

Chapter 11

Application of 2D-DIGE in Cancer Proteomics Toward Personalized Medicine

Tadashi Kondo and Setsuo Hirohashi

Summary

Two-dimensional difference gel electrophoresis (2D-DIGE) is an advanced variation of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE); protein samples are labeled with different fluorescent dyes, mixed and separated by 2D-PAGE. 2D-DIGE solves major inherent drawbacks of 2D-PAGE, demonstrating great utility in biomarker studies. Biomarker development requires quantitative, reproducible, highly sensitive and high-throughput experimental platforms, and 2D-DIGE meets these criteria. Here we demonstrate the advantages of 2D-DIGE and discuss the possibilities 2D-DIGE offers for further, more comprehensive proteome studies.

Key words: Cancer, Personalized medicine, Biomarker, Proteome, Database

1. Introduction

Biomarker identification is a key technology in the effort to improve the clinical outcome of patients with cancer. Cancer is a diverse disease. The response to treatment and prognosis after therapy vary between patients, and existing diagnostic technologies do not always demonstrate such important disease features accurately. The patients diagnosed as being at the same clinical stage often demonstrate different response to treatment and have varying survival periods. Pathologic grading does not always correlate with clinical outcome either. For these reasons, treatment with the best-optimized therapy for each individual patient, namely personalized medicine, has long required the development of the

next level of diagnostic tools. Cancer is a disease of the genome. Genomic aberrations result in the transformation of normal cells into fully malignant tumor cells, and the type of these aberrations determines cancer phenotypes. Many lines of evidence suggest that specific molecules and pathways govern the malignant behaviors of tumor cells, such as unexpected early metastasis after curative surgery and disease progress after chemotherapy. By monitoring such molecules or pathways, we can use them as biomarkers to predict life-threatening events after therapy and to individual treatment (**Fig. 1**).

Recent technological advances have enabled the performance of comprehensive studies at the genome, transcriptome, and proteome level. The application of novel technologies to DNA, RNA, and protein samples of tumor tissues provided us new insights into the molecular background of cancer. Global studies revealed that cancer phenotypes demonstrate distinctive genome, transcriptome, and proteome profiles, allowing the possibility of novel cancer classification (*1*). At the same time, these global studies shed light on curious relations between the genome, transcriptome, and proteome. Aberrations in the genomic content of cancer cells are not always reflected in the transcriptome, and, similarly, aberrations in the transcriptome are not necessarily reflected in the proteome (*2, 3*). Many lines of evidence suggest that the copy number of DNA sequences is not always

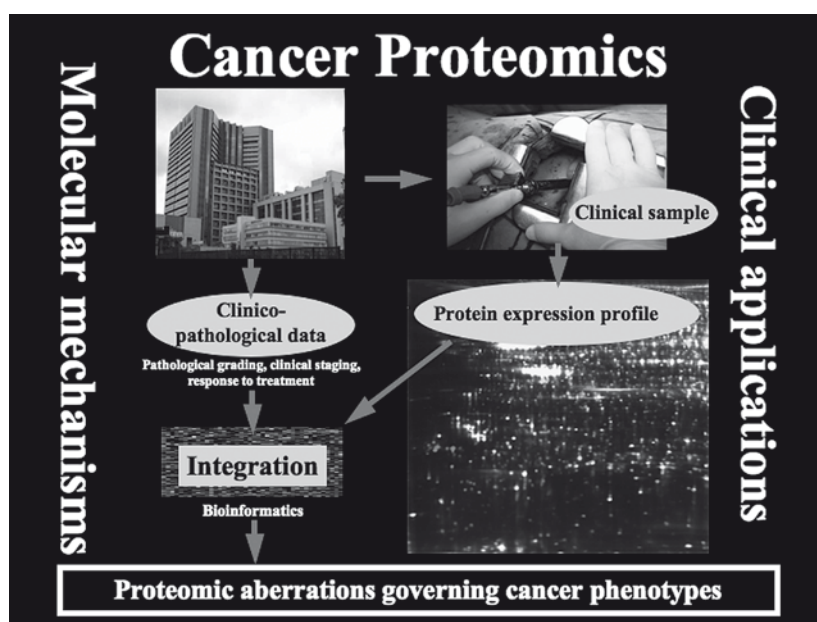


Fig. 1. Work flow and purpose of cancer proteomics. Proteomic profiles of tumor tissues are integrated with clinico-pathological data to identify the proteomic aberrations governing cancer phenotypes. The goals of cancer proteomics are understanding of the molecular mechanisms of cancer diversity and developing applications of potential clinical benefit.

parallel with the expression level of the corresponding mRNA, and that the expression level between mRNAs and corresponding proteins may show discordance (4, 5). In addition, proteins do not function in the form that they are translated from mRNA; they are post-translationally modified by small molecules such as phosphates, acetylates, lipids, and glycans. The functionality of proteins largely depends on their localization in the cells and tissues, and is determined by their association with other proteins or nucleic acids. These properties are aberrantly regulated in cancer cells. The aberrant proteome may also generate cancer-specific autoantibodies in the cancer patients (6–14). Considering that the functional translation of the genome is the proteome, these observations suggest that genomic aberrations do not directly determine the process of carcinogenesis and cancer phenotypes. As reading DNA sequences and measuring RNA levels alone do not presently predict the status of the proteome, it is proteomic studies that will provide the unique information about the molecular mechanisms underlying cancer. By examining the proteomic features, we may be able to identify the molecules directly regulating the phenotypes of individual cancers. Therefore, we consider that the proteome is a rich source for biomarkers for personalized medicine.

With this notion, we conduct cancer proteomics studies to develop biomarkers for personalized medicine. Using large-scale clinical sample sets and linking the acquired proteome data to clinicopathological data, we try to identify the proteomic aberrations which may govern cancer phenotypes (Fig. 1). We have found protein groups or specific proteins the expression of which correlates highly with response to treatment (15), the development of unexpected metastases after surgical operation (16), the number of lymph node metastases (17), and shorter survival (18), all of which should be good biomarker candidates for personalized medicine.

