Three combinations have shown excellent first-line efficacy in phase III trials – IFL with bevacizumab, FOLFOX, and FOLFIRI – however, neither of these combinations is clearly superior. Our clinical practice in Japan has been guided in a major way by extrapolation from the results of clinical trials conducted mainly in Western countries. To evaluate the value of FOLFOX regimens in the treatment of refractory or advanced CRC, a retrospective analysis study was designed to assess the feasibility (toxicities) and efficacy of combining oxaliplatin with the LV5FU2 schedule in a Japanese population. We herein report our experience with two FOLFOX regimens (FOLFOX4 and modified [m] FOLFOX6) in patients with advanced CRC, specifically, the toxicities and objective tumor response rates obtained.

Patients and methods

A retrospective analysis study was conducted at Kinki University Hospital in 51 consecutive patients with histologically confirmed metastatic colon or rectum cancer who were treated between April 2005 and March 2006. The primary objectives were to assess the feasibility (toxicities) and efficacy of two FOLFOX regimens (FOLFOX4 and mFOLFOX6) in a Japanese population. FOLFOX4 is a regimen comprising oxaliplatin 85 mg/m^2 as a 2-h infusion (day 1); LV 100/m^2 per day as a 2-h infusion (days 1 and 2); followed by a 5FU bolus 400 mg/m^2 per day and 5FU 600 mg/m^2 per day as a 22-h infusion (days 1 and 2). mFOLFOX6 is a regimen also comprising oxaliplatin 85 mg/m^2 as a 2-h infusion (day 1), LV 200 mg/m^{2}per day as a 2-h infusion (day 1), followed by a 5FU bolus 400 mg/m^2 (day 1) and 5FU 2400 mg/m^2 per day as a 46-h infusion (days 1 to 2). These therapies were administered on day 1 and repeated on day 2 of a 14-day treatment cycle. Routine antiemetic prophylaxis with a serotonin (5-HT3) antagonist (granisetron) and dexamethasone was given (Fig. 1). The use of implantable ports and infusion pumps allowed chemotherapy to be administered on an outpatient basis in some cases. Treatment was continued until either disease progression, the occurrence of unacceptable toxicity, or the patient refused further treatment. Toxicity was graded according to the Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 regarding toxicities other than peripheral sensory neuropathy and by following the oxaliplatin-specific scale (DEB-NTC). The definitions in the oxaliplatin-specific scale, which was developed as a specific scoring scale for oxaliplatin-inducing peripheral sensory neuropathy, are as follows: grade 1, transient dysesthesia and/or paresthesia lasting for less than 7 days; grade 2, transient dysesthesia and/or paresthesia lasting for 7 days or longer; and grade 3, dysesthesia and/or paresthesia with pain or function impairment that interferes with activities of daily living (such as difficulty with fastening buttons and writing). The response of measurable target lesions to treatment was objectively evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST) criteria after each four cycles of treatment. Complete response (CR) was defined as the disappearance of all disease. Partial response (PR) was defined as at least a 30% reduction in the sum of the longest diameters of all measured lesions by at least 4 weeks. Progressive disease (PD) was defined as an increase in lesions by 20%

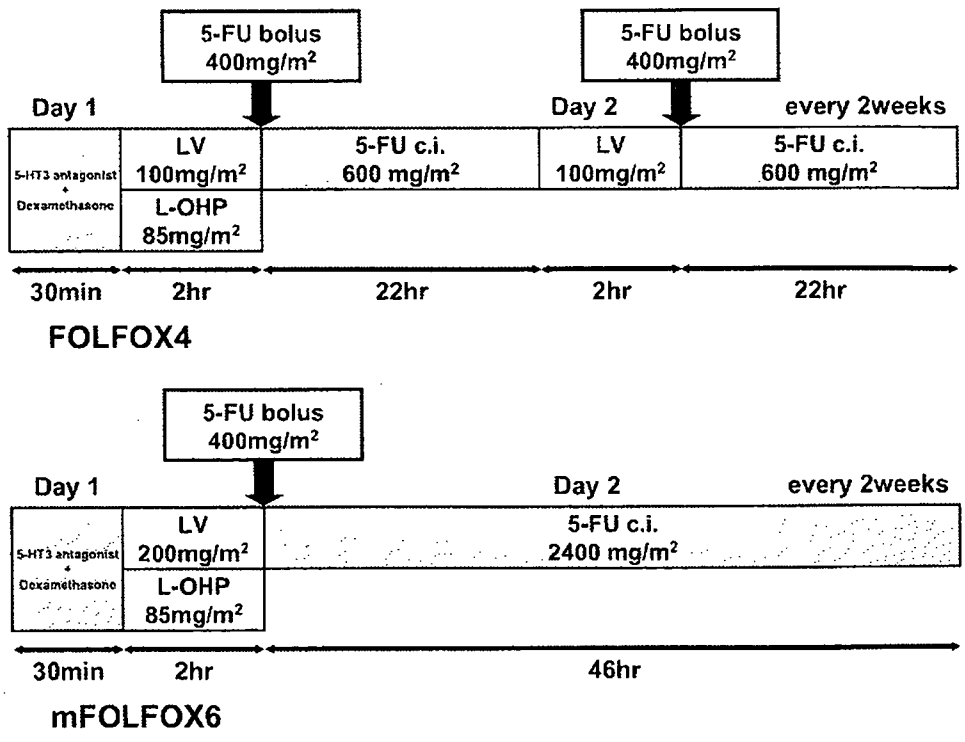

Fig. 1. Treatment schema for FOLFOX4 and mFOLFOX6 regimens. *LV*, leucovorin; *5-FU*, 5-fluorouracil; *L-OHP*, oxaliplatin; *c.i.*, continuous infusion

or greater, or the appearance of new lesions. Responses not falling into any of these categories were classified as stable disease (SD).

Results

Patient characteristics

A total of 51 patients with a median age of 61 years (range, 34–78 years) with refractory or advanced CRC were retrospectively analyzed. The patients' characteristics are listed in Table 1. Patients were treated with the FOLFOX4 (39%) or mFOLFOX6 regimens (61%). Twenty-eight patients (55%; FOLFOX4, 9; mFOLFOX6, 19) were treated in a first-line setting, and 23 patients (45%; FOLFOX4, 11; mFOLFOX6, 12) were treated in a second-line setting. Since April 2005, 4 months prior to beginning this study, we have used the FOLFOX4 regimen for inpatients, and from that time we selected the mFOLFOX6 regimen for all patients in the outpatient setting. The total number of chemotherapy cycles administered was 384, with a median of 8 cycles per patient (range, 1–12 cycles). The median dose intensity (actual/planned dose) was 93.4% for oxaliplatin and 100% for 5FU in the FOLFOX4 group and 91.2% for oxaliplatin and 94.8% for 5FU in the mFOLFOX6 group. The median dose intensity of oxaliplatin was 37 mg/m^2 per week (range, 31–42 mg/m^2 per week) in the FOLFOX4 group and 34 mg/m^2 per week (range, 23–42 mg/m^2 per week) in the mFOLFOX6 group.

Hematological toxicity

Several pertinent hematological toxicities are listed in Table 2, shown with numbers of patients who experienced them. The onset of neutropenia typically occurred between 10 and 14 days after treatment. Grade 3 and 4 neutropenia was observed in 20% of patients, with neutropenic fever being uncommon. Neutropenia often caused delay in the start of a subsequent treatment course. In all, 88 (23%) of 384 cycles were delayed due to toxicity, most commonly hematological: 64 (17%) for neutropenia and 24 (6%) for neurotoxicity. Using our administration schedule, no thrombocytopenia of over grade 3 was observed to develop. In addition, only one patient developed grade 3 anemia with transfusion.

Nonhematological toxicity

The most common nonhematological adverse effects of the FOLFOX regimens were peripheral neuropathy and lethargy (fatigue). These effects are listed in Table 3. Two patients experienced grade 2 hypersensitivity reactions (rash/hives, erythema, and one patient also experienced vomiting) during the administration of oxaliplatin. The symptoms rapidly resolved, in a few minutes, on symptomatic treatment (termination of infusion, use of steroids, and antagonists of type 1 and 2 histamine receptors). In using successful strategies over the next treatment courses (slowing the infusion rate, increasing the doses of steroids, and dose reduction of oxaliplatin), both patients were able to tolerate rechallenge of oxaliplatin, and one patient achieved a partial response. Peripheral neurotoxicity, characterized by paresthesia in a symmetric, glove-and-stocking distribution, occurred in 40 (78%) patients and there was no grade over 3.

Whenever the number of treatment cycles increases, neuropathy, within grade 2 level, tends to increase. The incidence of neurotoxicity along with the number of treatment cycles is listed in Table 4. Cold-related dysesthesia

Table 1. Characteristics of the study patients

Characteristics	$n = 51$
Sex, male/female	28/23
Age, years, median (range)	61 (34–78)
Performance status (ECOG), 0/1/2	19/25/7
Primary tumor, colon/rectum/rectosigmoid	28/20/3
Adjuvant therapy, +/–	18/33
Previous irinotecan therapy, +/–	15/36
Site of metastases, lung/liver/LN/peritoneum	22/21/18/7
FOLFOX4/mFOLFOX6	20/31
First-line/second-line	28/23
Dose reduction: +/–	7/44

ECOG, Eastern Cooperative Oncology Group; LN, distant lymph nodes

Table 2. Hematological toxicity (CTCAE V3.0)

	$n = 51$					
	Grade 1	Grade 2	Grade 3	Grade 4	All grades	Grade ≧3
Leucocytopenia	17	16	1	0	67%	2%
Neutropenia	18	15	9	1	84%	20%
Anemia	19	8	1	0	55%	2%
Thrombocytopenia	23	5	0	0	55%	0%

Table 3. Nonhematological toxicity (CTCAE V3.0)*

	$n = 51$					
	Grade 1	Grade 2	Grade 3	Grade 4	All grades	Grade ≧3
Anorexia	16	2	2	0	39%	4%
Nausea	13	2	2	0	33%	4%
Vomiting	5	2	2	0	18%	4%
Mucositis	10	1	2	0	25%	4%
Febrile neutropenia	–	–	0	0	0%	–
Hand-foot syndrome	2	0	0	0	4%	–
Pigmentation	4	0	–	–	7%	–
Allergy	0	2	–	–	4%	–
Lethargy	13	4	0	0	33%	–
AST/ALT elevation	22	2	0	0	47%	–
Diarrhea	8	2	2	0	23%	4%
Sensory neuropathy	31	9	0	0	78%	–

*Other than sensory neuropathy

Table 4. Incidence of neurotoxicity in FOLFOX regimens

	1–4 Cycles ($n = 9$)			5–8 Cycles ($n = 28$)			9–12 Cycles ($n = 14$)			All cycles ($n = 51$)		
Grade	1	2	3	1	2	3	1	2	3	1	2	3
Sensory neuropathy	5	0	0	19	2	0	7	7	0	31	9	0
	56%	0%	0%	68%	7%	0%	50%	50%	0%	60%	18%	0%

Grade	The oxaliplatin-specific scale (DEB-NTC)			
	0	1	2	3
Dysesthesia and/or paresthesia	No abnormality	Transient dysesthesia and/or paresthesia lasting less than 7 days	Transient dysesthesia and/or paresthesia lasting 7 days or more	Dysesthesia and/or paresthesia with pain or function impairment that interferes with activities of daily living

was reported in 31 patients (61%). Paresthesia lasting 7 days or longer (grade 2) occurred in 9 patients (18%). Peripheral neuropathy appeared in two forms. In the first form, an acute, transient, cold-exacerbated dysesthesia or paresthesia occurred shortly after the administration of oxaliplatin; it affected the hands, feet, perioral area, and throat; and typically lasted for several days after drug administration. In the second form, a delayed-onset, cumulative, dose-related peripheral neuropathy was characterized by paresthesias affecting the hands and feet that did not remit between cycles of treatment. Investigators also reported pharyngolaryngeal dysesthesia in only one patient; however, no patients had a laryngospasm-like syndrome.

Overall, 7 of the 51 patients (14%) required dose modification during treatment; dose reduction was required for oxaliplatin alone in 4 patients, for 5FU alone in 2 patients and for both agents in 1 patient. The majority of dose reductions were by one level (reduction to 65 mg/m^2 oxaliplatin and/or 75% of the starting dose of 5FU). No patients required a second-level dose reduction. The adverse events most commonly leading to dose reduction were neurotoxicity (1 patient in FOLFOX4 and 3 patients in mFOLFOX6) and diarrhea (2 patients in mFOLFOX6). In addition, 2 patients in the mFOLFOX6 setting underwent a dose reduction of oxaliplatin due to allergic reaction. The most common reason for treatment discontinuation was PD.

Antitumor activity

All 51 patients were able to be evaluated for response. Objective responses are listed in Table 5 and Table 6. There was no complete response. The overall objective response rates (in those who underwent first-line or second-line therapy) were 50.0% and 8.7%, respectively (Table 6). Stable disease was achieved in 49% of patients. The tumor control rate (PR + SD) was 80.4%.

