

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/26715947>

First insight into the human liver proteome from PROTEOME(SKY)-LIVER(Hu) 1.0, a publicly available database

ARTICLE in JOURNAL OF PROTEOME RESEARCH · SEPTEMBER 2009

Impact Factor: 4.25 · DOI: 10.1021/pr900532r · Source: PubMed

CITATIONS

29

READS

51

90 AUTHORS, INCLUDING:



Ming Chen

Thomas Jefferson University

26 PUBLICATIONS 214 CITATIONS

[SEE PROFILE](#)



Yunping Zhu

99 PUBLICATIONS 1,721 CITATIONS

[SEE PROFILE](#)



Huali Shen

Fudan University

35 PUBLICATIONS 330 CITATIONS

[SEE PROFILE](#)



Quanhu Sheng

Vanderbilt University

63 PUBLICATIONS 1,407 CITATIONS

[SEE PROFILE](#)

First Insight into the Human Liver Proteome from **PROTEOME^{SKY}-LIVER^{Hu} 1.0**, a Publicly Available Database

Chinese Human Liver Proteome Profiling Consortium*,†,‡,§,¶

State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, 33 Life Science Park Road, Changping District, Beijing 102206, P. R. China, Institutes of Biomedical Sciences, 130 DongAn Road, and Department of Chemistry, 220 Handan Road, Fudan University, Shanghai 200032, P. R. China, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing Airport Industrial Zone B-6, Beijing, 101300, P. R. China, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, 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, Eastern Hepatobiliary Surgery Hospital/Institute, Second Military Medical University, 225 Shanghai Road, Shanghai 200438, P. R. China, Shanghai-Ministry Key Laboratory of Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, 351 Guo Shou-Jing Road, Shanghai 201203, P. R. China, College of Life Sciences, Hunan Normal University, Lushan Road 14, Changsha, Hunan Province 410081, P. R. China, Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, P. R. China, Shanghai Center for Bioinformatics Technology, 100 Qinzhou Road, Shanghai 200235, P. R. China, First Affiliated Hospital of Harbin, Medical University, Heilongjiang Province 150001, P. R. China, and Ministry of Science and Technology, 15B, Fuxing Road, Beijing 100862, P. R. China

Received March 10, 2009

Herein, we report proteome and transcriptome profiles of the human adult liver and present an initial analysis. Overall, the human liver proteome (HLP) data set comprises 6788 identified proteins with at least two peptides matches at 95% confidence, including 3721 proteins newly identified in liver. The human liver transcriptome (HLT) data set consists of 11 205 expressed genes. The HLP is the largest proteome data set for a human organ and is the first direct association between a proteome and its transcriptome derived from the same sample. Although it is hard to approach complete coverage of the HLP currently, several conclusions based on this data set are clearly reached: (1) The 5816 protein-encoding genes (PEGs) represented by the HLP and the 11 104 PEGs represented in the HLT have been identified from 20 070 PEGs in IPI Human v3.07 and 19 478 PEGs in the integrated human transcriptome database, respectively. (2) The patterns of chromosomal distribution of the genes corresponding to the HLP are highly consistent with those of the HLT. Some chromosomal regions, such as 16p13.3, 19q13.31, 19q13.42, and Xq28, exhibit particularly high densities of liver-specific genes, which perform the important functions related to normal physiology or/and pathology in this organ. (3) The HLP spans 6 orders of magnitude in relative protein abundance and 78% of the proteins fall in the middle of this range. Of newly identified liver proteins, 82.5% are of low abundance. (4) Proteins involving in metabolism, transport, and coagulation and those containing active domains for metabolism, transport, and biosynthesis are significantly enriched in liver. (5) All 94 metabolic pathways in KEGG are touched to different extent. Of which, for 48 pathways, particularly those involved in metabolism of carbohydrates and amino acids, more than 80% of the component proteins have been detected. The liver-specific pathways, such as those participating in metabolism of bile acid and bilirubin and in biotransformation, are identified with remarkably high coverage. A total of 31 members of the cytochrome P450 family are identified, four of which have been observed for the first time in human liver. (6) Transport proteins involved in energy metabolism and secretion of both protein and bile acid are highly abundant. Three ion channels are described for the first time in liver. (7) The 800 proteins related to signal transduction and primarily involved in cellular recognition, localization, communication, and inflammation are present in the HLP data set. Insulin and adipocytokine pathways, which are involved in the regulation of glucose and fatty acids, are highly covered. (8) Transcription factors (309 in total) have been recognized at relatively low detection rates and abundance; however, transcription factors regulating gene expression related to transport, metabolism, and biosynthesis are detected at relatively higher coverage and the protein products of their target genes (100 in total), such as metabolic enzymes and plasma proteins, are also identified. (9) The overlap between the human liver and plasma proteomes is particularly noteworthy in the coagulation/anticoagulation/fibrinolysis and complement system. There is a significantly positive linear correlation between the abundance of coagulator proteins in liver and plasma.

Keywords: Chinese Human Liver Proteome Project • protein expression profile • transcriptome • plasma

Introduction

The completion of the human genome has sparked new interest in the study of human transcriptomes and proteomes. The proteome contains the complete complement of proteins encoded by the genome and is responsible for many of the important functions of an organism; further exploration of the proteome will be a primary focus of the postgenomic era. 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 in health, the variety, number, and abundance of liver proteins have not been extensively characterized. Because the liver is such a complex biological system, global analysis at the “omics” level is necessary to fully elucidate its functions. The Chinese Human Liver Proteome Project (CNHLPP) is one of the large-scale proteomic initiatives coordinated by the Chinese Human Proteome Organization and aims to describe the proteomic atlas of liver development, physiology, and pathology.^{1–3} The program of the CNHLPP was initiated in 2004 by the Ministry of Science and Technology of China, which was cofounded by the State Key Program for Basic Research and the State Key Program for High Technology. The objectives of the expression profiling pilot subproject are to (1) establish comprehensive and complementary platforms and develop new methods and technologies, with a particular focus on identifying low-abundance proteins; (2) construct the primary proteome profile of adult human liver; and (3) integrate and compare the proteome with its transcriptome and with the human plasma proteome. Eleven laboratories are participating in this subproject. To globally catalog the human liver proteome, a strategy was developed to deal effectively with its complexity. First, fresh samples of adult liver were carefully

* Correspondence and requests for materials should be addressed to Fuchu He (hefc@nic.bmi.ac.cn), Xiaohong Qian (qianxh@nic.bmi.ac.cn), Pengyuan Yang (pyyang@fudan.edu.cn), Siqi Liu (siqliu@genomics.org.cn), or Yunping Zhu (zhuyping@gmail.com).

[†] The Chinese Human Liver Proteome Profiling Consortium partial list of author and affiliations. State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine: Ying Jiang, Wantao Ying, Songfeng Wu, Ming Chen, Wei Guan, Dong Yang, Yanping Song, Xin Liu, Jianqi Li, Yunwei Hao, Aihua Sun, Chao Geng, Hao Li, Wei Mi, Yangjun Zhang, Jiyang Zhang, Xilin Chen, Lei Li, Yan Gong, Tao Li, Jie Ma, Dong Li, Xiaoyan Yuan, Xuequn Zhang, Xiaofang Xue, Yunping Zhu, Xiaohong Qian, Fuchu He. Institutes of Biomedical Sciences and Department of Chemistry, Fudan University: Fan Zhong, Huali Shen, Chengzhao Lin, Haojie Lu, Xiaohui Liu, Liming Wei, Jing Cao, Dong Yun, Jie Zhang, Mingxia Gao, Huizhi Fan, Yang Zhang, Gang Cheng, Yanyan Yu, Lili Xie, Hong Wang, Xiangmin Zhang, Fuchu He, Pengyuan Yang. Beijing Genomics Institute, Chinese Academy of Sciences: Liang Shi, Wei Tong, Xiaolei Li, Yuan Wang, Siqi Liu. Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences: Quanhui Sheng, Rong Zeng, Cancer Institute & Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College: Yulin Sun, Yang Xu, Jianqiang Cai, Ping He, Hongjun Gao, Xiaohang Zhao, Eastern Hepatobiliary Surgery Hospital/Institute, Second Military Medical University: Yexiong Tan, Hexin Yan, Yuan Yang, Hongyang Wang. Chinese National Human Genome Center at Shanghai: Jian Huang, Zeguang Han, College of Life Sciences, Hunan Normal University: Quanyuan He, Ping Chen, Songping Liang. College of Life Sciences, Peking University: Min Zhao, Xizeng Mao, Liping Wei. Shanghai Center for Bioinformatics Technology: Hong Yu, Zhiwei Cao, Yixue Li. First Affiliated Hospital of Harbin Medical University: Wenjie Dai, Hongchi Jiang. Scientific management: Beijing Proteome Research Center, Beijing Institute of Radiation Medicine: Donggen Wang, Junjie Zheng, Xue Gao, Ying Tang. Ministry of Science and Technology, P.R.C: Xueyong Li, Jinpei Cheng, Yanhua Liu, Xianen Zhang, Xiaofang Wang, Jingdun Jia, Daochang An, Zhen Wang, Qing Li, Tuo Cui.

[#] The following authors contributed equally: Ying Jiang, Wantao Ying, Songfeng Wu, Ming Chen, Fan Zhong, Huali Shen, and Chengzhao Lin.

⁺ Principal investigators of participating laboratories: Yunping Zhu, Xiaohong Qian, Fuchu He, Xiangmin Zhang, Pengyuan Yang, Siqi Liu, Rong Zeng, Xiaohang Zhao, Hongyang Wang, Zeguang Han, Songping Liang, Liping Wei, Yixue Li, Hongchi Jiang.

Chinese Human Liver Proteome Profiling Consortium

collected in strict accordance with the ethical rules of tissue collection and standard operating procedures (SOPs). In February 2005, livers of 10 healthy Chinese volunteers were prepared and distributed as reference specimens to the four main proteomics research centers and one genomics research center in China, for parallel analysis of proteome and transcriptome. Second, due to the size and complexity of the proteome and the lack of an established system for its comprehensive analysis, an integrated platform was established. Four major proteomic technologies were employed for profiling the liver proteome: 2DLC-ESI, 3DLC-ESI, 2DE-MALDI, and 1DE-LC-ESI. Finally, a powerful bioinformatics system was constructed to process the large amount of data generated. In November 2005, all data were submitted to the Beijing Proteome Research Center and, in May and June 2006, the data analysis jamboree involving the 11 participating laboratories took place in Beijing.

In this overview, the major findings from the pilot phase of the CNHLPP Expression Profiling Subproject will be summarized. Detailed descriptions of sample collection, proteomic technical strategies, data set infrastructure, data integration and annotation, and additional associated studies will be presented in this Special Issue of the “Chinese Human Liver Proteome Project”.

Materials and Methods

Sample Collection, Handling, and Distribution. The sample collection procedure followed the regulations of the ethics committee of the Chinese Human Liver Proteome Project. Human liver samples were collected from volunteers of the Chinese Han nationality, who had undergone hepatic hemangioma resection (neoplastic lesions were less than 5 cm in diameter). Volunteers were selected from a large pool of patients. Qualified donors were free from viral hepatitis, autoimmune liver disease, cystic liver disease, fatty liver disease, and other forms of liver disease. Before surgery, the donors gave informed consent and received systematic examinations, including biochemical (alanine aminotransferase, aspartate aminotransferase, g-glutamyl transferase, alkaline phosphatase, bilirubin, albumin, prothrombin time, and alpha-fetoprotein) and liver imaging (B-type ultrasound) examinations. Personal and clinical information on liver donors is shown in Supplementary Tables 1 and 2. Surgical samples were processed for proteomic analysis only if the volunteer’s diagnostic data was within normal range. Approximately 1 g of adult human liver tissue located 2 cm from the tumor was collected and divided into two parts: one part was used for histopathological analysis and the other for proteomic and transcriptomic analysis.

Fresh liver samples were immediately rinsed thoroughly with ice-cold PBS and preserved in liquid nitrogen until further transcriptomic and/or proteomic analysis. On the basis of earlier estimates of individual variation in proteomes, 10 individual samples were eventually used for this project.⁴ For transcriptomic analysis, RNA was extracted from each individual sample and equal amounts of these samples were pooled. For proteomic analysis, frozen samples of tissue from individual donor was separately ground into fine powder in liquid nitrogen and equal amounts of these samples were pooled; the pooled powders were distributed among the laboratories conducting proteomic analyses (Supporting Information).

