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

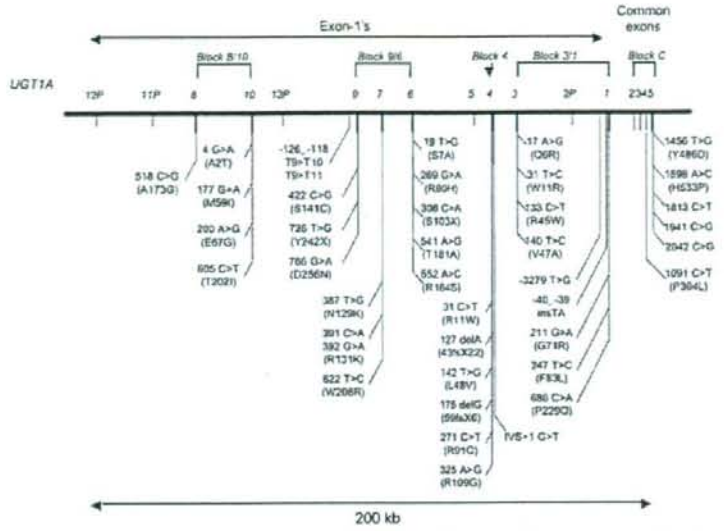

Figure 1 The organization of the human *UGT1A* gene complex and the polymorphisms used for genotyping. The human *UGT1A* gene complex (nine active and four inactive exon-1's, and common exons 2–5) spans approximately 200 kb. Four pseudogenes (*1A2p*, *1A11p*, *1A12p*, and *1A13p*), and *1A5*, which mRNA 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 χ^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,

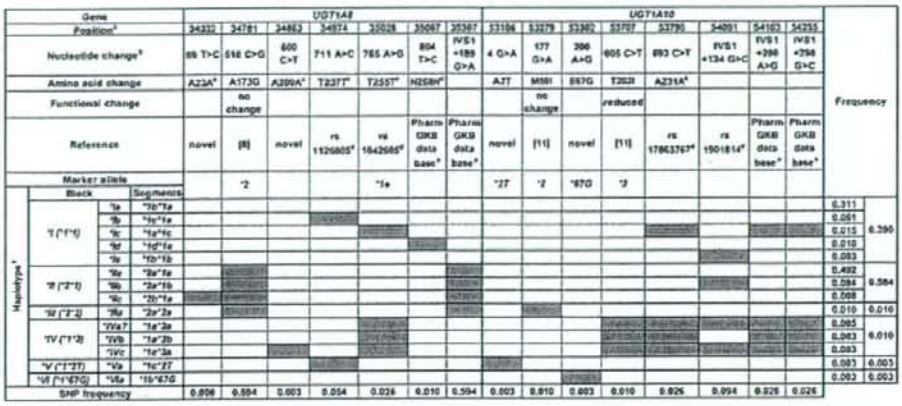

Figure 2 SNPs and haplotypes of *UGT1A8* and *1A10* (*Block 8/10*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number in the 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.

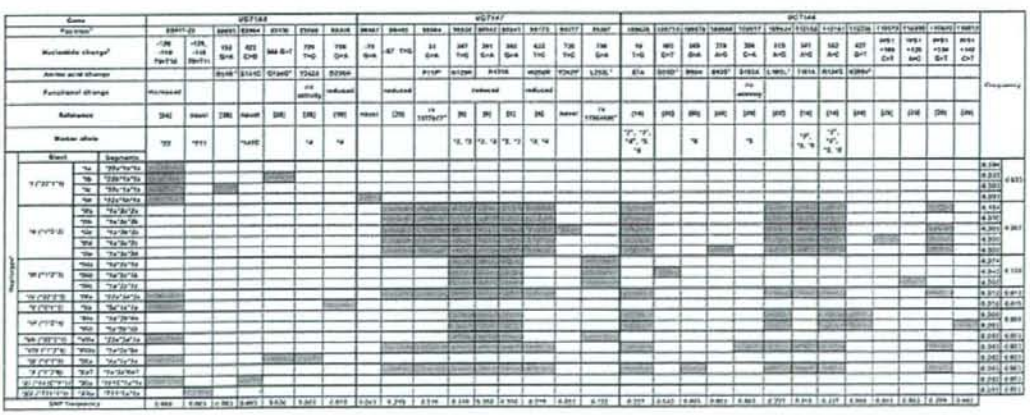

Figure 3 SNPs and haplotypes of *UGT1A9*, *1A7*, and *1A6* (*Block 9/6*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number 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.

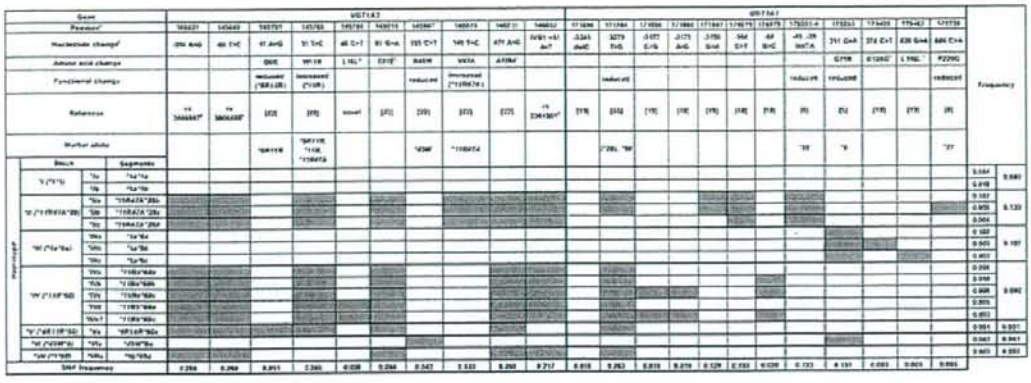

Figure 4 SNPs and haplotypes of *UGT1A3* and *1A1* (*Block 3/1*) in a Japanese population. [a]Position on AF297093.1. [b]A of the initiation codon in each gene segment is numbered 1. For intron SNPs, their positions were numbered from the nearest exon. [c]No amino-acid change. [d]dbSNP number in the 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 (χ^2), and $|D'|$ were obtained. Variations found in only one subject were excluded from the analysis. Since the data for r^2 and χ^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.

