A Tool for Biomarker Discovery in the Urinary Proteome: A Manually Curated Human and Animal Urine Protein Biomarker Database* [S]

Chen Shao‡, Menglin Li‡, Xundou Li‡, Lilong Wei‡, Lisi Zhu‡, Fan Yang‡, Lulu Jia‡, Yi Mu‡, Jiannong Wang‡, Zhengguang Guo‡, Dan Zhang‡, Jianrui Yin‡, Zhigang Wang§, Wei Sun¶, Zhengguo Zhang§, and Youhe Gao‡

Urine is an important source of biomarkers. A single proteomics assay can identify hundreds of differentially expressed proteins between disease and control samples; however, the ability to select biomarker candidates with the most promise for further validation study remains difficult. A bioinformatics tool that allows accurate and convenient comparison of all of the existing related studies can markedly aid the development of this area. In this study, we constructed the Urinary Protein Biomarker (UPB) database to collect existing studies of urinary protein biomarkers from published literature. To ensure the quality of data collection, all literature was manually curated. The website (http://122.70.220.102/biomarker) allows users to browse the database by disease categories and search by protein IDs in bulk. Researchers can easily determine whether a biomarker candidate has already been identified by another group for the same disease or for other diseases, which allows for the confidence and disease specificity of their biomarker candidate to be evaluated. Additionally, the pathophysiological processes of the diseases can be studied using our database with the hypothesis that diseases that share biomarkers may have the same pathophysiological processes. Because of the natural relationship between urinary proteins and the urinary system, this database may be especially suitable for studying the pathogenesis of urological diseases. Currently, the database contains 553 and 275 records compiled from 174 and 31 publications of human and animal studies, respectively. We found that biomarkers identified by different proteomic methods had a poor overlap with each other. The differences between sample preparation and separation methods, mass spectrometers, and data analysis algorithms may be influencing factors. Biomarkers identified from animal models also overlapped poorly with those from human samples, but the overlap rate was not lower than that of human proteomics studies. Therefore, it is not clear how well the animal models mimic human diseases. Molecular & Cellular Proteomics 10: 10.1074/mcp.M111.010975, 1–8, 2011.

Urine is an ideal source of biomarkers because it can be noninvasively obtained. Proteins in the urine are mainly composed of plasma proteins that pass through the glomerular filtration barrier as well as proteins secreted from the kidney and urinary tract. In comparison to plasma, urine has some unique advantages that make it a suitable source for both physiological research and disease biomarker discovery. First, urine can be collected continuously and noninvasively. Second, the urinary proteome directly reflects the condition of the urinary system. Third, because the urinary proteome contains a number of plasma proteins, some changes of the plasma proteome can also be found in urine. Therefore, urine is not only a good source for the study of urological diseases, but can also reflect the status of the entire body.

Considerable achievements have been made in the study of urinary proteomics during recent years. With the development of proteomic techniques, over 1500 proteins have been identified in normal human urine (1). The protein list can be found in several web-based databases, such as the Max-Planck Unified Proteome database (MAPU; http://mapuproteome.com/) and the Human Kidney and Urine Proteome Project (HKUPP; http://hkupp.org/). In addition, the Sys-BodyFluid (http://www.biosino.org/bodyfluid/) (2) collected 1941 normal human urinary proteins from nine peer-reviewed publications.

Biomarker discovery is a hot area of urinary proteomics. Current studies focus on urogenital diseases, such as various chronic and acute renal injuries (3), rejection after renal transplantation (4), bladder cancer (5), and prostate cancer (6). In addition, changes in the urinary proteome were also shown to be related to some systematic diseases, such as diabetes (7), graft-versus-host-disease (GVHD) after allogeneic hematopoietic stem cell transplantation (8, 9), and coronary artery disease (10).