RNA Extraction, Microarray Hybridization, and Massively Parallel Signature Sequencing (MPSS) Analysis. Total RNAs were extracted by TRIzol reagent and DNA contaminants were

removed by digestion with RNase-free DNase I. After determination of concentration and quality, equal amounts of RNA from the 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 MPSS⁵ (TaKaRa, Japan). Only those transcripts that were 'present' in two chips with a fluorescence intensity greater than 100 were considered. 'Nonsignificant' signatures, that is, those never observed at levels greater than 3 TPM (transcripts per million signatures) in any library, were discarded.⁶

Protein and Peptide Separation and Identification. Four technical approaches for proteome analysis were designed: (1) 2DLC-ESI, strong cation exchange (SCX) and reverse-phase liquid chromatography (RPLC) for separation of digested peptides coupled with electrospray ionization (ESI)-tandem mass spectrometry (MS/MS); (2) 3DLC-ESI, LC for protein prefractionation followed by SCX and RPLC for separation of digested peptides combined with ESI-MS/MS; (3) 2DE-MALDI, two-dimensional gel electrophoresis (2DE) for protein separation linked to matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry (MALDI-TOF/TOF-MS); (4) 1DE-LC-ESI, SDS-PAGE for protein prefractionation and RPLC for separation of digested peptides followed by ESI-MS/MS^{7–9} (Supplementary Table 3). Each approach was implemented at two centers. All determinations were run in parallel at least three times at each center to reduce the random aspect of peptide capture and to enhance the possibility of capturing low-abundance peptides.^{10,11}

Data Processing and Integration. A data processing and integration flowchart is shown in Supplementary Figure 1. According to the SOPs (Supplementary Tables 4–7) all qualified peak lists were accepted for protein search against both IPI Human v3.07 database and its reversed version¹² (Supplementary Methods). To evaluate the likelihood of obtaining false positives, an identification threshold of 95% confidence was established for each technical approach (Supplementary Table 8). All peptides identified using the four approaches were combined, and the unique peptides were mapped back to IPI Human v3.07 database to identify the corresponding proteins/groups. The protein groups were integrated by set operation, resulting in the minimized identifications from the whole set of peptides.¹³ The representative protein of a given group in the final list was defined based on database source and sequence length.^{8,14}

Data Quality Control. The SOPs for data quality control were established in concordance with the consensus in the literature and recommendations from the HUPO community¹⁵ (Supplementary Methods). According to the SOPs, a qualified protein must satisfy the following criteria: (1) all peptide data should meet a 95% confidence cutoff; (2) false-positive matching of peptides should be eliminated after searching of the reversed IPI database;^{16–18} (3) for shotgun sequencing, the qualified peptide should have an MS/MS sequence of more than six amino acids. All identified proteins should have two or more peptides matches;^{19–21} (4) for peptide mass fingerprinting (PMF), a protein should have at least five peptide matches with 20 ppm mass tolerance in database searching.²²

Storage and Presentation of the Reference Expression Profile. In this project, taking into consideration the special features of proteomic data and the requirements of the biologists, we developed new databases for liver proteomic expression profiles, PROTEOME^{SKY}-LIVER^{Hu} 1.0 Human Liver Expression Profile (dbLEP, <http://dbelep.hupo.org.cn>) and the

Liverbase (<http://liverbase.hupo.org.cn>). Both databases were designed to contain multiple data sets and to be traceable with multiple-access and detail annotation.

Protein Semiquantitation. Semiquantitation of protein abundance was carried out using the spectral counts (SC) method.^{17,23,24} The Spectral Count Index (SCI), a normalized SC used in estimating the theoretical number of tryptic peptides of a protein, was calculated according to previously reported methods.^{25,26} To integrate large-scale data from multiple sources, SCI values from different batches were normalized and the normalized SCI value was termed the Spectral Count Index Normalized (SCIN) (Supplementary Methods).

The Hypergeometric Distribution Model. The hypergeometric distribution model^{27,28} (<http://www.r-project.org/>) was used to compare the characteristics of the HLP and HLT against the Human Genome Encoding Proteome (HGEP), including domain annotation from Pfam,^{29,30} Gene Ontology (GO)³¹ function annotation with GOfact,³² and specific biological processes from transporter annotated by GO. When the value of standard hypergeometric probability (*P*) for a given protein, protein family, functional category, or domain was no more than 0.05, the analyzed targets were regarded as significantly enriched or depleted in the HLP or HLT compared to the HGEP.

Node Analysis Strategy. Pathways were divided into nodes based on signal transduction data from KEGG. Signals from extracellular to nuclear were conveyed by these nodes, which usually localized to different subcellular sites. A single node can include several protein members based on structural and functional parameters, with some belonging to the same protein family and others sharing the same function. Consequently, the integrity of each pathway in the HLP database is represented by the coverage of nodes, not by coverage of proteins. For example, in the MAPK signaling pathway, the RAS node includes six proteins according to KEGG;³³ as long as one of these is identified in the HLP database, this node will be considered to be covered.

Results

Data Sets and Features. The Human Liver Transcriptome (HLT). To compare the global expression patterns at the level of translation and transcription, the human liver transcriptome was profiled by massively parallel signature sequencing (MPSS) and Affymetrix high-density oligonucleotide arrays⁵ with the same samples used for proteomic analysis above. A total of 10 224 and 5422 genes were detected by microarray and MPSS, respectively. Following integration of the two data sets, a total of 11 205 genes were defined as the HLT core data set (Supplementary Table 9).

The Human Liver Proteome (HLP). Overall, 607 851 identifications corresponding to 62 117 peptides containing 45 781 peptides with tandem MS spectra (Supplementary Table 10) led to the identification of 23 345 proteins with a 95% confidence level (Supplementary Table 11). After removal of duplicates, the number of identified proteins was reduced to 12 951. By eliminating the proteins with only one peptide match, the number of identified proteins was further reduced to 6788. The 6788 proteins were taken as the core data set (proteins having two or more peptide matches with 95% confidence, HLP_95P2) and the 12 951 proteins were defined as the extended data set (HLP_95P1). Further analysis of the detection frequency for each protein showed that 82% of the proteins in HLP_95P2 data set were identified more than three times and 65% of the proteins in HLP_95P1 were identified more than twice. Besides

research articles

the normal 95% confidence data set, another higher confidence (99%) data set was also provided to confirm the biological conclusion. At the 99% confidence, 5454 unique proteins were identified (HLP_99P1), of which 3013 were identified with two or more unique peptides (HLP_99P2). (Supplementary Table 11).

Complementarity of Technical Approaches. Because of the limitations inherent in current proteomic technologies, a complementary strategy combining four analytical approaches was implemented in this project. The 2DLC-ESI³⁴ approach contributed the largest amount of data with 1391 uniquely identified proteins and a wide dynamic range with a SCIN of 0.001–0.1 (Figure 1a,b). The 3DLC-ESI³⁵ approach yielded 916 uniquely identified proteins and provided the largest number of identifications at the extremely low-abundance grade (SCIN of 0.001–0.01). The 2DE-MALDI approach was also fairly complementary to these two approaches and contributed 379 uniquely identified proteins. As an indication of the complementarity of the four techniques, 1112 proteins were identified by two or more different technical approaches, but each with only one peptide match.

Dynamic Range. Compared to the consistency of the genome, the variations in the levels of proteins and their RNA transcripts can be dramatic, and therefore, quantitative information is particularly critical to a proteomic survey. The algorithm based upon the near-linear relationship between SCIN values and relative protein molar abundance has been generally accepted for use in quantitative estimation of a proteome.²³ The SCIN values of 6641 proteins were calculated based on the mass spectra from seven runs of LC-based analysis (Supplementary Table 12). The relative abundance was separated into 6 orders of magnitude (Figure 1c) and exhibited a normal distribution, that is, most of the proteins occurred at the two medians.

The SCIN semiquantitation method was validated by results obtained by Western analysis (26 proteins), ELISA (6), and semiquantitative HPLC (1). The tested concentration ranges for these three methods were 0.1–10 fmol protein/mg liver sample for ELISA; 0.01–100 pmol/mg liver sample for Western blotting; and 0.1–1 nmol/mg liver sample for HPLC, respectively. The concentrations of five additional proteins determined by Western analysis and reported in references were also included (Supplementary Methods). The 30 data points were eventually used in a least-squares regression analysis between double logarithmic SCIN values and concentration (Figure 1d), with eight data points rejected statistically.

Specificity of the HLP and HLT Compared to the HGEP. A fraction of the genes in the human genome is expressed specifically in a given tissue or organ. The specificity of the HLP (5816 protein-encoding genes [PEGs] corresponding to 6788 proteins) and the HLT (11 104 PEGs corresponding to 11 205 genes) was shown qualitatively by the degree of overlap with the HGEP (represented by IPI Human v3.07 database with 20 070 PEGs corresponding to 50 207 proteins) or the transcriptome database (19 478 PEGs), respectively (Supplementary Methods, Supplementary Figure 2, and Supplementary Table 13). The extent of overlap of the HLP and HLT with the appropriate corresponding database was estimated to be 29.0% and 57.0%, respectively, indicating that liver indeed expresses only a subset of the human genome.

Newly Identified Proteins in Human Liver. Four reference proteome data sets, including the Integrated Liver tissue Proteome of mouse and human (ILP, 3011 proteins), the

Chinese Human Liver Proteome Profiling Consortium

Human normal Heart Proteome (HHP, 619 proteins), the Human Plasma Proteome (HPP, 3885 proteins), and the Liver Disease-related Genes and Proteins (LDGP, 228 proteins) were collected and constructed (Supplementary Methods) and used for comparison with the HLP data set. The overlap between HLP and ILP, HPP, and HHP was 28.6%, 23.1%, and 6.3%, respectively. Intriguingly, 46.1% of the LDGP data set was also identified in the HLP data set, implying that a significant number of liver disease-associated genes or proteins are expressed in normal liver. It was also noted that the medium- and low-abundance proteins occur more frequently in the HLP data set than in the other liver proteome data sets (ILP, Swiss-Prot, and Human Protein Reference Database liver proteins). In the HLP data set, 3721 proteins were newly identified (Supplementary Table 14), of which 82.5% were of relatively low abundance (Figure 1e), whereas 59.4% of the overlapping proteins in the ILP and HLP were of relatively high abundance (red line in Figure 1e). These proteins include 205 that are annotated by domain information, 360 without functional annotation, and 977 defined as hypothetical (Supplementary Figure 3). Twenty-five KEGG pathways, in which all the involved proteins were covered by the HLP, were listed in Supplementary Table 15. Recognition of four established pathways in liver was facilitated by knowledge of proteins detected for the first time in this study. Of the 8 highest newly identified node covered pathways, only the heparan sulfate biosynthesis pathway is known to be critical for liver. The remaining seven pathways appear to be outside the scope of normal liver function and are primarily related to pathological processes; almost all of those processes are related to the nervous system (Supplementary Table 16).

