

# The KUPKB: a novel Web application to access multiomics data on kidney disease

Julie Klein,<sup>\*,†,1</sup> Simon Jupp,<sup>‡</sup> Panagiotis Moulos,<sup>\*,†</sup> Myriem Fernandez,<sup>\*,†</sup>  
Bénédicte Buffin-Meyer,<sup>\*,†</sup> Audrey Casemayou,<sup>\*,†</sup> Rana Chaaya,<sup>\*,†</sup> Aristidis Charonis,<sup>§</sup>  
Jean-Loup Bascands,<sup>\*,†</sup> Robert Stevens,<sup>‡</sup> and Joost P. Schanstra<sup>\*,†,1</sup>

\*Institut National de la Santé et de la Recherche Médicale, U1048, Institut of Cardiovascular and Metabolic Disease, Toulouse, France; †Université Toulouse III Paul-Sabatier, Toulouse, France; ‡School of Computer Science, University of Manchester, Manchester, UK; and §Section of Histology, Center for Basic Research I, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

**ABSTRACT** The information gathered from the large number of omics experiments in renal biology is underexplored, as it is scattered over many publications or held in supplemental data. To address this, we have developed an open-source Kidney and Urinary Pathway Knowledge Base (KUPKB) that facilitates simple exploration of these omics data. The KUPKB currently comprises 220 data sets (miRNA, mRNA, proteins, and metabolites) extracted from existing publications or databases. Researchers can explore the integrated data using the iKUP browser, and a simple template is provided to submit new omics data sets to the knowledge base. As an example of iKUP's use, we show how we identified, *in silico*, calreticulin as a protein induced in human interstitial fibrosis and tubular atrophy (IFTA) in chronic kidney transplant rejection; a link that would have been difficult to establish using existing Web-based tools. Using immunohistochemistry, we validated *in vivo* this *in silico* result in human and rat biopsies of IFTA, thus identifying calreticulin as a potential new player in chronic kidney transplant rejection. The KUPKB provides a simple tool that enables users to quickly survey a wide range of omics data sets and has been shown to facilitate rapid hypothesis generation in the context of renal pathophysiology.—Klein, J., Jupp, S., Moulos, P., Fernandez, M., Buffin-Meyer, B., Casemayou, A., Chaaya, R., Charonis, A., Bascands, J.-L., Stevens, R., Schanstra, J. P. The KUPKB: a novel Web application to access multiomics data on kidney disease. *FASEB J.* 26, 000–000 (2012). [www.fasebj.org](http://www.fasebj.org)

*Key Words:* renal pathophysiology • genomics • proteomics • bioinformatics • experimental data

THE WEALTH OF information produced by omics experiments is still largely underexplored, as it is scattered across publications, comes in different formats and versions, and is hidden in figures and supplemental data. Moreover, most publications only show or focus on a limited part of the results, as the large numbers of genes or proteins involved far exceed those of direct or immediate interest to the researchers. Finally, the results of one experiment should be seen in the context of other experiments that may show the same or related entities exhibiting a variety of regulations. As a result, most of the omics data are inaccessible by search tools such as Medline (<http://www.ncbi.nlm.nih.gov/pubmed/>) or Google (<http://www.google.com>) and therefore cannot be readily reused. In addition, although data standardization efforts for reporting have been established [*e.g.*, Minimum Information about a Microarray Experiment (MIAME; <http://www.mged.org/Workgroups/MIAME/miame.html>), Minimum Information about a Proteomics Experiment (MIAPE; <http://www.psidev.info/index.php?q=node/91>), *etc.*], accessing and comparing these data for the everyday biologist is long, difficult, and error prone (1).

Research in the renal field is no exception. Large-scale experimental techniques, such as microarray transcriptomics, allow parallel analysis of modifications of large numbers of molecules in different parts of the kidney, across different diseases, and sometimes from rare samples, such as kidney biopsies. Several attempts have been made in the renal field to make some of the data available *via* databases on the Web, such as the

Abbreviations: EMT, epithelial to mesenchymal transition; HK-2, human kidney 2; HMDB, Human Metabolome Database; IFTA, interstitial fibrosis and tubular atrophy; iKUP, Kidney and Urinary Pathway Knowledge Base interface; KUPKB, Kidney and Urinary Pathway Knowledge Base; KUPO, Kidney and Urinary Pathway Ontology; MCNS, minimal change nephrotic syndrome; Na/K-ATPase, sodium/potassium-ATPase; OWL, Web Ontology Language; RDF, Resource Description Framework; SPARQL, SPARQL Protocol and RDF Query Language; TGF- $\beta$ , transforming growth factor- $\beta$ .

<sup>1</sup> Correspondence: INSERM U1048, Institut of Cardiovascular and Metabolic Disease, 1 Avenue Jean Poulhes, Toulouse Cedex 4, 31432 France. E-mail: J.P.S., [joost-peter.schanstra@inserm.fr](mailto:joost-peter.schanstra@inserm.fr); J.K., [julie.klein64@gmail.com](mailto:julie.klein64@gmail.com)  
doi: 10.1096/fj.11-194381

This article includes supplemental data. Please visit <http://www.fasebj.org> to obtain this information.

Human and Kidney Proteome Project (ref. 2; <http://www.hkupp.org/>), the GenitoUrinary Development Molecular Anatomy Project (ref. 3; <http://www.gudmap.org/>), the Inner Medullary Collecting Duct Proteome Database (ref. 4; <http://dir.nhlbi.nih.gov/papers/lkem/cdpd/>), or Nephromine (<http://www.nephromine.org/>). Despite these honorable efforts, these databases only provide access to subsets of information, are devoted to only one technology of one omics field, and are scattered across the Web.

