

Cancer Proteomics for Biomarker Development

Tadashi Kondo

Proteome Bioinformatics Project, National Cancer Center Research Institute

Corresponding author: Tadashi Kondo, MD, PhD, Proteome Bioinformatics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan,

Tel: +81-3-3542-2511; Fax: +81-3-3547-5298; E-mail: takondo@ncc.go.jp

Received August 09, 2008; Accepted September 19, 2008; Published December 05, 2008

Citation: Tadashi K(2008) Cancer Proteomics for Biomarker Development. J Proteomics Bioinform 1: 477-484. doi:10.4172/jpb.1000055

Copyright: © 2008 Tadashi K. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Cancer is a diverse disease, and biomarkers reflecting this diversity will lead to better therapeutic strategies. By linking proteomic data and clinico-pathological parameters, we are identifying proteins informative for features of clinical relevance to cancer patients. To obtain proteome data, we developed a large format system for two-dimensional difference gel electrophoresis and an application of highly sensitive fluorescent dye for laser microdissection. Following the comprehensive proteome study of more than 1,000 surgical specimens and corresponding clinico-pathological data, we concluded that the proteome reflects the major cancer phenotypes such as histological differentiation, poor prognosis, and response to treatment. Furthermore, we found that certain single proteins predict the clinical outcome in many types of malignancies including esophageal cancer, lung adenocarcinoma, gastrointestinal stromal tumor, hepatocellular carcinoma, Ewing's sarcoma and osteosarcoma. By monitoring biomarker proteins to predict the clinical outcome, we will be able to optimize the therapeutic strategy, either by intensifying treatment or by avoiding over-treatment. The results of the cancer proteomics studies are integrated into a proteome database named Genome Medicine Database of Japan Proteomics. The above establish proteomics as a primary tool in the development of novel diagnostic modalities for personalized medicine.

Introduction

Cancer is a diverse disease. Cancer patients exhibit different response to treatment, even when they are diagnosed at the same clinical stage. Existing diagnostic modalities have obvious limitations in predicting response to treatment; certain populations of patients may lose the opportunity to receive more intensive treatment that would improve their prognosis while others may receive over-treatment. The next level of molecular diagnostics is needed, by which the clinical outcome will be predicted before treatment initiation and therapeutic strategy will then be individually optimized on the basis of the prognosis.

The proteome is, in a sense, a functional translation of the genome, directly regulating tumor behavior. Genomic aberrations of cancer cells are transcribed to the transcriptome and translated to the proteome, thus determining cancer phenotypes. It is obvious that proteomic features can more directly reflect tumor characters than genomic contents. Many lines of evidence have demonstrated discordance between mRNA and protein expression (Chen et al., 2002). In addition, examining DNA sequences and measuring mRNA expression do not ac-

curately predict the status of posttranslational modifications such as phosphorylation and glycosylation, which play a key role in regulating the malignant behavior of cancer cells. Therefore, proteomic studies can generate unique data on the way the information contained in the genome is finally expressed. Taken together, the proteome can be a rich source for biomarker identification. With this notion, we are conducting cancer proteomics studies aimed at biomarker development.

Two-dimensional gel electrophoresis (2D-PAGE) has unique advantages over other proteomic modalities such as mass spectrometry and protein array. With 2D-PAGE, the overall features of the proteome are visualized due to the demonstration of the electrophoretic mobility of proteins, which in turn reflects the effects of posttranslational modifications on the physical properties of proteins, while at the same time quantitative and qualitative protein expression data are generated in such a way that multiple samples can be compared. 2D-PAGE data can be integrated with the corresponding biomedical information in a database that will thus allow a more comprehensive understanding of the proteomic background of

