

Proteome Expression Database of Lung Adenocarcinoma: a segment of the Genome Medicine Database of Japan Proteomics

Seiji Kosaihi^{1,2}, Yukako Tsunehiro¹, Koji Tsuta³, Naobumi Tochigi⁴,
Akihiko Gemma², Setsuo Hirohahsi¹, Tadashi Kondo^{1*}

¹Proteome Bioinformatics Project, National Cancer Center Research Institute

²Fourth Internal Department of Medicine, Nippon Medical School

³Clinical Laboratory Division, National Cancer Center Hospital

⁴Pathology Division, National Cancer Center Research Institute

Abstract

Lung cancer is a leading cause of cancer death worldwide, and lung cancer proteomics studies have been carried out to reveal the molecular background of cancer phenotypes and to develop clinically relevant applications. Here, we report an open-access proteome expression database derived from the study of 262 lung cancer cases using data extracted by two-dimensional difference gel electrophoresis (2D-DIGE) and mass spectrometry. Proteins extracted from primary tumor tissues were labeled with CyDye DIGE Fluor saturation dye, and separated using a large format electrophoresis device, generating 3179 protein spots. Mass spectrometry following in-gel digestion identified 487 proteins corresponding to 721 protein spots. Multiple proteins were observed from single protein spots, and single proteins generated multiple protein spots, suggesting diversity of the proteome. The results of 2D-DIGE and protein identification, and part of the corresponding clinico-pathological data are freely accessible in the public proteome database Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics, <http://gemdbj.nibio.go.jp/dgdb/DigeTop.do>).

Keywords: Lung cancer proteomics; GeMDBJ proteomics; Proteome database; Two-dimensional difference gel electrophoresis (2D-DIGE)

Lung cancer is a leading cause of cancer death in Japan, claiming 55,000 lives annually, and is a major health problem in many countries. Despite the modern therapeutic strategies, early recurrence is common and the prognosis for patients with lung cancer is generally poor, with an overall 5-year survival rate for patients receiving treatment of only 14% (Hoffman et al., 2000). A more detailed characterization of the molecular background of the carcinogenesis and progression of lung cancer is required for obtaining information relevant to early tumor detection and for the development of novel targeted therapeutics.

Lung cancer proteomics studies have been conducted to identify the proteins that correspond to certain clinico-pathological parameters of value in lung cancer. An open-access proteome database is a useful platform to integrate the proteome data derived from different patients and different malignancies, allowing the proteomics community to share the proteome data. However, there is no proteome expression database practically applicable in cancer proteomics studies to date. For this reason, we constructed a proteome database for lung cancer using 262

surgically resected frozen tissue samples, two-dimensional difference gel electrophoresis (2D-DIGE) using highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dye) and an original large format electrophoresis device.

In 2D-DIGE, different protein samples are labeled with fluorescent dyes with different emission and excitation wavelength, mixed together and separated by two-dimensional gel electrophoresis (Unlu et al., 1997). By including a common internal control sample labeled with a fluorescent dye different from that for the individual samples, the gel-to-gel variations can be canceled out, and reproducible results can be expected across a large number of samples. 2D-DIGE can improve the aspects of classical 2D-PAGE that place critical limitations and provide a platform for unique applications such as for the use of laser microdissected tissues (Kondo et al., 2003; Kondo and Hirohashi, 2006). We have extensively applied 2D-DIGE to study samples derived from surgical specimens with an aim of developing clinically relevant biomarkers (Kondo, 2008).

In this study, primary tumor tissues from 262 lung cancer patients were subjected to lung cancer proteomics. The corresponding clinico-pathological information is available in GeMDBJ Proteomics (Figure 1) and includes the tumor size, status of lymph node metastases and pathological staging. This project was approved by the ethical board of the National Cancer Center and written informed consent was obtained from all donors.


Protein samples were prepared by homogenizing frozen lung cancer tissues as previously described (Kondo and Hirohashi, 2006). In brief, proteins were extracted using a urea lysis buffer (2 M thiourea, 6 M urea, 3% CHAPS, 1% Triton X-100) from tumor tissue powdered by a Multi-beads shocker (Yasui-kikai, Osaka, Japan). For preparative purposes, 100 micrograms of the extracted proteins were labeled with a CyDye DIGE Fluor saturation dye according to the manufacturer's instructions. For

***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 ext.3004; Fax: +81-3-3547-5298; E-mail: takondo@ncc.go.jp

Received September 26, 2009; **Accepted** November 23, 2009; **Published** November 24, 2009

Citation: Kosaihi S, Tsunehiro Y, Tsuta K, Tochigi N, Gemma A, et al. (2009) Proteome Expression Database of Lung Adenocarcinoma: a segment of the Genome Medicine Database of Japan Proteomics. *J Proteomics Bioinform* 2: 463-465. doi:10.4172/jpb.1000106

Copyright: © 2009 Kosaihi S, et al. 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.



Search Proteome Database Version 1.1

[Login](#)

■ Search by Gene
■ Search by Disease
■ Search Proteome Database
■ Data File Download

HOME
SITE MAP
HELP

■ Search by Protein
If you have specific proteins, click [here](#) to start search.