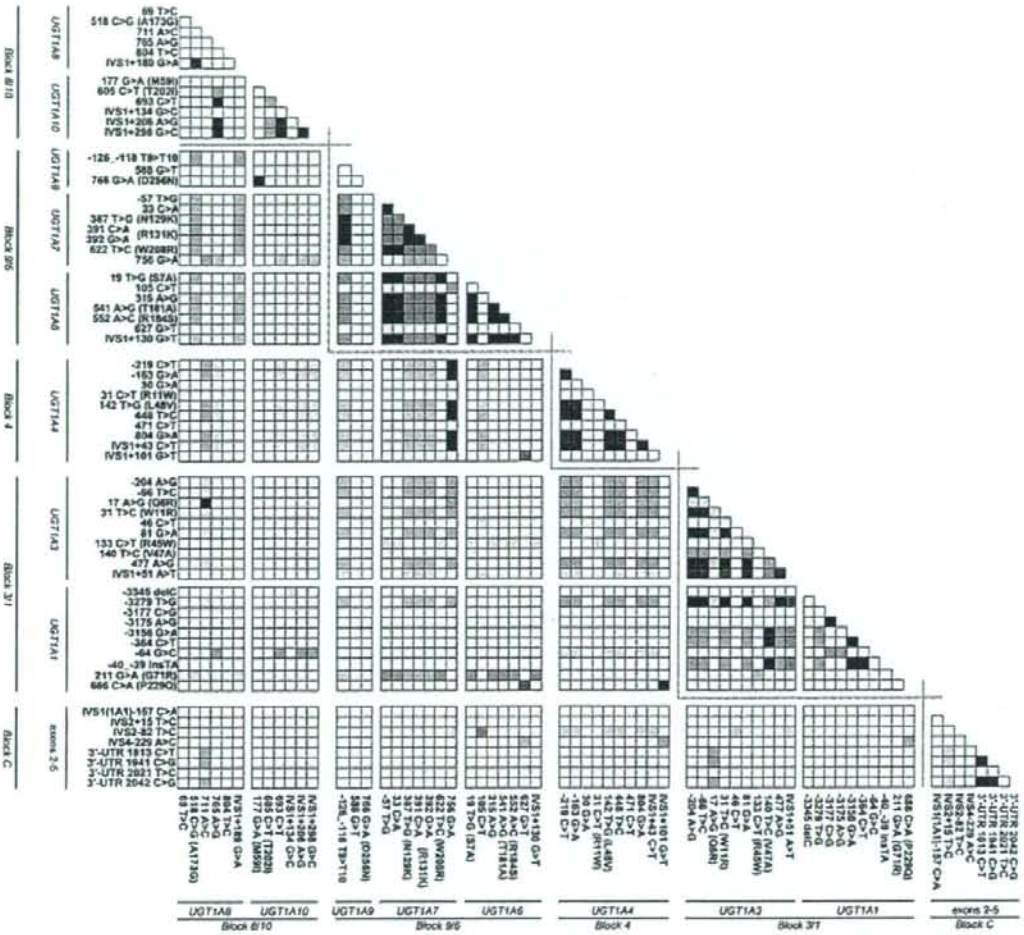

Figure 5 Linkage disequilibrium (LD) analysis of *UGT1A* SNPs. Pairwise LD is expressed as r^2 (from 0 to 1) by a 10-graded blue color for the polymorphisms ($n=2$ or greater) in *1A8, 1A10, 1A9, 1A7, 1A6, 1A4, 1A3, 1A1*, and common exons 2–5 located in the same order on the chromosome. The denser color 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 χ^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 χ^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 χ^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

Table 2 Primers for amplification and sequencing of exon-1's in *UGT1A8, 1A7, 1A3,* and 5'-flanking region of *1A9*

	Direction	Gene	Primer name	Sequences
1st amplification	Forward	*1A8*	UGT1A8ZF	5'-GTGGCTGGTACTCATTTTTCC-3'
	Reverse		UGT1A8ZR	5'-CTTCCAAACCCAACATCTCTAA-3'
	Forward	*1A9-7*	UGT1A9-7ZF	5'-TCTTGATTGTCCTCCATTGAGT-3'
	Reverse		UGT1A9-7ZR	5'-ACCAAGCAACCATACTCATAGG-3'
2nd amplification	Forward	*1A8*	UGT1A8-1stF	5'-AGAATGTGGAAGTAGAGCGG-3'
	Reverse		UGT1A8-1stR	5'-TTAGCAAAAAAGGAAAGTTCAA-3'
	Forward	*1A9* 5'-Flank	UGT1A9pro-248F1st	5'-TTGAGACAGAGTCGTGCTGTT-3'
	Reverse		UGT1A9pro-608R1st	5'-GCAAAGCCACAGGTCAGC-3'
	Forward	*1A7*	UGT1A71stF	5'-AAATGAATGAATAAGTACACGCC-3'
	Reverse		UGT1A71stR	5'-GGAAGTTTCATTTCTTACTGTGG-3'
	Forward	*1A3*	UGT1A3-1stF	5'-GTGAGCACAGGGTCAGACGTGT-3'
	Reverse		UGT1A3-1stR	5'-TTACAAACATTCGTGTCTACTT-3'
Sequencing	Forward	*1A8* Exon1	UGT1A8Ex1Fseq1	5'-TATGACAGGATAAATACACGCC-3'
	Forward	*1A8* Exon1	UGT1A8Ex1Fseq3	5'-ACTCAACCTCATACACTCTGGAG-3'
	Forward	*1A8* Exon1	UGT1A8Ex1Fseq2	5'-TGCTCCTCTTTCCTATGTCC-3'
	Reverse	*1A8* Exon1	UGT1A8Ex1Rseq1	5'-AACTTCGTACTTGTGCTTTCCA-3'
	Reverse	*1A8* Exon1	UGT1A8Ex1Rseq2	5'-ACTGGCAAAATAAATGTTCCTC-3'
	Reverse	*1A8* Exon1	UGT1A8Ex1Rseq3	5'-GCAACAAATGAAAATGTCAAATC-3'
	Forward	*1A9* 5'-Flank	UGT1A9pro-275Fseq	5'-GCTCTCGCAAGGATTGGG-3'
	Reverse	*1A9* 5'-Flank	UGT1A9pro-275Rseq	5'-CTTATGGTCTTTGCCTTGGG-3'
	Forward	*1A7* Exon1	UGT1A7F3-2	5'-TTTGAGGGCAGGTTCTATCTG-3'
	Reverse	*1A7* Exon1	UT1A71R3	5'-CAAAAACCATGAACTCCCGG-3'
	Forward	*1A7* Exon1	UT1A71F4	5'-TGGCAACTGGGAAGATCAC-3'
	Reverse	*1A7* Exon1	UT1A71R4	5'-GGACATAGGAAAGAGGAGCAG-3'
	Forward	*1A7* Exon1	UT1A71F5	5'-CTCCCTCCCCTCTGTGGTC-3'
	Reverse	*1A7* Exon1	UT1A71R5	5'-GCACTGGCTTTCCCTGATGAC-3'
	Forward	*1A7* Exon1	UGT1A7F6-2	5'-GAGGAACATTTATTTTGCCC-3'
	Reverse	*1A7* Exon1	UT1A71R6	5'-TACATATCAACAAGAGCTGC-3'
	Forward	*1A3* Exon1	UGT1A3seqF1	5'-GTGTTTTTCAAGATAGTC-3'
	Reverse	*1A3* Exon1	UGT1A3seqR1	5'-GCACATGGCGATCAAATTC-3'
	Forward	*1A3* Exon1	UGT1A3seqF2	5'-AGGCAGTGGTCCTCACCCCAGA-3'
	Reverse	*1A3* Exon1	UGT1A3seqR2	5'-AAGCATGGCAAATGTAGGACAGG-3'
	Forward	*1A3* Exon1	UGT1A3seqF3	5'-CCCTTCCTCCTATATTCCTAGA-3'
	Reverse	*1A3* Exon1	UGT1A3-1stR	5'-TTACAAACATTCGTGTCTACTT-3'

