

## Nuclear proteome profile of C57BL/6J mouse liver

ZHANG Yang<sup>1</sup>, FANG CaiYun<sup>2</sup>, BAO HuiMin<sup>2</sup>, FAN HuiZhi<sup>2</sup>, SHEN HuaLi<sup>1,2\*</sup>  
& YANG PengYuan<sup>1,2</sup>

<sup>1</sup>*School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China;*

<sup>2</sup>*Department of Chemistry, Fudan University, Shanghai 200433, China*

Received October 29, 2012; accepted April 19, 2013

The liver proteome can serve as a reference to better understand both disease mechanisms and possible therapeutics, since the liver is an important organ in the body that performs a large number of tasks. Here we identify the organelle proteome of C57BL/6J mouse liver nuclei as a promising strategy to enrich low abundance proteins, in the sense that analysis of whole liver cells is rather complex for current techniques and may not be suitable for proteins with low abundance. Evaluation of nucleus integrity and purity was performed to demonstrate the effectiveness of the optimized isolation procedure. The extracted nuclear proteins were identified by 2-DE MS analyses, and a total of 748 proteins were identified. Bioinformatic analyses were performed to demonstrate the physicochemical properties, cellular locations and functions of the proteins.

### C57 mouse, liver nuclei, proteome, bioinformatics

**Citation:** Zhang Y, Fang C Y, Bao H M, et al. Nuclear proteome profile of C57BL/6J mouse liver. *Sci China Life Sci*, 2013, 56: 513–523, doi: 10.1007/s11427-013-4488-y

Liver is the largest organ in the body and performs an astonishingly large number of tasks, influencing all body systems by its three main functions: vascular functions, metabolic achievements, and secretory and excretory functions. One consequence of this complexity is that hepatic disease has widespread effects on virtually all other organ systems. Expressional proteomics takes advantage of the physical and chemical properties of proteins to separate, quantify and catalog complex proteinaceous mixtures. These experimental methods offer a reproducible, qualitative and quantitative assessment of protein dynamics in a given sample. A proteomic approach can likely be enhanced by protein fractionation methods that focus on selected populations and less abundant species. This strategy has proven to be promising for blood plasma proteomics [1–3].

Subcellular fractionation prior to protein extraction emphasizes protein localization and compartment dynamics in

addition to improving the signal/noise. Foster and Mann mapped 1404 proteins to 10 subcellular locations in mouse liver. These localizations allowed assessment of specificity for published organellar proteomic inventories and demonstrated multiple locations for 39% of all organellar proteins [4]. Another study by Kislinger and Emilifour profiled proteins from major organellar compartments (cytosol, membranes, microsomes, mitochondria and nuclei) in six organs (brain, heart, kidney, liver, lung and placenta) of the laboratory mouse *Mus musculus*, and determined the subcellular localization for 3274 of the 4768 proteins identified [5]. The nucleus houses the majority of cellular genetic material DNA, although there is some contained in mitochondria, chloroplasts, and contains the components and enzymes necessary to maintain, pack and unpack, “transcribe” and process genetic material in a selective manner (i.e., produce mRNA) and “replicate” the DNA less selectively. The nucleus is also responsible for synthesis and/or assembly of components used to translate genetic material (ribosomes

\*Corresponding author (email: shenhuali@fudan.edu.cn)

and tRNA), although translation of the code contained in mRNA is performed in the cytoplasm.

In this work we focus on the nuclear proteome of C57BL/6J mouse liver tissue. Mouse models, either transgenic or xenograft, represent an invaluable experimental system for understanding human disease pathogenesis. Experimental artifacts related to genetic background and environment can be more carefully controlled using mouse models for research into human disease like cancer. Herein we analyzed the nuclear proteome of C57BL/6J mouse liver tissue after nuclear isolation; 748 non-redundant proteins were identified with the 2-DE MS method. The proteome data were analyzed with respect to localization and function in the nuclei.

## 1 Materials and methods

### 1.1 Materials

Analytical reagent grade chemicals were used throughout unless otherwise stated. Water was prepared using a Milli-Q system (Millipore, Bedford, MA, USA). Nycodenz, sodium orthovanadate ( $\text{Na}_3\text{VO}_4$ ), sodium fluoride (NaF), magnesium chloride hexahydrate ( $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ ), potassium chloride (KCl), AEBSF, pepstatin A, E-64, bestatin, leupeptin and thiourea were obtained from Sigma (St. Louis, MO, USA). Materials employed for 2-D electrophoresis were purchased from GE Healthcare (Uppsala, Sweden). EDTA, EGTA and phenylmethylsulfonyl fluoride (PMSF) were purchased from Amresco (Solon, OH, USA). Centrifuges used were the Himac CR 21G high-speed refrigerated centrifuge and Himac CP 80MX preparative ultracentrifuge, both from Hitachi Koki Co. Ltd. (Tokyo, Japan). Iodoacetamide was obtained from Fluka. Sequencing grade-modified trypsin was from Roche (Switzerland). Horse myoglobin was purchased from Sigma. Adult male C57BL/6J mice were purchased from the Shanghai Laboratory Animal Center (Jiuting, Shanghai, China).

### 1.2 Antibodies

Anti-nucleolin (C23), anti-GAPDH, anti-C-Jun and anti-lamin B2 (Santa Cruz Biotechnology, Santa Cruz, CA, USA) were used as recommended by the manufacturer.

### 1.3 Separation and purification of mouse liver nuclei

For separation by differential centrifugation of mouse liver nuclear fractions, C57BL/6J mice were sacrificed and their livers promptly removed and placed in ice-cold homogenization buffer consisting of 200  $\text{mmol L}^{-1}$  mannitol, 50  $\text{mmol L}^{-1}$  sucrose, 1  $\text{mmol L}^{-1}$  EDTA, 0.5  $\text{mmol L}^{-1}$  EGTA, 10  $\text{mmol L}^{-1}$  Tris-HCl (pH 7.4) and a mixture of protease and phosphatase inhibitors (1  $\text{mmol L}^{-1}$  PMSF, 0.2  $\text{mmol L}^{-1}$   $\text{Na}_3\text{VO}_4$  and 1  $\text{mmol L}^{-1}$  NaF). After mincing with

scissors and washing to remove blood, livers were homogenized in a Potter-Elvehjem homogenizer with a Teflon piston using 10 mL homogenization buffer per 2 g tissue. Centrifugation at successively higher speeds at 4°C yielded the following fractions: crude nuclear fraction at 1000×g for 10 min and mitochondria at 15000×g for 15 min.

