Original Article

Proteome reference map and regulation network of neonatal rat cardiomyocyte

Zi-jian LI^{1, #}, Ning LIU^{2, #}, Qi-de HAN¹, You-yi ZHANG^{1, *}

¹Institute of Vascular Medicine, Peking University Third Hospital and Key Laboratory of Cardiovascular Molecular Biology and Regulatory Peptides, Ministry of Health and Key Laboratory of Molecular Cardiology, Ministry of Education, Beijing 100191, China; ²Central Laboratory, Jilin University Second Hospital, Changchun 130041, China

Aim: To study and establish a proteome reference map and regulation network of neonatal rat cardiomyocyte.

Methods: Cultured cardiomyocytes of neonatal rats were used. All proteins expressed in the cardiomyocytes were separated and identified by two-dimensional polyacrylamide gel electrophoresis (2-DE) and matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). Biological networks and pathways of the neonatal rat cardiomyocytes were analyzed using the Ingenuity Pathway Analysis (IPA) program (www.ingenuity.com). A 2-DE database was made accessible on-line by Make2ddb package on a web server.

Results: More than 1000 proteins were separated on 2D gels, and 148 proteins were identified. The identified proteins were used for the construction of an extensible markup language-based database. Biological networks and pathways were constructed to analyze the functions associate with cardiomyocyte proteins in the database. The 2-DE database of rat cardiomyocyte proteins can be accessed at <http://2d.bjmu.edu.cn>.

Conclusion: A proteome reference map and regulation network of the neonatal rat cardiomyocytes have been established, which may serve as an international platform for storage, analysis and visualization of cardiomyocyte proteomic data.

Keywords: proteome; neonatal rat cardiomyocyte; regulation network; 2-DE database

Acta Pharmacologica Sinica (2011) 32: 1116–1127; doi: 10.1038/aps.2011.86; published online 15 Aug 2011

Introduction

Proteomic technology allows for the global analysis of gene products in various cells, tissues and organisms. In parallel with other burgeoning technologies and coupled with the rapidly growing needs of life scientists, proteomics has become one of the most utilized approaches in all of life science research. As the Human Proteome Project is set in place^[1], there will be an urgent need to store vast amounts of information in proteomic databases; these databases will become a powerful tool to utilize in the research of human diseases. Heart disease is the leading cause of death in many countries including the United States, England and Canada. Proteomics approaches have clearly grown in potential to provide global and holistic analysis of changes in protein expression during heart disease. The earliest establishment of myocardial two-dimensional gel electrophoresis (2-DE) protein databases,

beginning in the 1990s, focused on identifying proteins from human myocardium. In the database from the Jungblut laboratory, N-terminal sequencing, internal sequencing and amino acid analysis were used to identify twelve proteins of the human myocardial 2-DE pattern^[2]. During the same year, a database containing approximately fifty proteins from 2-DE gels of human myocardial tissue was characterized by the Baker laboratory^[3]. Subsequent studies have succeeded in establishing myocardial 2-DE protein databases in different species, such as rat, mouse, and dog, and thereby enriched myocardial 2-DE protein databases^[4–7]. These databases have allowed investigators to compare data and to establish reference standards.

Proteomic technology has been used extensively to study cardiovascular diseases and to identify candidate molecules for diagnosis and therapy in intact live animals. Compared with intact hearts, cell culture models would be very useful to distinguish the differences between myocytes *versus* fibroblasts, to eliminate hemodynamic and hormonal influences and to study cell phenotypes and signaling in more homogeneous populations in a defined environment. However, pro-

These authors contributed equally to this work.

* To whom correspondence should be addressed.

E-mail zhangyy@bjmu.edu.cn

Received 2010-10-16 Accepted 2011-05-18

teomic investigations of cardiomyocyte are scarce, and a 2-DE protein database for the rat cardiomyocyte has not been developed before. In this work, we have launched a proteomic study of neonatal rat cardiomyocytes and compiled a profile of proteins expressed in these cells 2-DE and matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). Using immobilized pH gradient isoelectric focusing in a linear gradient from pH 3 to 10 in 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), more than 1000 proteins were separated and displayed from cultured cardiomyocytes. Among those, 148 proteins spots have so far been identified and used for the construction of an extensible markup language-based database. In addition to the cardiomyocyte proteins stored in this database, we also characterized interaction-based biological networks of proteins, which allow us to model how proteins work together to mediate biological processes. Moreover, our network analysis has also revealed function maps and pathway maps potentially involved in the neonatal rat cardiomyocyte proteins database.

In summary, with the rapid development of modern proteomics, it is essential that annotated databases are constructed to store all of the vast data generated by these techniques. More important is that these databases can be interrogated effectively both within the laboratory and by other scientists worldwide through the use of the World Wide Web. The database presented in this work will serve as an international platform for the storage, analysis, and visualization of cardiomyocyte proteomic data that can contribute to a more holistic view of heart tissue. This database may contribute to understanding key cardiovascular functions and pathways and may offer the potential for new avenues of future therapeutic intervention.

Materials and methods

Cardiomyocyte culture

Neonatal rat cardiomyocytes were isolated and cultured as previously described^[8]. Briefly, cardiomyocytes were dissociated from ventricles of 1- to 2-d old neonatal Sprague-Dawley rats using 0.1% trypsin (Hyclone) and 80 units/mL collagenase (Worthington Biochemical Corp) in a Hank's balanced salt solution (calcium-free; Hyclone). To purify the cardiomyocytes from non-myocytes, isolated cells were pre-plated for 90 min. The enriched cardiomyocyte fractions were seeded into 150-cm culture dishes and cultured in DMEM (Sigma, 4500 mg/L D-glucose, and L-glutamine, sodium pyruvate) supplemented with 10% heat-inactivated fetal bovine serum for 24 h. The DNA synthesis inhibitor, bromodeoxyuridine (100 μ mol/L), was added during the first 48 h to prevent the proliferation of non-cardiomyocytes. The experiments were approved by the Institutional Animal Care and Use Committee of Peking University Health Science Center (LA2010-035) and were adhered to the American Physiological Society's "Guiding Principles in the Care and Use of Vertebrate Animals in Research and Training." All protocols were conducted in accordance with the Guidelines for Animal Experiments,

Peking University Health Science Center.

