Liverbase: A Comprehensive View of Human Liver Biology

Aihua Sun,† Ying Jiang,*† Xue Wang,† Qijun Liu,† Fan Zhong,§ Quanyuan He,† Wei Guan,† Hao Li,† Yulin Sun,⊥ Liang Shi,§ Hong Yu,† Dong Yang,† Yang Xu,⊥ Yanping Song,† Wei Tong,§ Dong Li,† Chengzhao Lin,§ Yunwei Hao,† Chao Geng,† Dong Yun,§ Xuequn Zhang,† Xiaoyan Yuan,† Ping Chen,† Yunping Zhu,⊥ Yixue Li,† Songping Liang,⊥ Xiaohang Zhao,⊥ Siqi Liu,¶ and Fuchu He*,†,§

State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 102206, P. R. China, National Laboratory for Parallel & Distributed Processing, National University of Defense and Technology, Changsha 410073, P. R. China, Institutes of Biomedical Sciences and Department of Chemistry, Fudan University, Shanghai 200032, P. R. China, College of Life Sciences, Hunan Normal University, Lushan Road 14, Changsha, Hunan Province, 410081, P. R. China, State Key Laboratory of Molecular Oncology, Cancer Institute & Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 17 Panjiayuan, Chaoyangqu, Beijing, 100021, P. R. China, Beijing Genomics Institute, Chinese Academy of Sciences, Beijing Airport Industrial Zone B-6, Beijing, 101300, P. R. China, and Shanghai Center for Bioinformatics Technology, 100 Qinzhou Road, Shanghai 200235, P. R. China

Received February 28, 2009

The Liverbase (http://liverbase.hupo.org.cn) integrates information on the human liver proteome, including the function, abundance, and subcellular localization of proteins as well as associated disease information. The overall objective of the Liverbase is to provide a unique public resource for the liver community by providing comprehensive functional annotation of proteins implicated in liver development and disease. The central database features are manually annotated proteins localized in or functionally associated with human liver. In this first version of Liverbase, the associated data includes the human liver proteome (6788 proteins) and transcriptome (11205 significantly expressed genes: 10224 from CHIP and 5422 from MPSS, respectively) from the Chinese human liver proteome project (CNHLPP). As a database made publicly available through the Web site, Liverbase provides browsing and searching capabilities and a compilation of external links to other databases and homepages. Liverbase enables (i) the establishment of liver GO slim with 51 nonredundant items; (ii) systematic searches for proteins within specific functional or metabolic pathways; (iii) systematic searches that aim to find the proteins that underlie common and rare liver diseases; and (iv) the integration of detailed protein annotations derived from the literature. Liverbase also contains an external links page with links to other biological databases and homepages, including GO, KEGG, pfam, SWISS-PROT, and GNF databases. Liverbase users can utilize all these information to conduct systems biology research on liver.

Keywords: liver • transcriptome • proteome • quantification • localization

1. Introduction

The liver plays a significant role in metabolism and is the primary source of plasma proteins. Liver diseases are some of the most life-threatening illnesses worldwide. Despite the importance of the liver to health, the variety, number, and abundance of liver proteins have not been extensively described. Because liver is such a complex biological system, global analysis at the “omics” level is necessary to fully elucidate its functions. The Human Liver Proteome Project (HLPP), as one of the large-scale proteomic initiatives coordinated by the Human Proteome Organization, aims to uncover the proteomic atlas of liver development, physiology, and pathology.1−3 Under the project’s umbrella, analyses of the proteomes of human fetal liver,4 mouse liver organelles,5 and rat liver secretory system6 have made inspiring progress. In addition, the proteomes of human liver carcinoma cells7,8 have been described as well.