From the ‡National Key Laboratory of Medical Molecular Biology, Department of Physiology and Pathophysiology; ¶Department of Biomedical Engineering; §Core Instrument Facility, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, Beijing, China
Received May 11, 2011, and in revised form, July 25, 2011
Published, MCP Papers in Press, August 29, 2011, DOI 10.1074/mcp.M111.010975
Studies focused on urinary biomarker discovery face two major problems. The first involves the disease specificity of the biomarker. Because an important clinical aim is to differentiate diseases with similar signs and symptoms in order to help physicians choose proper treatments, the identification of disease-specific biomarkers is crucial. Considering that a number of previous studies have only compared the urinary proteome of patients with a specific disease to healthy controls, the disease specificity of previously identified biomarkers or biomarker candidates is not conclusive.

The second problem involves the confidence of the biomarker. Every experimental technique has advantages and limitations, their results are hardly to be completely consistent. In addition to the limitation of experimental methods, the existence of individual differences is another factor that can affect the results. It has been reported that the urinary proteome has significant differences, even among healthy individuals, and varies in the same individual at different time points because of the effects of exercise, diet, lifestyle, or other factors (11, 12). These physiological variations make it more difficult to discover biomarkers from human urine. Candidate biomarkers identified by proteomics approaches based on small amounts of samples have to be verified in larger sample sets to determine whether their changes are caused by diseases or by the physiological variations.

An ideal way to solve the problems described above is to compare all related diseases at the same time with a sufficient sample size. However, it is almost impossible for a single research group to perform this type of study because of limitations in the availability of patient samples and experimental costs. Currently, a single proteomics assay can identify hundreds of differentially expressed proteins between disease and control groups, but only a portion of them have enough disease specificity and confidence to serve as biomarkers.

To identify potential biomarkers for further validation, researchers need as much information as possible to help them assess the disease specificity and confidence of all of the differentially expressed proteins that are identified. For example, researchers may want to know whether these proteins have been reported to be biomarkers or biomarker candidates for the same disease by other groups, and if so, whether the fold changes they observed agree or conflict with the existing results. In addition, it would be useful to know whether the proteins were previously identified as biomarkers or biomarker candidates for other diseases, which would indicate the poor disease specificity of these proteins, or whether orthologous proteins have been studied in animal models but still lack validation in human samples.

The construction of a database that compiles all of the previously published biomarkers or biomarker candidates in urine may help elucidate answers to the questions above. Currently, the Human Urinary Proteome Database (http://mosaiques-diagnostics.de/diapatpcms/mosaiquescms/front_content.php?idcat = 257) compiled by Coon et al. (13) includes 3926 Capillary Electrophoresis-Mass Spectrometry (CE-MS) results for 34 diseases and is the only database to store information regarding urinary peptide and protein biomarkers. CE-MS is highly reproducible, but does not provide sequence information on peptides, and therefore the protein names of the biomarkers cannot be inferred. Thus, it is difficult to study the function of the biomarkers as well as the pathogenesis of diseases that can be inferred from their respective biomarkers.

In this study, we focused on proteins identified by proteomic approaches with sequence information or by small-scale experiments in which the protein names of the biomarkers could be identified, such as ELISA and Western blot. Knowledge regarding peptide sequences or protein names allows for the functions of these proteins to be analyzed by further studies. The database was constructed by querying all literature on urinary biomarker discovery in PubMed and manually curating the studies to retrieve disease and biomarker information.

The database described in this study is not only useful for researchers who study urinary proteomics, but also for physicians and pathologists who are interested in renal diseases or other diseases that may cause changes in urinary proteins. The pathophysiology of diseases may be related by biomarker similarities. Because of the natural relationship between urinary proteins and the urinary system, this database may be especially suitable for studying the pathophysiological processes of urological diseases.

**EXPERIMENTAL PROCEDURES**

**Data Collection**—Information about urinary biomarkers or biomarker candidates was compiled from literature listed in PubMed. Proteomic and nonproteomic studies on the urine obtained from both humans and animals were included. To ensure the quality of the database, all publications were reviewed manually to extract disease and protein information.

