apy was graded according to the Radiation Therapy Oncology Group/European Organization for Research and Treatment of Cancer late radiation morbidity scoring scheme. Vinorelbine administration on day 8 was omitted if any of the following were noted: WBC count $<3.0 \times 10^9$/liter, neutrophil count $<1.5 \times 10^9$/liter, platelet count $<100 \times 10^9$/liter, elevated hepatic transaminase level or total serum bilirubin of at least grade 2, fever ≥38°C, or PS ≥2. Subsequent cycles of cisplatin and vinorelbine chemotherapy were delayed if any of the following toxicities were noted on day 1: WBC count $<3.0 \times 10^9$/liter, neutrophil count $<1.5 \times 10^9$/liter, platelet count $<100 \times 10^9$/liter, serum creatinine level ≥1.6 mg/dl, elevated hepatic transaminase level or total serum bilirubin of at least grade 2, fever ≥38°C, or PS ≥2. The dose of cisplatin was reduced by 25% in all subsequent cycles if the serum creatinine level rose to 2.0 mg/dl or higher. The dose of vinorelbine or docetaxel was reduced by 25% in all subsequent cycles if any of the following toxicities were noted: WBC count $<1.0 \times 10^9$/liter, platelet count $<10 \times 10^9$/liter, or grade 3 or 4 infection or liver dysfunction. Thoracic radiotherapy was suspended if any of the following were noted: fever ≥38°C, grade 3 esophagitis, PS of 3, or $PaO_2$ <70 torr. Thoracic radiotherapy was terminated if any of the following were noted: grade 4 esophagitis, grade 3 or 4 pneumonitis, PS of 4, or duration of radiotherapy of over 60 days. The use of granulocyte colony-stimulating factor during radiotherapy was not permitted unless radiotherapy was on hold. The criteria for termination of docetaxel consolidation were not defined in the protocol.

## Response Evaluation

Objective tumor response was evaluated according to the Response Evaluation Criteria in Solid Tumor.[11] Local recurrence was defined as tumor progression in the primary site and in the hilar, mediastinal, and supraclavicular lymph nodes after a partial or complete response; regional recurrence as the development of malignant pleural and pericardial effusions; and distant recurrence as the appearance of a distant metastasis.

## Study Design, Data Management, and Statistical Considerations

This study was conducted at three institutions: the National Cancer Center Hospital, National Cancer Center Hospital East, and Tochigi Cancer Center. The protocol and consent form were approved by the institutional review board of each institution. Registration was conducted at the registration center. Data management, periodic monitoring, and the final analysis were performed by the study coordinator.

The primary objective of the current study was to evaluate the feasibility of docetaxel consolidation therapy. The secondary endpoints were toxicity observed during chemoradiotherapy and consolidation therapy, the best response, and overall survival in all patients eligible to participate in this study. Because no standard method to evaluate consolidation chemotherapy after chemoradiotherapy has been established, we arbitrarily defined the primary endpoint of this study as a ratio (R) of the number of patients receiving docetaxel without grade 4 nonhematological toxicity or treatment-related death to the total number of patients receiving docetaxel. The sample size was initially estimated to be 34 patients with a power of 0.80 at a significance level of 0.05, under the assumption that a R of 0.95 would indicate potential usefulness, whereas a R of 0.8 would be the lower limit of interest, and that 85% of patients would move into the consolidation phase. An analysis of the first 13 patients, however, showed that only 8 (61%) patients advanced into the consolidation phase. The reasons for not receiving docetaxel were disease progression in one, delay in completion of chemoradiotherapy in two, grade 3 esophagitis in one, and death due to hemoptysis in one patient. Considering that the SWOG trial S9504 included 83 patients, we decided to revise the number of patients in the current study. According to Simon's two-stage minimax design, the required number of patients was calculated to be 59 with a power of 0.80 at a significance level of 0.05, under the assumption that a R of 0.85 would indicate potential usefulness, whereas a R of 0.7 would be the lower limit of interest.[12] Assuming that 61% of registered patients would move into the consolidation phase, the sample size was determined to be 97 patients.

Overall survival time and progression-free survival time were estimated by the Kaplan–Meier method, and confidence intervals (CI) were based on Greenwood's formula.[13] Overall survival time was measured from the date of registration to the date of death (from any cause) or to the last follow-up. Progression-free survival time was measured from the date of registration to the date of disease progression, death (from any cause), or the last follow-up. Patients who were lost to follow-up without event were censored at the date of their last known follow-up. A CI for RR was calculated using methods for exact binomial CIs. The Dr. SPSS II 11.0 for Windows software package (SPSS Japan Inc., Tokyo, Japan) was used for statistical analyses.

# RESULTS

## Registration and Characteristics of the Patients

A total of 97 patients were enrolled in this study between April 2001 and June 2003. Four patients were excluded from this study before the treatment was started because the radiation treatment planning disclosed that their tumors were too advanced for curative thoracic radiotherapy. Thus, 93 patients who received the protocol-defined treatment were the subjects of this analysis (Figure 2). There were 76 males and 17 females, with a median age of 60 (range 31–74). Body weight loss was less than 5% in 77 patients; adenocarcinoma histology was noted in 57 patients, and stage IIIA disease was noted in 41 patients (Table 1).

## Treatment Delivery

Treatment delivery was generally well maintained in the chemoradiotherapy phase (Table 2). Full cycles of cisplatin and vinorelbine and the full dose of thoracic radiotherapy were administered in 80 (86%) and 87 (94%) patients, respectively. Delay in radiotherapy was less than 5 days in 61 (66%) patients. In contrast, the delivery of docetaxel was poor (Table 2). A total of 59 (63%) patients could enter the consolidation phase, and only 34 (37%) patients completed three cycles of docetaxel chemotherapy. The reasons for not

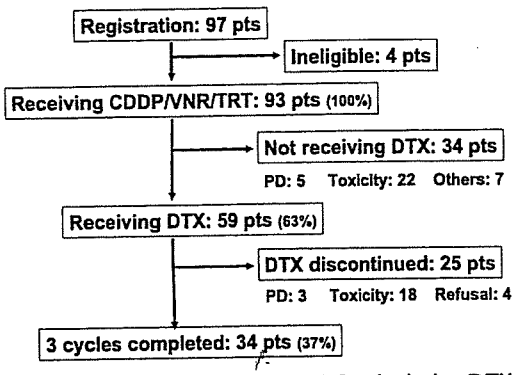


FIGURE 2. Patient registration. CDDP, cisplatin; DTX, docetaxel; TRT, thoracic radiotherapy; VNR, vinorelbine.

receiving consolidation were toxicity in 22 (65%) patients including pneumonitis in seven patients, myelosuppression in five patients, esophagitis in four patients, liver dysfunction in two patients, infection in two patients, other toxicity in two patients, progressive disease in five (15%) patients, patient refusal in three (9%) patients, early death due to hemoptysis in one (3%) patient, and other reasons in three (9%) patients. Of the 59 patients, 18 (31%) discontinued docetaxel consolidation because of toxicity, including pneumonitis ($n$ = 14) and esophagitis, infection, gastric ulcer, and allergic reaction ($n$ = 1 each), four (7%) because of patient refusal, and three (5%) because of progressive disease.

## Toxicity

Acute severe toxicity in the chemoradiotherapy phase was mainly leukopenia and neutropenia, whereas grade 3 or 4 thrombocytopenia was not noted (Table 3). Severe nonhematological toxicity was sporadic, and grade 3 esophagitis and pneumonitis were observed in only 11 (12%) and 3 (3%) patients, respectively. Acute severe toxicity in the consolidation phase also consisted of neutropenia and associated infection (Table 4). In addition, grade 3 or 4 pneumonitis developed in 4 (7%) patients. The R observed in this study was 0.05 (3 out of 57 patients), which was much lower than the hypothetical value. Grade 3 or 4 late toxicities were included lung toxicity in four patients, esophageal toxicity in two patients, renal toxicity in one patient, and a second esophageal cancer that developed 35.4 months after the start of the chemoradiotherapy in one patient. Treatment-related

TABLE 1. Patient Characteristics

| Characteristics | $n$ | % |
|---|---|---|
| Gender | | |
| Male | 76 | 82 |
| Female | 17 | 18 |
| Age median (range) | 60 | 31–74 |
| Weight loss | | |
| <5% | 76 | 81 |
| 5–9% | 12 | 13 |
| ≥10% | 3 | 3 |
| Unknown | 2 | 2 |
| Histology | | |
| Adenocarcinoma | 57 | 61 |
| Squamous cell carcinoma | 23 | 25 |
| Large cell carcinoma | 12 | 13 |
| Others | 1 | 1 |
| Stage | | |
| IIIA | 41 | 44 |
| IIIB | 52 | 56 |

TABLE 2. Treatment Delivery

| Variables | $n$ | % |
|---|---|---|
| Cisplatin and vinorelbine chemotherapy | | |
| Total number of cycles | | |
| 3 | 80 | 86 |
| 2 | 10 | 11 |
| 1 | 3 | 3 |
| Number of vinorelbine skips | | |
| 0 | 63 | 68 |
| 1 | 25 | 27 |
| 2–3 | 5 | 5 |
| Thoracic radiotherapy | | |
| Total dose (Gy) | | |
| 60 | 87 | 94 |
| 50–59 | 4 | 4 |
| <50 | 2 | 2 |
| Delay (days) | | |
| <5 | 61 | 66 |
| 5–9 | 20 | 22 |
| 10–16 | 6 | 6 |
| Not evaluable (<60 Gy) | 6 | 6 |
| Docetaxel consolidation | | |
| Number of cycles | | |
| 3 | 34 | 37 |
| 2 | 12 | 13 |
| 1 | 13 | 14 |
| 0 | 34 | 34 |

TABLE 3. Acute Toxicity in Chemoradiotherapy ($n$ = 93)

| | Grade | | | |
|---|---|---|---|---|
| Toxicity | 3 | 4 | 3 + 4 | % |
| Leukopenia | 54 | 18 | 72 | 77 |
| Neutropenia | 33 | 29 | 62 | 67 |
| Anemia | 21 | 0 | 21 | 23 |
| Infection | 15 | 1 | 16 | 17 |
| Esophagitis | 11 | 0 | 11 | 12 |
| Hyponatremia | 11 | 0 | 11 | 12 |
| Anorexia | 9 | 1 | 10 | 11 |
| Nausea | 5 | — | 5 | 5 |
| Pneumonitis | 3 | 0 | 3 | 3 |
| Syncope | 2 | 0 | 2 | 2 |
| Hyperkalemia | 2 | 0 | 2 | 2 |
| Ileus | 0 | 1 | 1 | 1 |
| Cardiac ischemia | 1 | 0 | 1 | 1 |

TABLE 4. Acute Toxicity in Consolidation Therapy ($n$ = 57)

| | Grade | | | |
|---|---|---|---|---|
| Toxicity | 3 | 4 | 3 + 4 | % |
| Leukopenia | 33 | 11 | 44 | 77 |
| Neutropenia | 24 | 26 | 50 | 88 |
| Anemia | 5 | 0 | 5 | 9 |
| Infection | 5 | 1 | 6 | 11 |
| Esophagitis | 2 | 0 | 2 | 3 |
| Anorexia | 1 | 0 | 1 | 2 |
| Pneumonitis | 2 | 2 | 4 | 7 |

death was observed in four (4%) patients. Of these, three received docetaxel, and one did not. The reason for death was pneumonitis in all patients. We calculated a $V_{20}$ (the percent volume of the normal lung receiving 20 Gy or more) on a dose–volume histogram in 25 patients. Of these, five patients developed grade 3 or severer radiation pneumonitis. A median $V_{20}$ for these five patients was 35% (range, 26–40%), whereas that for the remaining 20 patients was 30% (range, 17–35%) ($p$ = 0.035 by a Mann–Whitney test).

## Objective Responses, Relapse Pattern, and Survival

All 93 patients were included in the analyses of tumor response and survival. Complete and partial responses were obtained in 5 (5%) and 71 patients (76%), respectively, for an overall RR of 81.7% (95% CI, 72.7–88.0%). Stable and progressive diseases occurred in 12 (13%) and 5 (5%) patients, respectively. With a median follow-up period of 29.7 months, 38 patients developed locoregional recurrence, 32 developed distant recurrence, 4 developed both locoregional and distant recurrences, and 19 did not. The median progression-free survival time was 12.8 (95% CI, 10.2–15.4) months (Figure 3). Two patients underwent salvage surgery for a recurrent primary tumors. Conventional chemotherapy and gefitinib monotherapy were administered after recurrence in 20 and 25 patients, respectively. The median overall survival time was 30.4 (95% CI, 24.5–36.3) months. The 1-, 2-, and 3-year survival rates were 80.7, 60.2, and 42.6%, respectively. (Figure 4).

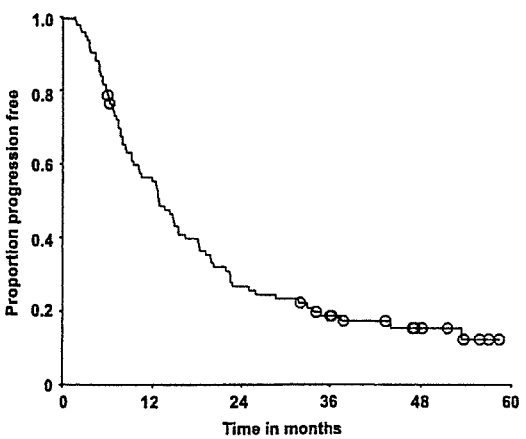


FIGURE 3. Progression-free survival ($n$ = 93). The median progression-free survival time was 12.8 (95% CI, 10.2–15.4) months.