Chromosomal Distribution of the Genes Represented in the HLP and HLT. A total of 6331 genes represented in the HLP and 10 933 genes in the HLT were mapped to the corresponding sites on chromosomes with the relative abundance of the transcripts and proteins (Figure 2, Supplementary Table 17). On the basis of the chromosomal distribution of genes in HGEP and estimated by the chi-square test, the complete pattern of HLP gene distribution was compatible with that of HLT (9.5×10^{-13} and 2.2×10^{-16} , respectively). Some chromosomal regions exhibited significantly high gene density (Figure 2), such as Xq28 (loss of heterozygosity [LOH] in this region is crucial in tumor progression),³⁶ 16p13.3, 19q13.31 (associated with lipid level in plasma through LDL-C and ApoB),³⁷ and 19q13.42 (associated with human prostate cancer)³⁸ (Supplementary Table 18). The genomic interval of 16p13.3 is extremely gene-rich and contains at least 20 genes,³⁹ including the human disease genes *PKD1* (polycystic kidney disease-related),⁴⁰ *TSC2* (tuberous sclerosis-related),⁴¹ and *AXIN1* (Axis inhibition protein 1, hepatocellular carcinoma-related),⁴² and physiological regulators such as *ALR* (augmenter of liver regeneration)/*HPO* (hepatopoietin), which is responsible for the extraordinary regenerative capacity of mammalian liver.^{43,44} Some gene-enriched regions contain known gene clusters encoding histones, the cytochrome P450 superfamily, transferases, and enzymes involved in amino acid/derivative metabolism.^{45–48}

Functional Annotations. Characteristics of HLP from Liver GO Slim and Protein Domains. On the basis of the GO infrastructure, liver GO slim with 51 nonredundant items was established for the liver proteome functional overview (Supplementary Tables 19 and 20). In a comparison of the

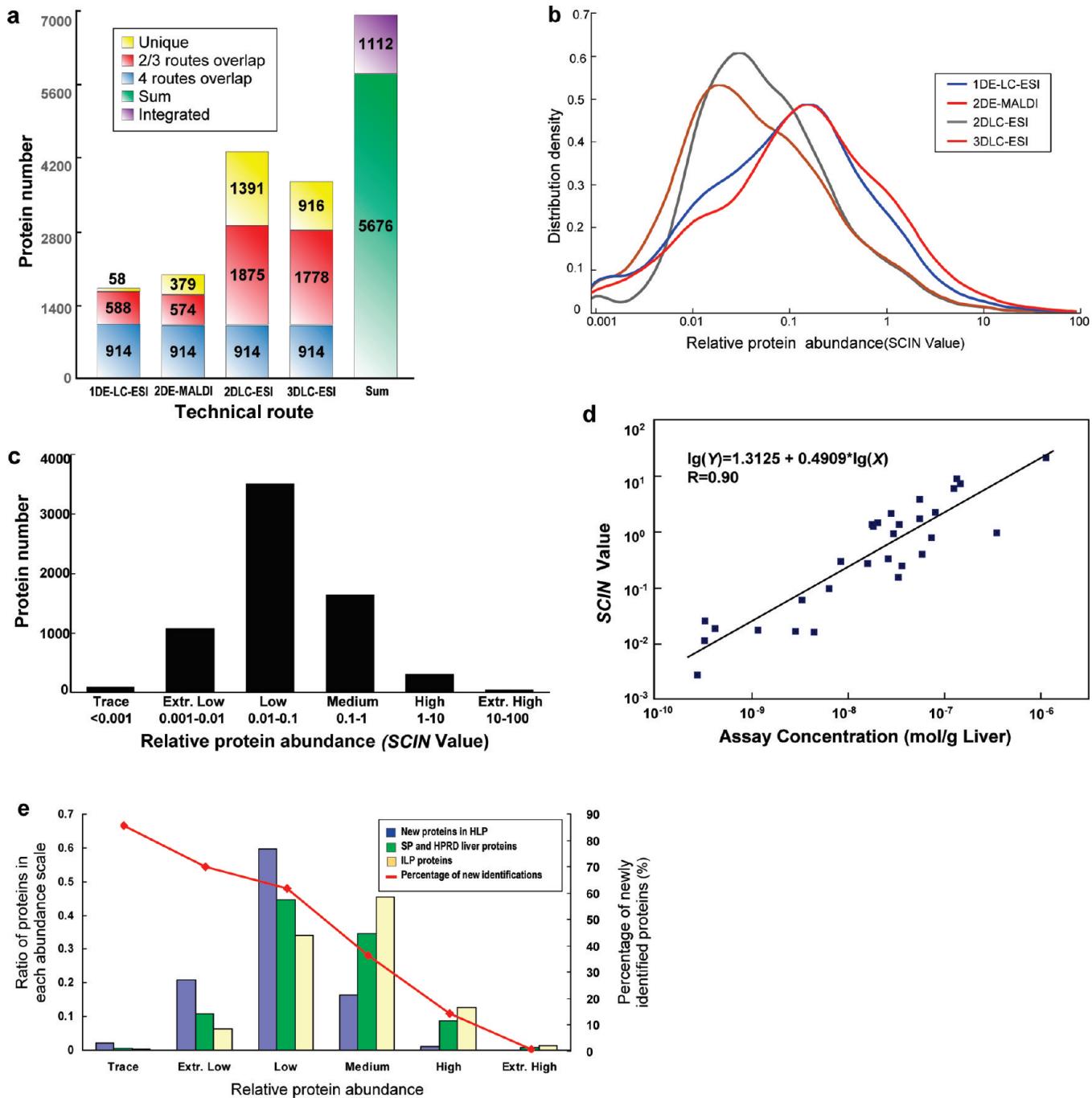


Figure 1. Complementary, comprehensive, and quantitative illustrations of the HLP data set. (a) Complementarity of four technical approaches: 1DE-LC-ESI, 2DE-MALDI, 2DLC-ESI, and 3DLC-ESI. The level of complementarity is indicated by the pattern of protein identification: Unique, proteins identified only by one approach; 2/3 routes overlap, identified by two or three approaches; 4 routes overlap, identified by all four approaches; Sum, total number of identified nonredundant proteins; Integrated, proteins identified by two or more different technical approaches but each with only one peptide match. (b) Density distribution of protein abundance for four technical approaches. (c) Abundance distribution of identified proteins. (d) Quantitative illustrations of the HLP data set. Linear correlation between logarithmic SCIN value and logarithmic protein concentration, validated by ELISA (0.1–10 fmol protein/mg), Western blotting (0.01–100 pmol protein/mg), and HPLC (0.1–1 nmol protein/mg). (e) Discovery of new liver proteins. SP and HPRD liver proteins, proteins annotated as expressed in liver by Swiss-Prot and the HPRD database. ILP, the integrated liver proteome data set from this study (Supplementary Methods). New Proteins in the HLP, proteins in the HLP data set identified as new liver proteins following comparison with ILP, HPRD, and SP liver protein data sets.

HLP with the HGEP based on the hypergeometric model (Supplementary Methods), the degree of enrichment or depletion of proteins was statistically estimated. The classes related to metabolism (carbohydrates, lipids, amino acids

and derivatives, and proteins), transport, and coagulation were significantly enriched. Conversely, the classes of protein related to signal transduction, RNA metabolism, and the category “responsive to stimulus” were significantly depleted

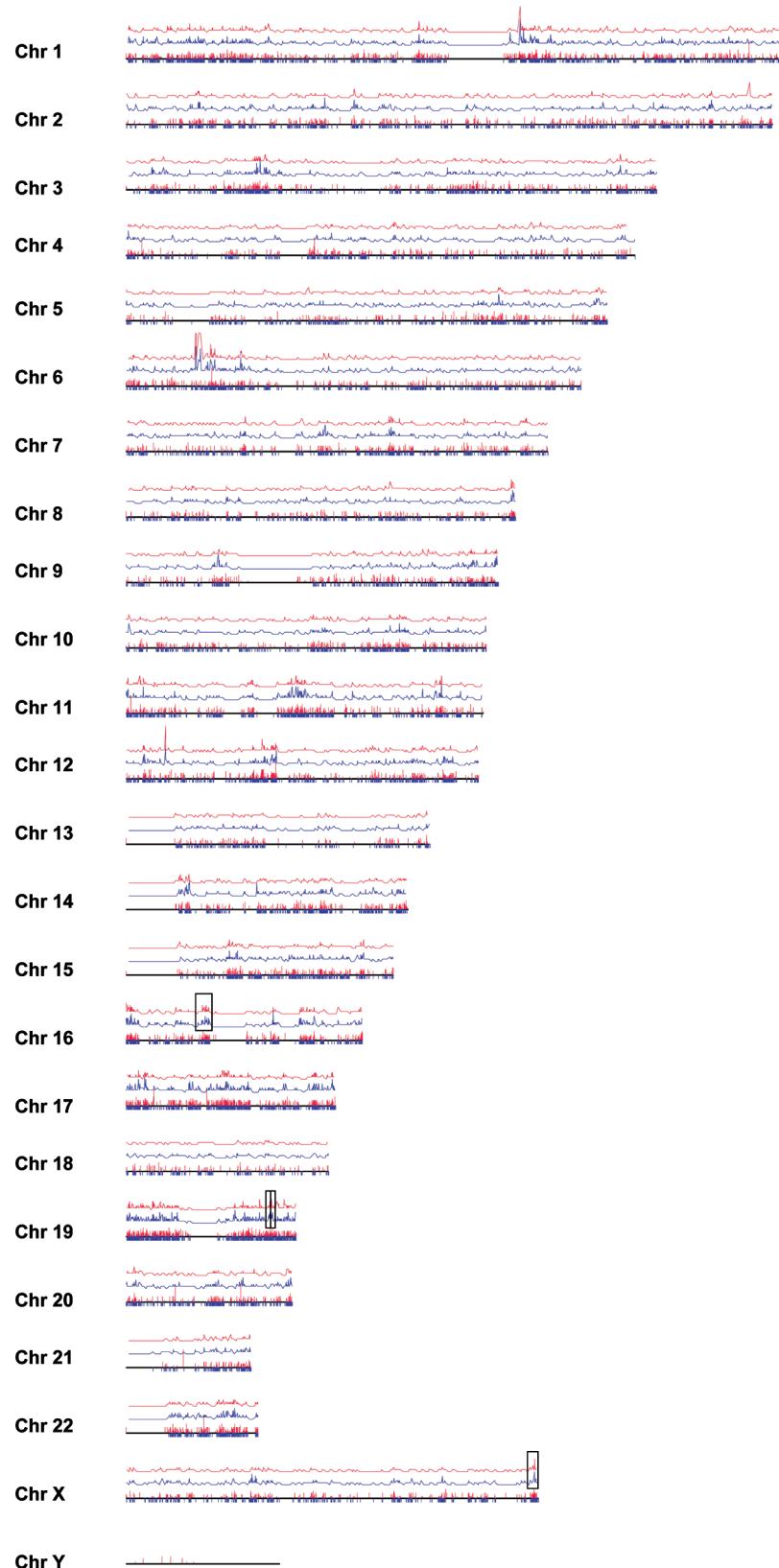


Figure 2. Chromosomal localization and density distribution of the genes represented in the HLP and HLT. Black lines indicate the human chromosomes. Red and blue lines on the chromosomal lines indicate chromosomal localization of the genes related to the HLP and HLT, respectively. The height of the red lines indicates the quantitative values of the proteins in the HLP. The red and blue curves above the chromosomal lines represent the density distribution of corresponding genes related to the HLP or HLT in each 100-kb chromosomal window. The four rectangles represent the four highest gene density chromosomal areas including 16p13.3, 19q13.31, 19q13.42, and Xq28, respectively.

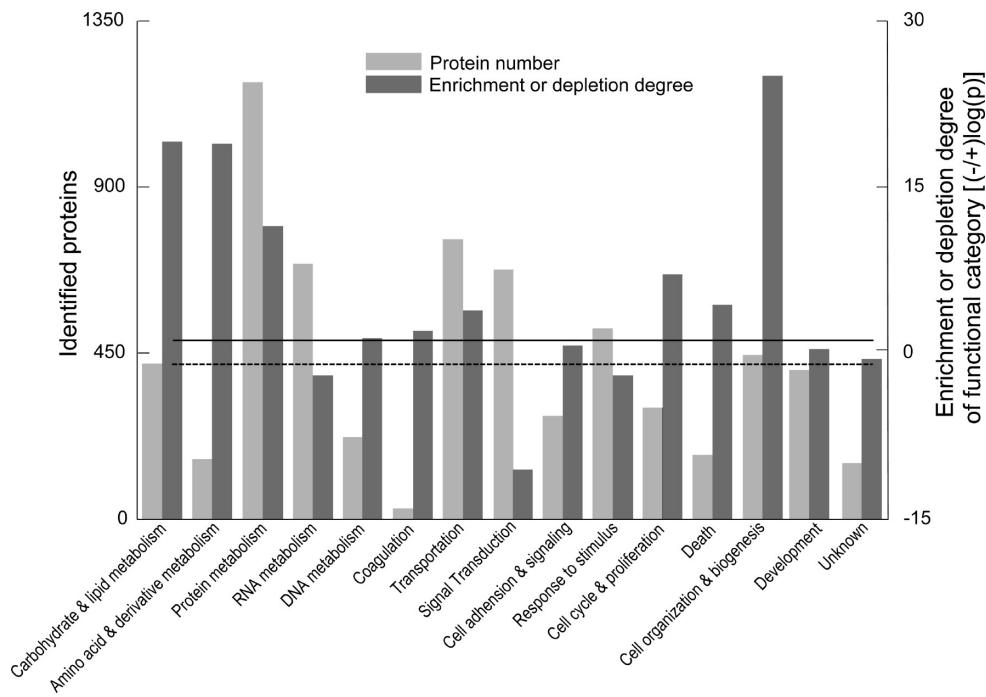


Figure 3. Functional categorization in the HLP. The degree of enrichment or depletion of identified proteins in a given functional category compared to the HGEP is represented as $(+/-)\log(P)$. The horizontal or dashed line across the bars indicates the cutoff value of $-\log(0.05)$ or $+\log(0.05)$. The bars above the line or below the dashed line indicate the significantly enriched or depleted category; those between the two lines indicate that there is no significant enrichment or depletion.