For each protein in the database, the UniProtKB-Swiss-Prot/TrEmBL (14) ID and International Protein Index (IPI) (15) ID were identified, if not mentioned in the literature, from the IPI cross-reference data files (IPI human release 3.74, IPI rat release 3.74, and IPI mouse release 3.74), the UniProt online ID mapping server, or by querying the protein name directly in the UniProt website.

A human plasma proteome data set comprised of 3020 proteins with two or more distinct peptide identifications was downloaded from the Human Proteome Project (HUPO) Plasma Proteome Project website (http://www.bioinformatics.med.umich.edu/hupo/ppp) (16). This data set was used to determine whether a urinary biomarker candidate was a plasma protein. Additionally, putative orthologous protein clusters were downloaded from the IntegR8 web portal (17).

**Network Analysis**—To cluster the human core disease-disease network, the free software CFinder was used to first identify densely connected subgroups based on the weighted Clique Percolation Method (CPMw) (18, 19). Additional manual adjustments were then...
made. The clustering process was completely based on the topological structure of the network without any consideration of the real world relationship between the diseases. The network was plotted using Cytoscape 2.6.0 (20).

RESULTS

General Information—By manually curating the data obtained from all literature related to urinary biomarker studies, two data sets were setup: a human data set and an animal data set. Any time a protein or its fragment peptide was reported to have differential expression between normal and disease conditions or among different disease stages, it was defined as a biomarker and embodied in the database.

In the human data set, five “negative” records were included. In these records, no statistically significant changes in protein expression level were observed between the disease and control groups. These records were included because the same proteins were identified as biomarkers for the same diseases in other studies.

In each record, the following information was collected (if available): the definition of the disease and control groups, the number of samples, detection methods and fold change of the protein and its fragment, any variants, post-translational modifications, experimental molecular weight, and experimental pl. The UniProtKB ID and IPI ID of the proteins were also identified as described in the Experimental Procedures section. In addition, the plasma proteome list from the HUPO Plasma Proteome Project was also used to determine the origin of these proteins. The website now allows users to browse the records by navigating the disease trees and search proteins by their ID numbers in bulk.

| TABLE I
| Human | Animal |
|-------|-------|-------|
|       | Rat   | Mouse | Cat  |
| Literatures | 174   | 22    | 8    | 1    |
| Records   | 553   | 224   | 50   | 1    |
| Diseases  | 56    | 20    | 8    | 1    |
| Biomarkers| 496   | 197   | 49   | 1    |
| Proteins  | 320   | 140   | 48   | 1    |

Data Statistics—The human data set consists of biomarker information for 10 cancers, 31 urological diseases, and 15 nonurological diseases (supplemental Table S1a); whereas the animal data set consists of biomarker information for 21 urological and four nonurological disease models (supplemental Table S1b). Detailed statistical data are shown in Table I.

The UPB database compiled biomarkers identified by both proteomic methods (two-dimensional electrophoresis, liquid chromatography-tandem MS, etc) and nonproteomics methods (Western blot, ELISA, etc). There were 50 and 17 proteomic studies in the human and animal data set, respectively. Fig. 1 classified the records in the two data sets by detection methods. Of these, 46 and 61% of the biomarkers in the human and animal data set were identified by proteomic methods without any validation, respectively, whereas only 33 and 11 biomarkers (88 and 32 records) were identified by more than one study, respectively.

Overlap Among Biomarkers Identified by Different Proteomic Methods—An interesting question is how well the biomarkers that were identified by different proteomic methods overlapped. In the human data set of the UPB database, eight diseases had more than one proteomic study. To evaluate the overlap rate, biomarkers from all studies related to the same disease were compared pairwise. A total of 66 literature pairs and 370 records were involved in the comparison. Remarkably, only 17 biomarkers were co-identified in seven literature pairs. The overlap rate was defined as the number of co-identified biomarkers divided by the average number of biomarkers identified in the two literature pairs. The average overlap rate was 0.02 ± 0.07, indicating that biomarkers identified by different publications of proteomics studies overlapped poorly in the human data set. In the animal data set, two diseases were studied twice, respectively. Because of a lack of data for comparison, the similarity of biomarkers identified by different proteomic studies based on animal models could not be evaluated.

