bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title Page

Regional differences of the urinary proteomes in healthy Chinese individuals

Jianqiang Wu [1], Weiwei Qin [2], Li Pan [3], Fanshuang Zhang [1], Xiaorong Wang [1], Biao Zhang [3], Guangliang Shan [3*], Youhe Gao [1,2*]

Affiliations:

[1]Department of Pathophysiology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, No.5 Dongdansantiao, 100005, Beijing, China;

[2]Department of Biochemistry and Molecular Biology, Gene Engineering and Biotechnology Beijing Key Laboratory, Beijing Normal University, No.19 Xinjiekouwai Street, 100875, Beijing, China

[3]Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, No.5 Dongdansantiao, 100005, Beijing, China

* **Corresponding author**: Youhe Gao & Guangliang Shan

Email addresses: ; gaoyouhe@bnu.edu.cn; guangliang_shan@163.com

Phone: 86-10-5880-4382; Fax: 86-10-6521-2284

List of abbreviations

LC-MS/MS, liquid chromatography tandem mass spectrometry; SBP, systolic blood pressure; DBP, diastolic blood pressure; Urimem, urine protein-bound membrane; NC, nitrocellulose membrane; total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), immunoglobulin G (IgG), immunoglobulin A (IgA), immunoglobulin M (IgM), alanine aminotransferase (ALT), serum total protein (TP), serum albumin (Alb), gamma glutamyl transferase (GGT), alkaline phosphatase (ALP), aspartate aminotransferase (AST), calcium (Ca), phosphorous (P), creatinine, urea nitrogen, fasting venous blood glucose (Glu) and uric acid (UA).

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Regional differences of the urinary proteomes in healthy Chinese individuals

Abstract

Urine is a promising biomarker source for clinical proteomics studies. Although regional physiological differences are common in multi-center clinical studies, the presence of significant differences in the urinary proteomes of individuals from different regions remains unknown. In this study, morning urine samples were collected from healthy urban residents in three regions of China and urinary proteins were preserved using a membrane-based method (Urimem). The urine proteomes of 27 normal samples were analyzed using LC-MS/MS and compared among the three regions. We identified 1,898 proteins from Urimem samples using label-free proteome quantification, of which 62 urine proteins were differentially expressed among the three regions. Hierarchical clustering analysis showed that inter-regional differences caused less significant changes in the urine proteome than inter-sex differences. Of the 62 differentially expressed proteins, 10 have been reported to be disease biomarkers in previous clinical studies. Urimem facilitates urinary protein storage for large-scale urine sample collection, and thus accelerates biobank development and urine biomarker studies employing proteomics approaches. Regional differences are a confounding factor influencing the urine proteome and should be considered in future multi-center biomarker studies.

Keywords: urine; proteomics; Urimem; biobank; biomarkers; regional differences; confounding factors

Introduction

Biomarkers are measurable changes associated with specific pathophysiological conditions. Unlike blood, which is controlled by homeostatic mechanisms, urine reflects systemic changes in the body and can be a more promising source of biomarkers (Gao, 2013; Wu and Gao, 2015). Currently, in addition to its application in urogenital diseases, urinary proteomics has been used in systematic diseases, such as cancers, cardiovascular diseases, diabetes, and mental disorders, to identify urinary disease biomarkers (Shao et al., 2011). More importantly, urine has potential uses in detecting small and early pathological changes in the very early stages of disease (Wu et al., 2017a; Wu et al., 2017b; Yin et al., 2017; Yuan et al., 2015; Zhang et al., 2017). For example, several urinary proteins change significantly even before a tumor mass is palpable, indicating that urine is an excellent source of biomarkers for early cancer detection (Wu et al., 2017a). Another study suggested that early changes in urinary proteins can be detected in a lung fibrosis model, which may lead to early treatment and a better prognosis (Wu et al., 2017b).

In the last decade, advances in proteomics techniques have accelerated the pace of biomarker research. Urinary proteome changes, some of which have been reported to be associated with disease, have been found in various systematic diseases both in animal models and patients. However, candidate biomarkers or biomarker panels need

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to be validated in large-scale, multi-center patient cohorts. Previous studies showed that several factors can influence the urine proteome (Bakun et al., 2014; Castagna et al., 2011; Kohler et al., 2010; Wu and Gao, 2015; Zhao et al., 2016). Regional differences are a common phenomenon in multi-center clinical studies. These differences may consist of genetic background, food consumption, lifestyle and environmental factors. However, whether there are significant differences in the urinary proteome among different regions remains unknown.