Subcellular fractionation of nuclei was performed according to the procedure of Berezney and Coffey [6] with minor modifications. The crude nuclear fraction was resuspended in ice-cold TM buffer (0.25  $\text{mol L}^{-1}$  sucrose, 0.05  $\text{mol L}^{-1}$  Tris pH 7.4 and 5  $\text{mmol L}^{-1}$   $\text{MgCl}_2$ ) containing 10  $\mu\text{g mL}^{-1}$  AEBSF, 5  $\mu\text{g mL}^{-1}$  pepstatin A, 5  $\mu\text{g mL}^{-1}$  E-64, 5  $\mu\text{g mL}^{-1}$  bestatin, 5  $\mu\text{g mL}^{-1}$  leupeptin, 0.2  $\text{mmol L}^{-1}$   $\text{Na}_3\text{VO}_4$  and 1  $\text{mmol L}^{-1}$  NaF. The homogenate was centrifuged at 1000×g for 10 min at 4°C to yield a crude nuclear pellet (CN, crude nuclei). The pellet was resuspended in 2.2  $\text{mol L}^{-1}$  sucrose TM buffer (2.2  $\text{mol L}^{-1}$  sucrose, 0.05  $\text{mol L}^{-1}$  Tris pH 7.4, 5  $\text{mmol L}^{-1}$   $\text{MgCl}_2$ ) containing 10  $\mu\text{g mL}^{-1}$  AEBSF, 5  $\mu\text{g mL}^{-1}$  pepstatin A, 5  $\mu\text{g mL}^{-1}$  E-64, 5  $\mu\text{g mL}^{-1}$  bestatin, 5  $\mu\text{g mL}^{-1}$  leupeptin, 0.2  $\text{mmol L}^{-1}$   $\text{Na}_3\text{VO}_4$  and 1  $\text{mmol L}^{-1}$  NaF, and centrifuged at 40000×g for 90 min at 4°C to yield a primarily purified nuclear pellet. The primarily purified nuclear pellet was washed two times using TM buffer at 40000×g for 10 min.

Nuclei were purified by Nycodenz density gradient centrifugation. The procedures recommended by Axis-Shield PoC were followed (<http://www.axis-shield-poc.com/>). Nycodenz was dissolved at 50% (w/v) in 10  $\text{mmol L}^{-1}$  Tris-HCl, pH 7.4 containing 25  $\text{mmol L}^{-1}$  KCl and 5  $\text{mmol L}^{-1}$   $\text{MgCl}_2$ . This stock solution was diluted with buffer containing 0.25  $\text{mol L}^{-1}$  sucrose, 10  $\text{mmol L}^{-1}$  Tris-HCl, 25  $\text{mmol L}^{-1}$  KCl and 5  $\text{mmol L}^{-1}$   $\text{MgCl}_2$  at pH 7.4. The primarily purified nuclei (PPN) pellet was suspended in 25% Nycodenz and 9 mL of the suspension was then overlaid onto the following discontinuous Nycodenz gradient: 3.6 mL at 50%, 3.6 mL at 40%, 5.4 mL at 34% and 7.2 mL at 30%. The sample in this gradient was topped off with 7.2 mL 20% Nycodenz. The tubes were centrifuged for 90 min at 52000×g at 48°C. The particle bands after centrifugation have been identified by Nycomed Pharma and Invitrogen Life Technologies as follows: nuclei at the 40%/50% interface and mitochondria at the 25%/30% interface. The band at the 40%/50% interface was collected and diluted with the same volume of homogenization buffer and then centrifuged at 15000×g for 30 min. The purified nuclear pellet was then collected.

### 1.4 Transmission electron microscopy and hematoxylin-eosin staining of purified nuclei

Isolated nuclei were processed for transmission electron microscopy using standard methods. In brief, the pelleted nuclei were washed briefly in PBS, fixed in 3.7% paraformaldehyde in PBS, contrasted with 3% uranyl acetate in 70%

ethanol, then further dehydrated with 90% ethanol, 100% ethanol and propylene oxide before embedding in epoxy resin. Sections were cut using a Reichart ultracut ultramicrotome and visualized in a JOEL 1200EX TEM.

To examine nuclear integrity, purified nuclei or cells were deposited on glass slides and centrifuged for 10 min at  $100\times g$ . The samples were fixed with 4% paraformaldehyde (Merck, Whitehouse, Station, NJ, USA) for 20 min and washed with PBS. Cells were permeabilized with 0.5% Triton X-100 for 15 min and washed with PBS. Nuclei or cells were stained for 20 min with 1% hematoxylin and washed extensively with water. Sodium acetate (1%) was added for 10 min, and then samples were stained with 1% eosin.

### 1.5 Protein preparation

Purified nuclei (PN) were suspended in lysis buffer consisting of 7 mol L<sup>-1</sup> urea, 2% 3-[(3-cholamidopropyl)dime-thylammonio]-1-propanesulfonate (CHAPS), 65 mmol L<sup>-1</sup> dithiothreitol and 40 mmol L<sup>-1</sup> Tris, vortexed for 30 min in an ice bath, and centrifuged at  $25000\times g$  for 1 h. Supernatants were collected and a clean-up protocol was performed on the nuclear protein sample to deplete DNA, RNA and other contaminants. The protein concentration was determined by the Bradford assay.

### 1.6 Western blotting

Proteins were separated on a small gel (8 cm $\times$ 7 cm) by 1-D SDS-PAGE and electrophoretically transferred onto nitrocellulose sheets as described [7]. Membranes were blocked in 5% nonfat dry milk in PBS/0.05% Tween 20 and probed at room temperature successively with primary and secondary antibodies. Immune complexes were visualized using ECL reagents (Amersham Biosciences).

### 1.7 2-D gel electrophoresis and MS identification

2-D Gel Electrophoresis IPG strips (length 24 cm, thickness 0.5 mm; GE Healthcare), with a non-linear pH gradient ranging from 3 to 10 and linear pH gradient ranging from 4 to 7, were passively rehydrated with 500  $\mu$ g (for silver stain) or 1000  $\mu$ g (for CCB stain) of sample. IEF was performed using an Electrophoresis Unit (GE Healthcare) at 20°C, with a maximum current setting of 50  $\mu$ A/strip. IEF running conditions were the following: linear voltage up to 200 V in 1 h, up to 1000 V in 1 h, up to 1500 V in 1 h, up to 3500 V in 1 h, constant voltage at 3500 V for 3 h, linear voltage up to 5000 V in 1 h, constant voltage at 5000 V for 4 h, linear voltage up to 8000 V for 1 h and constant voltage at 8000 V until the total product time $\times$ voltage applied was 64000 Vh for each strip. The strips were equilibrated, loaded and run in 24 cm $\times$ 20 cm 12.5% polyacrylamide pre-cast gels (GE Healthcare). The stained gels were scanned using an ImageScanner (Amersham Pharmacia) and analyzed with Im-

ageMaster software (Amersham Pharmacia).