Thus, the Liverbase integrates diverse information on liver (e.g., relating to the proteome and transcriptome) from the Chinese Human Liver Proteome Project (CNHLPP) and pro-
low-abundance peptides. All qualified peak lists were accepted of peptide capture and to enhance the possibility of capturing parallel at least two times to reduce the random component MALDI, and (4) 1DE-LC-ESI. All determinations were run in analysis were designed: (1) 2DLC-ESI, (2) 3DLC-ESI, (3) 2DE-

theoretical number of tryptic peptides of a protein, was

Spectral Count Index (SCI), a normalized SC in accounting for

more than six amino acids. All identified proteins should have two

obtaining false positives, the identification thresholds of peptide

zol reagent and DNA contaminants were removed with RNase-

Overall, the human liver proteome (HLP) data set comprises

provides a good foundation resource for this information. The overall objective of the Liverbase is to provide a unique public resource for the international liver community by providing comprehensive functional annotation for proteins implicated in liver development and disease.

2. Database Content and Structure

2.1. Database Content. The Liverbase is publicly available at http://liverbase.hupo.org.cn. The Web site uses JSP technology and Office Access 2003. JSP technology, which is based on powerful Java language, has some advantages, such as good scalability, system multiplatform support, diversification, and development of powerful support tools; it is also closely integrated with Java Enterprise API.

The database consists of four parts: browsing, searching, results, and a compilation of external links to other databases and homepages. The list of proteins is based on the proteome and transcriptome profiles of the healthy human liver, which were generated as the result of a collaborative effort of the CNHLPP. Table 1 provides an overview of the current contents of Liverbase.

2.1.1. Proteome. Four technical approaches for proteome analysis were designed: (1) 2DLC-ESI, (2) 3DLC-ESI, (3) 2DEMALDI, and (4) 1DE-LC-ESI. All determinations were run in parallel at least two times to reduce the random component of peptide capture and to enhance the possibility of capturing low-abundance peptides. All qualified peak lists were accepted for protein search against databases of both IPI Human v3.07 and its reversed version. To evaluate the likelihood of obtaining false positives, the identification thresholds of peptide were established at 95% confidence. For the shotgun strategy, qualified peptides should have an MS/MS sequence with more than six amino acids. All identified proteins should have two or more peptide matches. Semiquantitation of protein abundance was estimated by the spectral counts (SC) method. The Spectral Count Index (SCI), a normalized SC in accounting for theoretical number of tryptic peptides of a protein, was calculated according to previously reported methods. To integrate large-scale data from multiple sources, SCI values from different batches are normalized and this normalized SCI value was termed the Spectral Count Index Normalized (SCIN). Overall, the human liver proteome (HLP) data set comprises 6788 identified proteins.

2.1.2. Transcriptome. Total RNAs were extracted with TRI-zol reagent and DNA contaminants were removed with RNase-free DNase I. After assessing the concentration and quality of the RNA, equal amounts of RNAs from 10 donors were pooled, labeled, and hybridized to HG-U133 plus 2.0 high-density oligonucleotide arrays (Affymetrix). The same pooled RNAs were profiled by massively parallel signature sequencing (MPSS) (TaKaRa, Japan). Only those genes that were declared ‘present’ in two chips with fluorescence intensity greater than 100 were taken into account. A total of 10224 and 5422 genes were detected by microarray and MPSS, respectively. When the two data sets were integrated, a total of 11205 genes were identified as the human liver transcriptome (HLT) core data set and included in Liverbase.

Gene Ontology (GO) can provide a simplified summary of the function of each gene. GO slims are created by users according to their needs. In the study, 375 GO terms and 51 nonredundant items were established as liver GO slim and could be widely applied in liver research. The Liver GO slims are available from the GOFact Web site at http://61.50.138.118/gofact, which is linked to Liverbase at “Liver GO slim”.

Liver is a metabolically active organ responsible for many vital life functions. All 94 metabolic pathways in human were extracted from KEGG http://www.genome.jp/kegg/ and used for data browsing in Liverbase. Within the metabolic pathways, 1783 proteins (including 912 enzymes) were detected through HLP determinations and listed in Liverbase. Through this database, users can easily see all proteins involved in liver metabolism and in specific functional categories.

A total of 938 transporters were identified in the HLP data set. An overview of the possible transport pathways was integrated from the identified transporter proteins in liver (Supporting Information Figure S1). Transcription factors are critical for regulation of gene expression and are commonly expressed at low abundance in cells. In the TransFac database, 1670 transcription factors have been collected from human tissues. Of these, 67.6% (1129) were identified in the HLT data set and 18.5% (309) were recognized in the HLP data set (Supporting Information Table S1).