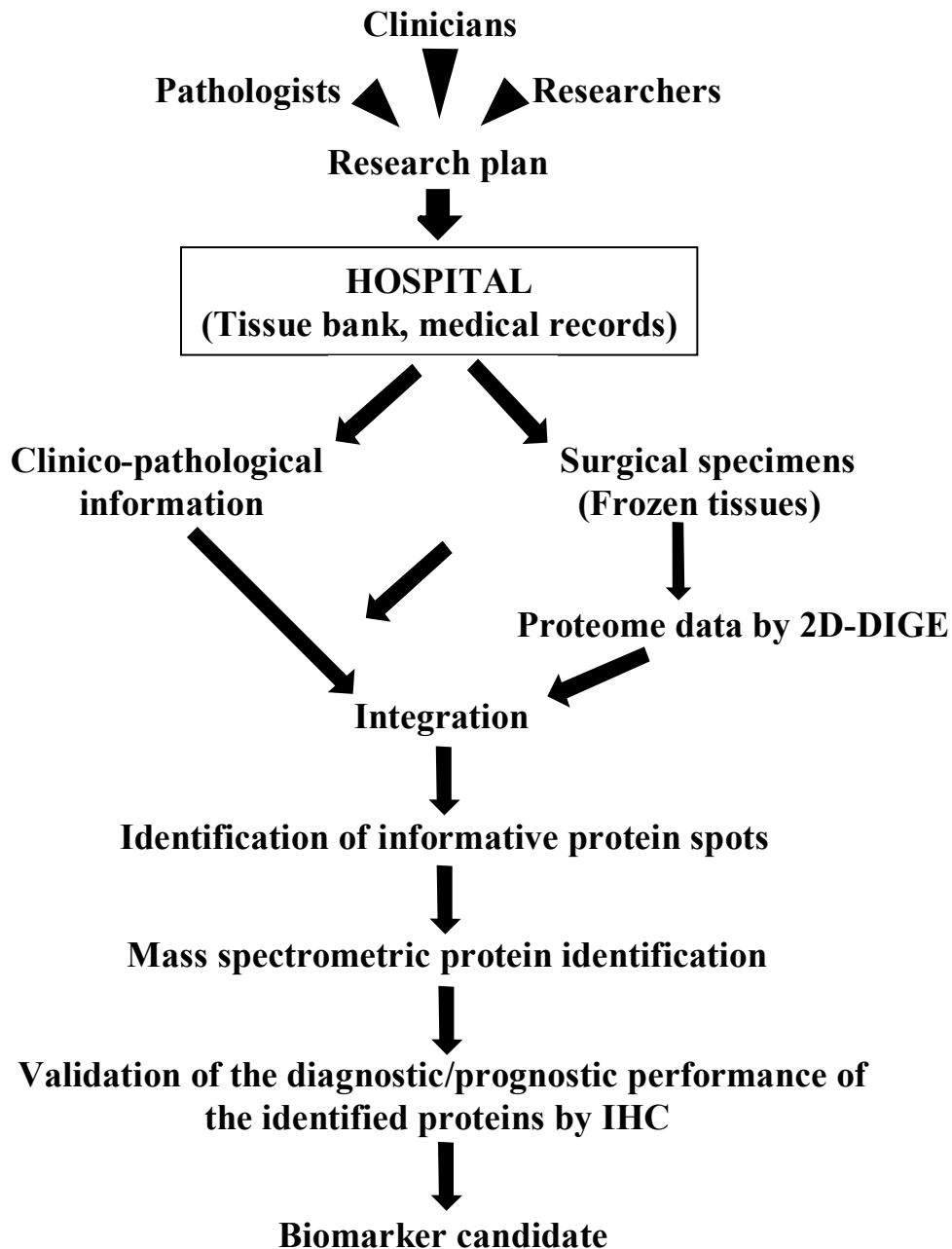


Figure 1: The workflow of cancer proteomics studies for biomarker development. 2D-DIGE generates the proteome data from the clinical samples, which are then examined in relation to the clinico-pathological information of the donors. Informative protein spots are identified using a data-mining approach, and the corresponding proteins are determined by mass spectrometry. Specific antibodies against the identified proteins are used in subsequent validation studies, and will also be used as part of the clinical examination.

