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.

ƒ Institute of Environmental Medicine, Karolinska Institute.

⊥ 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.

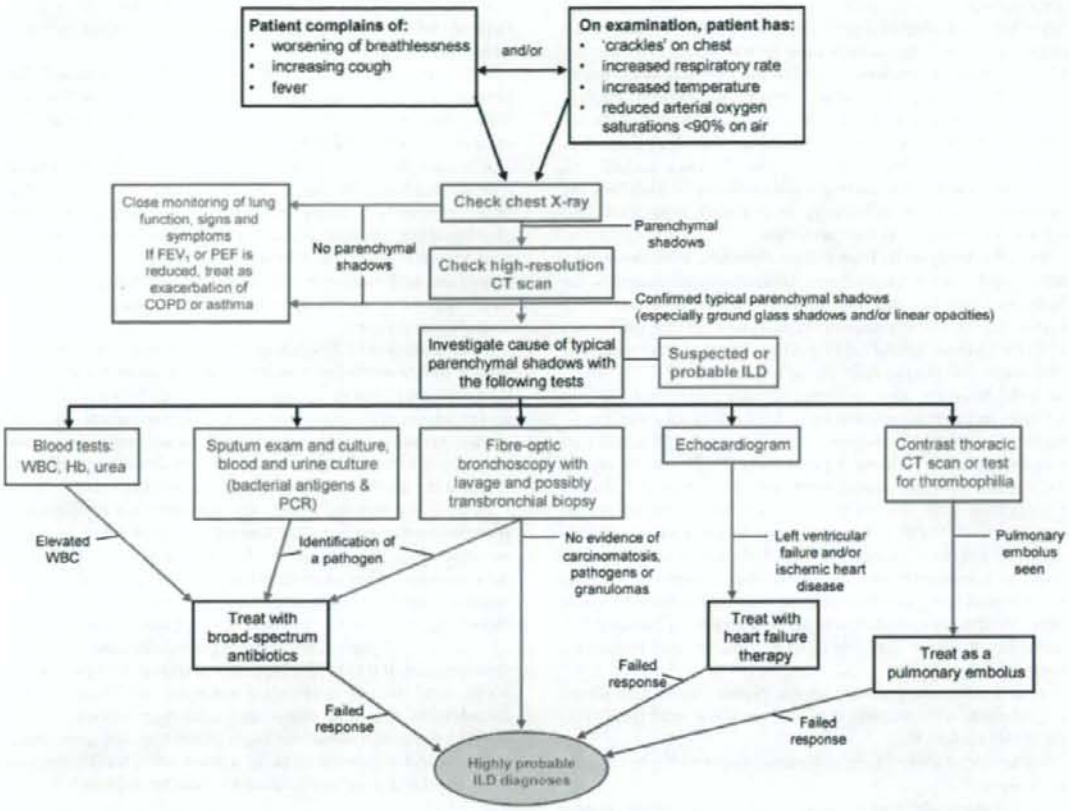

Figure 1. Algorithm for diagnosis of interstitial lung disease (ILD) in non-small cell lung cancer (NSCLC) patients.

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,[36,54-56] 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.[49]
- 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.[47,49] 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.,

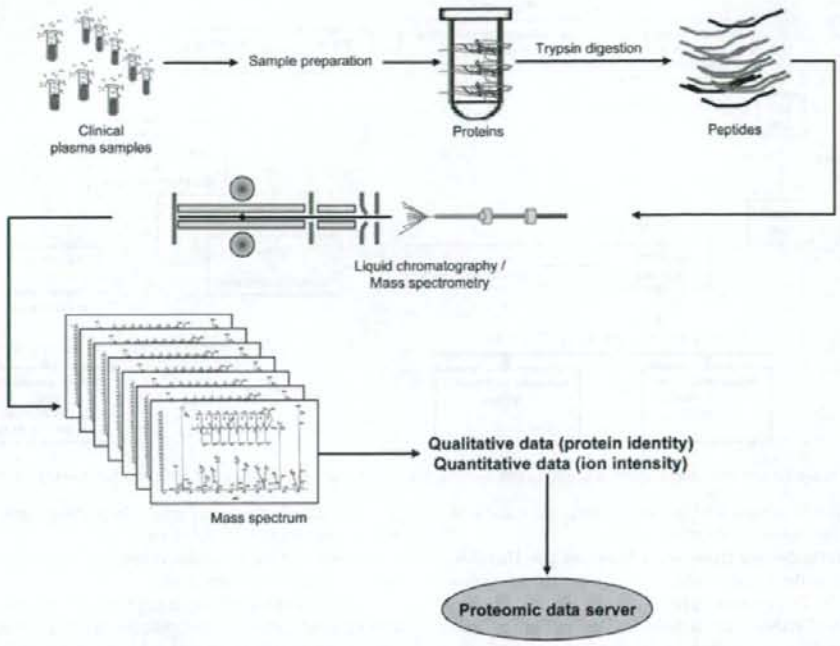

Figure 2. Schematic illustration of the clinical proteomics screening process.

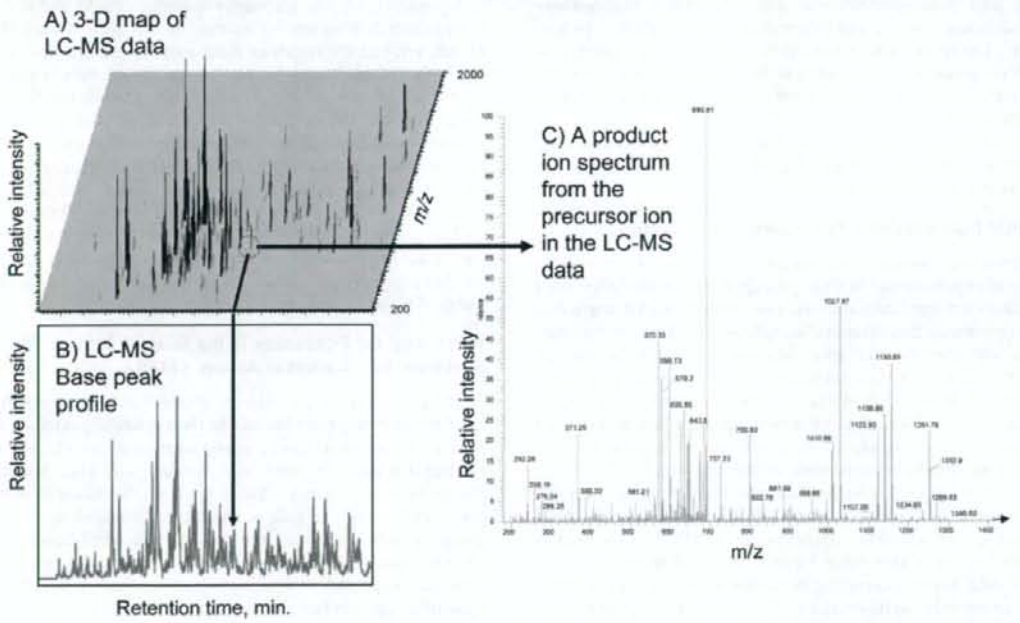

Figure 3. Profile of LC–MS data: (a) the three-dimensional view of LC-MS data, (b) the base-peak mass chromatogram, and (c) a product ion spectrum measured for a precursor ion in data-dependent acquisition mode (with MS acquisition operational *m/z* range set at 450–2000).