sequences). The PCR primers for the 2nd amplification of *1A10* and *1A9* were described previously.[37,38] Exon 1 in *1A3* and the promoter region of *1A9* were first directly amplified from genomic DNA (100 ng) using Ex-Taq as in the 2nd round of PCR described below. The second round of PCR consisted of one cycle at 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 55°C for 1 min, and 72°C for 2 min, and then a final extension for 7 min at 72°C. These PCR products were then treated with a PCR Product Pre-Sequencing Kit (USB Co., Cleveland, OH, USA) and directly sequenced on both strands using an ABI BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and the primers listed in Table 2 (see 'Sequencing'). The excess dye was removed by a DyeEx96 kit (Qiagen, Hilden, Germany), and the eluates were analyzed on an ABI Prism 3700 DNA Analyzer (Applied Biosystems). All the SNPs were confirmed by repeating the PCR on genomic DNA and sequencing the newly generated PCR products.

Genotyping (pyrosequencing)

Genotyping was performed by pyrosequencing, except for *UGT1A4* 31C>T (R11W), 127delA (43fsX22: frameshift from codon 43 resulting in the termination at the 22nd codon, codon 65), 142T>G (L48V), 175delG (59fsX6), 271C>T (R91C), 325A>G (R109G), and IVS+1G>T, and *1A9* −126_−118 T9>T10 or T11, which were genotyped by direct sequencing because these polymorphisms were not clearly determined by pyrosequencing. Fragments were directly amplified from genomic DNA (10–15 ng) by Ex-Taq (1 U) with amplification primer pairs (either primer was biotinylated) (Table 3). The PCR conditions consisted of 1 cycle at 94°C for 5 min, followed by 50 cycles of 94°C for 30 s, 55°C for 30 s (except for *UGT1A* 1598A>C (H533P), in which annealing was carried out at 58°C for 45 s), and 72°C for 30 s. Primers for *1A1* −3279T>G, −40_−39 insTA, 211G>A (G71R), 247T>C (F83L), 686C>A (P229Q), and 1456T>G (Y486D) in common exon 5 were described