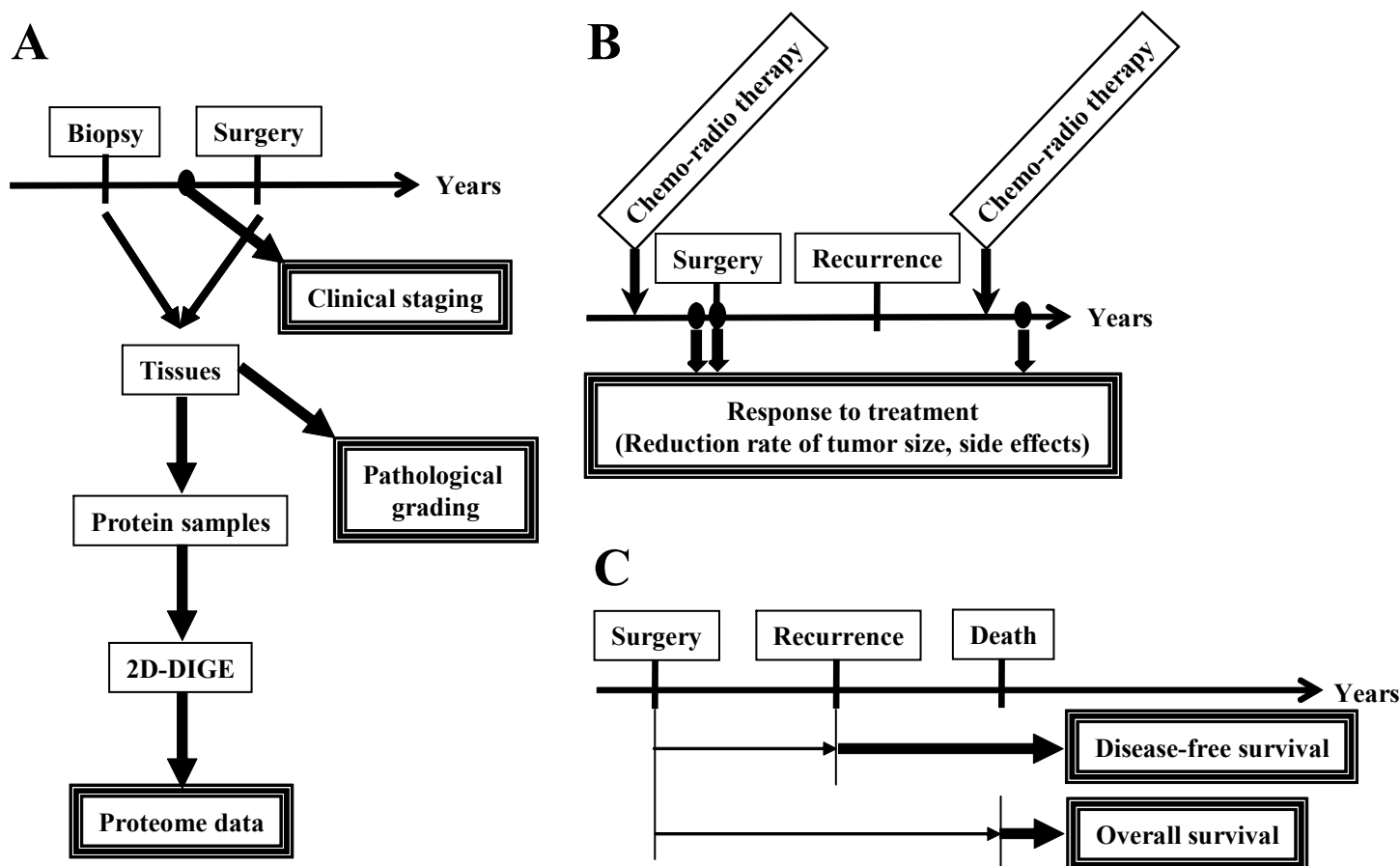


Figure 2: The clinico-pathological information for cancer proteomics for biomarker development. A. Following clinical staging and pathological grading of the surgically resected tumor tissues, protein samples extracted from them are used to obtain proteome data. B. The response to chemo-radio therapy can be monitored by measuring the reduction rate of the tumor size and the degree of necrosis, and by monitoring the side effects. C. The prognosis of the patients is transformed to numerical data by measuring the time to recurrence and the survival period.