In this study, urine samples were collected from healthy Han residents in three urban areas of China (Haikou, Xi'an and Xining). These three cities are the capitals of Hainan, Shaanxi and Qinghai Provinces, respectively. Urimem is a product that can store urine proteins simply and economically, enabling the large-scale storage of clinical samples (Jia et al., 2014). In the current study, a large number of urine samples were collected at sites in the three regions, and urine proteins were stored using Urimem. Twenty-seven normal Urimem samples were selected and compared based on region to investigate the regional differences in the urinary proteome. Using label-free proteome quantification, a total of 1,898 urine proteins were identified from Urimem samples. A total of 62 urine proteins were differentially expressed among the three regions, of which 10 have been reported to be disease biomarkers. Our findings suggested that regional differences should be taken into consideration in biomarker identification and validation, particularly in multi-region and multi-center clinical studies. This study will not only benefit biobank development but also improve our understanding of biomarker discovery.

Results

Characteristics of the participants from the three regions of China

In this study, we recruited healthy urban residents from the capitals of three Chinese provinces (Hainan, Shaanxi, and Qinghai). The geographic distribution of these three regions is shown in **Figure 1**. Haikou is the capital of Hainan Province, located in southern China (longitude: 110.35, latitude: 20.02, elevation: 8 m); Xi'an is the capital of Shaanxi Province, located in central China (longitude: 108.95, latitude: 34.27, elevation: 415 m); and Xining is the capital of Qinghai Province, located in northwest China (longitude: 101.74, latitude: 36.56, elevation: 2250 m). The latitudes of Xi'an and Xining are similar but significantly higher than that of Haikou. The elevations of these three cities are significantly different.

To avoid potential interference from confounding factors, 27 healthy urban Han Chinese adults between 30 and 36 years of age with no chronic or metabolic diseases and no drug treatment within 2 the past weeks were randomly selected. Fasting blood samples were used for laboratory biochemical analyses. Detailed information on the 27 participants is listed in **Supplemental Table S1.** Age, sex, blood pressure and most blood biochemical indicators were similar among the healthy participants from the three areas. However, total protein (TP), albumin (ALB), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), fasting venous blood glucose (glucose level) and calcium (Ca) were significantly different ($p < 0.05$) in the blood of the

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

participants from these three cities (**Figure 2**). Serum protein concentrations, including TP and ALB, were lower in Xi'an residents than in residents from the other cities. The concentrations of serum lipids, including TC and LDL-C, were higher in the Haikou residents than in residents from the other cities. Additionally, glucose level in Xining residents was higher than in residents from Haikou. Meanwhile, serum concentration of Ca was significantly lower in residents from Xi'an than in residents from the other cities.

Urimem for proteomics analysis

The system of enriching and preserving urinary proteins on membranes using a vacuum filter bottle was first described in our previous study (Jia et al., 2014). Nitrocellulose (NC) membranes were used to absorb the urine proteins in the current study because their reproducibility is similar to protein isolation, and proteins can be easily eluted from them (Wang et al., 2014b). A total of 27 Urimem samples (9 samples from each geographic area) were randomly selected from healthy participants for proteome analysis. During the protein elution, acetone completely dissolved the NC membranes and precipitated the proteins at a low temperature. Urinary proteins were eluted from Urimem, digested with trypsin and then analyzed using one-hour LC-MS/MS. For quantification, Progenesis LC-MS software was used to process the MS data (ion detection, accurate mass retention time pair clustering, normalization and statistically analysis). Finally, a total of 1,898 urine proteins were identified from the Urimem samples using label-free identification (**Supplemental Table S2**).