Two-dimensional electrophoresis (2-DE)

Proteins were separated by two-dimensional electrophoresis. Samples containing 200 μ g protein for analytical gels or up to 1.0 mg for micropreparative gels were diluted to 300 μ L with rehydration solution (6 mol/L urea, 2 mol/L thiourea, 2% CHAPS, 65 mmol/L DTT, 0.5% *v/v* pH 3-10 Bio-lyte, trace bromophenol blue) and applied to IPG strips (Bio-Rad) for 12-14 h in a passive mode. The isoelectric focusing was performed at 20 °C over 24 h for a total of 70 000 V-h. After equilibration of the IPG strips, the second-dimensional SDS electrophoresis was performed on 12% gels at a current setting of 7.5 mA/gel for the initial 2 h and 15 mA/gel until the tracking dye reached the cathode.

Protein visualization and image analysis

After two-dimensional gel electrophoresis, proteins were stained with silver or with G-250 for subsequent mass spectrometry. In order to increase the sensitivity of Coomassie staining, we improved the prescription of Coomassie staining as follows: 20% methanol, 2% phosphoric acid, 10% ammonium sulfate and 0.1% Coomassie Brilliant Blue G-250. Gels were fixed in 12.5% TCA for 60 min and rinsed in Milli-Q water for 20 min. The gels were then stained in the improved Coomassie staining solution for 24 h. After the stained gels were scanned with a high-resolution scanner (Umax 1120), the gel images were analyzed using the PDQuest software (Version 7.1.1; Bio-Rad) according to the protocols provided by the manufacturer.

Identification of protein spots

Protein spots of interest visualized with colloidal Coomassie blue G-250 staining were excised and transferred to 1.5 mL siliconized Eppendorf tubes. The gel spots were then washed and destained by 50% ACN, then dried in a vacuum centrifuge. The dried gel-pieces were incubated in the digestion solution containing 50 mmol/L NH_4HCO_3 and 0.1 g/L TPCK-trypsin for 12 h at 37 °C. The resulting peptides were extracted three times by 50 μ L aliquots of 5% trifluoroacetic acid in 60% acetonitrile. Combined extracts were concentrated in a Speed Vac to 3-5 μ L.

The concentrated tryptic peptide mixture was mixed with a saturated CHCA matrix solution and vortexed gently. A volume (1 μ L) of the mixture containing CHCA matrix was loaded on a 96 $\times$ 2 well hydrophobic plastic surface sample plate (Applied Biosystems) and air-dried. The samples were analyzed with a Voyager DE STR MALDI-TOF MS (Applied Biosystems) fitted with a 337-nm nitrogen laser. Spectra were acquired using the instrument in reflectron mode and calibrated using a standard peptide mixture. Database searching with the monoisotopic peptide masses was performed against the NCBIInr database by using the peptide search engine ProFound (<http://prowl.rockefeller.edu/cgi-bin/ProFound>) by using the following conditions: partial methionine oxidation, complete cysteine carbamidomethylation, mass tolerance of

0.2 Da, and one missing cleavage.

Biological networks and pathway analysis

The biological network data were generated through the use of Ingenuity Pathways Analysis (IPA) software (www.ingenuity.com), a web-delivered application that evaluates biological networks. A data set of proteins from cardiomyocytes was uploaded into IPA. Each protein identifier was mapped to its corresponding protein object in the Ingenuity Pathways Knowledge Base (IPKB). These proteins were then used as the starting point for generating biological networks based upon the identities of the focus proteins and interactions with genes/proteins that were reported in the literature. IPA calculated a significance score for each network. The score was generated using a *P*-value calculation and was displayed as the negative log of that *P*-value, indicating the likelihood that the assembly of a set of focus proteins in a network could be explained by random chance alone. A score of 2 indicates that there is a 1 in 100 chance that the focus proteins are together in a network due to random chance. Therefore, networks with scores of 2 or higher have at least a 99% confidence of not being generated by random chance alone. Biological functions or canonical pathways were then calculated and assigned to each network by using findings that had been extracted from scientific literature and stored in the IPKB. The biological functions assigned to each network were ranked according to the significance of that biological function to the network.

Construction of an on-line database

One 2-DE gel of neonatal rat cardiomyocyte proteins was analyzed by PDQuest software (Version 7.1.1; Bio-Rad). The identified protein spots were stored in a flat file database that was made accessible on-line by the Make2ddb package on a web server with search functions, such as by accession number, description, or author or by clicking on a protein spot. The individual protein entries were hyperlinked to the relevant spots on an image map created from the reference gel.

Results

Data management, analysis, and presentation

A schematic representation of our proteomic informatics approach for the research of neonatal rat cardiomyocytes is illustrated in Figure 1. It mainly consists of three parts: data acquisition, data analysis, and web-accessible data presentation. The laboratory information management system (LIMS) is applied in this study for the acquisition of proteomic data. Results of gel image analysis performed by PDQuest can also be transmitted into the LIMS. A Java graphical user interface (Java-GUI) enables transfer of selected data from the LIMS into the database system consisting of 2D-PAGE, network/function analysis and protein information. The database system is interconnected with public and other knowledge databases. Furthermore, continuative explorative data analysis for significant networks and canonical pathways relevant to the dataset were calculated and analyzed using IPA.

Figure 1. A schematic representation of our proteome informatics approach for the database research of neonatal rat cardiomyocytes.