cancer phenotypes. Recently, a new application of 2D-PAGE, two-dimensional difference gel electrophoresis (2D-DIGE) (Unlu et al., 1997) (GE Healthcare Biosciences), was established. In 2D-DIGE, multiple protein samples are labeled with different fluorescent dyes, mixed together and co-separated on single 2D-PAGE gels. By running different samples in single gels, gel-to-gel variations can be canceled out, and the proteome data can be obtained in a reproducible way. 2D-DIGE is one of the most dramatic innovations of 2D-PAGE in its more than 30 years' history.

In this paper, we introduce our proteomics approach for biomarker development at the National Cancer Center (Kondo and Hirohashi, 2006), and demonstrate the recent results in our laboratory. The application of a highly sensitive fluorescent dye for laser microdissected tissues (Kondo et al., 2003) and a large format 2D system (Kondo and Hirohashi, 2006) for 2D-DIGE are our

advantageous methods for cancer proteomics. We investigated several types of malignancies and developed practical biomarkers that were found to be applicable in the clinical setting with specific antibodies.

The research goal and strategy of cancer proteomics

The purpose of our proteome studies is to reveal the proteomic background of cancer diversity and develop biomarkers to predict response to therapy. Using 2D-DIGE, we generate protein expression profiles of surgical specimens and determine the sets of protein spots that are most significantly associated with clinico-pathological features, using a data-mining approach (Figure 1).

In our biomarker development strategy, proteome data are generated from frozen tissues obtained before treatment is initiated, and integrated with clinical data acquired after treatment. The clinico-pathological informa-

tion concerning the tumor tissues and patients is critical in cancer proteomics. Figure 2 summarizes the types of data that can be used in cancer proteomics. Experiments are always designed in a way that will allow the application of promising biomarkers in practice by medical doctors and pathologists. The prioritization of the clinico-pathological data for data-mining and the construction of patient groups for comparative biomarker studies require clinical and pathology experience. Because the expected results are almost determined when the sample

groups are created, the experimental design is most critical in biomarker development. With this notion, and with the clinical problems in mind, we have established a productive collaboration between basic researchers, clinicians and pathologists.

Protein expression profiling by 2D-DIGE

We extensively use 2D-DIGE for protein expression profiling. In 2D-DIGE, the different samples are labeled with different fluorescent dyes, mixed together and co-