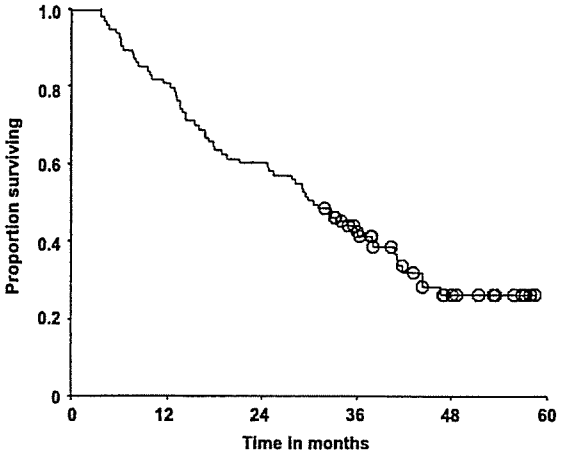


FIGURE 4. Overall survival ($n$ = 93). The median overall survival time was 30.4 (95% CI, 25.4–35.4) months. The 1-, 2-, and 3-year survival rates were 80, 60, and 40%, respectively.

## DISCUSSION

This study showed that concurrent chemoradiotherapy with cisplatin, vinorelbine, and standard thoracic radiotherapy was well tolerated, with a high completion rate exceeding 80%. The incidence of acute toxicity, including 67% (62/93) of grade 3 or 4 neutropenia, 12% (11/93) of grade 3 esophagitis, and 3% (3/93) of grade 3 pneumonitis, were comparable with other reports of concurrent chemoradiotherapy.[3,4,10] In contrast, consolidation docetaxel could be administered in only 59 of 93 (63%) patients eligible to participate in this study. Of the remaining 34 patients, 22 (65%) patients did not receive consolidation chemotherapy because of toxicities affecting various organs. Other studies also showed that not all patients proceeded to the consolidation phase after completion of concurrent chemoradiotherapy: 61 to 78% of patients after two cycles of cisplatin and etoposide with radiotherapy,[3,10] and 54 to 75% of patients after weekly carboplatin and paclitaxel with radiotherapy.[14,15] Thus, for 20 to 40% of the patients, concurrent chemoradiotherapy was as much as they could undergo, and the additional chemotherapy was not practical.

Furthermore, the number of patients who fulfilled the three cycles of consolidation docetaxel was only 34 (58%) of the 59 patients, which corresponded to only 37% of those eligible in this study. The reason for the termination of docetaxel in the 25 patients was toxicity in 18 (72%) patients, especially pneumonitis in 14 (56%) patients. The grade of pneumonitis during the consolidation phase was within grade 2 in most cases, and this was probably because docetaxel was discontinued early. Considering that pneumonitis associated with cancer treatment is more common in Japan, docetaxel consolidation is not thought to be feasible in the Japanese population. The MST and the 3-year survival rate in all eligible patients were 33 months and 44% in this study, but docetaxel consolidation was unlikely to contribute to these promising results because only 37% of patients received full cycles of docetaxel. This contrasts clearly with the result of

the SWOG study S9504, a phase II trial of two cycles of cisplatin and etoposide with thoracic radiation followed by three cycles of docetaxel. In this trial, 75% of patients starting consolidation and 59% of those entering the trial received full cycles. In addition, docetaxel consolidation seemed to prolong survival, although this was drawn from a retrospective comparison of the results between the two SWOG studies S9504 and S9019.[10]

There is no widely used definition of consolidation therapy following chemoradiotherapy. Given that consolidation therapy is arbitrarily defined as chemotherapy with three cycles or more after the completion of concurrent chemoradiotherapy, only one randomized trial is available in the literature. The randomized phase III trial of standard chemoradiotherapy with carboplatin and paclitaxel followed by either weekly paclitaxel or observation in patients with stage III NSCLC showed that only 54% of patients proceeded to randomization, and overall survival was worse in the consolidation arm (MST, 16 versus 27 months).[15] Thus, there have been no data supporting the use of consolidation therapy, especially when a third-generation cytotoxic agent such as paclitaxel and vinorelbine is incorporated into concurrent chemoradiation therapy.

The low complete-response rate of 5% in this study may be explained partly by an inability to distinguish between inactive scarring or necrotic tumor and active tumor after radiotherapy. Positron emission tomography (PET) using 18F-fluorodeoxyglucose showed a much higher rate of complete response than conventional CT scanning and provided a better correlation of the response assessment using PET with patterns of failure and patient survival.[16] In addition, the high locoregional relapse rate in this study clearly showed that the conventional total dose of 60 Gy was insufficient. Three-dimensional treatment planning, omission of elective nodal irradiation, and precise evaluation of the gross tumor volume by PET may facilitate the escalation of the total radiation dose without enhanced toxicity.

In conclusion, cisplatin and vinorelbine chemotherapy concurrently combined with standard thoracic radiotherapy and followed by docetaxel consolidation produced promising overall survival in patients with stage III NSCLC, but the vast majority of patients could not continue with the docetaxel consolidation because of toxicity.

## ACKNOWLEDGMENTS

*We thank residents and staff doctors in the National Cancer Center Hospital, National Cancer Center Hospital East, and Tochigi Cancer Center for their care of patients and valuable suggestions and comments on this study. We would also like to thank Fumiko Koh, Yuko Yabe, and Mika Nagai for preparation of the manuscript.*

*This study was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan.*

## REFERENCES

1. Vokes EE, Crawford J, Bogart J, et al. Concurrent chemoradiotherapy for unresectable stage III non-small cell lung cancer. *Clin Cancer Res* 2006;11:5045s–5050s.
2. Auperin A, Le Pechoux C, Pignon JP, et al. Concomitant radio-chemotherapy based on platin compounds in patients with locally advanced non-small cell lung cancer (NSCLC): a meta-analysis of individual data from 1764 patients. *Ann Oncol* 2006;17:473–483.
3. Fournel P, Robinet G, Thomas P, et al. Randomized phase III trial of sequential chemoradiotherapy compared with concurrent chemoradiotherapy in locally advanced non-small-cell lung cancer: Groupe Lyon-Saint-Etienne d'Oncologie Thoracique-Groupeqqfrancaisqqqq de Pneumo-Cancerologie NPC 95-01 Study. *J Clin Oncol* 2006;23:5910–5917.
4. Furuse K, Fukuoka M, Kawahara M, et al. Phase III study of concurrent versus sequential thoracic radiotherapy in combination with mitomycin, vindesine, and cisplatin in unresectable stage III non-small-cell lung cancer. *J Clin Oncol* 1999;17:2692–2699.
5. Curran W, Scott CJ, Langer C, et al. Long-term benefit is observed in a phase III comparison of sequential vs concurrent chemo-radiation for patients with unresected stage III NSCLC: RTOG 9410. *Proc Am Soc Clin Oncol* 2003;22:621 (abstr 2499).
6. Sekine I, Noda K, Oshita F, et al. Phase I study of cisplatin, vinorelbine, and concurrent thoracic radiotherapy for unresectable stage III non-small cell lung cancer. *Cancer Sci* 2004;95:691–695.
7. Fossella FV, DeVore R, Kerr RN, et al. Randomized phase III trial of docetaxel versus vinorelbine or ifosfamide in patients with advanced non-small-cell lung cancer previously treated with platinum-containing chemotherapy regimens. The TAX 320 Non-Small Cell Lung Cancer Study Group. *J Clin Oncol* 2000;18:2354–2362.
8. Shepherd FA, Dancey J, Ramlau R, et al. Prospective randomized trial of docetaxel versus best supportive care in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy. *J Clin Oncol* 2000;18:2095–2103.
9. Fossella FV, Lee JS, Shin DM, et al. Phase II study of docetaxel for advanced or metastatic platinum-refractory non-small-cell lung cancer. *J Clin Oncol* 1995;13:645–651.
10. Gandara DR, Chansky K, Albain KS, et al. Consolidation docetaxel after concurrent chemoradiotherapy in stage IIIB non-small-cell lung cancer: phase II Southwest Oncology Group Study S9504. *J Clin Oncol* 2003; 21:2004–2010.
11. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92: 205–216.
12. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;10:1–10.
13. Armitage P, Berry G, Matthews J. Survival analysis. In Armitage P, Berry G, Matthews J (eds.), *Statistical Methods in Medical Research* (4th ed.). Oxford: Blackwell Science Ltd, 2002, pp. 568–590.
14. Belani CP, Choy H, Bonomi P, et al. Combined chemoradiotherapy regimens of paclitaxel and carboplatin for locally advanced non-small-cell lung cancer: a randomized phase II locally advanced multi-modality protocol. *J Clin Oncol* 2006;23:5883–5891.
15. Carter D, Keller A, Tolley R, et al. A randomized phase III trial of combined paclitaxel, carboplatin, and radiation therapy followed by either weekly paclitaxel or observation in patients with stage III non-small cell lung cancer. *Proc Am Soc Clin Oncol* 2006;22:635s (abstr 7076).
16. Mac Manus MP, Hicks RJ, Matthews JP, et al. Metabolic (FDG-PET) response after radical radiotherapy/chemoradiotherapy for non-small cell lung cancer correlates with patterns of failure. *Lung Cancer* 2006; 49:95–108.

# A Literature Review of Molecular Markers Predictive of Clinical Response to Cytotoxic Chemotherapy in Patients with Lung Cancer

*Ikuo Sekine, MD,* John D. Minna, MD,† Kazuto Nishio, MD,‡ Tomohide Tamura, MD,* and Nagahiro Saijo, MD**

**Background:** To find candidate genes for a predictive chemosensitivity test in patients with lung cancer by using a literature review.
**Methods:** Using MEDLINE searches, "in vitro chemosensitivity associated genes" and articles on association of the gene alteration with clinical chemosensitivity in lung cancer patients were selected. We calculated odds ratios (ORs) and their 95% confidence intervals (95% CIs) of response rates for patients who had tumors with or without gene alteration. Combined ORs and 95% CIs were estimated using the DerSimonian-Laird method.
**Results:** Of the 80 in vitro chemosensitivity-associated genes identified, 13 genes were evaluated for association with clinical chemosensitivity in 27 studies. The median (range) number of patients in each study was 50 (range, 28-108). The response rates of lung cancer with high and low P-glycoprotein expression were 0% and 73% to 85%, respectively ($p < 0.001$). Glutathione S-transferase pi expression (OR 0.22, 95% CI 0.06-0.79), excision repair cross-complementing 1 alterations (combined OR 0.53, 95% CI 0.28-1.01; $p = 0.055$), and tumor suppressor p53 mutation (combined OR 0.25, 95% CI 0.12-0.52) were associated with clinical chemosensitivity.
**Conclusion:** In total, 80 in vitro chemosensitivity-associated genes were identified in the literature, and high and low P-glycoprotein, glutathione S-transferase pi expression, excision repair cross-complementing 1 alterations, and tumor suppressor p53 mutation were candidates for future clinical trials of chemosensitivity tests in lung cancer patients.

**Key Words:** chemotherapy, drug response, molecular markers, prediction, lung cancer

(*J Thorac Oncol.* 2006;1: 31–37)

Lung cancer is the leading cause of death in many countries despite extensive basic research and clinical trials. Approximately 80% of patients with lung cancer have developed distant metastases either by the time of initial diagnosis or during recurrence after surgery for local disease. Systemic chemotherapy against lung cancer, however, has limitations in efficacy such that patients with distant metastases rarely live long.[1]

Tumor response to chemotherapy varies among patients, and objective tumor response rates to standard chemotherapy regimens are approximately 20 to 40% in patients with non—small-cell lung cancer and 60 to 90% in patients with small-cell lung cancer. Thus, it would be extremely useful to know in advance whether patients have tumors that respond to chemotherapy agents and whether the tumors would be resistant to such therapy. For this purpose, cell culture-based chemosensitivity tests have been investigated for more than 20 years, but they are not widely accepted because of technical problems such as the large amount of material required, a low success rate for the primary culture, length of time required, and poor correlation with the clinical response.[2–5]

To overcome these obstacles, DNA-, RNA-, and protein-based chemosensitivity tests have been created, but gene alterations that are predictive of the clinical drug response are not established. Recently, as many as 400 genes whose expression was associated with drug response were identified by cDNA microarray studies, but their functions do not seem to be related to drug sensitivity or resistance.[6–10] In addition, the genes identified by microarray studies were highly unstable and depended on the selection of patients used for gene identification.[11,12] The purpose of this study was to provide an overview of gene alterations in lung cancer that are associated with chemotherapy drug response to identify candidate genes for predictive chemosensitivity tests in patients with lung cancer.

## MATERIALS AND METHODS

Because one set of genes associated with chemosensitivity is those directly involved in drug resistance mechanisms, we conducted a MEDLINE search for articles on tumor drug resistance published in the years 2001–2003. This search yielded 112 studies, including several review articles. By searching manually through these articles, we identified 134 genes or gene families that may be involved in drug resistance based on their function. We conducted the second MEDLINE searches for papers of in vitro studies on the 134 genes or gene families by using their names as a keyword.

*Divisions of Thoracic Oncology and Internal Medicine, National Cancer Center Hospital, Tokyo, Japan; †Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center at Dallas, Dallas, TX; ‡Pharmacology Division, National Cancer Center Research Institute, Tokyo, Japan.