Many factors may cause the poor overlap rate seen with proteomic experiments, such as different sample preparation and separation methods, mass spectrometers, and data anal-
ysis algorithms. Therefore, the poor overlap rate may not necessarily indicate that the biomarkers identified by proteomic methods are not reliable. However, the poor overlap rate does cause difficulty in the validation of the proteomic results. Biomarker candidates identified by high throughput proteomic methods have to be validated by methods of lower throughputs, and therefore the entire efficiency of the biomarker discovery process is reduced.

Overlap Among Biomarkers Identified from Human and Animal Samples—To study how well animal models can mimic human diseases, biomarkers from the human and animal data sets were compared. Surprisingly, although 90% of rat genes possessed strict orthologs in both mouse and human genomes (21), the similarity of their proteomes was low. Putative orthologous protein clusters were downloaded from the Integr8 web portal (17). Among all of the 140 rat proteins and 48 mouse proteins in our database, only 80 and 25 of them had orthologous proteins in human, respectively. Table II shows that very few of the biomarkers identified by animal models were also identified in the human data set.

It was also interesting to elucidate how biomarkers identified from rat and mouse models overlapped. As shown in Tables III and IV diseases were studied by both animal models, and only a few biomarkers were identified for each disease in each data set. No significant difference was observed between the human-animal and the rat-mouse overlap rates.

Considering that the overlap rate among different proteomic studies inside the human data set was not higher than the interorganism overlap rates, no clear conclusion can be made on how well these animal models mimic real human diseases or on how similar the rat and mouse models are to each other.

Disease Specificity of Biomarkers—As discussed in the previous section, the reliability of transforming animal model studies to human diseases is difficult to assess. Therefore, we only used the human data set to study the disease specificity of biomarkers.

A total of 242 proteins were found to be associated with only one disease. However, 143 of these proteins were reported in only one publication and identified by proteomic methods only. These potential biomarkers were in relatively low confidence, because experimental or interindividual variations of the samples cannot be ruled out. Therefore, we focused on the disease specific biomarkers that were identified in at least two different

| Disease                          | Biomarkers (human) | Biomarkers (animal) | Orthologs in human | Co-identified biomarkers | Overlap rate |
|----------------------------------|--------------------|---------------------|--------------------|--------------------------|--------------|
| Acute kidney injury              | 17                 | 21                  | 11                 | 1                        | 0.052        |
| Chronic allograft nephropathy    | 1                  | 1                   | 0                  | 0                        | 0            |
| Diabetic nephropathy             | 50                 | 2                   | 2                  | 1                        | 0.038        |
| Interstitial cystitis            | 6                  | 14                  | 9                  | 1                        | 0.1          |
| Nephrotoxicity                   | 4                  | 30                  | 20                 | 2                        | 0.105        |

| Disease                           | Biomarkers (rat) | Biomarkers (mouse) | Co-identified biomarkers | Overlap rate |
|-----------------------------------|------------------|--------------------|--------------------------|--------------|
| Acute kidney injury               | 3 (2)            | 2 (0)              | 0                        | 0            |
| Cisplatin nephrotoxicity          | 3 (3)            | 1 (1)              | 0                        | 0            |
| Diabetic nephropathy              | 2 (2)            | 1 (1)              | 1                        | 0.667        |
| FSGS                              | 20 (10)          | 36 (16)            | 1                        | 0.036        |