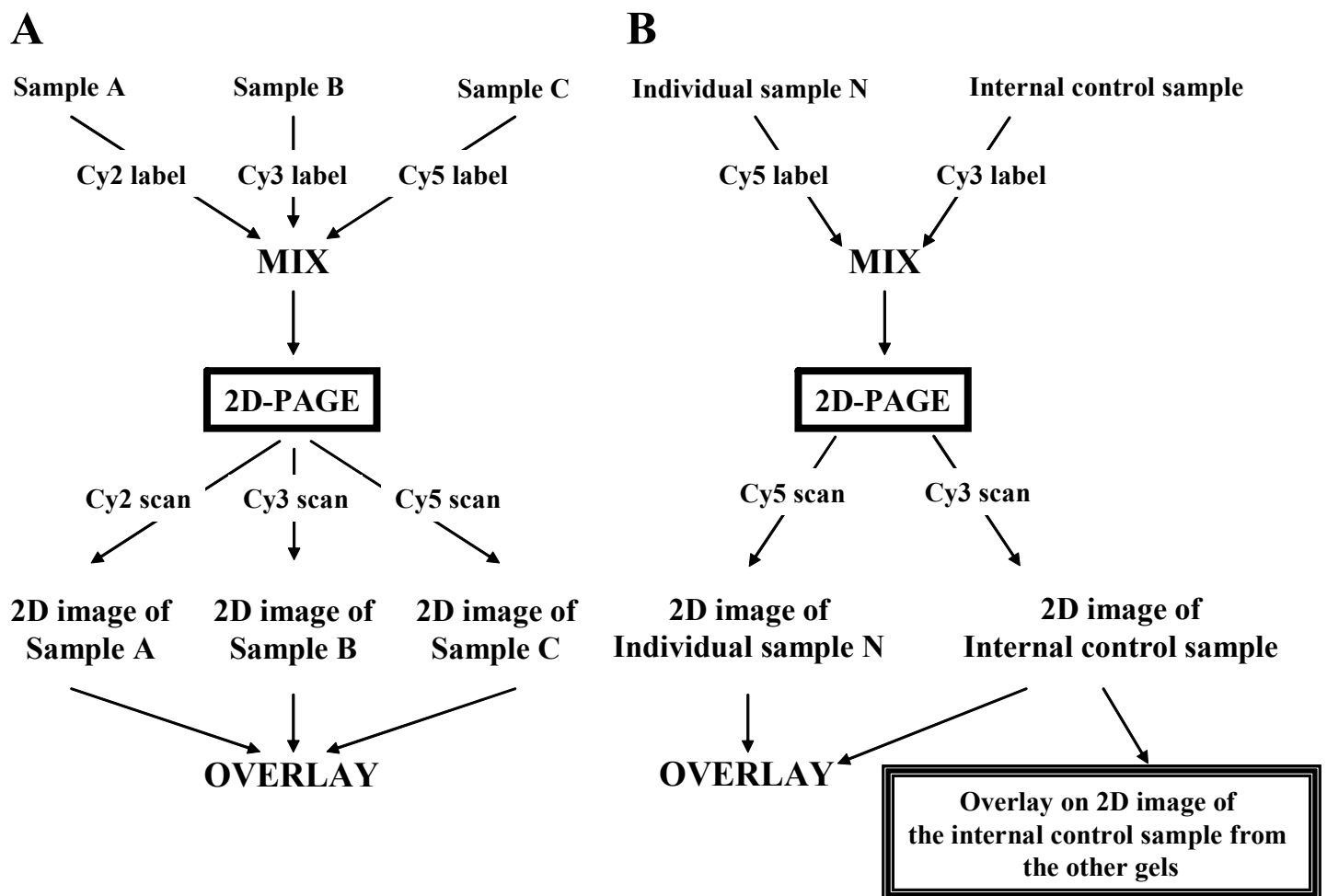


Figure 3: Methodology employed for 2D-DIGE experiments using multiple samples. A. Three samples are labeled with Cy2, Cy3, and Cy5 respectively, mixed together, then separated by 2D-PAGE. Laser scanning for each sample generates the 2D images, which are then overlaid so that they can be examined comparatively. B. The individual sample and the internal control sample are labeled with Cy5 and Cy3 respectively, mixed together and separated by 2D-PAGE. The same procedure is repeated for all individual samples. The Cy3 images of the internal control sample in the different gels are compared to normalize gel-to-gel variations. The ratio of the Cy5 and Cy3 spot intensity in the same gels is considered as the normalized intensity of the protein spots.

separated in the same 2D gel (Unlu et al., 1997) (Figure 3A). The gel is then scanned using a laser scanner at the appropriate wavelength generating a 2D image for each individual sample. Because multiple samples are separated in the same gel, gel-to-gel variations are evened out so that consistent results are generated allowing comparative studies. The disadvantage of this method is that we need as many fluorescent dyes as the number of samples; in biomarker development, we have to examine many samples to obtain conclusive results, while the number of fluorescent dyes currently available for 2D-DIGE is limited.

To solve this problem, we employed the following protocol (Kondo and Hirohashi, 2006) (Figure 3B) which generates 2D-DIGE data from multiple samples using only two fluorescent dyes. In this protocol, an internal control sample is created by mixing a small portion of all individual samples used in the experiment. This internal control sample is labeled with Cy3 fluorescent dye, while the individual samples are labeled with Cy5 fluorescent dye. The differently labeled samples are mixed together, and separated in individual gels for each sample. 2D images are then generated for the individual samples and the common internal control sample using a laser scanner. Because all 2D gels generate the same 2D image of the common internal control sample, by normalizing Cy5 intensity with Cy3 intensity, gel-to-gel variations are compensated.

