients in selected model. Note that, out of the top 100 genes, the same model was selected for 98 genes based on AIC, as expected by the equality of the criterion on R^2_{DA} with that on AIC as noted in Section 2. Seventeen genes were related to only stage via the model M1; 9 (8) had negative (positive) estimates of the regression coefficient β_1 for stage in M1. In other words, the 9 (8) genes were under-(over-) expressed for invasive cancer. Thirteen genes were related to only grade via the model M2; 10 (3) were under-(over-) expressed for high grade cancer. Interestingly, the rest 70 genes were related to both stage and grade; 39 genes via the model M3 without the interaction term of stage and grade and 31 genes via the model M4 with the interaction term. Of the 39 genes for which the model M3 was selected, negative values for both β_1 and β_2 (under-expression for invasive or high grade cancer) were suggested for 37 genes, while positive values for both β_1 and β_2 (over-expression for invasive or high grade tumours) for two genes. Of the 31 genes for which the model M4 was selected, 15 (16) genes had negative (positive) estimates for β_3, i.e. the interaction of stage and grade.

Figure 1 shows typical mean expression profiles for the selected genes with negative (Fig. 1A and 1C) or positive interaction (Fig. 1B and 1D). Figures 1A and 1B indicate differential expression between invasive cancer with high grade (the worst type) and the others, while Figures 1C and D indicate differential expression between superficial cancer with low grade (the best type) and the others. In general, selected genes for which under-expression was linked to more progressed cancer, i.e. invasive and/or high grade, involved putative tumour suppressor genes. A gene related to tumour protein *p53*, which is a well-known tumour suppressor, was related to grade via the model M2 and the estimate for β_1 (the regression coefficient for grade) was negative. *Glutathione peroxidase* (*GPX*) *2*, which is a member of oxidoreductase and the destruction of it increased the genesis of bacteria-induced intestinal cancer (Chu *et al.*, 2004), was related to both stage and grade via the model M3 and the estimates for β_1 and β_2 (the regression coefficient for stage and grade, respectively) were negative. With respect to genes for which the model M4 was selected, putative tumour suppressor genes, *RNA binding motif protein 3* (*RBM3*) (Sutherland *et al.*, 2005) in Figure 1A and *CD-81* (Koenig-Hoffmann *et al.*, 2005) in Figure 1D were under-expressed for the worst type of cancer (invasive and high grade) and for cancers other than the best type (superficial and low grade), respectively.

On the other hand, selected genes for which over-expression was linked to more progressed cancer are suggested to act like as oncogenes. *Amphiregulin* in Figure 1B acts as a protease of the extracellular matrix (ECM) with induction of extracellular matrix metalloproteinase (MMP) –2 and –9, and the expression level was associated with survival of bladder cancer patients (Menashi *et al.*, 2003; Thogersen *et al.*, 2001). Highly aggressive invasive cancers are characterized by their activity of inducing break down of cell– cell or cell– matrix adhesion, modulation of ECM proteolysis and induction of angiogenesis. A family of human endogenous retroviruses (HERVs), *HERV-H*, in Figure 1C was highly expressed in bladder cancer (Stauffer *et al.*, 2004). *Amphiregulin* in Figure 1B and *HERV-H* in Figure 1C were over-expressed for the worst type of cancer (invasive and high grade) and for cancers other than the best type (superficial and low grade), respectively.

Table 2. Estimates of FDR for various cut-offs for the maximum R^2_{DA}

Cut-off	Number of significant genes	Estimate of expected number of false positives	Estimate of FDR (%)
0.10	1351	470.8	34.8
0.20	288	47.3	16.4
0.25	99	14.7	14.8
0.30	37	4.5	12.2

Table 3. Sign of estimated regression coefficients in the selected models for the 100 significant genes

Model	Stage	Grade	Interaction	Number of genes	Total
M1	Negative	–	–	9	17
	Positive	–	–	8	
M2	–	Negative	–	10	13
	–	Positive	–	3	
M3	Negative	Negative	–	37	39
		Positive	–	0	
	Positive	Negative	–	0	
		Positive	–	2	
M4	Negative	Negative	Negative	1	31
			Positive	10	
		Positive	Negative	0	
			Positive	2	
	Positive	Negative	Negative	7	
			Positive	3	
		Positive	Negative	7	
			Positive	1	
Total					100