To provide a more comprehensive functional annotation for genes implicated in liver disease, a Liver Disease-related Genes and Proteins reference data set (LDGP, Supporting Information Table S2), composed of 240 nonredundant liver disease related genes and proteins, was compiled. The diseases covered alcoholic liver disease, cirrhosis, hepatitis, and tumor, and these disease related informations were annotated to the Liverbase proteins.

2.2. Search Page. Liverbase is a fully searchable database. Users can access the ‘Search’ page by clicking on the term ‘Search’ located on the home page (http://liverbase.hupo.org.cn/liverbase). From there (Figure 1), users may search for specific data using six parameters for queries: KEGG pathways, functional categories, liver disease category, OMIM, and keywords. By combining these different search parameters, more complex searches can be performed. Leaving the input boxes empty retrieves all entries. To choose a KEGG pathway or functional category, users must specify one from the pull-down menu. Currently, Liverbase encompasses 94 metabolic pathways and 14 functional categories. Upon request from submitters and users, we will include additional functional categories in this pull-down menu.

2.3. Annotation Page. In Liverbase, detailed annotation for each protein is shown (Figure 2) which contains the Swiss-Prot name, subcellular localization, mRNA/protein abundance, KEGG pathway, E.C. number, OMIM, and other basic annotation information (e.g., Mw, pI and Hp).
2.3.1. Abundance. Many liver researchers are interested in protein abundance and the protein’s specific subcellular localization(s). It is impossible to measure the concentrations of all liver proteins using currently available techniques. Because Liverbase integrates the information on human liver transcriptome and proteome, we provide mRNA/protein abundance and abundance rank for all Liverbase proteins.

Protein abundance usually indicates the activity level of the corresponding function. Of the 25 extremely high-abundance proteins (Supporting Information Table S3), the two most abundant proteins are hemoglobin, seven are mitochondrial proteins, six are related to alcohol metabolism, three are related to antioxidation, and two are involved in glycolysis. The mean abundance of enzymes was higher than that of proteins with nonenzymatic functions, which is consistent with liver’s physiological role as a center of metabolism.

Even for two proteins with the same function, protein abundance is an indicator of protein activity. For instance, there are 938 (GO: 741) transporters in the HLP data set and listed in Liverbase. An overview of the possible transport pathways was integrated from the HLP list of identified transporter proteins in liver (Supporting Information Figure S1 and Table S4). Mitochondrial membrane ATP synthase (or Complex V) produces ATP from ADP in the presence of a proton gradient across the membrane that is generated by electron transport complexes of the respiratory chain. The high abundance of ATP5A1 (ATP synthase subunit alpha, mitochondrial precursor), which is one of the catalytic subunits of ATP synthase, is consistent with the active process of energy metabolism in liver cells. Further, the transporters in the protein secretory pathway had relatively higher abundance than other classes of proteins, which are compatible with one of the main functions in liver, that is, secretion of large amounts of hepatic proteins into blood.

2.3.2. Protein Domain. Protein domain information (Pfam)\textsuperscript{29,30} is extracted from the IPI (International Protein Index)\textsuperscript{31} database (http://www.ebi.ac.uk/IPI). Domain would be an indicator of protein function. Because the list of Liverbase proteins comes from the proteome and transcriptome profiles, it will include some new or poorly characterized proteins, for example, XP_371754, which is described as “predicted: hypothetical protein”. From its domain (PF01101; HMG14_17), the function could be deduced to be a protein of relatively low molecular weight, possibly a nonhistone component of chromatin.

Proteins containing some specific domains are remarkably enriched in liver, while other domains are notably lacking. Fisher’s exact test was used to measure the degree of enrichment of proteins containing certain domains in human liver. Comparative analysis of every enriched or depleted domain in HLP was carried out. Comparison to the HGEP (Human Genome Encoding Proteome) showed that 188 significantly enriched and 47 significantly deleted Pfam domains exist in the HLP ($P < 0.01$). Domains were categorized according to involvement in several cellular processes based on the detailed description of every Pfam domain in the Pfam database and other related databases or references.\textsuperscript{29,32–54} The result shows that most of the highly enriched domains in the HLP are primarily involved in liver-specific physiological processes such as metabolism, transport, and biosynthesis. In contrast, the relatively depleted domains primarily participate in sense receptor-mediated signaling, neural processing, and cellular immunity.