Grouping the different omics data sets (different biological levels, different cells, and different diseases or models) is difficult without an adequate annotation mechanism to reduce the semantic heterogeneities between them (5, 6). To facilitate the reuse and exploration of omics data in the field of renal research, we present a publicly accessible and updatable repository that we have called the Kidney and Urinary Pathway Knowledge Base (KUPKB; <http://www.kupkb.org>) that currently contains 220 annotated omics data sets. We provide a search tool to these data *via* the KUPKB interface (iKUP) website, allowing researchers to query and explore these data. We encourage the community to contribute new data and take advantage of this use of novel technology to facilitate data integration and promote future reuse of these data. To demonstrate that the KUPKB clearly helps to develop new hypotheses in the renal research domain, we present 2 examples. The first example explains how the KUPKB helped to identify for the first time calreticulin as a new potential player in chronic allograft nephropathy [also known as interstitial fibrosis and tubular atrophy (IFTA)]. Calreticulin has been recently identified as a potential actor in the fibrotic process, as Kypreou *et al.* (7) have shown that gene and protein expression was induced in the rat model of unilateral ureteral obstruction. However, until now, access to the expression of calreticulin in omics experiments in other animal models or in human nephropathies was not available through classic research tools. We thus used the KUPKB to find out whether modified expression of calreticulin was observed in renal pathologies with fibrosis. In the second example, the KUPKB, in combination with Medline, allowed us to understand atypical sodium/potassium-ATPase (Na/K-ATPase) localization in our experimental conditions. *In vivo*, renal epithelial cells possess an apical and a basolateral domain, and this polarization state is tightly linked to their function and modifies cellular response to stimuli. For example, the Na/K-ATPase pump, which transports Na<sup>+</sup> out and K<sup>+</sup> into the cell, is strictly localized at the basolateral membrane of renal polarized cells. Renal cell lines such as human kidney 2 (HK-2) cells are widely used in *in vitro* experiments. We demonstrate here that culture conditions (*i.e.*, with or without gelatin coating) influence the distribution of Na/K-ATPase and that the KUPKB helped us to explain this result.

Through the KUPKB and iKUP browser, we have been able to generate new *in silico* hypotheses in the renal field that have been validated in the laboratory. It

is the integration of various –omic data sets and the access through a Web interface that has allowed these hypotheses to be made.

The KUPKB has been developed using a novel approach to publishing data on the Web. So-called semantic web technologies (8) are used to ensure that the data are described in a way that enables reuse and integration of the data with future applications in the KUP domain. This approach provides a showcase for the technology and a working framework that could be readily adapted for other biological domains (*e.g.*, cardiovascular research and cancer).

## MATERIALS AND METHODS

### Ontologies

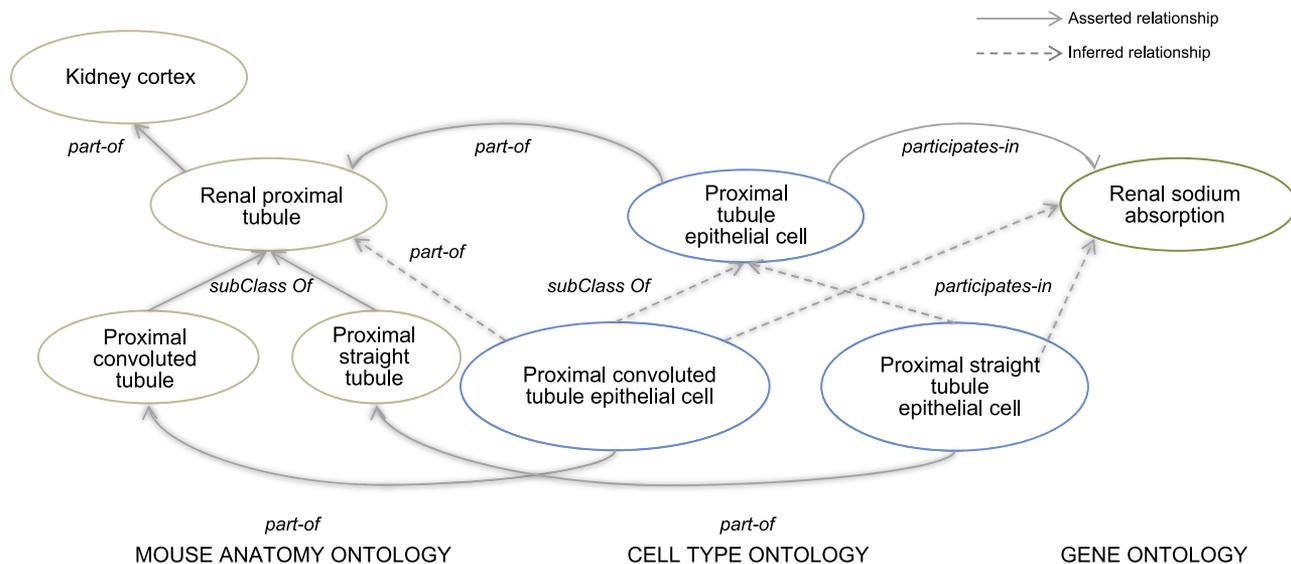
The KUPKB uses an ontology written in the Web Ontology Language (OWL) called the KUP Ontology (KUPO) as a schema to organize the data about KUP entities and from experiments in the KUP domain (Fig. 1 and ref. 9). Ontologies describe the types of entity in a domain and the relationships between those entities (10). Ontologies for the life sciences are continuing to grow, and coverage for specialized domains like the KUP domain is ongoing through efforts like the Renal Gene Ontology Working Group (11). The KUPO reuses many existing ontologies, thus capitalizing on the work of others and reducing the effort involved in creating the KUPKB. The KUPO uses the Adult Mouse Gross Anatomy Ontology (12), the Cell Type Ontology (13), and the Gene Ontology (14), as well as fragments of ontologies for disease and experiments (transcriptomic, proteomic, and metabolomic). A detailed method of the development of the KUPO can be found in the study of Jupp *et al.* (9).

The KUPKB is modeled in OWL and stored in the Resource Description Framework (RDF). RDF is a World Wide Web Consortium standard for publishing data on the Web (15). Many life science data sets are becoming available in RDF, offering greater opportunities for data integration (16, 17). Using standard formats for data exchange, such as RDF and OWL, offers new possibilities for data integration within the KUP community.

To provide a user-friendly interface to browse the RDF data in the KUPKB, we developed the iKUP browser. iKUP allows users to browse through the KUPKB, narrowing their search to focus on particular proteins or genes in particular locations, with particular functional attributes. More complex queries can be executed using SPARQL Protocol and RDF Query Language (SPARQL).

### KUPKB background knowledge

The KUPKB has integrated external resources, such as Human Metabolome Database (HMDB), National Center for Biotechnology Information (NCBI) Gene (18), Universal Protein Resource (UniProt; ref. 19), MicroCosm (20), and HomoloGene (21). Various mappings of identifiers were done to integrate these different data sets into a single queryable resource. This additional knowledge layer provides an opportunity to expand user queries on a given gene symbol to include both the genes and the proteins but also to link the orthologs across different species.



**Figure 1.** Example of entity relationships and organization in KUPO. Cells are related to anatomy and biological process concepts. By using an ontology language like OWL, we can infer new relationships between concepts based on asserted information; this ability to perform inference over the data can be exploited when querying the KUPKB.