Table 3 Primers in pyrosequencing

	Gene	Detected variation	Amino-acid change	Primer name	Direction	Length	Sequences
Amplification	*UGT1A8*	518C>G	A173G	UGT1A8PyroF	Forward	387 bp	5′-bTACGAAGTTTGTTTTCTCTATTTCT-3′
				UGT1A8PyroR2b[a]	Reverse		5′-CTATTTCTAGGGCATTTTTGG-3′
	UGT1A10	4G>A, 177G>A, 200A>G	A2T, M59I, E67G	UGT1A10Pyro1F2-2	Forward	378 bp	5′-GGTCAGGTTTTGTGCCTGTAC-3′
				UGT1A10Pyro1R2b[a]	Reverse		5′-bATTGAGCATGGGCGAAAA-3′
	UGT1A10	605C>T	T202I	UGT1A10Pyro2F	Forward	123 bp	5′-CTTTCCTATGTCCCCAATGA-3′
				UGT1A10Pyro2Rb[a]	Reverse		5′-bGGCATTTCTAAAAAGATACTGG-3′
	UGT1A9	422C>G, 726T>G, 766G>A	S141C, Y242X, D256N	UGT1A9PyroF	Forward	589 bp	5′-TATTAATGGGTTCATACAATGACA-3′
				UGT1A9PyroRb[a]	Reverse		5′-bTTTTTTAATTTCCAAAGGTGAAG-3′
	UGT1A7	387T>G, 391C>A, 392G>A, 622T>C	N129K, R131K, W208R	UGT1A7PyroF	Forward	364 bp	5′-TATTAACAAGTTCATCCAATGGTA-3′
				UGT1A7PyroRb[a]	Reverse		5′-bTCTAAGACATTTTTGAAAAAATAGG-3′
	UGT1A6	19T>G, 269G>A, 308C>A	S7A, R90H, S103X	UGT1A6-1pyroF2	Forward	354 bp	5′-TTTAACTCTTTCCAGGATGGC-3′
				UGT1A6-1pyroR1b[a]	Reverse		5′-bCTGTACTCTGTCTGAGGAGCAGT-3′
	UGT1A6	541A>G, 552A>C	T181A, R184S	UGT1A6-2pyroF1b[a]	Forward	180 bp	5′-bACCATCTGTGTACCTCTTCAGG-3′
				UGT1A6-2pyroR1	Reverse		5′-AGACAATAAAATAGATAGGGCTCC-3′
	UGT1A3	17A>G, 31T>C	Q6R, W11R	UGT1A3pyroF1	Forward	379 bp	5′-CACTCTGTCTTCCAATTACACGT-3′
				UGT1A3pyroR3b[a]	Reverse		5′-bTCCACGAAATGGCATAGGT-3′
	UGT1A3	133C>T, 140T>C	R45W, V47A	UGT1A3pyroF1b[a]	Forward	379 bp	5′-bCACTCTGTCTTCCAATTACACGT-3′
				UGT1A3pyroR3	Reverse		5′-TCCACGAAATGGCATAGGT-3′
	UGT1A	1091C>T[b]	P364L[b]	UT1AP364L-F	Forward	226 bp	5′-TGAAACTCAGAGATGTAACTGCTG-3′
				UT1AP364L-Rb[a]	Reverse		5′-bCAGAAGTCATTTCCAGAACATTC-3′
	UGT1A	1598A>C[b]	H533P[b]	exon5-F	Forward	242 bp	5′-TGGGTGGAGTTTGTGATGA-3′
				bExon5R-H533P[a]	Reverse		5′-bCCTTATTTCCCACCCACTT-3′
	UGT1A 3′-UTR[c]	1813C>T[b], 1941C>G[b], 2042C>G[b]	*1B*[d]	UGT1AEx5-1BPyroFb[a]	Forward	332 bp	5′-bATTTGAATATGTATCGTGCCC-3′
				UGT1AEx5-1BPyroR	Reverse		5′-CATTCATTCATTTCACCTACACT-3′
Pyrosequencing	*UGT1A8*	518C>G	A173G	UGT1A8c518gPF			5′-CGCCAGGGGAATAG-3′
	UGT1A10	4G>A	A2T	UGT1A10g4aPF			5′-GGCTGCAGTTCTCTCA-3′
	UGT1A10	177G>A	M59I	UGT1A10g177aPF			5′-GCATGAGGTGGTTGTAG-3′
	UGT1A10	200A>G	E67G	UGT1A10a200gPF			5′-TGAGTTGGCAACTGG-3′
	UGT1A10	605C>T	T202I	UGT1A10c605tPF			5′-GGTTCTCAGATGCCATG-3′
	UGT1A9	422C>G	S141C	UGT1A9c422gPF			5′-AGAATACTTAAAGGAGAGT-3′
	UGT1A9	726T>G	Y242X	UGT1A9t726gPF			5′-AAACACCTGTTACGGAG-3′
	UGT1A9	766G>A	D256N	UGT1A9g766aPF			5′-GGTTGTTGCGAACG-3′
	UGT1A7	387T>G, 391C>A, 392G>A	N129K, R131K	UGT1A7387-392PF			5′-ATTGCAGGAGTTTGTTTA-3′
	UGT1A7	622T>C	W208R	UGT1A7c622tPF			5′-TTCAAGGAGAGAGTA-3′
	UGT1A6	19T>G	S7A	UGT1A6-1t19gPseq-F			5′-CCTGCCTCCTTCGC-3′
	UGT1A6	269G>A	R90H	UGT1A6-1g269aPseq-F			5′-AAGAAGAGCTGAAGAACC-3′
	UGT1A6	308C>A	S103X	UGT1A6-1c308aPseq-F			5′-AACAATCACTTTGCTGAG-3′
	UGT1A6	541A>G, 552A>C	T181A, R184S	UGT1A6-2a552cPseq-R			5′-GGACACAGGGTCTGG-3′
	UGT1A3	17A>G, 31T>C	Q6R, W11R	UGT1A3a17g-seqF			5′-TGGCCACAGGACTC-3′
	UGT1A3	133C>T, 140T>C	R45W, V47A	UGT1A3t140cPseq-R			5′-ATGGAGCTCCCGCAA-3′
	UGT1A	1091C>T[b]	P364L[b]	UGT1A1P364LPseqF			5′-TTTGCATCTCAGGTCAC-3′
	UGT1A	1598A>C[b]	H533P[b]	Exon5-1598F-Page			5′-CACAAATCCAAGACCC-3′
	UGT1A 3′-UTR[c]	1813C>T[b]	*1B*[d]	UGT1AEx5c1813tPseqR			5′-TCCTGATCAAAGACACC-3′
	UGT1A 3′-UTR[c]	1941C>G[b]	*1B*[d]	UGT1AEx5c1941gPseqR			5′-CAGTAGGGGCAGC-3′
	UGT1A 3′-UTR[c]	2042C>G[b]	*1B*[d]	UGT1AEx5c2042gPseqR			5′-TGGCACCTATGAAGCA-3′

[a]b: biotinylated.
[b]The positions of nucleotide and protein were numbered as those of *UGT1A1*.
[c]UTR: untranslated region.
[d]1813C>T, 1941C>G, and 2042C>G found in the 3′-UTR were named haplotype **1B*.[19]

previously.[39] Biotinylated single-stranded DNA fragments were generated as described previously.[39] Briefly, PCR products were mixed with streptavidin beads for 10 min. The beads were transferred to a MultiScreen-HV Plate (Millipore Corporation, Billerica, MA, USA), and the buffer was removed by vacuum. DNA attached to the beads was denatured, washed twice, and then suspended in 20 mM Tris-acetate containing 2 mM Mg-acetate (pH 7.6). After transferring to a 96-well PSQ plate (Pyrosequencing AB, Uppsala, Sweden), 10 pmol of the sequencing primer (PAGE-purified grade) (Table 3) for SNP analysis was added to the single-stranded fragments. The mixture was incubated at 95°C for 2 min, and then cooled to room temperature for annealing. An automated pyrosequencing instrument, the PSQ™96MA (Pyrosequencing AB), and the PSQ 96 SNP reagent kit (Pyrosequencing AB) were used to perform the genotyping. To validate the typing methods, the results for 48 samples were confirmed to be identical to those obtained by direct sequencing (data not shown).