2.3.3. Liver GO Slim. Gene Ontology (GO) is an accepted standard for ontological annotation of gene products (www.GeneOntology.org) and can provide a simplified summary of the function of each gene. GO slims are created by users according to their needs and may be specific to species or to particular areas of the ontology. In this study, liver GO slim was first established. The categories were chosen to provide a broad representation of the distribution of liver proteins with three major categories (biological process, cellular component, and molecular function) and intended to be nonoverlapping. In liver’s GO slim selection, four principles were carried out: (1) GO terms were inferred by IPI from the GOfact;\textsuperscript{55} (2) the housekeeping functional protein categories were to be reserved; (3) because the liver is a metabolic and energy center, items related to biosynthesis, metabolism, and transport were to be retained; (4) items specific to other organs (e.g., contractile fiber, muscle myosin) or which do not exist in the human (e.g., chloroplasts, melanosomes) were to be deleted. On the basis of these criteria, 375 GO terms and 51 nonredundant items were established as liver GO slim (Supporting Information Table S5). The Liver GO slim is available from the GOfact Web site at http://61.50.138.118/gofact (see: select the function category/GO slim-liver-v1), which is linked to Liverbase at “Liver GO slim”.

Liver GO slims could provide a useful way of finding all liver proteins with a common function. For example, LIPC_HUMAN (hepatic triacylglycerol lipase) is annotated with the GO process term for alcohol metabolism, GO: 0006066. Clicking that GO term navigates the user to a page that both provides the definition of that GO term and lists all other gene products (153) within Liverbase that have been annotated with that GO term.

As shown in Supporting Information Table S5, categories which have the larger numbers of identified proteins were consistent with liver functions, that is, the proteins with binding activity or catalytic capacity in the molecular function category; the proteins involved in metabolism, transport, and signal transduction in the biological process category; and the proteins localized in the nucleus, plasma membrane, mito-
chondrion, and cytoskeleton in the category of cellular components.

By comparing HLP data and the HGEP with the hypergeometric model,36–39 the degree of protein enrichment or deple-

Figure 2. (A) Examples for protein entry in Liverbase. As illustrated for some proteins, for each protein entry, Liverbase provides the UniProt entry name and description, the subcellular localization, mRNA/protein abundance, RIPpro, KEGG pathway, and so forth. (B) Graphic representation of the KEGG pathway. Each KEGG pathway was manually annotated based on protein identification and abundance information, in which proteins were identified from the HLP data set (blue) and/or the HLT data set (brown) with protein and/or transcript abundance. Proteins identified by different HLP databases were denoted here by four degrees of blue color. From deep to light blue, the colors represent 99P2 (proteins with two or more peptide matches with 99% confidence), 95P2 (proteins with two or more peptide matches with 95% confidence), 99P1 (proteins with one peptide match with 99% confidence), and 95P1 (proteins with one peptide match with 95% confidence). Proteins identified by different HLT databases were denoted by four degrees of brown color. From deep to light brown, the colors indicate (i) proteins covered by HLT core database (MPSS ∩ CHIP); (ii) proteins covered by MPSS or CHIP data but excluded from the HLT core database, or (iii) that the data were simultaneously supported by EST and SAGE in the public adult normal liver database; and (iv) proteins identified by either EST or SAGE in the public adult normal liver database.
tion could be statistically estimated (Supporting Information Table S5). Some classes of proteins, such as those involved in binding and catalytic activity, metabolism, transport, and coagulation, were highly enriched; these represent the major functions of human liver. However, other classes, such as those involved in signal transduction, signal transducer activity, transcription regulator activity, and plasma membrane, were depleted, which implies that these proteins might be tissue- or organ-specific and/or temporally regulated.