| Disease                           | Biomarkers       |
|-----------------------------------|------------------|
| Bladder cancer                    | BLCA-4, nuclear matrix protein 22, survivin, UBC, CYFRA21–1 |
| Prostate cancers                  | prostate specific antigen (PSA) |
| Pancreatitis                      | trypsin-2, trypsinogen activation peptide (TAP) |
| Diabetic nephropathy              | ceruloplasmin; connective tissue growth factor (CTGF); prostaglandin-H2 D-isomerase |
| Necrotizing enterocolitis         | intestinal fatty acid binding protein (I-FABP) |
minuria, compared with healthy controls. However, Narita et al. (25) reported that a protein increased in the urine of diabetes patients, which they suggested to be related to fibrotic renal disease. The up-regulation of other plasma proteins is not clear whether this process would affect the quantification of disease severity (28, 29). In type 2 diabetic patients with nephropathy; however, the results from separate studies did not agree. In a two-dimensional Difference Gel Electrophoresis (DIGE) study, Paturi et al. (22) reported that ceruloplasmin decreased in diabetic patients with normoalbuminuria, and that prostatic acid phosphatase decreased in diabetic patients with macroalbuminuria, compared with healthy controls. However, Narita et al. (23) and Jiang et al. (24) observed increases in these proteins using an immunoradiometric assay and two-dimensional DIGE method, respectively. The different results of these studies may have been caused by several factors, such as the criteria used for sample selection or the sample preparation methods. For example, Paturi et al. depleted six major plasma proteins before two-dimensional DIGE analysis, and it is not clear whether this process would affect the quantification of other plasma proteins.

Connective tissue growth factor (CTGF) has been reported to be related to fibrotic renal disease (25), and the up-regulation of CTGF can be induced by high glucose (26) and may contribute to chronic tubulointerstitial fibrosis (27). CTGF may be a reliable biomarker to detect kidney injury in patients with diabetes, and both publications stored in our database studied the expression level of CTGF in patients with type 1 diabetes and nephropathy.

In the case of necrotizing enterocolitis (NEC), I-FABP, which is a marker of intestinal mucosal cell damage, was reported as a promising biomarker for both the diagnosis and prediction of disease severity (28, 29).

We also sought to identify proteins that had been shown to be biomarkers for multiple diseases. A total of 76 proteins were found to be associated with more than one disease. Table V lists the top five biomarkers shared by multiple diseases. All of these proteins, with the exception of albumin, have molecular weights smaller than the glomerular filtration cut-off of ~45 kDa. In addition, all of the proteins, with the exception of β-2-microglobulin, are plasma proteins identified by the Human Plasma Proteome Project. β-2-microglobulin and neutrophil gelatinase-associated lipocalin (NGAL) are well-known specific biomarkers for proximal tubular injury (30, 31), but in our database, they were also associated with diseases not originating from the kidney. The relationship between NGAL and sepsis is easy to explain, because sepsis can induce acute kidney injury. In addition, because serum β-2-microglobulin has been reported to be elevated in a number of cancers (32), including colorectal cancer (33), the up-regulation of this protein in urine can be explained by the increase of its abundance in serum. To our knowledge, albumin, zinc-α-2-glycoprotein, and α-1-microglobulin/bikunin precursor (AMBP) proteins have not been reported to be produced by injury sites in the kidney, and because all of these proteins are plasma proteins, their changes in expression levels in urine may reflect changes in the serum, abnormal function of glomerular filtration, or abnormal tubular reabsorption in the kidney.
this network might include a number of false positive connections. Therefore, a human core network, without the biomarkers that were identified by only one proteomic method or based on less than 10 samples, was constructed. This data set presumably had a higher confidence level. The human core network included 40 diseases and 112 connections. The free software CFinder was used to explore subgroups in this network. A subgroup was defined as a group where nodes inside it were more densely connected to each other than to nodes outside the subgroup. A total of seven subgroups were identified and are shown in Fig. 2 with different colors. Edge line widths indicate the connection weights, i.e., the number of biomarkers shared by the two disease nodes.