2-DE pattern of cardiomyocytes

Proteins from cardiomyocytes were resolved by 2-DE methods. IEF was first conducted on a 17 cm, broad range pH 3–10 linear immobilized pH gradient (IPG) strip. However, the results showed that a high resolution 2-DE gel was not obtained due to an insufficient spatial resolution, difficult-to-reveal low copy number proteins in the presence of more abundant proteins and the fact that most of the proteins were distributed in a pH range from 5 to 8. Therefore, in subsequent investigations, we adopted multiple overlapping narrow ranges IPG strips of pH 3–6, pH 5–8, and pH 7–10 in the first dimension (zoom-in gel). In this way, the resolution is greatly improved, and the detection of spots with low intensity can be easily performed (Figure 2A). In addition, Coomassie blue staining was improved as described in Materials and methods. The result showed that the new method possessed the advantages of both Coomassie blue and silver stains, and has sensitivity almost equivalent to silver stain (Figure 2B).

2D-PAGE database

Proteins from cardiomyocytes were separated and stained with the above-mentioned methods. The 2-DE gel was used to create a gel image in the database mentioned above as shown in Figure 3. Protein spots were identified using the PMF method with MALDI-TOF MS. Identified proteins were summarized in Table 1. Furthermore, some selected proteins from the database were confirmed for validity in cardiomyocytes by Western blot (Figure 4). This 2-DE database is fully independent of the platform and can be grafted from one platform to another. The identified protein spots

Figure 2. 2-DE proteome map of the cardiomyocytes. (A) Shown are the gel of pI 3–10 (upper) and narrow pI range gels, 3–6, 5–8, and 7–10 (lower). Staining was by colloidal Coomassie blue G-250. (B) Representative 2-DE profiles of neonatal rat cardiomyocytes. Proteins on 2-DE gels were visualized by colloidal Coomassie blue G-250 (left) or silver staining (right).

Figure 3. A proteome reference map of neonatal rat cardiomyocytes. Protein spot information (annotation) on prohibitin and generated mass spectrum data.

are hyperlinked to individual protein entries, and the all identification information of each protein entry can be obtained through clickable images. These hyperlinks include protein

names, database accession numbers, theoretical molecular mass and pI values. Furthermore, the detailed information of certain protein spots can be obtained through several search-

Figure 4. The confirmation of partial proteins from the database in cardiomyocytes by Western blot.

ing tools provided by the database. Besides these, detailed protocols in all the experiments were also provided to ensure the reproducibility of data in the database. Some useful links to other proteomic tools and database can be found throughout the database. For further details of the identified proteins and correspondence of identified proteins to spots in the 2-DE gel image, please refer to our database at <http://2d.bjmu.edu.cn>.

Data analysis and visualization

In order to further explore how the cardiomyocyte proteins in the database were related, we used the Ingenuity pathway analysis (IPA) program and the Ingenuity Pathways Knowledge Base (IPKB). The IPA program analyzes a large genomic or proteomic data set to find the most significant networks and canonical pathways relevant to the data set based on a calculated probability score. The specificity of connections for each protein was calculated, as defined by the percentage of its direct connections to other proteins showing significant changes. Just as described in Materials and methods, networks with scores of 2 or higher have at least a 99% confidence of not being generated by random chance alone. Therefore, based on the computed scores, four direct networks with scores of 24, 20, 16, and 12 were found to be significant in the database (Figure 5). In addition, high-level functions were calculated and assigned to each direct network if the significance of the association between the network and the biological function had a $P < 0.05$ (Table 2). The functions displayed in the table represent the top 4 high-level functions from the local analysis, which provides an overview of the biological functions associated with a given network (A, network 1; B, network 2; C, network 3; and D, network 4). In order to obtain a more comprehensive regulation-relationship map of the cardiomyocyte proteins in the database, a merge network, which included above-mentioned four networks, was created by IPA (Supple-

mentary Figure 1). Not only does the IPA software analysis find novel interconnectivity but the “biological functions and canonical pathways” feature also enables us to further understand biological functions and signaling cascades engaged by each network. The functional analysis for the merge direct network revealed a function map (Supplementary Figure 2A). The top three functions in the network were related to cellular assembly and organization, cardiovascular disease and cell death. Moreover, pathway analysis showed a canonical pathways map (Supplementary Figure 2B). The top three canonical pathways in the network were citrate cycle, glyoxylate and dicarboxylate metabolism and valine, leucine and isoleucine degradation. Beside these, four indirect networks and a merge indirect network with their functional analysis and pathway analysis were also found in the database (Supplementary Figure 3 and 4, and Supplementary Table 1). More detailed information can be found at <http://2d.bjmu.edu.cn>.

Discussion

Proteomic technology can provide new insights into cellular mechanisms involved in cardiac dysfunction^[9–13]. The majority of proteomic investigations still use 2D gel electrophoresis (2-DE) with immobilized pH gradients to separate the proteins in a sample and combine this with mass spectrometry (MS) technologies to identify proteins. In spite of the development of novel gel-free technologies, 2-DE remains the only technique that can be routinely applied to parallel quantitative expression profiling of large sets of complex protein mixtures such as whole cell lysates. Current limitations to this technology include the resolution of the 2-DE and the detection of low-abundance proteins. In the present study, two approaches were explored to solve these drawbacks. The first was to apply multiple overlapping narrow range IPG strips (to increase the resolution of the 2-DE), and the second was to improve protein visualization (to achieve the highly sensitive detection of 2-DE separated proteins). With complex samples such as cardiomyocytes, 2-DE on a single wide-range pH gradient reveals only a small percentage of the whole proteome because of insufficient spatial resolution and difficulty to reveal low-copy proteins in the presence of the most abundant proteins. Our results showed that the resolution of 2-DE with multiple overlapping narrow range IPG strips (pH 3–6, pH 5–8, and pH 7–10) is about 2.5 times that of wide range (pH 3–10). Using these gel systems, more than 1000 proteins were resolved from our model. It was well-accepted that silver staining and Coomassie blue staining are currently the most popular methods of protein detection. Although silver staining, with a sensitivity approximately 10–50 times that of Coomassie blue, is the most sensitive staining method so far reported, there are a number of drawbacks and limitations associated with this method. A high background resulting from a number of variables may cause poor resolution of protein spots. Quantifying protein abundance is also difficult with silver-stained gels due to the poor linearity to protein concentration. In addition, silver stain does greatly interfere with subsequent protein identification by mass

Figure 5. The direct interaction functional networks map of cardiomyocyte proteins in the database. The direct interaction network scores were 24 (A, network 1), 20 (B, network 2), 16 (C, network 3), and 12 (D, network 4), accordingly. Nodes represent proteins, with their shape indicating the functional class of the protein and multiple edges indicating the biological relationships between the nodes. For the functional analyses of networks, see Table 2.

spectrometry. Unlike silver staining, Coomassie blue staining has a good compatibility with subsequent analytical techniques and yields higher dynamic range in determining the changes of protein expression levels. But the major issue of Coomassie staining is that its sensitivity is well below that of silver. In this study, we improved colloidal Coomassie brilliant blue G-250 (Coomassie blue dye) and increased its

sensitivity to be almost equivalent to that of silver stain.