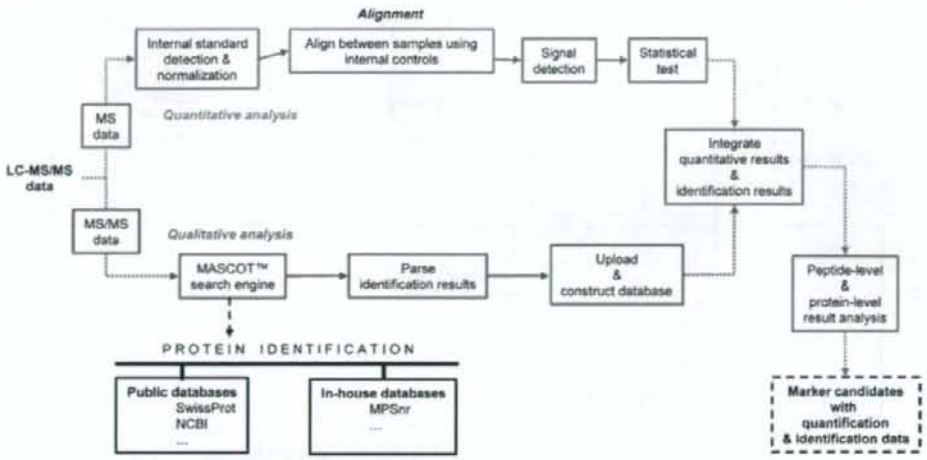

Figure 4. Overview of the data acquisition and database mining process developed within the gefitinib biomarker study.

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.

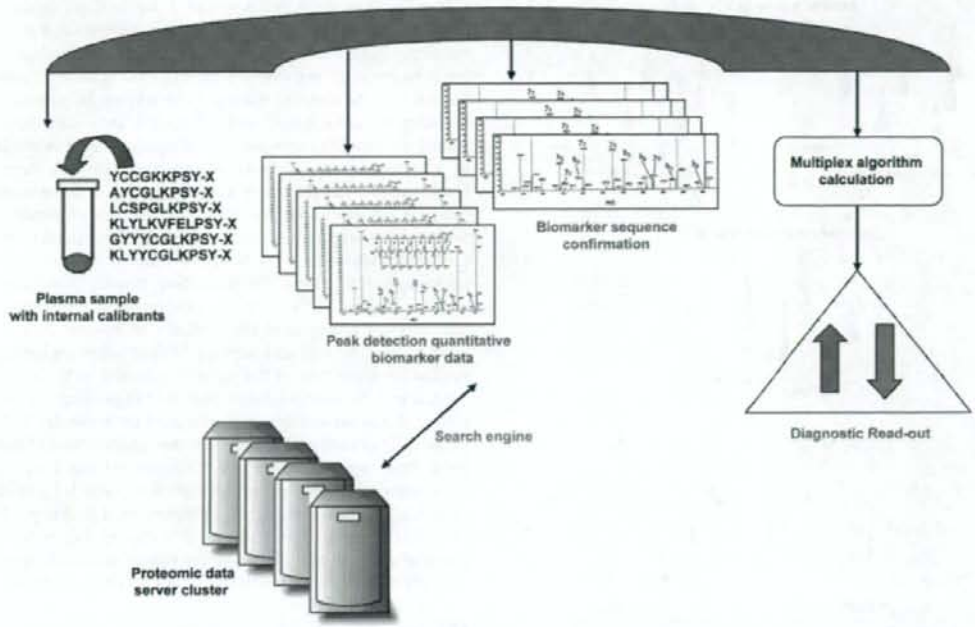

Figure 5. Entire flow of the operational components of Mass Spectrometric Biomarker Assays (MSBA).

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

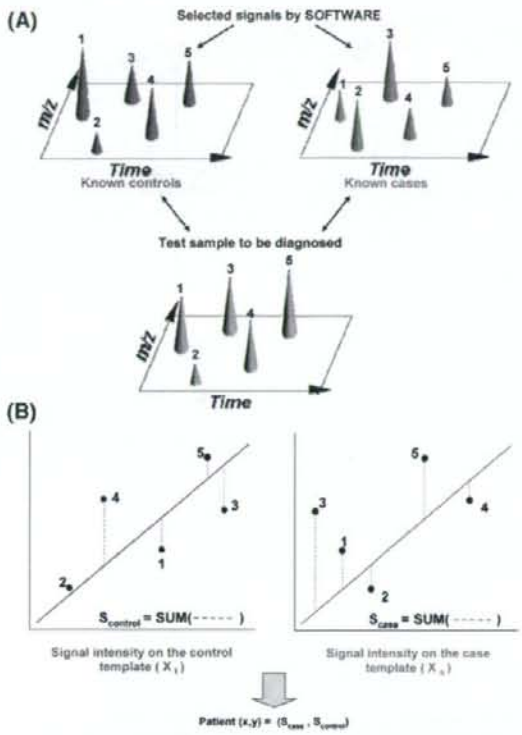

Figure 6. (A) Peptide signal comparison that MSBA (Mass Spectrometric Biomarker Assays) performs of the generated ions from the sample. The comparison is made both with the pattern of the controls and with the pattern of the case group for the corresponding signals. (B) Illustration of the regression 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-

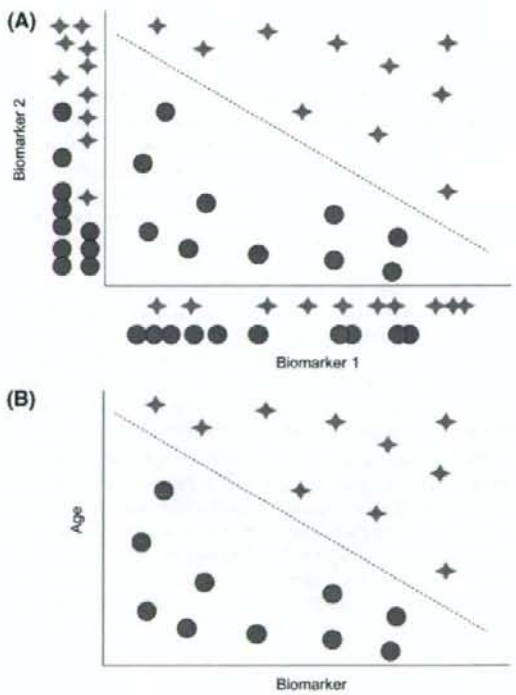

Figure 7. (A) Hypothetical example of the combined disease-linkage effect of two protein biomarkers. (Stars signify affected case individuals, circles non-affected control individuals). (B) Hypothetical example of the combined disease-linkage effect of a biomarker and a clinical variable. (Stars signify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

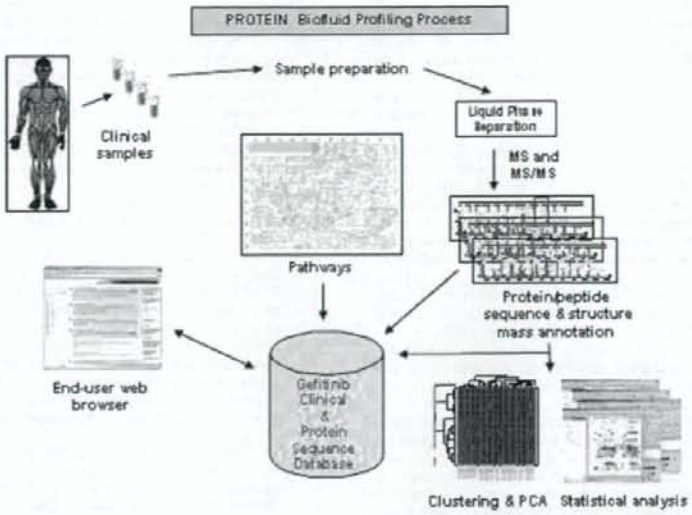

Figure 8. Illustration of the bioinformatics and data processing structure within which MSBA (Mass Spectrometric Biomarker Assays) data are captured, modified and analyzed.