Regional differences in the normal human urine proteome

To investigate the regional differences in the normal urine proteome, the Urimem samples from 27 healthy residents from three Chinese regions were analyzed by LC−MS/MS. The participants were matched by age, sex and ethnicity. A total of 62 differentially expressed proteins were identified among the three regions ($\geq$ 2 unique peptides, ANOVA $p < 0.05$, and fold change > 2). Proteins significantly changed among the three areas were further subjected to the following t-test analyses: Xi'an vs. Haikou, Xi'an vs. Xining and Haikou vs. Xining. There were 43 differential proteins between Xi'an and Haikou, 33 between Xi'an and Xining, and 6 between Haikou and Xining (**Supplemental Table S3**). Hierarchical clustering analysis of the differential proteins (ANOVA $p < 0.05$) also showed that samples from Xi'an were different from samples in the other two areas, and samples from Haikou and Xining could not be distinguished from one another (**Figure S1**). As shown in **Figure 3A**, there were 59 urine proteins that were differentially expressed between Xi'an and the other cities, 47 between Haikou and the other cities, and 38 between Xining and the other cities.

Sex is a confounding factor that may influence the urinary proteome. As shown in **Figure 3B**, a hierarchical clustering analysis of all identified proteins was performed to comprehensively define inter-regional and inter-sex differences in the urinary proteome. The results showed that all female samples were clustered into one group and that several samples in the same area could be clustered together. These results suggested that male and female urinary proteomes exhibited different patterns,

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and inter-regional differences were smaller than inter-sex differences. Between the male and female samples, a total of 53 urine proteins showed more than a 2-fold change and had a p-value < 0.05 (**Table S4**). Interestingly, several sex-specific proteins, such as highly expressed prostate-specific antigen and beta-microseminoprotein, were identified in the male samples.

Ten disease biomarkers are differentially expressed in the three regions

Of the 62 differentially expressed urine proteins among the three areas, 10 have been reported as disease biomarkers or biomarker candidates in clinical studies (**Table 1**). Peroxiredoxin-3 is involved in cell redox regulation and protects against oxidative stress. Recently, peroxiredoxin-3 was identified as a novel molecular biomarker for hepatocellular carcinoma (Ismail et al., 2015; Shi et al., 2014). Catalase protects cells from the toxic effects of hydrogen peroxide and is involved in oxidative damage. Catalase was reported as a candidate biomarker for colorectal cancer (Maffei et al., 2011), and this oxidative stress biomarker was also upregulated in the urine samples of diabetic patients (Kurutas et al., 2015). Fibrinogen beta chain has a major function in hemostasis, with increased expression reported in the urine samples of patients with deep vein thrombosis and pulmonary embolism (Klovaite et al., 2013; von Zur Muhlen et al., 2016). Alpha-1-acid glycoprotein 2 (Orosomucoid-2) is a candidate plasma biomarker for cholangiocarcinoma (Rucksaken et al., 2017), and is differentially expressed in the urine of pre-eclampsia and diabetes patients (Kronborg et al., 2007; Narita et al., 2004). Dihydropyrimidinase, neuropilin-1 and lysosome membrane protein 2 are disease biomarkers for cancers (Tse et al., 2017; Vasiljevic et al., 2014). Epidermal growth factor receptor kinase substrate 8-like protein 2 and syntenin-1 regulate migration, growth, proliferation, and cell cycle progression in a variety of cancers. These two proteins were reported to be differentially expressed in the urine samples of bladder cancer patients and to be potential cancer biomarkers (Smalley et al., 2008). N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase (AGA) was reported to be increased in the urine sample of prostatitis patients compared with healthy volunteers (Vermassen et al., 2015). Thus, in multi-center biomarker studies, it will be difficult to definitively determine whether potential disease biomarkers are truly associated with disease states or regional differences.

Functional analysis of differentially expressed proteins

To analyze the functions of the differentially expressed urine proteins, DAVID was used. Forty-three differential proteins between Xi'an and Haikou and thirty-three differential proteins between Xi'an and Xining were uploaded into the database respectively. The most frequently represented categories were biological processes, molecular functions and pathways. The results showed that the main biological processes were the oxidation-reduction process, glycolytic process and reactive oxygen species metabolic process. The main molecular functions included oxidoreductase activity and hydrolase activity, and main biological pathways included metabolic pathways and glycolysis/gluconeogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

In our present study, we investigated whether regional differences can significantly influence the human urinary proteome among different regions. Comparative proteomics analysis of Urimem samples from three areas in China was performed using label-free quantification. Our results indicate that a total of 62 urine proteins were differentially expressed among Xi'an, Hainan and Xining. Of which, 10 proteins were found to be disease biomarkers or biomarker candidates identified in previous clinical studies.