As mentioned earlier, the protein database will be important to serve as a basis for further studies of human disease. There are four 2-DE protein databases of cardiac proteins, established by three independent groups, which can be accessed via the World Wide Web^[7, 14-16]. These databases facilitate proteomic research into heart diseases containing information

Table 1. Identified proteins of neonatal rat cardiomyocytes.

No	Accession No	Protein information	Z score estimated by profound	Sequence coverage %	pl/ Theoretical MW (kDa)
1	gi 2894106	Collagen alpha1	2.35	25	5.8/138.9
2	gi 27688933	Similar to Collagen alpha1	2.43	16	5.8/138.9
3	gi 27688933	Similar to Collagen alpha1	2.22	16	5.8/138.9
4	gi 58865966	Tumor rejection antigen gp96	2.42	43	5.02/74.12
5	gi 58865966	Tumor rejection antigen gp96	2.42	43	5.02/74.12
6	gi 58865966	Tumor rejection antigen gp96	2.42	43	5.02/74.12
7	gi 38303969	Heat shock 70 kDa protein 5	2.4	26	5.1/72.50
8	gi 56385	Hsc70-ps1	2.24	24	5.4/71.14
9	gi 2119726	DnaK-type molecular chaperone grp75 precursor-rat	2.35	31	5.9/74.01
10	gi 19705459	poly(A) binding protein, cytoplasmic 1	1.75	15	9.6/70.91
11	gi 227429	Non muscle caldesmon	2.43	18	6.4/61.76
12	gi 2498205	Non-muscle caldesmon (CDM) (L-caldesmon)	2.43	20	6.3/60.68
13	gi 227429	Non muscle caldesmon	2.43	24	6.4/61.76
14	gi 10637996	Mitochondrial aconitase	1.83	12	7.87/85.42
15	gi 40538860	Aconitase 2, mitochondrial	2.23	22	7.87/85.38
16	gi 10637996	Mitochondrial aconitase	2.02	16	7.87/85.42
17	gi 10637996	Mitochondrial aconitase	1.93	20	7.87/85.42
18	gi 92654	Propionyl-CoA carboxylase (EC 6.4.1.3) alpha chain	2.43	12	6.3/78.22
19	gi 129684	Propionyl-CoA carboxylase alpha chain, mitochondrial	2.08	17	6.3/78.32
20	gi 47938978	Dars protein	1.72	10	6.0/57.56
21	gi 1346413	LAMA_RAT Lamin A	2.37	23	6.5/74.59
22	gi 1346413	LAMA_RAT Lamin A	2.43	26	6.5/74.59
23	gi 1334284	Unnamed protein product	2.35	21	5.3/58.08
24	gi 90207	Mitochondrial protein P1 precursor	2.25	32	5.3/58.08
25	gi 51036657	Prolyl 4-hydroxylase alpha subunit	1.79	18	5.6/61.22
26	gi 56090441	Ddx5	2.03	12	9.06/69.2
27	gi 34874349	Similar to Dihydropyrimidinase related protein-2 (DRP-2)	2.28	25	6.3/82.08
28	gi 453180	Lamin A	2.19	21	6.2/71.91
29	gi 1794160	Lamin C	2.32	37	6.4/65.49
30	gi 51702230	60 kDa heat shock protein, mitochondrial precursor (Hsp60)	1.94	20	5.35/57.89
31	gi 51702230	60 kDa heat shock protein, mitochondrial precursor (Hsp60)	2.37	23	5.35/57.89
32	gi 34875806	Hypothetical protein XP_212759	2.29	24	5.9/61.05
33	gi 34875806	Hypothetical protein XP_212759	2.40	37	5.9/61.05
34	gi 38197676	Desmin	2.26	41	5.2/53.46
35	gi 38382858	Glucose regulated protein, 58 kDa	2.34	21	5.9/57.06
36	gi 38382858	Glucose regulated protein, 58 kDa	2.43	43	5.9/57.06
37	gi 38382858	Glucose regulated protein, 58 kDa	2.25	26	5.9/57.06
38	gi 38382858	Glucose regulated protein, 58 kDa	2.36	20	5.9/57.06
39	gi 38382858	Glucose regulated protein, 58 kDa	2.43	40	5.9/57.06
40	gi 38303871	Dihydrolipoamide dehydrogenase	1.93	14	7.96/54
41	gi 34862189	ATP synthase, H ⁺ transporting, mitochondrial F1 complex	2.38	26	5.1/56.33
42	gi 34862189	ATP synthase, H ⁺ transporting, mitochondrial F1 complex	2.37	36	5.1/56.33
43	gi 34862189	ATP synthase, H ⁺ transporting, mitochondrial F1 complex	2.37	36	5.1/56.33
44	gi 25990263	Mitochondrial aldehyde dehydrogenase	2.43	20	5.7/53.29
45	gi 25742617	Epidermal growth factor receptor	1.84	28	6.66/134.8
46	gi 37589940	Heterogeneous nuclear ribonucleoprotein H1	2.43	25	5.9/49.47
47	gi 50926833	Eno1 protein	2.43	28	6.2/47.45
48	gi 16073616	Aldehyde dehydrogenase	2.43	31	6.1/48.65
49	gi 62644240	Dihydrolipoamide branched chain transacylase E2	2.12	24	7.64/33.89
50	gi 6729934	Chain A, Rat Liver F1-Atpase	2.06	20.0	8.28/55.24
51	gi 6729934	Chain A, Rat Liver F1-Atpase	1.83	20.0	8.28/55.24
52	gi 40538742	ATP synthase, H ⁺ transporting, mitochondrial F1 complex	1.98	24	9.22/58.79
53	gi 34876531	Similar to hypothetical protein FLJ10849	2.09	21	6.5/59.16
54	gi 114523	ATP synthase alpha chain, mitochondrial precursor	2.21	24	9.22/58.79