(Figure 3), suggesting that those categories might be tissue/organ specific and/or temporally regulated.

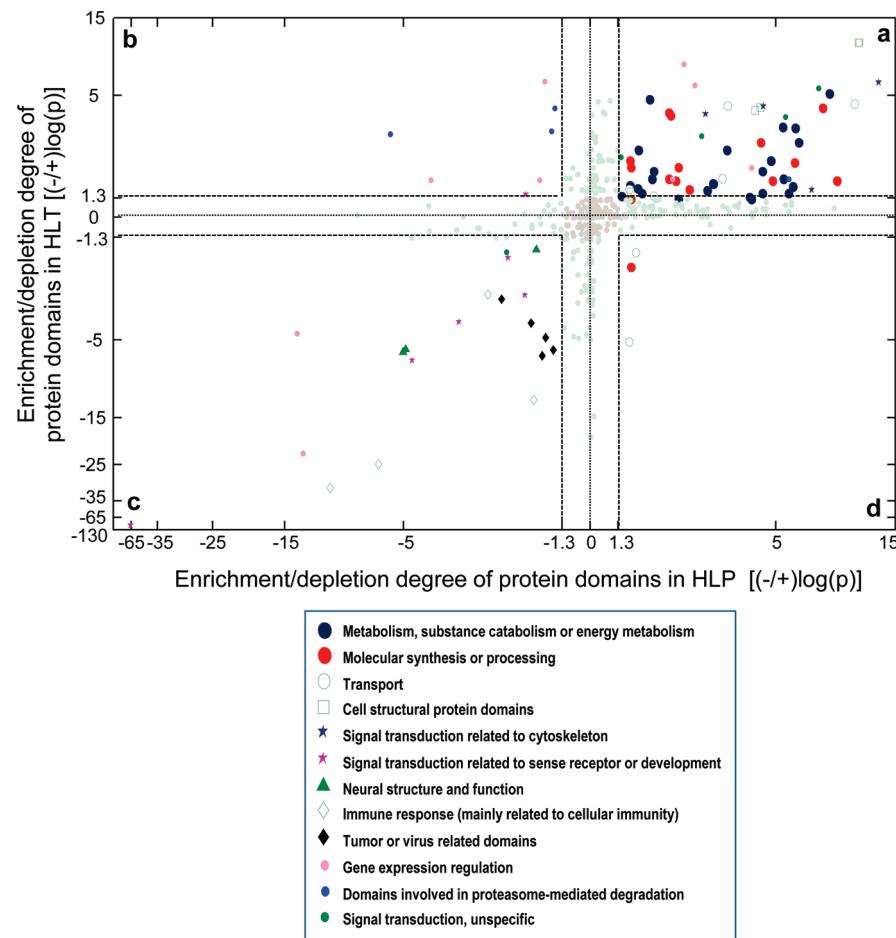
The protein functions correlate closely with their structural domains, in agreement with the general rule that tissue-specific domains are associated with tissue-specific physiological functions. Domains enriched in both the HLP and HLT are involved primarily in liver-specific physiological processes such as metabolism, transport, and biosynthesis (Figure 4 and Supplementary Table 21). In contrast, the depleted domains in both data sets participate primarily in sense receptor-mediated signaling, neural processes, and cellular immunity, consistent with the fact that those processes are usually absent in liver. Further supporting the results of the comparison with the HGEP, there is good agreement in terms of domain enrichment and depletion between the HLP and HLT.

Metabolic Pathways. Liver is a metabolically active organ responsible for many vital life functions, such as metabolism of carbohydrates, fats, and proteins; storage of glycogen; bile production and excretion; secretion of bilirubin and cholesterol; and synthesis and detoxification of plasma proteins. For this comprehensive survey, all 94 human metabolic pathways were extracted from KEGG and applied to the functional annotation of the HLP and HLT data sets. Within the metabolic pathways, 1040 proteins were detected through HLP determinations. Twenty-four pathways involved in the major liver metabolic functions (Table 1) were completely covered by HLP data set. In addition, more than 80% of the members of another 24 pathways, which also played essential roles in liver metabolism, were covered (Supplementary Tables 22–25). Thirty-one cytochromes of the P450 (CYP) family were identified, of which four (CYP4B1, CYP11B2, CYP19A1, and CYP24A1) were first observed in human liver (Supplementary Table 26 and Supplementary Figure 4).

Transport. The liver is responsible for production and secretion of a number of materials, such as bile salts and serum

proteins, so transport is an essential function of this organ. An overview of possible transport pathways was integrated from the transporter proteins identified in liver (Figure 5, Supplementary Table 27). A total of 938 transporters and associated proteins were identified in the HLP data set. Of these, 46.1% (432) were annotated as membrane or membrane-associated proteins and 31.9% (300) contained one or more transmembrane domains. GO analysis revealed that these proteins were primarily involved in three transport systems, that is, the secretory pathway, endoplasmic reticulum to Golgi transport, and protein transport. In addition, three ion channels (the sodium channel type V alpha subunit,⁴⁹ the alpha1A-voltage-dependent calcium channel,⁵⁰ and the voltage-gated potassium channel beta-3 subunit⁵¹) were first identified in liver. Transcripts of the sodium channel type V alpha subunit and alpha1A-voltage-dependent calcium channel were identified in the HLT data set, providing evidence of gene expression. Although the voltage-gated potassium channel beta-3 subunit was not included in the HLT data set, its identification was based upon four unique peptides with 15.1% sequence coverage, making it a high-confidence identification.

Signal Transduction. Although signaling molecules are always low-abundance proteins, 800 proteins related to signal transduction exist in the HLP data set. When the strategy of node analysis was used, the MAPK pathway, a typical cascade and pivotal pathway, was shown to have the largest number of nodes with 125; in contrast, the hedgehog pathway had the smallest number of nodes with just 18. The node coverage of all 16 pathways was calculated and eight pathways were covered up to 50% in the HLP data set (Figure 6a). Adherens junction (80%, 59/73) and focal adhesion (78%, 47/60) proteins and proteins of the calcium signaling pathway (72%, 28/39) were among the most frequently detected proteins; these are involved in cellular recognition, localization, communication, and inflammation.^{52–55} The pathways participating in meta-



bolic regulation, such as the insulin (66%, 41/62) and adipocytokine (59%, 22/37) pathways, were also highly covered in the HLP data set. Otherwise, the pathways relevant to liver disease, such as the TGF β ,^{56–58} Wnt,^{59,60} and Notch⁶¹ pathways, were covered at low rates, as expected (Supplementary Table 28). An overview, the Yin-Yang map, was derived from proteomic signal transduction data in liver (Figure 6b). This map represents the balance of proliferation and apoptosis in the liver, that is, the two fundamental states of life and death.

Transcription Factors. Transcription factors are critical for regulation of gene expression and are commonly expressed at low-abundance in cells.^{62,63} The TRANSFAC database⁶⁴ contains data on 1670 transcription factors from human tissues. Of these transcription factors, 67.6% (1129) were identified in the HLT data set and 18.5% (309) were recognized in the HLP data set (Supplementary Tables 29 and 30). Those identified transcription factors showed both low detection rates and low abundance compared to housekeeping proteins (Figure 7a). Several liver-enriched transcription factors were found in the HLP data set; these were involved in regulation of gene expression of liver enzymes and plasma proteins,^{65–69} such as hepatic nuclear factor 4 α and CCAAT/enhancer-binding protein (CEBP), as well as common nuclear receptors, including the glucocorticoid receptor, the androgen receptor, the estrogen receptor, and the peroxisome proliferator-activated receptor.

Upon application of GO analysis to the identified transcription factors, they were broadly divided into five categories: development, growth (including differentiation and proliferation), defense systems, signal transduction, and liver-related processes (e.g., transport, metabolism, and biosynthesis). On the basis of the total coverage and average abundance of each category of transcription factor, it appears that those involved in liver-related processes were detected at a high rate at both the transcript and protein levels (Figure 7b). The TRANSFAC database includes information on the target genes regulated by the 98 identified transcription factors. On the basis of this data, 100 products of the target genes of those transcription factors were found in the HLP data set and these could be classified into six groups: metabolic enzymes, plasma proteins, constitutive components, transcription factors, growth factors and receptors, and cell cycle regulators. Figure 7c shows that most of the target genes are related to the physiological functions of liver.

Comparison between the HLP and Human Plasma Proteome (HPP). Most blood components are synthesized in the liver. The HLP was compared with the HPP based on three pairwise data sets with three different levels of confidence of protein identification. The number of overlapping proteins varied from 4241 to 1214 to 216 (Figure 8a). There were 184 proteins consistently identified in the area of overlap between

Table 1. Coverage of Major Metabolism Pathways in Human Liver

metabolism category/name	pathway/process number	enzyme number			HLP coverage (%)		fully covered pathway number
		HGEP ^a	HLT	HLP	average	range	
Carbohydrate metabolism	14	171	154	151	88.3	61.9–100	5
Glycolysis/Gluconeogenesis				28		100	
Propanoate metabolism				22		100	
Citrate cycle (TCA cycle)				21		100	
Glyoxylate and dicarboxylate metabolism				11		100	
Inositol metabolism				2		100	
Lipid metabolism	12	147	128	102	69.4	44.4–100	3
Fatty acid metabolism				23		100	
Fatty acid biosynthesis				10		100	
Fatty acid elongation in mitochondria				8		100	
Amino acid metabolism	14	207	183	170	82.1	70.8–100	2
Phenylalanine metabolism				10		100	
Phenylalanine, tyrosine and tryptophan biosynthesis				7		100	
Energy metabolism	2	108	89	70	64.8	64.8–65.5	0
Bile acid metabolism	2	51	49	39	76.5	71.0–85.0	
Bilirubin metabolism	4	7	7	7	100.0	100–100	
Biotransformation	15	50	47	45	90.0	50.0–100	8
Benzoate degradation via CoA ligation				10		100	
Caprolactam degradation				8		100	
Gamma-Hexachlorocyclohexane degradation				6		100	
1- and 2-Methylnaphthalene degradation				4		100	
Bisphenol A degradation				3		100	
Ethylbenzene degradation				3		100	
Tetrachloroethene degradation				2		100	
Atrazine degradation				1		100	
Cytochrome P450	60	37	31	51.7			4
Metabolism of rare amino acids							
Selenoamino acid metabolism				11		100	
Glutathione metabolism				10		100	
Aminophosphonate metabolism				5		100	
Cyanoamino acid metabolism				4		100	
Metabolism of cofactors and vitamins							1
Vitamin B6 metabolism				5		100	
Biosynthesis of secondary metabolites							1
Monoterpeneoid biosynthesis				3		100	

^a HGEP: Human Genome Encoding Proteome.

HLP and HPP at all three confidence levels (Supplementary Table 31). Functional categorization with GO indicated that these 184 proteins are associated with immune response, transport, metabolism, cytosis, signal transduction, cell adhesion, coagulation, and so forth (Figure 8b, Supplementary Table 32). Quantitative comparison of the HLP and HPP exhibited a close correlation; that is, most of the liver-specific secreted proteins⁷⁰ appeared at high abundance and other proteins appeared at relatively low abundance in both liver and plasma (Figure 8c).