2.3.4. KEGG Pathway. Liver is a metabolically active organ responsible for many vital life functions, so most proteins listed in Liverbase are involved in metabolism. Thus, the KEGG pathway category and corresponding E.C. number were annotated for Liverbase proteins. On the Liverbase annotation page, clicking the KEGG path (hsaxxxxx) navigates the user to a page that provides both a graphic representation of the KEGG pathway in humans and an abundance annotation diagram for each protein within the graph. Through this database and external links page, users can easily view all proteins involved in liver special metabolism and functional categories.

The coverage of each pathway is shown in Supporting Information Table S6. In summary, all the 94 metabolism pathways are identified in different coverage. For 24 pathways, especially those for metabolism of carbohydrates, lipids, and amino acids, all pathway participants have been detected (Supporting Information Table S7). The liver-specific pathways, such as the pathways for metabolism of bile acid and bilirubin and biotransformation, are identified with higher coverage. As proteins involved in biotransformation, 31 cytochrome P450 proteins were identified and four of these were first observed in human liver.

Carbohydrate metabolism in liver plays a key role in maintaining blood glucose concentrations in the human body. Fourteen carbohydrate metabolic pathways were verified by HLP with high coverage of the relevant genes up to 93% (235/253). The enzymes participating in five pathways, including glycolysis/gluconeogenesis, tricarboxylic acid cycle, glyoxylate and dicarboxylate metabolism, propanoate metabolism, and inositol metabolism, were fully detected. More than 80% of the enzymes were found in eight other pathways. For inositol phosphate metabolism, member coverage was the lowest of the 14 pathways, but also reached 62% coverage.

There is high level of lipid metabolism in the liver. A total of 12 lipid metabolic pathways were identified in the HLP data set. Three fundamental pathways of fatty acid metabolism, that is, fatty acid biosynthesis, fatty acid elongation in mitochondria, and fatty acid metabolism, were completely covered. Among nine pathways with incomplete coverage, the enzymes in five pathways for the catalytic reactions of hydrophobic compounds, such as phospholipids, steroid hormones, and arachidonic acid, were detected at low coverage.

Liver is the major site in the body for synthesis of the nonessential amino acids, amino acid remodeling, and conversion of nonamino acid carbon skeletons into amino acids and other derivatives that contain nitrogen. Upon analysis of the HLP data set, 14 pathways related to amino acid metabolism were recognized with an average coverage of 86%. All the pathways in this metabolic area had relatively high detection rates (70.8–100%), compared with other classes of metabolic pathways.

The liver is capable of generating ATP by utilization of sugars, fats, and other chemical fuels. Unexpectedly, oxidative phosphorylation and ATP synthesis, two major pathways involved in energy metabolism, were identified by the HLP data set with relatively low coverage (64.8% and 65.5%, respectively), compared to all major metabolic areas of liver. Considering that most enzymes in the two pathways are tightly associated with the mitochondrial membrane, it was plausible that their hydrophobicity might lead to such results. The high coverage of the HLT data set for the pathways (82.4% and 89.7%, respectively) provided additional support for this hypothesis.

The functions of biological transformation are primarily to decrease or eliminate the toxicity of toxins by improving the polarity of non-nutritional materials and facilitating their elimination from the body. Of 71 enzymes in 15 xenobiotic biodegradation and metabolism pathways, 61 enzymes were recognized in the HLP data set and eight pathways were fully covered. The high detection rate of 83.2% in those pathways could be interpreted as evidence that the liver possesses powerful function in biotransformation.