ISSN: 1556-0864/06/0101-0031

From the 134 genes, we selected genes that met the following definition of "in vitro chemosensitivity associated genes": 1) alteration of the gene was identified in a human drug-induced resistant, solid tumor cell line; 2) transfection of the gene induced drug resistance; or 3) down-regulation of the gene or its encode protein increased drug sensitivity. In this latter category, we included studies in which the gene expression or function was suppressed by antisense RNA, hammerhead ribozyme, or an antibody against the gene product. We excluded studies in which drugs were used to inhibit function because the specificity of the drug against the target may not have been complete. We performed a third MEDLINE search for articles on the association between the gene alteration and chemosensitivity of lung cancer cell lines by using the name of the gene as a keyword. Articles in which the association was evaluated in 20 or more cell lines were included in this study. Finally, we searched MEDLINE for studies on the association between the gene alteration and clinical drug response in patients with lung cancer by using the name of the gene as a keyword. Articles in which the association was evaluated in 25 or more patients with advanced lung cancer were included in this study. Studies in which gene expression was evaluated with microarray were excluded because result analysis and interpretation of this technique have not been established, as indicated by the fact that the list of genes identified by microarray studies was highly variable without overlap between these gene sets.[11,12] Clinical studies on concurrent chemoradiotherapy were excluded. We constructed 2 × 2 tables from the response data and calculated odds ratios (ORs), their variances, and 95% confidence intervals (95% CIs) for the patients who had tumors with gene alteration relative to those who had tumors without gene alteration. Combined ORs and 95% CIs were estimated using the DerSimonian-Laird method.[13] When a response rate was 0, association with gene alteration was evaluated using the $\chi^2$ test because 95% CIs for ORs cannot be calculated. The name of each gene was standardized according to Human Gene Nomenclature Database of National Center for Biotechnology Information.

## RESULTS

Of the 134 genes or gene families found, a gene alteration in drug-induced resistant cells, an increased or decreased resistance in transfected cells, and an altered sensitivity in gene down-regulated cells were reported for 45, 57, and 32 genes, respectively. In total, 80 genes met the definition of "in vitro chemosensitivity associated gene" (Table 1).

Gene alteration was associated with in vitro chemosensitivity in 15 (50%) of 30 studies on 15 (56%) of 27 gene alterations (Table 2). Clinical drug response was evaluated in 27 studies on 13 gene alterations. The methods used to identify gene alteration included immunohistochemical protein expression analysis ($n$ = 18), polymerase chain reaction (PCR)-based mRNA expression analysis ($n$ = 3), and PCR-based mutation analysis ($n$ = 6). All but one clinical study was retrospective, and the median (range) number of patients in each study was 50 (28-108). Gene alteration was associated with clinical response in 8 of the 27 (30%) studies (Table 2).

**TABLE 1.** In Vitro Chemosensitivity-Associated Genes

| |
|---|
| Transporters: ABCA2, ABCB1, ABCB11, ABCC1, ABCC2, ABCC3, ABCC4, ABCC5, ABCG2, MVP, ATP7A, ATP7B, SLC29A1, SLC28A1, SLC19A1 |
| Drug targets: TUBB, TUBB4, TUBA, TYMS, TOP1, TOP2A, TOP2B, DHFR, |
| Target-associated proteins : MAP4, MAP7, STMN1, KIF5B, HSPA5, PSMD14, FPGS |
| Intracellular detoxifiers: GSTP1, GPX, GCLC, GGT2, MT, RRM2, AKR1B1 |
| DNA damage recognition and repair proteins: HMGB1, HMGB2, ERCC1, XPA, XPD, MSH2, MLH1, PMS2, APEX1, MGMT, BRCA1, GLO1 |
| Cell cycle regulators: RB1, GML, CDKN1A, CCND1, CDKN2A, CDKN1B |
| Mitogenic signal regulators: ERBB2, EGFR, KRAS2, HRAS, RAF1 |
| Survival signal regulators: AKT1, AKT2 |
| Integrin: ITGB1 |
| Transcription factors: JUN, FOS, MYC, NFKB1 |
| Apoptosis regulators: TP53, MDM2, TP73, BCL2, BCL2L1, MCL1, BAX, BIRC4, BIRC5, TNFRSF6, CASP3, CASP8, HSPB1 |

We evaluated the association between transporter P-glycoprotein/multidrug resistance 1 (ABCB1) expression and clinical chemosensitivity in four studies. The response rate of lung cancer with high ABCB1 expression was consistently 0%, whereas that for lung cancer with low ABCB1 expression was 73 to 85% (Table 3). Among drug targets, only topoisomerase II-beta (TOP2B) expression was associated with clinical drug response in patients with small-cell lung cancer (OR 0.29, 95% CI 0.09-0.95). The intracellular detoxifier glutathione s-transferase pi (GSTP1) was associated with both in vitro and clinical drug response (OR 0.22, 95% CI 0.06-0.79) (Table 4). DNA repair gene excision repair cross-complementing 1 (ERCC1) alterations were associated with drug response among patients with non—small-cell lung cancer with marginal statistical significance; the combined OR (95% CI) for ERCC1 alteration was 0.53 (0.28-1.01; $p$ = 0.055) (Table 5). Tumor suppressor p53 (TP53) mutation was the only alteration associated with drug response among patients with non—small-cell lung cancer among genes involved in cell cycle and apoptosis. A combined OR (95% CI) for TP53 among patients with non—small-cell lung cancer was 0.25 (0.12-0.52) (Table 6). B-cell CLL/lymphoma 2 (BCL2) and its family protein expression was not associated with clinical drug response (Table 7).

## DISCUSSION

We identified 80 in vitro chemosensitivity-associated genes in our literature search. Of these, 13 were evaluated clinically in 27 studies; ABCB1, TOP2B, GSTP1, and ERCC1 expression and TP53 mutation were associated with changes to drug responses among patients with lung cancer.

Classical drug resistance is believed to be the result of molecular changes inhibiting the drug-target interaction. ABCB1, an ATP-binding cassette protein, acts as an energy-dependent transmembrane efflux pump and decreases the intracellular accumulation of anticancer drugs, including anthracyclines, vinca alkaloids, taxanes, and epipodophyllotox-

TABLE 2. Chemosensitivity-Associated Genes and Association with Chemosensitivity

| | | Association with chemosensitivity | | | | | |
|---|---|---|---|---|---|---|---|
| | | In vitro studies (*n*) | | | Clinical studies (*n*) | | |
| Category | No of Genes | Total | Yes | % | Total | Yes | % |
| Transporter | 15 | 9 | 5 | 55 | 4 | 4 | 100 |
| Drug target | 8 | 2 | 1 | 50 | 5 | 1 | 20 |
| Target-associated protein | 7 | 0 | 0 | | 0 | 0 | |
| Intracellular detoxifier | 7 | 3 | 3 | 100 | 1 | 1 | 100 |
| DNA repair | 10 | 1 | 1 | 100 | 6 | 0 | 0 |
| Damage recognition protein | 2 | 0 | 0 | | 0 | 0 | |
| Cell cycle | 6 | 4 | 2 | 50 | 2 | 0 | 0 |
| Mitogenic signal | 5 | 3 | 1 | 33 | 1 | 0 | 0 |
| Survival signal | 2 | 0 | 0 | | 0 | 0 | |
| Transcription factor | 4 | 3 | 0 | 0 | 0 | 0 | |
| Cell adhesion-mediated drug-resistance protein | 1 | 0 | 0 | | 0 | 0 | |
| Apoptosis | 13 | 5 | 2 | 40 | 8 | 2 | 25 |
| Total | 80 | 30 | 15 | 50 | 27 | 8 | 30 |

TABLE 3. ABCB1 (P-Glycoprotein) and Clinical Response to Chemotherapy

| Author | Histology | Drugs | Method | Expression | Patients (*n*) | RR (%) | Odds ratio |
|---|---|---|---|---|---|---|---|
| Yeh et al.[30] | Non-small cell | Paclitaxel | IHC | Low | 35 | 80 | 0 |
| | | | | High | 15 | 0 | $p < 0.001$* |
| Kawasaki et al.[31] | Small cell | CAV or EP | IHC | Low | 26 | 85 | 0 |
| | | | | High | 4 | 0 | $p < 0.001$* |
| Hsia et al.[32] | Small cell | EP | IHC | Low | 37 | 73 | 0 |
| | | | | High | 13 | 0 | $p < 0.001$* |
| Savaraj et al.[33] | Small cell | CAV, CEV, or EP | RT-PCR | Low | 24 | 75 | 0 |
| | | | | High | 7 | 0 | $p < 0.001$* |
| Combined odds ratio for ABCB expression in patients with SCLC: 0 | | | | | | | |

IHC, Immunohistochemical analysis; RR, response rate; RT-PCR, reverse transcriptase-polymerase chain reaction.
*Calculated using the $\chi^2$ test because the confidence interval cannot be calculated.

ins. Overexpression of this protein gives tumor cells a multidrug resistance phenotype in vitro, which is thought to be associated with clinical chemoresistance.[14] Our review showed that the response rate of tumors with ABCB1 overexpression was 0 in all studies of lung cancer, whereas that for lung cancer tumors with low ABCB1 expression was 73 to 85% (Table 3).

There is a close relationship between drug sensitivity and quantitative and qualitative alterations of the drug's target, including tamoxifen sensitivity and estrogen receptor expression and trastuzumab response and Her-2/neu overexpression in breast cancer,[15] imatinib resistance and BCR-ABL gene amplification and mutations in Philadelphia chromosome-positive leukemias,[16] and imatinib response and KIT gene mutations in gastrointestinal stromal tumors.[17] In all of these cases, the target molecule is a receptor or a mutated tyrosine kinase located at the entry of growth-stimulating signal transduction pathways. Recently, gefitinib, a tyrosine kinase inhibitor of the epidermal growth factor receptor (EGFR), has been developed, and two large phase II trials showed a response rate of 18% and 12% in patients with non—small-cell lung cancer who were previously treated with conventional chemotherapy.[18,19] Responses to the drug have been unpredictable, but mutations of the EGFR gene were identified in patients with gefitinib-responsive lung cancer.[20,21] Furthermore, all mutations in these tumors were restricted to the activation loop of the kinase domain of EGFR, which are in distinct contrast to mutations in extracelluar and regulatory domains of EGFR in glioblastoma, which are unresponsive to gefitinib.[22] Thus, molecular developments of structure and function of the targets hold the promise of targeted cancer therapy. The target molecules of many anticancer cytotoxic agents have not been clearly defined; therefore, the relationship between the target molecule status and sensitivity to the agent has not been established. TOP2B expression was associated with drug response in patients with small-cell lung cancer, with a response rate of 71% for high TOP2B expression tumors versus 90% for low TOP2B expression tumors (OR 0.29, 95% CI 0.09-0.95).[23] This result, however, is in contrast with the idea that a higher

**TABLE 4.** Drug Targets, Intracellular Detoxifier, and Clinical Response to Chemotherapy

| Author | Histology | Drugs | Method | Expression | Patients (*n*) | RR (%) | Odds ratio (95% CI) |
|---|---|---|---|---|---|---|---|
| Beta-tubulin class III | | | | | | | |
| Rosell et al.[34] | Non-small cell | Paclitaxel, Vinorelbine | Real-time PCR | Low | 13 | 46 | 0.39 |
| | | | | High | 24 | 25 | (0.09-1.62) |
| Topoisomerase II-alpha | | | | | | | |
| Dingemans et al.[23] | Small cell | CEV or EP | IHC | Low | 65 | 85 | 0.65 |
| | | | | High | 23 | 80 | (0.20-2.17) |
| Dingemans et al.[35] | Non-small cell | Platinum-based | IHC | Low | 30 | 47 | 0.67 |
| | | | | High | 8 | 38 | (0.14-3.40) |
| Topoisomerase II-beta | | | | | | | |
| Dingemans et al.[23] | Small cell | CEV or EP | IHC | Low | 48 | 90 | 0.29 |
| | | | | High | 35 | 71 | (0.09-0.95) |
| Dingemans et al.[35] | Non-small cell | Platinum-based | IHC | Low | 18 | 50 | 0.86 |
| | | | | High | 13 | 46 | (0.21-3.58) |
| Glutathione s-transferase pi | | | | | | | |
| Nakanishi et al.[36] | Non-small cell | Cisplatin-based | IHC | Low | 17 | 47 | 0.22 |
| | | | | High | 37 | 16 | (0.06-0.79) |

CI, confidence interval; IHC, immunohistochemical analysis; PCR, polymerase chain reaction; RR, response rate; CEV, cyclophosphamide, etoposide, and vincristine; EP, etoposide and cisplatin.

**TABLE 5.** DNA Repair Genes and Clinical Response to Chemotherapy

| Author | Histology | Drugs | Method | Alteration | Patients (*n*) | RR (%) | Odd ratio (95% CI) |
|---|---|---|---|---|---|---|---|
| Excision repair cross-complementing 1 expression | | | | | | | |
| Lord et al. [37] | Non-small cell | Cisplatin, gemcitabine | Real-time PCR | Low | 23 | 52 | 0.38 |
| | | | | High | 24 | 36 | (0.11-1.26) |
| Excision repair cross-complementing 1 (ERCC1) polymorphism at codon 118 | | | | | | | |
| Ryu et al.[38] | Non-small cell | Cisplatin-based | PCR | C/C | 54 | 54 | 0.61 |
| | | | Hybridization | C/T or T/T | 53 | 42 | (0.28-1.31) |
| Combined odds ratio (95% C.I.) for ERCC1 alteration in patients with NSCLC0.53 (0.28-1.01, $p = 0.055$) | | | | | | | |
| Xeroderma pigmentosum group D polymorphism | | | | | | | |
| At codon 231 | | | | | | | |
| Ryu et al.[38] | Non-small cell | Cisplatin-based | PCR | G/G | 100 | 48 | 1.08 |
| | | | Hybridization | G/A or A/A | 8 | 50 | (0.26-4.57) |
| At codon 312 | | | | | | | |
| Camps et al.[39] | Non-small cell | Cisplatin, gemcitabine | PCR Sequencing | G/G | 18 | 17 | 3.33 |
| | | | | G/A or A/A | 15 | 40 | (0.66-16.7) |
| At codon 751 | | | | | | | |
| Camps et al.[39] | Non-small cell | Cisplatin, gemcitabine | PCR Sequencing | A/A | 22 | 23 | 2.04 |
| | | | | A/C or C/C | 16 | 38 | (0.49-8.45) |
| Ryu et al.[38] | Non-small cell | Cisplatin-based | PCR | A/A | 96 | 49 | 0.74 |
| | | | Hybridization | A/C | 12 | 42 | (0.22-2.51) |
| Combined odds ratio (95% CI) for XPD polymorphism in patients with NSCLC: 1.38 (0.68-2.78). | | | | | | | |

CI, confidence interval; PCR, polymerase chain reaction; RR, response rate; NSCLC, non—small-cell lung cancer; XPD, xeroderma pigmentosum group D.