Liver is a unique organ that produces different protein components involved in coagulation, anticoagulation, and fibrinolysis (CAF), and the complement system. The proteins participating in these systems were clearly recognized, that is, 67% (22/33) in CAF and 64% (16/25) in the complement system (Figure 8d). Of these, 11 CAF proteins and 14 complement proteins were consistently found across all the HLP and HPP data sets (Figure 8e, Supplementary Table 33). In addition, a highly positive correlation ($R = 0.978$) was observed between SCIN values of coagulation proteins in the HLP and the quantitative proteomic data reported by Mann's group⁷¹ (Figure 8f). Taken together, the quantitative comparison demonstrated that the expression levels of secreted proteins in liver

were in agreement with the concentrations of corresponding plasma proteins, particularly for coagulators.

We realize that it is almost impossible to complete a completely comprehensive study on this vast proteome data set in a short time. Consequently, here we have presented some enticing highlights from the analysis of the HLP that we think we will encourage future in-depth analyses.

Discussion

A comprehensive map of the human liver proteome is reported for the first time in this article. With an integrated strategy combining four distinct technical approaches, 6788 proteins (3721 newly identified in liver) were identified with confidence and compiled in the database PROTEOME^{SKY}-LIVER^{Hu} 1.0, which represents the largest proteomic data set of a human organ to date. To estimate the extent of coverage of the HLP data, the corresponding transcriptome was determined in parallel with the same sample; this analysis showed that 61.1% of the transcriptome was covered by proteome database. Semiquantitative analysis revealed a dynamic range of 6 orders of magnitude within the proteome. Of the newly identified liver proteins, 82.5% were of low abundance. The

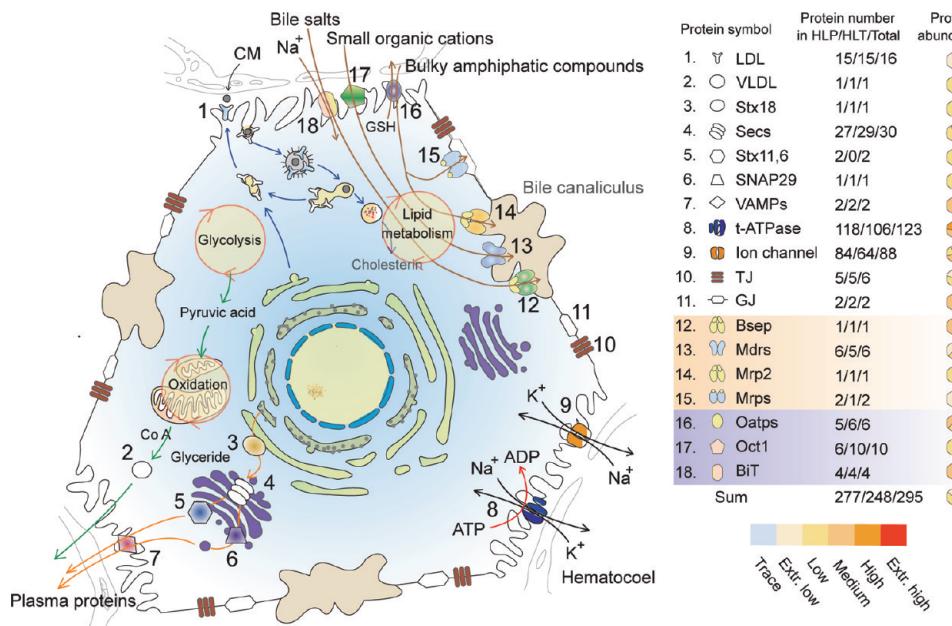


Figure 5. Overview of transport systems in the HLP. Abbreviations are as follows: CM, chylomicrometer; CoA, acetyl-CoA; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein; TJ, tight junction-related protein; GJ, gap junction-related protein; Secs, SEC related proteins; Stx, Syntaxin; SNAP, SnRNA-activating protein; VAMPs, vesicle-associated membrane proteins; t-ATPase, transportation related ATPase; Bsep, bile salt export pump; Mdr, multidrug resistance protein; Mrps, multidrug resistance-associated proteins for basolateral bile salt efflux; Oatps, ornithine aminotransferase proteins; Oct1, organic cation transporter 1; BiT, bile transporters. Additional detail about the data presented in the figure can be found in Supplementary Table 27. Transporters for carrying bile acid into or out of the cell are highlighted with light red or cyan background, respectively. The colored arrows in the overview map indicate the corresponding pathways/components for transporters: blue, LDL transport; green, junction components; yellow, VLDL transport; orange, blood-related protein secretory pathway; black, ion transport and ATPase-related pathway.

biological categories associated most closely with liver functions were characterized and highlighted in the proteome. The HLPP, as the first proteomic initiative to profile a human organ,⁷² has successfully demonstrated that the proteins in a human organ can be assessed quantitatively and qualitatively in a large-scale approach.

Confidence of the HLP Data Set. Currently, the individual analytical platforms can determine the fractions of proteins and peptides in a sample,²³ but a complementary technical strategy is necessary to extend coverage of the proteome. To address this, we constructed integrated technology platforms by combining separation techniques, mass spectrometry ionization modes, and search engines. Furthermore, to address the challenge of the extreme complexity and large dynamic range of the liver proteome, we performed the experiments employing a variety of replicates; these included analysis of the same sample using different platforms, implementation of the same platform in multiple laboratories, and duplication of runs of the same platform within a single lab.^{10,11} This technical strategy resulted in a comprehensive profile of human liver consisting of 6788 unique proteins; these proteins were shown to be associated with 60% of the PEGs acquired by analysis of the liver transcriptome in the same sample. Thus, the scale of the proteomic data in the HLP has demonstrated that our strategy was successful in a comprehensive effort to profile a tissue/organ proteome. The semiquantitative analysis indicated that most proteins newly identified in human liver were in relatively low abundance, demonstrating the sensitivity of the mass spectrometric techniques employed here. For example, three ion channel proteins with low abundance and high hydrophobicity were first identified in the HLP, that is, the

sodium channel type V alpha subunit, the alpha1A-voltage-dependent calcium channel, and the voltage-gated potassium channel beta-3 subunit. The dynamic range of protein expression in a tissue/organ has been estimated to span up to 9 orders of magnitude.⁷³ The technical approaches described here, therefore, must be further enhanced and refined to allow detection and characterization of extremely low-abundance proteins.

Because the experimental data were generated in multiple laboratories using different technical approaches, maintenance of high-confidence data as well as efficient and accurate integration of diverse data were of critical importance. Three approaches were adopted to process the large amount of data. First, the same reversed-shift database was used to evaluate all identified peptides as well as the false positive rate of peptide identification. This target-decoy database was found to be suitable for the assessment of both the PMF and tandem MS data. Second, the key parameters used in data quality control (e.g., the delta-Cn and Xcorr, least number of amino acids of the tandem mass spectra peptides, and the least number of peptides and the lowest mass error in PMF results) were determined by statistical analysis of real experimental data rather than by empirical judgments. Finally, identification of a protein was confirmed only with two or more peptides matches, which conferred a higher confidence level on the core data set and established a reliable data set for further data mining.

The HLP and Liver Physiology. The fundamental requirement of a proteome data set for a human organ is that the physiological functions of the organ are clearly represented in the component proteins. The HLP data set provides the

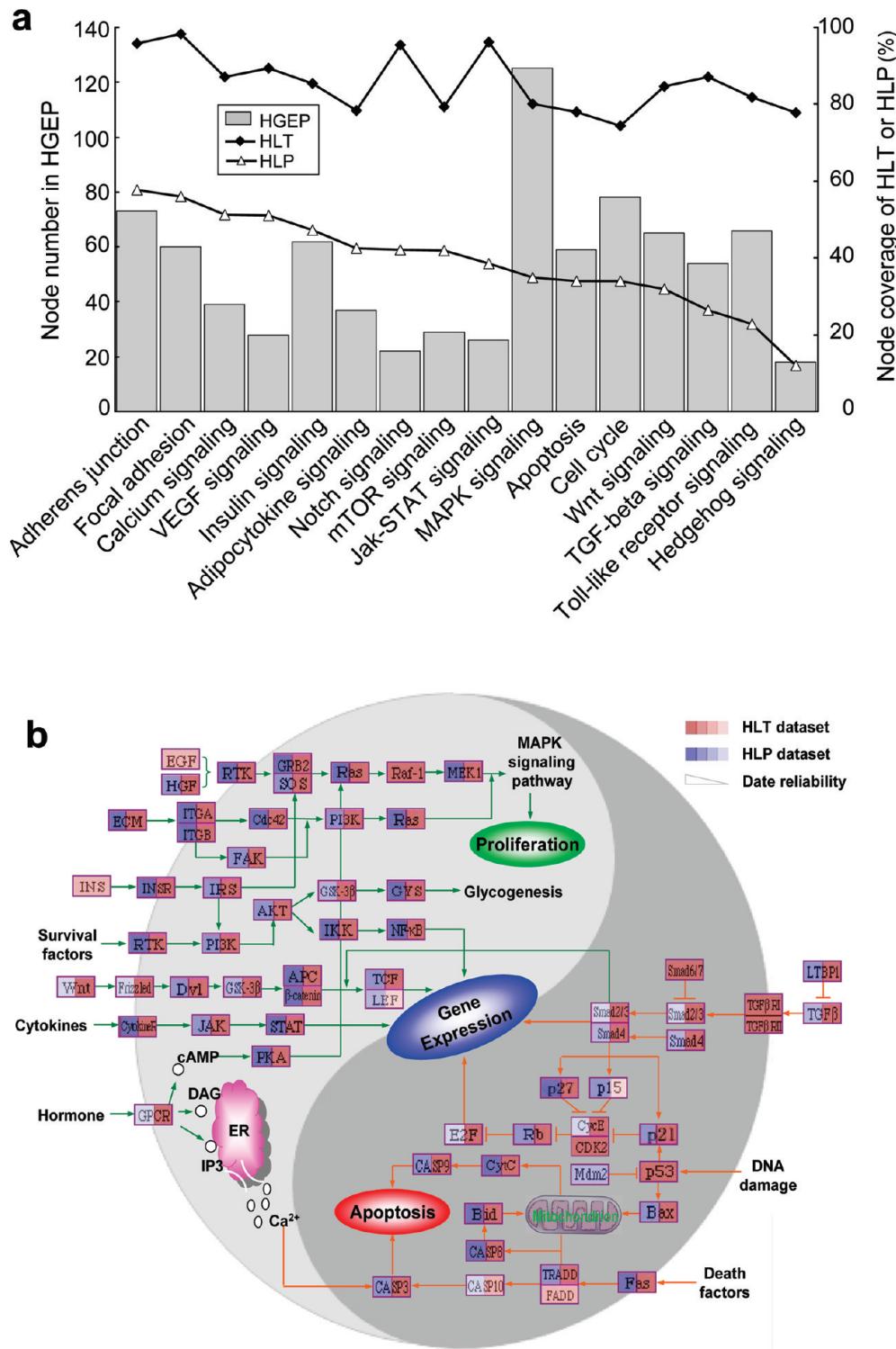


Figure 6. Overview of signal transduction pathways in the HLP. (a) Signal transduction pathways represented in the HLP and HLT. According to KEGG, the unit of each signal transduction pathway is defined as a node. One node includes several proteins which either belong to one family or have the same function in the pathway. The number of nodes for all 16 signal transduction pathways is represented by the bars. The node number was calculated according to KEGG and the coverage of each pathway in the HLP or HLT data set compared to HGEP, abstracted from IPI Human v3.07 database, was demonstrated by the percentage of node-coverage. (b) The overview of signal transduction network. All key signal transduction pathways were integrated into the Yin-Yang scheme, in which all components were identified from different HLP (blue) and/or HLT (brown) data sets. Four degrees of blue color from deep to light represent proteins HLP_99P2/HLP_95P2 (proteins with two or more peptide matches with 99%/95% confidence), HLP_99P1/HLP_95P1 (proteins with one peptide match with 99%/95% confidence). Four degrees of brown color, also from deep to light, represent transcripts identified in both MPSS/chip data sets, in either the MPSS or chip data set only, plus in both EST/SAGE data sets in the public adult normal liver database, and in either the EST or SAGE database only. According to their roles in cell apoptosis and proliferation, those pathways were categorized into two zones, "Yin" and "Yang", indicating that the two zones are different, but not mutually exclusive, and distinct but interchangeable from each other.