A large format 2D gel can generate a higher number of protein spots, and this has even been possible in conventional 2D-PAGE. However, due to the fragility of the large format 2D gels, it is extremely difficult to perform silver staining on them. In 2D-DIGE, we can obtain the gel image by scanning the gel between the glass plates with a laser scanner; the fragility of the gels is therefore not a concern, as far as the size of the gel does not exceed the scan area of the laser scanner. The dimensions of the working gel area are 24 x 36cm, as we use a 24cm length IPG gel for the first dimension separation. A 2D gel image generated from this device includes approximately 5,000 protein spots (Kondo and Hirohashi, 2006).

In practice, the number of 2D images that can be produced in one file by the image analysis software is limited. A large number of 2D images may cause software problems, such as unexpected shut-down/freezing or remarkably slow speed for each operation. To overcome this, we divide the images into several groups and analyze them in different files. After the image analysis is completed, the results are extracted in xml or excel file format, which can then be integrated into one file. Once the data are integrated into one file, any type of data-mining software, which are basically developed for DNA microarray studies, can be used. Thus, with this approach,

the limitations associated with the use of the image analysis software are overcome.

Proteomics tools for the study of surgical specimens

Laser microdissection is a critical tool for tissue proteomics. Tumor tissues may contain many types of non-tumor cells such as normal epithelial or mesenchymal cells, inflammatory cells, stromal cells, as well as plasma contained in vascular structures. When these tissues are homogenized together in a sample for the purpose of protein extraction, the protein expression profile will reflect both the ratio of the number of cells of the different populations and the different protein contents in the individual cells. Because both the ratio of the different types of cells and the degree of vascularity are variable even within tumors of the same type, the results of protein expression studies cannot be reproducible when the heterogeneity of the tumor tissues studied is higher than a certain level.

Laser microdissection (LMD) is one solution to this problem (Emmert-Buck et al., 1996). With LMD, the desired population of cells is recovered under microscopic observation with the use of a laser beam. The idea to use 2D-PAGE for LMD samples was published in 1999 (Banks et al., 1999). But due to the low sensitivity of silver staining, several hours or days were required to recover an adequate number of cells for 2D-PAGE by LMD (Craven et al., 2002). Therefore, the combined use of LMD with 2D-PAGE was not practical for biomarker studies, where more than 100 samples are required to be examined.

We found a solution to this problem when, in 2001, we obtained an ultra highly sensitive fluorescent dye from the R&D team in Amersham Bioscience to develop a novel application of laser microdissection. The sensitivity of the dye was much higher than that of conventional silver staining, so we considered it may be applicable for laser microdissected tissues. We subsequently successfully developed a novel application, reported it to Amersham Biosciences, and published the results in 2003 (Kondo et al., 2003). Since then, LMD is a routine method in our laboratory. The dye is now commercially available with the name of CyDy DIGE Fluor saturation dye from GE Healthcare Biosciences. The detailed protocol was published in one of our recent publications (Kondo and Hirohashi, 2006). Other research groups have since used the application of CyDy DIGE Fluor saturation dye for several types of malignancies, including pancreatic cancer (Sitek et al., 2005) and cervical cancer (Greengauz-Roberts et al., 2005).

Other tips for proteomics studies using 2D-DIGE

One of the rate limiting-steps in 2D-DIGE experiments

is protein identification by mass spectrometry. Protein identification requires multi-step procedures including running preparative gels, recovering target protein spots, extracting peptides from the gel by in-gel digestion, and running mass spectrometry. To increase the success rate of protein identification, we optimized the in-gel digestion protocol as previously described (Kondo and Hirohashi, 2006).

With previous protocols, once proteins were labeled with CyDy DIGE Fluor saturation dye, the success rate of protein identification by MALDI TOF MS was dramatically decreased, since CyDy DIGE Fluor saturation dye labels all the reduced cysteine residues of proteins. We assumed that the decreased success rate is probably due to the fact that the dye inhibits the ionization of peptides that have cysteine residues and are labeled by the dye. In addition, the ionization of cysteine-negative peptides may also be inhibited because these peptides are surrounded by the dyes that label cysteine positive peptides on the same spot on the MALDI target plates. After several trials, we found that LC-MS/MS was definitely required for effective protein identification for the proteins labeled with CyDy DIGE Fluor saturation dye. Liquid chromatography prior to mass spectrometric examination may separate the dye-free peptides from the labeled ones, resulting in effective protein identification.