■ Search by Gel Image
If you would like to start search by selecting a gel image from our gel image list, start [here](#).

■ Overview
If you would like to know the overview of GeMDBJ Proteomics, start [here](#).

■ Link to 2D Database
If you would like to know the 2D-based proteome databases on the World Wide Web, start [here](#).

What is GeMDBJ Proteomics?

Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics) is an integrated proteome database for use in cancer research.

The following information is available in the database:

- Expression levels of proteins identified in surgical specimens and cell lines by two-dimensional difference gel electrophoresis (2D-DIGE)
- Protein spots identified by liquid chromatography tandem mass spectrometry (LC-MS/MS), with corresponding peptide information
- Clinical and biological information pertaining to the samples

GeMDBJ Proteomics has been developed as part of the [Proteome Bioinformatics Project](#) of the [National Cancer Center Research Institute, Tokyo](#). This database aims to contribute to proteome-based research for the benefit of human health and welfare, and will keep expanding as new proteome information becomes available. [>more](#)

[Contact us](#) | [Privacy Policy](#) | [Credits](#) | [Terms of Use](#)

Copyright © 2003-2009, The Study Groups of Millennium Genome Project for Alzheimer's disease, Cancer, Diabetes Mellitus, Hypertension, Asthma and Pharmacogenetics (SGMGP).

Figure 1: Database structure of GeMDBJ Proteomics. The proteins are searchable by their name or their localization on the master 2D gel image.

analytical purposes, the internal control sample was prepared by mixing a small portion of all 262 individual samples. Five micrograms of the internal control sample and the individual samples were labeled with Cy3 and Cy5 respectively, and mixed together. Then the labeled protein samples were separated by 2D-PAGE using a large format electrophoresis device (Kondo and Hirohashi, 2006). The gel images were obtained by scanning the gels with a laser scanner (Typhoon Trio, GE Healthcare Biosciences, Uppsala, Sweden) at the appropriate wavelength. All protein spots were numbered by the Progenesis SameSpots software (Nonlinear Dynamics, Newcastle, UK) according to the spot numbers in the master gel image. A typical 2D image with the merged and numbered protein spots is exhibited in the GeMDBJ Proteomics. Proteins in the recovered protein spots were subjected to in-gel digestion and the trypsin digests were subjected to liquid chromatography coupled with tandem mass spectrometry, using a Finnigan LTQ linear ion trap mass spectrometer (Thermo Electron Co., San Jose, CA) equipped with a nano-electrospray ion (NSI) source (AMR Inc., Tokyo, Japan). The Mascot software (version 2.1, Matrix science, London, UK) was used to search for the mass of the peptide ion peaks against the SWISS-PROT database. Proteins with a Mascot score of 34 or more were subjected to protein identification. When multiple proteins were identified in a single spot, the proteins with the highest number of peptides were considered as those corresponding to the spot, while the proteins with lower but significant scores were also recorded in the database. All procedures for protein identification were reported in our previous report (Kondo and Hirohashi, 2006).

The system reproducibility of 2D-DIGE was significantly high when we ran the same lung cancer tissue sample twice; the correlation coefficient of the intensity of the 3170 protein spots detected was 0.85, and the intensity of 3029 of these protein

spots (ie. of 95.5% of all spots detected) was scattered within a range of two fold differences from the mean. We randomly selected protein spots, and resulted in the positive identification of the proteins contained in 721 protein spots by mass spectrometry. The two-dimensional gel image and the results of protein identification as well as the supporting mass spectrometric data are exhibited in GeMDBJ Proteomics.

Among the 721 protein spots identified, we found that 391 protein spots contained multiple proteins, accounting for 45.8% of all protein spots with annotations. Figure 2A demonstrates the number of proteins that may be observed in a single protein spot. We had similar results in our database study for the pancreatic cancer proteome (Yamada et al., 2008), which, in turn, are consistent with the results of previous proteome studies (Campostrini et al., 2005; Westbrook et al., 2001). Multiple proteins may be detected in single protein spots probably because of the limited separation performance of 2D gels, the relatively large number of protein spots with detectable intensity, and the high sensitivity of protein identification by mass spectrometry. However, the protein overlap may not be a serious problem when we use two-dimensional gels for semi-quantitative comparative studies, because, as discussed by Hunsucker et al., (2006), only a few proteins contained at each location may contribute to the detectable signal due to the fact that the intensity of the rest is lower than the detection limit. Indeed, in our experience, gels with longer separation distance have a higher number of protein spots, as the protein spots that overlap in gels of smaller size are separated. However, to construct further experiments based on 2D-DIGE results, western blotting may need to be employed to confirm the contribution of each individual identified protein to the differential intensity of the spots between the sample groups.

