considered as positive, and the proportion was scored as 0 = negative, 1 = less than 50% positive tumor cells and 2 = equal to or more than 50% positive tumor cells. Averaged scores were calculated, and each case was categorized into 'mostly positive' when the averaged score was more than or equal to 1.75, 'mostly negative' when less than or equal to 0.25 and 'heterogeneous' when the score was greater than 0.25 and less than 1.75. The expression status in all cases in the heterogeneous category and 15 cases each of the mostly positive and mostly negative category was verified with staining of the corresponding whole sections.

Relative quantification by real-time RT–PCR

Total RNA was extracted from 13 cell lines, and first-strand cDNAs were synthesized using Superscript II (Invitrogen, Carlsbad, CA, USA) and random hexamer primers (Roche Applied Science, Alameda, CA, USA). Real-time quantitative PCR amplifications were performed with the Smart Cycler system (SC-100, Cepheid, Sunnyvale, CA, USA). The reactions were carried out using the QuantiTect SYBR Green PCR kit (Qiagen, Valencia, CA, USA). In each reaction, standard samples were diluted up to 1/1000 with cDNA from a lung cancer cell line, selected in preliminary experiments for each gene, and were run with unknown tumor samples. Finally, relative quantitative values from each sample were compared to their 18s rRNA values.

*Mutation status of p53 and K-*ras

Frozen tissue of the tumor specimens was grossly dissected to enrich for the tumor cells and to extract total RNA with the RNeasy kit (Qiagen). Using a standard RT–PCR procedure, exon 4 to exon 10 of the p53 gene was amplified, and the products directly sequenced with an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). When no mutation signals were obtained, the result was confirmed by a functional assay in yeast (Ishioka *et al.*, 1993; Waridel *et al.*, 1997). When more than 10% of red colonies or significant deviation of the split assay using pWF35 and pFW34 were observed with the functional assay, RNA was re-extracted from tumor cells removed with a laser-captured microdissection system (PixCell-II, Arcturus, CA, USA), and the products of RT–PCR were sequenced. Using the same RNA as the p53 mutational analysis, the K-*ras* mutation was examined with direct sequencing.

Analysis of methylation status

Methylation status of the maspin promoter was examined in representative cases, including eight cases with uniformly positive expression, five cases with completely negative expression and three cases with heterogeneous expression, where the positive and negative portions were analysed separately. Genomic DNA was microdissected from frozen tissues mounted in OCT compound, to ensure that over 90% of the extract was derived from tumor cells, using a laser captured microdissection system. Then, bisulfite genomic sequencing was performed as previously described (Yatabe *et al.*, 2001). Briefly, converted DNA was amplified with maspin-specific primers based on the sequence obtained from Genebank (Accession NT_033907). The PCR product contained 13 CpG sites, was 153 bp downstream from an *Msp*I site (Zou *et al.*, 1994; Futscher *et al.*, 2002), and the product was located between the p53 binding site 1 and 2 (Zou *et al.*, 2000). The primer sequences were: forward, 5′-TGTTAAGAGGTTTGAGTAGGAGAGG-3′ and reverse, 5′-CCCACCTTACTTACCTAAAATCACAAT-3′. Amplified products were cloned (TOPO TA cloning kit, Invitrogen, Carlsbad, CA, USA), and five or more clones per case were sequenced. The averaged percentage of CpG sites was considered for methylation status of the case or tumor portion.

Statistical analysis

The χ^2 test for independence and unpaired *t*-test compared incidences of maspin expression and frequencies of clinicopathologic variables. A *P*-value below 0.05 was considered statistically significant.

Acknowledgements
We thank Richard D Iggo for the kind gift of the FASAY plasmids and yeast strains, Kaori Hayashi for excellent technical assistance with the molecular genetic experiments and Hiroji Ishida for assistance in constructing the tissue array. This study is supported by Grant-in-Aid for Encouragement of Young Scientists (B) from the Ministry of Education, Science, Sports and Culture, Japan.

References

Achiwa H, Yatabe Y, Hida T, Kuroishi T, Kozaki K, Nakamura S, Ogawa M, Sugiura T, Mitsudomi T and Takahashi T. (1999). *Clin. Cancer Res.*, **5,** 1001–1005.

Bass R, Fernandez AM and Ellis V. (2002). *J. Biol. Chem.*, **277,** 46845–46848.

Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ and Meyerson M. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 13790–13795.

Bieche I, Girault I, Sabourin JC, Tozlu S, Driouch K, Vidaud M and Lidereau R. (2003). *Br. J. Cancer*, **88,** 863–870.

Day ML, Zhao X, Vallorosi CJ, Putzi M, Powell CT, Lin C and Day KC. (1999). *J. Biol. Chem.*, **274,** 9656–9664.

Domann FE, Rice JC, Hendrix MJ and Futscher BW. (2000). *Int. J. Cancer*, **85,** 805–810.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H and Domann FE. (2002). *Nat. Genet.*, **31,** 175–179.

Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D and Petersen I. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 13784–13789.

Goldstein NS and Thomas M. (2001). *Am. J. Clin. Pathol.*, **116,** 319–325.

Graff JR, Gabrielson E, Fujii H, Baylin SB and Herman JG. (2000). *J. Biol. Chem.*, **275,** 2727–2732.

Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG and Morrow M (eds) (2002). *AJCC Cancer Staging Manual*, 6th edn. Springer-Verlag: New York.

Ishioka C, Frebourg T, Yan YX, Vidal M, Friend SH, Schmidt S and Iggo R. (1993). *Nat. Genet.*, **5,** 124–129.

Jiang N, Meng Y, Zhang S, Mensah-Osman E and Sheng S. (2002). *Oncogene*, **21,** 4089–4098.

Jones PA and Baylin SB. (2002). *Nat. Rev. Genet.*, **3,** 415–428.

Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, Guise TA and Massague J. (2003). *Cancer Cell*, **3,** 537–549.

Kawai J, Hirotsune S, Hirose K, Fushiki S, Watanabe S and Hayashizaki Y. (1993). *Nucleic Acids Res.*, **21,** 5604–5608.
Laird PW. (2003). *Nat. Rev. Cancer*, **3,** 253–266.
Lau SK, Desrochers MJ and Luthringer DJ. (2002). *Mod. Pathol.*, **15,** 538–542.
Li E, Beard C and Jaenisch R. (1993). *Nature*, **366,** 362–365.
Maass N, Biallek M, Rosel F, Schem C, Ohike N, Zhang M, Jonat W and Nagasaki K. (2002). *Biochem. Biophys. Res. Commun.*, **297,** 125–128.
Maass N, Hojo T, Ueding M, Luttges J, Kloppel G, Jonat W and Nagasaki K. (2001). *Clin. Cancer Res.*, **7,** 812–817.
Marchetti A, Buttitta F, Pellegrini S, Chella A, Bertacca G, Filardo A, Tognoni V, Ferreli F, Signorini E, Angeletti CA and Bevilacqua G. (1996). *J. Pathol.*, **179,** 254–259.
Markl ID, Cheng J, Liang G, Shibata D, Laird PW and Jones PA. (2001). *Cancer Res.*, **61,** 5875–5884.
Mohandas T, Sparkes RS and Shapiro LJ. (1981). *Science*, **211,** 393–396.
Nass SJ, Herman JG, Gabrielson E, Iversen PW, Parl FF, Davidson NE and Graff JR. (2000). *Cancer Res.*, **60,** 4346–4348.
Nishio M, Koshikawa T, Yatabe Y, Kuroishi T, Suyama M, Nagatake M, Sugiura T, Ariyoshi Y, Mitsudomi T and Takahashi T. (1997). *Clin. Cancer Res.*, **3,** 1051–1058.
Oshiro MM, Watts GS, Wozniak RJ, Junk DJ, Munoz-Rodriguez JL, Domann FE and Futscher BW. (2003). *Oncogene*, **22,** 3624–3634.
Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. (2000). *Nature*, **406,** 747–752.
Sheng S, Carey J, Seftor EA, Dias L, Hendrix MJ and Sager R. (1996). *Proc. Natl. Acad. Sci. USA*, **93,** 11669–11674.
Sheng S, Pemberton PA and Sager R. (1994). *J. Biol. Chem.*, **269,** 30988–30993.
Sood AK, Fletcher MS, Gruman LM, Coffin JE, Jabbari S, Khalkhali-Ellis Z, Arbour N, Seftor EA and Hendrix MJ. (2002). *Clin. Cancer Res.*, **8,** 2924–2932.
Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL and Botstein D. (2003). *Proc. Natl. Acad. Sci. USA*, **100,** 8418–8423.
Tsuchiya E, Furuta R, Wada N, Nakagawa K, Ishikawa Y, Kawabuchi B, Nakamura Y and Sugano H. (1995). *J. Cancer Res. Clin. Oncol.*, **121,** 577–581.
Umekita Y, Ohi Y, Sagara Y and Yoshida H. (2002). *Int. J. Cancer*, **100,** 452–455.
Waridel F, Estreicher A, Bron L, Flaman JM, Fontolliet C, Monnier P, Frebourg T and Iggo R. (1997). *Oncogene*, **14,** 163–169.
Yatabe Y, Hida T, Achiwa H, Muramatsu H, Kozaki K, Nakamura S, Ogawa M, Mitsudomi T, Sugiura T and Takahashi T. (1998a). *Cancer Res.*, **58,** 3761–3764.
Yatabe Y, Konishi H, Mitsudomi T, Nakamura S and Takahashi T. (2000). *Am. J. Pathol.*, **157,** 985–993.
Yatabe Y, Masuda A, Koshikawa T, Nakamura S, Kuroishi T, Osada H, Takahashi T, Mitsudomi T and Takahashi T. (1998b). *Cancer Res.*, **58,** 1042–1047.
Yatabe Y, Mitsudomi T and Takahashi T. (2002). *Am. J. Surg. Pathol.*, **26,** 767–773.
Yatabe Y, Tavare S and Shibata D. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 10839–10844.
Zhang M, Volpert O, Shi YH and Bouck N. (2000). *Nat. Med.*, **6,** 196–199.
Zou Z, Anisowicz A, Hendrix MJ, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E and Sager R. (1994). *Science*, **263,** 526–529.
Zou Z, Gao C, Nagaich AK, Connell T, Saito S, Moul JW, Seth P, Appella E and Srivastava S. (2000). *J. Biol. Chem.*, **275,** 6051–6054.

Oncogene (2004), 1–11

www.nature.com/onc

ORIGINAL PAPER

Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients

Shuta Tomida[1,6], Katsumi Koshikawa[1,6], Yasushi Yatabe[2], Tomoko Harano[1], Nobuhiko Ogura[3], Tetsuya Mitsudomi[4], Masato Some[3], Kiyoshi Yanagisawa[1], Toshitada Takahashi[5], Hirotaka Osada[1] and Takashi Takahashi*,[1]

[1]*Division of Molecular Oncology, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan;* [2]*Department of Pathology and Molecular Diagnostics, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan;* [3]*Miyanodai Technology Development Center, Fuji Photo Film Co., Ltd., Kanagawa 258-8538, Japan;* [4]*Department of Thoracic Surgery, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan;* [5]*Division of Immunology, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan*

Individualized outcome prediction classifiers were successfully constructed through expression profiling of a total of 8644 genes in 50 non-small-cell lung cancer (NSCLC) cases, which had been consecutively operated on within a defined short period of time and followed up for more than 5 years. The resultant classifier of NSCLCs yielded 82% accuracy for forecasting survival or death 5 years after surgery of a given patient. In addition, since two major histologic classes may differ in terms of outcome-related expression signatures, histologic-type-specific outcome classifiers were also constructed. The resultant highly predictive classifiers, designed specifically for nonsquamous cell carcinomas, showed a prediction accuracy of more than 90% independent of disease stage. In addition to the presence of heterogeneities in adenocarcinomas, our unsupervised hierarchical clustering analysis revealed for the first time the existence of clinicopathologically relevant subclasses of squamous cell carcinomas with marked differences in their invasive growth and prognosis. This finding clearly suggests that NSCLCs comprise distinct subclasses with considerable heterogeneities even within one histologic type. Overall, these findings should advance not only our understanding of the biology of lung cancer but also our ability to individualize postoperative therapies based on the predicted outcome.
Oncogene advance online publication, 5 April 2004; doi:10.1038/sj.onc.1207697

Keywords: lung cancer; gene expression profile; microarray; prognosis

*Correspondence: T Takahashi, Division of Molecular Oncology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan:
E-mail: tak@aichi-cc.jp
[6]ST and KK contributed equally to this work
Received 5 January 2004; revised 23 February 2004; accepted 1 March 2004

Introduction

Lung cancer is the most prevalent and deadly cancer in economically developed countries. In the US, more than 170 000 people die annually due to lung cancer, while Japan is facing a continuing and steep increase in lung cancer deaths with loss of more than 51 000 lives annually. Lung cancers are classified into two major groups, small-cell lung cancer and non-small-cell lung cancer (NSCLC), based on their clinicopathologic characteristics. Owing to the considerable differences in their etiologies, genetic and epigenetic changes and clinicopathologic features, NSCLCs are often further subgrouped into two major categories, that is, squamous cell carcinomas and nonsquamous cell carcinomas (Minna *et al.*, 1997; Osada and Takahashi, 2002).

Surgical resection gives the best hope for a cure for NSCLC cases, which comprise 80–85% of lung cancers. Unfortunately, however, their long-term survival rate remains unsatisfactory and no more than 50% of the cases that have successfully undergone potentially curative resection can survive for 5 years after operation. The TNM classification according to tumor size, extent of nodal involvement and the presence or absence of distant metastasis is routinely used as a prognostic tool, but it only provides information about what percentage of cases at a particular disease stage, which a given patient belongs to, can be expected to be alive at specific postoperative time points. Although various genetic and epigenetic changes of cancer-related genes have been identified and examined in the search for clinically relevant prognosticators, no single variable thus far evaluated has proven to be sufficiently predictive to accurately forecast a patient's outcome (Slebos *et al.*, 1990; Horio *et al.*, 1993; Mitsudomi *et al.*, 2000).