To identify biomarker candidates by global studies, we need a technology with the following technical characteristics. First, it must measure protein expression levels in a quantitative way. Most oncogene products and their downstream molecules exist in the cells of the different cancer types, and their expression level often determines the characteristics of malignant cells. Therefore, we need the quantitative data to determine the cut-off value to distinguish samples from certain tumor types from others. Second, the data should be generated in a reproducible way. We need to examine proteome data across many clinical samples to obtain conclusive results in a statistically valid way. The reproducibility of the proteome data is indispensable for the statistical analysis. Third, it should uncover as much proteome data as possible; the more proteins are detected, the more likely it is to succeed in identifying biomarkers. Fourth, data should be generated

in a high-throughput way. Measuring the expression level of proteins is only the very first phase of biomarker development, and thus may need to be concluded within a reasonable time frame. Fifth, biomarkers should be identified so that they can then be measured using other, simpler, and less costly tools, since comprehensive technologies are not always suitable for application in a clinical setting.

In this manuscript, we reviewed a novel gel-based proteomics technology, two-dimensional difference electrophoresis (2D-DIGE). 2D-DIGE is an advanced variation of two-dimensional gel electrophoresis (2D-PAGE). In 2D-DIGE, the protein samples are labeled with different fluorescent dyes, mixed together, and separated according to isoelectric points and molecular weights. 2D-DIGE solves many drawbacks of gel-based proteomics and facilitates cancer proteomics. We found that 2D-DIGE meets the above-mentioned criteria for a biomarker development tool. In the National Cancer Center Research Institute, approximately 2,000 large format 2D-DIGE gels are annually ran to identify biomarker candidates. The detailed protocols were published in our previous report (19). We will describe the advantages that, based on our experience, we believe 2D-DIGE offers with respect to biomarker identification studies, and discuss the critical issues relating to biomarker development.

2. 2D-PAGE in Cancer Proteomics

2D-PAGE is the most popular proteomics tool (20–22). In 2D-PAGE, proteins are separated according to the individual physiological characteristics of proteins, namely their isoelectric point and molecular weight. Following the colorimetric staining of gels, the expression level of proteins is quantified by measuring the staining intensity of the corresponding protein spots. Alternatively, the cellular proteins are labeled with either ^{35}S -methionine (23), ^{32}P -phosphate (24), ^{14}C -containing amino acids (25) or ^{125}I -Na (26) in tissue culture conditions, and the isotope-labeled proteins are subjected to 2D-PAGE and then detected by exposing the gel to an X-ray film. Protein expression can be quantified by measuring the intensity of the protein spots on X-ray film. The identification of the proteins is nowadays achieved by mass spectrometry and database search. A series of proteomic experiments using 2D-PAGE has been established with innovative improvements such as the application of immobilized pH gradient gels, detergents for membrane proteins and sample application methods (27–32), and many relevant protocols have been published (20, 33). There have also been many reports in the last three

decades regarding its application to cancer research (34–36). However, besides its popular use and fruitful academic results, there are a number of limitations inherent in 2D-PAGE.

First, in 2D-PAGE, single gels separate single samples, so that gel-to-gel variations can affect the apparent expression level of proteins. Because the intensity of protein spots reflects both the amount of corresponding proteins and experimental variations, it may not be exactly the same for identical proteins in different gels. We can somehow compensate for the experimental variations by running multiple gels. However, the compensation is not perfect, and running multiple gels multiplies labor intensity. Second, silver staining was known to be the most sensitive detection method next to radioisotope labeling method, and widely used in 2D-PAGE experiments. However, the protein detection by silver staining is the most rate-limiting and labor intense step in 2D-PAGE. Silver staining takes at least several hours, uses large quantities of water, and requires well-trained operators and a considerable amount of laboratory space for staining trays. Moreover, to store the gel images for further analysis, the gel needs to be scanned by an optical scanner or photographed, requiring additional time. Silver staining enhances the other inherent problems of 2D-PAGE as follows. Third, 2D-PAGE does not uncover the entire proteome; similar to other proteomic tools, and despite many efforts to increase the number of observable proteins (20) only a limited number of proteins can be visualized. The number of protein spots is almost parallel to the area of gel; the larger the gel, the higher the number of protein spots on the gel. Thus, the large format gel is one of the most powerful solutions to increase proteome coverage (37). However, polyacrylamide gels are very fragile and are often damaged during the multistep procedures of the colorimetric protein detection methods such as silver staining. Multiple fractionations (38) and the use of narrow range isoelectric focusing gels (39) also increase the number of observable proteins, but these approaches multiply the number of gels used and thus labor intensity, particularly when colorimetric gel staining is used. Fourth, in 2D-PAGE with conventional silver staining, 100 μ g of protein are usually needed for single gels; the study of the proteome of tumor tissues requires higher sensitivity. Tumor tissues consist of heterogenous populations of cells that include nontumor cells, and each cell population probably has different proteome content. Once they are homogenized together for the purpose of protein extraction, the expected proteome pattern would reflect both the ratio of the number of cells of the different populations and the different protein contents in the individual cells. It is hard to know which factors more dominantly affect the intensity of protein spots, and accurate protein profiling requires the separate collection of specific cell populations before protein extraction. One possible remedy is laser

microdissection (LMD), by which the cells are recovered under microscopic observation (40). LMD has been used in conjunction with 2D-PAGE to specifically study tumor cells in tumor tissues (41). However, because of the limited sensitivity of silver staining, several hours or days were required in order to recover adequate numbers of cells by LMD (42), suggesting that LMD was not a practical tool for biomarker studies in which a relatively large number of tumor tissues was examined to generate conclusive results. When the tissues are histologically homogeneous and the use of LMD is not required, samples may be homogenized for protein extraction. However, in many cases, the amount of tumor tissues obtained from the hospital is very limited, and higher sensitivity is required. The isotope-labeling method has much higher sensitivity than silver staining, but requires a special laboratory set up to reduce the risk of exposure to hazardous materials and may not be suitable for routine experiments. In addition, metabolic labeling needs living cells, meaning that frozen tissues cannot be used. Fifth, the identification of proteins corresponding to protein spots has been sometimes very difficult using Edman degradation. However, this limitation has largely been solved with the recent use of mass spectrometry and database searching. Taken all together, classical 2D-PAGE had obvious limitations, mainly in the associated spot detection methods. We have thus long needed a novel technology to address the problems inherent in 2D-PAGE.