As shown in Figure 2, several blood biochemical indicators were significantly different in the blood of the participants from the three regions, which was consistent with the results of several previous studies. For example, the serum TP and ALB of a healthy population in Xi'an were significantly different from those of some other cities in China (Yin, 2008). Another study suggested that several serum indicators changed with increasing elevation in four areas of Yunnan Province, including BUN, Cr, and uric acid (UA) (Zi, 2014). In another recent study, significant differences in serum immunological indexes and the coagulation index of healthy people between Xi'an and Xining were reported, including a higher expression of IgG, IgM and IgA in Xining than in Xi'an (Gu, 2016). Our results in Table S1 are consistent with this study. Because blood and Urimem samples come from the same participants, and the urinary proteome contains a number of plasma proteins, changes in the blood composition are likely to have a potential impact on the urinary proteome.

In this study, a total of 62 differentially expressed urine proteins were identified among the three areas. The results also suggested that samples in Xi'an were significantly different from samples in the other two cities. Because sex has been reported to influence the urine proteome (Guo et al., 2015), we compared the effects of inter-regional and inter-sex differences on the urinary proteome. A total of 53 differentially expressed urine proteins were identified between the male and female samples. As shown in Figure 3B, the result of the unsupervised clustering analysis indicated that the inter-regional differences were smaller than the inter-gender differences. Previous studies demonstrated that some intrinsic factors (such as sex and age) and extrinsic influences (such as diet, exercise, and environmental factors) may influence the urine proteome, and intrinsic factors likely cause more significant changes to the urine proteome than extrinsic influences (Wu and Gao, 2015). Regional differences are attributable to genetic background, food consumption, lifestyle and environmental factors. In this study, the participants were matched by age, sex and ethnicity. Thus, regional differences in our study were likely due to extrinsic influences, and the result of the clustering analysis is reasonable.

In the present study, Urimem was used to preserve the urine proteins. The system of preserving urinary proteins on membranes using a vacuum filter bottle was first described in our previous study (Jia et al., 2014). The feasibility and validity of Urimem were proven in flow up studies (Qin and Gao, 2015; Wang et al., 2014a; Wang et al., 2014b). The findings showed that for protein absorption, NC membrane had a similar reproducibility to that of PVDF membranes. The proteome coverage of urinary proteins absorbed onto NC membranes has also been evaluated using

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LC-MS/MS analysis (Wang et al., 2014b). Recently, membrane-based urine protein absorption using a syringe-push system was also described as a simple rapid method for preparing urine for proteomics analysis (Chutipongtanate et al., 2015). These studies demonstrated that dry membranes can be stored at room temperature for an extended period of time are stable for further proteomic analysis. The long-term stability of membranes with dried adsorbed urine proteins allows specimen collection from areas without freezers and overcomes sample shipping problems, such as protein stability during sample shipping and shipping expenses due to weight. Membrane-based methods for urine protein preservation are quick and suitable for batch sample processing in large-scale clinical studies. Additionally, membranes with different properties can be used to selectively retain different substances from urine, such as DNA, metabolites and microRNAs (Zhang et al., 2015). Massive sample storage (biobanking) in a simple and economical manner is the foundation of future large-scale clinical precision medicine studies, and this membrane-based method may become one good option for urine sample processing and storage (Gao, 2015). Urimem is a membrane that can preserve urinary proteins simply and economically. This is the first application of Urimem in epidemiological studies. Urimem is a timesaving and suitable tool for large-scale and rapid sample preservation, especially in places without centrifugation or refrigeration facilities. The extensive use of Urimem in and out of hospitals will accelerate biobank development and future urine biomarker studies using proteomics approaches.

In this study, regional differences in normal urine proteome were evaluated in the urine samples of urban residents from three Chinese regions. Some urinary proteins were differentially expressed in different regions, and several proteins were reported in previous clinical studies to be disease biomarkers. The potential application of our findings in biomarker discovery is that regional differences should be taken into consideration in biomarker identification and validation, particularly in multi-region and multi-center clinical studies. The object of this study was to assess the presence of significant differences in the urinary proteomes of individuals from different regions. Due to the low throughput of MS, only 27 Urimem samples are used for the proteomics analysis, and the results should be confirmed in future studies with a larger sample size and additional populations in other regions.