### KUPKB experimental data

An initial set of 220 experiments has been included in the KUPKB. These experiments span different biological levels (miRNA, mRNA, proteins, and metabolites), different techniques, different species (mice, rats, and humans), and different sample types (urine or kidney tissue). Data were derived from published articles in Medline and have been extracted manually from the figures or from the supplementary data when possible and then programmatically added to the KUPKB. No further selection of the data has been performed before integration. Each gene, miRNA, protein, and metabolite name has been converted to a stable identifier, *i.e.*, Human Metabolome DataBase HMDB ID for metabolites, NCBI Gene ID for genes, UniProt ID for proteins, and MicroCosm ID for miRNA, using the open-source MadGene (ref. 22; <http://cardioserve.nantes.inserm.fr/mad/madgene/>). Some raw microarray data sets have also been downloaded from the Gene Expression Omnibus and analyzed by the open-source software Gene Armada (ref. 23; <http://www.grissom.gr/armada/>) and Bioconductor.

### Na/K-ATPase localization

HK-2 cells (American Type Culture Collection, Manassas, VA, USA) were cultured on an iPCellCulture membrane (it4ip S.A., Seneffe, Belgium) with or without 1% gelatin coating for 12 d in DMEM/F12 (1:1; Life Technologies, Carlsbad, CA, USA), 1% penicillin/streptomycin, 10 ng/ml EGF, 5 µg/ml insulin, 36 ng/ml hydrocortisone, 4 pg/ml T3, and 10% fetal calf serum (Life Technologies). For immunofluorescence staining, cells were washed with PBS, fixed with 4% PFA, and permeabilized with 0.3% Triton/0.2% BSA. After saturation of nonspecific antigenic sites, anti-Na/K-ATPase primary antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) was incubated for 1 h at room temperature.

### Animal model of IFTA

Male Sprague-Dawley rats (8 wk old, weighing ~250 g; Harlan ZI Du, Moulon, France) with free access to normal salt diet and water *ad libitum* were used. Experiments were conducted

in accordance with the European Communities Council Directive (86/609/EEC) for experimental animal care and were approved by the local animal care and use committee. (UMS US006, Institut National de la Santé et de la Recherche Médicale, Rangueil, Purpan, PreCREFRE, Toulouse, France).

Surgery was performed as described previously (24). Briefly, while the rats were under anesthesia with isoflurane/oxygen inhalation (3%/97%), they were subjected to unilateral nephrectomy. At 7 d after unilateral nephrectomy, rats were anesthetized with sodium pentobarbital (60 mg/kg ip), and ischemia was induced by clamping the right renal pedicle for 45 min using atraumatic vascular microclamps (Arex, Palaiseau, France). After clamp removal, the right kidney was inspected for restoration of blood flow. At 24-h after reperfusion, all rats were treated daily with ciclosporin A for 28 d. Sham-treated animals were subjected to the same surgical procedure (uninephrectomy and ciclosporin A treatment) without clamping the renal pedicle.

### Human kidney specimens

We retrospectively analyzed paraffin sections of renal biopsies from patients with either minimal change nephrotic syndrome (MCNS) or kidney transplantation (postunclamping biopsy or grade 2 IFTA). Informed consent was given by the patients or their parents for use of part of the biopsy for scientific purposes. All procedures and use of tissue were performed according to national ethical guidelines and were in accordance with the Declaration of Helsinki.

### Histological analysis and immunohistochemistry

Routine histology was performed using paraffin-embedded biopsy samples. Sections (4 µm) were cut and used for routine staining (hematoxylin-eosin and periodic acid-Schiff staining) and immunohistochemistry. All human biopsies were scored (from 0+ to 3+) by an expert pathologist masked to the clinical data to evaluate the score of IFTA. For immunohistochemistry, human and rat renal tissues were first dewaxed in toluene and rehydrated through a series of graded ethanol washes before endogenous peroxidase blockage. Specific primary antibody was incubated (45 min at room

temperature) for the detection of calreticulin (Abcam, Cambridge, MA, USA). The Dako Envision system (Dako, Glostrup, Denmark) was used for visualization. Sections were finally counterstained with hematoxylin. The percentage of fibrosis was estimated by Sirius red. Histo-morphometric analyses were performed as described previously (25), using a commercially available image-analysis software that allows rebuilding of a kidney section from adjacent individual captures (Mosaic software; Explora Nova, La Rochelle, France). The number of stained pixels was quantified for each section.

### Statistical analyses

Data are expressed as means  $\pm$  SD. Mann-Whitney test was performed for comparison between 2 groups. Values of  $P < 0.05$  were considered statistically significant.

## RESULTS

### Populating the KUPKB

At the time of writing, the KUPKB comprises 220 miRNA, mRNA, protein, phosphoprotein, and metabolite expression data sets ([http://www.kupkb.org/etc/summary\\_KUPKBcontent.html](http://www.kupkb.org/etc/summary_KUPKBcontent.html)) extracted from KUP-related publications (figures, tables, and supplemental data); from specific KUP-related databases, including EureGene, the Human Glomerulus Protein Database, or the Urinary Exosome Protein Database; or from the generalistic database repository of high-throughput gene expression data, the Gene Expression Omnibus. Experiments include expression data in specific nephron segments (*e.g.*, glomerulus), kidney cells (*e.g.*, podocyte), or urine; across different kidney diseases or animal models (*e.g.*, diabetic nephropathy, IFTA, and polycystic kidney disease); and across different *in vitro* models [*e.g.*, transforming growth factor- $\beta$  (TGF- $\beta$ ) treatment, hypoxia, and high-glucose treatment].

To allow users to submit data for inclusion in the KUPKB, we developed a spreadsheet template where cell content is limited to specific terms of KUPO (the user can choose from a list of appropriate terms) so the data can be annotated in a straightforward and reproducible way (see Submit Data tab at <http://www.kupkb.org/>). These data can then be automatically added to the KUPKB *via* its standard pipelines (9).

### Querying the KUPKB using the iKUP browser

The user can search the KUPKB using the iKUP browser for a molecule or a list of molecules of interest. The result of the query is returned in a tabular format to view, at a glance, the current information in the KUPKB that has been drawn from omics experiments. The KUPKB is publicly available and will remain so, and users are encouraged to submit their published data using the spreadsheet template (see preceding section and Fig. 2A).