LD and haplotype analysis

Hardy–Weinberg equilibrium analysis and LD analysis were performed using SNPAlyze software (version 3.2). (Dynacom Co. Ltd., Yokohama, Japan), and pairwise two-dimensional maps between SNPs were obtained for the $|D'|$, χ^2, and r^2 values. Some of the haplotypes were unambiguously determined from the subjects with homozygous SNPs at all sites or a heterozygous SNP at only one site. Separately, the diplotype configurations (combination of haplotypes) were inferred by an expectation-maximization-based program, LDSUPPORT, which determines the posterior probability distribution of the diplotype configuration for each subject based on the estimated haplotype frequencies.[40]

The diplotype configurations of the subjects were inferred with probability (certainty) values over 0.96 for 184, 191, and 188 out of 196 subjects in the *UGT1A8-1A10* block (*Block 8/10*), the *1A9-1A7-1A6* block (*Block 9/6*), and the *1A3-1A1* block (*Block 3/1*), respectively. The *Block 4* (*1A4*) haplotypes were described previously.[21] Note that the predictability for the extremely rare haplotypes inferred from only one subject is known to be low in some cases. Haplotype analysis was also performed among the representative SNPs in *Block 9/6*, *Block 4*, and *Block 3/1* by LDSUPPORT software.

Abbreviations

LD	*linkage disequilibrium*
PCR	*polymerase chain reaction*
SN-38	*7-ethyl-10-hydroxycamptothecin*
SNP	*single nucleotide polymorphism*
UGT	*UDP-glucuronosyltransferase*
UTR	*untranslated region*

Acknowledgments

This study was supported in part by the Program for Promotion of Fundamental Studies in Health Sciences and by the Health and Labour Sciences Research Grants from Ministry of Health, Labour and Welfare, and the grant from the Japan Health Sciences Foundation. We thank Ms Chie Sudo for her secretarial assistance.

DUALITY OF INTEREST

None declared.

References

1 Radominska-Pandya A, Czernik PJ, Little JM. Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab Rev* 1999; **31**: 817–899.
2 Tukey RH, Strassburg CP. Genetic multiplicity of the human UDP-glucuronosyltransferases and regulation in the gastrointestinal tract. *Mol Pharmacol* 2001; **59**: 405–414.
3 Gong QH, Cho JW, Huang T, Potter C, Gholami N, Basu NK *et al.* Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus. *Pharmacogenetics* 2001; **11**: 357–368.
4 Metz RP, Ritter JK. Transcriptional activation of the UDP-glucuronosyltransferase 1A7 gene in rat liver by aryl hydrocarbon receptor ligands and oltipraz. *J Biol Chem* 1998; **273**: 5607–5614.
5 Guillemette C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J* 2003; **3**: 136–158.
6 Guillemette C, Ritter JK, Auyeung DJ, Kessler FK, Housman DE. Structural heterogeneity at the UDP-glucuronosyltransferase 1 locus: functional consequences of three novel missense mutations in the human UGT1A7 gene. *Pharmacogenetics* 2000; **10**: 629–644.
7 Gagne JF, Montminy V, Belanger P, Journault K, Gaucher G, Guillemette C. Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol Pharmacol* 2002; **62**: 608–617.
8 Huang YH, Galijatovic A, Nguyen N, Geske D, Beaton D, Green J *et al.* Identification and functional characterization of UDP-glucuronosyltransferases UGT1A8*1, UGT1A8*2 and UGT1A8*3. *Pharmacogenetics* 2002; **12**: 287–297.
9 Elahi A, Bendaly J, Zheng Z, Muscat JE, Richie Jr JP, Schantz SP *et al.* Detection of UGT1A10 polymorphisms and their association with orolaryngeal carcinoma risk. *Cancer* 2003; **98**: 872–880.
10 Jinno H, Saeki M, Saito Y, Tanaka-Kagawa T, Hanioka N, Sai K *et al.* Functional characterization of human UDP-glucuronosyltransferase 1A9 variant, D256N, found in Japanese cancer patients. *J Pharmacol Exp Ther* 2003; **306**: 688–693.
11 Jinno H, Saeki M, Tanaka-Kagawa T, Hanioka N, Saito Y, Ozawa S *et al.* Functional characterization of wild-type and variant (T202I and M59I) human UDP-glucuronosyltransferase 1A10. *Drug Metab Dispos* 2003; **31**: 528–532.
12 Villeneuve L, Girard H, Fortier LC, Gagne JF, Guillemette C. Novel functional polymorphisms in the UGT1A7 and UGT1A9 glucuronidating enzymes in Caucasian and African-American subjects and their impact on the metabolism of 7-ethyl-10-hydroxycamptothecin and flavopiridol anticancer drugs. *J Pharmacol Exp Ther* 2003; **307**: 117–128.
13 Ehmer U, Vogel A, Schutte JK, Krone B, Manns MP, Strassburg CP. Variation of hepatic glucuronidation: novel functional polymorphisms of the UDP-glucuronosyltransferase UGT1A4. *Hepatology* 2004; **39**: 970–977.
14 Nagar S, Zalatoris JJ, Blanchard RL. Human UGT1A6 pharmacogenetics: identification of a novel SNP, characterization of allele frequencies and functional analysis of recombinant allozymes in human liver tissue and in cultured cells. *Pharmacogenetics* 2004; **14**: 487–499.
15 Jinno H, Tanaka-Kagawa T, Hanioka N, Saeki M, Ishida S, Nishimura T *et al.* Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38), an active metabolite of irinotecan (CPT-11), by human UGT1A1 variants, G71R, P229Q, and Y486D. *Drug Metab Dispos* 2003; **31**: 108–113.
16 Judson R, Stephens JC, Windemuth A. The predictive power of haplotypes in clinical response. *Pharmacogenomics* 2000; **1**: 15–26.
17 Kohle C, Mohrle B, Munzel PA, Schwab M, Wernet D, Badary OA *et al.* Frequent co-occurrence of the TATA box mutation associated with Gilbert's syndrome (UGT1A1*28) with other polymorphisms of the