In-Gel Digestion and MS Analysis Gels were washed twice with water before spots were excised. All detected spots were then excised from gels using an Ettan Spot Picker (Amersham Pharmacia). Excised spots were twice destained with 60 mL 50 mmol L<sup>-1</sup> NH<sub>4</sub>HCO<sub>3</sub> and 50% ACN then dried twice with 60 mL ACN. Trypsin (3 mL sequencing grade at 12.5 ng mL<sup>-1</sup>), freshly diluted in 25 mmol L<sup>-1</sup> NH<sub>4</sub>HCO<sub>3</sub>, was added to the dried gel spots. The digestion was performed at 37°C overnight. Sixty microliters of 50% ACN/0.1% TFA was added to the gel and incubated for 20 min. The supernatant with extracted peptides was recovered and the operation was repeated once. All digestion steps were performed with the Ettan TA Digester (Amersham Sciences). The peptide solution was then dried under N<sub>2</sub>. 0.8 mL of matrix (5 mg mL<sup>-1</sup> CHCA diluted in 50% ACN/0.1% TFA) was added and mixed several times to extract the peptides, and the mixture was spotted onto MALDI plate. MS measurements were carried out on an ABI 4700 TOF-TOF MS. Trypsin-digested peptides from horse myoglobin were used as mass standards to calibrate the mass accuracy of the instrument before MS acquisition. After MS acquisition, the five strongest peptides per spot were selected automatically for MS/MS analysis. Data were searched with the search engine MASCOT (Matrix Science, London, UK) against the IPI non-redundant protein sequence database 3.27 version. The mass tolerance was set at 0.3 Da and MS/MS tolerance was 0.4 Da. Proteins with a protein score of more than 60 were accepted. Automated data analysis and database searching were performed by GPS Explorer software.

### 1.8 Data mining to annotate and classify identified proteins

The sub-cellular localizations for our data, especially intranuclear localizations, were obtained by a combined search of GO (<http://www.geneontology.org>), LOCATE and NoPdb [8,9] as well as predicted by Proteome Analyst [10,11], NucPred [12], pTarget [13,14] and WoLFPSORT [15].

DAVID tools (Database for Annotation, Visualization and Integrated Discovery, <http://david.abcc.ncifcrf.gov/>) and IPA (Ingenuity Pathways Analysis, <http://www.ingenuity.com/>) were used to perform gene enrichment analysis on the functions, biological processes and interaction networks for the identified proteins. Results were ranked by the ascending *P*-value of Fisher's test and cut off by a default median threshold. Furthermore, enrichment categories which have most common genes will be grouped in one cluster to represent higher functional categories.

## 2 Results

### 2.1 Isolation of nuclei

It is essential to have highly purified nuclei prepared as a

source for direct nuclear protein analysis. To ensure the integrity and purity of nucleus fractions, the organelle isolation procedure should be carefully considered and selected. In the homogenization process, liver tissue was pre-homogenized with two strokes in a Potter-type homogenizer to maximize exposure of cells to the hypotonic lysis buffer. After 5 min incubation in hypotonic buffer to render the cells fragile, only four strokes were applied to release the subcellular fractions from the cells. Differential centrifugation based on the inherent density of the nucleus was then applied to obtain the PPN fraction. The percentage yield of PPN (the percentage of the PPN weight constituting the initial tissue weight) was below 10%, and the average percentage yield was 8.5% from eight experiment trials. To purify further the PPN, Nycodenz density gradient centrifugation was performed to obtain the purified nuclei (PN) fraction. The density gradient was configured as described in Section 2.5. After centrifugation the nuclear band at the 40/50% interface was collected. The 4.5% average PN percentage yield was obtained from eight repeated extractions. The percentage yield of protein extracted from nuclei was about 2.5%, which meant that only 1 mg of nuclear protein could

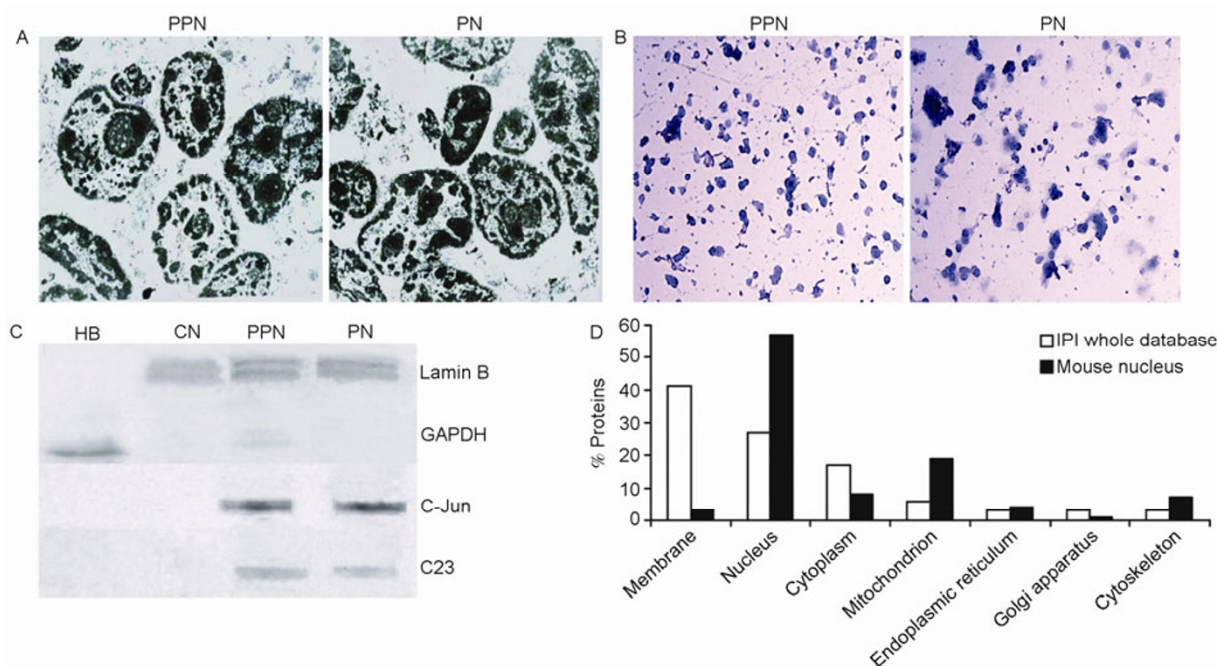
be extracted from 1 g of liver tissue (Table 1).

## 2.2 Evaluation of nuclear integrity and nuclear protein enrichment due to purification

The purity and integrity of PN were analyzed by several methods. TEM and H-E staining analysis (Figure 1A and B) showed that the purified nuclei had a typical nuclear morphology. Western immunoblot analyses were performed on protein samples from extracts of mouse liver, crude nuclei, primarily purified nuclei and purified nuclei. Three nuclear marker proteins (Lamin B2, C23 and C-jun) and a reference protein located in the cytoplasm (GAPDH) were used to evaluate the efficiency of the purification strategy. It was noted that immunoblot signals for Lamin B2, C23 and C-jun (Figure 1C) could not be observed in crude homogenate extracts of mouse liver tissue and the CN fraction, whereas immunoblot signals were observed in the PPN and PN protein extracts. The absence of signal was due to the low concentrations of these two markers in the protein extracts from liver tissue and CN. The signal presence in PPN and PN

**Table 1** Recovery from each step during nuclear purification

Tissue	24 g	100%	16 g	100%	24 g	100%	24 g	100%
CN	14 g	58.70%	8.7 g	54.00%	12 g	50.00%	13 g	54.00%
PPN	1.42 g	5.90%	1.68 g	10.50%	3 g	12.50%	2.4 g	10.00%
PN	0.628 g	2.60%	0.473 g	2.90%	0.607 g	2.50%	0.650 g	2.70%
Protein			15 mg	0.10%	30 mg	0.13%		



**Figure 1** Evaluation of nuclear purification. A, TEM images of isolated primary purified nuclei (PPN) and purified nuclei (PN) from C57BL/6J mouse liver. B, H-E stain of isolated PPN and PN from C57BL/6J mouse liver. C, Western blotting of four protein fractions: total proteome (HB), crude nuclear (CN), primary purified nuclear (PPN) and purified nuclear (PN) using a cytoplasmic marker (GAPDH) and three nuclear markers (C23, C-Jun and Lamin B). D, GO analysis for cellular components of the total proteins in the IPI mouse database and purified nuclear proteome from this work.

protein extracts demonstrated that these nuclear proteins had been enriched following purification of nuclei. Conversely, the signal intensity of GAPDH decreased in PPN and PN, which indicated depletion of non-nuclear proteins during purification.