3. Advanced 2D-PAGE, 2D-DIGE, and its Application for Biomarker Development

Recently, two-dimensional difference gel electrophoresis (2D-DIGE) (43) was introduced as an innovative 2D-PAGE technology that can solve its afore-mentioned problems. **Figure 2** demonstrates the basic application of 2D-DIGE in the comparison of two protein samples. In this application, each individual protein sample is labeled with Cy3 and Cy5 dye respectively. The fluorescent dyes are designed so that the electrophoretic mobility of proteins labeled with different fluorescent dyes is almost identical. After stopping the labeling reaction, the samples are mixed together and separated in one 2D-PAGE gel. After gel electrophoresis, the 2D-PAGE images of the two samples are obtained by scanning the gel with a laser for Cy3 and Cy5 respectively. As these protein samples are separated on the same gel, there are no gel-to-gel variations. This protocol can be used for multiple protein samples depending on the number of different fluorescent dyes used. Three fluorescent dyes with different emission and excitation wavelengths are currently commercially available for

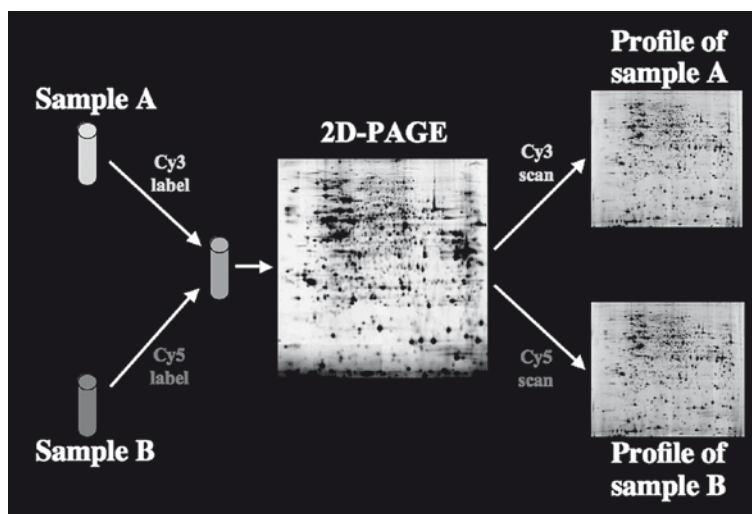


Fig. 2. Basic protocol of 2D-DIGE for two protein samples. The different protein samples are labeled with different fluorescent dyes (Cy3 and Cy5), mixed together and separated on the same 2D gel. After gel electrophoresis, the gels are scanned with laser at the appropriate wavelength for Cy3 and Cy5. A single gel can generate two 2D images, so that gel-to-gel variations are canceled out. We can compare as many protein samples as the number of available fluorescent dyes.

Table 1
Fluorescent dyes used in 2D-DIGE

Name of dye	Dyes commercially available ¹	Amino acids labeled	Sensitivity ^{2,3}
CyDye DIGE Fluor minimal dye ²	Cy2, Cy3, Cy5	A small % of lysine residues	Equivalent to silver staining
CyDye DIGE Fluor saturation dye ³	Cy3, Cy5	100% of cystein residues	One hundred times higher than silver staining

¹GE Healthcare

²Ref. 43

³Ref. 45

2D-DIGE, (Cy2, Cy3, and Cy5; **Table 1**), meaning that three samples can be compared using this 2D-DIGE application. This way, 2D-DIGE can address the most major problem inherent in 2D-PAGE, gel-to-gel variations. In addition, the spot intensity is measured as a fluorescent signal, the nature of fluorescence allowing the data obtained to have wide dynamic range.

3.1. 2D-DIGE Allows Multiple Sample Comparisons

Cancer is, however, a genetically diverse disease and therefore more than three samples need to be examined in cancer proteomics to compensate for the genetic variations. In general, results from one-by-one comparisons do not provide any meaningful information in clinical studies. The application shown in **Fig. 3** can be used to examine more protein samples than the number of fluorescent dyes used. In this application, a mixture of a small portion of all protein samples is used as the internal control sample. The internal control sample is aliquoted into small tubes and stored in a deep freezer until use. This internal control sample is labeled with Cy3, while the individual samples are labeled with Cy5. These two differently labeled protein samples are mixed together and separated by 2D-PAGE. All gels thus generate the 2D image of the internal control sample as the Cy3 image. Therefore, by normalizing Cy5 intensity with that of Cy3 intensity for all protein spots, we can cancel out gel-to-gel variations. **Figure 3** shows that 100 protein samples can be examined using the same number of gels. Gel-to-gel variations can be further decreased by running each sample in multiple gels and then using the mean or median of the obtained normalized intensity values. Using this