The human cytochrome P450 (CYP) superfamily is divided into 18 families and 43 subfamilies with 60 enzymes, which can oxidize xenobiotics and endogenous lipids. The latter include cholesterol, vitamin D, steroids, retinoids, eicosanoids, and fatty acids. According to a previous estimate, 39 CYPs are expressed in human liver (40% (Supporting Information Table S8) (www.expasy.org: www.ncbi.nlm.nih.gov). In Liverbase, CYPs, including CYP2, CYP3, and CYP4, were remarkably enriched. All three families of enzymes participate primarily in metabolism of xenobiotics, such as arachidonic acid, drugs, and toxins; this is consistent with their important role in non-nutritional materials. Among 31 human CYP enzymes listed in Liverbase, four had never been reported to be expressed in human liver; these were CYP4B1, CYP11B2, CYP19A1, and CYP24A1. The two new liver CYPs, CYP19A1 and CYP24A1, were confirmed by transcriptomic determinations. CYP19A1 was reported to be widely distributed in extragonadal tissues and involved in estrogen biosynthesis (41–43) and CYP24A1 (25-hydroxyvitamin D-24-hydroxylase) was reported to play a role in maintaining calcium homeostasis and in initiating degradation of [1a, 25(OH) 2D3] for prevention of vitamin D toxicity.

Signal transduction is another important component of KEGG. We analyzed the coverage of all 16 signal transduction pathways in the HLP data set (Supporting Information Table S9) using the strategy of node analysis. Every pathway was divided into nodes based on signal transduction figures from KEGG. Signals from extracellular to nuclear were conveyed by these nodes, which usually localized differently in the cell. One node can include several protein members, some belonging to one family and some exhibiting the same functions. The integrity of every pathway in the HLP database was then represented by the coverage of nodes, not by coverage of proteins. For example, in the MAPK signaling pathway, the RAS node included six proteins according to KEGG; as long as one of the proteins is identified in the HLP database, this node will be considered to be covered. With the strategy of node analysis, the typical cascade and pivotal pathway, MAPK, was demonstrated to have the highest number (125) of nodes; whereas, hedgehog pathway had the fewest, just 18 nodes. Further, the node coverage of all 16 pathways was calculated and eight were covered up to 50% in the HLP data set. Adherents junction (80%, 59/73), focal adhesion (78%, 47/60), and calcium signaling pathways (72%, 28/39) are required for cellular recognition, localization, communication, and inflammation and showed some of the highest detection rates. The pathways participating in metabolic regulation were also highlighted by the HLP data.
set; these include the insulin (66%, 41/62) and adipocytokine (59%, 22/37) pathways. The pathways related to liver disease had low levels of coverage, as expected; these include the TGF-β, wnt, and notch pathways (Supporting Information Table S9). An overview of the signal transduction network demonstrated that its two global highlights were proliferation and apoptosis, that is, the two fundamental states of life or death, represented by the Yin-Yang format (Supporting Information Figure S2). The two zones are different from but not exclusive to each other, distinct from but not unswitchable between each other.

Some proteins involved in the KEGG pathway are not identified or listed in Liverbase (Supporting Information Table S10); there are three possible explanations for this phenomenon: (1) the protein is not expressed in the normal human liver, which is the designated sample for the whole project; (2) the abundance of the protein is too low to be detected by the currently used technical platform; (3) the physical and the chemical properties of the protein are out of the detection range of the instruments used in the project. For example, only 23% of the 13 mtDNA encoding proteins were identified in our core data set. One possible explanation for this is the high hydrophobicity of these proteins.

2.3.5. Disease-Related Proteins. For genes implicated in a hereditary disease, Liverbase provides a link to the corresponding entry in the Online Mendelian Inheritance in Man database (OMIM; http://www.ncbi.nlm.nih.gov).55,56 To date, more than 19380 human proteins are known to be involved in a hereditary disease. A small independent data set of Liverbase, LDGP (Liver Disease-related Genes and Proteins), which is composed of 240 nonredundant liver disease-related genes and proteins, was set up from literature mining. In data collection, we first sorted liver diseases according to the textbook Hepatology57 and confirmed the keywords for each type of disease by searching MeSH terms on PubMed. The most common keywords included hepatitis, cirrhosis, liver tumor, infectious liver disease, vascular disease, inherited metabolic diseases, fatty liver, alcoholic liver diseases, and other peculiar liver diseases. Second, we set the table list, which included all the information we needed. The table list included types of disease, liver disease-associated genes, loci, liver disease-associated proteins, Unipro ID, score, and so forth. We then used the OMIM database55,56 and Genecards database58,59 as the sources of liver disease-associated genes and proteins. As for the data from OMIM, we read the search results carefully before accepting or rejecting the results. As for the data from Genecards, we read the search results with a score more greater than 30 and carefully considered whether to accept or reject. Finally, we confirmed that 240 genes and corresponding proteins were liver disease-associated genes and proteins. In LDGP, 60% were included in the Liverbase transcriptome data set and 43% were included in the Liverbase proteome data set. These data suggest that a majority of liver disease-associated genes are expressed in normal liver. Diseases occur with mutation, which result in changes in the quantity or structure of proteins. The data also suggest that the transcriptome and proteome data sets may be used as reference data to discover biomarkers of liver disease.