(to be continued)

No	Accession No	Protein information	Z score estimated by profound	Sequence coverage %	pI/ Theoretical MW (kDa)
55	gi 56200	Unnamed protein product	2.24	23	8.3/61.75
56	gi 6978441	Actin beta	2.32	26	5.3/42.26
57	gi 71620	Actin beta	1.82	26	5.2/42.08
58	gi 81890783	Suc1g2_predicted protein (fragment)	1.76	13	7.57/46.75
59	gi 6981112	Isovaleryl coenzyme a dehydrogenase	2.01	15	8.03/46.86
60	gi 11968102	Ornithine aminotransferase	1.82	12	6.53/48.30
61	gi 38181818	Ornithine aminotransferase	2.31	22	6.5/48.71
62	gi 34859187	Similar to RIKEN cDNA 2300002G02	2.39	22	7.3/49.9
63	gi 34859187	Similar to RIKEN cDNA 2300002G02	2.24	34	7.3/49.9
64	gi 34859187	Similar to RIKEN cDNA 2300002G02	1.83	23	7.3/49.9
65	gi 38303993	Cathepsin D [rattus norvegicus]	2.43	18	6.7/45.12
66	gi 30171809	NADH dehydrogenase 1 alpha subcomplex 10-like protein	1.82	15	7.2/40.81
67	gi 20150495	Chain B, crystal structure of rat short chain acyl-coa dehydrogenase complexed with Acetoacetyl-Coa	1.87	22	6.4/42.44
68	gi 20150495	Chain B, crystal structure of rat short chain acyl-coa dehydrogenase complexed with Acetoacetyl-Coa	1.90	36	6.4/42.44
69	gi 16758570	Cell division cycle 37 homolog; CDC37 (cell division cycle 37, <i>S cerevisiae</i> , homolog)	2.18	22	5.4/44.87
70	gi 12964644	NAD ⁺ -specific isocitrate dehydrogenase a-subunit	1.90	31	6.5/40.05
71	gi 38304046	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	1.85	26	5.4/31.85
72	gi 38541406	Acidic ribosomal protein P0	2.35	34	5.9/34.37
73	gi 511131	Alpha-actin cardiac	1.9	26	5.2/42.34
74	gi 145553966	Voltage-dependent calcium channel alpha 1E	1.82	21	8.25/252.1
75	gi 235879	Lipocortin I	2.43	13	7.0/39.14
76	gi 15100179	Malate dehydrogenase 1; Malate dehydrogenase-like enzyme	2.43	15	6.2/36.64
77	gi 38197394	Annexin 1	2.43	13	7.0/39.15
78	gi 37590235	Malate dehydrogenase 1	2.29	15	5.9/36.64
79	gi 37590785	Calpactin I heavy chain	1.7	27	7.7/38.95
80	gi 984553	G protein beta 1 subunit	2.16	14	5.5/38.17
81	gi 3122930	3-mercaptopyruvate sulfurtransferase (MST)	1.91	23	5.9/33.21
82	gi 4699609	Chain C, dienoyl-coa Isomerase	1.81	18	7.1/30.70
83	gi 49903126	Calcium modulating ligand	1.8	17	7.8/33.06
84	gi 396418	Actin-capping protein beta chain, splice form 2	2.4	20	5.69/30.95
85	gi 16758848	Endoplasmic reticulum protein 29	2.25	31	6.2/28.61
86	gi 2119779	Multidrug resistance protein 3	2.23	17	6.5/27.74
87	gi 13937353	Prohibitin	1.7	34	5.6/29.86
88	gi 13937353	Prohibitin	2.37	41	5.6/29.86
89	gi 37590233	Unknown (protein for MGC:72744)	1.75	25	6.2/31.22
90	gi 37590233	Unknown (protein for MGC:72744)	1.75	25	6.2/31.22
91	gi 20149810	Chain F, structure of enoyl-coa hydratase complexed with the substrate dac-coa	2.16	26	6.4/28.31
92	gi 1170367	Heat-shock protein beta-1 (HspB1) (HSP 27)	2.23	32	6.1/22.93
93	gi 1170367	Heat-shock protein beta-1 (HspB1) (HSP 27)	2.43	30	6.1/22.93
94	gi 34877098	Similar to spindlin	1.73	15	6.4/29.72
95	gi 34855918	Similar to thioether S-methyltransferase	2.22	11	5.9/34.79
96	gi 1709429	Ornithine decarboxylase antizyme (ODC-Az)	2.19	33	6.7/25.67
97	gi 6981240	Myosin light polypeptide 3 (myosin light chain 1)	2.00	60	5.03/22.26
98	gi 6981240	Myosin light polypeptide 3 (myosin light chain 1)	2.12	38	5.03/22.01
99	gi 206681	Rieske Fe-S protein precursor	1.8	19	8.9/27.67
100	gi 136708	Ubiquinol-cytochrome c reductase iron-sulfur subunit, mitochondrial precursor (Rieske iron-sulfur protein) (RISP)	2.02	20	9.2/27.96
101	gi 38511598	Adenylate kinase 2	2.43	18	7.0/26.74
102	gi 30387799	Alpha B-crystallin	2.43	43	6.8/20.07
103	gi 30387799	Alpha B-crystallin	2.43	45	6.8/20.07
104	gb AAB37470.1	Myosin regulatory light chain 2	2.22	38	4.86/18.74
105	gi 818029	Dismutase	2.43	32	5.9/15.73
106	gi 30387800	Alpha B-crystallin	2.43	34	6.8/21.15