To demonstrate the use of the KUPKB and iKUP browser in general renal research, we have taken two

real-life examples from our laboratories: we are interested in the role of calreticulin in the development of renal fibrosis (7), and we used the KUPKB to find out whether modified expression of calreticulin was observed in renal pathologies with fibrosis; and we routinely use, as do many laboratories, HK-2 cells for mechanistic studies and observed atypical cellular localization of transporters during attempts to determine whether these cells can be polarized. We showed how we used the KUPKB to explain this atypical expression.

### Calreticulin in the KUPKB

Previous work by Kypreou *et al.* (7) demonstrated that calreticulin was induced *in vivo* in the unilateral ureteral obstruction model of renal fibrosis and *in vitro* in response to TGF- $\beta$ , a major profibrotic molecule. To find out whether modified expression of calreticulin was observed in other renal pathologies, we entered “calreticulin and chronic kidney disease” and “calreticulin and kidney fibrosis” in Medline. The search did not return any relevant result, except for the publication by Kypreou *et al.* (7). We next searched for “calreticulin” into the iKUP browser. Among other results, the search revealed that calreticulin mRNA was induced by TGF- $\beta$  (ref. 26 and Fig. 3A, line 1), confirming the result of Kypreou *et al.* (7). The protein phosphorylation status was reduced in the same conditions (ref. 27 and Fig. 3A, line 5), an effect associated with the profibrotic processes in the literature (28). Another result showed that glomerular calreticulin protein expression was observed to be unmodified in a mouse model of diabetic nephropathy (ref. 29 and Fig. 3A, line 4). Finally, the most striking result was that calreticulin appeared in the KUPKB to be increased in kidney biopsies of patients with IFTA (ref. 30 and Fig. 3A, line 2). It is important to point out that the original publications used for these examples were not aiming to study calreticulin. Information about the expression of calreticulin was embedded in the supplemental data among thousands of other proteins; thus, querying Medline alone for calreticulin and IFTA did not return this information. It was only after integrating this data set into the KUPKB and viewing the data alongside other similar data sets that we were able to unravel this finding.

We also verified the presence in the KUPKB of molecules known to be modified in IFTA (Supplemental Table S1). Eight molecules [TGF- $\beta_1$ , serpine 1 (PAI1), BMP7, COLA1, COL3A1, CD44, vimentin, and  $\beta_2$ -microglobulin] were found both in the KUPKB and in Medline. Five other molecules known to be involved in IFTA were not found in the KUPKB [S100A4, ACTA2 ( $\alpha$ SMA), HAVCR1 (KIM-1), CCL2, and ADAM17].

### Calreticulin protein is increased in human and rat biopsies of IFTA

To validate the *in silico* result that calreticulin was induced in IFTA, we decided to confirm the proteomics

A

**KUPKB** *Beta*  
The Kidney & Urinary Pathway Knowledge Base

IKUP Browser About Submit Data Contact Acknowledgements FAQ

Search: ATP1B1 ATP1B2

The KUPKB is a collection of omics datasets that have been extracted from scientific publications and other related renal databases. The IKUP browser provides a single point of entry for you to query and browse these datasets.

Simply enter your gene, protein or miRNA of interest into the query box and press search. You can search for multiple entities per line and we support a range of identifiers including entrez gene ids, gene names, uniprot ids and miRNA ids from MirBase DB. e.g. Search for TGFβ1 or transforming growth factor α 3172. We have currently collected over 160 experiments, a summary of all the experiments collected is available [here](#). If you would like to submit your own datasets please choose the submit data tab above.

**Results View**

The results table shows the KUPKB experiments that reference your search terms. You can sort the results table by clicking on the column headers. The navigation tree below gives you a summary of your results and can be used to filter the results table.

Add Filter Remove Filter Current Filter Set: TGFbeta in vitro model

Entity id	Species	Anatomy	Disease/Model	Expression	Experiment	Type
ATP1B2	Human	Kidney proximal tubule epithelial cell	TGFbeta in vitro model	Up	<a href="#">Hills, Mol Endocrinol. 2010</a>	mRNA
ATP1B1	Human	Kidney proximal tubule epithelial cell	TGFbeta in vitro model	Down	<a href="#">Hills, Mol Endocrinol. 2010</a>	mRNA

Cell (2)  
Somatic cell (2)  
Epithelial cell (2)  
Kidney cell (2)

Anatomy (0)

Disease/Model (2)  
In vitro model (2)

The Kidney and Urinary Pathway Knowledge Base has been developed in collaboration between the [Renal Fibrosis Laboratory](#) at INSERM, France and the [Bio-health Informatics Group](#) at the University of Manchester, UK. This work has been funded by [e-LICO project](#), an EU-FP7 Collaborative Project (2009-2012) Theme ICT-4.4: Intelligent Content and Semantics.

RF lab BHIG Inserm MANCHESTER IICO

B

Microarray analysis was performed to identify genes regulated by TGF- $\beta$ 1 (2ng/mL, 48h) in a human proximal tubular cell line, HK-2 (n=3/group). HillsCE et al.

Proinsulin C-Peptide Antagonizes the Profibrotic Effects of TGF- $\beta$ 1 via Up-Regulation of Retinoic Acid and HGF-Related Signaling Pathways. Molecular Endocrinology. 2010.

**Figure 2.** Screenshot of the iKUP browser. A) iKUP browser is a simple Web interface to query and retrieve information from the KUPKB. In this example, the user queried for 2 genes (ATP1B1 and ATP1B2), and the results came back in a tabular form that can be sorted based on the species, the anatomy/cell, the disease/model or the expression strength. Results can also be filtered by anatomy/cell and disease/model using the arborecence in the left panel (here the results have been filtered to keep only expression values in a “TGFbeta *in vitro* model”). Each result is linked to its publication, as the hyperlink in the Experiment column is connected to the Medline abstract. B) Short experiment description can be obtained by either mouseover or clicking on the ? button.

result with a targeted study of calreticulin protein expression in rat and human kidney biopsies of IFTA, using immunohistochemistry. As shown in Fig. 3B, calreticulin was significantly induced in the kidney cortex in a rat model of IFTA induced by the combination of ischemia/reperfusion and ciclosporin A treatment (24), and this effect was associated with increased renal fibrosis. We also studied calreticulin expression in renal biopsies of patients with grade 2 IFTA, and, although not significant due to the limited number of biopsies ( $P=0.07$ ), calreticulin expression was increased compared with controls (Fig. 3C).