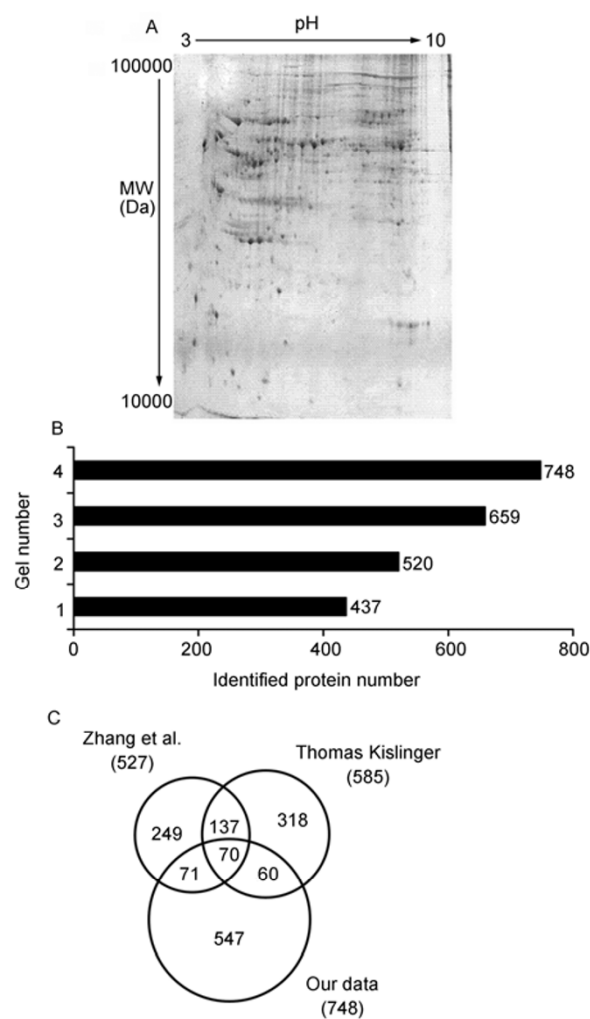
In Figure 1D, the proportions of nuclear proteins among total proteins in the IPI database and our identified data set were compared. The results show that nuclear proteins were greatly enriched. Meanwhile, proteins from other compartments such as mitochondria, cytoskeleton and endoplasmic reticulum (ER) were also enriched to varying degrees. One reason is that some nuclear proteins have multiple locations in the GO database, for example Lamin A/C, a nuclear protein, is also present in the cytoplasm and plasma membrane. Keratins, widespread in nature, were found in both nuclei and cytoplasm of almost all differentiated eukaryote cells [16]. Thus, along with enrichment of nuclei, other cellular compartments sharing proteins with nuclei (like the cytoskeleton) also seem to be enriched. Another possibility is that compartments with similar density cannot be separated completely.

### 2.3 Expression proteome profile of C57BL/6J mouse liver nuclei

There are 748 unique proteins found in the C57BL/6J mouse liver nuclear proteome from a total of 5987 spots cut from four gels (3–10 Ag; 3–10 CCB; 4–7 Ag; 4–7 CCB), of which 3825 spots were successfully identified by MS. Figure 2A shows an image of one gel (pI 3–10 Ag stain). Data from the gels complemented each other so that more proteins were successfully identified (Figure 2B). During the differential velocity centrifugation and density gradient centrifugation steps, many contaminants such as sucrose, Nycodenz and inorganic ions were retained in the protein sample. A clean-up protocol (Biorad) to deplete those contaminants was necessary for a successful separation. After sample clean-up, proteins were separated with 2-DE.

To evaluate the sensitivity of the 2-DE MS/MS strategy, we compared our data with two other datasets: Emilifour's study using *Mus musculus* [5] and Zhang's work [17] for the C57BL/6J strain in collaboration with us. A shotgun strategy was used in both works. Searching against the SwissProt database (SP), 824 and 462 nuclear proteins were identified respectively, and all SP identifiers were consequently converted into corresponding IPI identifiers for facility of comparison among datasets.

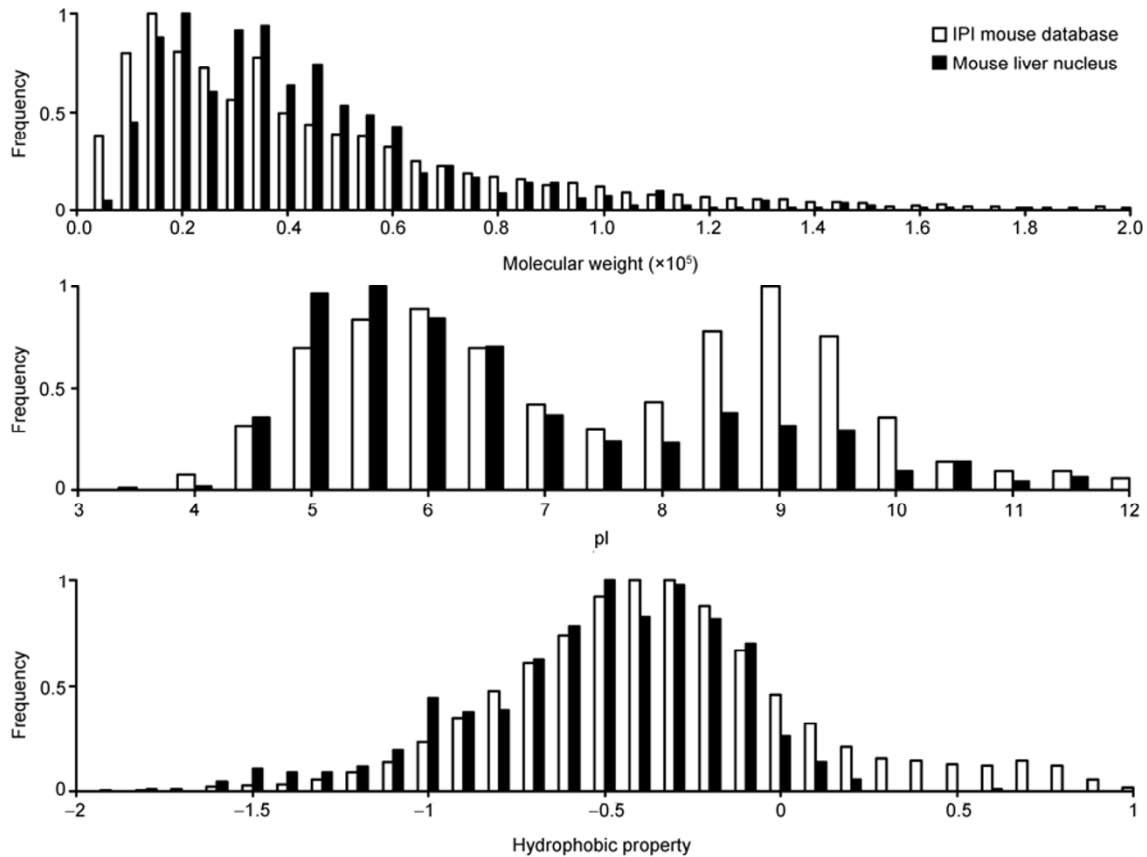
In comparison, our 2-DE MS/MS strategy showed comparable if not better sensitivity than the shotgun strategies. A total of 748 proteins were identified through the 2-DE MS/MS strategy, whereas 527 and 585 proteins (in the IPI database) were identified through the shotgun strategies. Another interesting observation is that, although using exactly the same sample and same nuclear purification method, Zhang's data share fewer proteins with our data than with