**TABLE 6.** Cell Cycle Regulators, Mitogenic Signals, Tumor Protein p53, and Clinical Response to Chemotherapy

| Author | Histology | Drugs | Method | Alteration | Patients (*n*) | RR (%) | Odds ratio (95% CI) |
|---|---|---|---|---|---|---|---|
| Retinoblastoma 1 expression | | | | | | | |
| Gregorc et al.[40] | Non-small cell | Cisplatin-based | IHC | Low | 61 | 51 | 0.45 |
| | | | | High | 41 | 32 | (0.20-1.03) |
| Cyclin-dependent kinase inhibitor 1A, p21 expression | | | | | | | |
| Dingemans et al.[23] | Small cell | CEV, EP | IHC | Low | 63 | 90 | 0.57 |
| | | | | High | 22 | 71 | (0.17-1.92) |
| Kirsten rat sarcoma 2 viral oncogene homolog mutation | | | | | | | |
| Rodenhuis et al. [41, *a*] | Aenocarcinoma | Ifosfamide, carboplatin | PCR-MSH | Normal | 46 | 26 | 0.65 |
| | | | | Mutated | 16 | 19 | (0.16-2.70) |
| Tumor protein p53 (P53) mutation | | | | | | | |
| Nakanishi et al.[36] | Non-small cell | Cisplatin-based | IHC | Normal | 11 | 45 | 0.19 |
| | | | | Mutated | 29 | 15 | (0.04-0.94) |
| Gregorc et al.[40] | Non-small cell | Cisplatin-based | IHC | Normal | 56 | 57 | 0.26 |
| | | | | Mutated | 46 | 26 | (0.11-0.62) |
| Combined odds ratio (95% CI) for P53 mutation in patients with NSCLC: 0.25 (0.12-0.52) | | | | | | | |
| Kawasaki et al.[31] | Small cell | CAV or EP | IHC | Normal | 10 | 70 | 1.3 |
| | | | | Mutated | 20 | 75 | (0.24-6.96) |
| Dingemans et al.[23] | Small cell | CEV or EP | IHC | Normal | 47 | 85 | 0.81 |
| | | | | Mutated | 45 | 82 | (0.27-2.45) |
| Combined odds ratio (95% C.I.) for P53 mutation in patients with SCLC: 0.93 (0.37-2.35). | | | | | | | |

CI, confidence interval; IHC, immunohistochemical analysis; PCR-MSH, polymerase chain reaction-mutation specific hybridization; RR, response rate; CEV, cyclophosphamide, etoposide, and vincristine; EP, etoposide and cisplatin.
[a]Prospective study.

**TABLE 7.** B-Cell CLL/Lymphoma 2 (BCL2) Family Expression and Clinical Response to Chemotherapy

| Author | Histology | Drugs | Method | Expression | Patients (*n*) | RR (%) | Odds ratio (95% CI) |
|---|---|---|---|---|---|---|---|
| BCL2 | | | | | | | |
| Krug et al.[42] | Non-small cell | Docetaxel, vinorelbine | IHC | Low | 26 | 46 | 1.75 |
| | | | | High | 5 | 60 | (0.25-12.3) |
| Dingemans et al.[23] | Small cell | CEV or EP | IHC | Low | 20 | 79 | 1.36 |
| | | | | High | 71 | 85 | (0.38-4.86) |
| Takayama et al.[43] | Small cell | CAV or EP | IHC | Low | 17 | 76 | 0.50 |
| | | | | High | 21 | 62 | (0.12-2.08) |
| Combined odds ratio (95% CI) for BCL2 expression in patients with SCLC: 0.87 (0.33-2.32) | | | | | | | |
| BAX (BCL2-associated X protein) | | | | | | | |
| Krug et al.[42] | Non-small cell | Docetaxel, vinorelbine | IHC | Low | 9 | 56 | 0.72 |
| | | | | High | 19 | 47 | (0.15-3.54) |

CI, confidence interval; IHC, immunohistochemical analysis; RR, response rate; CEV, cyclophosphamide, etoposide, and vincristine; EP, etoposide and cisplatin.

expression of topoisomerase II enzymes correlates with greater chemosensitivity in patients with breast cancer.[24]

In addition to genes involved in classical drug resistance, genes that act downstream of the initial damage induced by a drug-target complex are thought to play an important role in chemosensitivity.[25] ERCC1 is a key enzyme in nucleotide excision repair, one of the key pathways by which cells repair platinum-induced DNA damage. High levels of ERCC1 mRNA have been associated with platinum resistance in the treatment of ovarian and gastric cancer.[26,27] The codon 118 in exon 4 of ERCC1 gene is polymorphic with the nucleotide alteration AAC to AAT. Although this base change results in coding for the same amino acid, it may affect gene expression based on the usage frequency of synonymous codons.[28] The associations between drug response and both ERCC1 gene expression and polymorphism at codon 118 in patients with non—small-cell lung cancer have been reported in the literature. A combined OR (95%

CI) for these ERCC1 alterations was 0.53 (0.28-1.01, $p$ = 0.055), although each study failed to show statistical significant association. Thus, ERCC1 may be a candidate for evaluation of the predictability of drug response in future clinical trials.

TP53, which is mutated or deleted in more than half of lung cancer cells, has a remarkable number of biological activities, including cell-cycle checkpoints, DNA repair, apoptosis, senescence, and maintenance of genomic integrity. Because most anticancer cytotoxic agents induce apoptosis through either DNA damage or microtubule disruption, mutated TP53 may decrease chemosensitivity by inhibiting apoptosis or, in contrast, may increase chemosensitivity by impairing DNA repair after drug-induced DNA damage.[29] This review showed that mutated TP53 was associated with poor drug response in patients with non—small-cell lung cancer (Table 6).

No other genes located downstream (including xeroderma pigmentosum group D, retinoblastoma 1, cyclin-dependent kinase inhibitor 1A, Kirsten rat sarcoma 2 viral oncogene homolog, B-cell CLL/lymphoma 2, and B-cell CLL/lymphoma 2-associated X protein) were associated with clinical drug response (Tables 5-7). The association was evaluated for only 8 of 43 *in vitro* chemosensitivity-associated downstream genes; therefore, key genes may be among the remaining 35 genes. Most clinical studies included a limited number of patients with various background characteristics such as tumor stage and chemotherapy regimen administered, which resulted in low statistical power to identify the association. Finally, because all but one study was retrospective, the quality of tumor samples may vary, and it is therefore unclear whether the gene alteration was detected in all samples. Thus, in future prospective clinical studies, the method of tumor sample collection and preservation, as well as immunohistochemistry and polymerase chain reaction-based methods, should be standardized, and the sample size of patients should be determined with statistical consideration.

The recently developed microarray technique enables investigators analyze mRNA expression of more than 20,000 genes at once, and as many as 100 to 400 genes were selected statistically as chemosensitivity-related genes.[6–8,10] Among them, however, only a limited number of genes were functionally related to chemosensitivity, and only ABCB1 and BAX corresponded with the 80 chemosensitivity-associated genes identified in this literature review, which were picked because of their known function and contribution to in vitro chemosensitivity. Thus, it will be interesting to evaluate the role of expression profile of these genes using microarray analysis.

The association between the expression and alterations of genes and clinical drug responses should be studied further in prospective trials. ABCB1, GSTP1, ERCC1, and TP53, and other genes identified by exploratory microarray analyses should be evaluated in those trials. Simple methods to identify gene alterations, such as immunohistochemistry and polymerase chain reaction-based techniques, will be feasible in future clinical trials because of their simplicity, cost, and time. The median number of patients in retrospective studies analyzed in this review was 50 (range, 28-108). In future prospective trials, sample size consideration for statistical power will also be important.

In conclusion, we identified 80 in vitro chemosensitivity-associated genes in a review of the literature; ABCB1, GSTP1, and ERCC1 expression and TP53 mutation were associated with drug responses among patients with lung cancer.

## ACKNOWLEDGMENTS

*This study was supported in part by Lung Cancer SPORE Grant P50CA70907, ATP010019, and Grants-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan. We thank Yuko Yabe and Mika Nagai for their collection and arrangement of numerous papers.*

## REFERENCES

1. Sekine I, Saijo N. Novel combination chemotherapy in the treatment of non-small cell lung cancer. *Expert Opin Pharmacother* 2000;1:1131–1161.
2. Gazdar AF, Steinberg SM, Russell EK, et al. Correlation of in vitro drug-sensitivity testing results with response to chemotherapy and survival in extensive-stage small cell lung cancer: a prospective clinical trial. *J Natl Cancer Inst* 1990;82:117–124.
3. Cortazar P, Johnson BE. Review of the efficacy of individualized chemotherapy selected by in vitro drug sensitivity testing for patients with cancer. *J Clin Oncol.* 1999;17:1625–1631.
4. Cortazar P, Gazdar AF, Woods E, et al. Survival of patients with limited-stage small cell lung cancer treated with individualized chemotherapy selected by in vitro drug sensitivity testing. *Clin Cancer Res* 1997;3:741–747.
5. Shaw GL, Gazdar AF, Phelps R, et al. Individualized chemotherapy for patients with non-small cell lung cancer determined by prospective identification of neuroendocrine markers and in vitro drug sensitivity testing. *Cancer Res* 1993;53:5181–5187.
6. Mariadason JM, Arango D, Shi Q, et al. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res* 2003;63:8791–8812.
7. Kikuchi T, Daigo Y, Katagiri T, et al. Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. *Oncogene* 2003;22:2192–2205.
8. Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;362:362–369.
9. Dan S, Tsunoda T, Kitahara O, et al. An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res* 2002;62:1139–1147.
10. Girard L, Sekine I, Shah J, et al. Correlation between in vitro drug sensitivity and microarray-based gene expression signatures in lung and breast cancer. In Proceedings of the 95th Annual Meeting of American Association for Cancer Research, Orlando, FL, March 27–31, 2004.Pp. 1098.
11. Ein-Dor L, Kela I, Getz G, et al. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 2006;21:171–178.
12. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2006;365: 488–492.
13. Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research, 4th ed. Oxford: Blackwell Science Ltd, 2002.
14. Sekine I, Saijo N. Polymorphisms of metabolizing enzymes and transporter proteins involved in the clearance of anticancer agents. *Ann Oncol* 2001;12:1515–1525.
15. Ellis M, Hayes DF, Lippman ME. Treatment of metastatic breast cancer. In Harris J, Lippman ME, Morrow M, Osborne CK (Eds.), Diseases of

 

the Breast, 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2003.Pp. 1101-1159.
16. Gambacorti-Passerini CB, Gunby RH, Piazza R, et al. Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias. *Lancet Oncol* 2003;4:75–85.
17. Heinrich MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol.* 2003;21:4342–4349.
18. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol* 2003;21:2237–2246.
19. Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003;290:2149–2158.
20. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
21. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304: 1497–1500.
22. Minna JD, Gazdar AF, Sprang SR, et al. Cancer: a bull's eye for targeted lung cancer therapy. *Science* 2004;304:1458–1461.
23. Dingemans AM, Witlox MA, Stallaert RA, et al. Expression of DNA topoisomerase IIalpha and topoisomerase IIbeta genes predicts survival and response to chemotherapy in patients with small cell lung cancer. *Clin Cancer Res* 1999;5:2048–2058.
24. Di Leo A, Isola J. Topoisomerase II alpha as a marker predicting the efficacy of anthracyclines in breast cancer: are we at the end of the beginning? *Clin Breast Cancer* 2003;4:179–186.
25. Johnstone RW, Ruefli AA, Lowe SW. Apoptosis: a link between cancer genetics and chemotherapy. *Cell* 2002;108:153–164.
26. Dabholkar M, Vionnet J, Bostick-Bruton F, et al. Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy. *J Clin Invest* 1994;94:703–708.
27. Metzger R, Leichman CG, Danenberg KD, et al. ERCC1 mRNA levels complement thymidylate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy. *J Clin Oncol* 1998;16:309–316.
28. Yu JJ, Mu C, Lee KB, et al. A nucleotide polymorphism in ERCC1 in human ovarian cancer cell lines and tumor tissues. *Mutat Res* 1997;382: 13–20.
29. Brown JM, Wouters BG. Apoptosis, p53, and tumor cell sensitivity to anticancer agents. *Cancer Res* 1999;59:1391–1399.
30. Yeh JJ, Hsu WH, Wang JJ, et al. Predicting chemotherapy response to paclitaxel-based therapy in advanced non-small-cell lung cancer with P-glycoprotein expression. *Respiration* 2003;70:32–35.
31. Kawasaki M, Nakanishi Y, Kuwano K, et al. Immunohistochemically detected p53 and P-glycoprotein predict the response to chemotherapy in lung cancer. *Eur J Cancer* 1998;34:1352–1357.
32. Hsia TC, Lin CC, Wang JJ, et al. Relationship between chemotherapy response of small cell lung cancer and P-glycoprotein or multidrug resistance-related protein expression. *Lung* 2002;180:173–179.
33. Savaraj N, Wu CJ, Xu R, et al. Multidrug-resistant gene expression in small-cell lung cancer. *Am J Clin Oncol* 1997;20:398–403.
34. Rosell R, Scagliotti G, Danenberg KD, et al. Transcripts in pretreatment biopsies from a three-arm randomized trial in metastatic non-small-cell lung cancer. *Oncogene* 2003;22:3548–3553.
35. Dingemans AC, van Ark-Otte J, Span S, et al. Topoisomerase IIalpha and other drug resistance markers in advanced non-small cell lung cancer. *Lung Cancer* 2001;32:117–128.
36. Nakanishi Y, Kawasaki M, Bai F, et al. Expression of p53 and glutathione S-transferase-pi relates to clinical drug resistance in non-small cell lung cancer. *Oncology* 1999;57:318–323.
37. Lord RV, Brabender J, Gandara D, et al. Low ERCC1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. *Clin Cancer Res* 2002;8:2286–2291.
38. Ryu JS, Hong YC, Han HS, et al. Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer* 2004; 44:311–316.
39. Camps C, Sarries C, Roig B, et al. Assessment of nucleotide excision repair XPD polymorphisms in the peripheral blood of gemcitabine/cisplatin-treated advanced non-small-cell lung cancer patients. *Clin Lung Cancer* 2003;4:237–241.
40. Gregorc V, Ludovini V, Pistola L, et al. Relevance of p53, bcl-2 and Rb expression on resistance to cisplatin-based chemotherapy in advanced non-small cell lung cancer. *Lung Cancer* 2003;39:41–48.
41. Rodenhuis S, Boerrigter L, Top B, et al. Mutational activation of the K-ras oncogene and the effect of chemotherapy in advanced adenocarcinoma of the lung: a prospective study. *J Clin Oncol* 1997;15:285–291.
42. Krug LM, Miller VA, Filippa DA, et al. Bcl-2 and bax expression in advanced non-small cell lung cancer: lack of correlation with chemotherapy response or survival in patients treated with docetaxel plus vinorelbine. *Lung Cancer* 2003;39:139–143.
43. Takayama K, Ogata K, Nakanishi Y, et al. Bcl-2 expression as a predictor of chemosensitivities and survival in small cell lung cancer. *Cancer J Sci Am* 1996;2:212.