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) Rahbar, 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–L23.
(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

Alcohol dehydrogenase and *aldehyde dehydrogenase* polymorphisms and colorectal cancer: The Fukuoka Colorectal Cancer Study

Guang Yin,[1,12] Suminori Kono,[1] Kengo Toyomura,[1] Malcolm A. Moore,[1] June Nagano,[1] Tetsuya Mizoue,[1] Ryuichi Mibu,[2] Masao Tanaka,[2] Yoshihiro Kakeji,[3] Yoshihiko Maehara,[3] Takeshi Okamura,[4] Koji Ikejiri,[5] Kitaroh Futami,[6] Yohichi Yasunami,[7] Takafumi Maekawa,[8] Kenji Takenaka,[9] Hitoshi Ichimiya[10] and Nobutoshi Imaizumi[11]

Departments of [1]Preventive Medicine, [2]Surgery and Oncology, and [3]Surgery and Science, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582; [4]Department of Gastroenterological Surgery, National Kyushu Cancer Center, 3-1-1 Notame, Minami-ku, Fukuoka 811-1395; [5]Division of Surgery, National Kyushu Medical Center, 1-8-1 Jigyohama, Chuo-ku, Fukuoka 810-8563; [6]Department of Surgery, Fukuoka University Chikushi Hospital, 377-1 Oaza-zokumyoin, Chikushino-shi 818–0067; The [7]First and [8]Second Departments of Surgery, Fukuoka University School of Medicine, 4-45-1 Nanakuma, Jonan-ku, Fukuoka 814-0180; [9]Division of Surgery, Fukuoka City Hospital, 13-1 Yoshizuka-honmachi, Hakata-ku, Fukuoka 812-0046; [10]Division of Surgery, Hamanomachi General Hospital, 3-5-27 Maizuru, Chuo-ku, Fukuoka 810–8539; and [11]Division of Surgery, Fukuoka Red Cross Hospital, 3-1-1 Ogusu, Minami-ku, Fukuoka 815-8555

(Received January 12, 2007/Revised April 9, 2007/Accepted April 19, 2007/Online publication May 22, 2007)

Alcohol dehydrogenase and aldehyde dehydrogenase are key enzymes in alcohol metabolism and therefore may be of importance to colorectal cancer development. The present case–control study was conducted to determine the influence of *ADH2*, *ADH3* and *ALDH2* polymorphisms in Fukuoka, Japan, with 685 incident cases of histologically confirmed colorectal adenocarcinomas and 778 community controls selected randomly from the study area. Alcohol use was ascertained by in-person interview. Statistical adjustment was made for sex, age class, area, and alcohol use. Individuals with the allele *47Arg* of the *ADH2* polymorphism (slow metabolizers) had a statistically significant increase in risk, with an adjusted OR of 1.32 (95% CI = 1.07–1.63), compared with those having the *ADH2*47His/His* genotype. This association was not affected by the level of alcohol consumption. The *ADH3* polymorphism showed no measurable association with the risk of colorectal cancer on either overall analysis or stratified analysis with alcohol use. The heterozygous *ALDH2*487Glu/Lys* genotype was not associated with an increase in the risk of colorectal cancer (adjusted OR 0.89, 95% CI = 0.71–1.13) compared with the *ALDH2*487Glu/Glu* genotype. Rather unexpectedly, the homozygous *ALDH2*487Lys/Lys* genotype was related to a statistically significantly decreased risk of colorectal cancer (adjusted OR 0.55, 95% CI = 0.33–0.93). It is unlikely that acetaldehyde metabolism determined by *ALDH2* polymorphism contributes to the risk of colorectal cancer, whereas the role of *ADH2* polymorphism deserves further investigation. (*Cancer Sci* 2007; 98: 1248–1253)

Alcohol consumption has fairly consistently been related to an increased risk of colorectal cancer.[1] In a pooled analysis of eight cohort studies in North America and Europe, a consumption of ≥45 g of alcohol per day was associated with a 1.4-fold increase in the risk of colorectal cancer.[2] A positive association between alcohol and colon or colorectal cancer has also been observed in Asian countries,[3–7] with few exceptions.[8] However, uncertainty remains as to the biological mechanisms for the association between alcohol use and colorectal cancer.

Ethanol is first oxidized to acetaldehyde by ADH, and acetaldehyde is further metabolized to acetate by ALDH. Human ADH exhibits several isoenzymes, and functional polymorphisms are known for the *ADH2* and *ADH3* genes.[9] A polymorphism in exon 3 of the *ADH2* gene, resulting in an arginine to histidine substitution in codon 47, affects the enzyme activity substantially. Individuals that are homozygous for the *ADH2*47His* allele (previously called *ADH2*2*) metabolize ethanol 40 times faster than those homozygous for the *ADH2*47Arg* allele (previously called *ADH2*1*).[10] The enzyme activity of *ADH2*47His/Arg* genotype is in the intermediate range between the two homozygous genotypes.[11] The polymorphic site for the *ADH3* gene is *Ile349Val* in exon 8. Maximal velocity is 2.5-fold greater in individuals homozygous for the *ADH3*349Ile* allele (previously called *ADH3*1*) than in those homozygous for the *ADH3*349Val* allele (previously called *ADH3*2*).[10] The *ADH2*47His* allele is fairly common in Asian populations and rare in Caucasians, while the *ADH3*349Val* allele is more frequent in Caucasians than in Asians.[12] *ALDH2* is the gene encoding mitochondrial ALDH, which contributes the majority of acetaldehyde oxidation in human liver and contains a functional polymorphism of *Glu487Lys*, with the variant *ALDH2*487Lys* (previously called *ALDH2*2*) allele resulting in an inactive form. The *ALDH2*487Lys* allele is mainly found in Asian populations.[12,13]

Several studies have investigated the relation of genetic polymorphisms of these alcohol-metabolizing enzymes to colorectal cancer and adenomas. As regards the *ADH2* polymorphism and colorectal cancer, a moderate increase in the risk of colorectal cancer was observed for each of the *Arg/His* and *Arg/Arg* genotypes compared with the *His/His* genotype in Japan,[14] but not in Spain.[15] The *ADH3* polymorphism was unrelated to colorectal cancer in two studies of Caucasians,[16,17] but one of these suggested an effect modification of alcohol consumption.[16] Two studies have examined the relation between *ADH3* polymorphism and colorectal adenomas in Caucasians, producing inconsistent results.[18,19] Although there was no difference in the distribution of *ADH3* genotypes between adenoma cases and controls in these studies, one showed a moderate increase in the risk of adenoma in men and women with the *ADH3*349Ile/Ile* genotype compared with those with the *ADH3*349Ile/Val* or *ADH3*349Val/Val* genotypes when alcohol consumption was high,[18] whereas the other reported an increased risk of adenoma associated with the *ADH3*349Val* allele for men with high alcohol consumption.[19] Studies regarding the *ALDH2* polymorphism and colorectal cancer or adenomas have all been done in Japan.[14,20–23] An approximately 3-fold increase in the risk of colorectal cancer has been observed for *ALDH2*487Glu/Lys* versus *ALDH2*487Glu/Glu* among alcoholics.[20] Another study

[12]To whom correspondence should be addressed.
E-mail: yinguan@phealth. med.kyushu-u.ac.jp
Abbreviations: ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; CI, confidence interval; HDL, high-density lipoprotein; OR, odds ratio; PCR–RFLP, polymerase chain reaction–restriction fragment length polymorphism.

doi: 10.1111/j.1349-7006.2007.00519.x

suggested a greater increase in the risk of colon cancer, not of rectal cancer, associated with high alcohol consumption among individuals with the *ALDH2*487Glu/Lys* genotype.[21] A small case–control study showed a positive interaction between high alcohol consumption and the *ALDH2*487Glu/Lys* genotype, particularly on the risk of rectal cancer.[22] In contrast, the *ALDH2* polymorphism did not show any measurable association with either colorectal cancer or adenomas in recent studies.[14,23]

The present paper examines the relation of the *ADH2*, *ADH3* and *ALDH2* polymorphisms to colorectal cancer in a case–control study in Japan, focusing on effect modification of alcohol consumption and gene–gene interaction.

Materials and Methods

The Fukuoka Colorectal Cancer Study is a case–control study of incident cases and community controls, with Fukuoka City and three adjacent areas as the catchment area. Details have been reported previously.[24] Described below are methods relevant to the present analysis. The study protocol was approved by the ethical committees of Kyushu University and of all but two of the participating hospitals. There was no ethical committee at the two hospitals, and the survey was done at these hospitals with permission from the director of each hospital.