**Figure 2** Identified protein data. A, 2-DE image of nuclear proteins (silver stain). 400  $\mu$ g protein with 2050 protein spots detected. B, Identified protein numbers increased with the gel number. C, Comparison with other data.

Emilifour's. This indicates that 2-DE MS/MS and shotgun strategies complement each other remarkably. We also noticed that proteins shared by two or three datasets were quite rare. The different origins of the mouse sample and the transfer from Swissprot to IPI may affect the consistency of the three datasets. Another possible reason is that the mouse nuclear proteome has far more proteins than those identified, such that each data set may only represent a portion of it. For all we know, there are 6387 proteins in the IPI database (IPI mouse version 3.44) located in nucleus as annotated by the GO Slim approach.

Statistics for the pI, MW and hydrophobic properties of the identified proteins are shown in Figure 3. Cellular localization classification of the identified proteins is shown in Figure 4. Our data are similar to the IPI mouse database regarding physical and chemical properties; however, basic proteins were rarer in our data. One reason may be that there were discriminatory effects in the protein extraction



**Figure 3** Molecular weight, pI, and hydrophobic properties of identified proteins (using the IPI mouse database as background).

and 2-DE separation. On the other hand, the data set used as background (IPI mouse database) may have different properties than mouse liver nuclear proteins. Another interesting observation is the tissue expression of identified proteins, seen in Figure 5A. Unsurprisingly most proteins are expressed in liver. Proteins expressed in other tissues such as bone marrow were also greatly enriched, indicating that liver shares more proteins with the enriched tissue.

### 3 Discussion

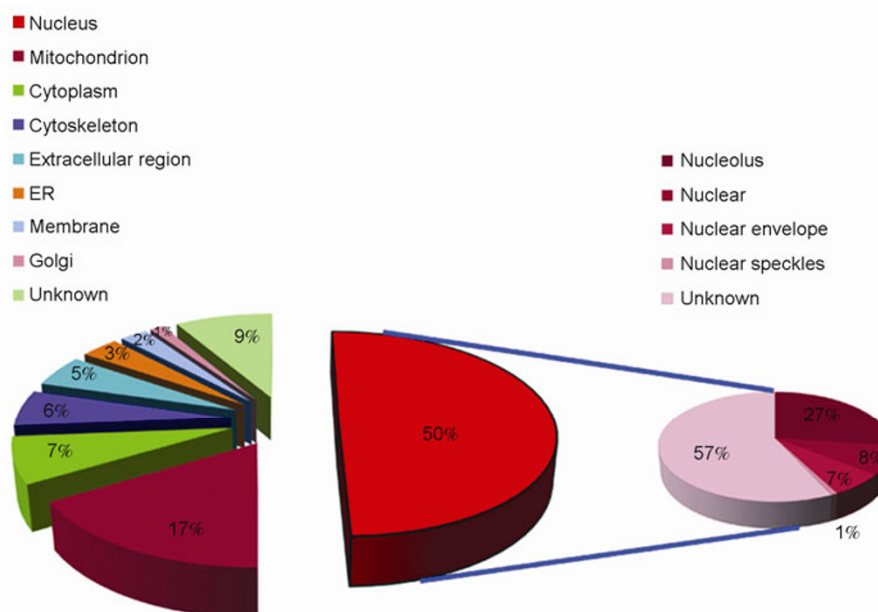
#### 3.1 Description of nuclear proteome

We classified the proteins identified according to their cellular component using a combination of various methods (see Section 2.8), and some nuclear proteins were shown with their intranuclear locations as displayed in Figure 4. Based on the DAVID and IPA bioinformatics analysis and literature searches, the functions, biological processes and networks of the proteins were also studied, as shown in Figures 5–7. The structure, function, and dynamic regulation of liver nuclei are discussed below, based on the nuclear protein data set provided by our study. The data set acquired from C57BL/6J mouse liver nuclei is representative for the liver nuclear proteome.

#### 3.2 Sub-nuclear localization of identified proteins

The main structures making up the nucleus are the nuclear envelope, a double membrane that encloses the entire organelle and separates its contents from the cellular cytoplasm, and the nuclear lamina, a meshwork within the nucleus that adds mechanical support much like the cytoskeleton supports the cell as a whole. Nuclear pores span both membranes of the nuclear envelope, providing a channel that allows free movement of small molecules and ions. Although the interior of the nucleus does not contain any membrane-bound sub-compartments, its contents are not uniform and a number of sub-nuclear bodies exist that are made up of unique proteins, RNA molecules and distinct regions of the chromosomes. The best known of these is the nucleolus, which is mainly involved in assembly of ribosomes.

Figure 4 shows the intranuclear distribution of identified nuclear proteins. There are 369 proteins located in nuclei with GO analysis. More than 40% (159 proteins) were located into nuclear compartments based on references and Locate [18]. The most abundant intranuclear compartments are the nucleolus and nuclear envelope. Many proteins still lack sub-organellar localization data, which are closely related with their functions and need to be studied.



**Figure 4** Localization analysis of C57BL/6J liver nuclear proteome data versus intranuclear localization.

### 3.3 Primary functions of the nuclear proteome

The primary functions of the cell nucleus are to control gene expression and mediate replication of DNA during the cell cycle. The nucleus provides a site for genetic transcription that is segregated from the location of translation in the cytoplasm, allowing modes of gene regulation that are not available to prokaryotes. DAVID bioinformatics analysis and IPA analysis were used to study the main functions of the liver nuclear proteome. DAVID bioinformatics analysis can perform Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping and other analyses. Using the mouse genome as background, DAVID analysis shows enriched GO entries, protein domains, functions and pathways in a more scientifically relevant way.