UDP-glucuronosyltransferase-1 locus (UGT1A6*2 and UGT1A7*3) in Caucasians and Egyptians. *Biochem Pharmacol* 2003; **65**: 1521–1527.
18 Carlini LE, Meropol NJ, Bever J, Andria ML, Hill T, Gold P *et al.* UGT1A7 and UGT1A9 polymorphisms predict response and toxicity in colorectal cancer patients treated with capecitabine/irinotecan. *Clin Cancer Res* 2005; **11**: 1226–1236.
19 Sai K, Saeki M, Saito Y, Ozawa S, Katori N, Jinno H *et al.* UGT1A1 Haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese cancer patients. *Clin Pharmacol Ther* 2004; **75**: 501–515.
20 Saeki M, Saito Y, Jinno H, Sai K, Kaniwa N, Ozawa S *et al.* Genetic polymorphisms of *UGT1A6* in a Japanese population. *Drug Metab Pharmacokinet* 2005; **20**: 85–90.
21 Saeki M, Saito Y, Jinno H, Sai K, Hachisuka A, Kaniwa N *et al.* Genetic variations and haplotypes of *UGT1A4* in a Japanese population. *Drug Metab Pharmacokinet* 2005; **20**: 144–151.
22 Iwai M, Maruo Y, Ito M, Yamamoto K, Sato H, Takeuchi Y. Six novel UDP-glucuronosyltransferase (UGT1A3) polymorphisms with varying activity. *J Hum Genet* 2004; **49**: 123–128.
23 Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J* 1997; **16**: 2482–2492.
24 Yamanaka H, Nakajima M, Katoh M, Hara Y, Tachibana O, Yamashita J *et al.* A novel polymorphism in the promoter region of human UGT1A9 gene (UGT1A9*22) and its effects on the transcriptional activity. *Pharmacogenetics* 2004; **14**: 329–332.
25 Lankisch TO, Vogel A, Eilermann S, Fiebeler A, Krone B, Barut A *et al.* Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. *Mol Pharmacol* 2005; **67**: 1732–1739.
26 Ando M, Ando Y, Sekido Y, Ando M, Shimokata K, Hasegawa Y. Genetic polymorphisms of the UDP-glucuronosyltransferase 1A7 gene and irinotecan toxicity in Japanese cancer patients. *Jpn J Cancer Res* 2002; **93**: 591–597.
27 Huang MJ, Yang SS, Lin MS, Huang CS. Polymorphisms of uridine-diphosphoglucuronosyltransferase 1A7 gene in Taiwan Chinese. *World J Gastroenterol* 2005; **11**: 797–802.
28 Innocenti F, Liu W, Chen P, Desai AA, Das S, Ratain MJ. Haplotypes of variants in the UDP-glucuronosyltransferase1A9 and 1A1 genes. *Pharmacogenet Genomics* 2005; **15**: 295–301.
29 Ciotti M, Basu N, Brangi M, Owens IS. Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38) by the human UDP-glucuronosyltransferases encoded at the UGT1 locus. *Biochem Biophys Res Commun* 1999; **260**: 199–202.
30 Lepine J, Bernard O, Plante M, Tetu B, Pelletier G, Labrie F *et al.* Specificity and regioselectivity of the conjugation of estradiol, estrone, and their catecholestrogen and methoxyestrogen metabolites by human uridine diphospho-glucuronosyltransferases expressed in endometrium. *J Clin Endocrinol Metab* 2004; **89**: 5222–5232.
31 Basu NK, Kubota S, Meselhy MR, Ciotti M, Chowdhury B, Hartori M *et al.* Gastrointestinally distributed UDP-glucuronosyltransferase 1A10, which metabolizes estrogens and nonsteroidal anti-inflammatory drugs, depends upon phosphorylation. *J Biol Chem* 2004; **279**: 28320–28329.
32 Turgeon D, Chouinard S, Belanger P, Picard S, Labbe JF, Borgeat P *et al.* Glucuronidation of arachidonic and linoleic acid metabolites by human UDP-glucuronosyltransferases. *J Lipid Res* 2003; **44**: 1182–1191.
33 Little JM, Kurkela M, Sonka J, Jantti S, Ketola R, Bratton S *et al.* Glucuronidation of oxidized fatty acids and prostaglandins B1 and E2 by human hepatic and recombinant UDP-glucuronosyltransferases. *J Lipid Res* 2004; **45**: 1694–1703.
34 Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci USA* 1998; **95**: 8170–8174.
35 Sugatani J, Yamakawa K, Yoshinari K, Machida T, Takagi H, Mori M *et al.* Identification of a defect in the UGT1A1 gene promoter and its association with hyperbilirubinemia. *Biochem Biophys Res Commun* 2002; **292**: 492–497.
36 Oguri T, Takahashi T, Miyazaki M, Isobe T, Kohno N, Mackenzie PI *et al.* UGT1A10 is responsible for SN-38 glucuronidation and its expression in human lung cancers. *Anticancer Res* 2004; **24**: 2893–2896.
37 Saeki M, Ozawa S, Saito Y, Jinno H, Hamaguchi T, Nokihara H *et al.* Three novel single nucleotide polymorphisms in UGT1A10. *Drug Metab Pharmacokinet* 2002; **17**: 488–490.
38 Saeki M, Saito Y, Jinno H, Sai K, Komamura K, Ueno K *et al.* Three novel single nucleotide polymorphisms in UGT1A9. *Drug Metab Pharmacokinet* 2003; **18**: 146–149.
39 Saeki M, Saito Y, Jinno H, Tohkin M, Kurose K, Kaniwa N *et al.* Comprehensive UGT1A1 genotyping in a Japanese population by pyrosequencing. *Clin Chem* 2003; **49**: 1182–1185.
40 Kitamura Y, Moriguchi M, Kaneko H, Morisaki H, Morisaki T, Toyama K *et al.* Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm. *Ann Hum Genet* 2002; **66**: 183–193.

The FASEB Journal express article 10.1096/fj.05-4034fje. Published online December 22, 2005.