Recent rapid progress in microarray technology has made it possible to analyse gene expression profiles on a genome-wide basis in order to search for molecular markers for cancer classification and outcome prediction (Golub *et al.*, 1999; Alizadeh *et al.*, 2000; Perou *et al.*, 2000; Khan *et al.*, 2001; Pomeroy *et al.*, 2002; Ye *et al.*, 2003). While lung cancers are known to be very

heterogeneous in various aspects as clearly mentioned, recent expression profiling studies have shown the presence of several subclasses with distinct expression profiles not only among NSCLC but also and specifically among adenocarcinoma (Bhattacharjee *et al.*, 2001; Garber *et al.*, 2001; Beer *et al.*, 2002; Virtanen *et al.*, 2002). However, the existence of clinically relevant subclasses of squamous cell carcinomas in terms of expression profiles has not been detected thus far, while development of an individualized outcome classifier is eagerly awaited in the hope of implementing tailor-made treatments for this fatal disease.

In the study presented here, a consecutively operated, well-defined cohort of 50 NSCLC cases, followed up for more than 5 years, was used to acquire expression profiles of a total of 8644 unique genes, leading to the successful construction of supervised learning method-based individualized outcome prediction classifiers with high accuracy. In addition, the identification by means of unsupervised hierarchical clustering analysis of two distinct subclasses of squamous cell carcinomas with interesting histologic distinctions and a significant difference in their prognosis is also reported.

Results

Unsupervised hierarchical clustering of NSCLCs

We first applied unsupervised hierarchical clustering algorithm to a comparison of gene expression profiles of 50 NSCLC cases using the filtered 900 spots corresponding to 829 unique genes that had passed our filtering procedures. The resulting clusters recapitulated the distinctions between well-established histologic classes of NSCLCs, that is, squamous cell and nonsquamous cell carcinomas (Figures 1a–f), and the results of immunohistochemical analysis of TTF-1 and cytokeratin 7 (KRT7) expression, which used tissue microarrays containing all of the cases examined in this study, were consistent with those of the cDNA microarray analysis (data not shown). We noted that squamous cell carcinomas formed a more discrete cluster than nonsquamous cell carcinomas, demonstrating the generally strong expression of several well-accepted differentiation markers of squamous epithelium including cytokeratins 5 and 13, S100A2 and galectin 7 (LGALS7) (Figure 1d) (Magnaldo *et al.*, 1998; Chu and Weiss, 2002). Markedly high-level expression of small proline-rich proteins such as SPRR1B and SPRR2C was also generally observed in squamous cell carcinomas (Hu *et al.*, 1998). While squamous cell carcinomas were clustered as a large single cluster, there appeared to be two major subclasses of squamous cell carcinomas in relation to the gene expression profiles, and the distinctions between the two subclasses could be even more clearly illustrated in the separate hierarchical clustering analysis of squamous cell carcinomas (data not shown). Squamous cell carcinoma cluster 1 (SQ1) expressed a characteristic set of genes that are related to extracellular matrix proteins, including fibronectin and various collagen subunits such as α2(I) and α1(III) (Figure 1f), as well as those related to DNA replication and cell proliferation including ARAF1 and EPHA1 (Figure 1e) (Shelton *et al.*, 2003; Walker-Daniels *et al.*, 2003). Interestingly, patients belonging to the SQ1 cluster had a significantly poorer prognosis after potentially curative resection than those of SQ2 (Figure 2a, $P=0.013$ by log-rank test). Another prominent distinction between SQ1 and SQ2 was their growth patterns. Seven of the eight (88%) cases of SQ1 showed an infiltrative growth pattern with dense fibrous stroma in contrast to lack of these features in six of the seven SQ2 tumors ($P=0.010$ by Fisher's exact test), which showed a rather well-circumscribed, expansive growth pattern with inconspicuous keratinization (Figure 2b).

Adenocarcinomas, a major fraction of nonsquamous cell carcinomas, also appeared to be subclassifiable into two major groups according to their expression profiles: adenocarcinoma cluster 1 (AD1) clustered together with squamous cell carcinomas, and AD2, 3 and 4, which formed another distinct cluster. Of the latter subclass, AD3 most typically showed strong expression of peripheral lung cell markers including TTF-1 (TITF1), SP-C (SFTPC) and SP-A (SERPA2) (Figure 1b) (Yatabe *et al.*, 2002). Patients belonging to the AD3 cluster were predominantly female nonsmokers with a well-differentiated adenocarcinoma. In addition, 80% (eight of 10) of the AD3 cases were female subjects in contrast to 29% (six of 21) in non-AD3 adenocarcinoma clusters ($P=0.018$, Fisher's exact test), while well-differentiated tumors were more frequent in AD3 (60%) cases than other adenocarcinoma clusters (14%) ($P=0.015$ by Fisher's exact test). A total of 70% of the AD3 cases were found to be nonsmokers, and this ratio was considerably higher than that for non-AD3 clusters (38%), which in fact contained significantly more heavy smokers (>20 pack years) than AD3 ($P=0.046$ by Fisher's exact test). In marked contrast, AD1 showed low-level expression of peripheral lung markers such as TTF-1 and SP-C and to some extent shared gene expression profiles with squamous cell carcinomas, although they were still clearly distinct. Tumors corresponding to AD2 characteristically showed a strong expression of a number of expressed sequence tags (ESTs), whereas expression of the same set of ESTs was found to be distinctly reduced in AD4 as well as SQ1, suggesting that characterization of these ESTs may shed light on the biology of these tumor subclasses (Figure 1c).

Construction of a survival prediction model for NSCLCs

In order to develop an individualized outcome prediction classifier, we used a signal-to-noise metric first to select genes that most clearly distinguished prognosis-favorable from fatal patients. The unsupervised hierarchical clustering algorithm using the top 100 spots corresponding to 98 unique genes yielded two major branches representing those with favorable and fatal prognosis (Figure 3a). Of the 21 patients in the left or

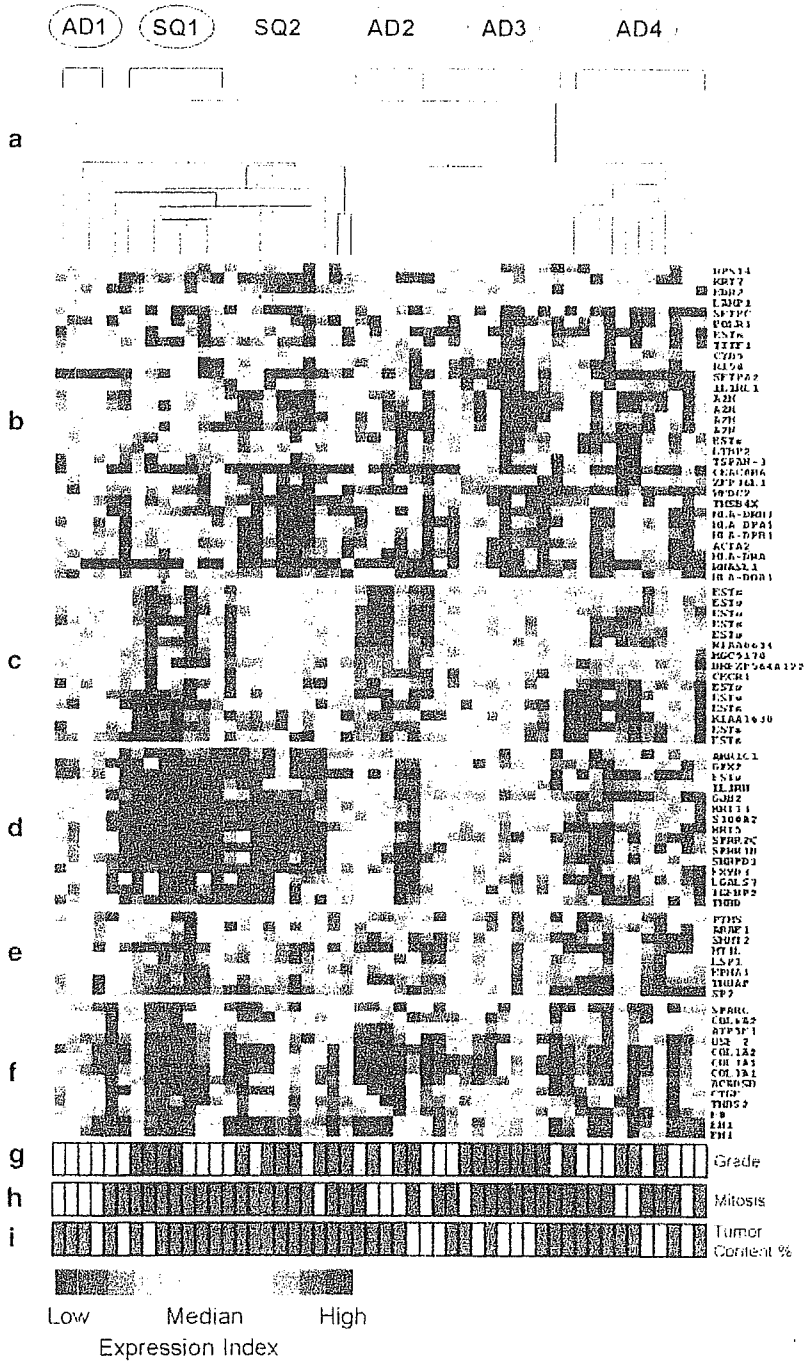

Figure 1 Hierarchical clustering defining subclasses of NSCLCs. (**a**) Dendrogram of two-dimensional hierarchical clustering analysis of the 900 filtered genes across 50 samples of NSCLCs. The expression index for the transcript sequences (rows) in the samples (columns) is indicated by a color code (see expression index in the bar at the bottom of this figure). (**b**) A cluster of genes with high relative expression in adenocarcinomas. (**c**) A gene cluster relevant to AD2. (**d**) A cluster of genes highly expressed in squamous cell lung carcinomas. (**e**, **f**) Cluster of genes with high-level expression in SQ1. (**g**) Histopathologic grade (orange, poor; yellow, moderate; green, well). (**h**) Mitotic indices (green, less than 10 in 10 high-power field; yellow, 10–40; orange, over 40). (**i**) Estimated nucleated tumor content (yellow, 50–74%; orange, more than 75%)

'favorable' branch, 19 had survived 5 years after surgery, while 15 of the 29 cases in the right or 'fatal' branch had died within 5 years after surgery. Their Kaplan–Meier survival curves showed a statistically significant difference (Figure 3b, $P=0.0025$, log-rank test).

Since the ultimate aim was the development of an individualized outcome prediction classifier, we next applied a supervised learning method. To this purpose, a weighted-voting outcome classifier was constructed based on the predictive genes, which were preselected using a signal-to-noise metric. The learning errors for

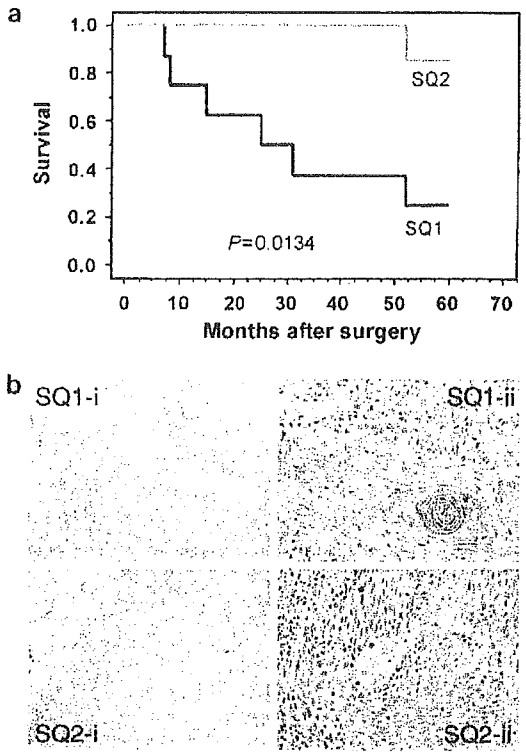

Figure 2 Clinicopathologic distinctions between the SQ1 and SQ2 clusters. (**a**) Kaplan–Meier survival curves of SQ1 and SQ2 showing a significantly different prognosis. (**b**) Representative morphologies of tumors in the SQ1 and SQ2 clusters. Tumors belonging to the SQ1 cluster frequently exhibited invasive growth to surrounding normal lung tissues with active stromal reaction (SQ1-i), while prominent keratinization was a characteristic of SQ1 (SQ1-ii). Tumors of the SQ2 cluster were generally well circumscribed due to their expansive growth (SQ2-i), and minimal stromal reaction is observed in this type. In addition, keratinization and nuclear pleomorphisms were inconspicuous in tumors of SQ2 (SQ2-ii)

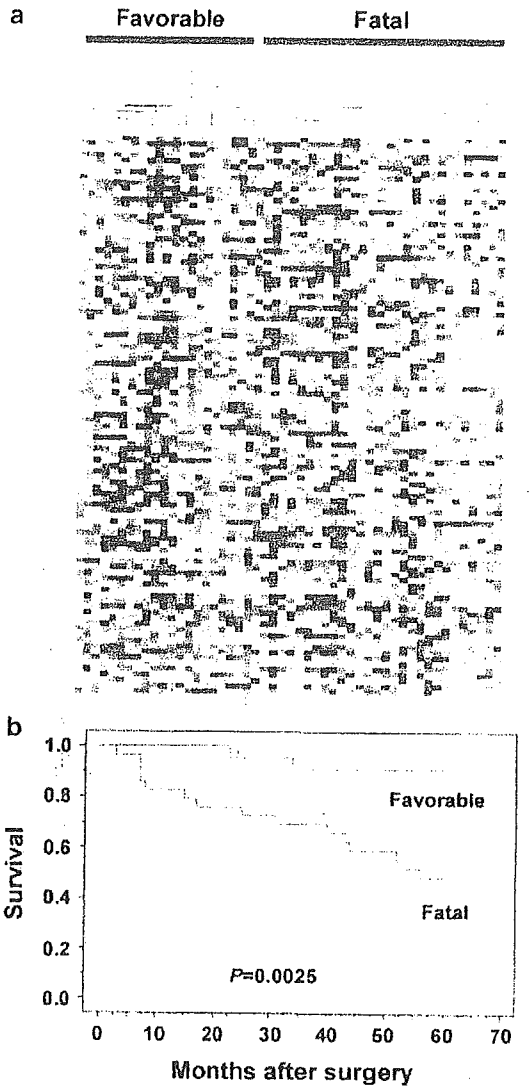

Figure 3 Unsupervised hierarchical clustering of 50 NSCLCs according to the expression of top 100 genes by signal-to-noise metric, which distinguish fatal and favorable prognosis. (**a**) Two-dimensional hierarchical with the top 100 predictive genes. (**b**) Kaplan–Meier survival curves for favorable *versus* fatal groups of 50 samples characterized by hierarchical clustering using the top 100 predictive genes

each model, to which increasing numbers of the preselected predictive genes were applied, calculated with the leave-one-out cross-validation method. The weighted-voting model using 25 predictive genes showed the highest accuracy for the prediction of an individual's outcome (Figure 4a), that is, in 41 of 50 (82%) cases. The 25 genes used for constructing the outcome classifier of NSCLCs are listed in Table 1. With this classifier, 27 of 33 (82%) cases, who actually had survived longer than 5 years after surgery, were judged to have a 'favorable' prognosis, while 14 of the 17 (82%) patients who had died within 5 years were correctly predicted as having a 'fatal' outcome (Figure 4b). Survival curves of the patients with 'favorable' and 'fatal' predictions are plotted in Figure 4c, showing a significant difference between the groups ($P = 6.0 \times 10^{-6}$). We also employed other supervised learning algorithms, including the support vector machine and k-nearest neighbors, using increasing numbers of genes as mentioned above. Although accuracies of these models were comparable to that of the weighted-voting outcome classifier, the latter achieved the highest accuracy (data not shown).