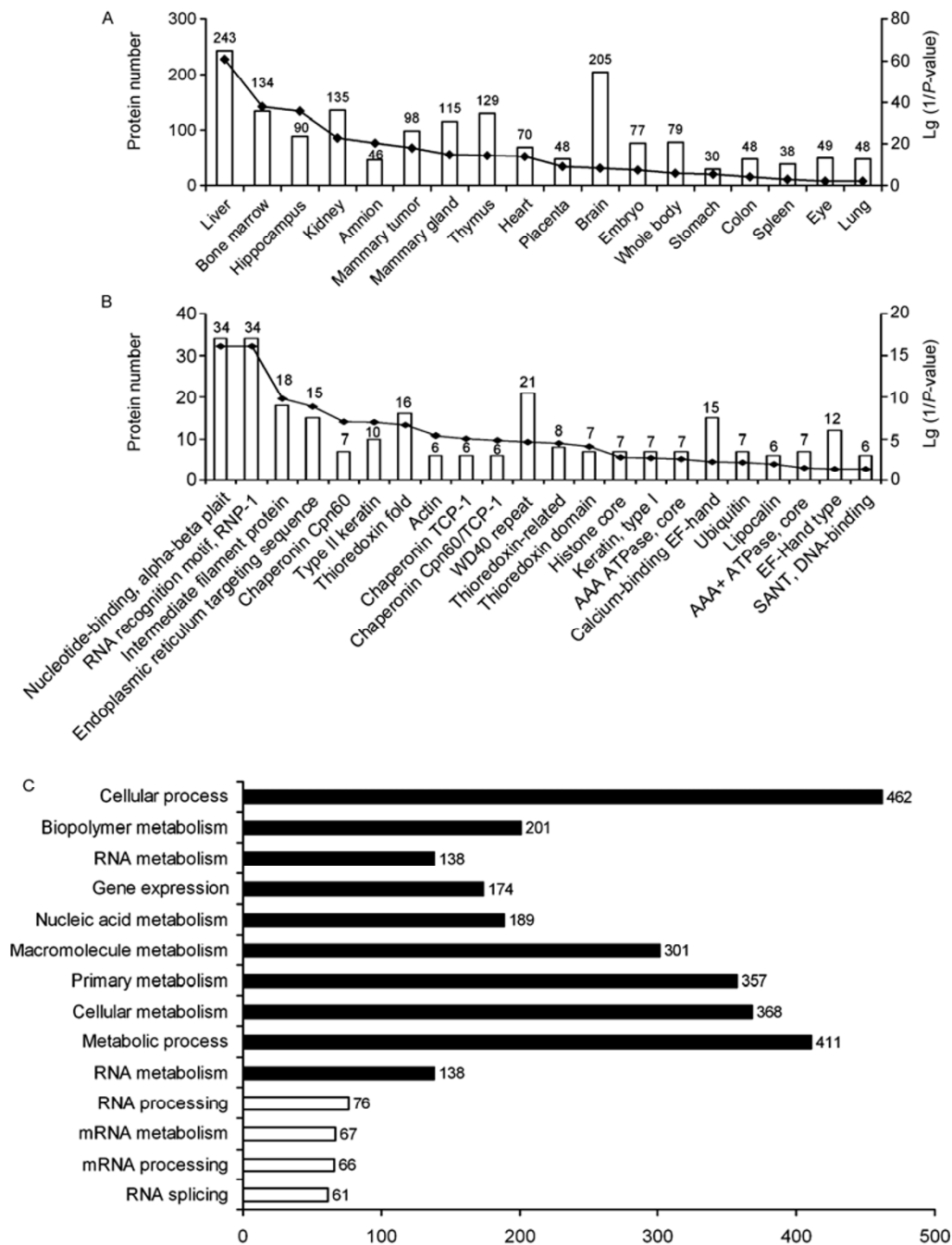
Figure 5B shows the protein domains appearing most frequently in the liver nuclear proteome. The most enriched domains are the nucleotide-binding alpha-beta plait and the RNA recognition motif RNP-1. The first domain has an alpha-beta plait structure, which consists of either a ferredoxin-like beta-alpha-beta(2) fold, such as that found in RNA-binding domains of various ribonucleoproteins or in viral DNA-binding domains [19,20], or a beta-(alpha)-beta-alpha-beta(2) fold like that found in the ribosomal protein L23 [21]. RNP-1 is an eight amino-acid consensus sequence [22,23] found in the largest group of single stranded RNA-binding proteins, the eukaryotic RNA recognition motif (RRM) family, which has a variety of RNA binding preferences and functions [24–26]. As shown in Figure 5C, the most enriched biological processes are biopolymer metabolism, especially of RNA, nucleosides, nucleotides and nucleic acids. Another enriched process is gene expression.

Figure 6 shows IPA analysis results regarding the main functions of the identified proteins. The molecular functions enriched were mainly focused on processing of DNA and RNA, such as splicing and modification of RNA or mRNA and expression, transcription and modification of DNA.

Topological results of protein-protein interaction networks with IPA analysis were also obtained, such that our main sub-networks and the overall networks are shown in Figure 7. The main sub-networks are (i) gene expression, DNA replication, recombination and repair and RNA post-transcriptional modification; (ii) RNA post-transcriptional modification, gene expression, cellular assembly and organization; (iii) RNA post-transcriptional modification, gene expression and cell cycle; and (iv) cellular assembly and organization, molecular transport and cardiovascular disease. These results mostly agree with the molecular function and biological process analysis.

### 3.4 Cell compartmentalization

The nuclear envelope allows the nucleus to control its contents and separate them from the rest of the cytoplasm where necessary. This is important for controlling processes on either side of the nuclear membrane. In cases where a cytoplasmic process needs to be restricted, a key participant is removed to the nucleus where it interacts with transcription factors to downregulate production of certain enzymes in the pathway. To control which genes are being transcribed, the cell prevents some transcription factors responsible for regulating gene expression from physically accessing the DNA until they are activated by other signaling pathways. This prevents even low levels of inappropriate gene expression. Furthermore, compartmentalization allows



**Figure 5** DAVID analysis of C57BL/6J liver nuclear proteome data. A, Tissue expression. B, Enriched protein domains. C, Main biological process.

the cell to prevent translation of unspliced mRNA [27] and avoid translating newly transcribed (unprocessed) mRNA that would result in misformed or nonfunctional proteins. We found 25 proteins that are located in the nuclear envelope. Formation of the nuclear envelope is one of the enriched molecular functions (Figure 4).

### 3.5 Gene expression

Gene expression first involves transcription, in which DNA

is used as a template to produce RNA. In the case of genes encoding proteins, the RNA produced from this process is messenger RNA (mRNA), which then is translated by ribosomes to form a protein. Because ribosomes are located outside the nucleus, the mRNA produced needs to be exported [28]. This is also one of the most enriched functions or processes indicated in our data.

Because the nucleus is the site of transcription, it also contains a variety of proteins that either directly mediate transcription or are involved in regulating this process.