(to be continued)

N _o	Accession N _o	Protein information	Z score estimated by profound	Sequence coverage %	pl/ Theoretical MW (kDa)
107	gi 34876505	Similar to RIKEN cDNA 4921537P18	2.24	11	5.9/20.97
108	gi 3790238	MHC class II RT1D beta1 chain antigen	2.00	31	6.9/14.83
109	gi 34883131	Similar to RIKEN cDNA 1700081022	2.43	22	7.8/19.17
110	gi 57580	Alpha B-crystallin	1.81	22	6.9/19.94
111	gi 8393823	Neurofilament 3, medium	1.92	23	9.57/110
112	gi 818021	Cytochrome c oxidase subunit Via (85 AA)	1.82	35	6.3/9.25
113	gi 12055450	Immunoglobulin heavy chain variable region	1.92	37	8.1/9.25
114	gi 511131	Alpha-actin cardiac	2.10	28	5.2/42.34
115	gi 1717770	Troponin T, cardiac muscle isoforms (TnTC) (cTnT)	2.33	24	5/38.81
116	gi 34864883	Similar to chaperonin containing TCP-1 beta subunit	2.21	39	6/57.78
117	gi 33943091	Serine/threonine kinase 25	2.05	23	6.3/48.37
118	gi 51948476	Ubiquinol-cytochrome c reductase core protein I	2.42	31	5.6/53.52
119	gi 51948476	Ubiquinol-cytochrome c reductase core protein I	2.25	31	5.6/53.52
120	gi 16758274	Peroxiredoxin 4	2.43	25	6.2/31.22
121	gi 34870174	Similar to RIKEN cDNA 0610040B21	1.83	33	6.1/27.37
122	gi 55992	Cytochrome c oxidase subunit Via (AA 1-118)	2.12	41	6.5/12.96
123	gi 38079128	RIKEN cDNA 4930473A06	2.43	25	5.2/42.69
124	gi 27229195	RIKEN cDNA 2310032D16; hypothetical glycerophosphoryl diester phosphodiesterase	2.23	22	5.2/51.8
125	gi 2253159	Peripherin	2.43	18	5.4/52.78
126	gi 6981240	Myosin light polypeptide 3 (Myosin light chain 1)	2.42	60	5.03/22.26
127	gi 34875200	Similar to expressed sequence ALO22641	2.06	21	5.6/47.02
128	gi 191618	Alpha cardiac myosin heavy chain	2.42	36	5.57/22.4
129	gi 21707954	Cytochrome c oxidase, subunit Va	1.87	27	6.1/16.31
130	gi 511131	Alpha-actin cardiac	2.10	28	5.2/42.34
131	gi 5869934	Hypothetical protein	2.08	25	6.3/145.71
132	gi 33859486	4-hydroxyphenylpyruvic acid dioxygenase	1.90	21	6.6/45.26
133	gi 1083311	Protein disulfide-isomerase (EC 5.3.4.1) ERp61 precursor	2.00	22	6.1/56.95
134	gi 13506795	LASP-1	2.16	23	6.6/30.35
135	gi 31542341	Caspase 1	2.00	20	4.3/45.8
136	gi 34784670	Cd1d2 protein	2.11	21	6.4/34.07
137	gi 33563276	SRY-box containing gene 8	2.43	19	6.6/50.03
138	gi 29468270	Hemopoietic lineage switch protein 5	2.38	18	6.7/58.64
139	gi 33859486	4-hydroxyphenylpyruvic acid dioxygenase	2.43	28	6.6/45.26
140	gi 109803	Phosphopyruvate hydratase (EC 4.2.1.11) alpha	2.43	24	6.4/47.45
141	gi 1717770	Troponin T, cardiac muscle isoforms (TnTC)	2.33	24	5/38.81
142	gi 1083311	Protein disulfide-isomerase (EC 5.3.4.1) ERp61 precursor	2.30	27	6.1/56.95
143	gi 12832696	Unnamed protein product	2.15	43	7.8/13.68
144	gi 6753530	Crystallin, alpha B; crystallin, alpha 2; alpha B-crystallin	2.42	43	6.8/20.05
145	gi 6753530	Crystallin, alpha B; crystallin, alpha 2; alpha B-crystallin	2.09	28	6.8/20.05
146	gi 27370092	RIKEN cDNA 2300002G02	2.43	40	7.3/49.89
147	gi 38075421	Similar to taxilin	2.43	27	6.1/57.18
148	gi 1346413	Lamin A	2.38	25	6.5/74.59

on several hundred cardiac proteins that have been identified by protein chemical methods. In addition, 2-DE protein databases and proteomic maps for other mammals are also under construction to support work on animal models of heart disease^[17-20].

As we know, cultured cardiac myocytes are attractive models for detailed proteomic investigations of cardiovascular diseases. Unlike the intact heart, cell culture models are very useful to distinguish effects on myocytes *versus* fibroblasts, to eliminate hemodynamic and hormonal influences and especially to analyze and integrate results of high throughput-

omic approaches. However, there is not, so far as we know, a proteome database for cardiomyocytes. In this study, a two-dimensional electrophoresis database for rat cardiomyocyte proteins was constructed using our improved methods. Our data show that the largest class of proteins in cardiomyocytes is energy and mass metabolism proteins (20%). For instance, the glucose-regulated proteins play important roles in glucose metabolism. Pyruvate decarboxylase, one of the key enzymes in gluconeogenesis, is closely involved in the process of carbohydrate metabolism. In addition, adenylate kinase also plays a crucial role in maintaining the energy balance in cells. The

Table 2. The top functions table for direct interaction networks.