Discussion

The recent advent of several new agents for the treatment of metastatic CRC has markedly enhanced the therapeutic armamentarium for this disease. Oxaliplatin in combination with infusional 5FU in the FOLFOX regimens has been shown to be effective in achieving improved response rate, time to progression, and survival time compared with 5FU/LV. In addition, recent large clinical phase III studies (N9741, EFC4584, GERCOR) showed that combination chemotherapy regimens, including irinotecan and oxaliplatin, markedly improved response rates and prolonged median survival over those seen with 5FU/LV),[11–13] and these combination chemotherapy regimens have supplanted 5FU/LV as a standard systemic approach for metastatic CRC. The median survival time (MST) has been gradually pro-

Table 5. Objective responses – (1)

FOLFOX4 (n = 20)				
CR	PR	SD	PD	NE
0	5 (25%) First-line, 3; second-line, 2	12 (60%)	3 (15%)	0
mFOLFOX6 (n = 31)				
0	11 (35.5%) First-line, 11; second-line, 0	13 (41.9%)	7 (22.6%)	0

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NE, not evaluable

Table 6. Objective responses – (2)

First-line (n = 28)				
CR	PR	SD	PD	NE
0	14 (50%) FOLFOX4, 3; mFOLFOX6, 11	11 (39.3%)	3 (10.7%)	0
Second-line (n = 23)				
0	2 (8.7%) FOLFOX4, 2	14 (60.9%)	7 (30.4%)	0

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NE, not evaluable

longed through the use of 5FU/LV with irinotecan and oxaliplatin. Currently, with the addition of molecular targeted agents, an MST of over 20 months has been reported.

Since April 2005, and the approval of oxaliplatin in Japan, clinical practice in this country has been conducted in a major way by extrapolation from the results of clinical trials conducted mainly in large Western phase III studies. The results of the present retrospective study demonstrate the efficacy and feasibility of FOLFOX regimens (FOLFOX4 and mFOLFOX6) as treatment for patients with advanced CRC in the Japanese population, as has been shown in Western populations. In this retrospective analysis study in a Japanese population, neutropenia grade 3/4 occurred in 20% of the patients who were assigned to receive oxaliplatin, but it was nonfebrile, whereas grade 3/4 vomiting and mucositis affected only 4% of the patients, while diarrhea affected 4%.

Lethargy has been described as the most frequent adverse event of the mFOLFOX6 regimen in a recent report by Braun et al.[14] In our study, 17 (33%) patients experienced lethargy similar to general fatigue symptoms.

The cumulative dose-limiting toxicity of oxaliplatin is peripheral sensory neuropathy, which reportedly occurs in about 70%–80% of patients; it typically resolves a few months after discontinuation of treatment, and may be exacerbated by cold stimulation. In our series, paresthesia lasting 7 days or longer was observed in 18% of patients and led to an oxaliplatin dose reduction for four patients after they had received a minimum of seven cycles (or at least 4 months) of chemotherapy.

The mechanism of this neurotoxicity has been elucidated to be as follows: the increased neuronal excitability is due to the action of oxaliplatin on voltage-gated sodium channels through the chelation of calcium by the oxaliplatin metabolite. The prevention of this neurotoxicity is a major goal, taking in to account the wide indications of this drug. Various different approaches have been either previously studied or are now being evaluated, based on pathogenic or practical concepts: (1) modification of the administration schedule; (2) substances acting upon sodium channels, such as calcium-magnesium, carbamazepine, gabapentine, venlafaxine; (3) detoxifying agents and antioxidants, such as glutathione, amifostine, alphalipoic acid, tocopherol; (4) substances used in other kinds of neuropathy, such as glutamine and alphalipoic acid; (5) neurotrophic factors, such as nerve growth factor (NGF), LIF; and (6) oxaliplatin analogs, with a DACH platin, without oxalate. Calcium-magnesium infusion appears to be an efficient and safe approach.

In this study, after September 2005, 32 patients (63%) were administered calcium-magnesium infusion for the prevention of the oxaliplatin-related neurotoxicity. Further studies are necessary for a better understanding and prevention of this potentially severe neurotoxicity.

In terms of antitumor activity, although the response rate (RR) in our population was slightly lower in comparison to that in previous Western clinical studies,[11–13] both of the oxaliplatin-based regimens demonstrated a promising objective RR in the first-line setting (50.0%) and in the tumor control rate (80.4%).

In a GERCOR study, the median survival was 21.5 months in 109 patients allocated to FOLFIRI then FOLFOX6 versus 20.6 months in 111 patients allocated to FOLFOX6 then FOLFIRI (P = 0.99). In first-line therapy, FOLFIRI achieved a 56% RR and 8.5-month median PFS, versus FOLFOX6, which achieved a 54% RR and 8.0-month median PFS (P = 0.26). Second-line FOLFIRI achieved a 4% RR and 2.5-month median PFS, versus

FOLFOX6, which achieved a 15% RR and 4.2-month PFS.[13] Although our study could not evaluate enough data for PFS and MST due to the short observation period after the approval of oxaliplatin in Japan, both the FOLFOX regimens we used seem to be beneficial as first-line and second-line therapy for refractory or advanced CRC in a Japanese population, with an overall response rate which is comparable to Western figures regarding first-line and second-line therapy. FOLFOX6 is the most useful of the FOLFOX regimens because it is simple and can be administered on an outpatient basis. When we use oxaliplatin in FOLFOX regimens, because 85 mg/m^2 is the approved dose for usage in Japan, the treatment is adapted for this dose, even in the mFOLFOX6 regimen.

In conclusion, the FOLFOX regimens we used were found to demonstrate good efficacy as chemotherapy regimens in our population, with an acceptable overall toxicity profile. However, attention must be paid to the occurrence of peripheral sensory neuropathy, which may influence a patient's quality of life, while also limiting the continuation of such treatment.

References

1. Poon MA, O'Connell MJ, Moertel CG, et al. (1989) Biochemical modulation of fluorouracil: evidence of significant improvement of survival and quality of life in patients with advanced colorectal carcinoma. J Clin Oncol 7:1407–1418
2. Levi F, Zidani R, Misset JL (1997) Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer: International Organization for Cancer Chronotherapy. Lancet 350:681–686
3. Giacchetti S, Perpoint B, Zidani R, et al. (2000) Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil-leucovorin as first-line treatment of metastatic colorectal cancer. J Clin Oncol 18:136–147
4. Douillard JY, Cunningham D, Roth AD, et al. (2000) Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. Lancet 355:1041–1047
5. Machover D, Diaz-Rubio E, de Gramont A, et al. (1996) Two consecutive phase II studies of oxaliplatin (L-OHP) for treatment of patients with advanced colorectal carcinoma who were resistant to previous treatment with fluoropyrimidines. Ann Oncol 7:95–98
6. Levi F, Perpoint B, Garufi C, et al. (1993) Oxaliplatin activity against metastatic colorectal cancer: a phase II study of 5-day continuous venous infusion at circadian rhythm modulated rate. Eur J Cancer 29A:1280–1284
7. Diaz-Rubio E, Sastre J, Zaniboni A, et al. (1998) Oxaliplatin as single agent in previously untreated colorectal carcinoma patients: a phase II multicentric study. Ann Oncol 9:105–108
8. Becouarn Y, Ychou M, Ducreux M, et al. (1998) A phase II trial of oxaliplatin as first-line chemotherapy in metastatic colorectal cancer patients. J Clin Oncol 8:2739–2744
9. de Gramont A, Vignoud J, Tournigand C, et al. (1997) Oxaliplatin with high-dose leucovorin and 5-fluorouracil 48-hour continuous infusion in pretreated metastatic colorectal cancer. Eur J Cancer 33:214–219
10. Raymond E, Buquet-Fagot C, Djelloul S, et al. (1997) Antitumor activity of oxaliplatin in combination with 5-fluorouracil and the thymidylate synthase inhibitor AG337 in human colon, breast, and ovarian cancers. Anticancer Drugs 8:876–885
11. Goldberg RM, Sargent DJ, Alberts SR, et al. (2004) A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. J Clin Oncol 22:23–30
12. de Gramont A, Figer A, Bonetti A, et al. (2000) Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. J Clin Oncol 18:2938–2947
13. Tournigand C, Andre T, de Gramont A, et al. (2004) FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: a randomized GERCOR study. J Clin Oncol 22:229–237
14. Braun MS, Adab F, Seymour MT, et al. (2003) Modified de Gramont with oxaliplatin in the first-line treatment of advanced colorectal cancer. Br J Cancer 89:1155–1158

perspectives Journal of proteome research

Personalized Medicine and Proteomics: Lessons from Non-Small Cell Lung Cancer

György Marko-Varga,[†,‡] Atsushi Ogiwara,[†,§,∥] Toshihide Nishimura,[§,∥] Takeshi Kawamura,[§,∥] Kiyonaga Fujii,[§,∥] Takao Kawakami,[§,∥] Yutaka Kyono,[∥] Hsiao-kun Tu,[∥] Hisae Anyoji,[∥] Mitsuhiro Kanazawa,[∥] Shingo Akimoto,[∥] Takashi Hirano,[⊥] Masahiro Tsuboi,[⊥] Kazuto Nishio,[§] Shuji Hada,[#] Haiyi Jiang,[×] Masahiro Fukuoka,[^] Kouichiro Nakata,[◆] Yutaka Nishiwaki,[÷] Hideo Kunito,[§] Ian S. Peers,[◊] Chris G. Harbron,[◊] Marie C. South,[▽] Tim Higenbottam,[∇,£] Fredrik Nyberg,[*,∀,○] Shoji Kudoh,[ƒ] and Harubumi Kato[÷]

Respiratory Biological Sciences, AstraZeneca R&D Lund, SE-221 87 Lund, Sweden, Clinical Proteome Center, Tokyo Medical University, Shinjuku Sumitomo Building 17F, 2-6-1 Nishishinjuku, Shinjuku, Tokyo 163-0217, Japan, Medical ProteoScope Company, Limited, Shinjuku Sumitomo Building 17F, 2-6-1 Nishishinjuku, Shinjuku, Tokyo 163-0217, Japan, Department of Surgery, Tokyo Medical University, 6-7-1 Nishishinjuku, Shinjuku, Tokyo 160-0023, Japan, Clinical Division, Research & Development, AstraZeneca K.K., Umeda Sky Building Tower East, 1-88, 1-chome, Ohyodo-naka, Kita-ku, Osaka 531-0076, Japan, Clinical Science Department, Research & Development, AstraZeneca K.K., Umeda Sky Building Tower East, 1-88, 1-chome, Ohyodo-naka, Kita-ku, Osaka 531-0076, Japan, Department of Medical Oncology, Kinki University School of Medicine, 377-2,Ohno-higashi, Osakasayama-city 589-8511, Osaka, Japan, Department of Respiratory Diseases, Toho University School of Medicine, 6-11-1, Omori-nishi, Ota-ku, Tokyo 143-8541, Japan, Dept. of Thoracic Oncology, National Cancer Centre Hospital East, 6-5-1, Kashiwanoha, Kashiwa-city, Chiba 277-8577, Japan, Statistical Sciences, AstraZeneca R&D Alderley Park, Cheshire, UK, Cancer & Infection Statistics, AstraZeneca R&D Alderley Park, Cheshire, UK, Medicine & Science, AstraZeneca R&D Charnwood, Loughborough LE11 5RH, Leicestershire, UK, Sheffield University, Sheffield, UK, Epidemiology, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden, Institute of Environmental Medicine, Karolinska Institute, Box 210, SE-171 77 Stockholm, Sweden, and 4th Department of Internal Medicine, Nippon Medical School, 1-1-5, Sendagi, Bunkyo-ku, Tokyo 113-8603, Japan

Received January 26, 2007

Personalized medicine allows the selection of treatments best suited to an individual patient and disease phenotype. To implement personalized medicine, effective tests predictive of response to treatment or susceptibility to adverse events are needed, and to develop a personalized medicine test, both high quality samples and reliable data are required. We review key features of state-of-the-art proteomic profiling and introduce further analytic developments to build a proteomic toolkit for use in personalized medicine approaches. The combination of novel analytical approaches in proteomic data generation, alignment and comparison permit translation of identified biomarkers into practical assays. We further propose an expanded statistical analysis to understand the sources of variability between individuals in terms of both protein expression and clinical variables and utilize this understanding in a predictive test.

Keywords: personalized medicine • gefitinib • therapy • interstitial lung disease • non-small cell lung cancer • biomarkers • predictive test • mass spectrometry • statistical analysis • proteomics

* To whom correspondence should be addressed. Epidemiology, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden; Tel. +46 31 706 5203; Fax, +46 31 776 3828; E-mail, Fredrik.Nyberg@astrazeneca.com.

† György Marko-Varga and Atsushi Ogiwara made equal contributions to this manuscript.

‡ Respiratory Biological Sciences, AstraZeneca R&D Lund.

§ Clinical Proteome Center, Tokyo Medical University.

∥ Medical ProteoScope Co., Ltd.

⊥ Department of Surgery, Tokyo Medical University.

Clinical Division, Research & Development, AstraZeneca K.K.

× Clinical Science Department, Research & Development, AstraZeneca K.K.

^ Department of Medical Oncology, Kinki University School of Medicine.

◆ Department of Respiratory Diseases, Toho University School of Medicine.

÷ Dept. of Thoracic Oncology, National Cancer Centre Hospital East.

◊ Statistical Sciences, AstraZeneca R&D Alderley Park.

▽ Cancer & Infection Statistics, AstraZeneca R&D Alderley Park.