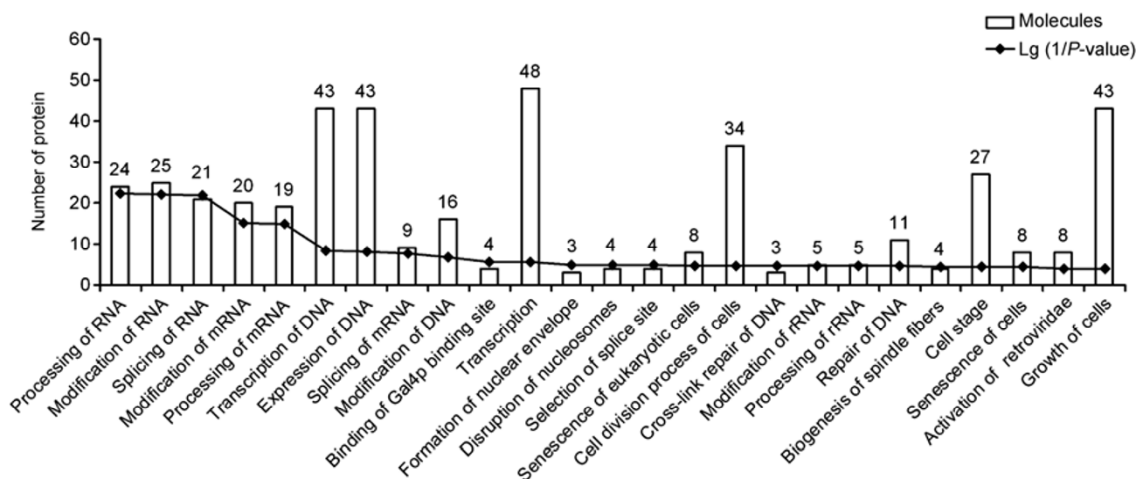


Figure 6 IPA enrichment analysis of molecular functions.

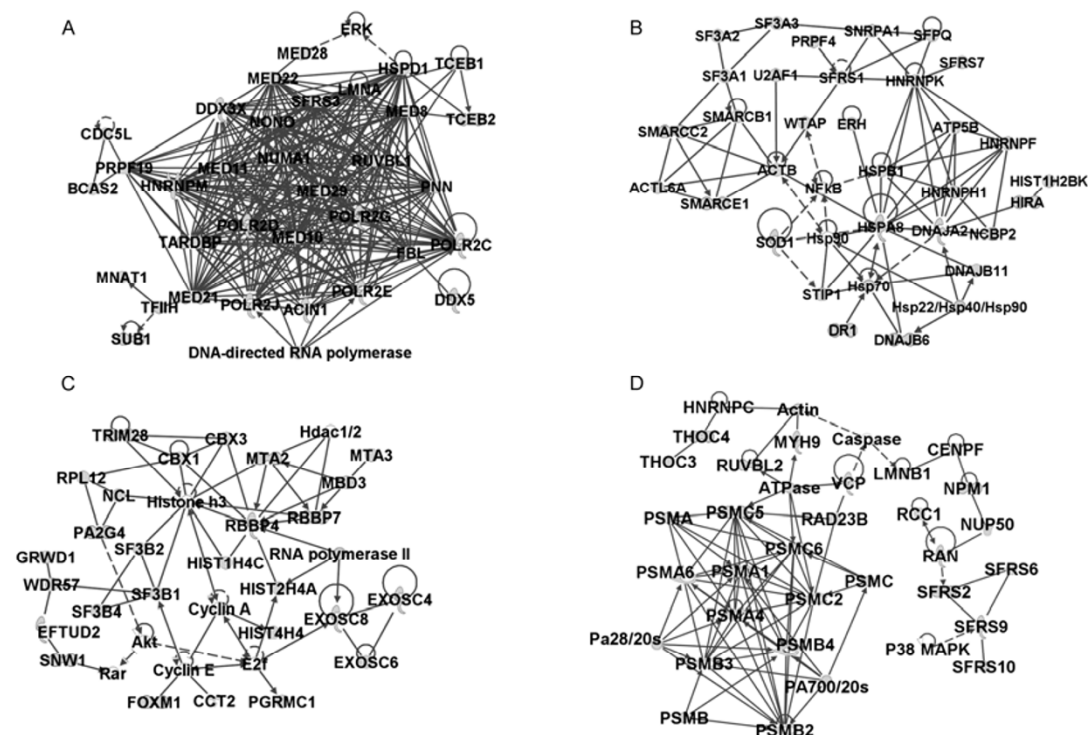


Figure 7 IPA enrichment analysis of the protein-protein interaction network. Four main sub-networks were identified: (A) gene expression, DNA replication, recombination, and repair and RNA post-transcriptional modification; (B) RNA post-transcriptional modification, gene expression, cellular assembly and organization; (C) RNA post-transcriptional modification, gene expression, cell cycle; (D) cellular assembly and organization, molecular transport and cardiovascular disease.

These proteins include helicases (five were found in this work) that unwind double-stranded DNA molecule to facilitate access, RNA polymerases (seven RNA polymerases and ten mediators of RNA polymerases were in our data set) that synthesize the growing RNA molecule, topoisomerases that change the amount of supercoiling in DNA and thereby help it wind and unwind, as well as a large variety of transcription factors that regulate expression [29].

### 3.6 Processing of pre-mRNA

Newly synthesized mRNA molecules are known as primary transcripts or pre-mRNA. They must undergo post-transcriptional modification in the nucleus before being exported to the cytoplasm. The three main modifications are 5' capping, 3' polyadenylation and RNA splicing. While in the nucleus, pre-mRNA is associated with a variety of proteins in complexes known as heterogeneous ribonucleopro-

tein particles (hnRNPs). Addition of the 5' cap occurs co-transcriptionally and is the first step in post-transcriptional modification. The 3' poly-adenosine tail is added only after transcription is complete.

There are 15 hnRNP proteins and their isoforms in our data set, including hnRNP A2/B1, C1/C2, H, M, K, etc. These proteins associate with pre-mRNAs to form hnRNP complexes, in which pre-mRNAs are processed into mature mRNAs [6,7,30–32]. Although the function of hnRNPs has not been fully unraveled, they are assumed to have a dynamic activity designed to bring RNAs into required configurations for precise splicing [7,32–35]. The variety of hnRNPs [36], their modular structure based on RNA- and protein-binding domains, their sequence-specific binding to pre-mRNA [37] and also the structural relationships between species most likely allow cells to tune splicing precisely.

### 3.7 Dynamics and regulation

#### 3.7.1 Nuclear transport

The entry and exit of large molecules from the nucleus is tightly controlled by nuclear pore complexes. Ran (RAs-related Nuclear protein) is a small GTP-binding protein belonging to the Ras superfamily that is essential for translocation of RNA and proteins through the nuclear pore complex [16,19,20]. The Ran protein is also involved in control of DNA synthesis and cell cycle progression. Nuclear localization of Ran requires the presence of regulator of chromosome condensation 1 (RCC1) and mutations in Ran disrupt DNA synthesis. Because of its many functions, Ran likely interacts with several other proteins [21]. In this work, Ran was also found among the identified proteins and its main protein-protein interaction network is molecular transport (Figure 7A–D).

#### 3.7.2 Assembly and disassembly

During its lifetime a nucleus may be broken down, either in the process of cell division or as a consequence of apoptosis. During these events, the structural components of the nucleus—the envelope and lamina—are systematically degraded. During the cell cycle, assembly and disassembly of the nuclear lamina are regulated by dephosphorylation and phosphorylation of Lamin (six were found among our data).