Subgroup 1 consists of seven renal diseases together with sepsis, a disease that may cause kidney injuries. Several biomarkers of tubular injury, such as NGAL, interleukin-18, and β2-microglobulin (30, 31) are shared in this group, indicating that these renal diseases may all have tubular injuries. Subgroup 2 is the biggest subgroup in this network, consisting of 8 renal diseases and ovarian cancer. Biomarkers shared by ovarian cancer and renal diseases include albumin, kininogen-1, and osteopontin. To date, the roles of these proteins in the disease process are still not clear, and therefore it is difficult to explain why ovarian cancer is densely connected to the clusters of renal diseases. Subgroup 3 is another subgroup that consists of renal diseases and one nonurological disease, acute viral hepatitis E. Albumin is the only biomarker shared by all of the diseases, and the expression level of albumin increases in renal diseases and decreases in acute viral hepatitis E. Because albumin is produced in the liver, its expression level in serum decreases during liver damage, which subsequently causes a decrease in the urinary expression level. Because the function of albumin in acute viral hepatitis E is so different from that in renal diseases, this subgroup is only meaningful in the topology of the network, but not in the biological relationship among diseases. Subgroup 4 consists of four renal diseases. The shared biomarker is kidney injury molecule-1 (KIM-1), another well-known biomarker for tubular injury (30, 31). Subgroup 5 consists of three renal diseases and colorectal cancer, with β2-microglobulin being the shared biomarker. As mentioned in the previous section, β2-microglobulin is not only a urinary biomarker for tubular injuries, but also a serum biomarker for cancers, and its changes in serum can also be reflected in urine. Subgroup 6 and 7 consist of urological diseases that are not limited to renal diseases. Biomarkers shared by these subgroups include chemokines (mcp-1 in subgroup 6; IL-6 and IL-8 in subgroup 7), indicating that inflammation is the common biological process of these diseases.

Kidney transplantation is a large category in the human core data set, consisting of six diseases. However, unlike the relationship among all types of acute kidney injuries or other renal diseases, these nodes rarely connect to each other, but do connect to other renal or nonrenal diseases. Note that although kidney transplantation is divided into six subclasses, they have not yet been studied in great detail and only a small number of biomarkers have been identified to date. It is pos-
sible that the topological structure of diseases in this category will change when these diseases are extensively studied.

**DISCUSSION**

*The Study of Pathophysiological Processes of Diseases Using This Database*—The database presented in this study is also a useful bioinformatics tool for the study of the pathophysiology of diseases, because diseases that share biomarkers may have the same injury sites or pathophysiological processes. For a “new” disease where the pathogenesis or injury sites are not clear (for example, a new drug with unknown toxicity), if the fold changes of urinary proteins caused by this disease are known, researchers can query the protein list in the database to link this disease to other diseases that cause similar fold changes in these proteins. The injury site, pathophysiological process, and severity of the “new” disease can then be inferred by its relationship to the other diseases.

**Database Confidence**—In this database, all of the urinary proteins that have been reported to be differentially expressed in disease conditions were defined as biomarkers. Approximately half of the records in this database were identified by proteomic methods used only once without any validation, which are usually thought to exhibit some false positives because of the limitation of proteomics quantification technologies or small sample size. In the analysis of the human disease-disease network, we found that nodes in the human network were too densely connected to each other, which made it difficult to cluster the entire network into subgroups whereby the diseases within the subgroups had a large probability of having similar pathophysiological processes. Therefore, a human_core data set was subsequently constructed by eliminating records that were identified by proteomic methods only once or records that were based on less than 10 samples. The disease-disease relationships observed in the human_core networks were more reasonable than in the entire human network. This phenomenon may be in agreement with the hypothesis that the results from proteomic methods without validation have relatively low confidence; however, these observations do not provide direct evidence for this hypothesis. Nevertheless, no filters of biomarker confidence were used because we wanted to preserve the original results from the literature and make the database as comprehensive as possible. Information such as detection methods and sample sizes in the database may help users assess the data confidence in their own unique way.

**Manual Curation or Text Mining?**—All of the disease and protein information in the database were curated manually to ensure that minimum mistakes were made in the process of data collection. No text mining methods were used, because to the best of our knowledge, no text mining methods are adequate for the retrieval of detailed descriptive information compiled in our database, such as the definitions and sample sizes of disease and control groups. One problem with manual curation is that it requires substantial time and labor. It is possible that proper text mining methods could be used to speed up future updates. Nevertheless, all of the records in our database will be either manually curated or validated to ensure the highest quality in the data collection process.