Construction of the survival prediction models specific for squamous or nonsquamous cell carcinoma cases

It is generally accepted that squamous cell and nonsquamous cell carcinomas are distinct disease entities in terms of their clinicopathologic features as well as their etiologies (Minna *et al.*, 1997; Osada and Takahashi, 2002). This concept has been supported by the presence of clearly discriminating expression profiles for these two tumor types (Bhattacharjee *et al.*, 2001; Garber *et al.*, 2001; Kikuchi *et al.*, 2003). We therefore constructed outcome prediction classifiers specific for each tumor subtype by using the weighted-voting algorithm and the predictive genes for each subtype selected by signal-to-noise metric. Leave-one-out cross-validation of the classifiers with increasing numbers of

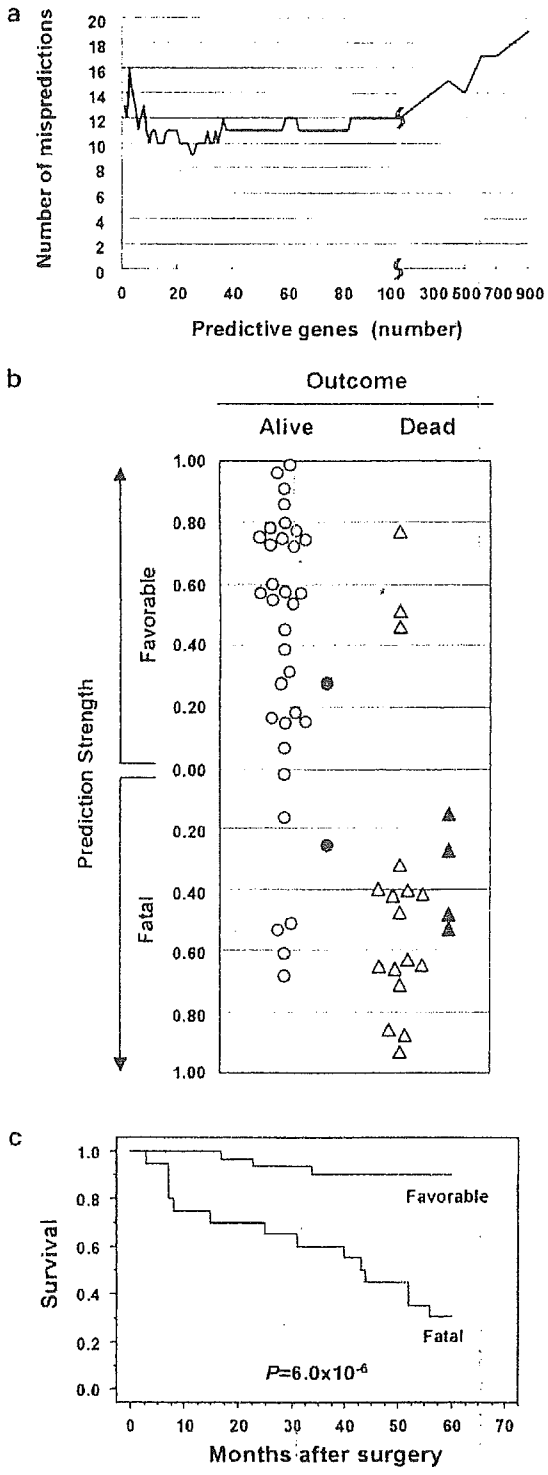

Figure 4 Results of outcome classification by the weighted-voting algorithm using genes preselected with the signal-to-noise metric. (**a**) Search for optimum number of genes for outcome classification. (**b**) Assessment of the outcome classifier using weighted voting algorithm. Open (○) and closed circles (●) indicate patients alive 5 years after surgery, while open (△) and closed triangles (▲) correspond to patients dead within 5 years after surgery. Open marks (○ and △) indicate 50 samples used for leave-one-out cross-validation, while closed marks (● and ▲) indicate additional blinded samples used for independent validation. (**c**) Kaplan–Meier survival curves for the favorable and fatal groups in 50 NSCLC cases predicted by the weighted-voting outcome classifier

the higher-ranked predictive genes yielded the highest accuracies with 12 genes for nonsquamous and 19 genes for squamous cell carcinomas (Tables 2 and 3). Results showed that patients' outcome 5 years after surgery were correctly predicted for 31 of 34 (91%) nonsquamous cell carcinoma cases (Figure 5a). Of the 25 cases judged as having 'favorable' prognosis, 23 (92%) had indeed survived 5 years after surgery, whereas only one of nine (11%) patients predicted as being 'fatal' had survived for 5 years. The differences between survival curves of the 25 patients with the 'favorable' and the nine cases with the 'fatal' prediction were highly significant (Figure 5b, $P = 2.4 \times 10^{-7}$). In the case of squamous cell carcinomas, the weighted-voting model successfully predicted 5-year outcome for 15 of 16 (94%) patients (Figure 5c). Eight of eight (100%) samples judged as being 'favorable' had survived 5 years after surgery, while seven of eight (88%) samples predicted as being 'fatal' had actually died within 5 years. Survival curves based on these predictions showed highly significant differences between the favorable and fatal groups (Figure 5d, $P = 3.3 \times 10^{-5}$).

It was noted that the 12 and 19 genes, which were used to construct the respective outcome classifiers for nonsquamous and squamous cell carcinomas, did not show any overlaps, and histologic-type-specific outcome classifiers contained only two genes each of the top 100, signal-to-noise-preselected genes of the other type of NSCLCs.

Confirmation of the robustness of the classifiers by using an independent blinded validation dataset as well as by random permutation tests

To validate the prognosis classifiers described here, we analysed an additional independent set of six NSCLC cases, three nonsquamous and squamous cell carcinomas each, in a completely blinded fashion. The outcome classifier for both types of NSCLCs correctly predicted the disease outcome in five of the six cases, including the death of a patient with a stage I tumor (Figure 4b). The outcome classifier for nonsquamous cell carcinomas accurately predicted both deaths and one survival (Figure 5a). The outcome classifier for squamous cell carcinomas correctly predicted the outcome for one of the three cases (Figure 5c).

We also performed permutation tests to determine whether the gene sets used in the outcome classifiers were robust, and found highly significant results ($P < 0.0001$ in all three classifiers). Through permutation tests analysing 10 000 datasets with permutated class labels, it was shown that outcome classifiers performing with equal or higher accuracy for the initial dataset as well as for the validation dataset could not be obtained by chance association in either NSCLCs ($P = 0.0104$) or in nonsquamous cell carcinomas ($P = 0.0055$). This type of estimate, however, did not support the statistical significance of the results obtained with squamous cell carcinoma classifier.

Table 1 A total of 25 predictive genes selected for the weighted-voting outcome classifier of NSCLCs

Rank	*Gene*	*Description*	*UniGene ID*	*Class correlation*	P*	*Rank* *Non-SQ*[a]	*SQ*[a]
1	WEE1	WEE1 homolog	Hs.75188	Fatal	0.0027	2	148
2	MYC	v-myc viral oncogene homolog	Hs.79070	Fatal	0.0057	14	55
3	TITF1	Thyroid transcription factor 1	Hs.197764	Favorable	0.0085	31	65
4	FOSL1	FOS-like antigen 1 (Fra-1)	Hs.283565	Fatal	0.0062	30	135
5	LYPLA1	Lysophospholipase I	Hs.393360	Fatal	0.0081	26	123
6	SSBP1	Single-stranded DNA binding protein	Hs.923	Fatal	0.0199	3	641
7	SFTPC	Surfactant, pulmonary-associated protein C	Hs.1074	Favorable	0.0113	12	290
8	THBD	Thrombomodulin	Hs.2030	Fatal	0.0099	51	14
9	NICE-4	NICE-4 protein	Hs.8127	Fatal	0.0099	1	294
10	PTN	Pleiotrophin (heparin binding growth factor 8)	Hs.44	Fatal	0.0100	38	139
11	SNRPB	Small nuclear ribonucleoprotein polypeptides B and B1	Hs.83753	Fatal	0.0115	10	621
12	NAP1L1	Nucleosome assembly protein 1-like 1	Hs.302649	Fatal	0.0131	46	116
13	CTNND1	Catenin delta 1	Hs.166011	Fatal	0.0120	50	149
14	CCT3	Chaperonin containing TCP1, subunit 3	Hs.1708	Fatal	0.0186	16	515
15	ESTs		Hs.95612	Fatal	0.0160	112	25
16	SPRR1B	Small proline-rich protein 1B (cornifin)	Hs.1076	Fatal	0.0209	70	6
17	COPB	Coatomer protein complex, subunit beta	Hs.3059	Fatal	0.0195	151	17
18	ARG1	Arginase type I (liver)	Hs.332405	Fatal	0.0193	28	435
19	ARCN1	Archain 1 (coatomer protein complex, subunit delta)	Hs.33642	Fatal	0.0169	33	174
20	MST1	Macrophage stimulating 1	Hs.349110	Fatal	0.0193	7	684
21	SERPINE1	Serine (or cysteine) proteinase inhibitor, clade E member 1	Hs.82085	Fatal	0.0194	17	532
22	SERPINB1	Serine (or cysteine) proteinase inhibitor, clade B member 1	Hs.183583	Fatal	0.0205	188	31
23	ESTs		Hs.215113	Favorable	0.0205	36	254
24	ACTR3	Actin-related protein 3 homolog (ARP3)	Hs.380096	Fatal	0.0229	99	67
25	PTP4A3	Protein tyrosine phosphatase type 4A, member 3	Hs.43666	Fatal	0.0199	171	5

*P-values were assigned based upon the frequency, with which signal-to-noise statistics, which were calculated for the 10 000 permutations of the sample labels in each of the nonsquamous cell and squamous cell carcinoma datasets, yielded better results than the actual signal-to-noise statistic. [a]Ranking of the signal-to-noise statistics for the genes in the nonsquamous cell and squamous cell carcinoma datasets

Table 2 A total of 12 predictive genes selected for the weighted-voting outcome classifier of nonsquamous cell carcinomas

Rank	*Gene*	*Description*	*UniGene ID*	*Class correlation*	P*	*Rank in SQ*[a]
1	NICE-4	NICE-4 protein	Hs.8127	Fatal	0.0036	294
2	WEE1	WEE1 homolog	Hs.75188	Fatal	0.0039	148
3	SSBP1	Single-stranded DNA binding protein	Hs.923	Fatal	0.0122	641
4	WFDC2	WAP four-disulfide core domain 2	Hs.2719	Favorable	0.0155	559
5	ACTA2	Actin, alpha 2, smooth muscle, aorta	Hs.195851	Favorable	0.0149	50
6	G22P1	Thyroid autoantigen 70kDa (Ku70)	Hs.197345	Fatal	0.0176	256
7	MST1	Macrophage stimulating 1	Hs.349110	Fatal	0.0153	684
8	PHB	Prohibitin	Hs.75323	Fatal	0.0219	85
9	DRPLA	Dentatorubral-pallidoluysian atrophy	Hs.169488	Fatal	0.0238	749
10	SNRPB	Small nuclear ribonucleoprotein polypeptides B and B1	Hs.83753	Fatal	0.0192	621
11	GJA	Gap junction protein, alpha 1	Hs.74471	Fatal	0.0268	295
12	SFTPC	Surfactant, pulmonary-associated protein C	Hs.1074	Favorable	0.0313	290

*P-values were assigned based upon the frequency, with which signal-to-noise statistics, which were calculated for the 10 000 permutations of the sample labels in the nonsquamous cell carcinoma dataset, yielded better results than the actual signal-to-noise statistic. [a]Ranking of the signal-to-noise statistics for the genes in the squamous cell carcinoma dataset

Discussion

A patient's survival is currently estimated on the basis of empirical information in the TNM classification system about survival distributions according to disease stages. Since survival or death is a matter of all or nothing, currently available information regarding what percentage of those at a certain disease stage are likely to survive after a certain period of time is insufficient in many respects. The study presented here described the successful construction of individualized outcome classifiers as a result of using the gene expression profiling data- and weighted-voting algorithm-based approach, which allowed us to predict outcomes for surgically treated lung cancer patients on an individual basis. Although our outcome classifiers were constructed by using a modestly sized cohort, highly robust prediction models were obtained especially for the prediction of outcome for nonsquamous cell carcinomas, which constitute about 60% of NSCLCs. In fact, patients' 5-year postoperative survival status was correctly predicted for 91% of the leave-one-out cross-validation