Dimerization and the signal transduction pathway of a small in-frame deletion in the epidermal growth factor receptor

Kazuko Sakai,*[,§] Tokuzo Arao,* Tatsu Shimoyama,* Kimiko Murofushi,[§] Masaru Sekijima,[‖] Naoko Kaji,[‖] Tomohide Tamura,[†] Nagahiro Saijo,[†] and Kazuto Nishio*[,‡]

*Shien-Lab, Medical Oncology, and [†]National Cancer Center Hospital and [‡]Pharmacology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo, Japan; and [§]Department of Biology, Faculty of Science, Ochanomizu University, Tokyo, Japan; and [‖]Mitsubishi Chemical Safety Institute, Ibaraki, Japan

Corresponding author: Kazuto Nishio, Shien-Lab, Medical Oncology, National Cancer Center Hospital, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, Japan. E-mail: knishio@gan2.res.ncc.go.jp

ABSTRACT

A short, in-frame deletional mutant (E746-A750del) is one of the major mutant forms of epidermal growth factor receptor (EGFR) and has been reported to be a determinant of response to EGFR tyrosine kinase inhibitors such as gefitinib and erlotinib. However, the biological and pharmacological functions of mutational EGFR remain unclear. To clarify these biological functions of deletional EGFR, we examined the cellular response to EGF ligand stimulation. Dimerization and phosphorylation of EGFR were observed without any ligand stimulation in the 293(D) cells transfected with deletional EGFR as compared with those transfected with wild-type EGFR (293(W) cells). When the 293(D) cells were exposed to gefitinib, an immunoblotting analysis revealed remarkable inhibition of AKT phosphorylation but not phospho-p44/42 MAPK. To examine the cellular response in a lung cancer cell line intrinsically expressing deletional EGFR, phospho-EGFR, and downstream reactions were monitored under EGF stimulation with a beads-based mulitiplex assay. EGFR and its downstream proteins were constitutively phosphorylated in the PC-9 cells without any ligand stimulation as compared with A549 lung cancer cells expressing wild-type EGFR. In conclusion, deletional EGFR is constitutively active and phosphorylates p44/42 MAPK and AKT in the cells, although the fact that the EGFR phosphorylation in the PC-9 cells is still modulated by EGF stimulation cannot be ignored. Gefitinib-inhibited phospho-AKT predominantly in deletional EGFR expressing cells.

Key words: mutation • gefitinib • tyrosine kinase

Epidermal growth factor receptor (EGFR) belongs to the ErbB family (1) and contains an extracellular ligand-binding domain, a transmembrane domain, and a tyrosine kinase domain. Wild-type EGFR is unphosphorylated and exists as a monomer in the unstimulated conditions. Binding of ligands such as EGF and TGF-α leads to the dimerization of EGFR, phosphorylation of tyrosine residues (2), and stimulation of the phosphorylation pathway downstream. This signaling pathway is considered to be closely related to cellular growth, differentiation, and the development of malignant phenotypes of cancer cells (1, 2). Increased

expression of EGFR and gene amplification of EGFR are often observed in several types of tumors such as lung and breast cancers.

Recently, many small anticancer molecules have been developed that target EGFR. Gefitinib (Iressa) is an orally available EGFR tyrosine kinase inhibitor. Previous clinical studies have demonstrated that EGFR expression levels in tumors did not correlate with the clinical response to gefitinib (3). On the other hand, EGFR mutation in adenocarcinoma of the lung was reported to be a determinant of sensitivity for EGFR tyrosine kinase inhibitors such as gefitinib and erlotinib (4, 5). To date, over 30 types of EGFR mutation have been reported in lung cancer. The in-frame, 15 base deletional mutation (delE746-A750 type deletion) is one of the most common of these EGFR mutations. We previously demonstrated that overexpression of deletional EGFR increased the cellular sensitivity to tyrosine kinase inhibitors targeting EGFR in human HEK293 cells in vitro (6). However, it remains unclear how deletional EGFR alters dimerization and the downstream signaling pathways from these heterodimers. Recently, Tracy et al. demonstrated that another major mutant EGFR (L858R) altered signal transduction downstream (7). In addition, it has been suggested that the L858R mutation is a hyper-response to ligand stimulation (7). We hypothesized that the deletional EGFR is constitutively active. The aim of this study was to clarify the downstream of the signaling and its function of the deletional EGFR.

Technically, a beads-based multiplex assay (Bio-Plex phosphoprotein assay) (8) allowed us to analyze numbers of phosphoproteins simultaneously after ligand stimulation.

MATERIALS AND METHODS

Reagents

Gefitinib (Iressa, ZD1839) was provided by AstraZeneca (Cheshire, UK).

Cell culture

The human embryonic kidney HEK293 cell line and human nonsmall-cell lung cancer (NSCLC) cell line A549 and cervix epitheloid cancer cell line HeLa were obtained from the American Type Culture Collection (Manassas, VA) and were cultured in RPMI 1640 medium supplemented with 10% FBS. The human NSCLC cell line PC-9 was established at the Tokyo Medical University (9, 10), and was maintained in RPMI 1640 medium (Sigma, St. Louis, MO) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Life Technologies, Rockville, MD).

Plasmid construction and transfection

The construction of the mock expression plasmid vector (empty vector) and of the wild-type of EGFR and the 15-bp deletional EGFR (delE746-A750 type deletion) vectors that posses the same deletion site observed in PC-9 cells has been described in another paper in detail (6, 11, 12). The plasmids were transfected into the HEK293 cells, and the transfectants were selected by Zeosin (Sigma). The stable transfectants (pooled cultures) of the empty vector, wild type of EGFR, and its deletion mutant were designated as 293(M), 293(W), and 293 (D) cells, respectively.

Reverse-transcription PCR

Five micrograms of total RNA from each cultured cell line was converted to cDNA with a GeneAmp RNA-PCR kit (Applied Biosystems, Foster City, CA). The primers used for the PCR were EGFR (forward), 5′-AAGTTAAAATTCCCGTCGCTATCA-3′ and (reverse) 5′-GAGCCAATATTGTCTTTGTGTTCC-3′. PCR amplification consisted of 28 cycles (95°C for 45 s, 57.5°C for 30 s, and 72°C for 45 s) followed by incubation at 72°C for 7 min. The RT-PCR products were analyzed with a 2100 Bioanalyzer and DNA 500 kit (Agilent Technologies, Waldbronn, Germany).