∇ Medicine & Science, AstraZeneca R&D Charnwood.

£ Medical School, Sheffield University.

∀ Epidemiology, AstraZeneca R&D Mölndal.

○ Institute of Environmental Medicine, Karolinska Institute.

ƒ Fourth Department of Internal Medicine, Nippon Medical School.

10.1021/pr070046s CCC: $37.00

Published on Web 07/17/2007

Introduction

A personalized medicine approach uses appropriate biomarkers to select treatments best suited for an individual patient and disease phenotype. A multiple biomarker approach (e.g., proteomics) has the advantage over conventional single biomarkers of combining many different pieces of information. Here, we review the key features of state-of-the-art proteomic profiling and introduce recent analytic developments to build a proteomic toolkit for use in personalized medicine, and we describe how these may be applied in a viable method for exploiting predictive proteomic fingerprints in the clinic. The potential of our proteomics toolkit hopefully brings us one step closer to a practical personalized medicine.

Cancer therapy is moving toward individually selected treatments, chosen not only according to tumor cell type but also based on the patient's predicted responsiveness to different classes of therapy or susceptibility to therapeutic adverse events. This emerging personalized medicine approach draws on both genotype and phenotype information, including protein expression. To implement personalized medicine, we need to develop effective biomarker tests predictive of response to treatment or susceptibility to adverse events. The benefits of personalized medicine are exemplified by considering interstitial lung disease (ILD) among non-small cell lung cancer (NSCLC) patients, which is associated with various kinds of chemotherapy treatment. A personalized medicine approach, using a simple blood test to predict those NSCLC patients at risk of developing ILD, would clearly be of great value.

We review current thinking and present some novel developments in a number of areas that have to be integrated to develop and then practically apply such tests in a clinical setting:

- The large scale collection of reliable and high quality phenotypic and clinical data and blood samples.
- Protein analysis in blood.
- Data acquisition, handling, combining and analysis.
- Interpretation and utilization of results in a clinical setting.

Clinical Background

A Motivating Example: Gefitinib (IRESSA) Treatment of NSCLC. The concepts of proteomics-based personalized medicine discussed in this article are very generally applicable. A motivating example that we will refer to in order to illustrate the potential benefits of personalized medicine is ongoing work in attempting to develop a simple blood test to address the potential occurrence of ILD in seriously ill NSCLC patients, the target group for the NSCLC treatment gefitinib.

Gefitinib is a "small molecule" inhibitor of the enzyme tyrosine kinase of the epidermal growth factor receptor (EGFR) family, such as erbB1. It is an approved therapy for advanced NSCLC in many countries and offers important clinical benefits (tumor shrinkage and improvement in disease-related symptoms) for "end-stage" patients. The large phase III ISEL (IRESSA Survival Evaluation in Lung Cancer) trial demonstrated some improvement in survival with gefitinib which failed to reach statistical significance compared with placebo in the overall population and in patients with adenocarcinoma.[1] However, in preplanned subgroup analyses, a significant increase in survival was shown with gefitinib in patients of Asian ethnicity and in patients who had never smoked.[1]

Analysis of the biomarker data from a subset of patients in the ISEL study suggested that patients with pretreated advanced NSCLC who have tumors with a high EGFR gene copy number (detected by fluorescent in situ hybridization [FISH]) have a higher likelihood of increased survival when treated with gefitinib compared with placebo.[2] Increased HER2 gene copy number has also been seen in tumors from patients who are responsive to gefitinib.[3] Somatic-activating mutations of EGFR in tumor tissue have also been associated with increased gefitinib responsiveness in patients with NSCLC.[4-7] Such mutations are more commonly found in tumor samples from patients of Asian origin and non-smokers.[8]

Following the ISEL subgroup analyses, and the biomarker evidence, it has become important to clarify which patients are more suitable for treatment with gefitinib. Analyses for both somatic-activating mutations and gene copy number require tumor tissue, which is not always available from the time of diagnosis; therefore, a blood test may represent a more versatile option and be of great value to clinicians.

With respect to tolerability, the search for a blood test that might include both genetic and proteomic biomarkers to define patients at risk of adverse effects from a drug, for example interstitial lung disease with gefitinib, is a focus of research.

Interstitial Lung Disease as a Complication in NSCLC Patients. ILD is a disease that afflicts the parenchyma or alveolar region of the lungs.[9] The alveolar septa (the walls of the alveoli) become thickened with fibrotic tissue. Associated with drug use, it can present precipitously with acute diffuse alveolar damage (DAD). The lungs show so-called "ground glass" shadowing on chest radiology, and patients complain of severe breathlessness. There are no effective treatments but patients can be supported by oxygen supplementation, corticosteroid therapy, or assisted ventilation. The process of alveolar damage is however fatal in some patients. ILD is a comorbidity in patients with NSCLC.[10-16] Both diseases are associated with cigarette smoking,[17-20] and ILD is also considered to be associated with various kinds of lung cancer chemotherapy.[21-26]

In the ISEL study of gefitinib in NSCLC mentioned above, ILD-type events occurred in 1% of both placebo and gefitinib-treated patients.[1] Most ILD-type events occurred in patients of Asian origin, where placebo and treated patients had similar prevalences of respectively 4% and 3%. The rate observed in the gefitinib-treated arm was in line with earlier safety data from Japan and a large gefitinib post-marketing surveillance study in Japan (3322 patients), where the reported rate of ILD-type events was 5.8%.[27]

A simple blood test to predict the potential occurrence of ILD in seriously ill NSCLC patients before initiating treatments would clearly be of great value. This article describes the personalized medicine approach, which could be used to provide such a test. Consequently, the proteomics objectives of the preliminary phase of the study we describe were to verify the protein expression alterations in blood plasma from case patients (who developed ILD) and control patients (without ILD) treated by gefitinib, using a liquid chromatography–mass spectrometry/mass spectrometry (LC–MS/MS) proteomics platform.

Data and Sample Collection

To develop a personalized medicine test, it is essential to have access to an adequately sized collection of high quality tissue samples on which to perform proteomics analysis, with corresponding reliable diagnostic and clinical data.

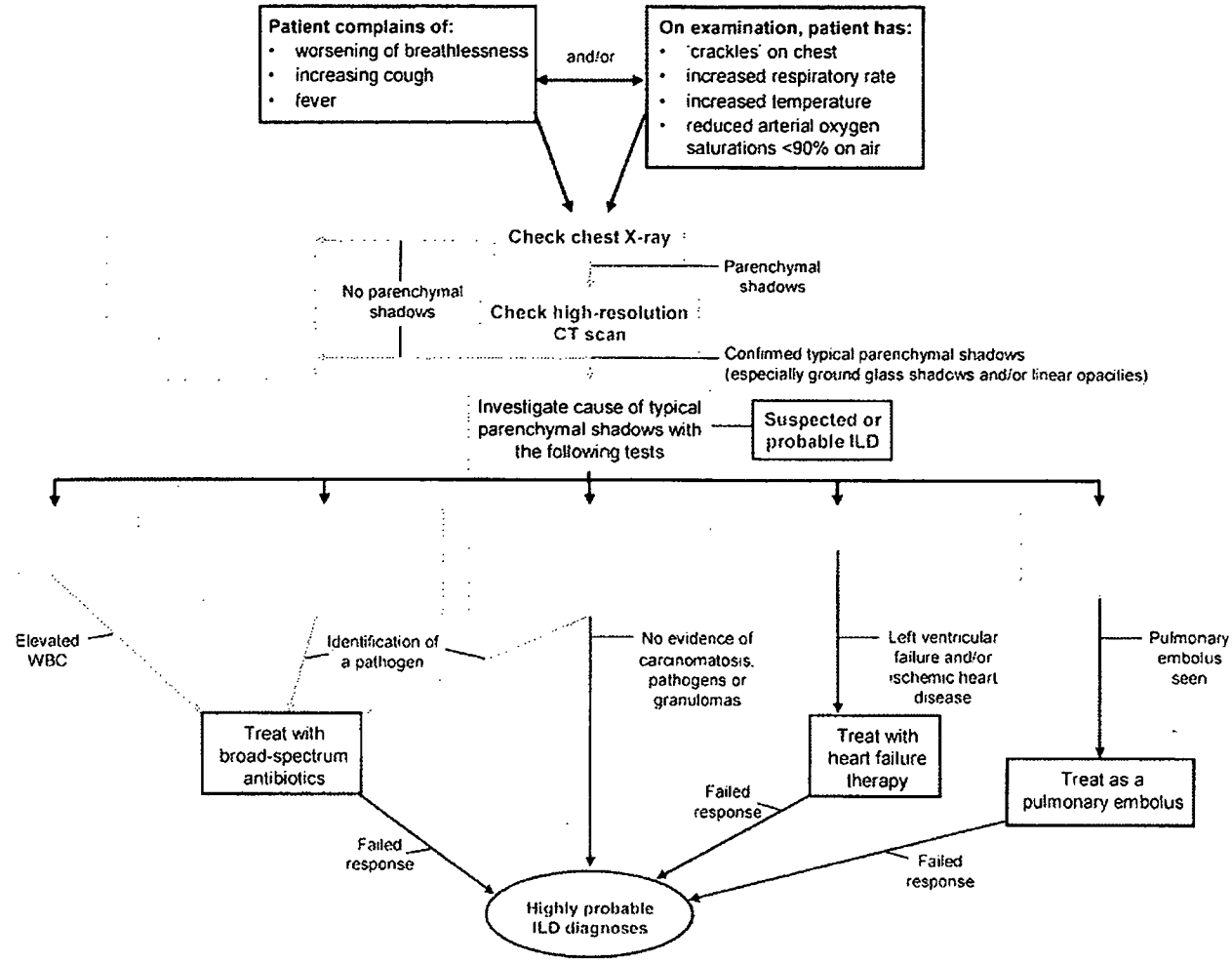

Figure 1. Algorithm for diagnosis of interstitial lung disease (ILD) in non-small cell lung cancer (NSCLC) patients.

As an example, in our work with gefitinib, samples were taken after obtaining informed consent from a nested case-control study, i.e., a case-control study performed within a prospective pharmacoepidemiological cohort of several thousand patients with advanced or recurring NSCLC who had received at least one prior chemotherapy regimen, and who were to be treated with gefitinib or chemotherapy. The main objective of this study was to measure the relative risk of ILD in Japanese patients with NSCLC using gefitinib compared with conventional therapy, with the associated aims of determining the incidence rate of ILD in late stage NSCLC patients and the principal risk factors for this complication.

Central to both the case-control study and the proteomics analysis was the use of internationally agreed criteria for the diagnosis of ILD and an algorithm of diagnostic tests to exclude alternative diseases.[28] Principal investigators in the study were asked to assess all patients for possible ILD using the diagnostic algorithm (Figure 1). Two case review boards of experts from oncology, radiology, and pulmonary medicine were set up to independently establish a consistent final diagnosis of ILD. In addition, extensive standard clinical and demographic risk factor data were collected on all registered cases and controls.

This degree of rigor in establishing accurate phenotypic diagnosis is critical to develop a robust and reliable personalized medicine test, as inaccuracies at this stage will affect all subsequent data analyses. The availability of clinical and risk factor data, and a rigorous epidemiological study design setting for the collection of proteomics samples is also of great value to fine-tune the statistical analysis.

Is Proteomics Ready for Personalized Medicine Applications?

The Human Proteome Map in Plasma. The impetus to develop personalized medicine based on blood samples has encouraged proteomic profiling that identifies individual proteins and multiple "fingerprint" protein patterns. A remaining limitation has been the lack of integration of the technology of protein separation with bioinformatics and statistical methods. Extensive national and international[29,30] collaborations are being implemented to address some of these needs. An important component in this development is the Human Proteome Organization (HUPO; www.HUPO.org), a scientific consortium that supports various programmes to map the proteins expressed in various human tissues, disease states, etc.[31–33] One of these is the Plasma Proteome initiative started in 2002, aiming to annotate and catalog the many thousands of proteins and peptides[34–37] of the human plasma proteome. Recently results from the pilot phase with 35 collaborating laboratories from 13 countries[38–42] and multiple analytical

groups were made publicly available on the Internet (www.bioinformatics.med.umich.edu/hupo/ppp; www.ebi.ac.uk/pride). The combined efforts have generated 15 710 different MS/MS datasets that were linked to the International Protein Index (IPI) protein IDs, and an integration algorithm applied to multiple matches of peptide sequences yielded 9504 IPI proteins identified with one or more peptides[40] and characterized by Gene Ontology, InterPro, Novartis Atlas, and OMIM. Such advances provide an important platform for transforming proteomics from a technology to a useful biomarker tool applicable to personalized medicine.