Table 3 A total of 12 predictive genes selected for the weighted-voting outcome classifier of squamous cell carcinomas

Rank	*Gene*	*Description*	*UniGene ID*	*Class correlation*	P*	*Rank in non-SQ*[a]
1	ESTs		Hs.28399	Favorable	0.0068	396
2	ESTs		Hs.98269	Favorable	0.0087	178
3	ESTs		Hs.35552	Favorable	0.0034	272
4	KRT5	Keratin 5	Hs.433845	Fatal	0.0046	240
5	PTP4A3	Protein tyrosine phosphatase type 4A, member 3	Hs.43666	Fatal	0.0104	171
6	SPRR1B	Small proline-rich protein 1B	Hs.1076	Fatal	0.0147	70
7	LOC339324	Hypothetical protein LOC339324	Hs.18103	Favorable	0.0171	361
8	MYST4	MYST histone acetyltransferase 4	Hs.27590	Fatal	0.0188	529
9	SPARCL1	SPARC-like 1	Hs.75445	Fatal	0.0210	846
10	IGJ	Immunoglobulin J polypeptide	Hs.76325	Fatal	0.0143	265
11	EIF4A2	Eucaryotic translation initiation factor 4A, isoform 2	Hs.173912	Favorable	0.0233	382
12	ESTs		Hs.261314	Fatal	0.0226	822
13	ID2	Inhibitor of DNA binding 2	Hs.180919	Fatal	0.0214	329
14	THBD	Thrombomodulin	Hs.2030	Fatal	0.0077	51
15	MGC15476	Thymus expressed gene 3-like	Hs.134185	Fatal	0.0231	298
16	ZFP	Zinc-finger protein	Hs.1148	Favorable	0.0217	297
17	COPB	Coatomer protein complex, subunit beta	Hs.3059	Fatal	0.0272	151
18	ZYG	ZYG homolog	Hs.29285	Fatal	0.0237	632
19	CACNA1I	Calcium channel, voltage-dependent, alpha 1I subunit	Hs.125116	Fatal	0.0312	492

*P-values were assigned based upon the frequency, with which the signal-to-noise statistics, which were calculated for the 10 000 permutations of the sample labels of the squamous cell carcinoma dataset, yielded better results than the actual signal-to-noise statistic. [a]Ranking of the signal-to-noise statistics for the genes in the nonsquamous cell carcinoma dataset

dataset, and this prediction accuracy was confirmed for the independent validation dataset. It should be noted that the outcome classifier appeared to perform well independently of disease stage; among 24 stage I/II nonsquamous cell carcinoma cases, 17 survivals and five deaths were correctly predicted with just a single misclassification for each, while accurate prediction was also obtained for nine of the 10 stage III cases. Beer *et al.* (2002) recently described a prediction model, which classifies adenocarcinoma cases into high- and low-risk groups, although their model was not intended to predict 5-year survival for individual cases. The difference in 5-year survival between the high- and low-risk groups can be estimated as 40% based on the Kaplan–Meier survival curves shown in Figure 3d, whereas our nonsquamous cell carcinoma outcome classifier yielded an 81% difference in 5-year survival between the fatal and favorable classifications (Figure 5b). At present, intensive chemotherapy is given only to patients with an advanced metastatic disease. However, the development of highly accurate, individualized outcome classifiers should make it possible to select patients, who are at high risk of future failure and thus most eligible for such intensive adjuvant therapy with the intention of eradicating undetectable micrometastases, sources of future recurrence.

The gene sets used for the outcome classifiers contained various functionally relevant genes (Table 1). For example, MYC and FOSL1 are known to be involved in the process of cellular transformation (Vennstrom *et al.*, 1982; Mechta *et al.*, 1997). The MYC oncogene, frequently overexpressed with or without amplification in lung cancers, is associated with a poor prognosis (Johnson *et al.*, 1987), as is SERPINE1. (Robert *et al.*, 1999) FOSL1 has been shown to play a role as a predominant component of the AP-1 complex in ras-induced transformation (Mechta *et al.*, 1997), and it also mediates induction of SPRR1B in response to epithelial injury caused by a variety of carcinogens (Patterson *et al.*, 2001; Vuong *et al.*, 2002). Both COPB and ARCN1 constitute COPI-coated vesicles and are involved in the transport of G protein-coupled receptors, while the binding of Ras-related GTP-binding protein Cdc42 to another COPI subunit, COPG, has been shown to be important for transmitting transforming signals (Wu *et al.*, 2000). PTN and MST1 are ligands of tyrosine kinase receptors, while it has been suggested that MST1 and its receptor RON form an autocrine/paracrine system involved in cell migration of NSCLC cells (Willett *et al.*, 1998). CCT3 and ACTR3 may also be involved in motile and invasive features, since ACTR3 is a major constituent of the ARP2/3 complex required for protrusion of the lamellipodia (Craig and Chen, 2003), while CCT3 is involved in actin and tubulin folding (Martin-Benito *et al.*, 2002). Increased WEE1 expression in patients with fatal outcome may be a reflection of higher proliferative activity. In this connection, high-level expression of WEE1 in combination with the Cyclin and CDK genes has been reported in hepatocellular carcinoma (Masaki *et al.*, 2003).

Ramaswamy *et al.* (2003) recently suggested the presence of a metastatic signature based on their analysis of lung adenocarcinoma cases and identification of a 17-gene set associated with metastatic capability. It is interesting to note that two closely related genes, that is, members of the actin and small nuclear ribonucleoprotein families, are included as predictors of concordant directions in both the 12-gene set of our nonsquamous cell carcinoma outcome classifier and their 17-gene set predicting metastasis of adenocarcinomas (Tables 2 and 3). In addition to SNRPB, three other genes (EIF4A1, EIF2S2 and HNRPA1) involved in translation apparatus are also included as fatal outcome predictors in our list of top 100 genes, while

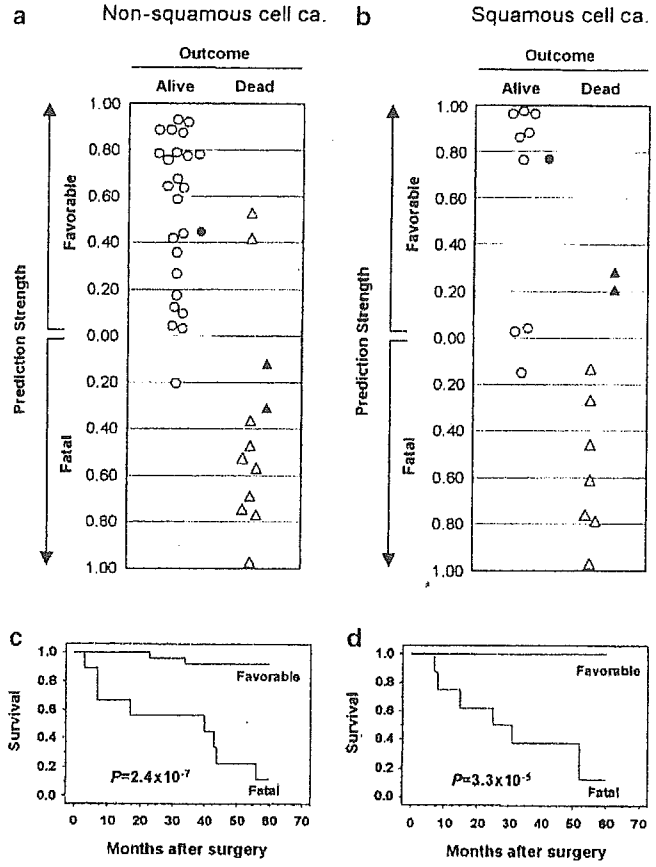

Figure 5 Results of histologic-type-specific outcome classification of nonsquamous cell and squamous cell lung cancers. The weighted-voting supervised learning was conducted using histologic-type-specific sets of genes preselected with the signal-to-noise metric. Circles and triangles indicate the same items as those in Figure 3. (**a**) Results for outcome classifier of nonsquamous cell lung cancer cases used for leave-one-out cross-validation (○ and △) as well as for an additional blinded validation set (● and ▲). (**b**) Kaplan–Meier survival curves of the 34 nonsquamous cell carcinoma cases based on the favorable and fatal outcomes predicted by the weighted-voting outcome classifier. (**c**) Results of outcome classification of squamous cell lung cancer cases used for leave-one-out cross-validation (○ and △) as well as for independent validation with an additional blinded sample set (● and ▲). (**d**) Kaplan–Meier survival curves of 16 squamous cell carcinoma cases based on the favorable and fatal outcomes predicted by the weighted-voting outcome classifier

Ramaswamy *et al.*'s 17-gene set contained two functionally related genes, EIF4EL3 and HNRPAB, suggesting the involvement of their upregulation in tumor progression. Moreover, HLA class II genes are included in both gene sets as favorable predictors. Although the use of different array platforms makes it difficult to evaluate frequencies of overlap of the selected genes in the present study and others, it is interesting that four of the five genes, which are shared by the top 100 genes of the Michigan group (Beer *et al.*, 2002) and ours, appear to be predictive in the same direction.

The outcome classifiers described here were constructed with a relatively modest number of cases, but the use of nonbiased, consecutive cases collected within a short period of time with complete follow-up data may have been a factor in their high accuracy, especially of the outcome classifier for nonsquamous cell carcinomas. This needs to be confirmed, however, in a large, independent cohort of patients, preferably in a prospective manner, before these classifiers can be put to clinical use for the selection of candidates for intensive adjuvant therapies. Nevertheless, the strategy presented here appears to be promising, and suggests that it may also be possible to develop similar predictive classifiers for predicting patients' response to conventional chemotherapy and/or molecular-targeted therapeutics such as those recently introduced involving EGF receptor inhibition. In the future, lung cancer patients who are classified as having fatal but therapeutic agent-responsive expression signatures can be expected to be selectively and intensively treated with the most appropriate active agents in an individualized manner, ultimately leading to significant improvement in the present dismal survival rate.

Our unsupervised hierarchical clustering analysis underscores previous observations that adenocarcinomas and squamous cell carcinomas are very distinct in terms of gene expression profiles (Bhattacharjee *et al.*, 2001; Garber *et al.*, 2001; Virtanen *et al.*, 2002; Kikuchi *et al.*, 2003). In addition, we observed that genes useful for constructing nonsquamous cell and squamous cell carcinoma outcome classifiers did not show any overlap, providing further support to the notion of two distinct entities. Interestingly, the unsupervised hierarchical clustering analysis used in our study demonstrated for the first time the existence of two clinically relevant subsets of squamous cell carcinomas with distinct gene expression signatures and markedly different clinicopathologic features. SQ1 showed a characteristic invasive growth pattern and a significantly worse prognosis than SQ2 (Figure 2). The most noticeable feature of the expression signature of SQ1 was high-level expression of extracellular matrix proteins, including fibronectin, SPARC and various collagen subunits such as α2(I) and α1(III) (Figure 1f). By using an expression profiling analysis of a *in vivo*-selected, highly metastatic melanoma model, Clark *et al.* (2000) found that enhanced expression of genes involved in the extracellular matrix assembly, including fibronectin, collagen α2(I) and α1(III), was associated with the acquisition of a metastatic phenotype during the *in vivo* selection, a finding consistent with the invasive growth pattern and poor prognosis of SQ1 in our study. Ramaswamy *et al.* (2003) also found a similar association between metastatic potential and collagen gene expression in the molecular signature analysis of solid tumors. In addition, expression of a set of ESTs was found to be distinctly reduced in SQ1 when compared with SQ2 (Figure 1c), suggesting that the reduced expression of genes corresponding to these ESTs may be related to the aggressive nature of SQ1. Although the gene expression-based subclasses of squamous cell carcinomas were identified by using a small cohort of squamous cell carcinomas, the present findings warrant further validation and characterization.

In view of their distinct molecular signatures, the existence of a few subclasses of adenocarcinomas has been reported (Bhattacharjee *et al.*, 2001; Garber *et al.*, 2001; Beer *et al.*, 2002; Virtanen *et al.*, 2002), although the expression signatures identified in these studies did not show complete correspondence with each other or with those identified by us. However, there is a high degree of consistency in the fact that all these studies have identified a subset of adenocarcinomas with a high-level expression of genes related to differentiation of normal peripheral lung epithelial cells such as TTF-1 and SP-C. This subclass of adenocarcinomas is most typically represented by AD3 in our study, which appears to correspond to AD group 2 in the Stanford study and to C4 in the MIT study. Consistently notable features of this subclass are significantly higher proportions of females, non/light smokers and well-differentiated tumors. These findings are consistent with our previous observations that there is a subset of adenocarcinomas which may represent those arising from cells committed to become peripheral lung epithelial cells under weaker influence of smoking (Yatabe *et al.*, 2002).

In conclusion, we have been successfully constructing an individualized outcome prediction classifier, which appears to be useful especially for nonsquamous cell carcinomas. In addition, we have been able to demonstrate the presence of marked heterogeneities of NSCLCs by identifying clinically relevant subclasses not only for adenocarcinomas but also for squamous cell carcinomas. Further studies appear to be warranted on the basis of these findings, with the aim to improve both understanding and survival of this fatal disease.

Materials and methods

Tissue sample

A total of 50 NSCLC tissue specimens (30 adenocarcinomas, 16 squamous cell carcinomas and four large cell carcinomas) were obtained from 15 female and 35 male patients, who had consecutively undergone potentially curative resection at Aichi Cancer Center Hospital (Nagoya, Japan). The vast majority of these lung cancer specimens were acquired within an 11-month period between December 1995 and October 1996, while squamous cell and large cell carcinomas were also obtained during an additional few months around this period of time. An additional three cases each of adenocarcinomas and squamous cell carcinomas, which were collected just before or after the specimen collection period, were used for an independent validation dataset. The tumor specimens were embedded in OCT compound and stored at −80°C until use with the approval of the institutional review board. The median age of this cohort was 63 (range: 43–76), and there were 23 pStage I tumors, 11 pStage II, and 16 pStage III. Approvals for this study were obtained from the institutional review board of the Aichi Cancer Center.