**CONCLUSION**

A database of urinary protein biomarkers was constructed by manually curating literature from PubMed. We strived to embody all of the existing studies regarding urinary biomarkers in this database and did not filter the data by the confidence level. Approximately half of the records in this database were identified by proteomic methods and were only reported once. The biomarkers that were identified by different proteomic methods overlapped poorly with each other. The interorganism (human vs. animal and rat vs. mouse) overlap rates were also very low. Network analysis method was used to explore some interesting relationships and shared biomarkers between diseases. Presently, only 13 disease-specific biomarkers were studied by multiple reports in the literature, while others still lack in-depth study and validation. We believe that with the advancement of urinary biomarker discovery, more comprehensive and accurate data will be compiled in this database and additional biomarkers with enough confidence and disease specificity will be identified for clinical use.

* This work was supported by the National Basic Research Program of China (973 Program, 2012CB517606 and 2011CB649401), National Science Fund for Distinguished Young Scholar (30725009), National High Technology Research and Development Program of China (863 Program, 2011AA020116), Program for Changjiang Scholars and Innovative Research Team in University-PCSIRT (IRT0909), 111 project (B08007), National Natural Science Foundation (30870502) and the Dean’s Grant from Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/School of Basic Medicine, Peking Union Medical College (2010PY220, 2009PY05).

This article contains supplemental Fig. S1 and Table S1.

To whom correspondence should be addressed: Chinese Academy of Medical Sciences/School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, Beijing, China. Tel: +86-10-6529-6493; Fax: +86-10-6521-2284; E-mail: gaoyouhe@pumc.edu.cn.

**REFERENCES**

1. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol.* 7, R80
2. Li, S. J., Peng, M., Li, H., Liu, B. S., Wang, C., Wu, J. R., Li, Y. X., and Zeng, R. (2009) Sys-BodyFluid: a systematical database for human body fluid proteome research. *Nucleic Acids Res.* 37, D807–912
3. Rosner, M. H. (2009) Urinary biomarkers for the detection of renal injury. *Adv. Clin. Chem.* 49, 73–97
4. Schaub, S., Wilkins, J. A., and Nickerson, P. (2008) Proteomics and renal transplantation: searching for novel biomarkers and therapeutic targets. *Contrinv. Nephrol.* 160, 65–75
5. Vrooman, O. P., and Witjes, J. A. (2008) Urinary markers in bladder cancer. *Eur. Urol.* 53, 909–916
6. Montagut, C., Albanell, J., and Bellmunt, J. (2009) Prostate cancer. Multidisciplinary approach: a key to success. *Crit. Rev. Oncol. Hematol.* 68, S32–36
7. Rossing, K., Mischak, H., Dakka, M., Züribig, P., Novak, J., Julian, B. A., Good, D. M., Coon, J. J., Tarnow, L., and Rossing, P. (2008) Urinary proteomics in diabetes and CKD. *J. Am. Soc. Nephrol.* 19, 1283–1290
8. Weissinger, E. M., Hertenstein, B., Mischak, H., and Ganser, A. (2005)
Online coupling of capillary electrophoresis with mass spectrometry for the identification of biomarkers for clinical diagnosis. *Expert Rev. Proteomics* 2, 639–647

9. Kaiser, T., Kamal, H., Rank, A., Kolb, H. J., Holler, E., Ganser, A., Hertenstein, B., Mischak, H., and Weissinger, E. M. (2004) Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation. *Blood* 104, 340–349

10. Zimmerli, L. U., Schiffer, E., Zürbig, P., Good, D. M., Kellmann, M., Moulis, L., Pitt, A. R., Coon, J. J., Schmieder, R. E., Peter, K. H., Mischak, H., Kolch, W., Delles, C., and Dominiczak, A. F. (2008) Urinary proteomic biomarkers in coronary artery disease. *Mol. Cell Proteomics* 7, 290–298