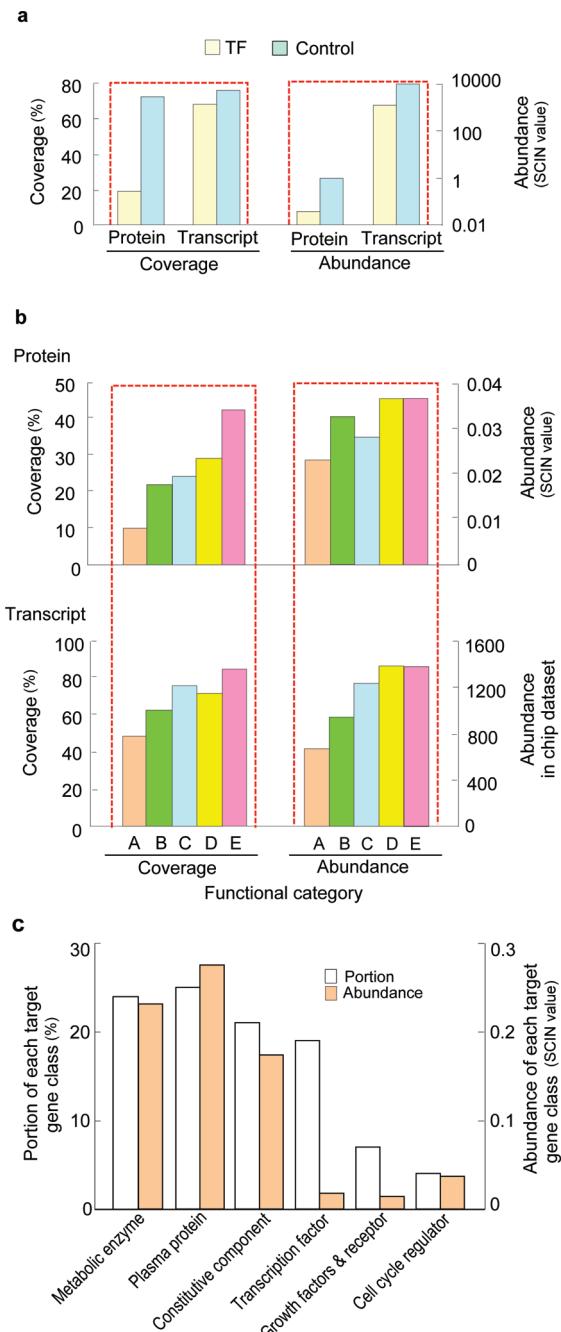


Figure 7. Overview of transcription factors in the HLP and HLT. (a) Coverage and abundance of transcription factors (TF). Alcohol and aldehyde dehydrogenases were used as controls for abundant liver proteins. Abundance of transcripts was determined from the chip data set. (b) Comparison of protein (above) or transcript (below) coverage (left) and abundance (right) among functional categories (A–E) of transcription factors. Functional categories: A, development; B, growth, including differentiation and proliferation; C, defense systems; D, signal transduction; E, liver-related processes (such as transport, metabolism, and biosynthesis). (c) Proportion and abundance of each target-gene class. Proportion indicates the percentage of a given class of target genes out of the total number of target genes. Abundance indicates proteins' average abundance of target genes in a given class.

proteomic basis of physiological liver functions from at least five perspectives: (1) Analysis of GO slim and protein domains showed that the proteins participating in metabolic pathways

Chinese Human Liver Proteome Profiling Consortium

and molecular delivery systems were significantly enriched in liver. For instance, consistent with liver's role as the unique site of metabolism and excretion of many compounds (endogenous or exogenous) into bile, the HLP data set includes most proteins involved in bile and bilirubin metabolism, that is, 71% of the known bile transporters, 85% of the proteins known to be involved in bile acid synthesis, and 100% of the proteins known to be participate in bilirubin metabolism. (2) Qualitative and quantitative comparison of the HLP and HPP show a close correlation, which is consistent with the fact that hepatocytes synthesize and secrete more than 90% of circulating plasma proteins.⁷⁴ (3) In the HLP, 83% of proteins participating in xenobiotic metabolism and degradation pathways were identified; of these, eight of 15 pathways were fully covered, demonstrating the high capacity of the liver for biotransformation.⁷⁵ CYPs were also identified with high coverage. Of these, CYP8B, which catalyzes the oxygenation of sterols, was previously shown at the mRNA level to be expressed specifically in rat liver.⁷⁶ (4) In the signal transduction pathways surveyed in the HLP, the coverage of protein functional categories related to cellular recognition, localization, communication, and inflammation were ranked at the top. Furthermore, insulin and adipocytokine signaling pathways participating in glucose and fatty acid metabolism regulation were also highly covered in the HLP data set. (5) Although transcription factors were generally recognized at relatively low coverage and abundance in the HLP, those with relatively high detection rates were involved in the regulation of expression of genes for typical liver functions such as transport, metabolism, and biosynthesis. Moreover, many liver-enriched transcription factors were identified in the HLP, such as hepatic nuclear factor 4 and CEBP.

The HLP and Disease. On the basis of the chromosomal distribution of PEGs in the HLP, we identified several disease-gene-enriched regions; these included Xq28, which is associated with tumor progression;³⁶ 19q13.31, associated with coronary heart disease;³⁷ and 16p13.3, associated with polycystic kidney disease, tuberous sclerosis, hepatocellular carcinomas, and liver regeneration.^{39–43} In addition, qualitative and quantitative comparisons of the HLP and HPP data sets provide an incentive to pursue the relationship between the HLP and disease and to develop diagnostic biomarkers.^{77,78} Furthermore, the data of the Liver Disease-related Genes and Proteins were carefully collected so as to systematically generate the LDGP database (228 gene products). Approximately 46.1% of the proteins in the LDGP were also present in the HLP data set, implying that a significant number of genes associated with disease are expressed in normal liver.

Novel Liver Proteins in the HLP. A number of research efforts have contributed to the progressive accumulation of liver proteome data derived from cells and tissues and from species ranging from rodents to human.^{79–89} Even so, more than half of the proteins identified in this study were discovered here for the first time. It is noteworthy that 82.5% of the newly identified proteins were of relatively low abundance. For example, four CYP members, CYP4B1 (tissue- and species-specific regulation of oxidation of steroids, fatty acids, and xenobiotics),⁹⁰ CYP11B2 (preferential catalysis of the conversion of 11-deoxycorticosterone to aldosterone),⁹¹ CYP19A1 (estrogen biosynthesis in extragonadal tissues),^{82–94} and CYP24A1 (maintenance of calcium homeostasis and prevention of vitamin D toxicity),^{95,96} were first identified in human liver. In addition, three ion channel proteins were also identified for the first time

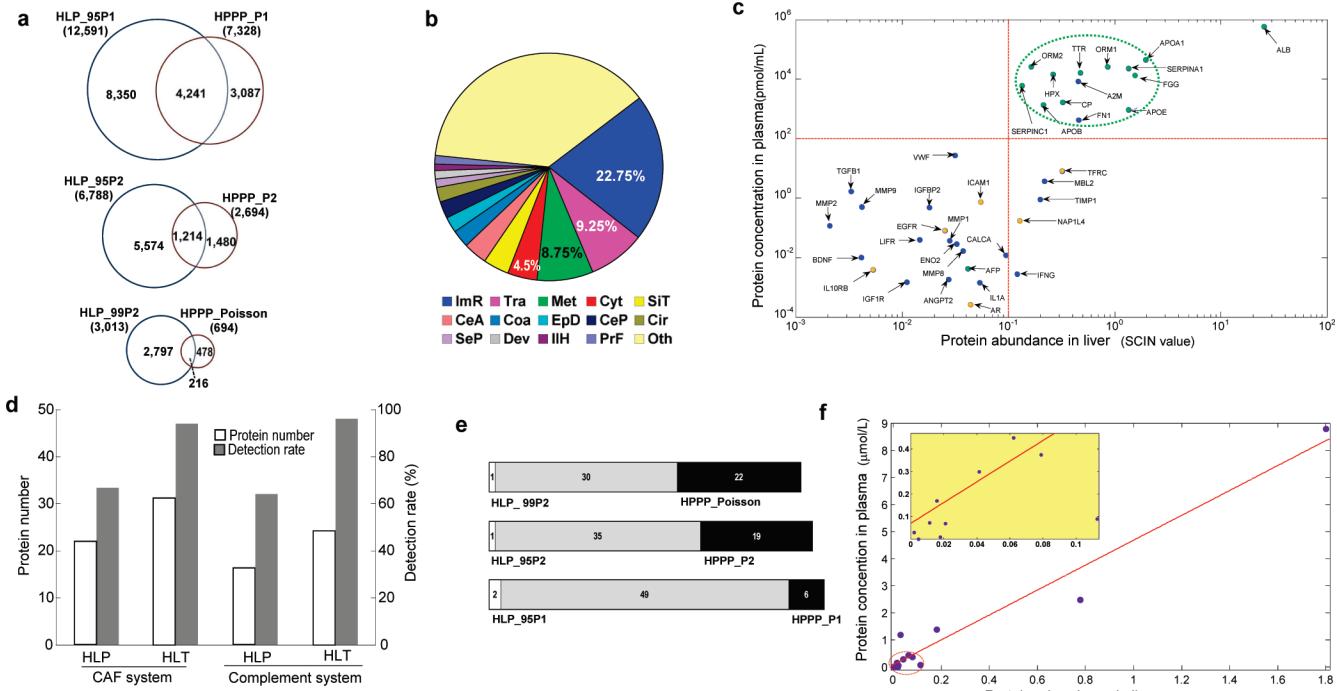


Figure 8. Comparison between the HLP and HPP. (a) Overlap of human liver and plasma proteins. HPPP, Human Plasma Proteome Project; HPPP_P1, P2, Poisson, human plasma proteome data sets at different confidence levels.²³ (b) Functional categories of the 184 proteins consistently overlapping between the HLP and HPPP data sets at three confidence levels. ImR, immune response; Tra, transport; Met, metabolism; Cyt, cytosis; SiT, signal transduction; CeA, cell adhesion; Coa, coagulation; EpD, epidermal development; CeP: cell proliferation; Cir: circulation; SeP: sensory perception; Dev: development; IIH: iron ion homoestasis; PrF: protein folding; Oth, other. (c) Quantitative comparison of the HLP (*SCIN* value) and HPP (immunoassay). Liver-specific secretory proteins, ubiquitous secretory proteins, and nonsecretory proteins were labeled in green, blue, and yellow, respectively. Two orthogonal red dashed lines partition the area into four regions by abundance grades (full protein names are listed in Supplementary Table 32). (d) Numbers of identified proteins and rates of coverage in CAF and complement systems in the HLP and HLT data sets. (e) Comparison of CAF and complement systems between the HLP and HPP data sets. White and black bars represent the number of unique proteins in the HLP and HPP data sets, respectively; the gray bar represents the number of overlapping proteins between the two data sets. (f) Concentration correlation of coagulators between liver (*SCIN* values) and plasma.⁷⁰ The inset is a higher magnification of the purple circular area.

in liver, in which active gene expression of the sodium channel type V alpha subunit and the alpha1A-voltage-dependent calcium channel was further supported by existence of the corresponding transcripts in the HLT data set. Although the voltage-gated potassium channel beta-3 subunit35 was not represented in the HLT data set, the identification of this protein was based on four unique peptides with 15.1% sequence coverage, which confers a high level of confidence on its identification. Approximately 25% of the newly identified proteins were previously hypothetical and were first confirmed at the protein level here, providing insight into their functions and demonstrating the power of the data set as a resource for annotating the human genome. Evidence for the existence of a large number of important biochemical pathways, descriptions of which had previously been fragmented, was provided by the identification of component proteins in these data sets. Intriguingly, most of the pathways associated with newly identified liver proteins have rarely been considered to be associated with normal physiological liver function. This interesting development may lead to the identification of new functions of these pathways in liver or of previously unidentified physiological functions of the liver.

The HLP and HLT. The genetic information flux from genome to transcriptome to proteome is a dynamic and precisely regulated biological process. The comparative analysis of the HLT and HLP data sets offers a wealth of knowledge for use in deciphering this complex process in human liver. It is

established that the transcriptome and proteome have related but distinct expression patterns.^{82,88} However, chromosomal localization analysis showed that the genes encoding the transcripts and proteins of the HLT and HLP have quite similar chromosomal distribution patterns, implying that the regulation of expression of both the transcriptome and the proteome might take place at the level of the genome. Domain analysis revealed that the domains of proteins in the HLP and in proteins encoded by transcripts in the HLT were enriched for functional categories characteristic of liver, such as metabolism, transport, and biosynthesis. These data support the hypothesis that the transcriptome and proteome are qualitatively similar and are consistent with the genome serving as the blueprint for liver structure and physiological function.