RNA extraction and microdissection

On average, 50 7 μm-thick cryostat sections were prepared for the extraction of total RNA from tumor cell-rich areas, which were identified by a surgical pathologist (YY) using every 10th section stained by May–Giemsa. Careful attention was paid to the microdissection of only such tumor cell-rich areas, yielding an average of 75.4% tumor cell content (Figure 1i). RNAs were then isolated using RNAeasy (Quiagen, Valencia, CA, USA) according to the manufacturer's instruction, and their quality was checked with the RNA 6000 Nano Assay kit and the 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA).

Microarray experiments and acquisition of datasets

We used two sets of membrane cDNA microarrays (GeneFilter Human Microarrays Release I and Release II; Invitrogen, Carlsbad, CA, USA), which contained a total of 11 168 spots corresponding to 8644 independent genes. A measure of 5 μg of total RNA was reverse-transcribed using oligo-dT primers (Invitrogen) and SuperScript II reverse transcriptase (Invitrogen) in the presence of 100 μCi of [^{33}P]dCTP (Amersham Bioscience, Piscataway, NJ, USA) according to the instructions for the GeneFilters (Invitrogen) with slight modification. The GeneFilters were prehybridized for 2 h at 51°C with 0.5 μg/ml of poly-dA (Invitrogen) and 0.5 μg/ml Cot-1 DNA (Invitrogen) in 10 ml of AlkPhos DIRECT hybridization buffer (Amersham Bioscience) and then hybridized for 17 h at 51°C with the denatured radiolabeled probe. Hybridization of the arrays was followed by two washings with a solution containing 2 M of urea, 0.1 % of SDS, 50 mM of Na phosphate buffer (pH 7.0), 150 mM of NaCl, 1 mM of $MgCl_2$ and 0.2% of AlkPhos DIRECT blocking reagent (Amersham Bioscience). The arrays were then washed twice with a solution containing 2 mM of $MgCl_2$, 50 mM of Tris and 100 mM of NaCl, and with a solution containing 2 mM of $MgCl_2$, 50 mM of Tris and 15 mM of NaCl. All procedures were carried out with the aid of custom-made AutoHybridizers (Fuji Photo Film, Tokyo, Japan). The arrays were then exposed for 2 h to an Imaging Plate and scanned at 25-μm resolution with a BAS5000 phosphoimager (Fuji Photo Film), images of the hybridized arrays were processed with L Process (Fuji Photo Film) and signal intensities quantified with ArrayGauge software (Fuji Photo Film). After each hybridization, the arrays were stripped by boiling them in 0.5% SDS solution for 60 min and scanned for residual hybridization before the next round of hybridization with a newly labeled probe of the same sample to acquire their expression profiles in duplicate.

Data rescaling and preprocessing

The raw data were rescaled to account for the differences in individual hybridization intensities. We employed a rank-invariant scaling method to select genes (Tseng *et al.*, 2001), which were then used for fitting of a nonlinear normalization curve. After normalization, scatter plots of the 50 pairs of replicate data points for each of the genes were generated, the reproducibility of expression between the replicate pairs was assessed, and genes showing a Pearson correlation coefficient higher than 0.85 were selected. Averaged values of the first and the second hybridizations were used for further analyses. In addition, genes whose expression levels did not vary by a factor of at least 2 across the 50 samples were eliminated, because they were unlikely to be informative. Genes with a median intensity lower than 0.3 were also filtered out from the following analyses.

Hierarchical clustering and construction of outcome classifiers

Average linkage hierarchical clustering was performed by using the Cluster program following log transformation and median centering, and the results were visualized using the TreeView program (Eisen *et al.*, 1998). Predictive genes that most effectively distinguished prognosis-favorable patients

from those with fatal outcome were selected by means of signal-to-noise metrics (Golub *et al.*, 1999). Briefly, when prognosis-favorable and fatal patients are defined as class 0 and 1, signal-to-noise statistic (S_x) is calculated as $S_x=(\mu_{class0}-\mu_{class1}/\sigma_{class0}+\sigma_{class1})$, where, for each gene, μ_{class0} represents the mean value and σ_{class0} represents the standard deviation for that gene in all samples of class 0. We selected top-ranked genes based on the absolute values for S_x of each gene. A weighted-voting classification algorithm was employed to predict outcome by using the genes selected as described above and the resulting outcome classifiers were tested by means of 'leave-one-out' cross-validation (Golub *et al.*, 1999). In this scheme, the algorithm can also be used to find the decision boundaries between the class means: $b_x=(\mu_{class0}+\mu_{class1})/2$ for each gene in addition to computing S_x. To predict the class of a test sample γ, each gene x in the predictive gene set has a vote based on the expression in this sample (g_x^γ) and b_x: $V_x=S_x\ (g_x^\gamma-b_x)$ and the final vote for class 0 or 1 is *sign* $(\sum_x V_x)$. The votes were summed to determine the winning class, as well as a 'prediction strength' (PS), which is a measure of the margin of victory that ranges from 0 to 1. The PS is defined as $PS=(V_{win}-V_{lose})/(V_{win}+V_{lose})$, where V_{win} and V_{lose} are the total votes for the winning and losing classes. The measure PS reflects the relative margin of victory of the vote.

Permutation tests

For the permutation tests of statistical significance of class-specific markers, 10000 random permutations of the sample labels (dead or alive) were performed with the dataset, and the signal-to-noise ratio was recalculated for each gene and for each class label permutation. Each gene was thus assigned a *P*-value based upon the frequency by which the permutated signal-to-noise statistics were better than the actual signal-to-noise ratio.

A total of 10000 random sets of genes of the same numbers used for the actual outcome classifiers were generated from the pool of 900 informative genes, and each of these permuted gene sets was then used for prediction with an independent weighted-voting algorithm by using the leave-one-out cross validation strategy. The *P*-values for each outcome classifier were assigned based on the frequency (N) with which random models performed better than models generated by using the significant genes selected by signal-to-noise, or $P=N/10000$. The probability that outcome classifiers constructed by chance might outperform the actual classifiers was also examined. In this test, the outcome labels were randomly permutated 10000 times and for each instance, a set of models was built by using the same set of parameters, for example, the number of predictive genes, as the ones used for finding the actual outcome classifiers. The probabilities of chance construction were calculated based on the frequencies with which random models performed better than the actual models in leave-one-out cross-validation as well as in further validation with independent datasets.

Acknowledgements

We thank Curtis C Harris and Kiyoshi Yanagisawa for their valuable discussions and suggestions. This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan and a Grant-in-Aid for the Second-Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare, Japan.

References

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM. (2000). *Nature*, **403**, 503–511.

Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB and Hanash S. (2002). *Nat. Med.*, **8**, 816–824.

Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ and Meyerson M. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 13790–13795.

Chu PG and Weiss LM. (2002). *Histopathology*, **40**, 403–439.

Clark EA, Golub TR, Lander ES and Hynes RO. (2000). *Nature*, **406**, 532–535.

Craig SW and Chen H. (2003). *Curr. Biol.*, **13**, R236–R238.

Eisen MB, Spellman PT, Brown PO and Botstein D. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 14863–14868.

Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D and Petersen I. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 13784–13789.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science*, **286**, 531–537.

Horio Y, Takahashi T, Kuroishi T, Hibi K, Suyama M, Niimi T, Shimokata K, Yamakawa K, Nakamura Y, Ueda R and Takahashi T. (1993). *Cancer Res.*, **53**, 1–4.

Hu R, Wu R, Deng J and Lau D. (1998). *Lung Cancer*, **20**, 25–30.

Johnson BE, Ihde DC, Makuch RW, Gazdar AF, Carney DN, Oie H, Russell E, Nau MM and Minna JD. (1987). *J. Clin. Invest.*, **79**, 1629–1634.

Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS. (2001). *Nat. Med.*, **7**, 673–679.

Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, Imai K and Nakamura Y. (2003). *Oncogene*, **22**, 2192–2205.

Magnaldo T, Fowlis D and Darmon M. (1998). *Differentiation*, **63**, 159–168.

Martin-Benito J, Boskovic J, Gomez-Puertas P, Carrascosa JL, Simons CT, Lewis SA, Bartolini F, Cowan NJ and Valpuesta JM. (2002). *EMBO J.*, **21**, 6377–6386.

Masaki T, Shiratori Y, Rengifo W, Igarashi K, Yamagata M, Kurokohchi K, Uchida N, Miyauchi Y, Yoshiji H, Watanabe S, Omata M and Kuriyama S. (2003). *Hepatology*, **37**, 534–543.

Mechta F, Lallemand D, Pfarr CM and Yaniv M. (1997). *Oncogene*, **14**, 837–847.

Minna JD, Sekido Y, Fong KM and Gazdar AF. (1997). *Molecular Biology of Lung Cancer: Cancer Principles & Practice of Oncology, 5th edn*, DeVita Jr VT, Hellman S and Rosenberg SA (ed). Lippincott-Raven: Philadelphia, pp. 849–857.

Mitsudomi T, Hamajima N, Ogawa M and Takahashi T. (2000). *Clin. Cancer Res.*, **6**, 4055–4063.
Osada H and Takahashi T. (2002). *Oncogene*, **21**, 7421–7434.
Patterson T, Vuong H, Liaw YS, Wu R, Kalvakolanu DV and Reddy SP. (2001). *Oncogene*, **20**, 634–644.
Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. (2000). *Nature*, **406**, 747–752.
Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES and Golub TR. (2002). *Nature*, **415**, 436–442.
Ramaswamy S, Ross KN, Lander ES and Golub TR. (2003). *Nat. Genet*, **33**, 49–54.
Robert C, Bolon I, Gazzeri S, Veyrenc S, Brambilla C and Brambilla E. (1999). *Clin. Cancer Res.*, **5**, 2094–2102.
Shelton JG, Steelman LS, Lee JT, Knapp SL, Blalock WL, Moye PW, Franklin RA, Pohnert SC, Mirza AM, McMahon M and McCubrey JA. (2003). *Oncogene*, **22**, 2478–2492.
Slebos RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, Wagenaar SS, Vanderschueren RG, van Zandwijk N, Mooi WJ, Bos JL and Rodenhuis S. (1990). *N. Engl. J. Med.*, **323**, 561–565.
Tseng GC, Oh MK, Rohlin L, Liao JC and Wong WH. (2001). *Nucleic Acids Res.*, **29**, 2549–2557.
Vennstrom B, Sheiness D, Zabielski J and Bishop JM. (1982). *J. Virol.*, **42**, 773–779.
Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, Miyoshi T, Nomura H and Jones MH. (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 12357–12362.
Vuong H, Patterson T, Adiseshaiah P, Shapiro P, Kalvakolanu DV and Reddy SP. (2002). *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **282**, L215–L225.
Walker-Daniels J, Hess AR, Hendrix MJ and Kinch MS. (2003). *Am. J. Pathol.*, **162**, 1037–1042.
Willett CG, Wang MH, Emanuel RL, Graham SA, Smith DI, Shridhar V, Sugarbaker DJ and Sunday ME. (1998). *Am. J. Respir. Cell Mol. Biol.*, **18**, 489–496.
Wu WJ, Erickson JW, Lin R and Cerione RA. (2000). *Nature*, **405**, 800–804.
Yatabe Y, Mitsudomi T and Takahashi T. (2002). *Am. J. Surg. Pathol.*, **26**, 767–773.
Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, Li Y, Robles AI, Chen Y, Ma ZC, Wu ZQ, Ye SL, Liu YK, Tang ZY and Wang XW. (2003). *Nat. Med.*, **9**, 416–423.

[CANCER RESEARCH 64, 3753–3756, June 1, 2004]

Advances in Brief

Reduced Expression of the *let-7* MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival

Junichi Takamizawa,[1,4] Hiroyuki Konishi,[1] Kiyoshi Yanagisawa,[1] Shuta Tomida,[1] Hirotaka Osada,[1] Hideki Endoh,[3] Tomoko Harano,[1] Yasushi Yatabe,[2] Masato Nagino,[4] Yuji Nimura,[4] Tetsuya Mitsudomi,[3] and Takashi Takahashi[1]

[1]*Division of Molecular Oncology, Aichi Cancer Center Research Institute, Nagoya, Japan; Departments of* [2]*Anatomic and Molecular Diagnostic Pathology and* [3]*Thoracic Surgery, Aichi Cancer Center Hospital, Nagoya, Japan;* [4]*Division of Surgical Oncology, Department of Surgery, Nagoya University Graduate School of Medicine, Nagoya, Japan*

Abstract

In this study, we report for the first time reduced expression of the *let-7* microRNA in human lung cancers. Interestingly, 143 lung cancer cases that had undergone potentially curative resection could be classified into two major groups according to *let-7* expression in unsupervised hierarchical analysis, showing significantly shorter survival after potentially curative resection in cases with reduced *let-7* expression ($P = 0.0003$). Multivariate COX regression analysis showed this prognostic impact to be independent of disease stage (hazard ratio = 2.17; $P = 0.009$). In addition, overexpression of *let-7* in A549 lung adenocarcinoma cell line inhibited lung cancer cell growth *in vitro*. This study represents the first report of reduced expression of *let-7* and the potential clinical and biological effects of such a microRNA alteration.

Introduction

Cells contain a variety of noncoding RNAs, which perform a multitude of functions. Recently, microRNAs (miRNAs), an abundant class of small noncoding RNAs of about 22 nucleotides in length, have been recognized as being numerous and phylogenetically well conserved (1). The miRNA species are encoded by genes that are presumably transcribed into single or clustered miRNA precursors, which are converted to mature forms of miRNAs through stepwise processing including generation of ~70 nucleotide pre-miRNA with a characteristic hairpin structure from the longer nascent transcripts (pri-miRNA) and the following Dicer-mediated processing into mature forms (2–5). Although thus far over 300 miRNA genes have been discovered in various organisms (6–10), including humans, their precise physiological functions are largely unknown except for a handful of miRNAs (11–17), and their potential pathological involvement including oncogenesis is yet to be explored.