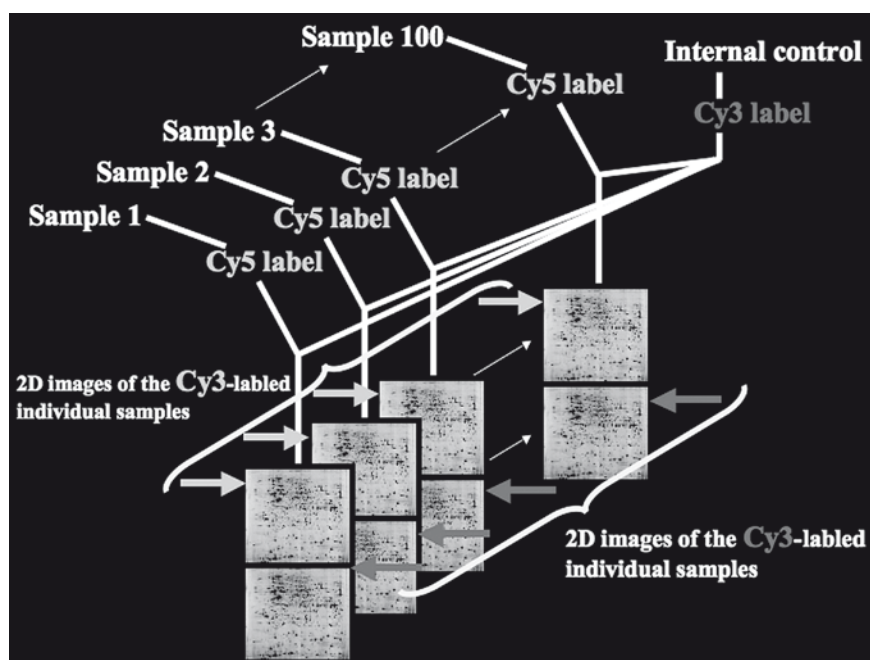


Fig. 3. Advanced protocol of 2D-DIGE for a number of protein samples larger than the available fluorescent dyes. The internal control sample and individual samples are labeled with Cy3 and Cy5, respectively. The differently labeled protein samples are mixed together and separated on individual gels. After gel electrophoresis, the gels are scanned with laser at the appropriate wavelength for Cy3 and Cy5. As all gel scans generate the Cy3 image that represents the proteomic profile of the internal control sample, gel-to-gel variations are compensated by normalizing Cy5 images with Cy3 images for each gel. We can compare more protein samples than the available fluorescent dyes.

protocol, we can perform 2D-DIGE using two fluorescent dyes and as many protein samples as necessary (19).

**3.2. 2D-DIGE
Facilitates High-
Throughput Gel-Based
Proteomics
Applications**

Studies in which several hundred clinical samples need to be examined by 2D-PAGE require a large number of gels to be run. The most popular method to detect proteins separated by conventional 2D-PAGE is silver staining, that is time consuming and labor intensive. In contrast, in 2D-DIGE, because all proteins are labeled with fluorescent dyes, simply laser scanning the gel generates a gel image within 1 h and much less laboriously. By using multiple laser scanners in a parallel way, many gels are run in a high-throughput way. At the National Cancer Center, we achieve such high-throughput proteomics results by using multiple electrophoresis devices and six laser scanners (Typhoon Trio, GE Healthcare), enabling us to annually run approximately 2,000 large format 2D gels for biomarker studies (19).

We found that this high-throughput facilitated the use of not only a large number of surgical specimens, but also of fractionated samples (44). We fractionated plasma samples in order to detect low-abundance plasma proteins, by subjecting them to multiple chromatograms such as the immuno-affinity column and the ion-exchange column, and then examined the fractionated samples by 2D-DIGE. Multidimensional fractionation is widely used in plasma proteomics to reduce the complexity of protein samples and increase the number of detectable proteins. Single samples were separated into eight fractions, and then separated on 2D-DIGE gels containing 3,890 protein spots. In this experiment, we found that laser scanning enabled high-throughput experiments while also decreasing labor intensity.

**3.3. 2D-DIGE Enables
More Comprehensive
Proteomics
Applications**

The fragility of gels is not of substantial consequence in 2D-DIGE, because the gels are scanned sandwiched between the two low-fluorescent glass plates that were used for electrophoresis. Therefore, we can run a gel as large as the scanning area of the laser scanner. We found that the number of protein spots detected was parallel with that of the gel area, probably because the increased resolution increases the visualization of the protein spots that were behind neighboring protein spots in small format 2D gels. We constructed a large format 2D electrophoresis device for the purpose of increasing the number of protein spots. The gel area of this device is twice that of the second largest electrophoretic device, EttanDalt II (GE Healthcare) and allows us to observe approximately 5,000 protein spots on single gel images using the DeCyder software (GE Healthcare) (19).

**3.4. 2D-DIGE Allows
Highly Sensitive
Proteomics
Applications**

Currently, two types of fluorescent dyes are available from GE Healthcare (Table 1) (43, 45). We first reported that the use of an ultra highly sensitive fluorescent dye (CyDye DIGE Fluor saturation dye, GE Healthcare) enabled protein expression profiling even

when using samples with scarce protein amounts such as those from laser microdissected tissues (46). As proteins cannot be amplified like DNAs, highly sensitive detection systems such as protein-labeling with highly sensitive fluorescent dyes are the only remedy for samples with low-protein content. The high sensitivity of the CyDye DIGE Fluor saturation dye means that few cells are needed for proteomics applications, and microdissection does not take a long time. We applied this method to study adenoma in min mice (46), lung cancer (47), esophageal cancer (17) and Uemra et al, manuscript in preparation) and hepatocellular carcinoma (Orimo et al, manuscript in preparation), while other research groups later followed with studies using tissues of pancreatic cancer (48), gastric cancer (49), and transgenic mice (50). The manufacturer (GE Healthcare) has released a basic protocol for protein-labeling, and we have published the detailed protocol for laser microdissection, protein extraction from microdissected tissues, and labeling of the extracted proteins (19). Currently, two types of CyDye DIGE Fluor saturation dyes are commercially available, Cy3 and Cy5 (Table 1). By using the protocol for multiple samples as mentioned above (Fig. 3), we can compare multiple microdissected samples using these two dyes (17). One noticeable character of the CyDye DIGE Fluor saturation dye is that it changes the electrophoretic mobility of proteins after labeling. Therefore, we cannot compare the 2D image generated by the CyDye DIGE Fluor saturation dye with that by the CyDye DIGE Fluore minimal dye or silver staining.