Chemical cross-linking

After treatment with or without EGF (Sigma), monolayer cells were washed twice with ice-cold phosphate buffered saline containing 0.33 mM $MgCl_2$ and 0.9 mM $CaCl_2$ (PBS(+)) and chemically cross-linked for 20 min at room temperature with freshly prepared 1.5 mM bis(sulfosuccinimidyl) suberate (BS^3, Pierce, Rockford, IL). To terminate the reaction, a final concentration of 20 mM glycine was added for an additional 5 min. For immunoblot analysis, the cells were washed twice with ice-cold PBS(+) and lysed with M-PER mammalian protein extraction reagent (Pierce). The lysate was centrifuged at 20,000 *g* for 10 min, and the protein concentration of the supernatant was measured with a BCA (bicinchoninic acid) protein assay (Pierce).

Ligand stimulation

After reaching 70–80% confluence, cultured cells were stimulated with EGF, TGF-α, and HB-EGF for 10 min under nonstarved conditions or serum-starved conditions. The cells were washed twice with ice-cold PBS(+), and lysed for immunoblotting.

Drug treatment

After reaching 70–80% confluence, cultured cells, were exposed to various concentrations of gefitinib and stimulated or not stimulated with EGF (100 ng/ml) for 10 min under nonstarved conditions or serum-starved conditions. The cells were washed twice with ice-cold PBS(+) and lysed for immunoblotting.

Immunoblot analysis

Immunoblot analysis was performed as described previously (12). Equivalent amounts of protein were separated by electrophoresis on a SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore, Bedford, MA). The membrane was probed with a mouse monoclonal antibody against EGFR (Transduction Lab, San Diego, CA), a phospho-EGFR antibody (specific for Tyr1068), p44/42 mitogen-activated protein kinase (p44/42 MAPK), phospho-p44/42 MAPK, AKT, phospho-AKT, nuclear factor-κB (NF-κB) inhibitor α (IκB-α), and phospho-IκB-α antibody (Cell Signaling Technology, Beverly, MA) as the first antibody, followed by a horseradish peroxidase-conjugated secondary antibody. The bands were visualized with electrochemiluminescence (ECL, Amersham, Piscataway, NJ), and images of blotted patterns were analyzed with NIH image software (National Institutes of Health, Bethesda, MD).

Phosphoprotein assay

A panel of phosphoproteins was measured in duplicate using a bead-based mulitiplex assay (Bio-Plex phosphoprotein assay, Bio-Rad, Hercules, CA), according to the manufacturer's instructions (8, 13). The EGFR-transfected 293 cells and NSCLC cells cultured in the serum-free medium for 24 h were stimulated by the addition of EGF at a final concentration of 100 ng/ml for the indicated time intervals. After incubation, the cells were rinsed with ice-cold Cell Wash Buffer and collected. The lysate was centrifuged at 1,700 *g* for 20 min. The protein concentration was calculated with a DC (detergent compatible) protein assay (Bio-Rad). The Bio-Plex assay was customized to detect and quantify phosphoproteins of EGFR, p44/42 MAPK, activating transcription factor 2 (ATF-2), c-Jun N-terminal kinase (JNK), p38 mitogen-activated protein kinase (p38 MAPK), IκB-α, and signal transducer and activator of transcription 3 (STAT-3). The prepared first antibody with coupled beads was captured under 96-well plates, and samples (17.5 μg each) were incubated overnight at room temperature. Samples were incubated with biotin-labeled detection antibodies followed by further incubation with the fluorescence-labeled avidin reporter. The level of phosphoproteins bound to the beads was indicated by the intensity of the reporter signal. The signal was measured with Bio-Plex Manager software (Bio-Rad) interfaced with a Bio-Plex Reader (Bio-Rad). In this assay, we used the lysates of untreated HeLa cells as the background control for all phosphoprotein assays provided by the Bio-Plex phosphoprotein assay (14). This experiment was repeated in duplicate.

RESULTS

Dimerization and phosphorylation of wild-type EGFR and deletional EGFR

The EGFR expression level in the 293 cells transfected with the empty vector (293(M)), wild-type EGFR (293(W)), and deletional EGFR (293(D)), in the A549 and PC-9 NSCLC cells, and in the HeLa cervix epitheloid cancer cells were determined by RT-PCR and immunoblot analysis (Fig. 1). The PCR products were separated into wild-type EGFR (upper band) and deletion mutant EGFR (lower band) by the different lengths of the sequences. Overexpression of EGFR was detected in the 293(W) and 293(D) cells. Only a small amount of intrinsically EGFR was detected in 293(M) cells by RT-PCR. We sequenced exon 19 and 20 of EGFR in HEK293 cells, but no mutations were detected (data not shown). A high level of EGFR expression was detected in the PC-9 cells, and a moderate level was detected in the A549 and HeLa cells. The EGFR protein levels closely matched the mRNA expression levels in all cells.

Dimerization of EGFR, the first step in the EGFR signaling pathway, was examined by chemical cross-linking. The 293(M), 293(W), and 293(D) cells were treated with EGF (10 ng/ml) for 10 min under nonstarved conditions (Fig. 2*A*). Cells were incubated with the cross-linking reagent BS^3. Dimerization and phosphorylation of EGFR were determined by immunoblot with anti-EGFR and anti-phospho EGFR antibodies. Dimerization and expression of EGFR were not detected in the 293(M) cells. Dimerized EGFR with a molecular weight of ~400 kDa was detected in the 293(W) cells. Increased phosphorylation and dimerization of the deletional EGFR were detected without EGF stimulation in the 293(D) cells by the chemical cross-linking and immunoblot assay, respectively. When the cells were stimulated with the EGF ligand (10 ng/ml for 10 min), increased phospho-EGFR dimers were observed in the 293(W) cells, whereas no response of EGFR to EGF was observed in the 293(D) cells. We quantified the levels of monomeric and dimerized EGFR in the 293(W) and 293(D) cells densitometrically under nonstarved conditions (Fig. 2*A*, right). In the