Protein Analysis in Blood—The Methods. With respect to automated studies, multidimensional chromatography is the main technology used for protein analysis in blood. It is coupled to mass spectrometry either by electrospray ionization (ESI) for analysis in solution or matrix assisted laser desorption/ionization (MALDI) in solid phase applications.[39,41,43-47] Alternatively, ion-trap mass spectrometers are gaining recognition for high-throughput sequencing.[46,48-53] Linking a Fourier transform ion cyclotrone resonance (FTICR) unit to the linear trap can increase the resolution profoundly,[36,54-58] one of several novel principles for strengthening the assignment of protein annotations with the most commonly used protein search engines.[36,47,54-61] For protein annotation, the recent development of a human protein reference database complements these technologies.[61] Studies of protein expression in a variety of biological compartments ranging from sub-cellular to whole organisms have been undertaken with these analytic approaches.[62-70] Some key findings from the HUPO initiatives that impact on methodology include:

- For studies using blood samples, plasma rather than serum is preferred, with ethylenediaminetetraacetic acid (EDTA) as an anticoagulant.[40]
- The abundant proteins in plasma should be depleted prior to analysis.[40]
- Acceptance of protein annotation, i.e., accepted protein identities[39,40] should use standard criteria. These include having two identified peptide sequences from each protein, both with a statistical significance score high enough to ensure a correct sequence confirmation when compared with the corresponding gene sequence entity.[39]

Despite the advances in methodology, important hurdles to using proteomics in a personalized medicine context remain.

Protein Expression Analysis in Blood—Some Important Hurdles. Although protein profiling technology is highly automated and interfaced with database search engines to relate peptide sequences to protein identities and function,[39,40] there are many practical reasons why determining the relative abundance of proteins relevant for prediction purposes is difficult:

- About 90% of proteins are believed to be present only in low copy numbers, i.e., at medium and low abundance levels.[42]
- There can be variation both in the quantity and form of protein expression within normal physiological function.
- Between 300 000 and 3 million human protein species exist as direct gene products or post-translational modifications.[44]
- The relative abundance of the post-translational modifications occurring within the cell is called a Cell-Protein-Index Number (CPIN).[29,30] As an example, if one considers that there are 30 types of phosphorylation variants of a single phosphoprotein, and a hundred possible fold forms of glycosylation of a single glycoprotein, the theoretical CPIN varies considerably depending on the sample complexity.
- The dynamic range of protein expression within cells, between levels of most and least abundant proteins, is in the order of 10^8-10^{10}.[34-36]
- In a typical clinical proteomics study the total cellular protein material in a sample seldom exceeds 10–20 milligrams. Therefore, the least abundant proteins would be present at starting levels not exceeding picograms.
- Recent studies use technology that can identify several thousand proteins in plasma samples,[29] but this still probably only represents a small fraction of the intermediate and processed protein forms. This is due to the current limitation of mass spectrometry not being able to ionize all amino acid sequences and protein modifications with equal efficiency. In most situations, a limited region of the full length protein is sequence annotated.
- The detection of differences in protein expression between groups of interest (e.g., cases and controls) takes place against a background of high variation between individuals within a group, within individuals over time and possible analytic run-to-run variation. Any method used to address this hurdle (which will involve "alignment" for spectral methods) directly impacts the ability to find good protein biomarkers.

Beyond the hurdles above, the fundamental challenge of protein biomarkers is to link the relative abundance of single markers or a fingerprint to clinically important biological processes based on some direct or indirect cause-effect link[29] related to normal or aberrant biological pathways.[17,49] In the following sections, we present the approach used for the identification of protein biomarkers potentially associated with development of ILD in NSCLC patients within the case-control study used as our motivating example. We build on the foundations described above and introduce further analytic developments and ideas relating to proteomic data generation, assaying and alignment to build a proteomics toolkit that can be applied today for personalized medicine approaches.

A State of the Art Clinical Biomarker Analysis System

In the previous section, we described several challenges in proteomic analysis. Here we describe a system and analysis approaches that we have successfully implemented to address some of these issues.

The Components of the Analysis System. The analysis system (Figure 2) uses liquid chromatography-based high-resolution separation of peptides with an interface to tandem MS/MS, a technology which has been attracting great attention as the "shotgun" method of proteome analysis.[44,68-70] With this technology, after depletion of albumin and immunoglobulin G (IgG), all extracted plasma proteins are digested into their specific peptide components by proteolytic enzyme treatment.

The generated peptides are subjected to capillary reverse-phase submicro- to micro-flow liquid chromatography (capillary RP μLC), separated by retention times due to their physicochemical properties, and then detected and sequenced by a linear ion-trap tandem mass spectrometer[71] (LTQ, Thermo Fisher Scientific, San Jose, CA) interfaced with a spray needle tip for ESI of peptides.[70] A two-dimensional quadrupole ion trap mass spectrometer[71] is used, operated in a data-dependent acquisition mode with operational m/z range limits set at 450–2000 (Figure 3, graphs A and B). Automatic switching to MS/MS acquisition mode is made in 1-second scanning cycles, controlled by the XCalibur software. The actual differences between annotated peptide fragment peaks shown in Figure 3, graph C, correspond to the amino acid residue mass, i.e.,

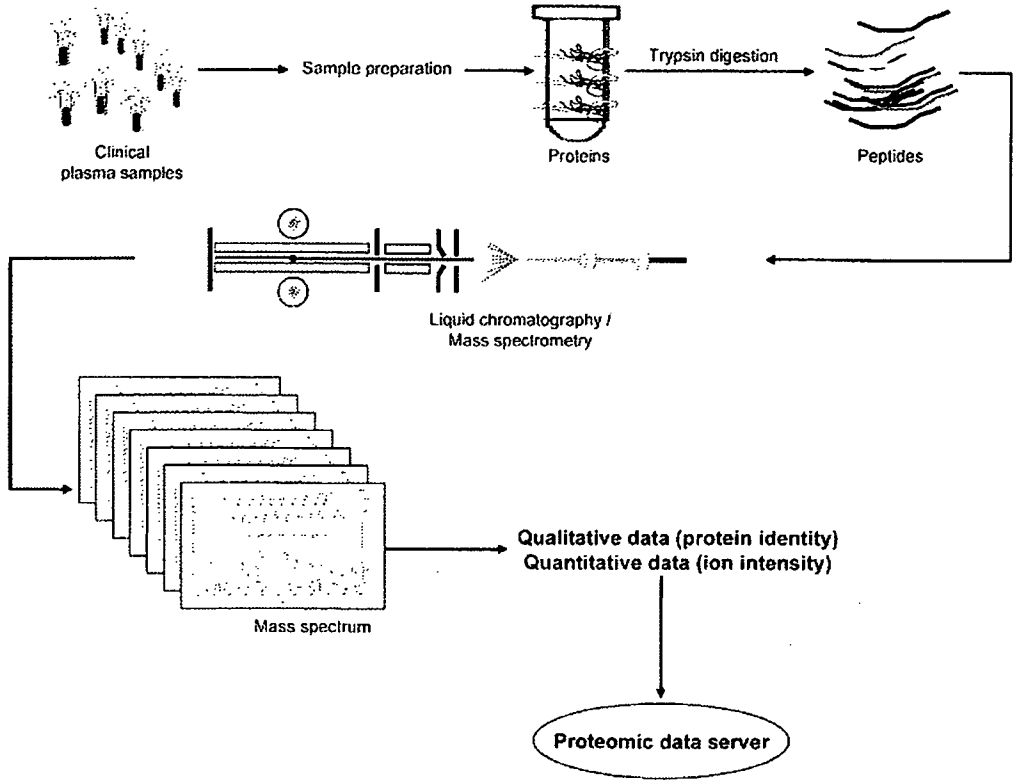

Figure 2. Schematic illustration of the clinical proteomics screening process.

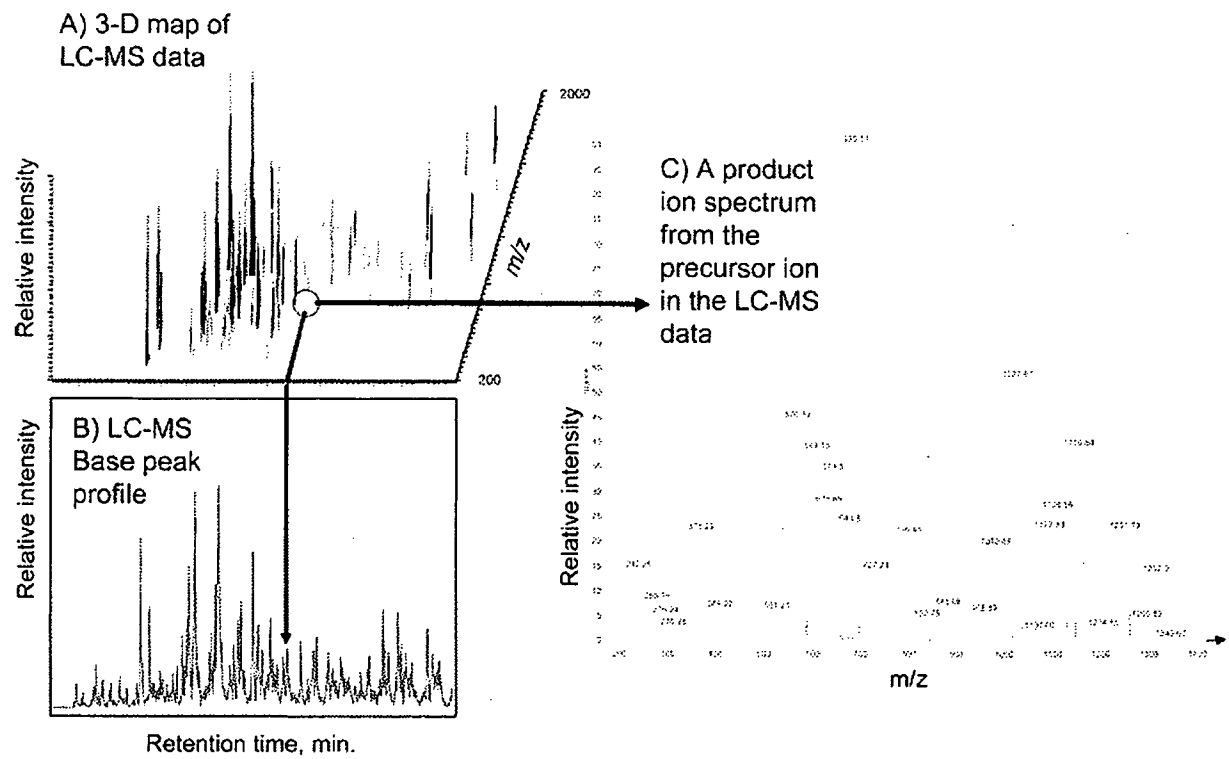

Figure 3. Profile of LC–MS data: (a) the three-dimensional view of LC-MS data, (b) the base-peak mass chromatogram, and (c) a product ion spectrum measured for a precursor ion in data-dependent acquisition mode (with MS acquisition operational *m/z* range set at 450–2000).

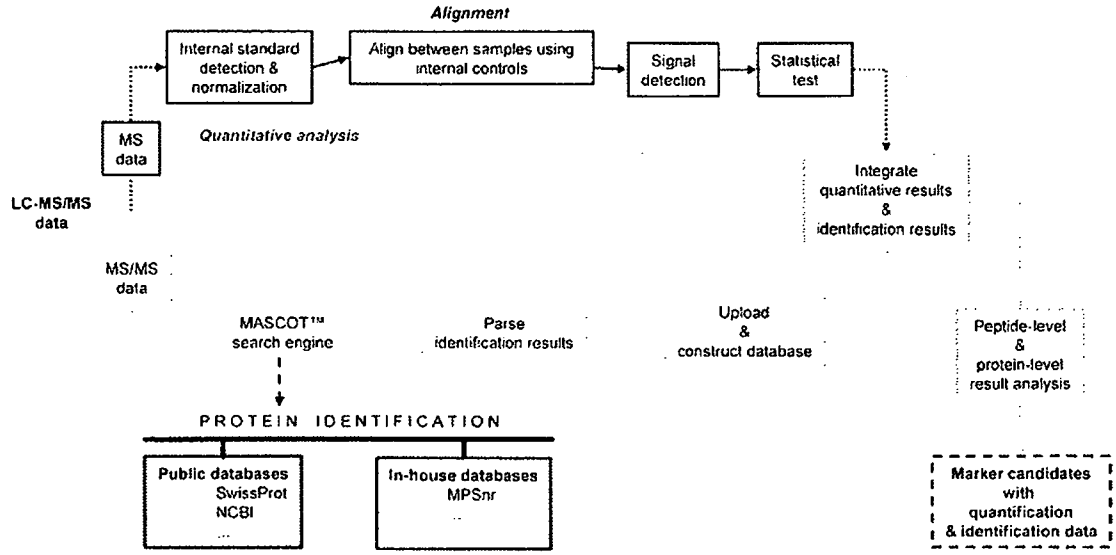

Figure 4. Overview of the data acquisition and database mining process developed within the gefitinib biomarker study.

identify the correct amino acid sequence. Internal standards are used for alignment of retention-times.

How the Methodology Overcomes Some of the Hurdles. The system described above addresses some of the hurdles noted previously. The digestion of all extracted plasma proteins into peptides will reduce the complexity by combining high-resolution nanoflow chromatographic fractionation with the separation power of modern mass spectrometry, performing automated and unattended shotgun sequencing in plasma.[35] Peptides are also more soluble and easier to handle than intact proteins. In addition, the two-dimensional quadrupole ion trap mass spectrometer[71] operates with a high-volume quadrupole electric field that makes it highly efficient to trap ions. The result is high sensitivity, high scanning speed, and better quantification over a wide dynamic range in comparison with the conventional three-dimensional ion-trap instruments.

Finding signals against a background of high variation is a further challenge, and the next section describes some approaches for addressing these.