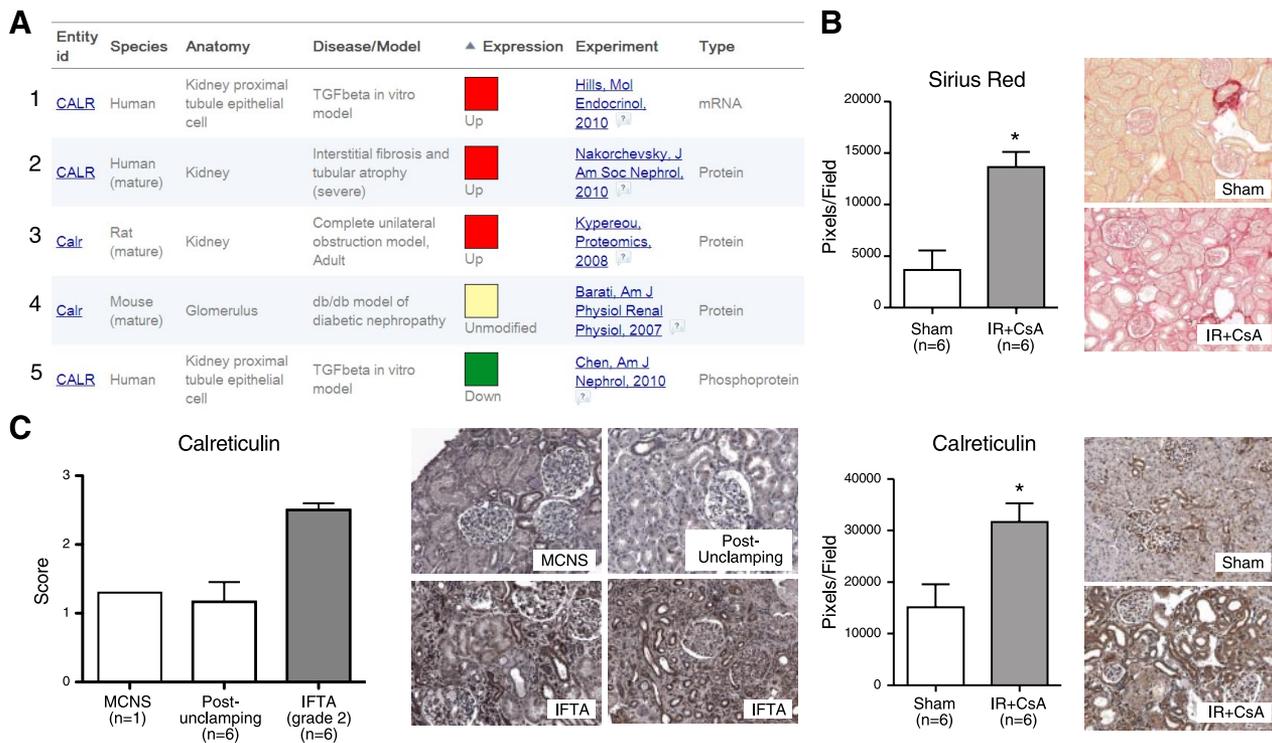
We therefore confirmed the *in silico* data obtained using the KUPKB on both rat and human biopsies of IFTA. The use of the KUPKB thus unburied the knowledge that calreticulin is induced in IFTA.

### Na/K-ATPase in the KUPKB

We next investigated whether *in vitro* culture conditions (*i.e.*, with or without gelatin coating) might influence distribution of Na/K-ATPase and how this could be

explained using the KUPKB. We first cultured HK-2 cells in gelatin-coated plates. Surprisingly, in these conditions, the cells expressed an Na/K-ATPase localized at the apical membrane of the cells (Fig. 4A), indicating an inappropriate polarization state of the HK-2 cells. Na/K-ATPase is a heterodimeric complex formed by the assembly of 2  $\alpha$  (catalytic) and 2  $\beta$  (chaperone) subunits (31). It has been shown that atypical localization of the Na/K-ATPase to the renal cell apical membrane could be induced by expression of the alternative  $\beta_2$  subunit (ATP1B2) instead of the common  $\beta_1$  isoform (ATP1B1; ref. 31). To understand what stimuli might induce such an imbalance, we entered a search for ATP1B1 and ATP1B2 into the iKUP browser (Fig. 2A).

The search revealed, among other data, that the  $\beta_1$  isoform mRNA was down-regulated, whereas the  $\beta_2$  was up-regulated in “kidney proximal tubule epithelial cell” in a “TGFbeta *in vitro* model” (ref. 26 and Fig. 2). The experiment description (Fig. 2B) showed that the data were obtained after a microarray experiment from HK-2 cells treated with 2 ng/ml TGF- $\beta$  for 48 h. These



**Figure 3.** Calreticulin expression is induced during IFTA. *A*) Query results returned by the iKUP browser and sorted for expression strength are summarized. *B*) Collagen accumulation (*i.e.*, fibrosis) and calreticulin expression were studied using Sirius Red staining and calreticulin immunohistochemistry, respectively, in a rat IFTA model. Representative pictures for sham-operated animals (uninephrectomy+ciclosporin A) and animals with uninephrectomy+ischemia/reperfusion+ciclosporin A for 28 d (IR+CsA) are shown for each staining, together with histological quantification ( $n=6$  animals/group). *C*) Calreticulin expression was studied by immunohistochemistry. Representative pictures for patients with grade 2 IFTA and controls (*i.e.*, biopsy of patient with MCNS and postunclamping/time 0 graft biopsies) are shown, together with histological quantification: MCNS,  $n=1$ ; postunclamping,  $n=3$ ; IFTA,  $n=3$ . Control staining using a primary, nonrelevant anti-HA rabbit antibody did not result in significant staining in either the rat or human samples (see Supplemental Fig. S1).  $*P < 0.05$  vs. sham treatment.

conditions were highly indicative of epithelial to mesenchymal transition (EMT; ref. 32) and suggest that the  $\beta$ -subunit isoform shifted from  $\beta_1$  to  $\beta_2$  during this process; a hypothesis confirmed by others (33). In the literature, it is known that renal cells exposed to culture plates coated with type I collagen are more prone to undergo EMT (34). It is noteworthy that gelatin used to coat the plates is mainly composed of type I collagen. Altogether, these results suggested that gelatin might be responsible for the atypical localization of the Na/K-ATPase by inducing molecular modifications associated with EMT. Indeed, we observed the first signs of the acquisition of a mesenchymal phenotype when HK-2 cells were grown on gelatin for 48 h (*e.g.*, increased vimentin and decreased E-cadherin expression; see Supplemental Fig. S2).