We also found that 254 proteins were identified in multiple

protein spots, accounting for 65.3% of all identified proteins. We had similar results in our previous study on the pancreatic cancer proteome (Yamada et al., 2008). Figure 2B demonstrates the number of proteins only detected in single protein spots. Actin, vimentin, and enolase A, among others, appeared repeatedly in different protein spots, a finding that may be explained by the presence of alternative splicing or posttranslational modifications. These observations may indicate some unique advantages of gel-based proteomics; the proteins are separated reflecting their whole structure. Once proteins are cleaved with protease for mass spectrometric studies, such data would be otherwise missed. These observations may also reveal the presence of a critical problem in employing 2D gel-based proteomics for biomarker studies. In our experience, 2D-DIGE and SDS-PAGE/western blotting often exhibited discordant results. Such discordance may in part be due to the presence of protein spots other than those identified as containing biomarker candidates.

We are planning to up-load the proteome data derived from a range of malignancies to the GeMDBJ Proteomics. Presently, the GeMDBJ Proteomics includes the 2D-DIGE proteome data derived from pancreatic cancer cell lines, esophageal cancer, Ewing's sarcoma, lung adenocarcinoma, and malignant pleural mesothelioma tissues. Proteome data of other malignancies such

as colorectal cancer, hepatocellular carcinoma, cholangiocarcinoma, synovial sarcoma, osteosarcoma, rhabdomyosarcoma, and gastrointestinal stromal tumor and will be included soon. The GeMDBJ Proteomics will thus be the largest proteome expression database containing data from a wide range and number of clinical cases. The integration of proteome data of different malignancies will be our next challenge.

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. Campostrini N, Areces LB, Rappsilber J, Pietrogrande MC, Dondi F, et al., (2005) Spot overlapping in two-dimensional maps: a serious problem ignored for much too long. *Proteomics* 5: 2385-2395. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Hoffman PC, Mauer AM, Vokes EE (2000) Lung cancer. *Lancet* 355: 479-485. » [CrossRef](#) » [Pubmed](#)
3. Hunsucker SW, Duncan MW (2006) Is protein overlap in two-dimensional gels a serious practical problem. *Proteomics* 6: 1374-1375. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Kondo T (2008) Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE. *BMB Rep* 41: 626-634. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. 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](#)
6. 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](#)
7. Nawrocki A, Larsen MR, Podtelejnikov AV, Jensen ON, Mann M, et al., (1998) Correlation of acidic and basic carrier ampholyte and immobilized pH gradient two-dimensional gel electrophoresis patterns based on mass spectrometric protein identification. *Electrophoresis* 19: 1024-1035. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. 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](#)
9. Yamada M, Fujii K, Koyama K, Hirohashi S, Kondo T (2008) The proteomic profile of pancreatic cancer cell lines corresponding to carcinogenesis and metastasis. *J Proteomics Bioinform* 2: 1-18.
10. Westbrook JA, Yan JX, Wait R, Welson SY, Dunn MJ (2001) Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms. *Electrophoresis* 22: 2865-2871. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)

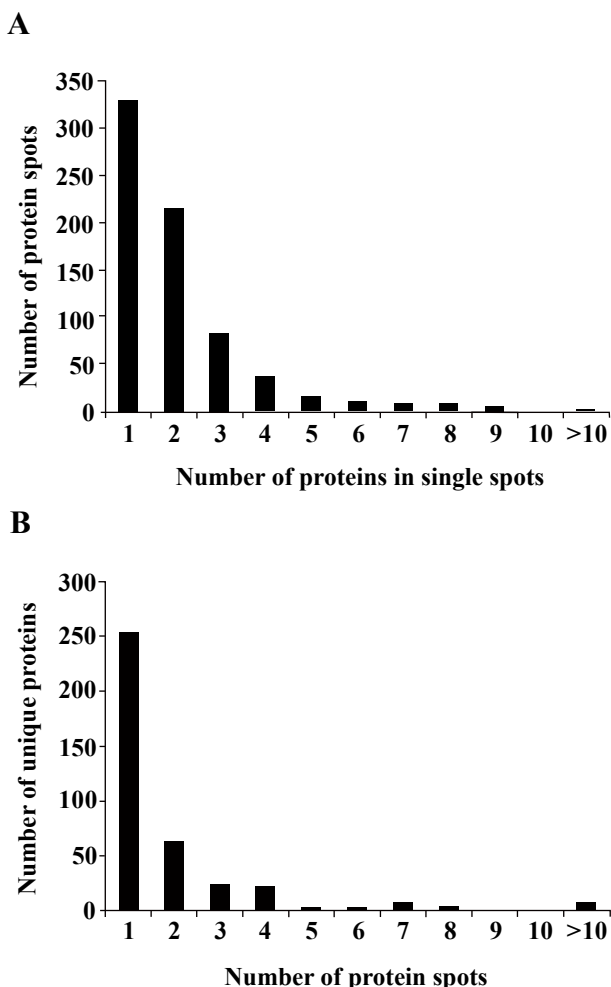


Figure 2: The characteristics of 2D-DIGE data. A. Single spots contain multiple proteins. The number of proteins included in the single spots is demonstrated. B. Single proteins may appear in multiple protein spots. The number of protein spots representing the same protein is demonstrated.