Subjects. Cases comprised a consecutive series of patients with histologically confirmed incident colorectal adenocarcinomas who were admitted to two university hospitals or six affiliated hospitals for surgical treatment during the period from October 2000 to December 2003. Other eligibility criteria included the following characteristics: age of 20–74 years at the time of diagnosis, residence in the study area, no prior history of partial or total removal of the colorectum, familial adenomatous polyposis or inflammatory bowel disease. Of the total of 1053 eligible cases, 840 cases (80%) participated in the interview, and 685 (65%) gave informed consent to genotyping.

Eligibility criteria for controls were the same as described for cases except for two items, that is, having no diagnosis of colorectal cancer and age of 20–74 years at the time of selection. A total of 1500 persons were selected as control candidates using two-stage random sampling from among residents living in 15 small areas. A total of 833 persons participated in the survey, and 778 gave informed consent to genotyping. Reasons for exclusion and non-participation were death ($n = 7$), migration from the study area ($n = 22$), undelivered mail ($n = 44$), mental incompetence ($n = 19$), history of partial or total removal of the colorectum ($n = 21$), diagnosis of colorectal cancer after the survey ($n = 5$), no response ($n = 158$), and refusal ($n = 391$). After exclusion of the first six categories of outcomes ($n = 118$), net participation rates were calculated as 60% (833/1382) for the interview and 56% (778/1382) for genotyping.

Interview. Research nurses interviewed cases and controls in person regarding physical activity, smoking, alcohol use, and other factors using a uniform questionnaire. Habitual alcohol consumption at the time 5 years prior to the onset of disease in cases or the interview in controls was ascertained. Individuals reported the average number of days per week that alcohol was consumed and the average amount of alcohol per day of drinking. The amount of alcohol was expressed by the conventional unit; one *go* (180 mL) of *sake*, one large bottle (633 mL) of beer, and half a *go* (90 mL) of *shochu* were each expressed as one unit; and one drink (30 mL) of whisky or brandy and one glass (100 mL) of wine were each converted to half a unit. The reproducibility of the questionnaire was tested on 29 control subjects (14 men and 15 women) with an interval of approximately 1 year, and the reported alcohol intake was highly reproducible (Spearman's $r = 0.82$).

Genotyping. A venous blood sample of 5 mL was taken after the interview. DNA was extracted from the buffy coat using a commercial kit (QIAGEN GmbH, Hilden, Germany) and genotyping was performed using the PCR–RFLP method. The PCR was performed in a reaction mixture of 10 μL containing 0.5 IU of Taq and 1 μL of template DNA with a concentration of approximately 50–150 ng/μL. The *ADH2 Arg47His* and *ADH3 Ile349Val* genotypes were determined according to the methods described by Osier *et al.*[25] Primers for the *ADH2 Arg47His* genotypes were 5′-ATT CTA AAT TGT TTA ATT CAA GAA g-3′ (sense) and 5′-ACT AAC ACA GAA TTA CTG GAC-3′ (antisense). PCR products were digested with 20 IU of *MslI* for 16 h at 37°C in a mixture of 20 μL, resulting in fragments of 443 bp and 242 bp for the *47His* allele and 685 bp for the *47Arg* allele. The *ADH3 Ile349Val* genotypes were determined using primers of 5′-TTG TTT ATC TGT GAT TTT TTT TGT-3′ (sense) and 5′-CGT TAC TGT AGA ATA CAA AGC-3′ (antisense). The PCR product of 378 bp fragments was digested with 5 IU of *SspI* in a reaction mixture of 20 μL for 3 h at 37°C, resulting in fragments of 274 bp and 104 bp for the *349Ile* allele and 378 bp for the *349Val* allele. The *ALDH2 Glu487Lys* genotypes were determined, as described by Goedde *et al.*,[26] using primers that were 5′-CAA ATT ACA GGG TCA ACT GCT-3′ (sense) and 5′-CCA CAC TCA CAG TTT TCT CTT-3′ (antisense). The PCR product was digested with *Ksp632I* (10 IU) or *EarI* (10 IU) for 12 h at 37°C in a mixture of 20 μL, resulting in fragments of 112 bp for the *ALDH2*487Glu* allele and 135 bp for the *ALDH2*487Lys* allele. The digested PCR products were separated using electrophoresis on 3% agarose gels (NuiSieve GTG, Rockland, ME, USA), and visualized with ethidium bromide.

Statistical analysis. The association of the genetic polymorphisms with risk of colorectal cancer was examined using multiple logistic regression analysis including indicator variables for sex, 5-year age class (the lowest class was <40 years), resident area (Fukuoka City or adjacent areas), and alcohol intake (0, 0.1–1.9, or ≥2.0 units per day) as covariates. Adjusted OR and 95% CI were obtained from the logistic regression coefficient and the standard error for the corresponding indicator variable. Statistical significance for the interaction was tested using the likelihood ratio test comparing the logistic models with and without interaction terms for the genotype and alcohol category. Statistical significance was concluded if the two-sided P-value was less than 0.05 or if the 95% CI did not include unity. All statistical analyses were performed using SAS version 8.2 (SAS Institute Inc., Cary, NC, USA).

Results

The number of men among the 685 cases and 778 controls was 426 (62%) and 490 (63%), respectively. The mean age of the cases was 60 years (range 27–74), and that of the controls was 59 years (range 22–75). More than half of the cases (61%) and controls (64%) were residents of Fukuoka City. All of the distributions of genotypes for the *ADH2 Arg47His*, *ADH3 Ile349Val*, and *ALDH2 Glu487Lys* polymorphisms were in agreement with the Hardy–Weinberg equilibrium in both cases and controls. The alcohol-drinking pattern differed strikingly by *ALDH2* polymorphism and slightly so with respect to the *ADH2* polymorphism (Fig. 1). Alcohol use was progressively less frequent with increasing numbers of the *ALDH2*487Lys* allele, and was slightly more frequent with increasing numbers of the *ADH2*47Arg* allele. There was no variation in the proportion of alcohol drinking according to the *ADH3 Ile349Val* polymorphism (data not shown).

Regarding the *ADH2* polymorphism, the *47Arg* allele was slightly more frequent in cases than in controls, and the adjusted OR for the *Arg/His* and *Arg/Arg* genotypes as compared with the *His/His* genotype were each greater than unity, the increase for the heterozygote being statistically significant (Table 1). The adjusted OR for those with the *ADH2*47Arg* allele compared

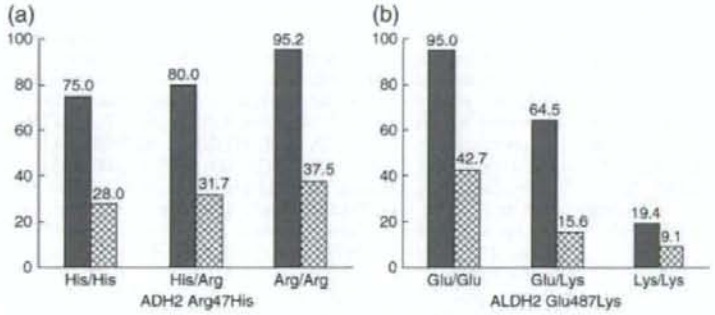

Fig. 1. Proportions (%) of alcohol drinkers for men (gray bar) and women (hatched bar) according to the *ADH2 Arg47His* and *ALDH2 Glu487Lys* polymorphisms in the control group. Values shown at the top of each bar are percentages of alcohol drinkers. Trend *P*-values were 0.03 in men and 0.35 in women for the *ADH2* polymorphism, and <0.0001 in both sexes for the *ALDH2* polymorphism. The trend *P*-value was based to the Mantel-Haenszel method with scores of 0, 1, and 2 assigned for the number of the variant allele.