We found that the use of samples with minute protein content can improve the quality of 2D gel images. In routine experiments with silver staining, 100 μg of protein are applied to 2D-PAGE. Protein samples include the substances that may interfere with 2D-PAGE; those include lipids, glycans, nucleic acids and salts. As they hinder the reproducibility of the 2D image, their amount should not exceed a critical interference threshold. Although they can be removed by precipitating proteins, low-abundance proteins are lost during the precipitation procedure. In 2D-DIGE with CyDye DIGE Fluor saturation dye, only 1 μg of protein is enough when using EttanDalt II size gels (24 cm $\times$ 20 cm). Because 1 μg of protein sample includes 1/100 of the interfering substances contained in the protein amount used for 2D-PAGE, we obtained high-quality 2D images in a constant way.

The problems inherent to 2D-PAGE that are solved with the use of 2D-DIGE are summarized in Table 2.

Table 2
Drawbacks of classical 2D-PAGE addressed by 2D-DIGE

Drawbacks of classical 2D-PAGE	Solution by 2D-DIGE
Low reproducibility due to gel-to-gel variations	Mixing the differently labeled protein samples cancels the experimental variations
Time-consuming and labor-intensive spot detection method	The gel image is produced by laser scanning within 1 h in a less labor-intensive way
Limited proteome coverage	Large-format gel, multi-fractionation, combined narrow range pI gels
Requires large protein content	Saturation dye is 100 times more sensitive than silver staining requiring significantly lower protein content
Difficult protein identification	Fluorescently labeled proteins are compatible with mass spectrometry

4. Data-Mining for Cancer Proteomics Using 2D-DIGE Data and Clinicopathological Information

2D-DIGE generates a huge amount of proteome data that have to be collected and analyzed. For instance, for a lung cancer proteomics study, we are currently examining 250 tumor tissues to develop biomarkers to predict lymph node metastasis and survival of lung cancer patients; each sample is applied on gels in triplicate, and each gel generates approximately 5,000 protein spots that produce quantitative data. These protein expression data are to be examined in relation to clinicopathological parameters such as TNM grading, histological classification, and the patients' response to treatment and survival period. With this method, visual inspection of the gels would not be of any benefit.

Image-analysis software for 2D-DIGE, such as DeCyder (GE Healthcare) and Progenesis SameSpot (Nonlinear Dynamics) are commercially available. The software is used to normalize Cy5 images with Cy3 images, then export the image data as numerical data in the format of an xml file or an xls file. Data-mining software that were basically developed to study DNA microarray data, such as Expressionist (GeneData, Switzerland), are then used to identify proteomic signatures that correlate with certain clinicopathological parameters and to rank these protein spots for further validation studies. In addition, proteome-based cancer classification can be achieved using clustering algorithms (51).

The use of these tools enables the collection of novel information that could not be achieved via visual inspection of the gels.

Image analysis is one of the rate-limiting steps of 2D-DIGE experiments as the different gel images have to be matched by manual inspection. We found that the use of the Progenesis SameSpot software dramatically shortens the image analysis process. Progenesis SameSpots transforms the gel images so that focal gel distortions are corrected and the groups of protein spots in every small area are aligned between the different gels. For more details the reader is referred to the Nonlinear Dynamics homepage (<http://www.nonlinear.com>). The image analysis software for 2D-DIGE has improved significantly in the last several years, providing an added advantage to the use of 2D-DIGE, as it enables the quantitative study of a large number of 2D gels with ease. Inclusion of optimized-multivariate analysis tools will be the next challenge in the refinement of image analysis software.

5. Protein Identification from 2D-DIGE Gels

Proteins corresponding to the protein spots of interest are identified by mass spectrometry (19). By comparing the 2D image generated for analytical purpose using a sample with low-protein content with that generated for preparative purposes using a sample containing 100 µg of protein, we can identify the target protein spots on the 2D image from a preparative gel. The target protein spots are then collected into 96-well PCR plates using an automated spot recovery machine. Before subjecting the recovered proteins to mass spectrometry, the proteins in the gel are digested by specific protease into peptides. It is generally hard to extract proteins from polyacrylamide gels. However, once the proteins are digested to peptides, it is easy to extract them from the gel. The protocol of digesting the proteins in a gel matrix and extracting the digested peptides was established with the name of “in-gel digestion” (52). In this protocol, the recovered gels are extensively washed with a buffer to remove the remaining detergent, and are then repeatedly shrunk and reswollen with treatment with organic solvent and buffer. By overnight treatment with trypsin, the proteins in the gel plug are digested to peptides. The peptides are then extracted with an organic solvent. We have extensively optimized the in-gel digestion protocol so that protein identification can be successfully achieved for most protein spots (19).