id	Genes	Score	Focus genes	Top functions
1	ACADL, ACADS, ALDH1A1, AQP7, ARNT2, CASP1, CEBPB, COX5A, COX6A1, CRK, DUT (includes EG:1854), E2F1, EGFR, ESR1, FHIT, GRAP, HSP90B1, IDH3A, INMT, LASP1, LRRFIP1, MAP4K1, MYH6, MYL3, OAZ1, PPARA, PPARGC1A, RUNX2, SERPINB9, SNIP, SOX8, SRF (includes EG:6722), STK25, SULT2A1 (includes EG:6822), TNNT2	24	15	Connective tissue development and function, tissue morphology, cancer
2	AK2, CAPZB, CRYAB, CSDE1, CYCS, DARS, DDX5, DDX17, EPRS, HNRPH1, HSP90AA1, HSP90AB1, HSPB1, HSPD1, IARS (includes EG:3376), JTV1, MAPKAPK2, MDH1, MYC, NR3C1, PABPC1, PCBP2, PCBP1 (includes EG:5093), PCK1, PHB, PPID, PRPH, PTGES3, SERPINB5, ST13, STIP1, TADA2L, TAT, TGFB111, UNC45A	20	13	Drug metabolism, endocrine system development and function, lipid metabolism
3	AEBP1, ASS, BHLHB3, CAD, CDC37, CHGA, CREB1, CTSD, DES (includes EG:13346), DLAT, DLD, ERP29, GNAI2, GNB1, HSPA5, LOR, MAPK1, MARCKS, MYLPF, MYOD1, NDUFA10 (includes EG:4705), NEF3, NEFH, PCYT1A, PDHA2, PDHB, PDHX, PLA2G4A, RECK, SLC2A3, SMPD1, SOD1, SP1, TG, TRH	16	11	Genetic disorder, neurological disease, cell morphology
4	ACTA1, ACTA2, ACTB, ACTC, ACTG1, ACTG2, AKAP1, ANXA1, ANXA2, APC, ATP5A1, ATP5B, CDC25A, CDC25C, CLIC4, DYSF, ECHS1, EEF1A1, EIF2AK2, GAPDH, HNRPA1, HNRPK, LIMK1, LMNA, MARK2, NCL, NFYB, OAT, PABPC4, RPL19, RPLP2, RPLP0 (includes EG:6175), STAU1, TUBB2A, YWHAZ	12	9	Cardiovascular system development and function, organ morphology, protein synthesis

other two main groups of proteins are cytoskeletal proteins and oxidative stress proteins/heat shock proteins. The former, including actin, myosin, lamin, *etc*, play a central role in the creation and maintenance of cell shapes in cardiomyocytes. Meanwhile, cytoskeletal proteins also play important roles in both intracellular transport (the movement of vesicles and organelles) and cellular division. In addition, well-known heat-shock proteins and oxidative-stress proteins play pivotal roles during cardiac ischemia, ischemia preconditioning and cardiac hypertrophy. The functions of all those proteins are consistent with a heart that is a myogenic muscular organ.

Based on those identified proteins, this paper describes the creation of an on-line accessible 2-DE reference database for the cardiomyocyte proteome and provides protein identification data for a restricted number of reference spots. Such a 2-DE protein database reduces the need for routine protein identification, which is often difficult to achieve from an organism of which the genome is not sequenced. This Java-based database is fully independent on the platform and can be grafted from one platform to another. The customers can view the information of any protein spots in 2D images simply by clicking them. That is, 2-DE images can be provided on a World Wide Web server and, as a response to a mouse click on any identified spot on the image, the user can obtain the database entry for the corresponding protein. This method allows a user to easily identify proteins on a 2-DE image. At the same time, the detailed information of certain protein spots can also be obtained through several searching tools provided by the database, such as searching by accession number, searching by theoretical molecular weight (kDa) and searching by theo-

retical pI. In addition, the detailed protocols in all the experiments were also provided to ensure the reproducibility of data in the database and enable other laboratory to repeat the results of those experiments. The database is linked to other databases through active hypertext cross-references; that is to say, the user automatically gets connected to the corresponding web site through a simple mouse click on a cross-reference.

Although it is important to identify the individual proteins expressed in cultured cardiac myocytes, there is an increasing need to move beyond this level of analysis. Cluster and principal component analyses describe overall changes in apparent protein expression but provide few insights into the biological processes and regulatory networks involving in cultured cardiac myocytes. In this study, we combine large-scale analysis of protein expression with knowledge-based functional network analyses. A complex network involved in this database was mapped, and the biological function and biologically meaningful pathways that yield novel, predictive insights were computed and analyzed by IPA. IPA is a software/database search tool for finding function and pathway for specific biological states. It is a web-delivered application that makes use of the Ingenuity Pathways Knowledge Base, the curated database consisting of millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, drugs, and diseases. Sets of interaction data can be viewed as graphs or maps in which each protein is a node and each interaction is a line connecting two nodes. The importance of this view has led to use of the term "interaction map" to refer generically to interaction datasets. The map view provides not only an intuitive interface for biologists to explore the data

but also a formal mathematical framework for computational biologists to explore the properties of interaction networks. These networks include both direct and indirect interactions. Direct interactions are characterized by a well-defined information flow (eg, from a transcription factor to the gene it regulates). Indirect interactions do not have an assigned direction (eg, mutual binding relationships). Furthermore, as we know, cardiovascular diseases are complicated diseases that are controlled by complex regulation network. In these regulation networks, most proteins function in collaboration with other proteins. In this study, we established, for the first time, a proteome reference map and regulation network of the neonatal rat cardiomyocyte. This is especially useful in understanding dysfunction of complicated diseases. Furthermore, integrated analysis such as that provided by IPA enables analysis of all interactions and may offer additional insights into how such an extensive map contributes to cardiovascular diseases.

In our cardiomyocyte protein database, protein regulatory networks and pathway analysis are created to show how protein-protein interaction may eventually participate in the physiological process of heart. The functional analysis and pathways analysis show that the majority of proteins from cardiomyocytes are involved in cellular assembly and organization, cardiovascular system development, cell morphology, cardiovascular disease and plenty of metabolism pathway and oxidative-stress pathway (such as citrate cycle, glyoxylate and dicarboxylate metabolism, valine and leucine degradation, fat acid metabolism, oxidative phosphorylation and endoplasmic reticulum stress pathway). These results from network analysis also are consistent with the above-mentioned protein components of cardiomyocytes. More importantly, if a user finds certain interesting proteins in this database, the detail regulatory relationship about these proteins can be further found in the protein regulatory network, which will serve as a useful predictive tool for further research.