11. Oh, J., Pyo, J. H., Jo, E. H., Hwang, S. I., Kang, S. C., Jung, J. H., Park, E. K., Kim, S. Y., Choi, J. Y., and Lim, J. (2004) Establishment of a near-standard two-dimensional human urinary proteome map. *Proteomics* 4, 3485–3497

12. Khan, A., and Packer, N. H. (2006) Simple urinary sample preparation for proteomic analysis. *J. Proteome Res.* 5, 2824–2838

13. Coon, J. J., Züürbig, P., Dakna, M., Dominiczak, A. F., Decramer, S., Fliesser, D., Frommerberger, M., Golovko, I., Good, D. M., Herget-Rosenthal, S., Jankowski, J., Julian, B. A., Kellmann, M., Kolch, W., Massy, Z., Novak, J., Rossing, K., Sandran, J. P., Schiffer, E., Theodorescu, D., van Halder, R., Weissinger, E. M., Mischak, H., and Schmitt-Kopplin, P. (2008) CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin. Appl.* 2, 964

14. (2009) *The Universal Protein Resource (UniProt)* 2009. *Nucleic Acids Res.* 37, D169–174

15. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) The International Protein Index: an integrated data-handling system for unifying protein sequence information. *Nature* 435, 2824–2838

16. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R. A., McPherson, J., Zhao, S., Fraser, C. M., Shetty, J., Shatsman, S., Geer, L., Nekrutenko, O., Amanatides, P. G., Baden-Tillson, H., Barnstead, M., Kasif, S., Ohlrich, J., Bransford, J., Amos, J., Eberhart, A., Hishinuma, A., Soller, S., Ostrander, E., White, B., Beck, M., Bittner, M., Slatkin, M., and McKusick, V. A. (1999) The CEPH linkage map 1999. *Science* 286, 2233–2236

17. Apweiler, R. (2004) The International Protein Index: an integrated data-handling system for unifying protein sequence information. *Nature* 435, 2824–2838

18. Palla, G., Dereńyi, I., Farkas, I., and Vicsek, T. (2005) Uncovering the *Mol. Cell Proteomics* 2, 639–647

19. Farkas, I., Belá, D., Palla, G., and Vicsek, T. (2007) Weighted network modules. *New J. Phys.* 9, 180

20. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Lesiak, B., Singer, A., Yu, J., Lu, P., Wufing, P., Zhu, Y., Zhang, J., Sun, H., Zhang, S., Sanganakar, A., Tew, K. D. (2001) Glutathione S-transferases as emerging therapeutic targets. *Drug Discov. Today* 6, 297–302

21. Palla, G., Dereńyi, I., Farkas, I., and Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818

22. Waanders, F., Navis, G., and van Goor, H. (2010) Urinary tubular biomarkers of kidney damage: potential value in clinical practice. *J. Cancer* 94, 1898–1905

23. van Goor, H., Navis, G., and van der Giezen, M. (2010) Urinary tubular biomarkers of kidney disease: potential value in clinical practice. *Am. J. Kidney Dis.* 55, 813–816

24. Ferguson, M. A., Vaidya, V. S., and Bonventre, J. V. (2008) Biomarkers of nephrotic syndrome: emerging as a promising cancer therapeutic target. *Drug Discov. Today* 14, 25–29

25. Ward, D. G., Suggett, N., Cheng, Y., Wei, W., Johnson, H., Billingham, L. J., Ismail, T., Wakeman, M., Johnson, P. J., and Martin, A. (2006) Identification of serum biomarkers for colon cancer by proteomic analysis. *Br. J. Cancer* 94, 1898–1905
In order to cite this article properly, please include all of the following information: Shao, C., Li, M., Li, X., Wei, L., Zhu, L., Yang, F., Jia, L., Mu, Y., Wang, J., Guo, Z., Zhang, D., Yin, J., Wang, Z., Sun, W., Zhang, Z., and Gao, Y. (2011) A Tool for Biomarker Discovery in the Urinary Proteome: A Manually Curated Human and Animal Urine Protein Biomarker Database. *Mol. Cell. Proteomics* 10(11):M111.010975. DOI: 10.1074/mcp.M111.010975.