2.3.6. External Links Page. Liverbase also contains an external links page with links to other biological databases and homepages; links are currently limited to GO, KEGG, pfam, and Swiss-Prot. Links to GNF databases (http://symatlas.gnf.org/SymAtlas/)60,61 are included in this page as well. The links page provides users with more comprehensive information. For example, as a gene atlas of protein-encoding transcriptomes from 79 human and 61 mouse tissues, GNF databases provide the expression patterns for thousands of predicted genes among different tissues. The tissue-specific patterns of mRNA/protein expression, which could not be deduced from data from a single transcriptome or proteome, can indicate important clues about gene function for known and poorly characterized genes.

3. Discussion

A comprehensive reference map of the human liver proteome is reported for the first time in this resource. With the integrated strategy, 6788 proteins were identified with confidence and exquisitely composed into a database, Liverbase, which represents the largest data set of human liver proteome to date. The database is a curated, open-source, Web-accessible resource for functional analysis of human liver. The primary focus of Liverbase is to provide a resource that facilitates functional analysis in human liver.

The Human Liver Proteome Project is the first initiative of the human proteome project for human organs/tissues. Its global scientific objectives are to reveal the “solar system” of the human liver proteome, including expression profiles, modification profiles, a protein linkage (protein–protein interaction) map, and a proteome localization map. Further research in human liver organelle expression profiles and differential expression profiles of liver disease are also in progress. As a next step, we will continue to integrate this information into Liverbase. Our long-term goal is to serve the needs of the liver community by providing more comprehensive functional annotation for genes implicated in liver development and disease.

Similar to all databases, Liverbase is an ongoing project and interaction with the user community is vital for its development and refinement. We encourage the submission of data, correction of errors, and suggestions for making Liverbase of greater use.

4. Conclusions and Perspective

In summary, we have developed a protein database for human liver. User-friendly Web interfaces were designed to easily query the database and access the various features. To our knowledge, this is the only liver protein database that incorporates mRNA/protein abundance information.

Although a large amount of important data is published each year, occasionally, some of this information is lost in the mass of literature available to researchers. As a professional liver database, in order to have a large impact in this area of biology, it is important that Liverbase includes more comprehensively reviewed and correctly summarized annotation for each protein. We are continuing to write mini-reviews for the most important proteins, which should include function, citations to the experimental literature, and whether or not the protein is related to disease. As a manual annotation, ‘minireviews’ of proteins are provided by highly trained curators reading published literature, evaluating the available experimental evidence, associating the appropriate function to the protein, and providing a detailed, information-rich summary of the knowledge about a liver protein. Also, scientists are encouraged to review the curators’ annotation progress, to comment on the annotations made, and to suggest published papers and
information to improve the protein annotations. Scientists could supply the curators with key experimental publications or other detailed information, and they are welcome to point out any experimental data that might be missing, wrong, or controversial.

Supporting Information Available: Supplementary figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgment. This work was partially supported by Chinese State Key Projects for Basic Research (973) (nos. 2006CB910401, 2006CB910801, and 2006CB910600), Chinese State High-tech Program (863) (2006AA02A308), National Natural Science Foundation of China (30700988, 30700356), National Natural Science Foundation of China for Creative Research Groups (30621063), and Chinese State Key Project Specialized for Infectious Diseases (2008ZX10002-016, 2009ZX10004-103, 2009ZX9301-002).

References
Liverbase: A Comprehensive View of Human Liver Biology

research articles
research articles


PR900191P