In conclusion, we have presented an improved methodology for cultured cardiomyocyte proteomics. The on-line database with reference gel and identified proteins serves as a stable framework that facilitates the study of the heart proteome and the exchange of heart proteome data. In addition, biological regulatory networks were for the first time constructed into 2-DE protein databases. Our network analysis has also revealed function and pathway maps potentially involved in the neonatal rat cardiomyocyte protein database. Together, this allowed us to better understand the mechanism, biological processes and specific regulatory networks of the heart. Our study also provides the basis for the development of new therapeutic strategies. Finally, the database we established can serve as an international platform for storage, exchange and analysis of proteomic data. The informational management for experimental data and the standardization, sharing and integration of this database will certainly provide a basis for further cardiovascular study.

Acknowledgements

This work was supported by grants from Major State Basic

Research Development Program of China (973 Program, 2011CB503903), the National Natural Science Foundation of China (81070078 and 81030001) and Beijing Municipal Natural Science Foundation (7102158).

Author contribution

Zi-jian LI designed and performed the research, analyzed the data, and wrote the manuscript; Ning LIU performed the research and analyzed the data; Qi-de HAN designed the research; and You-yi ZHANG designed the research and analyzed the data.

Supplementary information

Supplementary figures are available at Acta Pharmacologica Sinica website of NPG.

References

- 1 Peters KF, Menaker TJ, Wilson PL, Hadley DW. The Human Genome Project: an update. *Cancer Nurs* 2001; 24: 287–92.
- 2 Jungblut P, Otto A, Regitz V, Fleck E, Wittmann-Liebold B. Identification of human myocard proteins separated by two-dimensional electrophoresis. *Electrophoresis* 1992; 13: 739–41.
- 3 Baker CS, Corbett JM, May AJ, Yacoub MH, Dunn MJ. A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting. *Electrophoresis* 1992; 13: 723–6.
- 4 Dunn MJ, Corbett JM, Wheeler CH. HSC-2DPAGE and the two-dimensional gel electrophoresis database of dog heart proteins. *Electrophoresis* 1997; 18: 2795–802.
- 5 Corbett JM, Wheeler CH, Dunn MJ. Coelectrophoresis of cardiac tissue from human, dog, rat and mouse. Towards the establishment of an integrated two-dimensional protein database. *Electrophoresis* 1995; 16: 1524–9.
- 6 Pleissner KP, Sander S, Oswald H, Regitz-Zagrosek V, Fleck E. Towards design and comparison of World Wide Web-accessible myocardial two-dimensional gel electrophoresis protein databases. *Electrophoresis* 1997; 18: 480–3.
- 7 Li XP, Pleissner KP, Scheler C, Regitz-Zagrosek V, Salnikow J, Jungblut PR. A two dimensional electrophoresis database of rat heart proteins. *Electrophoresis* 1999; 20: 891–7.
- 8 Ito H, Hirata Y, Hiroe M, Tsujino M, Adachi S, Takamoto T. Endothelin-1 induces hypertrophy with enhanced expression of muscle-specific genes in cultured neonatal rat cardiomyocytes. *Circ Res* 1991; 69: 209–15.
- 9 Kullo IJ, Cooper LT. Early identification of cardiovascular risk using genomics and proteomics. *Nat Rev Cardiol* 2010; 7: 309–17.
- 10 Wang J, Bai L, Li J, Sun C, Zhao J, Cui C, et al. Proteomic analysis of mitochondria reveals a metabolic switch from fatty acid oxidation to glycolysis in the failing heart. *Sci China C Life Sci* 2009; 52: 1003–10.
- 11 Zamilpa R, Lopez EF, Chiao YA, Dai Q, Escobar GP, Hakala K, et al. Proteomic analysis identifies *in vivo* candidate matrix metalloproteinase-9 substrates in the left ventricle post-myocardial infarction. *Proteomics* 2010; 10: 2214–23.
- 12 Hare JM, Beigi F, Tziomalos K. Nitric oxide and cardiobiology-methods for intact hearts and isolated myocytes. *Methods Enzymol* 2008; 441: 369–92.
- 13 Didangelos A, Simper D, Monaco C, Mayr M. Proteomics of acute coronary syndromes. *Curr Atheroscler Rep* 2009; 11: 188–95.
- 14 Evans G, Wheeler CH, Corbett JM, Dunn MJ. Construction of HSC-

- 2DPAGE: a two-dimensional gel electrophoresis database of heart proteins. *Electrophoresis* 1997; 18: 471–9.
- 15 Muller EC, Thiede B, Zimny-Arndt U, Scheler C, Prehm J, Muller-Werdan U, et al. High-performance human myocardial two-dimensional electrophoresis database: edition 1996. *Electrophoresis* 1996; 17: 1700–12.
- 16 Pleissner KP, Sander S, Oswald H, Regitz-Zagrosek V, Fleck E. The construction of the World Wide Web-accessible myocardial twodimensional gel electrophoresis protein database “HEART-2DPAGE”: a practical approach. *Electrophoresis* 1996; 17: 1386–92.
- 17 McGregor E, Dunn MJ. Proteomics of the heart: unraveling disease. *Circ Res* 2006; 98: 309–21.
- 18 Raddatz K, Albrecht D, Hochgräfe F, Hecker M, Gotthardt M. A proteome map of murine heart and skeletal muscle. *Proteomics* 2008; 8: 1885–97.
- 19 Robichaud S, Lalu M, Udenberg T, Schulz R, Sawicki G. Proteomics analysis of changes in myocardial proteins during endotoxemia. *J Proteomics* 2009; 72: 648–55.
- 20 Lam L, Arthur J, Semsarian C. Proteome map of the normal murine ventricular myocardium. *Proteomics* 2007; 7: 3629–33.