### Na/K-ATPase localization is restored by gelatin suppression

To test this hypothesis, we next studied Na/K-ATPase in HK-2 cells cultured without gelatin. Confocal microscopy results showed that in these conditions, the Na/K-ATPase was correctly localized to the basolateral membrane and in intracellular reticulum vesicles (Fig. 4C).

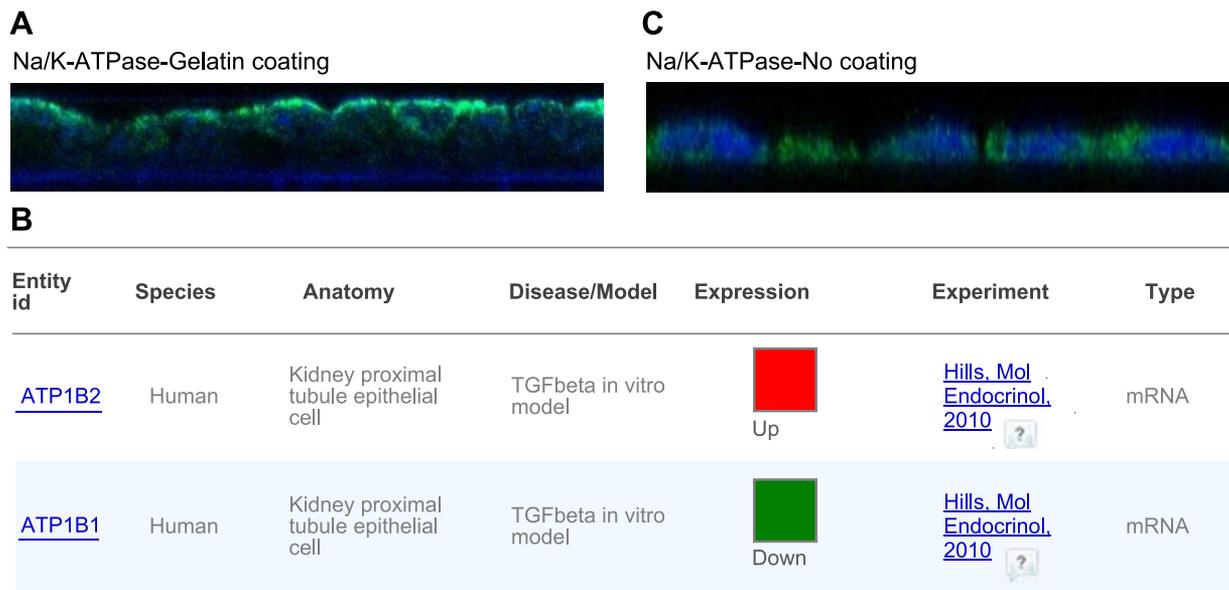
This example shows that the combination of Med-

line, KUPKB, and general knowledge of the field allowed the explanation of the atypical *in vitro* results and correcting targeting of Na/K-ATPase to the basolateral membrane by modification of culture conditions.

## DISCUSSION

The wealth of precious omics information is still largely underexploited, as it is often used only once for the experimenter's primary goal and then overlooked, as data are hidden in figures and supplementary data, making it harder for others to reuse these data. We thus developed, in the domain of renal research, a tool, the KUPKB, to gather, accurately annotate, and integrate scattered KUP data. The KUPKB, *via* the iKUP browser, is the first publicly available bioinformatics tool that can help researchers to exploit biological results to understand renal pathophysiology and assist biomarker discovery and molecular pathway modeling for diseases affecting the kidney and urinary pathways.

iKUP currently allows users to search for single or multiple genes, proteins, and miRNAs and rapidly survey a rich collection of experimental outcomes to see where and how these entities are expressed in terms of renal



**Figure 4.** Na/K-ATPase localization is modified by gelatin-induced EMT. A) Confocal microscopy analysis of Na/K-ATPase expression was performed on HK-2 cells cultured for 12 d on gelatin coating. B) Query results returned by the iKUP browser and filtered for “TGFbeta *in vitro* model.” C) Confocal microscopy analysis of Na/K-ATPase expression was performed on HK-2 cells cultured for 12 d without coating.

location and renal disease. We show in the KUPKB that by integrating these data, we can begin to find new discoveries in the data that previously would have been lost or at least very difficult to extract by hand.

We also integrated available metabolomics studies in the KUPKB, and search is available in iKUP for HMDB IDs or metabolite names. Metabolome analysis is still a novel approach in the KUP domain. Therefore, metabolomics data are scarce in kidney diseases, and the KUPKB only comprises 24 metabolomics data sets (coming from 5 publications) at this time. However, we expect that this technology will evolve rapidly, as urinary metabolomics is considered a very promising strategy for diagnosis or comprehension of kidney diseases (35).

Here we have used 2 examples to demonstrate how the iKUP browser may help researchers to develop easily new hypothesis in the renal research domain that were either not possible or more difficult to formulate using only classic tools, such as Medline. The first example demonstrated how calreticulin was confirmed *in vivo* as a new potential actor in IFTA, and this would not have been possible without the use of the KUPKB. It is beyond the scope of this article to discuss these results further, but altogether they suggest that a better understanding of the role of calreticulin might be of potential use in the context of IFTA. To our knowledge, this is the first time that calreticulin has been shown to be potentially involved in a human renal disease.

The second multisource example combined the KUPKB, Medline, and general knowledge from the field to hypothesize that gelatin-induced EMT might be also responsible for abnormal Na/K-ATPase localization, a result further confirmed *in vitro*. Here also, this hypothesis would not have been formulated and validated without the help of the KUPKB.

These two examples show how the KUPKB and iKUP browser can find their place in the renal research cycle. By allowing exploration of all the outputs of many types of high-throughput studies, the KUPKB and iKUP browser thus clearly represent significant progress in the state of the art in the field of nephrology.

Yet, the KUPKB is not complete. As shown in the example of IFTA (Supplemental Table S1), some of the molecules known to be involved in IFTA are seen in the KUPKB, while others are not. The reason might be either that these molecules are difficult to detect with omics approaches (in KUPKB) compared with targeted techniques (in Medline; *e.g.*, enzyme-linked immunosorbent assay *vs.* mass spectrometry, real-time polymerase chain reaction *vs.* microarrays) or simply that data are currently missing in the KUPKB and should be added over time. Note that some of the molecules that are defined as biomarkers of IFTA in Medline also appear in other diseases in the KUPKB. The KUPKB can thus provide valuable information on the specificity of a biomarker for a specific disease.