Conclusions and Perspective

We have presented here a draft of the human liver proteome, which may be considered a first step in the construction of a highly accurate and comprehensive human liver proteome. From a technological perspective, the quality of this proteome will be improved by the development of higher-performance methods of protein separation, by more sensitive and accurate methods of protein identification, and advances in data mining. Ultimately, the proteomes of the individual cell types that make up the liver, as well as their organelles, will be profiled.

Abbreviations: 2DE, two-dimensional gel electrophoresis; CEBP, CCAAT/enhancer-binding protein; CNHLPP, Chinese

Human Liver Proteome Project; CYP, CYtochromes of P450; ESI, electrospray ionization; GO, Gene Ontology; HGEP, human genome encoding proteome; HHP, human normal heart proteome; HLP, human liver proteome; HLT, human liver transcriptome; HPP, human plasma proteome; ILP, integrated liver tissue proteome; LDGP, liver disease related genes and proteins; MALDI-TOF-TOF-MS, matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry; MS/MS, tandem mass spectrometry; PEGs, protein-encoding genes; RPLC, reverse-phase liquid chromatography; SC, spectral counts; SCI, spectral count index; SCIN, spectral count index normalized; SCX, strong cation exchange; SOPs, standard operation procedures.

Acknowledgment. This work was supported by China Human Liver Proteome Project cofunded by Chinese State Key Projects for Basic Research (973) (2006CB910401), Chinese State High-tech Program (863) (2006AA02A308), National Natural Science Foundation of China for Creative Research Groups (30621063).

Supporting Information Available: Supplementary materials and methods, figures and tables. This material is available free of charge via the Internet at <http://pubs.acs.org> or <http://www.hupo.org.cn/paper/09hlp>.

References

- Cyranoski, D. China takes centre stage for liver proteome. *Nature* **2003**, *425*, 441.
- Service, R. F. Proteomics. Public projects gear up to chart the protein landscape. *Science* **2003**, *302*, 1316–8.
- He, F. Human liver proteome project: plan, progress, and perspectives. *Mol. Cell. Proteomics* **2005**, *4*, 1841–8.
- Zhang, X.; Guo, Y.; Song, Y.; Sun, W.; Yu, C.; Zhao, X.; Wang, H.; Jiang, H.; Li, Y.; Qian, X.; Jiang, Y.; He, F. Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis. *Proteomics* **2006**, *6*, 5260–8.
- Brenner, S.; Johnson, M.; Bridgham, J.; Golda, G.; Lloyd, D. H.; Johnson, D.; Luo, S.; McCurdy, S.; Foy, M.; Ewan, M.; Roth, R.; George, D.; Eletr, S.; Albrecht, G.; Vermaas, E.; Williams, S. R.; Moon, K.; Burcham, T.; Pallas, M.; DuBridge, R. B.; Kirchner, J.; Fearon, K.; Mao, J.; Corcoran, K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **2000**, *18*, 630–4.
- Jongeneel, C. V.; Delorenzi, M.; Iseli, C.; Zhou, D.; Haudenschild, C. D.; Khrebtukova, I.; Kuznetsov, D.; Stevenson, B. J.; Strausberg, R. L.; Simpson, A. J.; Vasicek, T. J. An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Res.* **2005**, *15*, 1007–14.
- Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–7.
- He, P.; He, H. Z.; Dai, J.; Wang, Y.; Sheng, Q. H.; Zhou, L. P.; Zhang, Z. S.; Sun, Y. L.; Liu, F.; Wang, K.; Zhang, J. S.; Wang, H. X.; Song, Z. M.; Zhang, H. R.; Zeng, R.; Zhao, X. The human plasma proteome: analysis of Chinese serum using shotgun strategy. *Proteomics* **2005**, *5*, 3442–53.
- Gong, Y.; Li, X.; Yang, B.; Ying, W.; Li, D.; Zhang, Y.; Dai, S.; Cai, Y.; Wang, J.; He, F.; Qian, X. Different immunoaffinity fractionation strategies to characterize the human plasma proteome. *J. Proteome Res.* **2006**, *5*, 1379–87.
- Liu, H.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193–201.
- Durr, E.; Yu, J.; Krasinska, K. M.; Carver, L. A.; Yates, J. R.; Testa, J. E.; Oh, P.; Schnitzler, J. E. Direct proteomic mapping of the lung microvascular endothelial cell surface *in vivo* and in cell culture. *Nat. Biotechnol.* **2004**, *22*, 985–92.
- Kersey, P. J.; Duarte, J.; Williams, A.; Karavidopoulou, Y.; Birney, E.; Apweiler, R. The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **2004**, *4*, 1985–8.
- Li, X.; Gong, Y.; Wang, Y.; Wu, S.; Cai, Y.; He, P.; Lu, Z.; Ying, W.; Zhang, Y.; Jiao, L.; He, H.; Zhang, Z.; He, F.; Zhao, X.; Qian, X. Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project. *Proteomics* **2005**, *5*, 3423–41.
- Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., III. A method for the comprehensive analysis of membrane proteins. *Nat. Biotechnol.* **2003**, *21*, 532–8.
- Bradshaw, R. A. Revised draft guidelines for proteomic data publication. *Mol. Cell. Proteomics* **2005**, *4*, 1223–5.
- Foster, L. J.; de Hoog, C. L.; Zhang, Y.; Zhang, Y.; Xie, X.; Mootha, V. K.; Mann, M. A mammalian organelle map by protein correlation profiling. *Cell* **2006**, *125*, 187–99.
- Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T.; Rossant, J.; Hughes, T. R.; Frey, B.; Emili, A. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **2006**, *125*, 173–86.
- Moore, R. E.; Young, M. K.; Lee, T. D. Qscore: an algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 378–86.
- Ram, R. J.; Verberkmoes, N. C.; Thelen, M. P.; Tyson, G. W.; Baker, B. J.; Blake, R. C., II; Shah, M.; Hettich, R. L.; Banfield, J. F. Community proteomics of a natural microbial biofilm. *Science* **2005**, *308*, 1915–20.
- Andersen, J. S.; Lam, Y. W.; Leung, A. K.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M. Nucleolar proteome dynamics. *Nature* **2005**, *433*, 77–83.
- Xue, X.; Wu, S.; Wang, Z.; Zhu, Y.; He, F. Protein probabilities in shotgun proteomics: evaluating different estimation methods using a semi-random sampling model. *Proteomics* **2006**, *6*, 6134–45.
- Egelhofer, V.; Gobom, J.; Seitz, H.; Giavalisco, P.; Lehrach, H.; Nordhoff, E. Protein identification by MALDI-TOF-MS peptide mapping: a new strategy. *Anal. Chem.* **2002**, *74*, 1760–71.
- States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat. Biotechnol.* **2006**, *24*, 333–8.
- Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4*, 1487–502.
- Rappaport, J.; Ryder, U.; Lamond, A. I.; Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **2002**, *12*, 1231–45.
- Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–58.
- Draghici, S.; Khatri, P.; Bhavsar, P.; Shah, A.; Krawetz, S. A.; Tainsky, M. A. Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. *Nucleic Acids Res.* **2003**, *31*, 3775–81.
- Coppe, A.; Danieli, G. A.; Bortoluzzi, S. REEF: searching REgionally Enriched Features in genomes. *BMC Bioinf.* **2006**, *7*, 453.
- Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonhammer, E. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. The Pfam protein families database. *Nucleic Acids Res.* **2004**, *32*, D138–41.
- Portugaly, E.; Harel, A.; Linial, N.; Linial, M. EVEREST: automatic identification and classification of protein domains in all protein sequences. *BMC Bioinf.* **2006**, *7*, 277.
- Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–9.
- Li, D.; Li, J.; Ouyang, S.; Wu, S.; Wang, J.; Xu, X.; Zhu, Y.; He, F. An integrated strategy for functional analysis in large-scale proteomic research by gene ontology. *Prog. Biochem. Biophys.* **2005**, *32*, 1026–9.
- Ogata, H.; Goto, S.; Fujibuchi, W.; Kanehisa, M. Computation with the KEGG pathway database. *BioSystems* **1998**, *47*, 119–28.
- Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676–82.
- Zhang, J.; Xu, X.; Gao, M.; Yang, P.; Zhang, X. Comparison of 2-D LC and 3-D LC with post- and pre-tryptic-digestion SEC fractionation for proteome analysis of normal human liver tissue. *Proteomics* **2007**, *7*, 500–12.
- Edelmann, J.; Richter, K.; Hanel, C.; Hering, S.; Horn, L. C. X chromosomal and autosomal loss of heterozygosity and micro-