In our experience, the identification of proteins with MALDI TOF MS (oMALDI Q-STAR, ABI) can be achieved when they

are labeled with the CyDye DIGE Fluor minimal dye but is very hard when they are labeled with the CyDye DIGE Fluor saturation dye. The two dyes have different characteristics; the former labels a small portion of lysine residue while the latter labels all reduced cystein residue. Therefore, the proteins labeled with the CyDye DIGE Fluor saturation dye include more fluorescent dye. We speculate that the fluorescent dye may have a suppressive effect on the ionization of peptides. The tryptic digests spotted on the MALDI plates include both labeled and unlabeled peptides, and the ionization of unlabeled peptides is also hindered by the presence of neighboring dyes. In the case of the CyDye DIGE Fluor minimal dye, protein identification can be achieved by MALDI TOF MS because only a limited number of lysine residues is labeled and the peptide samples may contain a lower amount of fluorescent dye. In contrast, the peptide samples labeled with CyDye DIGE Fluor saturation dye may contain a larger amount of fluorescent dye, resulting in reduced ionization efficiency. Alternative to MALDI TOF MS, we found that efficient protein identification can be achieved with the use of LC-MSMS even for proteins labeled with CyDye DIGE Fluor saturation dye. We speculate that the labeled peptides are separated from the unlabeled ones by LC separation, and then ionized by MS, so that the fluorescent dye does not hinder protein identification using the unlabeled peptides. Using the in-gel digestion protocol and LC-MSMS, we have already identified more than 3,000 protein spots labeled by the CyDye DIGE Fluor saturation dye (19).

6. Practical Biomarker Discovery Using 2D-DIGE Data

2D-DIGE is a powerful technology for biomarker discovery. We found that even beginners in basic research can master it in a short period following an adequate training program. However, it will be hard to optimize 2D-DIGE in terms of cost performance so that it is used as a clinical examination tool. After a small number of biomarker candidates is identified, one does not have to run large format gels; instead, we need a tool to survey these specific proteins across a large number of clinical samples in a reproducible, cost-effective, and less labor-intensive way. Therefore, desirably, the developed biomarkers can be used in hospitals using the existing equipment without or with minimal modifications. Such applications are also required at the validation phase in biomarker development. To establish the candidates as novel biomarkers, we may need to demonstrate their diagnostic or prognostic value in several hundred clinical samples, in collaboration with clinicians,

in which case we are expected to examine the candidates in a high throughput and more cost-effective way.

To validate the candidate biomarkers, specific antibodies against the identified proteins are used. As immunohistochemical examinations and enzyme-linked immuno assays (ELISA) are performed routinely in hospitals, once the relevant antibodies are obtained, the existing devices can be utilized for the examination. Indeed, we have successfully validated 2D-DIGE results in immunohistochemical (16) and ELISA (15) studies. However, the antibodies do not always work in the expected way. This is due to the fact that in 2D-DIGE experiments a specific protein may repeatedly appear in different protein spots owing to post-translational modifications. As a consequence, each protein spot may represent a particular protein isoform and not the total amount or all isoforms of each protein, and thus some isoforms may alone be identified as biomarker candidates, even when the total expression level of that particular protein is constant between sample groups. In this case, as the antibodies used are not specific to the protein isoform identified as the candidate but recognize all isoforms of the particular protein, the immunohistochemical and ELISA results are not consistent with those of 2D-DIGE. We are developing monoclonal antibodies to use our research results as diagnostic tools. The development of antibodies is now one of the most rate-limiting steps in biomarker studies and is a challenge that requires novel technologies or methods to be addressed.

7. 2D-DIGE Data Proteome Database

One of the unique characteristics of 2D-PAGE is that the data can be integrated in a database which in turn will facilitate biomarker development in the validation phase. In transcriptome studies, the expression data are deposited in public cyber space such as the Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo) and Oncomine databases (<http://www.oncomine.org>) (53) and are freely downloaded to validate the results of individual small-scale studies. This way, so-called metaanalysis enables large-scale expression studies (54). These research tools are compatible with studies that have common experimental platforms, such as DNA microarrays. In contrast, we do not have common experimental platforms in proteomic studies. Although there are many proteome databases using 2D-PAGE data (<http://www.expasy.org/world-2dpage>), they cannot be used as tools for biomarker development without problems. Most 2D databases include the intensity value for only a small number of protein spots. The number of

annotated protein spots is generally very small and is often less than 100. Each sample group usually consists of only one sample, and biological and clinical information is not provided in most cases. The experimental methods such as that for protein extraction, isoelectric focusing in the first dimension separation, size of gel in the second dimension separation, and staining protocol vary between the databases. Because of its high reproducibility, 2D-DIGE may provide a common experimental platform for biomarker development studies. We consider that the construction of a proteome database using 2D-DIGE data is the next challenge in our project.

For this reason, we are currently constructing a public proteome database named GeMDBJ Proteomics (Genome Medicine Database of Japan Proteomics, <http://gemdbj.nibio.go.jp>). GeMDBJ is an integrative database that includes genome and transcriptome data. Our database will include the quantitative proteome data generated by 2D-DIGE from tumor tissues of different cancer types such as lung, esophageal, liver, and colon cancer, soft-tissue sarcoma, bone tumors, and malignant mesothelioma. Annotation data acquired by LC-MS/MS will be added to as many protein spots as possible. We have published the beta-version, which includes 2D-DIGE data from nine pancreatic cancer cell lines and two normal pancreatic duct cell lines. The database includes annotations for approximately 1,100 protein spots. Proteome data from tumor tissues from studies conducted in our laboratory will be constantly up-loaded to the database.

8. Further Possibilities of 2D-DIGE

Using the present large format 2D gel device without prior fractionation, we can observe up to 5,000 protein spots with the DeCyder software, which may correspond to up to 2,500 unique proteins as identified by mass spectrometry (19). We have already identified many interesting proteins the expression of which correlates in a statistically significant way with important clinico-pathological parameters of tumors. However, the protocol still needs to be improved to expand the proteome coverage while keeping the advantageous characteristics of 2D-DIGE.

Such future modifications may include enlarging the gel size, fractioning the protein samples prior to electrophoresis, and using narrow range isoelectric focusing gels. These steps have been applied in the classical 2D-PAGE to increase the number of observable protein spots, and can therefore be employed to 2D-DIGE experiments. Because of the afore-mentioned advantages of 2D-DIGE, we believe it will be easier to carry out these modifications in 2D-DIGE compared with the classical 2D-PAGE.