Initial Data Handling, Processing, and Analysis

Proteomic data analysis process can be considered as consisting of two components (Figure 4). *Quantitative analysis* is used to discover significant differences in peptide signal intensities by comparing two (or more) sample groups. This process uses data collected from an entire MS run to quantify the amount of peptide ions by their respective ion signal intensity. *Qualitative analysis* is used to identify the amino acid sequence of each peptide ion, from the respective product ion spectra. To maximize their value, the results from the two component analyses should be considered in combination.

A typical quantitative analysis may consist of several steps:

1. Normalization: To account for differences in the original sample concentrations. Typically, the total signal intensity is scaled to a constant value for each analyzed sample.

2. Alignment: Correcting for nonlinear fluctuation in retention time between different samples. A variety of methodologies are available for aligning LC–MS data sets. We have found the i-OPAL algorithm (Patent # WO 2004/090526 AI), which is based on the single linkage clustering algorithm[72] and which makes use of internal standard signals, to perform well. Other alignment algorithms include xcms.[73]

3. Peak picking or signal detection: Identifying individual peptide ions within the data.

4. Identify discriminating peptides: A number of methods can be used, often in combination. A common approach is to apply a Student's *t*-test and select peptides which are significant, i.e., with a *p*-value less than the chosen cutoff, and which also show a fold-change or intensity ratio greater than another criterion. Further developments of this aspect are discussed in the Principled Statistical Analysis section.

A popular choice for qualitative analysis is the MASCOT MS/MS ion search program.[74] This may be run against a number of different peptide sequence databases, for example the NCBI Nr, Refseq, Gene Ontology, HUGO, and Swiss-Prot sequence databases. The results of the quantitative analysis can then be combined with the qualitative analysis so that, for example, a peptide must be both discriminating and have annotation—i.e., have achieved a high MASCOT score showing confidence in identification—to be considered a candidate biomarker.

The approaches we have discussed above are focused on finding potentially discriminating proteins of clinical utility. In the following section, we describe the next stage in our thinking, namely how we could rapidly deploy in the clinic a viable method for exploiting a predictive proteomic fingerprint.

A Proposal for Proteomics in the Clinical Setting: Mass Spectrometric Biomarker Assays - MSBA

Although today's technology allows for high-throughput analyses of many proteins rather than a single protein,[30] the details of how such multiplexing assays will be adapted for clinical use have not been well clarified. The Mass Spectrometric Biomarker Assay (MSBA) platform described here was conceived as one example of a rapid and seamless method to progress from identification of a diagnostic more directly to a clinically useful test. MSBA requires only a minute sample amount (5–20 μL) to obtain a read-out from a handful of quantified protein biomarkers (typically 3–35) and automatically analyzes proteins using liquid-phase separation and tandem mass spectrometry with simultaneous quantitation and identification.

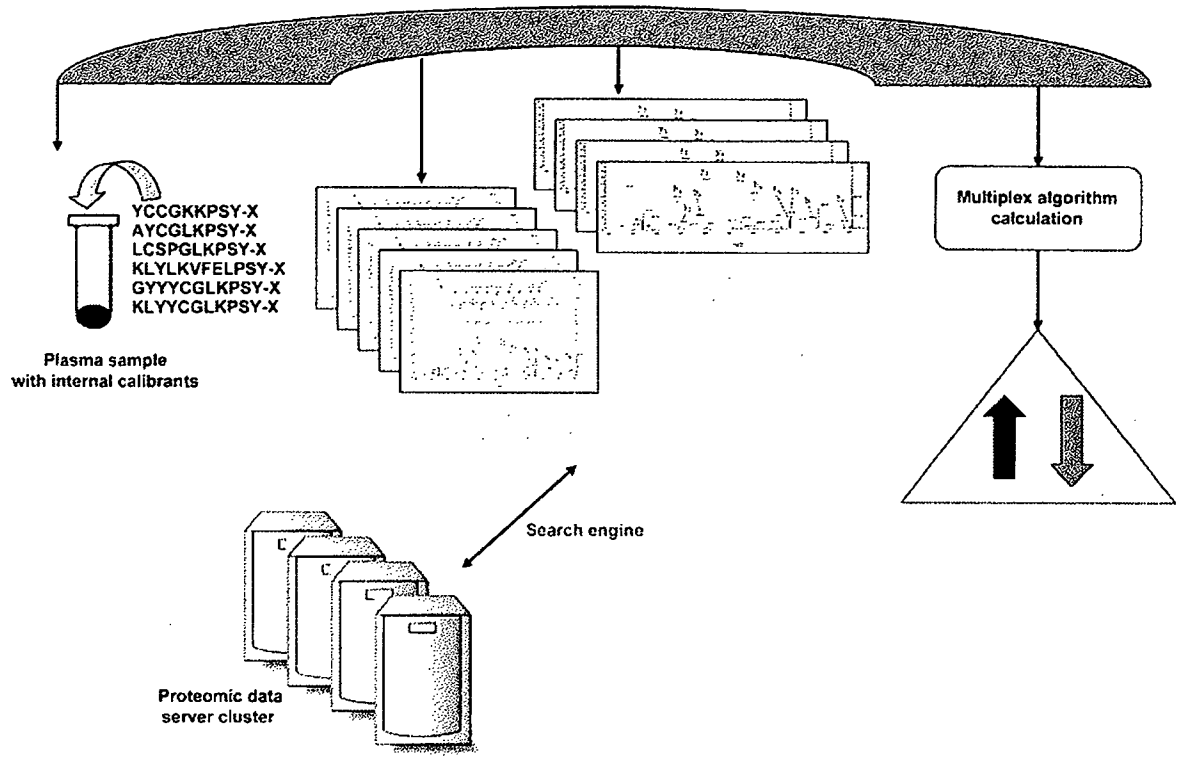

Figure 5. Entire flow of the operational components of Mass Spectrometric Biomarker Assays (MSBA).

The MSBA builds on a pre-defined Multiplex Biomarker list, which is stored within the MSBA database. Each marker entity has the values of masses and the relative retention time index with tolerance parameters. In running a patient sample, the predefined biomarker list is scanned to pick up patient sample signals that match with one of the predefined biomarker signals by satisfying the tolerance criteria (in general ±1 for m/z value and ±2% for relative retention time index). The selected candidate signals are further confirmed using the product ion spectrum. That is, the product ion spectrum is represented as a vector by binning (grouping) the m/z ratio values. Using the cosine correlation between the sample vectors and the reference vectors, we can confirm whether the selected candidate signals are truly assigned as target biomarkers. (A standard threshold value of the cosine correlation is 0.8.)

The process steps within the MSBA cycle are outlined in Figure 5. The calculation of the final multiplex biomarker assay read-out from all of the individual markers can be performed by a variety of applications, as discussed in more detail in the Principled Statistical Modeling Approach section. Figures 6A and B illustrate one approach, calculating a distance score which indicates to what extent a measured sample is distant from the case or control template in terms of predefined multiplex biomarkers.

$$S_{\substack{\text{case or}\\ \text{control}}} = \sqrt{\left[\frac{1}{n(n-2)}\right]\left[n\sum_i y_i^2 - \left(\sum_i y_i\right)^2 - \frac{\left[n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right)\right]^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\right]}$$

If the ratio of S_{case} and S_{control} exceeds an MSBA threshold parameter, then the test sample is predicted to be a patient susceptible to develop ILD (ILD case); if not, the test sample is predicted to be a non-susceptible patient (control). We are currently evaluating the MSBA approach in practice.

A Principled Statistical Modeling Approach

We have described an analytical approach based on proteomic data, with various novel developments. However, additional insight is needed to further improve model discrimination and to broaden the focus from the proteomic data to the ultimate goal of prediction using combinations of data. Statistical analysis can be used to provide further refinement by combining information from the full clinical and laboratory datasets.

An advantage of a multiple biomarker approach (e.g., proteomics) compared with standard single biomarker development is the capability to combine information from many different entities. An example is illustrated in Figure 7A. Considering each biomarker alone fails to separate the two groups of subjects, as there is considerable overlap for both biomarkers. Use of two biomarkers in combination completely separates the two groups.

We can also use clinical variables to advantage in the analysis of the peptide patterns. For example, the efficacy of gefitinib appears to be greater in non-smokers, women, patients of Asian origin, and patients with adenocarcinomas.[8] Figure 7B illustrates how, instead of two protein biomarkers, the combination of clinical data (e.g., age) and a proteomic biomarker is able to separate two groups.

On this basis, we propose using a principled statistical analysis approach to first explore and understand the data and

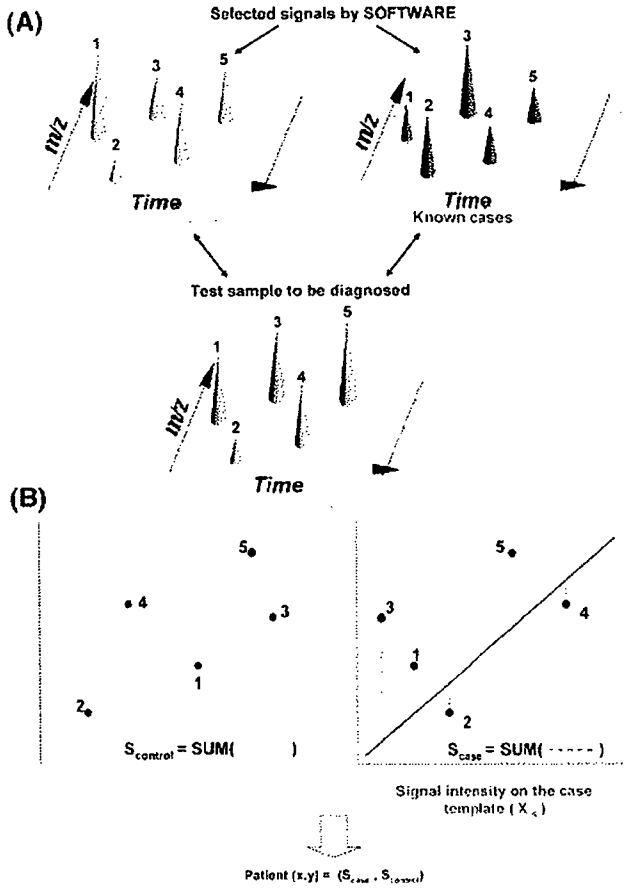

Figure 6. (A) Peptide signal comparison that MSBA (Mass Spectrometric Biomarker Assays) performs of the generated ions from the sample. The comparison is made both with the pattern of the controls and with the pattern of the case group for the corresponding signals. (B) Illustration of the regression model application of the MSBA where control templates and case templates are compared to that of the sample template generated in the analysis process.

then to model it and understand the quality of any models produced. A first step is to perform exploratory data analysis (EDA), for example using principal components analysis (PCA), to understand the major sources of data variation and the covariation between clinical parameters and protein intensity measures. The next step is univariate modeling for each protein marker individually, for example using analysis of covariance (ANCOVA), and an assessment of the effect of clinical parameters across the whole set of protein biomarkers using, for example, the False Discovery Rate as a tool.[75] This provides an understanding of key clinical variables and sources of variation within the data.

The next step is to perform multivariate predictive modeling using the proteins and clinical variables identified as being potentially important. There are a number of mathematical methods described in the literature for performing supervised classification, for example Support Vector Machines,[76] Random Forests,[77] PAM,[78] all of which have been successfully applied to high dimensional genomics data.[79] It remains an important unanswered question which modeling approach, or combination of modeling approaches, will generate the most predictive and robust models for data generated using this technology within a prospective study of this design.

Finally, to confirm that we have a practical prediction, the predictive power of a model must be assessed on a different set of patients from that used to generate the model. There are a number of approaches for external validation given a limited size dataset, for example the sequential approach of building a model based upon currently available data and testing on data from new patients when they become available, or withholding an arbitrary selection of subjects from the modeling as a test set and testing the model on these subjects. Internal validation approaches such as cross-validation or related bootstrapping methods may also be useful to assess the model selection *procedure*, but tend to overestimate the performance of a specific predictive model in subsequent external validation.[80,81] The key properties to consider when selecting an assessment method are to ensure that it will provide both precise and unbiased information regarding the prediction error rate of the potential model to be tested for clinical use. As well as assessing an overall predictive rate, it is also useful to separately assess the predictive rate for both the cases and controls and to consider the relative costs of making these false predictions within a clinical setting. Finally, the prevalence of the condition in question (here ILD) is also a critical factor in estimating what proportion of people predicted to be at risk are truly at risk, and this should also be borne in mind when evaluating a model for potential clinical use. The recently published FDA concept paper on drug-diagnostic co-

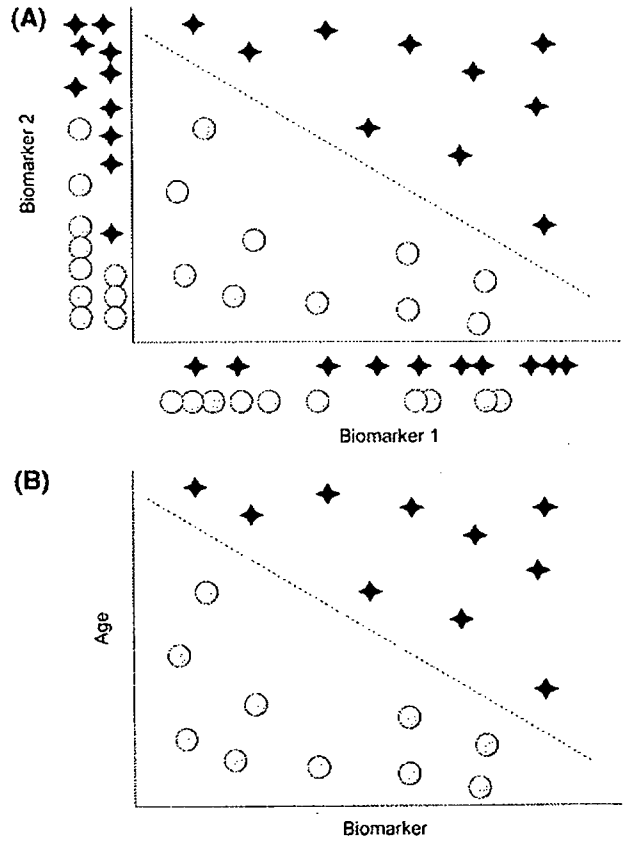

Figure 7. (A) Hypothetical example of the combined disease-linkage effect of two protein biomarkers. (Stars signify affected case individuals, circles non-affected control individuals). (B) Hypothetical example of the combined disease-linkage effect of a biomarker and a clinical variable. (Stars signify affected case individuals, circles non-affected control individuals).