The *Caenorhabditis elegans let-7*miRNA is to date the best-studied example along with *lin-4* of the same worm (11–15), both of which were initially identified by genetic analysis of the developmental timing defects of mutants. The *let-7* miRNA, which starts to be expressed during the late developmental stage, acts as a post-transcriptional repressor of *lin-41, hbl-1/lin-57* and perhaps other genes that contain sequences imprecisely complementary to the miRNA in their 3′ untranslated regions. The expression levels of the human *let-7* gene have been shown to vary among various adult tissues, lung being one of the tissues with most abundant expression of *let-7* (18).

In this study, we show for the first time that expression levels of *let-7* are frequently reduced in lung cancers both *in vitro* and *in vivo*. Furthermore, lung cancer patients with reduced *let-7* expression were found to have significantly worse prognosis after potentially curative resection, and this prognostic impact of reduced *let-7* expression appears to be independent of disease stage in multivariate COX regression analysis. In addition, we show that overexpression of *let-7* inhibits growth of lung cancer cells *in vitro*.

Materials and Methods

Study Population. This study dealt with 159 nonsmall cell lung carcinoma (NSCLC) tissue specimens collected with the approval of the institutional review board of the Aichi Cancer Center. The specimens from 143 cases (105 adenocarcinomas, 25 squamous cell carcinomas, 9 large cell carcinomas, and 4 adenosquamous cell carcinomas), which had been followed up for >5 years after potentially curative resection, were used specifically for studying the prognostic significance of *let-7*. These 143 cases consisted of 90 female and 53 male patients with a median age of 62 (range, 32–84), and with 75 in stage I, 19 in stage II, and 49 in stage III.

Preparation of Cell Line and Tissue Samples. All of the human NSCLC cell lines analyzed were cultured with 5% (v/v) FCS-containing RPMI 1640 at 37°C with 5% CO_2. BEAS-2B and HPL1D (19) cells were cultured with 1% (v/v) FCS-containing Ham's F-12 supplemented with bovine insulin (5 μg/ml), human transferrin (5 μg/ml), 10^{-7} M hydrocortisone, 2×10^{-10} M triiode thyronine, penicillin (100 IU/ml), and streptomycin (100 μg/ml) at 37°C with 5% CO_2. The tumor specimens were homogenized in guanidine isothiocyanate homogenization buffer immediately after resection and stored at −30°C until use with the approval of the institutional review board. Processing of all cell lines and tissue samples for RNA extraction were performed according to the standard procedures.

Northern Blotting. Ten μg of RNA were separated on a 15% denaturing polyacrylamide gel. The RNA was then transferred to Zeta-Probe GT Blotting Membranes electrophoretically overnight. Probes (let-7; 5′-TACTATACAACCTACTACCTCAATTTGCC and 5S; 5′-TTAGCTTCCGAGATCAGACGA) were generated by T4 polynucleotide kinase (New England Biolabs, Beverly, MA) mediated end-labeling of DNA oligonucleotides with [γ-^{32}P]ATP. Prehybridization and hybridization were carried out using hybridization buffer (0.25 M sodium phosphate (pH 7.2), 7% SDS, 0.5% sodium PP_i). The most stringent wash was carried out in 2× SSC and 1% SDS at 37.5°C.

Real-Time Reverse Transcription-PCR. Real-time reverse transcription-PCR was performed using an ABI Prism 7900 Sequence Detection System (Perkin-Elmer Applied Biosystems, Foster City, CA), the SYBR Green PCR Master Mix (Perkin-Elmer Applied Biosystems), and random-primed cDNAs (corresponding to 20 ng of total RNA extracted from tissue samples).

The primer pairs used were *let-7a-1S* (sense; 5′-CCTGGATGTTCTCTTCACTG) and *let-7a-1*AS (antisense; 5′-GCCTGGATGCAGACTTTTCT); *let-7a-2S* (sense; 5′-TTCCAGCCATTGTGACTGCA) and *let-7a-2*AS (antisense; 5′-CTCACCATGTTGTTTAGTGC); *let-7a-3S* (sense; 5′-ACCAAGACCGACTGCCCTTT) and *let-7a-3*AS (antisense; 5′-CTCTGTCCACCGCAGATATT); *let-7f-1S* (sense; 5′-TGTACTTTCCATTCCAGAAG) and *let-7f-1*AS (antisense; 5′-TAATGCAGCAAGTCTACTCC); *let-7f-2S* (sense; 5′-TGAAGATGGACACTGGTGCT) and *let-7f-2*AS (antisense; 5′-

Received 2/21/04; revised 4/9/04; accepted 4/19/04.

Grant support: This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan, a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science, and a Grant-in-Aid for the Second Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health and Welfare, Japan.

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: J. Takamizawa and H. Konishi contributed equally to the present study. H. Konishi is currently at the Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD.

Requests for reprints: Takashi Takahashi, Division of Molecular Oncology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya, 464-8681, Japan. Phone: 81-52-764-2983; Fax: 81-52-764-2983; E-mail: tak@aichi-cc.jp.

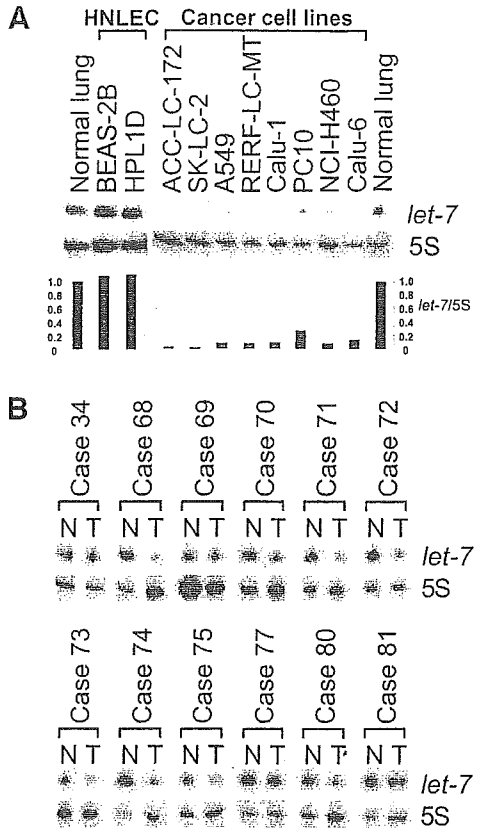

Fig. 1. Northern blot analysis of *let-7* expression in primary lung cancers. *A*, representative Northern blot analysis in lung cancer cell lines *in vitro*. BEAS-2B and HPL1D, immortalized human normal bronchial and peripheral lung epithelial cell lines, respectively. HNLEC, human normal lung epithelial cell lines. *B*, representative Northern blot analysis of primary lung cancer specimens *in vivo*. 5S rRNA served as a loading control. N, normal lung; T, lung cancer.

CAGTCGGAGAAGAAGTGTAC); and 5S–S (sense; 5′-TACGGCCATAC-CACCCTGAA) and 5S-AS (antisense; 5′-TAACCAGGCCCGACCCTGCT). To quantify the expression level of the *let-7* genes, standard curves were made using serially diluted pBluescriptIISK (−) inserted with each PCR product into the *Eco*RV site. PCR amplification consisted of 55 cycles (95°C for 30 s, 56°C to 60°C optimized for each primer set for 30 s and 72°C for 15 s) after the initial denaturation step (95°C for 10 min). Expression levels of the *let-7* genes were based on the amount of the target message relative to the 5S rRNA control, to normalize the initial input of total RNA.

Hierarchical Clustering. We used the Eisen CLUSTER and TREEVIEW programs for hierarchical clustering and visualization of data sets. Before applying the clustering algorithm, we log-transformed the fluorescence ratio for each expression and then average centered the data for all samples. Agglomerative hierarchical clustering was applied using the complete linkage method to investigate whether there was evidence for natural groupings of tumor samples based on correlations between gene-expression profiles.

Statistical Analysis. The Kaplan-Meier method was used to estimate survival as a function of time, and survival differences were analyzed by the log-rank test. Cox regression analysis of factors potentially related to survival was performed to identify which independent factors might jointly have a significant influence on survival.

Colony Formation Assay. The *let-7* expression construct and a control plasmid were constructed by the cloning of annealed oligonucleotides of *let-7a* (sense, 5′-GATCCCCTGAGGTAGTAGGTTGTATAGTTTTT and antisense, 5′-AGCTAAAAACTATACAACCTACTACCTCAGGG), *let-7f* (sense, 5′-GATCCCCTGAGGTAGTAGATTGTATAGTTTTT and antisense, 5′-AGCTA-AAAACTATACAATCTACTACCTCAGGG), or control (sense, 5′-GATC-CCCTTTTTTTTGGAAA and antisense, 5′-AGCTTTTCCAAAAAAAAGGG) into pH1-RNApuro, in which expression of a gene is under the control of the RNA polymerase III H1-RNA gene promoter prepared by PCR amplification of human genomic DNA. The *let-7a* and *-7f* expression constructs were transfected into A549 lung adenocarcinoma cell line using the FuGENE 6 reagent (Roche, Inc. Basel, Switzerland) according to the manufacturer's instructions. Cells were selected by the addition of puromycin (2 μg/ml) 3 days after the transfection and cultured at 37°C for 2 weeks. After 2 weeks of puromycin selection, the plates were stained with Giemsa and scored for the number of resistant colonies.

Results

Reduced Expression of *let-7* in Human Lung Cancers in Both *in Vitro* and *in Vivo*. Northern blot analysis was first performed to analyze *let-7* expression in 20 human lung cancer cell lines as well as in two immortalized human normal lung epithelial cell lines (Fig. 1*A*). The mature form of *let-7* miRNA was readily detectable in both immortalized lung epithelial cell lines at a level comparable with that in normal lung tissues. In marked contrast, a significant reduction (>80%) in the expression levels of *let-7* was observed in 60% (12 of 20) of lung cancer cell lines. Expression levels of *let-7* in primary human lung cancer tissues taken directly from surgically treated patients, in which sufficient RNA were available, were further analyzed by Northern blot analysis. Consequently, 44% (7

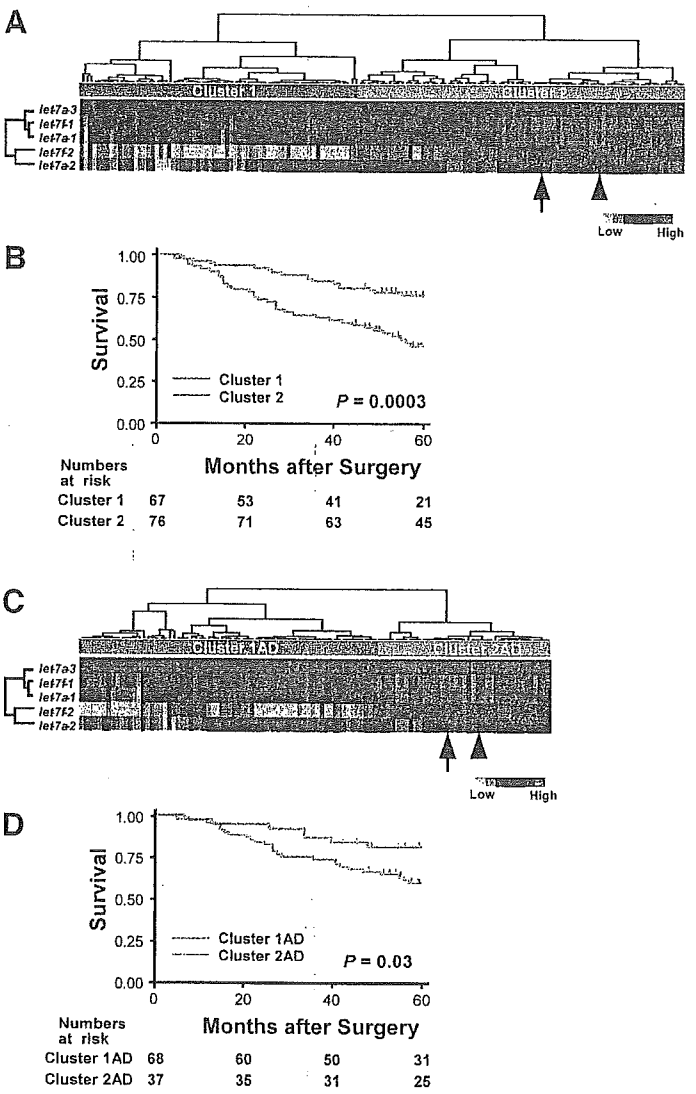

Fig. 2. Hierarchical clustering and Kaplan-Meier survival curves based on expression of *let-7* microRNA (miRNA) isoforms. *A*, results of unsupervised hierarchical clustering of the entire cohort of 143 nonsmall cell lung carcinoma (NSCLC) cases. *B*, Kaplan-Meier survival curves for NSCLC patients who were classified into clusters 1 and 2. The difference in postoperative survival between clusters 1 and 2 was highly significant ($P = 0.0003$ by log-rank test). *C*, results of unsupervised hierarchical clustering of the 105 adenocarcinoma cases. *D*, Kaplan-Meier survival curves for adenocarcinoma cases belonging to either cluster 1AD or 2AD. The difference in postoperative survival between clusters 1AD and 2AD was also statistically significant ($P = 0.03$ by log-rank test). *Arrows* and *arrow heads*, mixture of RNAs of 38 and 120 normal human peripheral lung tissues, respectively.

of 16) of the cases examined were found to exhibit >80% reduction in *let-7* expression when compared with that in the corresponding normal lung tissues (Fig. 1*B*). A more frequent occurrence of reduced *let-7* expression in cell lines *in vitro* may be related to the inevitable contamination of normal stromal/inflammatory cells in tumor tissues *in vivo* or, alternatively, this may reflect *in vitro* selection of cells with reduced *let-7* in the process of the establishment of cell lines. These findings thus clearly showed the frequent occurrence of a significant reduction in *let-7* miRNA expression in lung cancers.