Table 1. Relation of *ADH2*, *ADH3*, and *ALDH2* polymorphisms to colorectal cancer risk

Genotype	Cases (*n*, %)	Controls (*n*, %)	Adjusted OR (95% CI)†
ADH2 Arg47His‡			
His/His (fast)	345 (50.8)	452 (58.1)	1.00 (referent)
Arg/His	294 (43.3)	289 (37.1)	1.32 (1.06–1.64)
Arg/Arg (slow)	40 (5.9)	37 (4.8)	1.36 (0.84–2.20)
ADH3 Ile349Val§			
Ile/Ile (fast)	609 (88.9)	706 (90.9)	1.00 (referent)
Ile/Val	74 (10.8)	68 (8.7)	1.29 (0.91–1.83)
Val/Val (slow)	2 (0.3)	3 (0.4)	0.77 (0.13–4.70)
ALDH2 Glu487Lys			
Glu/Glu	400 (58.4)	416 (53.5)	1.00 (referent)
Glu/Lys	257 (37.5)	309 (39.7)	0.89 (0.71–1.13)
Lys/Lys (null activity)	28 (4.1)	53 (6.8)	0.55 (0.33–0.93)

†Adjusted for sex, 5-year age class, area, and alcohol use. ‡Six cases were excluded because of undetermined genotype. §One control was excluded because of undetermined genotype. CI, confidence interval; OR, odds ratio.

with those without was 1.32 (95% CI = 1.07–1.63). There was no measurable difference in the distribution of *ADH3 Ile349Val* genotypes between cases and controls. The *ALDH2*487Lys* allele was less frequent in cases than in controls, and the adjusted OR of colorectal cancer for the *Lys/Lys versus Glu/Glu* genotype was statistically significantly lower than unity. Analysis by sex showed similar results for men and women. For instance, the adjusted OR for *ADH2*47Arg/His* and *Arg/Arg* genotypes combined were 1.34 (95% CI = 1.03–1.75) in men and 1.35 (95% CI = 0.95–1.91) in women. Regarding the *ALDH2* polymorphism, the adjusted OR for the *Glu/Lys* and *Lys/Lys* genotypes compared with the *Glu/Glu* genotype in men were 0.77 (95% CI = 0.56–1.05) and 0.44 (95% CI = 0.22–0.91), respectively, while the corresponding values in women were 1.04 (95% CI = 0.71–1.53) and 0.67 (95% CI = 0.31–1.46), respectively.

Table 2 summarizes the results from the analysis regarding the interaction between alcohol intake and each genetic polymorphism. In this analysis, individuals heterozygous for the *ADH2* or *ADH3* polymorphism were each combined with those homozygous for the variant allele. Individuals homozygous for the variant allele of *ALDH2* (*487Lys/Lys*) were excluded because alcohol use was almost null in this group. There was no appreciable effect modification of each polymorphism on the relation between alcohol and colorectal cancer. High alcohol consumption was related to a moderate increase in the OR of colorectal cancer regardless of the genotype. Adjusted OR for high alcohol use (≥2 units/day) versus no use after control for the *ADH2*, *ADH3*, and *ALDH2* genotypes each were 1.34 (95% CI = 0.99–1.82), 1.37 (95% CI = 1.01–1.85) and 1.20 (95% CI = 0.85–1.69), respectively.

The joint effects of the *ADH2* or *ADH3* polymorphism in combination with the *ALDH2* polymorphism were examined among alcohol drinkers (Table 3). Individuals with the *ADH2*47His/His* genotype and the *ALDH2*487Lys* allele showed a statistically non-significant, small decrease in the OR for colorectal cancer. No such decrease was noted on analysis of the non-alcohol drinkers (data not shown). There was no clear interaction between the *ADH3* and *ALDH2* polymorphisms on the risk of colorectal cancer in either non-alcohol drinkers or alcohol drinkers.

Discussion

The present study addressed the relation between genetic polymorphisms in alcohol metabolism and colorectal cancer. Individuals with the *ADH2*47Arg* allele (slow metabolizers) showed a modest, but statistically significant, increase in the risk of colorectal cancer. In contrast, individuals homozygous for the *ALDH2* variant allele had a decreased risk of colorectal cancer. None of the polymorphisms affected the relation between alcohol and colorectal cancer. Because of the limited variation in the *ADH3* polymorphism, the present study did not provide useful information regarding the role of *ADH3* polymorphism in colorectal carcinogenesis.

Both *ADH2* and *ADH3* polymorphisms have been shown to affect the risk of various alcohol-related conditions. The slow alcohol metabolism of *ADH3* polymorphism has been related to increased risk of alcoholism and liver cirrhosis, elevated levels of serum HDL cholesterol, and decreased risk of myocardial infarction in Western populations.[27,28] Studies in Japan have reported that the slow metabolizers with the *ADH2* polymorphism had an increased risk of alcoholic liver disease,[29] and of cerebral infarction.[30] However, it remains uncertain whether these polymorphisms affect the risk of alcohol-related cancers. In a meta-analysis of seven case–control studies, no association between *ADH3*

doi: 10.1111/j.1349-7006.2007.00519.x

Table 2. Combined effects of *ADH2*, *ADH3*, and *ALDH2* polymorphisms with alcohol use on the risk of colorectal cancer

Genotype		Alcohol intake (unit/day)[†]		
		0	<2	≥2
ADH2 Arg47His				
His/His	No.[‡]	142/192	109/167	94/93
	OR (95% CI)[§]	1.00 (referent)	0.97 (0.69–1.37)	1.52 (1.02–2.25)
Arg/His + Arg/Arg	No.	128/119	109/123	97/84
	OR (95% CI)	1.46 (1.04–2.03)	1.30 (0.92–1.85)	1.69 (1.14–2.52)
		Interaction $P = 0.61$		
ADH3 Ile349Val				
Ile/Ile	No.	235/281	199/266	175/159
	OR (95% CI)	1.00 (referent)	0.98 (0.74–1.28)	1.44 (1.05–1.98)
Ile/Val + Val/Val	No.	37/30	22/23	17/18
	OR (95% CI)	1.49 (0.89–2.50)	1.29 (0.70–2.40)	1.27 (0.62–2.57)
		Interaction $P = 0.49$		
ALDH2 Glu487Lys[¶]				
Glu/Glu	No.	98/103	150/171	152/142
	OR (95% CI)	1.00 (referent)	1.04 (0.71–1.53)	1.24 (0.81–1.90)
Glu/Lys	No.	147/163	71/112	39/34
	OR (95% CI)	1.02 (0.71–1.48)	0.74 (0.47–1.17)	1.33 (0.74–2.39)
		Interaction $P = 0.30$		

[†]One unit of alcohol intake corresponded to 1 go (180 mL) of *sake*, 0.5 go (90 mL) of *shochu*, 1 large bottle (633 mL) of beer, 2 drinks (60 mL) of whiskey, or 2 glasses (200 mL) of wine. [‡]Numbers of cases/controls. [§]Adjusted for sex, 5-year age class, and area. [¶]Individuals with 487Lys/Lys genotype were excluded, because alcohol drinkers were few. CI, confidence interval; OR, odds ratio.