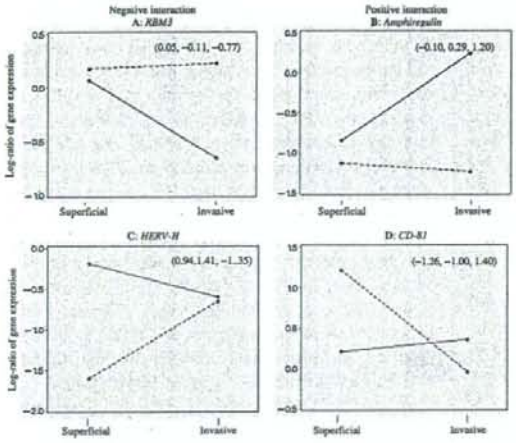

Fig. 1. Typical expression profiles for genes with negative or positive interaction. Each of four plots A–D shows the mean expressions of a gene of the 100 significant genes for the four types of cancer combined by stage, superficial and invasive, and grade, low (dashed line) and high (solid line). The estimates of parameters (β_1, β_2, β_3) for the model M4 are also given.

DISCUSSION

We have developed a method to select genes related to multiple clinical phenotypes via model selection based on a set of multivariate linear regression models. As indicated by the simulation study, our methods with model selection based on the doubly-adjusted R-square, R^2_{DA}, can substantially improve the power for various differential models for multiple phenotypes, compared with the conventional method that uses only a single model for gene selection. Also, the model selection indicates appropriate model for each selected gene, which may be helpful to gain insights on the relationship between gene expression and multiple phenotypes. Hence, our methods would be a useful tool for screening for selecting genes with various differential models for multiple phenotypes. In linear regression models, various types of phenotypic variables including nominal, ordinal and continuous variables with three or more levels can be accommodated. The R^2_{DA} can be generally adopted for both nested and non-nested models. Our methods are hence applicable to a wide variety of correlative analyses of global gene expression profiles from microarrays with multiple phenotypes. Although we have focused on gene selection, the model building process for prediction, e.g. for diagnostic prediction, preferably by generating hybrid models consisting of different kind of models based on a set of genes, is interesting and thus subject to future research.

Because of the large number of genes in microarray experiments, there will always be some genes with a very small sum of squares across samples, so that the value of an overall test statistic will be very large whether or not their averages are large. Tusher *et al.* (2001) have proposed a regularization that avoids this difficulty. For the overall F-test in (4), this regularization may be expressed as

$$F(\phi) = \frac{(TSS - RSS)/k}{RSS/(n-k-1)+\phi}$$

where ϕ is a fudge factor whose value is determined to reduce the dependence of $F(\phi)$ versus ϕ (Tusher *et al.*, 2001). As a simple modification of our methods, we obtain a regularized version of the doubly-adjusted R-square by replacing the usual F-statistic with this regularized F-statistic in (2), (3) and (5). This regularized statistic would be reasonable for gene selection because of its one-to-one correspondence with $F(\phi)$. However, for model selection, one may prefer the criterion of maximizing the original doubly-adjusted R-square to that of maximizing the regularized one because of the equality with popular criteria based on Cp or AIC. Hence, we first select the model for which the original doubly-adjusted R-square is maximized for each gene, followed by gene selection based on the regularized doubly-adjusted R-square for the selected model. We derived this regularization in a heuristic way and its theoretical justification is subject to future research. The approaches to borrow strength across genes via hierarchical modelling (e.g. Baldi and Long, 2001; Efron *et al.*, 2001; Wright and Simon, 2003) may provide a framework for this.

If clinical variables are highly correlated in a set of samples, then the models chosen for individual genes are likely to be highly unstable because of lack of power to discriminate between the different effects. One should select a set of samples to be well balanced especially when specimen banks from large cohort studies or clinical trials are available.

Determination of the set of candidate models is an important issue. One can assume, in principle, many candidate models, for example, by using different categorizing or higher order effects for phenotypic variables or by imposing constraints for regression coefficients to reflect a particular differential pattern such as the constraint in the model M4 described in the bladder cancer example (see Section 3.2). However, a large set of candidate models may suffer from error in model selection seriously. A good strategy would be to prepare a small number of plausible models with biologically different interpretations and one very flexible model to capture other differential patterns (such as the model M4 in the bladder cancer example).

We illustrated the methods using the bladder cancer data. We selected 100 genes with various patterns of differential expression for stage and/or grade (Table 3). Different models or differential expression patterns suggested from Table 3 have the potential to reflect different biologic aspects in stage and grade progressions of bladder cancer. Although some selected genes in our analyses have already been identified and studied for bladder cancer and other cancer, many genes were newly identified by our analyses. These genes have the potential to provide new insights on molecular alternation related to disease biology and progressiveness in bladder cancer. A fraction of these genes are now under investigation. As an initial step, we confirmed reproducibility of differential expression for six selected genes which were related to RNA processing and/or

translation (1, 3 and 2 genes for which the model M1, M3 and M4 was selected, respectively), but have not been studied for bladder cancer, for additional 74 samples by quantitative real time PCR assays (M. Ito, unpublished data).