Prognostic Impact of Reduced *let-7* Expression in Surgically Treated Lung Cancer Patients. We next wished to investigate whether reduced *let-7* expression has any relation to clinicopathological characteristics of lung cancers in an isoform-specific manner. To this end, 143 lung cancer cases, which had undergone potential curative resection of NSCLCs, were examined by real-time reverse transcription-PCR analysis using *let-7* isoforms-specific oligonucleotide primers. Expression levels of *let-7* pri-miRNAs were consequently shown to vary significantly among lung cancer cases, although they tended to be coordinately regulated. The most abundant species were *let-7a-1* and *let-7f-1*, which are known to be clustered within a few hundred bases in the human genome (6) and could be amplified together by reverse transcription-PCR (data not shown). We used unsupervised hierarchical clustering to classify the 143 resected human NSCLC cases in an unbiased manner without using any information on the identity of the samples. This procedure resulted in the classification of NSCLC cases into two major classes based on similarities in *let-7* expression (Fig. 2*A*). Except for a significant association between cluster 1 with low *let-7* expression and higher disease stages ($P = 0.004$ by the λ^2 test), no other significant associations were found between the clusters and various clinicopathological features including age, sex, histology, primary tumor status (pT), and differentiation grade. Of special interest was a striking difference in the postoperative survival of patients between the two clusters. The Kaplan-Meier survival curves demonstrated that patients belonging to cluster 1 were at a significantly greater risk of an earlier death than those classified as cluster 2 ($P = 0.0003$ by the log-rank test; Fig. 2*B*). A separate study analyzed the prognostic significance of *let-7* in adenocarcinomas, which constitute the major proportion of lung cancers in Japan as well as in other countries such as the United States. We found that adenocarcinoma cases can also be divided into two major clusters, again showing that patients in cluster 1AD with low *let-7* expression had significantly shorter survival than those in cluster 2AD with high *let-7* expression ($P = 0.03$ by the log-rank test; Fig. 2, C and D).

Univariate Cox regression analysis was then performed for the entire cohort and showed that, in addition to disease stage ($P < 0.001$; Table 1), classification into cluster 1 with characteristically low *let-7* expression is a significant predictive factor for poor prognosis ($P < 0.001$). Cox proportional hazards modeling was then conducted to identify which independent factors would jointly have a significant influence on survival (Table 1). The inter-relationship of possible prognostic factors and survival was analyzed, using age, sex, histological type, smoking history, disease stage, and the *let-7*-defined cluster as variables, resulting in the identification of *let-7*-defined cluster as a significant, independent prognostic factor in surgically treated NSCLC patients after potentially curative resection ($P = 0.009$) in addition to disease stage ($P < 0.001$). The hazard ratio of earlier death was 2.17 (95% confidence interval, 1.21–3.89) for clusters 1 *versus* 2 and 3.49 (95% confidence interval, 1.89–6.42) for pathological stage II/III *versus* pathological stage I. Taken together, expression levels of *let-7* seemed to have a significant impact on the postoperative survival of NSCLC patients.

Table 1 Cox regression analysis of various prognostic factors for postoperative survival of lung cancer patients

Variables	Hazard ratio (95% CI[a])	Unfavorable/favorable	*P*
Univariate analysis			
Age (yr)	1.70 (0.97–2.99)	≥62/<62	0.063
Sex	1.34 (0.75–2.38)	Male/female	0.323
Histology	1.30 (0.67–2.52)	Squamous/non-squamous	0.443
Smoking history	1.42 (0.80–2.51)	Smoker/non-smoker	0.233
Disease stage	3.89 (2.14–7.08)	II-III/I	<0.001
let-7	2.78 (1.56–4.89)	Cluster 1/cluster 2	<0.001
Multivariate analysis			
Age (yr)	1.68 (0.95–2.97)	≥62/<62	0.076
Sex	1.18 (0.44–3.13)	Male/female	0.741
Histology	1.03 (0.49–2.16)	Non-squamous/squamous	0.942
Smoking history	1.07 (0.41–2.82)	Non-smoker/smoker	0.889
Disease stage	3.49 (1.89–6.42)	II- III/I	<0.001
let-7	2.17 (1.21–3.89)	Cluster 1/cluster 2	0.009

[a] 95% CI, 95% confidence interval.

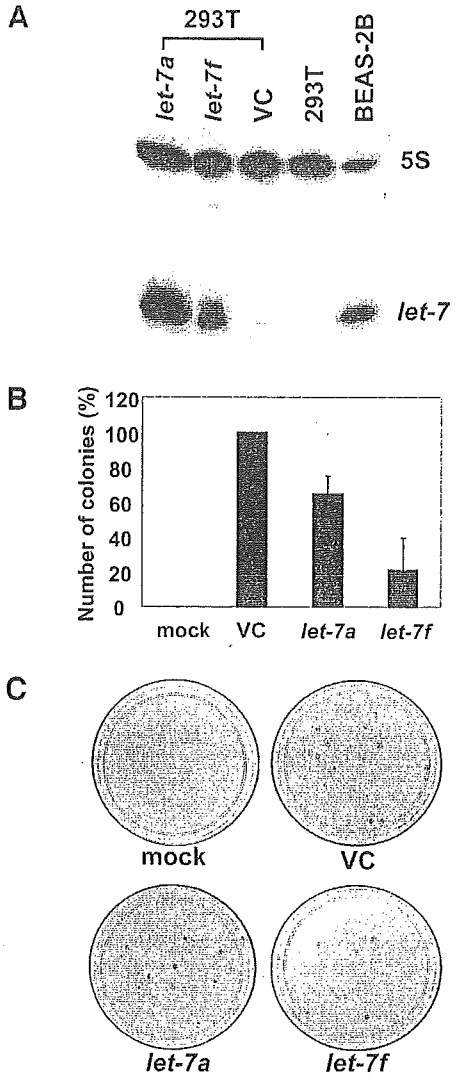

Fig. 3. Introduction of *let-7* into A549 lung adenocarcinoma cell line. *A*, results of Northern blot analysis confirming expression of *let-7a* and *let-7f* isoforms. *B*, graphic presentation of a representative colony formation assay by the introduction of exogenous *let-7*. Similar results were obtained in five independent assays done in triplicate. *C*, representative dishes showing reduced colony formation by overexpression of exogenously introduced *let-7*.

Growth Suppression of Lung Cancer Cells by Overexpression of Exogenous *let-7*. The identification of a reduced expression of *let-7* in lung cancers, in association with a shortened survival, prompted us to explore the possible biological significance of *let-7* in lung cancer development. As an initial step, we introduced *let-7* into a lung cancer cell line by using expression constructs, which were designed to synthesize mature miRNAs of two predominant *let-7* isoforms, *let-7a* and *let-7f*, under the control of the RNA polymerase-III H1-RNA gene promoter. We confirmed that these expression constructs could work as expected using 293T cells (Fig. 3*A*). Overexpression of *let-7f* in A549 lung adenocarcinoma cell line resulted in a 78.6% reduction in the number of colonies, whereas the introduction of *let-7a* also showed similar but a more modest growth-inhibitory effect (Figs. 3, *B* and *C*). Similar results were obtained in five independent experiments, which were done in triplicate using three independent preparations of plasmid DNAs.

Discussion

It has become apparent that genomic information for transcribing miRNAs is indeed implemented in the human genome (6, 9), but extremely little information is available regarding their physiological

and pathological roles. This is the first demonstration that expression levels of *let-7* miRNA, which to date is one of the best-studied miRNAs, are altered in human lung cancers. Furthermore, we have shown that reduced *let-7* expression is significantly associated with shortened postoperative survival and that overexpression of *let-7* results in the inhibition of lung cancer cell growth. Altogether, these findings suggest that reduced expression of *let-7* may play a role in the pathogenesis of lung cancers.

Very little information is available at the moment with regard to the potential pathological roles of miRNAs. Two proteins (Gemin 3 and Gemin 4), which are components of the protein complex related to spinal muscular atrophy, are also known to be components of a ribonucleoprotein complex containing miRNAs (microRNP; Ref. 9), whereas the *Drosophila* homologue of fragile X mental retardation protein has been shown to be a component of RNA-induced silencing complex/microRNPs (20, 21). This circumstantial evidence suggests the possibility of the involvement of miRNA machineries in these diseases. As for links between cancer and miRNA, Calin *et al.* (22) reported frequent down-regulation of *miR15* and *miR16* in chronic lymphocytic leukemia, whereas Michael *et al.* (23) recently reported reduced expression of *miR-143* and *miR-145* in human colon cancers. In contrast to these studies, which did not address the question of whether reduced expression of miRNAs has any influence on clinicopathological features, this study clearly shows that reduced *let-7* expression is indeed significantly associated with the shortened survival of patients. Because no changes in *let-7* expression were reported in colon cancers (23), it is possible that miRNAs may be distinctly involved in the pathogenesis of these two most common cancers of adults and possibly in other types of human cancers.

It has been shown that the *let-7* gene regulates developmental timing in *C. elegans* and that mutant worms lacking *let-7* fail to properly execute a larval-to-adult switch in hypodermal cell development (13). Although *lin-41* is known to be post-transcriptionally repressed by *let-7* (24), it is not inconceivable that other genes may also be targeted by *let-7*, because of the requirement of imprecise base-pairing for miRNA-mediated translational repression (1). Indeed, *hbl-1/lin-57* was recently reported to be targeted by *let-7* (14, 15), whereas a few additional genes have also been predicted to be a potential target for *let-7* (24, 25). Interestingly, such potential targets include *LIM kinase 2* (25), which belongs to a gene family having a role in the regulation of cell shape and motility as well as possibly in metastasis. One could speculate that the change in miRNA expression as is seen in this study might be an efficient strategy for cancer cells to simultaneously alter the expression profile of a series of genes. Alterations in miRNA expression may accordingly confer cancer cells with selective growth advantage, allowing them to form a distant metastasis and resulting in the consequential death of the patient. This scheme may be consistent with the present finding of the significant prognostic impact of *let-7* expression. One might argue that reduced expression of *let-7* in lung cancers may merely reflect its oncofetal regulation, because fetal lung exhibited considerably lower *let-7* expression than adult lung (data not shown). However, growth-inhibitory effects of overexpressed *let-7* in A549 adenocarcinoma cell line argue against this possibility. Taken together, these findings suggest the potential involvement of reduction in *let-7* expression in the pathogenesis of this fatal disease, although the results obtained with overexpression of mature miRNA need to be interpreted cautiously and await further experimental clarification.

In this study, we observed that various *let-7* pri-miRNA isoforms were coordinately regulated, *let-7a-1* and *let-7f-1* being the most predominant. In this connection, it should be noted that some of the *let-7* pri-miRNAs give rise to identical mature miRNA isoforms, and the others may also have presumably very similar, if not identical, spectra of the target genes (6). It is uncertain at the moment how the expression levels of various *let-7* isoforms are coordinated, and this remains an intriguing question awaiting further investigation.

In conclusion, we have shown for the first time that *let-7* expression is frequently reduced in lung cancers and that alterations in the miRNA expression may have a prognostic impact on the survival of surgically treated lung cancer patients. These findings warrant additional studies to investigate whether *let-7* alterations are also involved in other types of human cancers and how altered miRNA expression would manifest the biological and biochemical consequences in the development of human cancers. Accordingly, future identification of the downstream targets for *let-7* may provide clues to develop a novel therapeutic means. It is envisaged that such future studies may ultimately provide a foundation for a new paradigm of the involvement of noncoding small RNA species, miRNA, in human oncogenesis.

References

1. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 2003;113:673–6.
2. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science (Wash D C) 2001;293:834–8.
3. Grishok A, Pasquinelli AE, Conte D, et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 2001;106:23–34.
4. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev 2001;15:2654–9.
5. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 2002;21:4663–70.
6. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science (Wash D C) 2001;294:853–8.
7. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science (Wash D C) 2001;294: 858–62.
8. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science (Wash D C) 2001;294:862–4.
9. Mourelatos Z, Dostie J, Paushkin S, et al. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 2002;16:720–8.
10. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. Vertebrate microRNA genes. Science (Wash D C) 2003;299:1540.
11. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843–54.
12. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 1993;75:855–62.
13. Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature (Lond) 2000;403:901–6.
14. Abrahante JE, Daul AL, Li M, et al. The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. Dev Cell 2003;4:625–37.
15. Lin SY, Johnson SM, Abraham M, et al. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev Cell 2003;4:639–50.
16. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell 2003;113:25–36.
17. Xu P, Vernooy SY, Guo M, Hay BA. The Drosophila MicroRNA Mir-14 Suppresses Cell Death and Is Required for Normal Fat Metabolism. Curr Biol 2003;13:790–5.
18. Pasquinelli AE, Reinhart BJ, Slack F, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature (Lond) 2000;408:86–9.
19. Masuda A, Kondo M, Saito T, et al. Establishment of human peripheral lung epithelial cell lines (HPL1) retaining differentiated characteristics and responsiveness to epidermal growth factor, hepatocyte growth factor, and transforming growth factor β1. Cancer Res 1997;57:4898–904.
20. Caudy AA, Myers M, Hannon GJ, Hammond SM. Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev 2002;16:2491–6.
21. Ishizuka A, Siomi MC, Siomi H. A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev 2002;16:2497–508.
22. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci USA 2002;99:15524–9.
23. Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res 2003; 1:882–91.
24. Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell 2000;5:659–69.
25. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003;115:787–98.

K. Tajima
A. Demachi
Y. Ito
K. Nishida
Y. Akatsuka
K. Tsujimura
H. Kuwano
T. Mitsudomi
T. Takahashi
K. Kuzushima

Identification of an epitope from the epithelial cell adhesion molecule eliciting HLA-A*2402-restricted cytotoxic T-lymphocyte responses

Key words:
autoimmunity; CTL; dendritic cells; epithelial cell adhesion molecule; HLA-A24; immunotherapy

Acknowledgments:
Valuable discussions and suggestions by Dr Tatsuya Tsurumi, Dr Masao Seto and Dr Takashi Takahashi, Aichi Cancer Center Research Institute, Nagoya, Japan, are highly appreciated. We are very grateful to Ms Yasue Matsudaira and Ms Hiroko Tamaki for their secretarial assistance. This work was supported, in part, by the Japan Society for the Promotion of Science (15590429).

