Proteome Expression Database of Ewing Sarcoma: A Segment of the Genome Medicine Database of Japan Proteomics

Kazutaka Kikuta1,2, Yukako Tsunehiro1, Akihiko Yoshida1, Naobumi Tochigi1, Setsuo Hirohahsi1, Akira Kawai1, Tadashi Kondo1*

1Proteome Bioinformatics Project, National Cancer Center Research Institute
2Orthopedics Division, National Cancer Center Hospital
3Clinical Laboratory Division, National Cancer Center Hospital
4Pathology Division, National Cancer Center Research Institute

Abstract

Ewing sarcoma is the second most common primary malignant bone tumor in children and adolescents worldwide. Here, we report an open-access proteome expression database of eight Ewing sarcoma cases using proteome data obtained 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 2431 protein spots. Mass spectrometry following in-gel digestion identified 330 protein spots corresponding to 220 proteins. Multiple proteins were observed from single protein spots, and single proteins generated multiple protein spots, suggesting diversity of the proteome observed by 2D-DIGE. The results of 2D-DIGE and protein identification by mass spectrometry, and of the corresponding clinico-pathological data such as prognosis after treatments are freely accessible in the public proteome database Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics, https://gemdbj.nibio.go.jp/dgdb/DigeTop.do).

Keywords: Ewing sarcoma proteomics; GeMDBJ proteomics; Proteome database; Two-dimensional difference gel electrophoresis (2D-DIGE); Mass spectrometry

Ewing sarcoma is the second most common primary malignant bone tumor in children and adolescents. The prognosis of the patients with Ewing sarcoma remains dismal despite of progress of intensive chemotherapy and local control protocols; 30-40% of patients with localized tumor and 80% of patients with metastatic tumor at diagnosis die due to disease progression within five years (Cotterill et al., 2000). More intensified first-line chemotherapy regimens and combinations of chemotherapy agents demonstrated improved clinical outcome. However, as such modern therapies often result in serious toxicity, risk-adapted treatment strategies have been required (Atra et al., 1997; Diaz et al., 2000; Bernstein et al., 2006; Engelhardt et al., 2007; McTierman et al., 2006). The studies on molecular background of malignant features of Ewing sarcoma, followed by identification of biomarkers to predict the responses to treatment, were conducted by global expression studies (Ohali et al., 2004; Cheung et al., 2007; Armengol et al., 1997; Hattinger et al., 2002; Schaefer et al., 2008). By a proteomics approach, we previously reported nucleophosmin was observed in several types of cancers (Tanaka et al., 1992; Nozawa et al., 1996; Zhang et al., 2004; Tsui et al., 2004). However, the prognostic utility of nucleophosmin was not reported in Ewing sarcoma until our proteomic study (Ohali et al., 2004; Cheung et al., 2007; Armengol et al., 1997; Hattinger et al., 2002; Schaefer et al., 2008), suggesting the unique advantage of proteomics.

An open-access database is a useful platform to integrate the proteome data to share the proteome data in a proteomics community. Such database should be beneficial especially for the studies on rare cancers such as Ewing sarcoma. However, there was no proteome expression database practically applicable in studies on Ewing sarcoma. For this reason, we constructed a proteome database for Ewing sarcoma using eight surgically resected frozen tissue samples, two-dimensional difference gel electrophoresis (2D-DIGE) (Unlu et al.,1997), highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dye)( Shaw et al., 2003) and our original large format electrophoresis device(Kondo et al., 2006). 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. 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).

In this study, primary tumor tissues from eight incisional biopsy samples of Ewing sarcoma were subjected to proteomics. The corresponding clinico-pathological information is available in GeMDBJ Proteomics (Figure 1) including prognosis after chemotherapy and surgery. The sample numbers corresponded

*Corresponding author: Tadashi Kondo, MD, PhD, Proteome Bioinformatics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chu-o-ku, Tokyo 104-0045, Japan, Tel: +81-3-3542-2511 ext.3004; Fax: +81-3-3547-5298; E-mail: takondo@ncc.go.jp

Received December 07, 2009; Accepted December 19, 2009; Published December 20, 2009


Copyright: © 2009 Kikuta K, 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.
to those in our previous report, and more detailed clinico-pathological data are available there (Kikuta et al., 2009). 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 (Kikuta et al., 2009). In brief, proteins were extracted from snap frozen tissues using a urea lysis buffer and a Multi-beads shoker (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 analytical purposes, the internal control sample was prepared by mixing a small portion of all eight 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 our original large format electrophoresis device (Kondo et al., 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, which was shown 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. All procedures for protein identification were reported in our previous report (Kikuta et al., 2009).