Table 3. Combined effects of *ADH2* or *ADH3* polymorphism with the *ALDH2* polymorphism on the risk of colorectal cancer in alcohol drinkers

ADH2/ADH3		*ALDH2 Glu487Lys*	
		Glu/Glu	*Glu/Lys*
ADH2 Arg47His			
His/His	No.[†]	153/170	50/84
	OR (95% CI)[‡]	1.00 (referent)	0.70 (0.46–1.07)
Arg/His + Arg/Arg	No.	148/143	57/62
	OR (95% CI)	1.12 (0.81–1.55)	1.08 (0.70–1.67)
		Interaction $P = 0.30$	
ADH3 Ile349Val			
Ile/Ile	No.	274/283	100/135
	OR (95% CI)	1.00 (referent)	0.83 (0.60–1.14)
Ile/Val + Val/Val	No.	28/30	10/11
	OR (95% CI)	1.00 (0.58–1.74)	0.98 (0.40–2.38)
		Interaction $P = 0.76$	

[†]Numbers of cases/controls. [‡]Adjusted for sex, 5-year age class, area, and alcohol use low or high intake. CI, confidence interval; OR, odds ratio.

polymorphism and upper aerodigestive cancers was found, nor was any interaction between *ADH3* polymorphism and alcohol consumption on the risk of these cancers.[12] A study of Japanese alcoholics has showed increased risks of oral, laryngeal, and esophageal cancer for the *ADH2*47Arg/Arg* genotype,[31] while another study in Japan showed no effect modification of the *ADH2* polymorphism on the association between alcohol and esophageal cancer.[32]

A recent Japanese study reported a positive association between the *ADH2* polymorphism and colorectal cancer, showing a progressive increase in the risk with increasing numbers of the *ADH2*47Arg* allele.[14] No such progressive increase in the risk was observed in the present study, but the authors' findings are compatible with the previous observation in that the risk was elevated in individuals with the *ADH2*47Arg* allele. The authors have no clear explanation for the increased risk of colorectal cancer associated with the *ADH2* polymorphism, but the consistency in the two independent studies warrants further investigation regarding the role of the *ADH2* polymorphism in colorectal carcinogenesis.

The present study showed neither an increased risk of colorectal cancer associated with the *ALDH2*487Lys* allele nor interaction between the *ALDH2*487Lys* allele and alcohol consumption. These findings are at odds with results from previous studies of colorectal cancer,[20–22] but are consistent with the recent observations on colorectal cancer,[14] and adenomas.[23] The statistically significant decrease in the risk of colorectal cancer associated with the *ALDH2*487Lys/Lys* genotype was rather unexpected and difficult to interpret. A decrease in the OR associated with the *Lys/Lys* genotype was also observed in the analysis confined to non-drinkers of alcohol ($n = 583$), although the decrease was not statistically significant; adjusted OR for the *Glu/Glu*, *Glu/Lys*, and *Lys/Lys* genotypes were 1.00 (referent), 1.00 (95% CI = 0.68–1.46), and 0.64 (95% CI = 0.36–1.14), respectively. The decreased risk in individuals with the *ALDH2487Lys/Lys* genotype may have been due to residual confounding of lifestyle factors other than alcohol drinking. In the Fukuoka Colorectal Cancer Study, obesity and physical inactivity were related to increased risk,[33] and there was a protective association with intake of n-3 polyunsaturated fatty acids.[34] With further

adjustment for these factors as well as for dietary calcium and fiber using the variables and categories as defined previously,[34] the adjusted OR for the *Glu/Lys* and *Lys/Lys* genotypes versus the *Glu/Glu* genotype were 0.90 (95% CI = 0.70–1.14) and 0.52 (95% CI = 0.31–0.88), respectively, in the analysis excluding four cases and two controls with a total calorie intake estimated to be >20 929 kJ/day. While a similar, inverse association was noted for colorectal adenomas,[23] no such association was seen in another study of colorectal cancer in Japan.[14]

It is hypothesized that acetaldehyde is accumulated in individuals who are fast alcohol metabolizers and slow acetaldehyde metabolizers.[29] Thus the authors hypothesized that the combination of *ADH2*His/His* and *ALDH2*487Glu/Lys* genotype might be related to an increased risk of colorectal cancer, but the risk of colorectal cancer was decreased, rather than increased in alcohol drinkers with such composite genotypes.

This finding on the gene–gene interaction, together with the above-mentioned findings on the *ALDH2* polymorphism, suggests that acetaldehyde metabolism in the liver is not measurably linked to colorectal carcinogenesis. Bacterial production of acetaldehyde in the colon is an alternative mechanism by which alcohol may enhance colorectal carcinogenesis. Human colonic contents and isolated colonic microbes are capable of producing acetaldehyde when incubated with ethanol *in vitro*.[35,36] It has been demonstrated in piglets that high levels of acetaldehyde were produced in the colon during normal metabolism of alcohol.[37] ALDH activity is much lower in the colonic mucosa than in the liver,[38,39] and colonic epithelial cells are exposed to high concentrations of acetaldehyde in the colonic lumen. It is hypothesized that low folate status increases the risk of colorectal cancer by altering DNA methylation and DNA synthesis.[40,41] Alcohol and acetaldehyde exert adverse effects on folate metabolism.[42] High alcohol consumption results in inadequate folate status by decreasing intestinal absorption and increasing renal excretion. It is known that acetaldehyde rather than alcohol itself cleaves folate chemically. Ethanol ingestion has resulted in a substantial increase in the intracolonic concentration of acetaldehyde and decreased folate levels in the colonic mucosa in an experimental study of rats.[43]

The present study is probably the largest that has ever been reported regarding the *ADH2* or *ALDH2* polymorphism and colorectal cancer. Among the reported studies are those including 257 colorectal cancer cases and 771 controls,[14] 270 cases and 121 controls,[21] and 142 cases and 241 non-cancer controls,[22] in Japan. The size of a study is particularly important in investigating the role of rare genotypes in the gene–environment or gene–gene interaction. The participation rate in terms of genotyping was not so high in either cases (65%) or controls (56%). Because the *ADH2* and *ALDH2* polymorphisms affected alcohol drinking, a selection bias would be possible in the association with these polymorphisms if cases and controls participated in the study differentially with respect to alcohol drinking. Among those interviewed, however, the proportions of alcohol drinking in the cases and controls each did not differ by consent to genotyping.[44] Alcohol consumption 5 years prior to the referent date was used, and the authors have no data as to how valid the recalled alcohol consumption in the past was, although it was found to be highly reproducible.

In summary, a case–control study in Japan showed an increased risk of colorectal cancer associated with the *ADH2*47Arg* allele, but not with the *487Lys* allele of *ALDH2* polymorphism. None of the polymorphisms affected the relation between alcohol consumption and colorectal cancer risk. It is unlikely that acetaldehyde metabolism determined by *ALDH2* polymorphism contributes to the risk of colorectal cancer, whereas the role of *ADH2* polymorphism deserves further investigation.

Acknowledgments

This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology, Japan (18014022), and from a Grant-in-Aid in the year 2005, Fukuoka Cancer Society, Japan. The authors acknowledge support from Emeritus Professor Keizo Sugimachi; Professors Seiyo Ikeda, Takayuki Shirakusa, and Sumitaka Arima; and Drs Motonori Saku, Yoichi Ikeda, Soichiro Maekawa, Kazuo Tanoue, Kinjiro Sumiyoshi, and Shoichiro Saito in conducting the survey of cases. The following physicians kindly supervised the survey of controls at their clinics: Drs Hideaki Baba, Tomonori Endo, Hiroshi Hara, Yoichiro Hirokata, Motohisa Ikeda, Masayoshi Ishibashi, Fumiaki Itoh, Yasuhiro Iwanaga, Hideki Kaku, Shoshi Kaku, Minoru Kanazawa, Akira Kobayashi, Ryunosuke Kumashiro, Shinichi Matsumoto, Soukei Mioka, Umeji Miyakoda, Osamu Nakagaki, Nobuyoshi Nogawa, Nobuyuki Ogami, Toyoaki Okabayashi, Hironao Okabe, Nishiki Saku, Masafumi Tanaka, Masahiro Ueda, Bunichi Ushio, and Koheisho Yasunaga. The authors are grateful to research nurses: Nobuko Taguchi, Yuriko Moroe, Yuko Noda, Ryoko Tanaka, Hisako Nakagawa, and Yoko Mikasa; and research clerk Hiroko Mizuta; and the assistance of Masumi Koga and Akiko Koga, for their self-sacrificing work.