Abstract: Because the epithelial cell adhesion molecule (Ep-CAM) is expressed in almost all carcinomas and human leucocyte antigen (HLA)-A*2402 is the most common allele in many ethnic groups, including Japanese, the identification of peptide sequences, which elicit HLA-A*2402-restricted Ep-CAM-specific cytotoxic T-lymphocyte (CTL) responses, would facilitate specific immunotherapy for various histological types of carcinomas. An epitope was identified through the following steps: (i) computer-based epitope prediction from the amino acid sequence of Ep-CAM, (ii) major histocompatibility complex (MHC) stabilization assay to determine the affinity of the predicted peptide with HLA-A*2402 molecules, (iii) stimulation of $CD8^+$ T cells with peptide-pulsed dendritic cells and (iv) testing the CTL specificity by means of enzyme-linked immunospot (ELISPOT) assays, CTL assays and MHC/peptide-tetramer staining. Peripheral $CD8^+$ T cells of four of five healthy donors after three rounds of stimulation with the peptide $Ep\text{-}CAM_{173-181}$ (RYQLDPKFI) secreted interferon-γ in ELISPOT assays when exposed to the peptide. A CTL clone specific to the peptide efficiently lysed Ep-CAM-expressing cancer cell lines in an HLA-A*2402-restricted fashion. Endogenous processing and presentation of the peptide in a lung cancer cell line were confirmed by means of cold target inhibition assays. The CTL clone was also lytic to normal bronchial epithelial cells but to a lesser extent at low effector: target ratios. All these data suggest that the peptide-specific CTL responses may play some roles both in anti-cancer and autoimmune reactions. The peptide should prove useful to study anti-Ep-CAM CTL responses among population possessing HLA-A*2402.

Authors' affiliation:

K. Tajima[1*],
A. Demachi[1],
Y. Ito[1],
K. Nishida[1],
Y. Akatsuka[1],
K. Tsujimura[1],
H. Kuwano[2],
T. Mitsudomi[3],
T. Takahashi[1],
K. Kuzushima[1]

[1]Division of Immunology, Aichi Cancer Center Research Institute, Nagoya, Japan

[2]Department of Surgery I, Gunma University Faculty of Medicine, Maebashi, Japan

[3]Department of Thoracic Surgery, Aichi Cancer Center Hospital, Nagoya, Japan

Correspondence to:
K. Kuzushima
Division of Immunology
Aichi Cancer Center Research Institute
1-1 Kanokoden
Chikusa-ku
Nagoya 464-8681
Japan
Tel.: +81 52 762 6111
Fax: +81 52 764 2990
e-mail: kkuzushi@aichi-cc.jp

Cytotoxic T-lymphocytes (CTLs) have become widely accepted as important players in resistance to cancer. Although various CTL epitopes of tumour-associated antigens have been identified so far (1, 2), the search for additional epitopes continues, because the expression of tumour antigens is heterogeneous among tumours of various histological origins, various patients and between individual lesions. From the clinical point of view, molecular characterization of

Received 26 July 2004, revised 23 August 2004, accepted for publication 24 August 2004

doi: 10.1111/j.1399-0039.2004.00329.x

Tissue Antigens 2004: **64:** 650–659

*Present address: Department of Surgery I, Gunma University Faculty of Medicine, Maebashi 371-8511, Japan.

additional tumour antigens is crucial for successful immunotherapy, because immunoselection of antigen-negative tumour cell variants has been observed during peptide vaccination (3–5).

Epithelial cell adhesion molecule (Ep-CAM), also referred to as EGP-2, 17-1A, GA733-2, KSA or PE-35 (6–10), was originally reported as a serologically defined surface antigen, highly expressed on many carcinomas of diverse histological origins, such as colon (11), lung (12), head and neck (13) and breast tumours (14), but with limited expression by normal epithelial cells (15, 16). Its function is to mediate Ca^{2+}-independent homotypic cell–cell adhesion. Because of its intensive and uniform expression in a variety of human tumours, Ep-CAM has become one of the most attractive targets for immunotherapy with monoclonal antibodies, or even for gene therapy (17). Treatment of a series of patients suffering from Dukes' C colorectal carcinoma with a monoclonal antibody against Ep-CAM, namely 17-1A, has been found to reduce mortality and recurrence (18, 19). Recently, it was reported that HLA-A*0201-restricted Ep-CAM-derived peptide-specific CTLs can lyse epithelial tumour cells but not normal cells (20, 21). Immunotherapy using such epitope peptides has potential efficacy.

Using a bioinformatic approach, in the present study, we first predicted seven peptide sequences in Ep-CAM, which might bind to HLA-A*2402 molecules, the most common allele in Japanese (more than 60%) and also present in persons of European descent (nearly 20%). Specific CTL was successfully induced in four of five healthy donors by using Ep-$CAM_{173-181}$ (RYQLDPKFI) and a $CD8^+$ CTL clone specific to this peptide showed cytotoxicity against HLA-$A24^+$ Ep-CAM^+ but not HLA-$A24^-$ cancer cells. Cold target inhibition assays suggested that the peptide was naturally processed and was presented on the surfaces of HLA-$A24^+$ Ep-CAM^+ cancer cells. The fine specificity of the peptide-specific CTL was extensively studied and the results were discussed in the light of anti-cancer and anti-self cellular immunity.

Materials and methods

Donors and cell lines

The study design and purpose, which had been approved by the Institutional Review Board of Aichi Cancer Center, Nagoya, Japan, were explained fully to all donors. Peripheral blood was obtained from five HLA-A24-positive healthy donors and peripheral blood mononuclear cells (PBMCs) were isolated by means of centrifugation on a Ficoll-Paque (Pharmacia, Piscataway, NJ) density gradient.

Human cancer cell lines – LU99, HSC-2, MKN28, MKN45 and COLO320DM cells – were purchased from the Japanese Collection of Research Bioresources (Tokyo, Japan) and LC-1/sq from RIKEN Cell Bank (Tsukuba, Japan). LC-1/sq cells were maintained in 45% RPMI 1640 medium (Sigma, St Louis, MO) and 45% Ham's F12 medium (Sigma) supplemented with 10% fetal calf serum (FCS) (Life Technologies Limited, Auckland, New Zealand), L-glutamine, penicillin and streptomycin. COLO320DM and MKN28 were maintained in Dulbecco's modified Eagle medium (Sigma) with the same supplements. The other cancer cell lines were cultured in RPMI1640 medium with the same supplements (referred to as complete medium). HLA-A24-positive, normal human bronchial epithelial cells, designated as NHBE, were cultured according to the manufacturer's recommendations (CC2540, Clonetics Corp, BioWhittaker, Walkersville, MD). The HLA-A*2402 transfectants – T2-A24, QG56-A24 and A549-A24 – were established and were cultured as previously described (22, 23).

Reverse transcription polymerase chain reaction

Using a GenElute mRNA Miniprep kit (Sigma Chemical Co., St Louis, MO), total RNA was extracted from cultured cell lines. Gene-specific oligonucleotide primers were synthesized at Proligo (Kyoto, Japan) and were used in order to evaluate the mRNA expression of Ep-CAM. Forward and reverse primers used were as follows: ATG GCG CCC CCG CAG GTC CT and TTA TGC ATT GAG TTC CCT ATG CAT CTC ACC. Reverse transcription polymerase chain reaction (RT-PCR) was performed by using a thermal cycler (Perkin-Elmer, Wellesley, MA) and products were analysed by means of 1.5% agarose gel electrophoresis with ethidium bromide visualization.

Western blot analysis

Western blot analysis was performed as described previously (24) with slight modifications. Briefly, aliquots of 130-μg protein from the post-nuclear supernatant of the cell lysate were applied to 12% SDS-PAGE and were blotted onto Immobilon-P membranes (Millipore Corporation, Bedford, MA). After probing with a monoclonal antibody specific to Ep-CAM (clone 323/A3, Laboratory Vision, Fremont, CA), followed by peroxidase-conjugated goat anti-mouse immunoglobulin G (IgG) (Zymed, San Francisco, CA), proteins were visualized with the help of an ECL Western blot detection system (Amersham Biosciences, Buckinghamshire, UK).

Synthetic peptides

In order to identify potential HLA-A24-binding peptides within Ep-CAM (accession number M33011), we employed a computer-based program accessed through the World Wide Web site BioInformatics & Molecular Analysis Section (BIMAS) HLA peptide-binding predictions (available at http://bimas.dcrt.nih.gov/molbio/hla_bind/). Most peptides were synthesized with a Cleaved PepSet from Mimotope

(Melbourne, Australia), dissolved in 100 μl dimethyl sulfoxide and further diluted in 40% acetonitrile, 0.1 M HEPES (pH 7.4), where necessary. Characteristics of the seven synthetic peptides, designated as Ep_{31}, Ep_{173}, Ep_{185}, Ep_{250}, Ep_{225}, Ep_{296} and Ep_{304}, have been listed in Table 1. A human immunodeficiency virus-1 (HIV-1) envelope peptide RYLRDQQLL (25) (residues 584–592, designated as ENV_{584}) and an EBV (Epstein-Barr virus) latent membrane protein 2 peptide TYGPVFMCL (26) (residues 419–427, EBV-LMP2419) were synthesized by Toray Research Center (Kamakura, Japan).

Cell staining and flow cytometric analysis

Surface expression of HLA-A24 and Ep-CAM molecules was examined with the help of indirect immunofluorescence by using an anti-HLA-A24 monoclonal antibody (One Lambda, Inc., Canoga Park, CA), the anti-Ep-CAM monoclonal antibody and FITC-labelled (fluorescein isothiocyanate) anti-mouse IgG $F(ab')_2$ fragments (IMMUNOTECH, Marseille, France). MHC/peptide tetramers were produced as previously described (22, 27). The Ep-CAM-specific $CD8^+$ T cells were stained with PE-labelled HLA-A*2402 tetramers incorporating the Ep-CAM peptide, Ep_{173} (designated as the HLA-A24/Ep_{173} tetramer) or the HIV-1 peptide, ENV_{584} (HLA-A24/ENV_{584} tetramer). Flow cytometric analysis of the stained cells was performed by means of a FACSCalibur (Becton Dickinson, San Jose, CA) and the data were analysed with the help of CellQuest software (Becton Dickinson).

MHC stabilization assay

The seven synthesized peptides were used in an MHC stabilization assay by using T2-A24 cells as described earlier (22). Briefly, T2-A24 cells (2×10^5) were incubated with 200 μl of RPMI1640 containing 0.1% FCS and 5×10^{-5} M β-mercaptoethanol and each of the peptides at a concentration of 10 μM at 26°C for 16 h, followed by incubation at 37°C for 3 h. Surface HLA-A24 molecules were then stained with the anti-A24 monoclonal antibody and FITC-labelled anti-mouse IgG. Expression was measured in the FACSCalibur, and mean fluorescence intensity (MFI) was recorded. The percentage of MFI increase was calculated as follows: percentage of MFI increase = 100 × (MFI with the given peptide – MFI without the peptide)/(MFI without the peptide).

Characteristics of epithelial cell dhesion molecule (Ep-CAM) candidate peptides

Peptide designation	Amino acid sequence	Position	Sequence length	Score[a]	Percentage of MFI increase[b]
Ep_{31}	NYKLAVNCF	31–39	9	120	85
Ep_{173}	RYQLDPKFI	173–181	9	150	102
Ep_{185}	LYENNVITI	185–193	9	75	79
Ep_{225}	LFHSKKMDL	225–233	9	20	29
Ep_{250}	YYVDEKAPEF	250–259	10	198	57
Ep_{296}	KYEKAEIKEM	296–305	10	83	24
Ep_{304}	EMGEMHREL	304–312	9	5	16

[a]Estimated half-time of dissociation from HLA-A24 molecules (min), obtained with a computer program (World Wide Web site BioInformatics & Molecular Analysis Section (BIMAS) HLA peptide-binding predictions).
[b]Synthetic peptides were tested for binding to human HLA-A*2402 molecules in MHC stabilization assays as described in the section entitled 'Materials and methods.' MFI, mean fluorescence intensity.

Table 1

Generation of Ep-CAM peptide-specific CTL lines and clones

Peripheral blood monocyte-derived dendritic cells (DCs) were generated as described previously (28). Briefly, plastic adherent cells were isolated from PBMCs and were cultured in RPMI1640 medium supplemented with 5% heat-inactivated human serum, 10 ng/ml of recombinant human interleukin-4 (IL-4) (R&D Systems, Minneapolis, MN) and 50 ng/ml of recombinant human granulocyte-macrophage colony-stimulating factor (R&D Systems). On day 1 of incubation, 10 ng/ml of IL-1β PeproTech, Rocky Hill, NJ), 50 ng/ml of recombinant human tumour necrosis factor-α (TNF-α) (PeproTech) and 1 μM prostaglandin E_2 (Cayman Chemical Company, Ann Arbor, MI) was added for maturation. On days 2 or 3, the cells were harvested and were confirmed to express mature DC-associated antigens, such as CD1a, CD80, CD83, CD86 and HLA class-II molecules (data not shown). The DCs were pulsed with each of the synthetic peptides at a concentration of 10 μM in AIM-V medium (Gibco, Grand Island, NY) supplemented with 5×10^{-5} M β-mercaptoethanol for 2–4 h at room temperature and were irradiated (33 Gy). Thereafter, the DCs (1×10^5) were co-cultured with autologous $CD8^+$ T lymphocytes (1×10^6) purified with the aid of CD8 MicroBeads (Miltenyi Biotec, Bergisch Gladbach, Germany) in RPMI1640 medium supplemented with 10% pooled human serum, 25 ng/ml of recombinant human IL-7 (R&D Systems) and 5 ng/ml of recombinant human IL-12 (R&D Systems) in a culture tube. After culture for 7 days, the cells were stimulated again by adding 1×10^5 peptide-pulsed autologous DCs prepared as described above. After culture for 7 additional days, the cells were stimulated a third time in the same manner. One day after each restimulation, recombinant human IL-2 (Takeda Chemical Industries, Osaka, Japan) was added to a final concentration of 20 U/ml. If necessary, rapidly growing cells were split into two to three tubes and were fed with fresh culture medium containing 20 U/ml of IL-2.

In order to establish T-cell clones, limiting dilution of the polyclonal CTLs was performed (22). After 2-week culture in 96-well plates, the specificity of growing cells was examined with CTL–CTL killing assays as previously described (29). Clones that were