New data are published frequently. For this reason, we propose the KUPKB as a collaborative tool, as users can submit data for inclusion *via* a spreadsheet template. The populating of the KUPKB is a semiautomated process and is necessarily so. Where data already exist in a machine-readable form, such as a relational database, spreadsheet, or RDF store, they can be programmatically incorporated into the KUPKB. However, where the experimental data are in papers or some bespoke format as supplemental data, some manual intervention to render those data suitable for the automatic transformation into the KUPKB’s representation is necessary. For the KUP domain, the number of omic experiments is relatively small and has been

manageable in the development of the KUPKB. The need for manual intervention is unlikely to go away, but the spreadsheet template and the use of community crowd sourcing should ameliorate this problem.

The main limitation of the KUPKB is that there is no normalization across the data sets. Most of what can be seen in the KUPKB is based on trust with respect to what the authors published. While we could try to normalize across similar experiment types, *e.g.*, certain kinds of microarray experiments, it would be extremely difficult to do this kind of normalization for all experiment types. Our approach was to make as much information as possible available for a wide range of experiments and make clear links back to the original publications so the users can decide which results they want to trust. To help the users with this, we give some details about the analysis (extracted from the abstract or the materials and methods section) and the link to the full publication. Details on the analysis done by the authors can be found in the original publication, as each experiment has been linked to its respective PubMed identifier (PMID).

The iKUP browser will soon offer more options. We will provide an advanced query browser that will allow for a search of all the common genes, miRNAs, proteins, and metabolites that are differentially regulated in one or multiple diseases, locations, and/or biological process. Moreover, we aim to develop a visualization module that will enable the user to expose a selected set of molecules in a graphical way, displaying all the known interactions and relationships between these entities. In addition, we believe that although the KUPKB focuses on the kidney, the structure of the KUPKB and its easy-to-use interface will also be readily adaptable to any other biological discipline or organ. The core of the KUPKB can be easily populated with new data using the spreadsheet template. The KUPKB contains external resources, such as NCBI Gene, UniProt, HomoloGene, and MicroCosm. This additional knowledge layer provides opportunities to link miRNA, mRNA, and proteins when searching for a given entity but also to link the orthologs from different species. This information is universal and holds for any domain. Altogether, this makes the KUPKB easily adaptable for other biological domains.

In summary, the KUPKB combined with the iKUP browser provides a new resource for the KUP community where previously scattered information from the literature and existing databases, from different biological levels, different species, different portions of kidney anatomy, and different renal diseases has been integrated into a single repository for exploration. We expect this knowledge base to be of help to better understand complex results, to quickly retrieve information, and to generate new hypotheses in the context of renal pathophysiology. FJ

The authors thank EuroKUP members for participating in the KUPO day, in particular Antonia Vlahou, Jorgen Frokaier, Alberto Ortiz, and Harry Holthofer for useful suggestions; Michael Wicks and Duncan Davidson (Medical Research Council Human Genetics Unit, Edinburgh, UK) for supplying the

EuReGene data set; the RF lab (<http://www.renalfibrosis.fr>) for continuous support and constructive comments; and Auréja Steponavičiūtė and Aurélien Despierris for artistic participation in designing the logo and website. The authors thank Céline Gales [Institut National de la Santé et de la Recherche Médicale (INSERM) 1048, Team 8, Toulouse, France] for providing the rabbit anti-HA antibody and Laurent Sicard for help with immunohistology. This work is funded by the EU/FP7/ICT-2007.4.4 e-LICO project (<http://www.e-lico.eu>) and by the FP7-PEOPLE-2009-IAPP program Proclin (GA 251368). J.L.B. acknowledges support by INSERM and the Direction Régionale Clinique [Centre Hospitalier Universitaire (CHU) de Toulouse, Toulouse, France] under the Interface program. J.P.S. was supported by INSERM and the Direction de la Recherche Médicale et Innovation (CHU de Toulouse) under the Contrat Hospitalier de Recherche Translationnelle program. S.J., J.K., A.C., J.-L.B., and J.P.S. are members of COST action EuroKUP (<http://www.eurokup.org>). The KUPKB is available online (<http://www.kupkb.org>).

## REFERENCES

1. Stevens, R. D., Tipney, H. J., Wroe, C. J., Oinn, T. M., Senger, M., Lord, P. W., Goble, C. A., Brass, A., and Tassabehji, M. (2004) Exploring Williams-Beuren syndrome using myGrid. *Bioinformatics* **20**(Suppl. 1), i303–i310
2. Miyamoto, M., Yoshida, Y., Taguchi, I., Nagasaka, Y., Tasaki, M., Zhang, Y., Xu, B., Nameta, M., Sezaki, H., Cuellar, L. M., Osawa, T., Morishita, H., Sekiyama, S., Yaoita, E., Kimura, K., and Yamamoto, T. (2007) In-depth proteomic profiling of the normal human kidney glomerulus using two-dimensional protein prefractionation in combination with liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **6**, 3680–3690
3. Harding, S. D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-Macgilp, S., Pi, X., Roochun, Y., Sharghi, M., Tindal, C., McMahon, A. P., Gottesman, B., Little, M. H., Georgas, K., Aronow, B. J., Potter, S. S., Brunskill, E. W., Southard-Smith, E. M., Mendelsohn, C., Baldock, R. A., Davies, J. A., and Davidson, D. (2011) The GUDMAP database—an online resource for genitourinary research. *Development* **138**, 2845–2853
4. Pisitkun, T., Bieniek, J., Tchapyjnikov, D., Wang, G., Wu, W. W., Shen, R. F., and Knepper, M. A. (2006) High-throughput identification of IMCD proteins using LC-MS/MS. *Physiol. Genomics* **25**, 263–276
5. Davidson, S. B., Overton, C., and Buneman, P. (1995) Challenges in integrating biological data sources. *J. Comput. Biol.* **2**, 557–572
6. Goble, C., and Stevens, R. (2008) State of the nation in data integration for bioinformatics. *J. Biomed. Inform.* **41**, 687–693
7. Kypreou, K. P., Kavvadas, P., Karamessinis, P., Peroulis, M., Alberti, A., Sideras, P., Psarras, S., Capetanaki, Y., Politis, P. K., and Charonis, A. S. (2008) Altered expression of calreticulin during the development of fibrosis. *Proteomics* **8**, 2407–2419
8. Shadbolt, N. H. W., and Berners-Lee, T. (2006) The semantic web revisited. *Intelligent systems.* *IEEE* **21**, 96–101
9. Jupp, S., Klein, J., Schanstra, J., and Stevens, R. (2011) Developing a kidney and urinary pathway knowledge base. *J. Biomed. Semantics* **2**(Suppl. 2), S7
10. Stevens, R. A. M., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., and Rector, A. (2007) Using OWL to model biological knowledge. *Int. J. Human Comput. Studies* **65**, 583–594
11. Renal GO Working Group. *GO Annotation for the Renal System*. Available at: <http://www.geneontology.org/GO.renal.shtml>
12. Hayamizu, T. F., Mangan, M., Corradi, J. P., Kadin, J. A., and Ringwald, M. (2005) The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol.* **6**, R29
13. Bard, J., Rhee, S. Y., and Ashburner, M. (2005) An ontology for cell types. *Genome Biol.* **6**, R21
14. Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M.,

- Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., and White, R. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–261
15. World Wide Web Consortium. *W3C Recommendation*. Retrieved February 9, 2012, from <http://www.w3.org/TR/rdf-primer>
  16. Belleau, F., Nolin, M. A., Tourigny, N., Rigault, P., and Morissette, J. (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform.* **41**, 706–716
  17. Ruttenberg, A. R. J., Stephens, S., Samwald, M., and Cheung, K. H. A. (2008, June 4) *A Prototype Knowledge Base for the Life Sciences*. W3C Interest Group. Available at: <http://www.w3.org/TR/hcls-kb/>
  18. National Center for Biotechnology Information. *Entrez Gene*. Retrieved February 9, 2012, from <http://www.ncbi.nlm.nih.gov/gene>
  19. Universal Protein Resource Consortium. *UniProt Knowledge Base*. Retrieved February 9, 2012, from <http://www.uniprot.org>
  20. European Molecular Biology Laboratory–European Bioinformatics Institute. *MicroCosm Targets Version 5*. Retrieved February 9, 2012, from <http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/>
  21. National Center for Biotechnology Information. *HomoloGene*. Retrieved February 9, 2012, from <http://www.ncbi.nlm.nih.gov/homologene>
  22. Baron, D., Bihouee, A., Teusan, R., Dubois, E., Savagner, F., Steenman, M., Houlgatte, R., and Ramstein, G. (2011) MADGene: retrieval and processing of gene identifier lists for the analysis of heterogeneous microarray data sets. *Bioinformatics* **27**, 725–726
  23. Chatziioannou, A., Moulos, P., and Kolisis, F. N. (2009) Gene ARMADA: an integrated multi-analysis platform for microarray data implemented in MATLAB. *BMC Bioinformatics* **10**, 354
  24. Chaaya, R., Alfarano, C., Guilbeau-Frugier, C., Coatrieux, C., Kesteman, A. S., Parini, A., Fares, N., Gue, M., Schanstra, J. P., and Bascands, J. L. (2011) Pargyline reduces renal damage associated with ischaemia-reperfusion and cyclosporin. *Nephrol. Dial. Transplant.* **26**, 489–498
  25. Klein, J., Gonzalez, J., Decramer, S., Bandin, F., Neau, E., Salant, D. J., Heeringa, P., Pesquero, J. B., Schanstra, J. P., and Bascands, J. L. (2010) Blockade of the kinin B1 receptor ameliorates glomerulonephritis. *J. Am. Soc. Nephrol.* **21**, 1157–1164
  26. Hills, C. E., Willars, G. B., and Brunskill, N. J. (2010) Proinsulin C-peptide antagonizes the profibrotic effects of TGF-beta1 via up-regulation of retinoic acid and HGF-related signaling pathways. *Mol. Endocrinol.* **24**, 822–831
  27. Chen, Y. X., Li, Y., Wang, W. M., Zhang, W., Chen, X. N., Xie, Y. Y., Lu, J., Huang, Q. H., and Chen, N. (2010) Phosphoproteomic study of human tubular epithelial cell in response to transforming growth factor-beta1-induced epithelial-to-mesenchymal transition. *Am. J. Nephrol.* **31**, 24–35
  28. Mueller, C. F., Wassmann, K., Berger, A., Holz, S., Wassmann, S., and Nickenig, G. (2008) Differential phosphorylation of calreticulin affects AT1 receptor mRNA stability in VSMC. *Biochem. Biophys. Res. Commun.* **370**, 669–674
  29. Barati, M. T., Merchant, M. L., Kain, A. B., Jevans, A. W., McLeish, K. R., and Klein, J. B. (2007) Proteomic analysis defines altered cellular redox pathways and advanced glycation end-product metabolism in glomeruli of db/db diabetic mice. *Am. J. Physiol. Renal Physiol.* **293**, F1157–F1165
  30. Nakorchevsky, A., Hewel, J. A., Kurian, S. M., Mondala, T. S., Campbell, D., Head, S. R., Marsh, C. L., Yates, J. R., 3rd, and Salomon, D. R. (2010) Molecular mechanisms of chronic kidney transplant rejection via large-scale proteogenomic analysis of tissue biopsies. *J. Am. Soc. Nephrol.* **21**, 362–373
  31. Laughery, M. D., Clifford, R. J., Chi, Y., and Kaplan, J. H. (2007) Selective basolateral localization of overexpressed Na-K-ATPase beta1- and beta2- subunits is disrupted by butyrate treatment of MDCK cells. *Am. J. Physiol. Renal Physiol.* **292**, F1718–F1725
  32. Xu, Y., Wan, J., Jiang, D., and Wu, X. (2009) BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition in human renal proximal tubular epithelial cells. *J. Nephrol.* **22**, 403–410
  33. Rajasekaran, S. A., Huynh, T. P., Wolle, D. G., Espineda, C. E., Inge, L. J., Skay, A., Lassman, C., Nicholas, S. B., Harper, J. F., Reeves, A. E., Ahmed, M. M., Leatherman, J. M., Mullin, J. M., and Rajasekaran, A. K. (2010) Na,K-ATPase subunits as markers for epithelial-mesenchymal transition in cancer and fibrosis. *Mol. Cancer Ther.* **9**, 1515–1524
  34. Zeisberg, M., Bonner, G., Maeshima, Y., Colorado, P., Muller, G. A., Strutz, F., and Kalluri, R. (2001) Renal fibrosis: collagen composition and assembly regulates epithelial-mesenchymal transdifferentiation. *Am. J. Pathol.* **159**, 1313–1321
  35. Weiss, R. H., and Kim, K. (2011) Metabolomics in the study of kidney diseases. *Nat. Rev. Nephrol.* **8**, 22–33

Received for publication September 23, 2011.  
Accepted for publication January 30, 2012.