9. Collaboration Between Basic Researchers, Clinicians, Pathologists, and Industry is Critical in Biomarker Studies

Needless to say that technology alone does not develop biomarkers. We believe that collaboration between basic researchers, clinicians, and industry is critical in biomarker development. Basic researchers may know the details of the technological aspects of proteomics and how to investigate the proteome. They can develop novel technologies and establish novel concepts in biology using proteome data. However, they may not evaluate the results from a clinical perspective or select biomarkers that can best benefit cancer patients. Clinicians, in contrast, have a better understanding of what may improve the clinical outcome for the patients and can collect the required clinical samples and information. They can suggest the types of biomarkers that can optimize existing therapeutic protocols based on their experience. More importantly, they are the potential users of novel biomarkers, and biomarkers should be developed in the way that the final products can be accepted in a clinical setting. In addition to the medical benefits from the use of biomarkers, one may need to consider whether industry will be interested in commercializing the proposed novel diagnostic tools. Basic researchers and clinicians may not be familiar with the business aspects of biomarker development. Therefore, the involvement of industry partners in the project from an early phase is also critical to the successful development of biomarkers.

For these reasons, our research group includes basic researchers including a bioinformatics specialist, clinicians, pathologists, and industry partners, in a way that best-optimizes the use of 2D-DIGE related methods for biomarker development studies (Fig. 4).

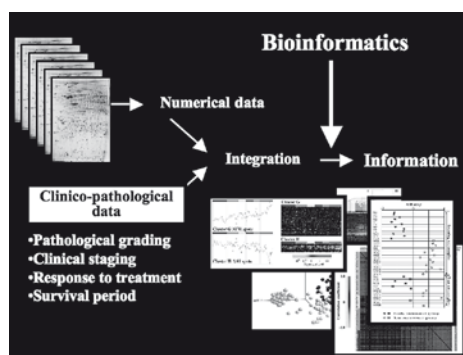


Fig. 4. Application of bioinformatics tools to 2D-DIGE data analysis. Proteome data consisting of the normalized spot intensity values are exported from the image analysis software and their correlation with clinicopathological data examined. Using informatics tools including clustering algorithms and machine-learning methods, a novel cancer classification based on proteome data is established, and key proteomic features and proteins corresponding to biomarker candidates are identified.

10. Conclusions

This chapter demonstrates the benefits of the use of 2D-DIGE in cancer proteomics for biomarker development. **Table 1** lists the commercially available fluorescent dyes for 2D-DIGE. **Table 2** summarizes the major drawbacks of the classical 2D-DIGE and how 2D-DIGE addresses them. **Table 3** shows how 2D-DIGE meets the criteria for a biomarker development tool. All detailed protocols were published in our previous paper (19). We should keep in mind that although 2D-DIGE is one of the most advanced versions of 2D-PAGE, additional modifications can further improve its performance. Finally, we would like to emphasize that a practical research strategy and a translational research mind are the most important factors to make the best use of 2D-DIGE for biomarker development.

Table 3
Advantageous characteristics of 2D-DIGE for biomarker studies

Technical requirements for proteomic tools to identify biomarker candidates	How 2D-DIGE meets the requirements
Quantitativity	1. The use of a common internal control sample enables normalization of gel-to-gel variations so that spot intensity directly reflects the protein expression level 2. The fluorescent signal that measures spot intensity has wide dynamic range
Reproducibility	The use of a common internal control cancels out the gel-to-gel variation, which hinders reproducibility
Coverage	The use of large format gels, pre-fractionation and multiple use of narrow range isoelectric focusing gels enable more comprehensive proteomic studies
Throughput	Gel images are produced by laser scanning within a short time and less laboriously
Transactivity	The proteins corresponding to protein spots are identified by mass spectrometry

Acknowledgments

This work was supported by a grant from the Ministry of Health, Labor, and Welfare and by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation of Japan. We appreciate Dr. Hisao Asamura (National Cancer Center Hospital) for an excellent photograph of surgical operation.

References

1. Rhodes DR, Chinnaiyan AM. (2005) Integrative analysis of the cancer transcriptome. *Nat Genet* **37**, S31–7.
2. Hanash SM. (2003) Operomics: molecular analysis of tissues from DNA to RNA to protein. *Clin Chem Lab Med* **38**, 805–13.
3. Orntoft TF, Thykjaer T, Waldman FM, Wolf H, Celis JE. (2002) Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas. *Mol Cell Proteomics* **1**, 37–45.
4. Varambally S, Yu J, Laxman B, et al. (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393–406.
5. Chen G, Gharib TG, Huang CC, et al. (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* **1**, 304–13.
6. Wang X, Yu J, Sreekumar A, et al. (2005) Autoantibody signatures in prostate cancer. *N Engl J Med* **353**, 1224–35.
7. Nam MJ, Madoz-Gurpide J, Wang H, et al. (2003) Molecular profiling of the immune response in colon cancer using protein microarrays: occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. *Proteomics* **3**, 2108–15.
8. Hanash S. (2003) Harnessing immunity for cancer marker discovery. *Nat Biotechnol* **21**, 37–8.
9. Shin BK, Wang H, Hanash S. (2002) Proteomics approaches to uncover the repertoire of circulating biomarkers for breast cancer. *J Mammary Gland Biol Neoplasia* **7**, 407–13.
10. Le Naour F, Brichory F, Misek DE, Brechot C, Hanash SM, Beretta L. (2002) A distinct repertoire of autoantibodies in hepatocellular carcinoma identified by proteomic analysis. *Mol Cell Proteomics* **1**, 197–203.
11. Le Naour F, Misek DE, Krause MC, et al. (2001) Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin Cancer Res* **7**, 3328–35.
12. Brichory FM, Misek DE, Yim AM, et al. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc Natl Acad Sci U S A* **98**, 9824–9.
13. Brichory F, Beer D, Le Naour F, Giordano T, Hanash S. (2001) Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. *Cancer Res* **61**, 7908–12.
14. Prasannan L, Misek DE, Hinderer R, Michon J, Geiger JD, Hanash SM. (2000) Identification of beta-tubulin isoforms as tumor antigens in neuroblastoma. *Clin Cancer Res* **6**, 3949–56.
15. Okano T, Kondo T, Fujii K, et al. (2007) Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, and mutation in EGFR in lung adenocarcinoma. *Clin Cancer Res* **13**, 799–805.
16. Suchara Y, Suchara Y, Kondo T, et al. (2008) Pftin as a prognostic biomarker of gastrointestinal stromal tumors revealed by proteomics. *Clin Cancer Res* **14**, 1707–17.
17. Hatakeyama H, Kondo T, Fujii K, et al. (2006) Protein clusters associated with carcinogenesis, histological differentiation and nodal metastasis in esophageal cancer. *Proteomics* **6**, 6300–16.
18. Yokoo H, Kondo T, Okano T, et al. (2007) Protein expression associated with early intrahepatic recurrence of hepatocellular carcinoma after curative surgery. *Cancer Sci* **98**, 665–73.