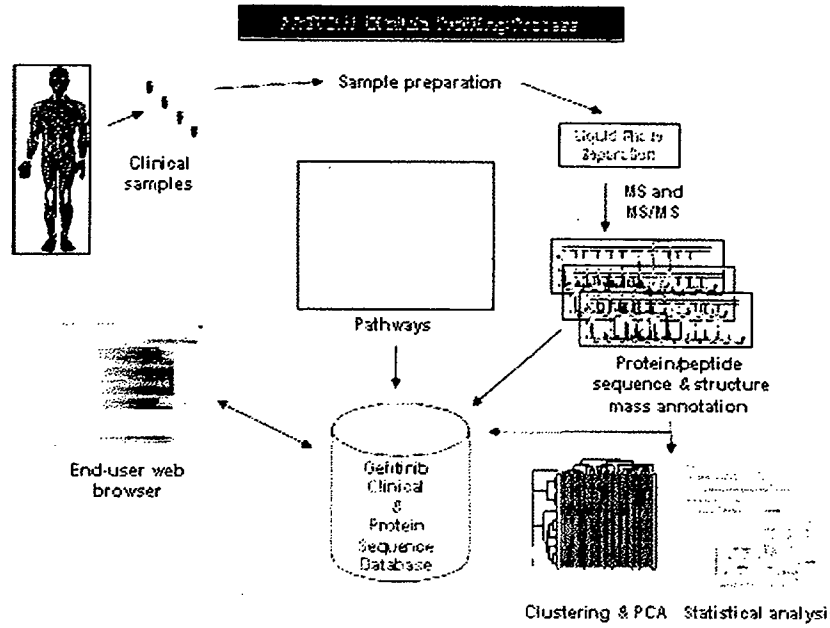

Figure 8. Illustration of the bioinformatics and data processing structure within which MSBA (Mass Spectrometric Biomarker Assays) data are captured, modified and analyzed.

development discusses many of the issues around validating predictive biomarkers.[82]

Finally, it is preferable to be able to assign a biological rationale to the biomarkers. Confidence in the reliability of a biomarker is greatly enhanced if we can correctly understand how it relates to the mechanism and progression of the disease of interest. Figure 8 illustrates a bioinformatics and data processing structure that we have developed to allow us to both conduct interactive exploratory and statistical analyses, and also investigate the disease and pathway linkage of discovered biomarker proteins through direct access to reference databases.

Future Perspectives

Within this paper we have discussed many of the issues that need to considered in developing a personalized medicine approach. A key starting point is that rigorous steps are taken to ensure accurate diagnosis and the careful gathering of both clinical and proteomic data to facilitate the search for peptide patterns.

There are many challenges in performing protein analysis in blood, but mass spectrometry equipment and methods can now be used to generate peptide data with high sensitivity, high scanning speed, and improved quantification. Data handling and processing techniques for steps such as peak alignment and the subsequent methodologies for statistical modeling and analysis are now far enough developed to generate high quality data and robustly analyze these data with confidence.

We have provided details of the MSBA method that can be used to easily translate protein intensities into a practical multiplex assay which can be exploited in the clinic without the need to develop anti-bodies for ELISA. We have also described how an expanded statistical analysis can be used to allow for the individual variance of protein expression to enable us to focus on the proteomic patterns that are actually related to ILD. Finally, we have emphasized the importance of validating the predictive power of a biomarker tool in a way that reflects the real-life setting of intended clinical use.

Hopefully, this combination of developments over a range of different areas brings us one step closer to a practical personalized medicine.

IRESSA is a trademark of the AstraZeneca group of companies.

Acknowledgments. We thank all involved in the Iressa study which provided the inspiration for this overview of personalized medicine approaches, including: the external Epidemiology Advisory Board (Kenneth J. Rothman, Jonathan M. Samet, Toshiro Takezaki, Kotaro Ozasa, Masahiko Ando) for their advice and scientific review of study design, conduct, and analysis; Professor Nestor Müller for his expert input into radiological aspects of ILD diagnosis; all Case Review Board members individually (M. Suga, T. Johkoh, M. Takahashi, Y. Ohno, S. Nagai, Y. Taguchi, Y. Inoue, T. Yana, M. Kusumoto, H. Arakawa, A. Yoshimura, M. Nishio, Y. Ohe, K. Yoshimura, H. Takahashi, Y. Sugiyama, M. Ebina) for their valuable work in blindly reviewing ILD diagnoses, as well as pre-study CT scans for pre-existing comorbidities, the Japan Thoracic Radiology Group, Shiga, Japan for their support of CRB work; and all Hospitals, Clinical Investigators, study monitors, nurses, data managers, other support staff, and the participating patients for providing and collecting the data in the study.

References

(1) Thatcher, N.; Chang, A.; Parikh, P.; Pereira, J. R.; Ciuleanu, T.; von Pawel, J.; et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* **2005**, *366*, 1527–1537.

(2) Hirsch, F. R.; Varella-Garcia, M.; McCoy, J.; West, H.; Xavier, A. C.; Gumerlock, P.; et al. Increased epidermal growth factor receptor gene copy number detected by fluorescence in situ hybridization associates with increased sensitivity to gefitinib in patients with bronchioloalveolar carcinoma subtypes: a Southwest Oncology Group study. *J. Clin. Oncol.* **2005**, *23*, 6838–6845.

(3) Cappuzzo, F.; Varella-Garcia, M.; Shigematsu, H.; Domenichini, I.; Bartolini, S.; Ceresoli, G. L.; et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J. Clin. Oncol.* **2005**, *23*, 5007−5018.
(4) Araki, J.; Okamoto, I.; Suto, R.; Ichikawa, Y.; Sasaki, J. Efficacy of the tyrosine kinase inhibitor gefitinib in a patient with metastatic small cell lung cancer. *Lung Cancer* **2005**, *48*, 141−144.
(5) Kim, K. S.; Jeong, J. Y.; Kim, Y. C.; Na, K. J.; Kim, Y. H.; Ahn, S. J.; et al. Predictors of the response to gefitinib in refractory non-small cell lung cancer. *Clin. Cancer Res.* **2005**, *11*, 2244−2251.
(6) Lynch, T. J.; Bell, D. W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R. A.; Brannigan, B. W.; 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.
(7) Paez, J. G.; Jänne, P. A.; Lee, J. C.; Tracy, S.; Greulich, H.; Gabriel, S.; et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **2004**, *304*, 1497−1500.
(8) Shigematsu, H.; Lin, L.; Takahashi, T.; Nomura, M.; Suzuki, M.; Wistuba II; et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* **2005**, *97*, 339−346.
(9) American Thoracic Society: American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS Board of Directors, June 2001 and by The ERS Executive Committee, June 2001. *Am. J. Respir. Crit. Care Med.* **2002**, *165*, 277−304.
(10) Raghu, G.; Nyberg, F.; Morgan, G. The epidemiology of interstitial lung disease and its association with lung cancer. *Br. J. Cancer* **2004**, *91* (Suppl. 2), S3−S10.
(11) Asada, K.; Mukai, J.; Ougushi, F. Characteristics and management of lung cancer in patients with idiopathic pneumonia. *Jap. J. Thor. Dis.* **1992**, *51*, 214−219.
(12) Hubbard, R.; Venn, A.; Lewis, S.; Britton, J. Lung cancer and cryptogenic fibrosing alveolitis. A population-based cohort study. *Am. J. Respir. Crit. Care Med.* **2000**, *161*, 5−8.
(13) Matsushita, H.; Tanaka, S.; Saiki, Y.; Hara, M.; Nakata, K.; Tanimura, S.; et al. Lung cancer associated with usual interstitial pneumonia. *Pathol. Int.* **1995**, *45*, 925−932.
(14) Ogura, T.; Kondo, A.; Sato, A.; Ando, M.; Tamura, M. Incidence and clinical features of lung cancer in patients with idiopathic interstitial pneumonia. *Nihon Kyobu Shikkan Gakkai Zasshi* **1997**, *35*, 294−299.
(15) Takeuchi, E.; Yamaguchi, T.; Mori, M.; Tanaka, S.; Nakagawa, M.; Yokota, S.; et al. Characteristics and management of patients with lung cancer and idiopathic interstitial pneumonia. *Nihon Kyobu Shikkan Gakkai Zasshi* **1996**, *34*, 653−658.
(16) Turner-Warwick, M.; Lebowitz, M.; Burrows, B.; Johnson, A. Cryptogenic fibrosing alveolitis and lung cancer. *Thorax* **1980**, *35*, 496−499.
(17) Baumgartner, K. B.; Samet, J. M.; Stidley, C. A.; Colby, T. V.; Waldron, J. A. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **1997**, *155*, 242−248.
(18) Britton, J.; Hubbard, R. Recent advances in the aetiology of cryptogenic fibrosing alveolitis. *Histopathology* **2000**, *37*, 387−392.
(19) Iwai, K.; Mori, T.; Yamada, N.; Yamaguchi, M.; Hosoda, Y. Idiopathic pulmonary fibrosis. Epidemiologic approaches to occupational exposure. *Am. J. Respir. Crit. Care Med.* **1994**, *150*, 670−675.
(20) Nagai, S.; Hoshino, Y.; Hayashi, M.; Ito, I. Smoking-related interstitial lung diseases. *Curr. Opin. Pulm. Med.* **2000**, *6*, 415−419.
(21) Lilly. Gemcitabine prescribing information. http://pi.lilly.com/gemzar.pdf, 2003.
(22) Kunitoh, H.; Watanabe, K.; Onoshi, T.; Furuse, K.; Niitani, H.; Taguchi, T. Phase II trial of docetaxel in previously untreated advanced non-small-cell lung cancer: a Japanese cooperative study. *J. Clin. Oncol.* **1996**, *14*, 1649−1655.
(23) Merad, M.; Le Cesne, A.; Baldeyrou, P.; Mesurolle, B.; Le Chevalier, T. Docetaxel and interstitial pulmonary injury. *Ann. Oncol.* **1997**, *8*, 191−194.
(24) Wang, G.-S.; Yan, K.-Y.; Perng, R.-P. Life-threatening hypersensitivity pneumonitis induced by docetaxel (taxotere). *Br. J. Cancer* **2001**, *85*, 1247−1250.
(25) Erasmus, J. J.; McAdams, H. P.; Rossi, S. E. Drug-induced lung injury. *Semin. Roentgenol.* **2002**, *37*, 72−81.
(26) Aviram, G.; Yu, E.; Tai, P.; Lefcoe, M. S. Computed tomography to assess pulmonary injury associated with concurrent chemoradiotherapy for inoperable non-small cell lung cancer. *Can. Assoc. Radiol. J.* **2001**, *52*, 385−391.
(27) Yoshida, S. The results of gefitinib prospective investigation. *Med. Drug J.* **2005**, *41*, 772−789.
(28) Mueller, N. L.; White, D. A.; Jiang, H.; Gemma, A. Diagnosis and management of drug-associated interstitial lung disease. *Br. J. Cancer* **2004**, *91*, S24−S30.
(29) Marko-Varga, G.; Fehniger, T. E. Proteomics and disease−the challenges for technology and discovery. *J. Proteome Res.* **2004**, *3*, 167−178.
(30) Marko-Varga, G.; Lindberg, H.; Lofdahl, C. G.; Jonsson, P. H. L.; Dahlback, M.; Lindquist, E.; et al. Discovery of biomarker candidates within disease by protein profiling: principles and concepts. *J. Proteome Res.* **2005**, *4*, 1200−1212.
(31) Omenn, G. S. The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. *Proteomics* **2004**, *4*, 1235−1240.
(32) Omenn, G. S. Advancement of biomarker discovery and validation through the HUPO plasma proteome project. *Dis. Markers* **2004**, *20*, 131−134.
(33) Orchard, S.; Hermjakob, H.; Binz, P. A.; Hoogland, C.; Taylor, C. F.; Zhu, W.; et al. Further steps towards data standardisation: the Proteomic Standards Initiative HUPO 3(rd) annual congress, Beijing 25−27(th) October, 2004. *Proteomics* **2005**, *5*, 337−339.
(34) Anderson, N. G.; Matheson, A.; Anderson, N. L. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **2001**, *1*, 3−12.
(35) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845−867.
(36) Jacobs, J. M.; Adkins, J. N.; Qian, W. J.; Liu, T.; Shen, Y.; Camp, D. G.; et al. Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* **2005**, *4*, 1073−1085.
(37) Anderson, N. G.; Anderson, L. The Human Protein Index. *Clin. Chem.* **1982**, *28*, 739−748.
(38) Haab, B. B.; Geierstanger, B. H.; Michailidis, G.; Vitzthum, F.; Forrester, S.; Okon, R.; et al. Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* **2005**, *5*, 3278−3291.
(39) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; et al. PRIDE: the proteomics identifications database. *Proteomics* **2005**, *5*, 3537−3545.
(40) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **2005**, *5*, 3226−3245.
(41) Patterson, S. D. Data analysis−the Achilles heel of proteomics. *Nat. Biotechnol.* **2003**, *21*, 221−222.
(42) Rabbar, A. M.; Fenselau, C. Integration of Jacobson's pellicle method into proteomic strategies for plasma membrane proteins. *J. Proteome Res.* **2004**, *3*, 1267−1277.
(43) Ho, Y.; Gruhler, A., Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002**, *415*, 180−183.
(44) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198−207.
(45) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **2004**, *3*, 311−326.
(46) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417−13422.
(47) Sadygov, R. G.; Liu, H.; Yates, J. R. Statistical models for protein validation using tandem mass spectral data and protein amino acid sequence databases. *Anal. Chem.* **2004**, *76*, 1664−1671.
(48) Fujii, K.; Nakano, T.; Kanazawa, M.; Akimoto, S.; Hirano, T.; Kato, H.; et al. Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics* **2005**, *5*, 1150−1159.