References

1 World Cancer Research Fund and American Institute for Cancer Research. *Food, nutrition and the prevention of cancer: a global perspective*. Washington, DC: American Institute for Cancer Research, 1997: 216–51.
2 Cho E, Smith-Warner SA, Ritz J *et al.* Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. *Ann Intern Med* 2004; **140**: 603–13.
3 Shimizu N, Nagata C, Shimizu H *et al.* Height, weight, and alcohol consumption in relation to the risk of colorectal cancer in Japan: a prospective study. *Br J Cancer* 2003; **88**: 1038–43.
4 Otani T, Iwasaki M, Yamamoto S *et al.* Alcohol consumption, smoking, and subsequent risk of colorectal cancer in middle-aged and elderly Japanese men and women: Japan Public Health Center-based prospective study. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 1492–500.
5 Wakai K, Kojima M, Tamakoshi K *et al.* Alcohol consumption and colorectal cancer risk: findings from the JACC Study. *J Epidemiol* 2005; **15**: S173–9.
6 Ji BT, Dai Q, Gao YT *et al.* Cigarette and alcohol consumption and the risk of colorectal cancer in Shanghai, China. *Eur J Cancer Prev* 2002; **11**: 237–44.
7 Ho JW, Lam TH, Tse CW *et al.* Smoking, drinking and colorectal cancer in Hong Kong Chinese: a case-control study. *Int J Cancer* 2004; **109**: 587–97.
8 Chen K, Jiang Q, Ma X *et al.* Alcohol drinking and colorectal cancer: a population-based prospective cohort study in China. *Eur J Epidemiol* 2005; **20**: 149–54.
9 Yoshida A, Hsu LC, Yasunami M. Genetics of human alcohol-metabolizing enzymes. *Prog Nucl Acid Res Mol Biol* 1991; **40**: 255–87.
10 Bosron W, Li TK. Genetic polymorphism of human liver alcohol and aldehyde dehydrogenase, and their relationship to alcohol metabolism and alcoholism. *Hepatology* 1986; **6**: 502–10.
11 Yin SJ, Bosron WF, Magnes LJ, Li TK. Human liver alcohol and dehydrogenase. Purification and kinetic characterization of the β2 β2, β2 β1, α β2, and β2γ1 'Oriental' isoenzymes. *Biochemistry* 1984; **23**: 5847–53.
12 Brennan P, Lewis S, Hashibe M *et al.* Pooled analysis of alcohol dehydrogenase genotype and head and neck cancer: a HuGE Review. *Am J Epidemiol* 2004; **159**: 1–16.
13 Takeshita T, Morimoto K, Mao XQ, Hashimoto T, Furuyama J. Characterization of three genotypes of low Km aldehyde dehydrogenase in a Japanese population. *Hum Genet* 1994; **94**: 217–23.
14 Matsuo K, Wakai K, Hirose K *et al.* A gene–gene interaction between ALDH2 Glu487Lys and ADH2 His47Arg polymorphisms regarding the risk of colorectal cancer in Japan. *Carcinogenesis* 2006; **27**: 1018–23.
15 Landi S, Gemignani F, Moreno V *et al.* A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of colorectal cancer. *Pharmacogenet Genomics* 2005; **15**: 535–46.
16 Chen J, Ma J, Stampfer MJ, Hines LM, Selhub J, Hunter DJ. Alcohol dehydrogenase 3 genotype is not predictive for risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 1303–4.
17 van der Logt EM, Bergevoet SM, Roelofs HM *et al.* Role of epoxide hydrolase, NAD(P)H: quinine oxidoreductase, cytochrome P450 2E1 or

doi: 10.1111/j.1349-7006.2007.00519.x

alcohol dehydrogenase genotypes in susceptibility to colorectal cancer. *Mutat Res* 2006; **593**: 39–49.
18 Tiemersma EW, Wark PA, Ocke MC *et al*. Alcohol consumption, alcohol dehydrogenese 3 polymorphism, and colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 419–25.
19 Giovannucci E, Chen J, Smith-Warner SA *et al*. Methylenetetrahydrofolate reductase, alcohol dehydrogenase, diet, and risk of colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 970–9.
20 Yokoyama A, Muramatsu T, Ohmori T *et al*. Alcohol-related cancers and aldehyde dehydrogenase-2 in Japanese alcoholics. *Carcinogenesis* 1998; **19**: 1383–7.
21 Murata M, Tagawa M, Watanabe S, Kimura H, Takeshita T, Morimoto K. Genotype difference of aldehyde dehydrogenase 2 gene in alcohol drinkers influences the incidence of Japanese colorectal cancer patients. *Jpn J Cancer Res* 1999; **90**: 711–19.
22 Matsuo K, Hamajima N, Hirai T *et al*. Aldehyde dehydrogenase 2 (ALDH2) genotype affects rectal cancer susceptibility due to alcohol consumption. *J Epidemiol* 2002; **12**: 70–6.
23 Hirose M, Kono S, Tabata S *et al*. Genetic polymorphisms of methylenetetrahydrofolate reductase and aldehyde dehydrogenase2, alcohol use and risk of colorectal adenomas: Self-Defense Forces Health Study. *Cancer Sci* 2005; **96**: 513–8.
24 Kono S, Toyomura K, Yin G, Nagano J, Mizoue T. A case-control study of colorectal cancer in relation to lifestyle factors and genetic polymorphisms. Design and conduct of The Fukuoka Colorectal Cancer Study. *Asian Pac J Cancer Prev* 2004; **5**: 393–400.
25 Osier MV, Pakstis AJ, Soodyall H *et al*. A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. *Am J Hum Genet* 2002; **71**: 84–99.
26 Goedde HW, Singh S, Agarwal DP, Fritze G, Stapel K, Paik YK. Genotyping of mitochondrial aldehyde oligonucleotides: comparison with phenotyping in hair roots. *Hum Genet* 1989; **81**: 305–7.
27 Cichoz-Lach H, Partycka J, Nesima I, Celinski K, Slomka M, Wojcierowski J. Genetic polymorophism of alcohol dehydrogenase 3 in alcohol liver cirrhosis and in alcohol chronic pancreatitis. *Alcohol Alcohol* 2006; **41**: 14–17.
28 Hines LM, Stampfer MJ, Ma J *et al*. Genetic variation in alcohol dehydrogenase and the beneficial effect of moderate alcohol consumption on myocardial infarction. *N Engl J Med* 2001; **344**: 549–55.
29 Tanaka F, Shiratori Y, Yokosuka O, Imazeki F, Tsukada Y, Omata M. High incidence of ADH2*1/ALDH2*1 genes among Japanese alcohol dependents and patients with alcoholic liver disease. *Hepatology* 1996; **23**: 234–9.
30 Suzuki Y, Fujisawa M, Ando F, Niino N, Ohsawa I. Alcohol dehydrogenase 2 variant is associated with cerebral infarction and lacunae. *Neurology* 2004; **63**: 1711–13.
31 Yokoyama A, Muramatsu T, Omori T *et al*. Alcohol and aldehyde dehydrogenases gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics. *Carcinogenesis* 2001; **22**: 433–9.
32 Yang CX, Matsuo K, Ito H *et al*. Esophageal cancer risk by ALDH2 and ADH2 polymorphisms and alcohol consumption: exploration of gene-environment and gene–gene interactions. *Asian Pac J Cancer Prev* 2005; **6**: 256–62.
33 Isomura K, Kono S, Moore MA *et al*. Physical activity and colorectal cancer: The Fukuoka Colorectal Cancer Study. *Cancer Sci* 2006; **97**: 1099–104.
34 Kimura Y, Kono S, Toyomura K *et al*. Meat, fish and fat intake in relation to subsite-specific risk of colorectal cancer: The Fukuoka Colorectal Cancer Study. *Cancer Sci* 2007; **98**: 590–7.
35 Jokelainen K, Roine RP, Vaananen H, Farkkila M, Salaspuro M. *In vitro* acetaldehyde formation by human colonic bacteria. *Gut* 1994; **35**: 1271–4.
36 Jokelainen K, Siitonen A, Jousimies-Somer H, Nosova T, Heine R, Salaspuro M. *In vitro* alcohol dehydrogenase-mediated acetaldehyde production by aerobic bacteria representing the normal colonic flora in man. *Alcohol Clin Exp Res* 1996; **20**: 967–72.
37 Jokelainen K, Matysiak-Budnik T, Mäkisalo H, Höckerstedt K, Salaspuro M. High intracolonic acetaldehyde values produced by a bacteriocolonic pathway for ethanol oxidation in piglets. *Gut* 1996; **39**: 100–4.
38 Koivisto T, Salaspuro M. Aldehyde dehydrogenases of the rat colon: Comparison with other tissues of the alimentary tract and the liver. *Alcohol Clin Exp Res* 1996; **20**: 551–5.
39 Visapää JP, Tillonen J, Salaspuro M. Microbes and mucosa in the regulation of intracolonic acetaldehyde concentration during ethanol challenge. *Alcohol Alcohol* 2002; **37**: 322–6.
40 Giovannucci E. Alcohol, one-carbon metabolism, and colorectal cancer: recent insights from molecular studies. *J Nutr* 2004; **134**: 2475S–81S.
41 Kono S, Chen K. Genetic polymorphisms of methylenetetrahydrofolate reductase and colorectal cancer and adenoma. *Cancer Sci* 2005; **96**: 535–42.
42 Mason JB, Choi SW. Effects of alcohol on folate metabolism: implications for carcinogenesis. *Alcohol* 2005; **35**: 235–41.
43 Homann N, Tillonen J, Salaspuro M. Microbially produced acetaldehyde from ethanol may increase the risk of colon cancer via folate deficiency. *Int J Cancer* 2000; **86**: 169–73.
44 Hagiwara T, Kono S, Yin G *et al*. Genetic polymorphism in cytochrome *P*450 7A1 and risk of colorectal cancer: The Fukuoka Colorectal Cancer Study. *Cancer Res* 2005; **65**: 2979–82.