## 4 Conclusion

This work provides an effective approach for a subcellular proteomic study based on enrichment of nuclear organelles and 2-DE MS protein separation and identification; a total of 748 proteins have been identified with this methodology. Analysis of the identified nuclear-expressed proteome demonstrates that subcellular proteomics is a promising strategy for maximizing proteomic coverage for a given organism.

The data set obtained in this work is shown to be representative of the liver nuclear proteome and can be used as a reference for research into liver nuclei from other species.

*This work was supported by the National Basic Research Program of China (2013CB910802, 2012CB910602, 2010CB912704), National High Technology Research and Development Program of China (2012AA020201, 2012AA020202), State Key Project Specialized for Infectious Diseases of China (2012ZX10002-012), and National Natural Science Foundation of China (30700990, 20975024, 31000379). The authors declare that they have no competing interest.*

- 1 Chromy B A, Gonzales A D, Perkins J, et al. Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J Proteome Res*, 2004, 3: 1120–1127
- 2 Freeman W M, Lull M E, Guilford M T, et al. Depletion of abundant proteins from non-human primate serum for biomarker studies. *Proteomics*, 2006, 6: 3109–3113
- 3 Lull M E, Freeman W M, Myers J L, et al. Plasma proteomics: a noninvasive window on pathology and pediatric cardiac surgery. *ASAIO J*, 2006, 52: 562–566
- 4 Foster L J, de Hoog C L, Zhang Y, et al. A mammalian organelle map by protein correlation profiling. *Cell*, 2006, 125: 187–199
- 5 Kislinger T, Cox B, Kannan A, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell*, 2006, 125: 173–186
- 6 Grigoryev S A, Bulynko Y A, Popova E Y. The end adjusts the means: heterochromatin remodelling during terminal cell differentiation. *Chromosome Res*, 2006, 14: 53–69
- 7 Schardin M, Cremer T, Hager H D, et al. Specific staining of human chromosomes in chinese hamster X man hybrid cell lines demonstrates interphase chromosome territories. *Hum Genet*, 1985, 71: 281–287
- 8 Andersen J S, Lam Y W, Leung A K, et al. Nucleolar proteome dynamics. *Nature*, 2005, 433: 77–83
- 9 Ahmad Y, Boisvert F M, Gregor P, et al. Nopdb: nucleolar proteome database—2008 update. *Nucleic Acids Res*, 2009, 37: D181–184
- 10 Szafron D, Lu P, Greiner R, et al. Proteome analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. *Nucleic Acids Res*, 2004, 32: W365–371
- 11 Lu P, Szafron D, Greiner R, et al. Pa-gosub: a searchable database of model organism protein sequences with their predicted gene ontology molecular function and subcellular localization. *Nucleic Acids Res*, 2005, 33: D147–153
- 12 Brameier M, Krings A, MacCallum R M. Nucpred—predicting nuclear localization of proteins. *Bioinformatics*, 2007, 23: 1159–1160
- 13 Guda C. Ptarget: a web server for predicting protein subcellular localization. *Nucleic Acids Res*, 2006, 34: W210–213
- 14 Guda C, Subramaniam S. Ptarget: a new method for predicting protein subcellular localization in eukaryotes. *Bioinformatics*, 2005, 21: 3963–3969
- 15 Horton P, Park K J, Obayashi T, et al. Wolf psort: protein localization predictor. *Nucleic Acids Res*, 2007, 35: W585–587
- 16 Plowman J E. The proteomics of keratin proteins. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2007, 849: 181–189
- 17 Zhang J, Xu X, Shen H, et al. Analysis of nuclear proteome in C57 mouse liver tissue by a nano-flow 2-D-LC-ESI-MS/MS approach. *J Sep Sci*, 2006, 29: 2635–2646
- 18 Sprenger J, Lynn Fink J, Karunaratne S, et al. Locate: a mammalian protein subcellular localization database. *Nucleic Acids Res*, 2008, 36: D230–233
- 19 Kielkopf C L, Lucke S, Green M R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev*, 2004, 18: 1513–1526
- 20 Bochkarev A, Barwell J A, Pfuetzner R A, et al. Crystal structure of the DNA-binding domain of the epstein-barr virus origin-binding

- protein, EBNA1, bound to DNA. *Cell*, 1996, 84: 791–800
- 21 Chenuil A, Solignac M, Bernard M. Evolution of the large-subunit ribosomal RNA binding site for protein L23/25. *Mol Biol Evol*, 1997, 14: 578–588
- 22 Bandziulis R J, Swanson M S, Dreyfuss G. RNA-binding proteins as developmental regulators. *Genes Dev*, 1989, 3: 431–437
- 23 Query C C, Bentley R C, Keene J D. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, 1989, 57: 89–101
- 24 Dreyfuss G, Swanson M S, Pinol-Roma S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mrna formation. *Trends Biochem Sci*, 1988, 13: 86–91
- 25 Chambers J C, Kenan D, Martin B J, et al. Genomic structure and amino acid sequence domains of the human Ia autoantigen. *J Biol Chem*, 1988, 263: 18043–18051
- 26 Sachs A B, Davis R W, Kornberg R D. A single domain of yeast poly(A)-binding protein is necessary and sufficient for rna binding and cell viability. *Mol Cell Biol*, 1987, 7: 3268–3276
- 27 Moreno F, Ahuatzli D, Riera A, et al. Glucose sensing through the HXK2-dependent signalling pathway. *Biochem Soc Trans*, 2005, 33: 265–268
- 28 Gorlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol*, 1999, 15: 607–660
- 29 Nierhaus K H, Wilson D H, eds. *Protein Synthesis and Ribosome Structure: Translating the Genome*. Weinheim: Wiley-VCH, 2004
- 30 Lamond A I, Earnshaw W C. Structure and function in the nucleus. *Science*, 1998, 280: 547–553
- 31 Kurz A, Lampel S, Nickolenko J E, et al. Active and inactive genes localize preferentially in the periphery of chromosome territories. *J Cell Biol*, 1996, 135: 1195–1205
- 32 Rothfield N F, Stollar B D. The relation of immunoglobulin class, pattern of anti-nuclear antibody, and complement-fixing antibodies to DNA in sera from patients with systemic lupus erythematosus. *J Clin Invest*, 1967, 46: 1785–1794
- 33 Bamed S, Goodman A D, Mattson D H. Frequency of anti-nuclear antibodies in multiple sclerosis. *Neurology*, 1995, 45: 384–385
- 34 Hernandez-Verdun D. Nucleolus: from structure to dynamics. *Histochem Cell Biol*, 2006, 125: 127–137
- 35 Lamond A I, Sleeman J E. Nuclear substructure and dynamics. *Curr Biol*, 2003, 13: R825–828
- 36 Cioce M, Lamond A I. Cajal bodies: a long history of discovery. *Annu Rev Cell Dev Biol*, 2005, 21: 105–131
- 37 Dundr M, Misteli T. Functional architecture in the cell nucleus. *Biochem J*, 2001, 356: 297–310

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.