(49) Campbell, J. M.; Collings, B. A.; Douglas, D. J. A new linear ion trap time-of-flight system with tandem mass spectrometry capabilities. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1463–1474.

(50) Cha, B. C.; Blades, M.; Douglas, D. J. An interface with a linear quadropole ion guide for an electrospray-ion trap mass spectrometer system. *Anal. Chem.* **2000**, *72*, 5647–5654.

(51) Hager, J. W. Product ion spectral simplification using time-delayed fragment ion capture with tandem linear ion traps. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1389–1398.

(52) Syka, J. E.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **2004**, *3*, 621–626.

(53) Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; et al. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* **2001**, *73*, 1766–1775.

(54) Wu, S. L.; Kim, J.; Hancock, W. S.; Karger, B. Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* **2005**, *4*, 1155–1170.

(55) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010–2021.

(56) Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* **2006**, *78*, 493–500.

(57) Anderson, D. C.; Li, W.; Payan, D. G.; Noble, W. S. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J. Proteome Res.* **2003**, *2*, 137–146.

(58) Carr, S.; Aebersold, R.; Baldwin, M.; Burlingame, A.; Clauser, K.; Nesvizhskii, A. The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data. *Mol. Cell. Proteomics* **2004**, *3*, 531–533.

(59) Fenyo, D.; Beavis, R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **2003**, *75*, 768–774.

(60) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658.

(61) Peri, S.; Navarro, J. D.; Kristiansen, T. Z.; Amanchy, R.; Surendranath, V.; Muthusamy, B.; et al. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **2004**, *32*, D497–D501.

(62) Kratchmarova, I.; Blagoev, B.; Haack-Sorensen, M.; Kassem, M.; Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **2005**, *308*, 1472–1477.

(63) Dreger, M.; Bengtsson, L.; Schöneberg, T.; Otto, H.; Hucho, F. Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11943–11948.

(64) Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; et al. A protein interaction map of Drosophilia melanogaster. *Science* **2003**, *302*, 1727–1736.

(65) Johnson, J. R.; Florens, L.; Carucci, D. J.; Yates, J. R., III. Proteomics in malaria. *J. Proteome Res.* **2004**, *3*, 296–306.

(66) Hirsch, J.; Hansen, K. C.; Burlingame, A. L.; Matthay, M. A. Proteomics: current techniques and potential applications to lung disease. *Am. J. Physiol. Lung Cell Mol. Physiol.* **2004**, *287*, L1–123.

(67) Malmström, J.; Larsen, K.; Hansson, L.; Löfdahl, C.-G.; Norregard-Jensen, O.; Marko-Varga, G.; et al. Proteoglycan and proteome profiling of central human pulmonary fibrotic tissue utilizing minaturized sample preparation: A feasibility study. *Proteomics* **2002**, *2*, 394–404.

(68) Malmström, J.; Larsen, K.; Malmström, L.; Tufvesson, E.; Parker, K.; Marchese, J.; et al. Proteome annotations and identifications of the human pulmonary fibroblast. *J. Proteome Res.* **2004**, *3*, 525–537.

(69) Oh, P.; Li, Y.; Yu, J.; Durr, E.; Krasinska, K. M.; Carver, L. A.; et al. Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* **2004**, *429*, 629–635.

(70) Fujii, K.; Nakano, T.; Kawamura, T.; Usui, F.; Bando, Y.; Wang, R.; et al. Multidimensional protein profiling technology and its application to human plasma proteome. *J. Proteome Res.* **2004**, *3*, 712–718.

(71) Schwartz, J. C.; Senko, M. W.; Syka, J. E. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659–669.

(72) Sneath, P. H. A.; Sokal, R. R. *Numerical Taxonomy. The principles and practice of numerical classification*; W. H. Freeman and Co.: San Francisco, 1973.

(73) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, *78*, 779–787.

(74) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence-databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.

(75) Storey, J. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 479.

(76) Vapnik, V. *Statistical Learning Theory*; Wiley: Chichester, UK, 1998.

(77) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

(78) Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6567–6572.

(79) Lee, J. W.; Lee, J. B.; Park, M.; Song, S. H. An extensive comparison of recent classification tools applied to microarray data. *Comp. Stat. Data Anal.* **2005**, *48*, 869–885.

(80) Steyerberg, E. W.; Harrell, F. E., Jr.; Borsboom, G. J.; Eijkemans, M. J.; Vergouwe, Y.; Habbema, J. D. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* **2001**, *54*, 774–781.

(81) Bleeker, S. E.; Moll, H. A.; Steyerberg, E. W.; Donders, A. R.; Derksen-Lubsen, G.; Grobbee, D. E.; et al. External validation is necessary in prediction research: a clinical example. *J. Clin. Epidemiol.* **2003**, *56*, 826–832.

(82) Food and Drug Administration (FDA): Drug-diagnostic co-development concept paper. Draft—not for implementation. http://www.fda.gov/cder/genomics/pharmacoconceptfn.pdf, 2005.

PR070046S

VOLUME 25 · NUMBER 1 · JANUARY 1 2007

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Pharmacokinetics of Gemcitabine in Japanese Cancer Patients: The Impact of a Cytidine Deaminase Polymorphism

Emiko Sugiyama, Nahoko Kaniwa, Su-Ryang Kim, Ruri Kikura-Hanajiri, Ryuichi Hasegawa, Keiko Maekawa, Yoshiro Saito, Shogo Ozawa, Jun-ichi Sawada, Naoyuki Kamatani, Junji Furuse, Hiroshi Ishii, Teruhiko Yoshida, Hideki Ueno, Takuji Okusaka, and Nagahiro Saijo

From the Project Team for Pharmacogenetics; Divisions of Medicinal Safety Sciences, Pharmacognosy and Phytochemistry, Biochemistry and Immunochemistry, and Pharmacology, National Institute of Health Sciences; Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University; Genetics Division, Research Institute, National Cancer Center, Tokyo; and Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital East, Kashiwa, Japan.

Submitted March 23, 2006; accepted August 25, 2006.

Supported by the Program for the Promotion of Fundamental Studies in Health Sciences (Grant No. MPJ6 and 05-25), and the Health and Labour Sciences Research Grant on Human Genome and Tissue Engineering (Grant No. H16-Genome-008) from the Ministry of Health, Labour, and Welfare of Japan.

Presented in part at the 41st Annual Meeting of the American Society of Clinical Oncology, May 13-17, 2005, Orlando, FL, and at the 13th Annual Meeting of the North American Society for the Study of Xenobiotics, October 22-27, 2005, Maui, HI.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Nahoko Kaniwa, PhD, Division of Medicinal Safety Sciences, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan; e-mail: nkaniwa@nihs.go.jp.

0732-183X/07/2501-32/$20.00

DOI: 10.1200/JCO.2006.06.7405

ABSTRACT

Purpose
Gemcitabine is rapidly metabolized to its inactive metabolite, 2′,2′-difluorodeoxyuridine (dFdU), by cytidine deaminase (CDA). We previously reported that a patient with homozygous 208A alleles of *CDA* showed severe adverse reactions with an increase in gemcitabine plasma level. This study extended the investigation of the effects of *CDA* genetic polymorphisms on gemcitabine pharmacokinetics and toxicities.

Patients and Methods
Genotyping of *CDA* was performed by a direct sequencing of DNA obtained from the peripheral blood of Japanese gemcitabine-naïve cancer patients (n = 256). The patients recruited to the association study received a 30-minute intravenous infusion of gemcitabine at a dose of either 800 or 1,000 mg/m^2, and eight blood samples were periodically collected (n = 250). Plasma levels of gemcitabine and dFdU were measured by high-performance liquid chromatography. Plasma CDA activities toward cytidine and gemcitabine were also measured (n = 121).

Results
Twenty-six genetic variations, including 14 novel ones and two known nonsynonymous single nucleotide polymorphisms (SNPs), were detected. Haplotypes harboring the nonsynonymous SNPs 79A>C (Lys27Gln) and 208G>A (Ala70Thr) were designated **2* and **3*, respectively. The allelic frequencies of the two SNPs were 0.207 and 0.037, respectively. Pharmacokinetic parameters of gemcitabine and plasma CDA activities significantly depended on the number of haplotype **3*. Haplotype **3* was also associated with increased incidences of grade 3 or higher neutropenia in the patients who were coadministered fluorouracil, cisplatin, or carboplatin. Haplotype **2* showed no significant effect on gemcitabine pharmacokinetics.

Conclusion
Haplotype **3* harboring a nonsynonymous SNP, 208G>A (Ala70Thr), decreased clearance of gemcitabine, and increased incidences of neutropenia when patients were coadministered platinum-containing drugs or fluorouracil.

J Clin Oncol 25:32-42. © 2007 by American Society of Clinical Oncology

INTRODUCTION

Gemcitabine (2′,2′-difluorodeoxycytidine) is a nucleoside anticancer drug that has a broad spectrum of antitumor activity against various solid tumors, such as non–small-cell lung cancer and pancreatic cancer.[1] In a randomized clinical trial, gemcitabine was confirmed to provide a survival advantage over fluorouracil in addition to symptom-relieving benefits in patients with advanced pancreatic cancer.[2] On the basis of these results, gemcitabine has generally been accepted as a standard chemotherapeutic agent for advanced pancreatic cancer.

Gemcitabine is transported into cells by concentrative and equilibrative nucleoside transporters,[3-8] where it is phosphorylated to its monophosphate form by deoxycytidine kinase. Gemcitabine triphosphate, an active form of gemcitabine, is incorporated into an elongating DNA strand, and is followed by the addition of another deoxynucleotide that leads to the halt of DNA synthesis.[9,10] Another mode of action in solid tumors, associated with the inhibition of ribonucleotide reductase, has also been suggested.[11]

Gemcitabine is rapidly metabolized to an inactive metabolite, 2′,2′-difluorodeoxyuridine (dFdU)

Table 1. *CDA* Haplotypes Estimated in This Study

Region		5'-Flanking			Exon 1 (5'-UTR)			Exon 1	Intron 1	Exon 2		Intron 2		
SNP ID		CDA001	CDA002	CDA003	CDA004	CDA005	CDA007	CDA009	CDA010	CDA011	CDA012	CDA014	CDA016	CDA017
Nucleotide change		−451C>T	−205C>G	−182G>A	−116G>A	−92A>G	−33_−31 delC	79A>C	IVS1+37 G>A	208G>A	210T>C	IVS2 +87_+88 insTCAT	IVS2+242 A>G	IVS2+296 T>A
Amino acid change								Lys27Gln		Ala70Thr	Ala70Ala			
Haplotypes														
*1	*1a													
	*1b													
	*1c													
	*1d													
	*1e													
	*1f													
	*1g													
	*1h													
	*1i													
	*1j													
	*1k													
	*1l													
	*1m													
	*1n													
	Other *1													
*2	*2a													
	*2b													
	*2c													
	*2d													
	Other *2													
*3	*3a													
	*3b													

(continued on next page)

NOTE. The haplotypes were described as a number plus a small alphabetical letter. Four single nucleotide polymorphisms (SNPs) (CDA006, 008, 013, 015) were found only in the very rare ambiguous *1 haplotypes. Since these ambiguous haplotypes were grouped and described as "Other *1" in this table, the four SNPs are not shown in the row of nucleotide change. White, major allele; gray, minor allele.

by cytidine deaminase (CDA),[9] and most of an administered dose is recovered as dFdU in the urine.[12] CDA is expressed at varying levels in the human tissues,[13] and the rapid clearance of gemcitabine can be attributed to its plentiful occurrence in the liver.[14] Two single nucleotide polymorphisms (SNPs), 79A>C (Lys27Gln) and 435T>C (Thr145Thr), have been discovered in *CDA*, the CDA-encoding gene in humans.[15,16] The 79A>C SNP reportedly reduces the deamination activity (maximum velocity/Km) toward 1-beta-D-arabinofuranosyl cytosine (cytarabine),[15] and increases Km toward gemcitabine,[17] in vitro. A recently discovered third SNP, 208G>A (Ala70Thr) displayed a decrease in deamination activity of 60% for cytidine and 68% for cytarabine when introduced into a *CDA*-null yeast strain.[18]

Toxicity of gemcitabine is generally mild,[19,20] but unpredictable severe toxicities such as myelosuppression are occasionally experienced.[21,22] Our previous case report described a patient with homozygous 208A alleles of the *CDA* gene who showed severe adverse reactions with increased plasma gemcitabine levels.[23] In addition, there has been controversy over the relationship between cellular CDA activity and the clinical effects of cytarabine.[24-27] This study examined the relationship between *CDA* polymorphisms, and the pharmacokinetics of gemcitabine, plasma CDA activity, or adverse reactions in Japanese cancer patients.