- satellite instability in human cervical carcinoma. *Int. J. Gynecol. Cancer* **2006**, *16*, 1248–53.
- (37) Heijmans, B. T.; Beekman, M.; Putter, H.; Lakenberg, N.; van der Wijk, H. J.; Whitfield, J. B.; Posthuma, D.; Pedersen, N. L.; Martin, N. G.; Boomsma, D. I.; Slagboom, P. E. Meta-analysis of four new genome scans for lipid parameters and analysis of positional candidates in positive linkage regions. *Eur. J. Hum. Genet.* **2005**, *13*, 1143–53.
- (38) Dumur, C. I.; Dechsukhum, C.; Ware, J. L.; Cofield, S. S.; Best, A. M.; Wilkinson, D. S.; Garrett, C. T.; Ferreira-Gonzalez, A. Genome-wide detection of LOH in prostate cancer using human SNP microarray technology. *Genomics* **2003**, *81*, 260–9.
- (39) Lisowsky, T.; Weinstat-Saslow, D. L.; Barton, N.; Reeders, S. T.; Schneider, M. C. A new human gene located in the PKD1 region of chromosome 16 is a functional homologue to ERV1 of yeast. *Genomics* **1995**, *29*, 690–7.
- (40) The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* **1994**, *77*, 881–94.
- (41) European Chromosome 16 Tuberous Sclerosis Consortium. Identification and characterization of the tuberous sclerosis gene on chromosome 16. *Cell* **1993**, *75*, 1305–15.
- (42) Satoh, S.; Daigo, Y.; Furukawa, Y.; Kato, T.; Miwa, N.; Nishiwaki, T.; Kawasoe, T.; Ishiguro, H.; Fujita, M.; Tokino, T.; Sasaki, Y.; Imaoka, S.; Murata, M.; Shimano, T.; Yamaoka, Y.; Nakamura, Y. AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat. Genet.* **2000**, *24*, 245–50.
- (43) Francavilla, A.; Hagiya, M.; Porter, K. A.; Polimeno, L.; Ihara, I.; Starzl, T. E. Augmenter of liver regeneration: its place in the universe of hepatic growth factors. *Hepatology* **1994**, *20*, 747–57.
- (44) He, F.; Wu, C. Human hepatoprotein and its relationship with liver regeneration. *Chin. Sci. Bull.* **1997**, *42*, 1355–60.
- (45) Ahn, J.; Gruen, J. R. The genomic organization of the histone clusters on human 6p21.3. *Mamm. Genome* **1999**, *10*, 768–70.
- (46) Hoffman, S. M.; Fernandez-Salguedo, P.; Gonzalez, F. J.; Mohrenweiser, H. W. Organization and evolution of the cytochrome P450 CYP2A-2B-2F subfamily gene cluster on human chromosome 19. *J. Mol. Evol.* **1995**, *41*, 894–900.
- (47) Morel, F.; Rauch, C.; Coles, B.; Le Ferrec, E.; Guillouzo, A. The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterization of the genetic polymorphism in the hGSTA1 promoter. *Pharmacogenetics* **2002**, *12*, 277–86.
- (48) Xu, S.; Wang, Y.; Roe, B.; Pearson, W. R. Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion. *J. Biol. Chem.* **1998**, *273*, 3517–27.
- (49) Ou, Y.; Gibbons, S. J.; Miller, S. M.; Streege, P. R.; Rich, A.; Distad, M. A.; Ackerman, M. J.; Rae, J. L.; Szurszewski, J. H.; Farrugia, G. SCN5A is expressed in human jejunal circular smooth muscle cells. *Neurogastroenterol. Motil.* **2002**, *14*, 477–86.
- (50) Toru, S.; Murakoshi, T.; Ishikawa, K.; Saegusa, H.; Fujigasaki, H.; Uchihara, T.; Nagayama, S.; Osanai, M.; Mizusawa, H.; Tanabe, T. Spinocerebellar ataxia type 6 mutation alters P-type calcium channel function. *J. Biol. Chem.* **2000**, *275*, 10893–8.
- (51) Leicher, T.; Bahring, R.; Isbrandt, D.; Pongs, O. Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel A-type potassium channel. *J. Biol. Chem.* **1998**, *273*, 35095–101.
- (52) Jamora, C.; Fuchs, E. Intercellular adhesion, signalling and the cytoskeleton. *Nat. Cell Biol.* **2002**, *4*, E101–8.
- (53) McCulloch, C. A.; Downey, G. P.; El-Gabalawy, H. Signalling platforms that modulate the inflammatory response: new targets for drug development. *Nat. Rev. Drug Discovery* **2006**, *5*, 864–76.
- (54) Jockusch, B. M.; Bubeck, P.; Giehl, K.; Kroemker, M.; Moschner, J.; Rothkegel, M.; Rüdiger, M.; Schlüter, K.; Stanke, G.; Winkler, J. The molecular architecture of focal adhesions. *Annu. Rev. Cell Dev. Biol.* **1995**, *11*, 379–416.
- (55) Burridge, K.; Chrzanowska-Wodnicka, M. Focal adhesions, contractility, and signaling. *Annu. Rev. Cell Dev. Biol.* **1996**, *12*, 463–518.
- (56) Bataller, R.; Brenner, D. A. Liver fibrosis. *J. Clin. Invest.* **2005**, *115*, 209–18.
- (57) Rossi, J. M.; Dunn, N. R.; Hogan, B. L.; Zaret, K. S. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev.* **2001**, *15*, 1998–2009.
- (58) Lemaigne, F.; Zaret, K. S. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr. Opin. Genet. Dev.* **2004**, *14*, 582–90.
- (59) Monga, S. P.; Monga, H. K.; Tan, X.; Mulé, K.; Pediaditakis, P.; Michalopoulos, G. K. Beta-catenin antisense studies in embryonic liver cultures: role in proliferation, apoptosis, and lineage specification. *Gastroenterology* **2003**, *124*, 202–16.
- (60) Suksaweang, S.; Lin, C. M.; Jiang, T. X.; Hughes, M. W.; Widelitz, R. B.; Chuong, C. M. Morphogenesis of chicken liver: identification of localized growth zones and the role of beta-catenin/Wnt in size regulation. *Dev. Biol.* **2004**, *266*, 109–22.
- (61) Tanimizu, N.; Miyajima, A. Notch signaling controls hepatoblast differentiation by altering the expression of liver-enriched transcription factors. *J. Cell Sci.* **2004**, *117*, 3165–74.
- (62) Yang, V. W. Eukaryotic transcription factors: identification, characterization and functions. *J. Nutr.* **1998**, *128*, 2045–51.
- (63) Mason, R. J.; Pan, T.; Edeen, K. E.; Nielsen, L. D.; Zhang, F.; Longphre, M.; Eckart, M. R.; Neben, S. Keratinocyte growth factor and the transcription factors C/EBP alpha, C/EBP delta, and SREBP-1c regulate fatty acid synthesis in alveolar type II cells. *J. Clin. Invest.* **2003**, *112*, 244–55.
- (64) Matys, V.; Kel-Margoulis, O. V.; Fricke, E.; Liebich, I.; Land, S.; Barre-Dirrie, A.; Reuter, I.; Chekmenev, D.; Krull, M.; Hornischer, K.; Voss, N.; Stegmaier, P.; Lewicki-Potapov, B.; Saxel, H.; Kel, A. E.; Wingender, E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**, *34* (Database issue), D108–10.
- (65) Schrem, H.; Klempnauer, J.; Borlak, J. Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol. Rev.* **2002**, *54*, 129–58.
- (66) Schrem, H.; Klempnauer, J.; Borlak, J. Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. *Pharmacol. Rev.* **2004**, *56*, 291–330.
- (67) Odom, D. T.; Zizlsperger, N.; Gordon, D. B.; Bell, G. W.; Rinaldi, N. J.; Murray, H. L.; Volkert, T. L.; Schreiber, J.; Rolfe, P. A.; Gifford, D. K.; Fraenkel, E.; Bell, G. I.; Young, R. A. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **2004**, *303*, 1378–81.
- (68) Chawla, A.; Repa, J. J.; Evans, R. M.; Mangelsdorf, D. J. Nuclear receptors and lipid physiology: opening the X-files. *Science* **2001**, *294*, 1866–70.
- (69) Tirona, R. G.; Kim, R. B. Nuclear receptors and drug disposition gene regulation. *J. Pharm. Sci.* **2005**, *94*, 1169–86.
- (70) Haab, B. B.; Geierstanger, B. H.; Michailidis, G.; Vitzthum, F.; Forrester, S.; Okon, R.; Saviranta, P.; Brinker, A.; Sorette, M.; Perlee, L.; Suresh, S.; Drwal, G.; Adkins, J. N.; Omenn, G. S. Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* **2005**, *5*, 3278–91.
- (71) Mann, K. G. Biochemistry and physiology of blood coagulation. *Thromb. Haemostasis* **1999**, *82*, 165–74.
- (72) He, F. Human liver proteome project: plan, progress, and perspectives. *Mol. Cell. Proteomics* **2005**, *4*, 1841–8.
- (73) Sanchez, J. C.; Corthals, G. L.; Hochstrasser, D. F. *Biomedical Applications of Proteomics*; Wiley-VCH Verlag: Weinheim, Germany, 2004.
- (74) MacSween, R. N. M., et al. *Pathology of the Liver*, 4th ed.; Churchill Livingstone: London, 2001.
- (75) Porter, T. D.; Coon, M. J. Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.* **1991**, *266*, 13469–72.
- (76) Ishida, H.; Kuruta, Y.; Gotoh, O.; Yamashita, C.; Yoshida, Y.; Noshiro, M. Structure, evolution, and liver-specific expression of sterol 12alpha-hydroxylase P450 (CYP8B). *J. Biochem. (Tokyo, Jpn.)* **1999**, *126*, 19–25.
- (77) Polanski, M.; Anderson, N. L. A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights* **2006**, *2*, 1–48.
- (78) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845–67.
- (79) Ying, W.; Jiang, Y.; Guo, L.; Hao, Y.; Zhang, Y.; Wu, S.; Zhong, F.; Wang, J.; Shi, R.; Li, D.; Wan, P.; Li, X.; Wei, H.; Li, J.; Wang, Z.; Xue, X.; Cai, Y.; Zhu, Y.; Qian, X.; He, F. A dataset of human fetal liver proteome identified by subcellular fractionation and multiple protein separation and identification technology. *Mol. Cell. Proteomics* **2006**, *5*, 1703–7.
- (80) Foster, L. J.; de Hoog, C. L.; Zhang, Y.; Zhang, Y.; Xie, X.; Mootha, V. K.; Mann, M. A mammalian organelle map by protein correlation profiling. *Cell* **2006**, *125*, 187–99.

research articles

- (81) Gilchrist, A.; Au, C. E.; Hiding, J.; Bell, A. W.; Fernandez-Rodriguez, J.; Lesimple, S.; Nagaya, H.; Roy, L.; Gosline, S. J.; Hallett, M.; Paiement, J.; Kearney, R. E.; Nilsson, T.; Bergeron, J. J. Quantitative proteomics analysis of the secretory pathway. *Cell* **2006**, *127*, 1265–81.
- (82) Mootha, V. K.; Bunkenborg, J.; Olsen, J. V.; Hjerrild, M.; Wisniewski, J. R.; Stahl, E.; Bolouri, M. S.; Ray, H. N.; Sihag, S.; Kamal, M.; Patterson, N.; Lander, E. S.; Mann, M. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **2003**, *115*, 629–40.
- (83) Kislinger, T.; Cox, B.; Kannan, A.; Chung, C.; Hu, P.; Ignatchenko, A.; Scott, M. S.; Gramolini, A. O.; Morris, Q.; Hallett, M. T.; Rossant, J.; Hughes, T. R.; Frey, B.; Emili, A. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **2006**, *125*, 173–86.
- (84) Yan, W.; Lee, H.; Deutsch, E. W.; Lazar, C. A.; Tang, W.; Chen, E.; Fausto, N.; Katze, M. G.; Aebersold, R. A dataset of human liver proteins identified by protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry. *Mol. Cell. Proteomics* **2004**, *3*, 1039–41.
- (85) Shen, H.; Cheng, G.; Fan, H.; Zhang, J.; Zhang, X.; Lu, H.; Liu, C.; Sun, F.; Jin, H.; Xu, X.; Xu, G.; Wang, S.; Fang, C.; Bao, H.; Wang, Y.; Wang, J.; Zhong, H.; Yu, Z.; Liu, Y.; Tang, Z.; Yang, P. Expressed proteome analysis of human hepatocellular carcinoma in nude mice (LCI-D20) with high metastasis potential. *Proteomics* **2006**, *6*, 528–37.
- (86) Schirmer, E. C.; Florens, L.; Guan, T.; Yates, J. R.; Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **2003**, *301*, 1380–2.
- (87) Kislinger, T.; Rahman, K.; Radulovic, D.; Cox, B.; Rossant, J.; Emili, A. PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol. Cell. Proteomics* **2003**, *2*, 96–106.
- (88) Da, Cruz, S.; Xenarios, I.; Langridge, J.; Vilbois, F.; Parone, P. A.; Martinou, J. C. Proteomic analysis of the mouse liver mitochondrial inner membrane. *J. Biol. Chem.* **2003**, *278*, 41566–71.
- (89) Cagney, G.; Park, S.; Chung, C.; Tong, B.; O'Dushlaine, C.; Shields, D. C.; Emili, A. Human tissue profiling with multidimensional protein identification technology. *J. Proteome Res.* **2005**, *4*, 1757–67.
- (90) McKinnon, R. A.; Burgess, W. M.; Gonzalez, F. J.; Gasser, R.; McManus, M. E. Species-specific expression of CYP4B1 in rabbit and human gastrointestinal tissues. *Pharmacogenetics* **1994**, *4*, 260–70.
- (91) Satoh, M.; Nakamura, M.; Saitoh, H.; Satoh, H.; Akatsu, T.; Iwasaka, J.; Masuda, T.; Hiramori, K. Aldosterone synthase (CYP11B2) expression and myocardial fibrosis in the failing human heart. *Clin. Sci.* **2002**, *102*, 381–6.
- (92) Harada, N.; Utsumi, T.; Takagi, Y. Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and promoters, and switching of tissue-specific exons 1 in carcinogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11312–6.
- (93) Mahendroo, M. S.; Mendelson, C. R.; Simpson, E. R. Tissue-specific and hormonally controlled alternative promoters regulate aromatase cytochrome P450 gene expression in human adipose tissue. *J. Biol. Chem.* **1993**, *268*, 19463–70.
- (94) Zhou, J.; Gurates, B.; Yang, S.; Sebastian, S.; Bulun, S. E. Malignant breast epithelial cells stimulate aromatase expression via promoter II in human adipose fibroblasts: an epithelial-stromal interaction in breast tumors mediated by CCAAT/enhancer binding protein beta. *Cancer Res.* **2001**, *61*, 2328–34.
- (95) Chen, K. S.; Prahl, J. M.; DeLuca, H. F. Isolation and expression of human 1,25-dihydroxyvitamin D₃ 24-hydroxylase cDNA. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4543–7.
- (96) Masuda, S.; Byford, V.; Arabian, A.; Sakai, Y.; Demay, M. B.; St-Arnaud, R.; Jones, G. Altered pharmacokinetics of 1alpha, 25-dihydroxyvitamin D₃ and 25-hydroxyvitamin D₃ in the blood and tissues of the 25-hydroxyvitamin D-24-hydroxylase (Cyp24a1) null mouse. *Endocrinology* **2005**, *146*, 825–34.

Chinese Human Liver Proteome Profiling Consortium

PR900532R