Our methods could also indicate candidate molecular markers for more accurate classification of cancer. For example, genes for which the linear model, M1 or M2, were selected could supplement traditional pathological assays in determining stage or grade. Besides, our methods could help in developing new approaches to cancer diagnosis incorporating multiple diagnostic factors. For example, genes for which the model M4 with the interaction effects of stage and grade was selected (e.g. the four genes in Fig. 1) could be candidate genes useful for identifying the best (superficial and low grade) or the worst (invasive and high grade) type of bladder cancer. In such attempts, link with outcome or prognosis information should also be investigated.

Conflict of Interest: none declared.

REFERENCES

Akaike,H. (1973) Information theory and an extension of the maximum likelihood principle. In Petrov B.N. and Caski F (eds.) *Second International Symposium on Information Theory*. Akademia Kiado, Budapest, pp. 267–281.

Baldi,P. and Long,A.D. (2001) A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, **17**, 509–519.

Chen,D. *et al.* (2005) Selecting genes by test statistics. *J. Biomed. Biotechnol.*, **2005**, 132–138.

Chu,F.F. *et al.* (2004) Bacteria-induced intestinal cancer in mice with disrupted Gpx1 and Gpx2 genes. *Cancer Res.*, **64**, 962–968.

Chu,T.M. *et al.* (2002) A systematic statistical linear modeling approach to oligonucleotide array experiments. *Math. Biosci.*, **176**, 35–51.

Dobbin,K. and Simon,R. (2002) Comparison of microarray designs for class comparison and class discovery. *Bioinformatics*, **18**, 1438–1445.

Dudoit,S. *et al.* (2003) Multiple hypothesis testing in microarray experiments. *Stat. Sci.*, **18**, 71–103.

Efron,B. *et al.* (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.*, **96**, 1151–1160.

Hocking,R.R. (1976) The analysis and selection of variables in linear regression. *Biometrics*, **32**, 1–49.

Kerr,M.K. and Churchill,G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.*, **77**, 123–128.

Kerr,M.K. *et al.* (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.

Koenig-Hoffmann,K. *et al.* (2005) High throughput functional genomics: identification of novel genes with tumor suppressor phenotypes. *Int. J. Cancer*, **113**, 434–439.

Mallows,C.L. (1973) Some comments on Cp. *Technometrics*, **15**, 661–676.

Matsui,S. (2006) Statistical applications using DNA microarrays for cancer diagnosis and prognosis. In Crowley J.J. and Ankerst D.P. (eds.) *Handbook of Statistics in Clinical Oncology*. 2nd edn. CRC Press, Boca Raton, pp. 419–436.

Menashi,S. *et al.* (2003) Regulation of extracellular matrix metalloproteinase inducer and matrix metalloproteinase expression by Amphiregulin in transformed human breast epithelial cells. *Cancer Res.*, **63**, 7575–7580.

Modlich,O. *et al.* (2004) Identifying superficial, muscle-invasive, and metastasizing transitional cell carcinoma of the bladder: use of cDNA array analysis of gene expression profiles. *Clin. Cancer. Res.*, **10**, 3410–3421.

Okuno,T. *et al.* (1977) Prediction sum of squares, Akaike's information criteria and doubly-adjusted multiple correlation coefficient. *Bull. Int. Stat. Inst.*, **47**, 370–373.

Sanchez-Carbayo,M. *et al.* (2003) Gene discovery in bladder cancer progression using cDNA microarrays. *Am. J. Pathol.*, **163**, 505–516.

Simon,R. *et al.* (2004) *Design and Analysis of DNA Microarray Investigations*. Springer, New York.

Stauffer,Y. *et al.* (2004) Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun.*, **4**, 2.

Storey,J.D. (2002) A direct approach to false discovery rates. *J. Roy. Stati. Soc.*, **B64**, 479–498.

Sutherland,L.C. *et al.* (2005) RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? *J. Cell Biochem.*, **94**, 5–24.

Thogersen,V.B. *et al.* (2001) A subclass of HER1 ligands are prognostic markers for survival in bladder cancer patients. *Cancer Res.*, **61**, 6227–6233.

Tusher,V.G. *et al.* (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5121.

Wright,G.W. and Simon,R.M. (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, **19**, 2448–2455.

Wolfinger,R.D. *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.*, **8**, 625–637.