 

The Pharmacogenomics Journal (2006) 6, 63–75

www.nature.com/tpj

ORIGINAL ARTICLE

# Haplotype structures of the *UGT1A* gene complex in a Japanese population

M Saeki[1], Y Saito[1,2], H Jinno[1,3], K Sai[1,4], S Ozawa[1,5], K Kurose[1,6], N Kaniwa[1,6], K Komamura[7,8], T Kotake[9], H Morishita[9], S Kamakura[7], M Kitakaze[7], H Tomoike[7], K Shirao[10], T Tamura[10], N Yamamoto[10], H Kunitoh[10], T Hamaguchi[10], T Yoshida[11], K Kubota[12], A Ohtsu[13], M Muto[12], H Minami[12], N Saijo[14], N Kamatani[15] and J-i Sawada[1,2]

[1]*Project Team for Pharmacogenetics, National Institute of Health Sciences, Tokyo, Japan;* [2]*Division of Biochemistry and Immunochemistry, National Institute of Health Sciences, Tokyo, Japan;* [3]*Division of Environmental Chemistry, National Institute of Health Sciences, Tokyo, Japan;* [4]*Division of Xenobiotic Metabolism and Disposition, National Institute of Health Sciences, Tokyo, Japan;* [5]*Division of Pharmacology, National Institute of Health Sciences, Tokyo, Japan;* [6]*Division of Medicinal Safety Science, National Institute of Health Sciences, Tokyo, Japan;* [7]*Division of Cardiology, National Cardiovascular Center, Osaka, Japan;* [8]*Department of Cardiovascular Dynamics Research Institute, National Cardiovascular Center, Osaka, Japan;* [9]*Department of Pharmacy, National Cardiovascular Center, Osaka, Japan;* [10]*Division of Internal Medicine, National Cancer Center Hospital, Tokyo, Japan;* [11]*Genetics Division, National Cancer Center Research Institute, National Cancer Center, Tokyo, Japan;* [12]*Division of Oncology/Hematology, National Cancer Center Hospital East, Chiba, Japan;* [13]*Division of GI Oncology/Digestive Endoscopy, National Cancer Center Hospital East, Chiba, Japan;* [14]*Deputy Director, National Cancer Center Hospital East, Chiba, Japan and* [15]*Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan*

**Correspondence:**
Dr Y Saito, Division of Biochemistry and Immunochemistry, National Institute of Health Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan.
E-mail: yoshiro@nihs.go.jp

**Received 1 July 2005; revised 12 August 2005; accepted 23 August 2005; published online 29 November 2005**

Genetic polymorphisms of UDP-glucuronosyltransferases (UGTs) are involved in individual and ethnic differences in drug metabolism. To reveal co-occurrence of the *UGT1A* polymorphisms, we first analyzed haplotype structures of the entire *UGT1A* gene complex using the polymorphisms from 196 Japanese subjects. Based on strong linkage disequilibrium between *UGT1A8* and *1A10*, among *1A9*, *1A7*, and *1A6*, and between *1A3* and *1A1*, the complex was divided into five blocks, *Block 8/10*, *Block 9/6*, *Block 4*, *Block 3/1*, and *Block C*, and the haplotypes for each block were subsequently determined/inferred. Second, using pyrosequencing or direct sequencing, additional 105 subjects were genotyped for 41 functionally tagged polymorphisms. The data from 301 subjects confirmed the robustness of block partitioning, but several linkages among the haplotypes with functional changes were found across the blocks. Thus, important haplotypes and their linkages were identified among the *UGT1A* gene blocks (and segments), which should be considered in pharmacogenetic studies.
*The Pharmacogenomics Journal* (2006) **6**, 63–75. doi:10.1038/sj.tpj.6500335; published online 29 November 2005

**Keywords:** UGT1A; single nucleotide polymorphism; linkage disequilibrium; haplotype

## Introduction

Glucuronidation, catalyzed by UDP-glucuronosyltransferases (UGTs), is one of the critical steps in the detoxification and elimination of various endogenous and exogenous compounds.[1,2] As for the genes coding UGTs, two subfamilies, *UGT1* and *UGT2*, have been identified in humans. The human *UGT1A* gene complex spans approximately 200 kb, is located on chromosome 2q37, and consists of nine active and four inactive exon 1 segments (in the following segment order: *UGT1A12P*, *1A11P*, *1A8*, *1A10*, *1A13P*, *1A9*, *1A7*, *1A6*, *1A5*, *1A4*, *1A3*, *1A2P*, and *1A1*) and common exons 2–5 (Figure 1). One of the nine active exon-1's (namely, *1A1* and *1A3–1A10*) can be used in conjunction with the common exons.[2,3] The UGT1A N-terminal domains (encoded by the exon-1's) determine the substrate-binding specificity and the C-terminal domain (encoded by exons 2–5) is important for binding to UDP-glucuronic acid.[1] Thus, the exon 1 segments confer the substrate specificity of UGT1A isoforms,[4] and the 5′-flanking region (and possibly the 3′-flanking region) of each exon 1 is acknowledged to independently regulate the expression of each isoform.[3,4]

A number of genetic polymorphisms including single nucleotide polymorphisms (SNPs) in *UGT1As* have been identified and published on the UDP glucuronosyltransferase home page (http://som.flinders.edu.au/FUSA/ClinPharm/UGT/allele_table.html). Some of these polymorphisms are known to affect glucuronidation rates.[5–14] Regarding *1A1*, a TATA box variant (−40_−39 insTA: **28* allele), increases the risk of irinotecan-induced toxicity via decrease

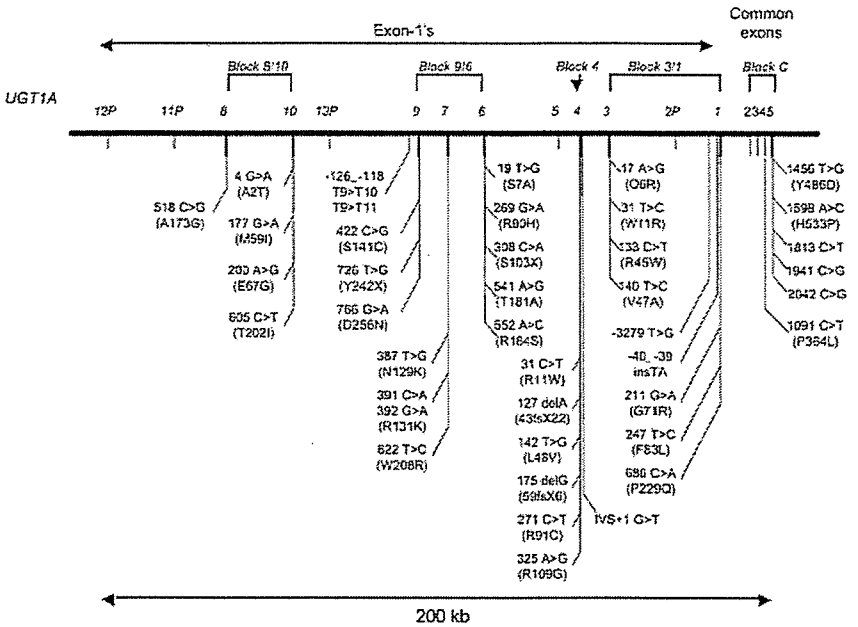


**Figure 1** The organization of the human *UGT1A* gene complex and the polymorphisms used for genotyping. The human *UGT1A* gene complex (nine active and four inactive exon-1's, and common exons 2–5) spans approximately 200 kb. Four pseudogenes (*1A2p*, *1A11p*, *1A12p*, and *1A13p*), and *1A5*, which mRNA expression has not yet been detected in any tissue, were not analyzed in this study and are depicted by the small and gray bars. The complex was divided into five blocks, *Block 8/10*, *Block 9/6*, *Block 4*, *Block 3/1*, and *Block C*.

in detoxicating glucuronidation of 7-ethyl-10-hydroxy-camptothecin (SN-38), an active metabolite of irinotecan.[5] Another *1A1* polymorphism, 211G>A (G71R: **6* allele), also shows reduced activity to SN-38.[7,15] In addition, it has been reported that the *1A7* alleles, **2* (387T>G (N129K), 391C>A and 392G>A (R131K)), **3* (N129K, R131K, and 622T>C (W208R)) and **4* (W208R) show reduced activities towards benzo(a)pyrene metabolites: for SN-38 glucuronidation, **3* and **4*, but not **2*, are less active.[6,7]

Haplotypes, linked combinations of SNPs on a chromosome, have the advantage of providing more useful information on phenotype–genotype links than individual SNPs.[16] Co-occurrence of the SNPs or segmental haplotypes with functional changes in the *UGT1A* complex could lead to a cooperative alteration in glucuronidation activity. Kohle *et al.*[17] reported close linkages among *1A1*28*, *1A6*2* (T181A/R184S), and *1A7*3* (N129K/R131K/W208R) in Caucasians and Egyptians. Moreover, a recent analysis has shown that the low-activity alleles, *1A7*2* and **3*, were completely associated with the *1A9* −126_−118 T9 allele, whereas the high-activity *1A7*1* allele was linked with the −126_−118 T10 allele (*1A9*22*: high expression) in Americans.[18] However, there is no haplotype analysis with a high SNP density for the entire *UGT1A* complex, especially for Asian populations, which includes the Japanese.

Previously, we have reported the segmental haplotype structures for *1A1*, *1A4*, and *1A6* exon-1's in Japanese subjects.[19–21] In this study, additional first exons (*1A3*, *1A7*, *1A8*, *1A9*, and *1A10*) and their surrounding promoter or intronic regions were sequenced for the same 196 Japanese subjects as used for the analysis of *1A1*, *1A4*, and *1A6*, and the haplotypes for the *UGT1A* complex were inferred in linkage disequilibrium (LD) blocks, *1A8-1A10*, *1A9-1A7-1A6*, and *1A3-1A1*. Then, the tagged polymorphisms with functional changes were genotyped for additional 105 Japanese subjects. Finally, several linkages among the block haplotypes were inferred in a total of 301 subjects and compared with those of other ethnic groups.

## Results

### UGT1A8, 1A10, 1A9, 1A7, *and* 1A3 *polymorphisms detected in a Japanese population*

All the exon-1's and their flanking regions of *UGT1A8*, *1A10*, *1A9*, *1A7*, and *1A3* were sequenced in 196 Japanese subjects (108 arrhythmic patients and 88 cancer patients). As for *1A6*, *1A4*, *1A1*, and common exons 2–5, their SNPs and segmental haplotypes have already been reported.[19–21] *UGT1A5* was omitted from the current analysis because the expression of *1A5* mRNA has not been shown in any tissue.[2] AF297093.1 (GenBank accession number) was used as the *UGT1* reference sequence. All the allele frequencies were in Hardy–Weinberg equilibrium. No statistically significant differences in allelic frequencies of the detected SNPs were found between the subjects with the different disease types ($P \geqslant 0.05$ by $\chi^2$ test or Fisher's exact test). Thus, the data for all subjects were analyzed as one group.