PATIENTS AND METHODS

Gemcitabine and dFdU for analytic standards were supplied by Eli Lilly Japan K.K. (Kobe, Japan). Tetrahydrouridine, 3'-deoxy-3'-fluoro-thymidine (3'-dFT), cytidine and uridine (Sigma-Aldrich Chemical Co, St Louis, MO) were purchased. All other chemicals were of highest grade available.

Patients

The participants in this study consisted of 256 Japanese patients with carcinoma, including six patients described in a previous report,[23] at the National Cancer Center Hospital (Tokyo, Japan) and National Cancer Center Hospital East (Kashiwa, Japan). Two hundred fifty-one patients received a 30-minute intravenous infusion of gemcitabine at a dose of either 800 or 1,000 mg/m^2, and five patients received a fixed dose-rate (10 mg/m^2/min) infusion at a dose between 1,000 and 1,500 mg/m^2. The eligibility criteria for the study were as previously reported.[23] The ethics committees of the National Cancer Center and the National Institutes of Health Sciences approved this study. Written informed consent was obtained from each participant.

Table 1. *CDA* Haplotypes Estimated in This Study (continued)

Intron 3					Exon 4	Exon 4 (3'-UTR)					
CDA018	CDA019	CDA020	CDA021	CDA022	CDA023	CDA024	CDA025	CDA026			
IVS3+71 T>C	IVS3 −194_−193 insAlu	IVS3-56 G>A	IVS3-36 G>A	IVS3-23 C>T	435C>T	510 (*69) G>T	637_638 (*196_*197) insC	676 (*235) A>G			
					Thr145Thr						
									No.	Frequency	
									175	0.342	
									63	0.123	
									52	0.102	
									17	0.033	
									13	0.025	
									12	0.023	
									12	0.023	
									11	0.021	0.756
									8	0.016	
									5	0.010	
									4	0.008	
									4	0.008	
									2	0.004	
									1	0.002	
									8	0.016	
									84	0.164	
									11	0.021	
									5	0.010	0.207
									3	0.006	
									3	0.006	
									18	0.035	0.037
									1	0.002	
									512	1.000	1.000

Monitoring and Toxicities

A complete medical history and data on physical examinations were recorded before the gemcitabine therapy. CBC and platelet counts, as well as blood chemistry, were measured once a week during the first 2 months of gemcitabine treatment. Toxicities were graded according to the National Cancer Institute Common Toxicity Criteria, version 2.

DNA Sequencing

All four exons and the 5'-upstream region (approximately 800 base pairs [bp] from the translation initiation codon) of *CDA* were amplified from 100 ng of DNA extracted from peripheral blood, and sequenced along both strands. Polymerase chain reaction (PCR) primers[23] and sequencing and PCR conditions[26] were described previously. For detection of an approximately 300-bp Alu insertion (IVS3-194_-193insAlu), PCR was performed using a specific primer set (5'- TTGTCATAGCAGAAGGAGGTT-3' and 5'- TCAG CTCTCCACACCATAAGG-3') and 100 ng of DNA as a template. Then, sizes of the amplified fragments were determined by 1% agarose gel electrophoresis. NT_004610.17 (GenBank, National Center for Biotechnology Information, Bethesda, MD) was used as the reference sequence.

Linkage Disequilibrium and Haplotype Analyses

Hardy-Weinberg equilibrium and linkage disequilibrium (LD) analyses were performed by SNPAlyze software (Dynacom Co, Yokohama, Japan). All of the detected variations were found to be in Hardy-Weinberg equilibrium ($P \geq .05$), except for the SNP IVS1+37G>A ($P = .002$). Some of the haplotypes were unambiguously assigned from subjects with homozygous variations at all sites or a heterozygous variation at only one site. The diplotype configurations (a combination of haplotypes) were separately inferred by LDSUPPORT software,[29] which determines the posterior probability distribution of the diplotype configuration for each subject based on the estimated haplotype frequencies. The diplotype configurations of all but 11 subjects were inferred with probability of more than 0.93. All haplotypes inferred in single subjects were gathered as the groups "Other *1" and "Other *2" in Table 1.

Pharmacokinetic Study

Five patients with fixed dose-rate infusion and one patient with interruption of infusion for more than 15 minutes were excluded from the pharmacokinetic analysis described herein. Heparinized blood was collected before administration of gemcitabine and used to measure plasma CDA activity. Five milliliters of heparinized blood was also sampled for pharmacokinetic analysis before the first gemcitabine administration, and at 0, 15, 30, 60, 90, 120, and 240 minutes after the termination of the infusion. Fifty microliters of 1% tetrahydrouridine was immediately added to these samples to prevent ex vivo deamination. Plasma levels of gemcitabine and dFdU were determined using the high-performance liquid chromatography method previously reported.[23] The area under the curve (AUC) and mean residence time from 0 to infinity, peak concentration (C_{max}), clearance (CL/m^2) and distribution volume based on the terminal phase (Vz/m^2) were calculated using WINNonlin (Scientific Consultant, Apex, NC) version 4.01 (Pharsight Corporation, Mountain View,

CA). AUC and C_{max} were corrected for dose, assuming that all patients received 1,000 mg/m^2 of gemcitabine.

CDA Activities in Plasma

Determination of CDA activities was performed using the method by Richards et al[30] with slight modifications (modifications are as follows: gemcitabine was used as a substrate as well as cytidine, internal standards for analysis [3'-dFT for gemcitabine or dFdU for cytidine] were added to the mixtures at the beginning of the reaction, and high-performance liquid chromatography was used for detection of reaction products). CDA activity was expressed by unit, and one unit of enzyme activity was defined as the concentration that produces 0.1 nmol of dFdU or uridine per minute per milliliter of plasma.[30]

Statistical Analysis

Kruskal-Wallis, Mann-Whitney, and Pearson's correlation tests were performed using the JMP software (SAS Institute Inc, Cary, NC). Two ordinally scaled categoric data were subjected to χ^2 analysis for a correlation test. A significance level of .05 was applied to all two-tailed and correlation tests. Multiplicity was adjusted by the false-discovery rate,[31] if necessary.

Results

Genetic Variations and Haplotype Structures of CDA

Twenty-six (14 novel) genetic variations were detected in the 256 Japanese cancer patients enrolled onto this study (Table 2). Three of the novel variations were found in the 5'-untranslated region, one in exon 2, three in the 3'-untranslated region and seven in the introns. Three known SNPs in the coding region of *CDA* were also detected. Among these, the nonsynonymous SNPs, 79A>C (Lys27Gln) and 208G>A (Ala70Thr), exhibited allelic frequencies of 0.207 and 0.037 (Table 2), respectively, and they were comparable to those reported previously.[18] One patient was found to be homozygous for the 208A polymorphism. A novel insertion of an approximately 320-bp Alu element (IVS3-194_-193insAlu) was newly found in intron 3.

The detected variations were used to analyze LD (Fig 1). Four novel variations (IVS3-56G>A, IVS3-36G>A, IVS3-23C>T and

Table 2. Variations of the *CDA* Gene Found

SNP ID				Position				
This Study	NCBI (dbSNP)	JSNP	Location	NT_004610.17	From the Translational Initiation Site or From the Nearest Exon	Nucleotide Change and Flanking Sequences (5' to 3')	Amino Acid Change	Allele Frequency
MPJ6_CDA001	rs532545	IMS-JST008767	5'-Flanking	3739514	−451‡	TGCCTCCTGCCTC/TGGGATGCCGCAG		0.199
MPJ6_CDA002	rs603412	IMS-JST008768	5'-Flanking	3739760	−205‡	CACACGTAGGCAC/GTGTCTTACACCA		0.256
MPJ6_CDA003	rs12726436		5'-Flanking	3739783	−182‡	CACACCTGCTGAG/ATCCAAACCATGG		0.061
MPJ6_CDA004*			Exon 1 (5'-UTR)	3739849	−116‡	CTGAGAGCCTGCG/AGTCTGGCTGCAG		0.059
MPJ6_CDA005	rs602950		Exon 1 (5'-UTR)	3739873	−92‡	GGGACACACCCAA/GGGGGAGGAGCTG		0.205
MPJ6_CDA006*			Exon 1 (5'-UTR)	3739884	−81‡	AAGGGGAGGAGCT/CGCAATCGTGTCT		0.002
MPJ6_CDA007	rs3215400	IMS-JST076939	Exon 1 (5'-UTR)	3739934	−33_−31‡	GCTCCTGTTTCCC/-GCTGCTCTGCTG		0.451
MPJ6_CDA008*			Exon 1 (5'-UTR)	3739957	−8‡	TGCCTGCCCGGGG/ATACCAACATGGC		0.002
MPJ6_CDA009†	rs2072671	IMS-JST008769	Exon 1	3740043	79‡	CAGGAGGCCAAGA/CAGTCAGCCTACT	Lys27Gln	0.207
MPJ6_CDA010	rs12059454		Intron 1	3740155	IVS1+37	CCCAGCCCAGCAG/ACCTGGGTGGTGG		0.184
MPJ6_CDA011†			Exon 2	3755816	208‡	GCTGAACGGACCG/ACTATCCAGAAGG	Ala70Thr	0.037
MPJ6_CDA012*			Exon 2	3755818	210‡	TGAACGGACCGCT/CATCCAGAAGGCC	Ala70Ala	0.004
MPJ6_CDA013*			Intron 2	3755932	IVS2+58	GCCAACATCTTCC/TTTACACATATTA		0.002
MPJ6_CDA014*			Intron 2	3755961_3755962	IVS2+87_+88	TCATTCATTCAT-/TCATCTGACATATGTT		0.135
MPJ6_CDA015*			Intron 2	3756043	IVS2+169	ATAAGGAGATAAA/GTAAGAAATGGAG		0.002
MPJ6_CDA016	rs10916825		Intron 2	3756116	IVS2+242	CATACAAGGGCCA/GGTATGCCCCTGT		0.289
MPJ6_CDA017	rs818194		Intron 2	3756170	IVS2+296	GTCCTACAAGATT/ATAACAGAAAGGC		0.217
MPJ6_CDA018	rs3738130	IMS-JST083844	Intron 3	3764805	IVS3+71	AGCCACGCCAAGT/CTGCAGGCATGGC		0.053
MPJ6_CDA019*			Intron 3	3769093_3769094	IVS3-194_−193	CTGTTCAGTTTC-/(Alu)§ACAGCATTCTTT		0.293
MPJ6_CDA020*			Intron 3	3769231	IVS3-56	CAGACCCAGTCCG/ATCTCAGCCCCCT		0.293
MPJ6_CDA021*			Intron 3	3769251	IVS3-36	CCCCTCAGCCACG/ACTGTGTCTCTCA		0.293
MPJ6_CDA022*			Intron 3	3769264	IVS3-23	CTGTGTCTCTCAC/TGCCAGCTTTGCC		0.293
MPJ6_CDA023†	rs17846527		Exon 4	3769397	435‡	CCTGCAGAAGACC/TCAGTGACAGCCA	Thr145Thr	0.293
MPJ6_CDA024*			Exon 4 (3'-UTR)	3769472	510 (*69)‡	CTCACAGCCCTGG/TTGGACACCTGCCC		0.002
MPJ6_CDA025*			Exon 4 (3'-UTR)	3769599_3769600	637_638 (*196_197)‡	ACCGCCGCCCCC-/CTGCCCCACCTTT		0.293
MPJ6_CDA026*			Exon 4 (3'-UTR)	3769638	676 (*235)‡	GGGCCCTCTTTCA/GAAGTCCAGCCTA		0.010

*Novel variations detected in this study.
†Yue et al.[18]
‡A of the translation initiation codon ATG is numbered 1, and the number with * in parentheses indicates the position from the termination codon TGA.
§The sequence of the Alu insertion was as follows: 5' - (T)nGAGACGGAGTCTCGCTGTCGCCCAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCACTGCAGGCTCCG CCCCCTGGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCGCCACCTCGCCCGGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTT TCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCCACTG TTCAGTTTC-3' (n = approximately 25).