JOP. J Pancreas (Online) 2006; 7(4):332-336.

EDITORIAL

Serum Tumor Markers for Pancreatic Cancer: The Dawn of New Era?

Takuji Okusaka[1], Tesshi Yamada[2], Masato Maekawa[3]

[1]Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital. Tokyo, Japan. [2]Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute. Tokyo, Japan. [3]Department of Laboratory Medicine, Hamamatsu University School of Medicine. Hamamatsu, Japan

Pancreatic cancer accounts for only 3% of all cancers, but it is the fifth leading cause of cancer death in both Western countries [1] and Japan [2]. The prognosis of patients with this disease is extremely poor with less than 5% of patients alive 5 years after diagnosis. Of all the treatment modalities for pancreatic cancer, only resection offers the opportunity for a cure. However, at the time of diagnosis, approximately half of the patients already have metastases and approximately one third of patients are diagnosed as having locally advanced disease, whereas only a small proportion of patients are eligible for surgery. Most symptoms related to this malignancy occur only after disease advancement to an unresectable stages and the early diagnosis of pancreatic cancer remains challenging. To increase the proportion of pancreatic cancer patients with a chance of a cure, there is an urgent need to develop an effective screening system for asymptomatic individuals and to improve the diagnostic accuracy for pancreatic cancer in its early stage. Serum is the most ideal biological specimen for assessing tumor markers in clinical practice because of its availability for repeated collection and reproducible quantification. Recent advancements in technology and an increasing understanding of molecular biology have facilitated research programs into serum markers for pancreatic cancer.

One of the most important roles for serum markers is as a tool for cancer screening in asymptomatic populations. High-quality evidence to justify population-based screening are present for only a few specific malignancies like breast and colorectal cancers but pancreatic cancer has insufficient prevalence in the preclinical population and little availability of adequate modalities for screening. With an estimated prevalence of pancreatic cancer in the population of 0.015%, which is comparable to the latest incidence rate in Japan [3], even a test with sensitivity and specificity of 95% would yield 350 false-positive individuals for every true-positive patient. This example indicates that the screening test needs an almost 100% specificity for this malignancy.

Accuracy in diagnosis for patients with symptoms suspicious of pancreatic cancer is also required for tumor markers in order to distinguish malignancy from benign or non-invasive pancreatic disorders. CA 19-9, the most widely used serum marker for pancreatic cancer diagnosis, had been reported to have a sensitivity of 70-90% and a specificity of 70-98% [4, 5, 6, 7, 8]. Although imaging tests play the main role in the diagnosis of pancreatic cancer, serum markers including CA 19-9 have a considerable predictive value to assist the differential diagnosis in patients with abdominal discomfort or jaundice.

However, currently available serum markers are inadequately sensitive for detection of resectable pancreatic cancer. The Pancreatic Cancer Registry in Japan demonstrated that only 48.4% of the patients with small pancreatic cancer less than 2 cm in diameter had elevated CA 19-9 values [9]. Furthermore, CA19-9 values are considered useless in distinguish neoplasms with high invasive potential, such as mucinous cystic tumors and intraductal papillary mucinous tumors, from those with benign feature [10].

The most commonly accepted uses of serum tumor markers in clinical practice are for assessing the prognosis of, and therapeutic monitoring for pancreatic cancer patients because tumor marker in these situations are more valuable than other modalities including imaging diagnosis. Various studies have demonstrated that CA 19-9 is one of the most significant prognostic factors for both patients with resectable and those with unresectable disease [11, 12, 13, 14, 15]. Measurement of tumor markers as a prognostic factor provides valuable information to assist in the therapeutic decision making especially for surgeons, because early recurrence can be expected in patients with high preoperative levels of the markers. An elevated tumor marker value even after resection indicates the high possibility of remnant disease [16]. The postoperative increase in the value often anticipates the presentation of recurrence in imaging studies or of clinical symptoms. Although the measurement of the tumor size on CT or MR images is standard for evaluation in response to non-surgical treatments such as chemotherapy and radiotherapy, serial change in tumor markers assist the evaluation practically because of the difficulty in accurate measurement of pancreatic mass with obscure margin in most patients, and because of high incidence of the clinically occult progression associated with this disease [17].

Although CA 19-9 is the most useful serum marker for pancreatic cancer, it has some weaknesses. Approximately 10% of the population with the Lewis negative genotype is not able to produce CA 19-9 due to the lack of the enzyme involved in its synthesis, even if they have advanced pancreatic cancer. The Lewis gene dosage positively affects CA 19-9 value, whereas the secretor gene dosage negatively affects it [18]. Patients with small pancreatic cancer often show false negative in the CA 19-9 values. Falsely positive CA 19-9 elevation is frequently observed in patients with benign disease such as chronic pancreatitis. CA 19-9 elevation is common in patients with obstructive jaundice regardless of its malignancy and those with hapatobiliary and gastrointestinal cancer other than pancreatic cancer. Various other serum markers have been developed, although they have not displaced CA 19-9 due to its diagnostic accuracy, especially in the early stage of the disease.

Recent advances in the understanding of the molecular biology of pancreatic cancer facilitate research programs to search for novel markers including tissue-based and circulating markers. Hundreds of over-expressed genes in pancreatic cancer tissues have been identified in investigations using global gene expression. The protein product of an overexpressed gene needs several indispensable characteristics before it can become a sensitive and specific serum-based marker for pancreatic cancer: for example, it should be a secreted protein; it should be overexpressed in pancreatic cancers, it should not be expressed in the nonneoplastic pancreas, and it should have a restricted pattern of expression in other organs and tissues [19]. Several protein products of overexpressed genes including macrophage inhibitory cytokine-1 (MIC-1), synuclein-gamma, mesothelin, and osteopontin have been investigated as potential markers for pancreatic cancer, but their efficacy as serum markers remain undetermined [20, 21, 22]. Detection of aberrantly methylated genes in serum may be a useful diagnostic strategy for pancreatic cancer. The hypermethylation of CpG islands in promoter region is frequently associated with the silencing of tumor-suppressor genes such as p16/CDKN2A, E-cadherin, and others in cancer cells [23, 24]. These abnormalities have been preliminary