In *1A8*, seven SNPs, including two novel synonymous ones, were detected (Figure 2). The known nonsynonymous

SNP, 518C>G (A173G, **2* allele), was found at a frequency of 0.594. As for *1A10*, eight SNPs were detected. Among them, two polymorphisms, 4G>A (A2T) and 200A>G (E67G), were novel (Figure 2). Previously reported SNPs 177G>A (M59I, **2* allele) and 605C>T (T202I, **3* allele) were both found at a 0.010 frequency. Seven polymorphisms were detected in *1A9*, and two of them, −126_−118 T9>T11 and 422C>G (S141C), were novel (Figure 3). The SNPs 726T>G (Y242X, **4* allele) and 766G>A (D256N, **5* allele) were found at frequencies of 0.003 and 0.010, respectively. In this study, the known insertion, −126_−118 T9>T10 (**22* allele), was also detected at a 0.666 frequency. A total of nine SNPs including two novel ones (−70G>A, and 726T>C (Y242Y)) were detected in *1A7* (Figure 3). The known nonsynonymous SNPs, 387T>G (N129K), 391C>A and 392G>A (R131K: the SNPs at 391 and 392 are completely associated), and 622T>C (W208R), were also detected at frequencies of 0.350, 0.350, and 0.219, respectively. In *1A3*, 10 SNPs were detected, and only 46C>T (L16L) was novel. The known nonsynonymous SNPs,

| Gene | | | *UGT1A8* | | | | | | | *UGT1A10* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Position[a] | | | 34332 | 34781 | 34863 | 34874 | 35028 | 35067 | 35367 | 53106 | 53279 | 53302 | 53707 | 53795 | 54091 | 54163 | 54255 | |
| Nucleotide change[b] | | | 69 T>C | 518 C>G | 600 C>T | 711 A>C | 765 A>G | 804 T>C | IVS1 +189 G>A | 4 G>A | 177 G>A | 200 A>G | 605 C>T | 693 C>T | IVS1 +134 G>C | IVS1 +205 A>G | IVS1 +298 G>C | |
| Amino acid change | | | A23A[c] | A173G | A200A[c] | T237T[c] | T255T[c] | N268N[c] | | A2T | M59I | E67G | T202I | A231A[c] | | | | |
| Functional change | | | | no change | | | | | | | no change | | reduced | | | | | Frequency |
| Reference | | | novel | [8] | novel | rs 1126805[d] | rs 1042605[d] | Pharm GKB data base[e] | Pharm GKB data base[e] | novel | [11] | novel | [11] | rs 17663767[d] | rs 1901614[d] | Pharm GKB data base[e] | Pharm GKB data base[e] | |
| Marker allele | | | | *2 | | | *1a | | | *2T | *2 | *67G | *3 | | | | | |
| Haplotype[f] | Block | Segments | | | | | | | | | | | | | | | | |
| | *1 (*1*1) | *1a | *1b*1a | | | | | | | | | | | | | | | 0.311 |
| | | *1b | *1c*1a | | | | | | | | | | | | | | | 0.051 |
| | | *1c | *1a*1c | | | | | | | | | | | | | | | 0.015 0.390 |
| | | *1d | *1d*1a | | | | | | | | | | | | | | | 0.010 |
| | | *1e | *1b*1b | | | | | | | | | | | | | | | 0.003 |
| | *II (*2*1) | *IIa | *2a*1a | | | | | | | | | | | | | | | 0.492 |
| | | *IIb | *2a*1b | | | | | | | | | | | | | | | 0.084 0.584 |
| | | *IIc | *2b*1a | | | | | | | | | | | | | | | 0.008 |
| | *III (*2*2) | *IIIa | *2a*2a | | | | | | | | | | | | | | | 0.010 0.010 |
| | *IV (*1*3) | *IVa? | *1a*3a | | | | | | | | | | | | | | | 0.005 |
| | | *IVb | *1a*3b | | | | | | | | | | | | | | | 0.003 0.010 |
| | | *IVc | *1c*3a | | | | | | | | | | | | | | | 0.003 |
| | *V (*1*2T) | *Va | *1c*2T | | | | | | | | | | | | | | | 0.003 0.003 |
| | *VI (*1*67G) | *VIa | *1b*67G | | | | | | | | | | | | | | | 0.003 0.003 |
| SNP frequency | | | 0.008 | 0.594 | 0.003 | 0.054 | 0.026 | 0.010 | 0.594 | 0.003 | 0.010 | 0.003 | 0.010 | 0.026 | 0.094 | 0.026 | 0.026 | |

**Figure 2** SNPs and haplotypes of *UGT1A8* and *1A10* (*Block 8/10*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number in the National Center for Biotechnology Information. [e]The SNPs included in the PharmGKB database (http://www.pharmgkb.org/). [f]The haplotypes are described as numbers plus alphabetical letters. The prediction of rare haplotypes that were inferred from only one subject (frequency was 0.003) is sometimes inaccurate.

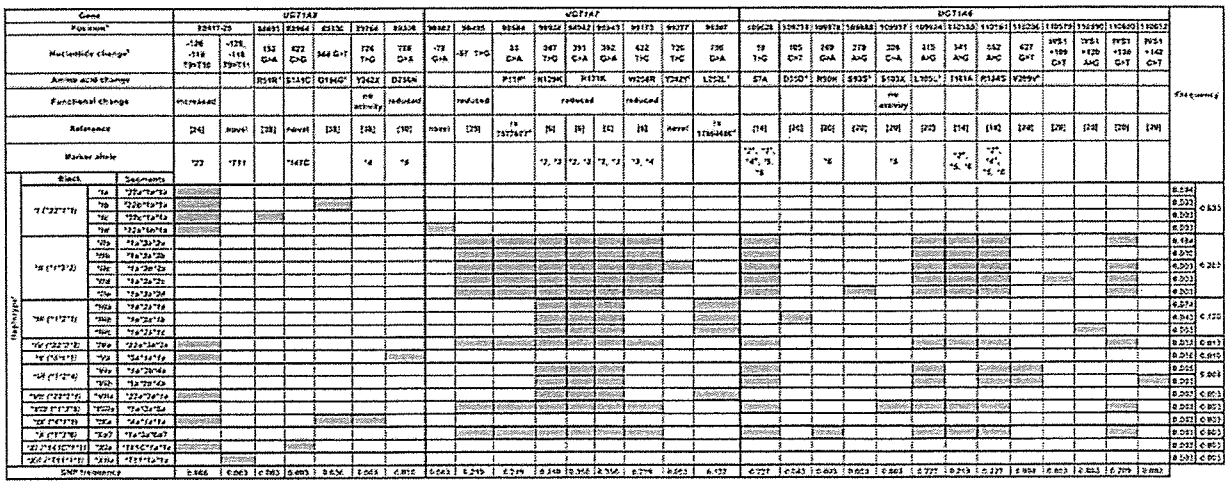

**Figure 3** SNPs and haplotypes of *UGT1A9*, *1A7*, and *1A6* (*Block 9/6*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number in the National Center for Biotechnology Information. [e]The haplotype numbering of *UGT1A6* basically followed the numbering by Nagar *et al.*[14] [f]The haplotypes are described as numbers plus alphabetical letters. The prediction of rare haplotypes that were inferred from only one subject (frequency was 0.003) is sometimes inaccurate.

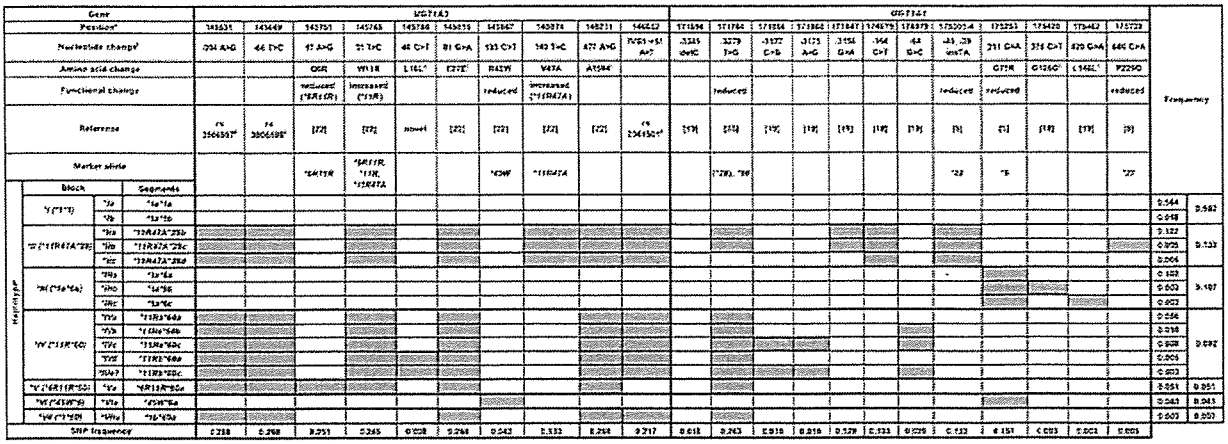

**Figure 4** SNPs and haplotypes of *UGT1A3* and *1A1* (*Block 3/1*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number in the National Center for Biotechnology Information. [e]The haplotypes are described as numbers plus alphabetical letters. The prediction of rare haplotypes that were inferred from only one subject (frequency was 0.003) is sometimes inaccurate.

17A>G (Q6R), 31T>C (W11R), 133C>T (R45W), and 140T>C (V47A), were detected at frequencies of 0.051, 0.265, 0.043, and 0.133, respectively (Figure 4).

*LD analysis*

Next, pairwise LD analysis was performed for the *UGT1A* genes using the polymorphisms detected in this study (*1A8*, *1A10*, *1A9*, *1A7*, and *1A3*) and previous studies (*1A6*, *1A4*, *1A1*, and common exons 2–5)[19–21] from the same 196 subjects, and the values for rho square ($r^2$), chi square ($\chi^2$), and $|D'|$ were obtained. Variations found in only one subject were excluded from the analysis. Since the data for $r^2$ and $\chi^2$ values were almost equivalent, only the data for $r^2$ is depicted in Figure 5. Several close linkages within each segment were seen between the variations of the *1A8*, *1A10*, *1A7*, and *1A3* gene segments, as seen in the *1A6*, *1A4*, and *1A1* segments and common exons that were previously analyzed. Furthermore, strong linkages were also observed across the segments, especially between *1A8* and *1A10*, among *1A9*, *1A7*, and *1A6*, and between *1A3* and *1A1*, where approximately 19, 35.5, 10, 11, 26, 10.5, and 29.5 kb separate *1A10* from *1A8*, *1A9* from *1A10*, *1A7* from *1A9*, *1A6* from *1A7*, *1A4* from *1A6*, *1A3* from *1A4*, and *1A1* from *1A3*, respectively. Thus, the region from *1A8* to *1A1* was divided into four LD blocks: *Block 8/10* (*1A8* and *1A10*), *Block 9/6* (*1A9*, *1A7*, and *1A6*), *Block 4* (*1A4*), and *Block 3/1* (*1A3* and *1A1*). The data from the $|D'|$ values supported this block partitioning (data not shown). In addition, a few exceptional strong linkages (over 0.7 for $r^2$) beyond the LD blocks were also observed. The perfect linkage ($r^2=1$) was detected between *1A10* 177G>A (M59I) and *1A9* 766G>A (D256N), and between *1A4* IVS1+101G>T and *1A1* 686C>A (P229Q). A strong LD ($r^2=0.95$) was also shown between *1A8* 711A>C (T237T) and *1A3* 17A>G (Q6R). In addition, *1A7* 756G>A (L252L) was strongly linked with *1A4* −219C>T, −163G>A, 142T>G (L48V), 448T>C (L150L), 804G>A (P268P), and IVS1+43C>T ($r^2=0.78$ or greater). The other linkages were less than 0.67 for $r^2$.

*Haplotype analysis*

Next, the haplotypes for *Block 8/10*, *Block 9/6*, and *Block 3/1* were analyzed. The haplotypes of *Block 4*, consisting of the *1A4* segment,[21] and *Block C*, covering common exons 2–5 (*Block 2* in a previous paper[19]), have already been reported. The block haplotypes were tentatively named with Roman numerals plus small alphabetical letters. The haplotypes were also estimated for each gene segment (segment haplotypes), where a group of haplotypes without amino-acid changes was defined as **1*.

As for *Block 8/10*, consisting of the two segments *1A8* and *1A10*, six haplotypes were first unambiguously assigned by the presence of homozygous SNPs at all sites (**Ia*, **Ib*, **IIa*, and **IIb*) or a heterozygous SNP at only one site (**IIc* and **IIIa*). We separately estimated the diplotype configuration (a combination of haplotypes) for each subject by LDSUPPORT software. The diplotype configurations of 184 subjects were inferred with a probability greater than 0.96, and the remaining diplotypes (12 subjects) had a probability greater than 0.54. The haplotype inferred in the diplotypes with low probabilities were described with a question mark in Figure 2 (i.e., **IVa?*). There were eight additionally inferred haplotypes. The block haplotypes were also described in the form of segment haplotype (*1A8* haplotype-*1A10* haplotype) combinations in Figure 2: in the *1A8* segment, the segment haplotype bearing the nonsynonymous A173G (**2* allele) was named **2*; in *1A10*, the haplotypes bearing M59I (**2* allele), T202I (**3* allele), A2T, and E67G were named **2*, **3*, **2T*, and **67G*, respectively. The most frequent block haplotype (segment haplotype combination in parenthesis) was **IIa* (**2a*1a*) (frequency: 0.492), followed by **Ia* (**1b*1a*) (0.311), **IIb* (**2a*1b*) (0.084), and **Ib* (**1c*1a*) (0.051). The frequencies of other block haplotypes were less than 0.05.


**Figure 5** Linkage disequilibrium (LD) analysis of *UGT1A* SNPs. Pairwise LD is expressed as $r^2$ (from 0 to 1) by a 10-graded blue color for the polymorphisms ($n=2$ or greater) in *1A8, 1A10, 1A9, 1A7, 1A6, 1A4, 1A3, 1A1*, and common exons 2–5 located in the same order on the chromosome. The denser color represents the higher linkage. *Block C*: the same as *Block 2* in the previous paper, and the nucleotide positions of polymorphisms in the exons 2–5 are numbered as in *UGT1A1*.[19] UTR: untranslated region.