Quantitative proteome database for 2D-DIGE data, clinico-pathological data, and mass spectrometry protein identification

In proteomic studies across different types of malignant tumors, we found that certain proteins were repeatedly identified as having significant correlation with clinico-pathological data. In contrast, the alteration or presence of other proteins was unique to certain malignancies. These observations led us to construct a proteome database using 2D-DIGE data. Our proteome database is part of the Genome Medicine Database of Japan (Genome Medicine Database of Japan), which was originally developed to include genome and transcriptome data. Our proteome database includes the normalized intensity of protein spots, the annotation for protein spots by mass spectrometry, and supporting information about protein identification. We have published the beta-version of the database, which includes the 2D-DIGE data of pancreatic cancer cell lines and mass spectrometric protein identification for more than 1,000 protein spots. The same set of clinical sample data is under preparation for publication.

Recent examples of identified practical biomarkers

We are investigating several types of malignancies including esophageal, lung, , pancreatic, and colorectal cancer,, hepatocellular and cholangiocellular carcinoma, ma-

lignant mesothelioma, and soft-tissue sarcomas (including gastrointestinal stromal tumor, osteosarcoma, Ewing's sarcoma, and rhabdomyosarcoma). These malignancies were selected for proteomic study considering the clinical importance, the expected practical output, and the availability of samples and clinical data. The commitment of medical doctors and pathologists to the experimental procedure at the early stage is also unique to our project. A number of medical doctors who are keen to develop novel diagnostic modalities participate in the proteomics experiments we are conducting for two or three years each. The success of the cancer proteomics projects is partly due to their practical ideas that are based on clinical experience, their enthusiasm to benefit the patients and their hard work.

Recently, we found that pftin expression significantly correlates with better prognosis of gastrointestinal stromal tumor (GIST) patients (Suehara et al., 2008). GIST is the most common type of mesenchymal tumor of the gastrointestinal tract, characterized by positive expression of c-kit (Hirota et al., 1998; Hirota et al., 2000). The first choice of treatment for GIST patients is surgical resection. Recent studies established that imatinib (Gleevec, Novartis), a tyrosine kinase inhibitor, has anti-tumor effects in GIST (Demetri et al., 2002). A large-scale clinical study revealed that postoperative treatment of the patients with imatinib had suppressive effects on metastasis (Nilsson et al., 2007). It is of practical interest to detect the patients who are going to have metastasis as early as possible, so that they can receive intensive treatment with imatinib, while the rest of the patients may avoid over-treatment.

We compared the 2D-DIGE data from primary tumors between the patients who survived more than 2 years post-surgery and those that died within 2 years (Suehara et al., 2006). The former had histopathological features denoting lower grade malignancy. The 2D-DIGE study demonstrated that 45 protein spots showed statistically significantly different intensity between these two sample groups ($p < 0.01$, more than 2 fold difference). Mass spectrometric protein identification revealed that 25 unique proteins corresponded to these 45 protein spots. Interestingly, pftin corresponded to eight of the 45 protein spots. Western blotting and immunohistochemistry using an antibody against pftin demonstrated perfect correlation between pftin expression and clinical outcome. A further immunohistochemical study on pftin on 210 GIST cases successfully validated these results; within a 10 year observation period, 95% of patients with pftin positive primary tumors survived without developing metastases post surgery, while 20% of the patients with pftin negative primary tumors developed metastases post surgery. These observations established pftin as a novel prognostic biomarker in GIST.

By measuring the expression level of pftin, we will be able to select the patients who are likely to develop metastases post surgery, who will benefit from adjuvant chemotherapy. Further clinical studies will clarify the utility of pftin for personalized medicine in GIST.