19. 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**, 2940–56.
20. Gorg A, Weiss W, Dunn MJ. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665–85.
21. O'Farrell PH. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007–21.
22. Klose J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231–43.
23. Bravo R, Celis JE. (1982) Human proteins sensitive to neoplastic transformation in cultured epithelial and fibroblast cells. *Clin Chem* **28**, 949–54.
24. Fey SJ, Larsen PM, Celis JE. (1983) Evidence for coordinated phosphorylation of keratins and vimentin during mitosis in transformed human amnion cells. Phosphate turnover of modified proteins. *FEBS Lett* **157**, 165–9.
25. Bravo R, Celis JE. (1982) Up-dated catalogue of HeLa cell proteins: percentages and characteristics of the major cell polypeptides labeled with a mixture of 16 ¹⁴C-labeled amino acids. *Clin Chem* **28**, 766–81.
26. Litin BS, Grimes WJ. (1979) Two-dimensional electrophoresis of membrane proteins from normal and transformed cells. *Cancer Res* **39**, 2595–603.
27. Gorg A, Postel W, Gunther S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531–46.
28. Righetti PG. (1990) Immobilized pH Gradients: Theory and Methodology. Amsterdam: Elsevier.
29. Molloy MP, Herbert BR, Walsh BJ, et al. (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis* **19**, 837–44.
30. Rabilloud T, Adessi C, Giraudel A, Lunardi J. (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**, 307–16.
31. Herbert B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660–3.
32. Rabilloud T, Valette C, Lawrence JJ. (1994) Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* **15**, 1552–8.
33. Carrette O, Burkhard PR, Sanchez JC, Hochstrasser DF. (2006) State-of-the-art two-dimensional gel electrophoresis: a key tool of proteomics research. *Nature Protocols* **1**, 812–23.
34. Chen G, Gharib TG, Huang CC, et al. (2002) Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. *Clin Cancer Res* **8**, 2298–305.
35. Gharib TG, Chen G, Wang H, et al. (2002) Proteomic analysis of cytokeratin isoforms uncovers association with survival in lung adenocarcinoma. *Neoplasia* **4**, 440–8.
36. Hanash S. (2001) 2-D or not 2-D – is there a future for 2-D gels in proteomics? Insights from the York proteomics meeting. *Proteomics* **1**, 635–7.
37. Young DA, Voris BP, Maytin EV, Colbert RA. (1983) Very-high-resolution two-dimensional electrophoretic separation of proteins on giant gels. *Methods Enzymol* **91**, 190–214.
38. Gorg A, Boguth G, Kopf A, Reil G, Parlar H, Weiss W. (2002) Sample prefractionation with Sephadex isoelectric focusing prior to narrow pH range two-dimensional gels. *Proteomics* **2**, 1652–7.
39. Wildgruber R, Harder A, Obermaier C, et al. (2002) Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* **21**, 2610–6.
40. Emmert-Buck MR, Bonner RF, Smith PD, et al. (1996) Laser capture microdissection. *Science* **274**, 998–1001.
41. Banks RE, Dunn MJ, Forbes MA, 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.
42. 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**, 815–22.
43. Unlu M, Morgan ME, Minden JS. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–7.
44. Okano T, Kondo T, Kakisaka T, et al. (2006) Plasma proteomics of lung cancer by a linkage

- of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis (2D-DIGE). *Proteomics* **6**, 3938–48.
45. Shaw J, Rowlinson R, Nickson J, et al. (2003) Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics* **3**, 1181–95.
 46. Kondo T, Seike M, Mori Y, Fujii K, Yamada T, Hirohashi S. (2003) Application of sensitive fluorescent dyes in linkage of laser microdissection and two-dimensional gel electrophoresis as a cancer proteomic study tool. *Proteomics* **3**, 1758–66.
 47. Seike M, Kondo T, Fujii K, et al. (2005) Proteomic signatures for histological types of lung cancer. *Proteomics* **5**, 2939–48.
 48. Sitek B, Luttges J, Marcus K, et al. (2005) Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* **5**, 2665–79.
 49. Greengauz-Roberts O, Stoppler H, Nomura S, 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**, 1746–57.
 50. Wilson KE, Marouga R, Prime JE, et al. (2005) Comparative proteomic analysis using samples obtained with laser microdissection and saturation dye labelling. *Proteomics* **5**, 3851–8.
 51. Suehara Y, Kondo T, Fujii K, et al. (2006) Proteomic signatures corresponding to histological classification and grading of soft-tissue sarcomas. *Proteomics* **6**, 4402–9.
 52. Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem* **203**, 173–9.
 53. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–80.
 54. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, et al. (2007) Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9**, 443–54.