It is noteworthy that the low-activity *1A10* haplotype **3* was completely linked with the *1A8*1* haplotype (*Block 8/10 *IV*).

Regarding *Block 9/6* (*1A9-1A7-1A6*), 14 haplotypes were first unambiguously assigned by homozygous SNPs at all sites (**Ia, *Ib, *IIa*, and **IIIa*) or a heterozygous SNP at only one site (**Ic, *Id, *IIc, *IId, *IIIb, *IVa, *Va, *VIIIa, *XIa*, and **XIIa*). Additionally, eight haplotypes (**IIb, *IIe, *IIIc, *VIa, *VIb, *VIIa, *IXa*, and **Xa?*) were inferred, and diplotype configurations of 191 subjects were inferred with a 1.00 probability by the software. The haplotype inferred in the diplotype with a low probability was **Xa?* (Figure 3). The block haplotypes were also described as combinations of segment haplotypes (*1A9* haplotype-*1A7* haplotype-*1A6* haplotype) in Figure 3: in the *1A9* segment, the segment haplotype bearing Y242X (**4* allele), D256N (**5* allele), −126_−118 T9>T10 (**22* allele), −126_−118 T9>T11, or S141C were named **4,*5, *22, *T11,* or **141C*, respectively; in the *1A7* segment, the haplotype bearing N129K/R131K (**2* allele) was named **2*, and the haplotype bearing N129K/R131K/W208R (**3* allele) was named **3*; in *1A6*, the haplotypes bearing S7A/T181A/R184S (**2* allele), S7A/R184S (**4* allele), S7A/S103X/T181A/R184S (**5* allele), and S7A/R90H/T181A/R184S (**6* allele) were named **2, *4, *5,*

and **6*, respectively, as described previously.[20] The most frequent haplotype of *Block 9/6* was **Ia* (**22a*1a*1a*) (0.594), followed by **IIa* (**1a*3a*2a*) (0.184), and **IIIa* (**1a*2a*1a*) (0.074) (Figure 3). The frequencies of the other haplotypes were under 0.05. Notably, most (97.6%) of the high-activity segment haplotype *1A9*22* was linked with *1A7*1* and *1A6*1* (*Block 9/6 *I*). The *1A7* low-activity haplotype **3* was mostly linked (97.7%) with *1A6*2* haplotype (**II* and **IVa* in Figure 3).

Regarding *Block 3/1* (*1A3-1A1*), six haplotypes were first unambiguously assigned by the presence of homozygous SNPs at all sites (**Ia*, **IIa*, **IIIa*, and **Va*) or a heterozygous SNP at only one site (**Ib* and **VIa*). The diplotype configurations of 188 subjects were inferred with a 1.00 probability. The additionally inferred haplotypes were **IIb*, **IIc*, **IIIb*, **IIIc*, **IVa–*IVe?*, and **VIIa*. The haplotype **IVe?* was inferred with a low probability (Figure 4). The combinations of segment haplotypes (*1A3* haplotype-*1A1* haplotype) were also described in Figure 4: in *1A3*, the group bearing the nonsynonymous variations Q6R/W11R, W11R, R45W, and W11R/V47A were named the **6R11R*, **11R*, **45W*, and **11R47A* haplotypes, respectively;[22] in *1A1*, the haplotypes bearing G71R (**6* allele), −40_−39 insTA (**28* allele with or without **60* allele), and −3279T>G (**60* allele without **28* allele) were named the **6*, **28*, and **60* haplotype groups as described previously.[19] The most frequent haplotype of *Block 3/1* was **Ia* (**1a*1a*) (frequency: 0.564), followed by **IIa* (**11R47A*28b*) (0.122), **IIIa* (**1a*6a*) (0.102), **IVa* (**11Ra*60a*) (0.056), and **Va* (**6R11R*60a*) (0.051). The frequencies of the other block haplotypes were less than 0.05. It is noteworthy that the high-activity segment haplotype *1A3*11R47A* was completely linked with the low-activity haplotype *1A1*28* (*Block 3/1 *II*). The low-activity haplotype *1A1*6* was mostly linked (71.3%) with the *1A3*1* haplotype (**III*). The high-activity *1A3*11R* haplotype was perfectly linked with the low-activity *1A1*60* haplotype (**IV*).

Finally, no statistically significant differences in haplotype frequencies were found between the subjects with the different disease types in *Block 8/10*, *Block 9/6*, and *Block3/1* ($P \geqslant 0.05$ by $\chi^2$ test or Fisher's exact test).

*Genotyping and haplotype analysis across the LD blocks*

A typing method was developed and additional 105 Japanese subjects (16 arrhythmic patients and 89 cancer patients) were genotyped, where direct sequencing (for nine polymorphisms in the *1A9* 5′-flanking region and *1A4*) and pyrosequencing (for the rest of the polymorphisms) were used for detection of 41 polymorphisms (see Materials and methods and the Table 1 legend) with (potentially) functional importance. The frequencies from 301 subjects in total are described in Table 1. Again, all the allele frequencies were in Hardy–Weinberg equilibrium, and statistically significant differences were not observed in any of the allelic frequencies between the two disease types ($P \geqslant 0.05$ by $\chi^2$ test or Fisher's exact test). Almost the same LD map as in Figure 5 was obtained between the 41 tagged variations (data not shown), indicating the robustness of our block partitioning. It is noteworthy that the known variation *1A1* 1456T>G (Y486D, **7* allele) was newly found in one subject (frequency: 0.002).

Several reports have shown that some polymorphisms in *1A9*, *1A7*, *1A6*, and *1A1* were closely linked,[17,18] and we also observed several weak linkages beyond the LD blocks (see Figure 5). Therefore, the block-haplotype combinations (whole complex haplotypes) were analyzed among *Block 8/10*, *Block 9/6*, *Block 4*, *Block 3/1*, and *Block C* (common exons 2–5) by LDSUPPORT software utilizing the polymorphisms. In *1A4* (*Block 4*), the haplotypes bearing L48V (**3* allele) and R11W (**4* allele) were named **3* and **4*, respectively, as described previously.[21] Polymorphisms found at a frequency less than 0.010, and subjects with these polymorphisms were excluded in this analysis. When *Block 8/10* or *Block C* was included in the analysis, the whole-complex haplotypes were highly complicated (data not shown). However, if *Block 8/10* and *Block C* were excluded, the diplotype configurations of 278 subjects were inferred with a probability greater than 0.91 (mostly>0.95) using the 18 tagged polymorphisms (see Figure 6 legend for polymorphisms). The haplotypes covering *Block 9/6*, *Block 4*, and *Block 3/1* are summarized in Figure 6. Again, we did not find any statistically significant differences in frequencies of haplotypes covering *Block 9/6*, *Block 4* and *Block 3/1* between the subjects with the different disease types ($P \geqslant 0.05$ by $\chi^2$ test or Fisher's exact test). The region from *1A9* to *1A1* is approximately 90 kb length. Since the 18 variations were used for haplotyping, the number of inferred haplotype combinations (only 26) is unexpectedly small compared to the theoretical ones (Figure 6).

Several functionally important linkages were found across the blocks. *Block 9/6 *VI* (*1A9*1-1A7*2-1A6*4*) and *Block 3/1 *IIb* (*1A3*11R47A-1A1*28c* containing the **60*, **28*, and **27* alleles) were perfectly linked (6/6 cases). Most of the *1A1*6*-containing haplotypes (*Block 3/1 *III* and **VI*) (69/85 cases) were associated with *Block 4* (*1A4*) **1* and *Block 9/6 *II* (harboring *1A7*3* and *1A6*2*). The *1A1*60*-harboring haplotypes (*Block 3/1 *IV* and **V*) were very closely linked with *Block 9/6 *III* (harboring *1A7*2*) and *Block 4 *3* (59/71 cases of *1A1*60*-harboring haplotypes). Most of *Block 3/1 *VI* (*1A3*45W-1A1*6*) (25/26 cases) was associated with *Block 9/6 *II* (*1A9*1-1A7*3-1A6*2*), and *Block 4 *4* was perfectly linked (4/4 cases) with both *Block 3/1 *VI* and *Block 9/6 *II*.

In addition, we found that *Block 8/10 *IV* (containing the low-activity allele *1A10*3* (T202I)) was strongly linked with *Block 9/6 *III* (*1A9*1-1A7*2-1A6*1*), *1A4*3*, and *Block 3/1 *IV* (*1A3*11R-1A1*60*) (4/5 cases of *Block 8/10 *IV*, data not shown). *Block 3/1 *V* (harboring *1A3*6R11R* and *1A1*60a*) was perfectly linked with *Block C *IB* (25/25 cases of *Block 3/1 *V*, data not shown).

## Discussion

Previously, we have reported the genetic variations of *UGT1A6*, *1A4*, *1A1* segments and common exons 2–5 found in 196 Japanese subjects.[19–21] In this study, we first directly

Table 1 Frequencies of variations in the *UGT1A* gene complex detected in 301 Japanese subjects

| *Location* | *Nucleotide change* | | *Amino-acid change* | *Number of subjects* | | | *Frequency* |
|---|---|---|---|---|---|---|---|
| | | | | *Wild type* | *Heterozygote* | *Homozygote* | |
| 1A8 Ex1 | 518[a] | C>G | A173G | 48 | 145 | 108 | 0.600 |
| 1A10 Ex1 | 4[a] | G>A | A2T | 300 | 1 | 0 | 0.002 |
| | 177[a] | G>A | M59I | 297 | 4 | 0 | 0.007 |
| | 200[a] | A>G | E67G | 300 | 1 | 0 | 0.002 |
| | 605[a] | C>T | T202I | 295 | 6 | 0 | 0.010 |
| *1A9* 5′-Flank | −126 to −118 | T9>T10 | | 34 | 130 | 136 | 0.668 |
| | | T9>T11 | | | 1 | 0 | 0.002 |
| 1A9 Ex1 | 422[a] | C>G | S141C | 300 | 1 | 0 | 0.002 |
| | 726[a] | T>G | Y242X | 300 | 1 | 0 | 0.002 |
| | 766[a] | G>A | D256N | 297 | 4 | 0 | 0.007 |
| 1A7 Ex1 | 387[a] | T>G | N129K | 128 | 135 | 38 | 0.350 |
| | 391[a] | C>A | | 128 | 135 | 38 | 0.350 |
| | 392[a] | G>A | R131K | 128 | 135 | 38 | 0.350 |
| | 622[a] | T>C | W208R | 186 | 101 | 14 | 0.214 |
| 1A6 Ex1 | 19[a] | T>G | S7A | 180 | 106 | 15 | 0.226 |
| | 269[a] | G>A | R90H | 300 | 1 | 0 | 0.002 |
| | 308[a] | C>A | S103X | 300 | 1 | 0 | 0.002 |
| | 541[a] | A>G | T181A | 186 | 101 | 14 | 0.214 |
| | 552[a] | A>C | R184S | 180 | 106 | 15 | 0.226 |
| 1A4 Ex1 | 31 | C>T | R11W | 295 | 6 | 0 | 0.010 |
| | 127 | delA | 43fsX22[b] | 300 | 1 | 0 | 0.002 |
| | 142 | T>G | L48V | 229 | 66 | 6 | 0.130 |
| | 175 | delG | 59fsX6[b] | 299 | 2 | 0 | 0.003 |
| | 271 | C>T | R91C | 300 | 1 | 0 | 0.002 |
| | 325 | A>G | R109G | 299 | 2 | 0 | 0.003 |
| *1A4* 3′-Flank | IVS+1 | G>T | | 300 | 1 | 0 | 0.002 |
| 1A3 Ex1 | 17[a] | A>G | Q6R | 276 | 24 | 1 | 0.043 |
| | 31[a] | T>C | W11R | 167 | 111 | 23 | 0.261 |
| | 133[a] | C>T | R45W | 274 | 26 | 1 | 0.047 |
| | 140[a] | T>C | V47A | 228 | 69 | 4 | 0.128 |
| *1A1* 5′-Flank | −3279[a] | T>G | | 167 | 110 | 24 | 0.262 |
| | −40 to −39[a] | insTA | | 227 | 70 | 4 | 0.130 |
| 1A1 Ex1 | 211[a] | G>A | G71R | 217 | 76 | 8 | 0.153 |
| | 247[a] | T>C | F83L | 301 | 0 | 0 | 0.000 |
| | 686[a] | C>A | P229Q | 295 | 6 | 0 | 0.010 |
| 1A Ex4 | 1091[a,c] | C>T | P364L | 298 | 3 | 0 | 0.005 |
| *1A* Ex5 | 1456[a,c] | T>G | Y486D | 300 | 1 | 0 | 0.002 |
| | 1598[a,c] | A>C | H533P | 300 | 1 | 0 | 0.002 |
| *1A* 3′-UTR | 1813[a,c] | C>T | | 236 | 63 | 2 | 0.111 |
| | 1941[a,c] | C>G | | 236 | 63 | 2 | 0.111 |
| | 2042[a,c] | C>G | | 236 | 63 | 2 | 0.111 |

[a]In all, 105 subjects were genotyped by pyrosequencing.
[b]43fsX22 represents frameshift from codon 43 resulting in the termination at the 22nd codon, codon 65. The same meaning for 59fsX6.
[c]The positions in *UGT1A1* were used.