Similar results are currently obtained for many of the malignancy types listed above. We consider that, albeit a powerful tool for biomarker discovery, 2D-DIGE is not suitable for routine clinical examination; running large format 2D gels manually in the hospitals is not a practical idea. The development of automated 2D machines has been repeatedly proposed in the past, but it is still doubtful whether such machines can be used in a clinical setting. Biomarkers can however be used as part of the clinical examination by using more cost-effective, easily operative, and already present and popular tests, such as immunohistochemistry and ELISA. With this notion, we make it a rule to validate the results of our 2D-DIGE studies using specific antibodies against the identified proteins that may be proposed as biomarkers.

Conclusions

We have established a proteomics laboratory at the National Cancer Center. The combined employment of 2D-DIGE and related applications, such as large format 2D-PAGE and laser microdissection, are technically unique to our proteomics approach. With the notion that involvement of medical doctors and pathologists is essentially needed for the development of practical biomarkers, we established a cancer biomarker team for individual malignant tumors. The research output is novel candidate biomarkers; their diagnostic and prognostic performance have been validated in several hundred cases using antibodies. We strongly believe that these efforts will benefit cancer patients in the very near future.

Acknowledgement

This work was supported by a grant from the Ministry of Health, Labor and Welfare and by the Program for the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation of Japan.

References

1. Banks RE, Dunn MJ, Forbes MA, Stanley A, Pappin D, et al. (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—preliminary findings. *Electrophoresis* 20: 689-700. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, et al. (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 1: 304-

313. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)

3. Craven RA, Totty N, Harnden P, Selby PJ, Banks RE (2002) Laser capture microdissection and two-dimensional polyacrylamide gel electrophoresis: evaluation of tissue preparation and sample limitations. *Am J Pathol* 160: 815-822. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Demetri GD, von Mehren M, Blanke CD, Van den Abbeele AD, Eisenberg B, et al. (2002) Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med* 347: 472-480. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, et al. (1996) Laser capture microdissection. *Science* 274: 998-1001. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Genome Medicine Database of Japan. <http://gemdbj.nibio.go.jp>
7. Greengauz-Roberts O, Stoppler H, Nomura S, Yamaguchi H, Goldenring JR, et al. (2005) Saturation labeling with cysteine-reactive cyanine fluorescent dyes provides increased sensitivity for protein expression profiling of laser-microdissected clinical specimens. *Proteomics* 5: 1746-1757. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, et al. (1998) Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* 279: 577-580. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Hirota S, Isozaki K, Nishida T, Kitamura Y (2000) Effects of loss-of-function and gain-of-function mutations of c-kit on the gastrointestinal tract. *J Gastroenterol* 12: 75-79. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. Kondo T, Hirohashi S (2006) Application of highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dyes) to laser microdissection and two-dimensional difference gel electrophoresis (2D-DIGE) for cancer proteomics. *Nat Protoc* 1: 2940-2956. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Kondo T, Seike M, Mori Y, Fujii K, Yamada T, et al. (2003) Application of sensitive fluorescent dyes in linkage of laser microdissection and two-dimensional gel electrophoresis as a cancer proteomic study tool. *Proteomics* 3: 1758-1766. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Nilsson B, Sjolund K, Kindblom LG, Meis-Kindblom JM, Bumming P, et al. (2007) Adjuvant imatinib treatment improves recurrence-free survival in patients with high-risk gastrointestinal stromal tumours (GIST). *Br J Cancer* 96: 1656-1658. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Sitek B, Luttes J, Marcus K, Kloppel G, Schmiegel W,

- et al. (2005) Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* 5: 2665-2679. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
14. Suehara Y, Kondo T, Fujii K, Hasegawa T, Kawai A, et al. (2006) Proteomic signatures corresponding to histological classification and grading of soft-tissue sarcomas. *Proteomics* 6: 4402-4409. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Suehara Y, Kondo T, Seki K, Shibata T, Fujii K, et al. (2008) Pftin as a prognostic biomarker of gastrointestinal stromal tumors revealed by proteomics. *Clin Cancer Res* 14: 1707-1717. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2071-2077. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)