The proteome data of Ewing sarcoma, which included the data of 2D-DIGE and the annotation of protein spots can be obtained in GeMDBJ Proteomics (Figure 1). Clicking “Search by Protein” opens the page for text search. Selecting “Search by Gel Image” leads the page of a list of projects including Ewing sarcoma proteomics. The clickable 2D image of Ewing sarcoma sample is appeared by clicking “Ewing sarcoma”. This page includes the annotation of protein spots, and the following page includes the intensity of protein spots, which is visualized by color spectrum, a heat map, across the eight samples. The mass spectrometric data supporting the protein identification and the prognostic information are available from this page. The number of protein spots in GeMDBJ Proteomics corresponds to that in our previous publication for Ewing sarcoma proteome (Kikuta et al., 2009). A page of “Project Overview” exhibits a list of projects, a brief description of projects, the number of observed or identified protein spots, samples, and identified proteins. “Link to 2D Database” summarizes the status of presently available proteome database using 2D-PAGE data.

Among the 330 protein spots identified, we found that 168 protein spots contained multiple proteins, accounting for 50.9% 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) and lung cancer proteome (Kosaihira et al., 2009), being consistent with the results of previous proteome studies (Campostrini et al., 2005; Nawrocki et al., 1998; Westbrook et al., 2001). 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, can cause the overlapping of protein spots. Although the extensive fractionation and the use of large format gel apparatus may solve this problem to some extent, the mass spectrometry with improved sensitivity will eventually detect multiple proteins in single protein spots. As 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 (Hunsucker et al., 2006), this feature of 2D-DIGE may not be practically a serious problem in comparative expression studies. However, the confirmation of differential expression proteins by western blotting should be achieved when the further experiments are considered. In case western blotting resulted in the discordant results with those by 2D-DIGE, the proteins secondly or thirdly ranked in protein identification studies should be considered as those contributed to the intensity differences. GeMDBJ Proteomics demonstrates all candidate proteins for positive protein identifications, which were low-ranked because of low but significant Mascot score. Such protein identification data will be helpful to interpret the data by western blotting and 2D-DIGE.

Among 220 unique proteins identified as those ranked top by mass spectrometry, we found that 58 unique proteins were identified in multiple protein spots, accounting for 26.4% of all identified unique proteins. We had similar results in our previous study on the pancreatic cancer proteome (Yamada et al., 2008) and lung cancer proteome (Kosaihira et al., 2009), a finding that may be explained by the presence of alternative splicing or posttranslational modifications. Figure 2B demonstrates the number of unique proteins as a function of the number of protein spots where they observed. Proteins repeatedly observed in seven protein spots include actin, alpha-1-antitrypsin, haptoglobin, serum albumin, vimentin; those in four protein spots include complement factor B, fibrinogen beta chain, haptoglobin-related

Figure 2: The characteristics of 2D-DIGE data of Ewing sarcoma. 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.
protein, heat shock protein HSP90-beta, nucleolin (protein C23), serotransferrin; those in three protein spots included alpha-1-antichymotripsin, angiotensinogen, antithrombin-III, heat shock cognate 71 kDa protein, hepatoma-derived growth factor (HDGF), Ig kappa chain C region, L-lactate dehydrogenase B chain, nucleophosmin, protein disulfide-isomerase A6, 60S acidic ribosomal protein P0, transaldolase, transitional endoplasmic reticulum ATPase, tubulin beta chain, tropomyosin alpha-3 chain and vinculin. Plasma proteins in tissues may tend to generate multiple protein spots, probably because of glycosylation. There are many examples of the proteins, the functions of which are appeared after they are digested by proteases or phosphorylated by kinases. Structural characterization of protein spots generated by identical genes will generate more proteome information from 2D-DIGE data. As the amount of proteins recovered from gel is limited and not enough for the mass spectrometric analysis for posttranslational modifications, molecular probes specific to postranslational modifications may be one of the solutions for this issue.

Presently, the GeMDBJ Proteomics includes the 2D-DIGE proteome data derived from pancreatic cancer cell lines, esophageal cancer, Ewing’s sarcoma, lung adenocarcinoma, malignant pleural mesothelioma tissues, colorectal cancer, and cholangiocarcinoma, and the proteome data of the other malignancies are under preparation. The GeMDBJ Proteomics will 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 for cancer research will be our next challenge.

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