| Block 9/6 | | | | Block 4 | Block 3/1 | | | Number of combination | Frequency |
|---|---|---|---|---|---|---|---|---|---|
| *UGT1A9* | *UGT1A7* | *UGT1A6* | | *UGT1A4* | *UGT1A3* | *UGT1A1* | | | |
| *22 | *1 | *1 | *I | *1 | *1 | *1 | *I | 323 | 0.581 |
| *1 | *3 | *2 | *II | *1 | *1 | *1 | *I | 3 | 0.005 |
| *1 | *2 | *1 | *III | *1 | *1 | *1 | *I | 3 | 0.005 |
| *22 | *1 | *1 | *I | *3 | *1 | *1 | *I | 1 | 0.002 |
| *22 | *3 | *2 | *IV | *1 | *1 | *1 | *I | 1 | 0.002 |
| *1 | *3 | *2 | *II | *1 | *1 | *6 | *III | 48 | 0.086 |
| *22 | *3 | *2 | *IV | *1 | *1 | *6 | *III | 7 | 0.013 |
| *22 | *1 | *1 | *I | *1 | *1 | *6 | *III | 3 | 0.005 |
| *1 | *2 | *1 | *III | *1 | *1 | *6 | *III | 1 | 0.002 |
| *1 | *3 | *2 | *II | *1 | *45W | *6 | *VI | 21 | 0.038 |
| *1 | *3 | *2 | *II | *4 | *45W | *6 | *VI | 4 | 0.007 |
| *22 | *1 | *1 | *I | *1 | *45W | *6 | *VI | 1 | 0.002 |
| *22 | *1 | *1 | *I | *1 | *11R47A | *28b | *IIa | 30 | 0.054 |
| *1 | *3 | *2 | *II | *1 | *11R47A | *28b | *IIa | 30 | 0.054 |
| *1 | *3 | *2 | *II | *3 | *11R47A | *28b | *IIa | 1 | 0.002 |
| *1 | *2 | *4 | *VI | *1 | *11R47A | *28c | *IIb | 6 | 0.011 |
| *22 | *1 | *1 | *I | *1 | *11R47A | *28d | *IIc | 1 | 0.002 |
| *1 | *3 | *2 | *II | *1 | *11R47A | *28d | *IIc | 1 | 0.002 |
| *1 | *2 | *1 | *III | *3 | *11R | *60 | *IV | 36 | 0.065 |
| *22 | *1 | *1 | *I | *3 | *11R | *60 | *IV | 3 | 0.005 |
| *1 | *3 | *2 | *II | *3 | *11R | *60 | *IV | 3 | 0.005 |
| *1 | *2 | *1 | *III | *1 | *11R | *60 | *IV | 3 | 0.005 |
| *22 | *1 | *1 | *I | *1 | *11R | *60 | *IV | 1 | 0.002 |
| *1 | *2 | *1 | *III | *3 | *6R11R | *60 | *V | 23 | 0.041 |
| *1 | *2 | *1 | *III | *1 | *6R11R | *60 | *V | 1 | 0.002 |
| *22 | *2 | *1 | *VII | *3 | *6R11R | *60 | *V | 1 | 0.002 |
| | | | | | | | | 556 | 1.000 |

**Figure 6** Combinations of *Block 9/6*, *Block 4*, and *Block 3/1* haplotypes in a Japanese population. The used variations were *UGT1A9* −126_−118 T9>T10, *1A7* 387T>G, 391C>A, 392G>A and 622T>C, *1A6* 19T>G, 541A>G and 552A>C, *1A4* 31C>T and 142T>G, *1A3* 17A>G, 31T>C, 133C>T and 140T>C, and *1A1* −3279T>G, −40_−39 insTA, 211G>A and 686C>A. In *1A4* (*Block 4*), the haplotypes bearing L48V (*3 allele), and R11W (*4 allele) were named *3 and *4, respectively.

sequenced *1A8*, *1A10*, *1A9*, *1A7*, and *1A3* using genomic DNA from the same Japanese subjects and detected 7, 8, 7, 9, and 10 genetic polymorphisms, respectively (Figures 2–4). Two and one novel nonsynonymous SNPs were found in *1A10* (4G>A, A2T; 200A>G, E67G) and *1A9* (422C>G, S141C), respectively. As for 1A9 S141C, our preliminary results have shown that this amino-acid substitution reduces the enzymatic activity against 7-hydroxy-4-trifluoromethylcoumarin *in vitro* (Jinno *et al.*, unpublished data). Since the guanine base at position +4 is important for translation initiation,[23] *1A10* 4G>A might decrease the translation rate. Moreover, the luciferase-reporter activity of *1A9* −126_−118 T10 (*1A9*22* allele) was reported to increase 2.6-fold as compared to that of *1A9* −126_−118 T9.[24] Therefore, the novel variation *1A9* −126_−118 T9>T11 may also affect transcriptional activity. Further studies are needed to ascertain these possibilities. Recently, *1A7* −57G was reported to reduce the luciferase activity by 70% of the wild-type −57T.[25] While this SNP is linked with either *1A7*3* (129K/131K/208R) or **4* (208R) in Germans, our study showed that −57G was completely linked with *1A7*3* due to the absence of *1A7*4* in Japanese.

For the *1A8* alleles, only **1* and **2* were detected. Our segment haplotypes **1a*, **1b*, and **2a* correspond to alleles **1a*, **1*, and **2*, respectively, in a previous study on Americans.[8] The frequencies obtained in the United States,[8] 0.282, 0.551, and 0.145, for **1a*, **1*, and **2*, respectively, are different from those obtained in this study, 0.023, 0.316, and 0.587 for **1a*, **1b*, and **2a*, respectively. The allele frequency of *1A9* −126_−118 T9>T10 (**22* allele) in our data (0.666) was similar to that reported previously in Japanese (0.60), but higher than those in Caucasians (0.39) and African-Americans (0.44).[24] For *1A7*, the frequencies of **1*, **2* (129K/131K), and **3* (129K/131K/208R) haplotypes were 0.651, 0.130, and 0.219, respectively. Our data are comparable to the previous data for a Japanese population,[26] but not to those on Caucasians (0.355, 0.280, and 0.365) and Egyptians (0.420, 0.200, and 0.380).[17] As for *1A3*, the frequencies of the haplotypes **1*, **11R*, **6R11R*, **11R47A*, and **45W* were 0.692, 0.082, 0.051, 0.133, and 0.043, respectively. These are similar to the previous data obtained from the Japanese.[22]

Recently, linkages among the SNPs in *1A9*, *1A7*, *1A6*, and *1A1* have been reported in Americans.[18] By our LD analysis, strong linkages were shown between the SNPs in *1A8* and *1A10*, among those in *1A9*, *1A7*, and *1A6*, and also between those in *1A3* and *1A1*. Moreover, this is the first report on the haplotype analysis using high-density SNPs for the entire *UGT1A* complex. By block haplotyping, several close linkages between the segmental haplotypes were observed: between the *1A8*1* and *1A10*3* haplotypes in *Block 8/10*; between the *1A7*3* and *1A6*2* in *Block 9/6*; between the *1A3*11R47A* and *1A1*28* and between the *1A3*11R* and *1A1*60* haplotypes in *Block 3/1*. Carlini *et al.*[18] reported that

*1A7* low-activity alleles (*1A7*2* and **3*) were perfectly linked to *1A9*1* in Americans (including Caucasians (83%) and African-American (14%)). Also in this study, most (95.6%) of the *1A7*2* or **3* alleles were linked to the *1A9*1* allele in Japanese (Figure 3).

We conducted additional typing of 105 subjects by pyrosequencing and direct sequencing, and confirmed the presence of several functionally important haplotype combinations beyond the blocks (Figure 6). In Americans (including Caucasians (83%) and African-American (14%)),[18] Caucasians, and Egyptians,[17] 75, 78, and 57%, respectively, of *1A7*3* were associated with the *1A1*28* allele, though only the **28* allele was genotyped in *1A1* in these analysis. In our more intensive analysis, most of the *1A7*3* haplotype was associated with either the *1A1*28b* haplotype (having *1A1*60* and **28* alleles) (26.1% of the *1A7*3* haplotype) or the *1A1*6* haplotype (67.2% of *UGT1A7*3*). Thus, different profiles for the linkage of *1A7*3* with the *1A1* polymorphisms between the Caucasians and Japanese reflect the facts that the frequency of the *1A1*6* haplotype in the Asian populations was relatively high, and that the *1A1*28* and **6* alleles were mutually exclusive.[19] In fact, linkage between *1A1*6* and *1A7*3* alleles was recently suggested in Taiwanese.[27] Innocenti *et al.* reported the three most common *1A9-1A1* haplotype combinations were *1A9*22-1A1*1* (36.4%), *1A9*1-1A1*28b* (28.0%), and *1A9*1-1A1*1* (18.6%) for Caucasians, and *1A9*22-1A1*1* (45.3%), *1A9*1-1A1*60* (22.3%), and *1A9*1-1A1*6* (12.7%) for Asians.[28] In this study for Japanese, *1A9*22-1A1*1*, *1A9*1-1A1*60*, and *1A9*1-1A1*6* (58.5, 11.9, and 13.3%, respectively) were also the most common three combinations. Furthermore, we revealed that most (98.2%) of the *1A1*1* haplotype was linked with *1A9*22*, and 87.1% of *1A1*6*, 100% of *1A1*28c*, and 93.0% of *1A1*60* were associated with *1A9*1*. Collectively, haplotype combinations are suggested to be different between Caucasians and Asians. In addition, several interesting linkages were found between the segmental haplotypes as shown in the Results. For example, the segment haplotypes *1A6*4* (S7A/R184S) and *1A1*28c* were strongly linked in Japanese subjects.

These linkages might be crucial for the metabolism of a certain drug for which two or more UGT1A isoforms significantly contribute to its metabolism. In fact, multiple UGT isoforms are involved in glucuronidation of several compounds, for example SN-38,[7,29] estrogens and their metabolites (estron, estradiol, 2-hydroxyestrone, and others),[30,31] and arachidonic acid and its metabolites.[32,33] UGT1A1, 1A9, and 1A7 play important roles in SN-38 glucuronidation.[7,29] The *1A1*60*, **28b*, and **6* haplotypes are associated with reduced UGT1A1 activity to SN-38.[15,19,34,35] Since the *1A9* high-activity (high transcription) haplotype **22* was dominant in Japanese (0.666), *1A9*1* can be considered (relative to **22*) as a low-activity haplotype. The *1A7*3*, but not **2*, haplotype has a reduced glucuronidation activity (by 59%) to SN-38.[7] A more recent report has shown that UGT1A10 is also responsible for SN-38 glucuronidation,[36] and that the *1A10 *3* (T202I) is a low-activity allele.[11] We found that the *Block 8/10 *IV* haplotype (harboring *1A10*3*) was closely linked with *Block 9/6 *III* (harboring *1A9*1*) and *Block 3/1 *IV* (harboring *1A1*60*). Furthermore, most of *Block 9/6 *II* (harboring *1A9*1* and *1A7*3*) were estimated to be linked with *Block 3/1 *III* or **VI* (having *1A1*6*), or *Block 3/1 *IIa* (having *1A1*28b*). Though the functional significance of 1A10 T202I toward SN-38 is currently unknown, it is possible that the concurrently reduced activities of UGT1A10, 1A9, 1A7, and 1A1 may influence SN-38 glucuronidation.

Arachidonic acid and its metabolites prostaglandins were conjugated with UGT1A1, 1A3, 1A9, 1A10, and 2B7.[33] UGT1A1, 1A3, and 1A4 also had catalytic activities toward a hydroxylated metabolite of arachidonic acid, 12- and 15-hydroxyeicosatetraenoic acid.[32] Furthermore, glucuronidation of leucotriene B4, another arachidonic acid metabolite that mediates the inflammation process, can be catalyzed by UGT1A1, 1A3, 1A8, and 2B7.[32] Thus, co-occurrence of the functionally less active haplotypes (such as *Block 9/6 *II* (including *1A9*1*)-*Block 3/1 *VI* (harboring *1A3*45W* and *1A1*6*)), might cooperatively influence the metabolism of several important compounds in the arachidonic acid cascade.

Since plural UGT isoforms are often involved in the glucuronidation of 'one' compound, co-occurrence of the functionally less active haplotypes in the entire *UGT1A* gene complex needs to be carefully considered in studies on the association of genetic polymorphisms with pharmacokinetic parameters and clinical and epidemiologic data. Our findings would provide fundamental and useful information for genotyping or haplotyping of *UGT1As* in the Japanese and probably other Asian populations.

## Materials and methods

### *Human genomic DNA samples*

The 301 Japanese subjects consisted of 124 arrhythmic patients, who were administered β-blockers, and 177 cancer patients, who were administered irinotecan. Genomic DNA was extracted directly from blood leukocytes. The ethical review boards of the National Cancer Center, the National Cardiovascular Center, and the National Institute of Health Sciences approved this study. Written informed consent was obtained from all patients.

### *Polymerase chain reaction (PCR) conditions for DNA sequencing*

First, the fragments were amplified from genomic DNA (150 ng) using 2.5 U of Z-Taq (Takara Bio Inc., Shiga, Japan) with 0.2 μM primers (see '1st amplification' in Table 2 for primer sequences). The exon-1's of *UGT1A8* and *1A10* were simultaneously amplified by mixed primers for each gene, and those of *1A9* and *1A7* were amplified as one fragment. The primer sequences for the 1st amplification of *1A10* were described previously.[37] The first PCR conditions consisted of 30 cycles of 98°C for 5 s, 55°C for 5 s, and 72°C for 190 s. Then, each exon 1 was amplified by Ex-Taq (0.625 U) (Takara Bio Inc.) using the first PCR products as templates with the 2nd amplification primers (0.2 μM) that were designed in the introns (see '2nd amplification' in Table 2 for primer