Materials and Methods

Participants

The China National Health Survey (CNHS) was conducted to evaluate physiological constants and health conditions of the Chinese population. Research on the normal urine proteome of the healthy Chinese population is a subproject of CNHS. The study was approved by the local Ethics Committee of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (No. 029-2013), and signed informed consent was obtained from all participants. A total of 754 healthy volunteers between 10 and 80 years of age were enrolled in this study. Urine samples from 283 healthy urban residents (164 males and 119 females) in Haikou, Hainan Province,

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

China, were collected on December 27-29, 2013; urine samples from 255 healthy urban residents (121 males and 134 females) in Xi'an, Shaanxi Province, China, were collected on July 8-10, 2014; and urine samples from 216 healthy urban residents (128 males and 88 females) in Xining, Qinghai Province, China, were collected on August 15-17, 2015.

Data collection

A questionnaire was administered to obtain information from the participants, including sex, age, ethnicity, personal medical history, education level, physical activity, smoking status and drinking status.

Blood pressure was measured on seated participants using an Omron digital blood pressure measuring device. After a donor had fasted overnight, a venous blood sample was collected at the time of the interview. Blood samples were processed within 4 h of collection, shipped to a laboratory in Beijing, and stored at −80°C until analysis.

Blood biochemical and immune indicators, including total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), immunoglobulin G (IgG), immunoglobulin A (IgA), immunoglobulin M (IgM), alanine aminotransferase (ALT), serum total protein (TP), serum albumin (Alb), the ratio of serum albumin to globulin (AG), gamma glutamyl transferase (GGT), alkaline phosphatase (ALP), aspartate aminotransferase (AST), calcium (Ca), phosphorous (P), creatinine, urea nitrogen, fasting venous blood glucose (Glu) and uric acid (UA), were assessed at the Peking Union Medical College Hospital and Chinese PLA General Hospital.

Urinary protein preservation on a membrane

Approximately 50-100 mL of morning midstream urine was collected from each healthy participant. The urinary protein concentration was determined using routine urine tests. The urinary proteins were preserved on nitrocellulose (NC) membranes using a filtering device as previously reported (Jia et al., 2014). Briefly, after urine collection, 20 mL of each urine sample was mixed with 10 mL of phosphate buffer (1 M Na_2HPO_4, 1 M NaH_2PO_4, pH = 6). The diluted urine sample was then added to a vacuum filter bottle (10-cm^2 filter area). In this device, a polyvinylidene fluoride (PVDF) membrane (0.45 µm HV, Millipore, Tullagreen, Carrigtwohill, Ireland) with ultra-low protein-binding capacity was used to filter cell debris from the urine. Another 0.2-µm NC membrane (GE Healthcare, Freiburg im Breisgau, Germany) was used to adsorb urine proteins. Three sheets of filter paper were wetted with pure water and placed below the NC membrane. The vacuum filter bottle was connected to a vacuum pump, and the urine in the bottle passed through the PVDF membrane, while the urine protein in the flow-through was saved on the NC membrane. Then, the protein-bound NC membrane was dried with a hair dryer at room temperature and sealed in a sterile plastic bag with a label. The medical record number, sex, age, date of urine collection and blood pressure were recorded on the label. Urine samples from each participant were preserved on 2 or 3 NC membranes (Urimem).

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Elution of urinary proteins from the membranes

To avoid potential interference from confounding factors, Urimem membranes from healthy urban Han Chinese adults aged between 30 and 36 with no chronic or metabolic diseases (hypertension, diabetes, hepatitis and cancers) and no drug treatment within 2 weeks were selected. Additionally, blood biochemical indicators from each participant were reviewed to select healthy individuals whose blood indicators fell within the reference ranges for Chinese populations. To rule out potential kidney disease, participants without severe proteinuria were selected after a routine urine protein test. Finally, 9 participants (5 males and 4 females) from each geographic area were randomly selected for further proteome analysis.

Urinary proteins were eluted from the membranes according to a previously described method (Qin and Gao, 2015). The NC membrane with bound proteins was cut into small pieces and placed in a 2-mL Eppendorf tube. Then, 1.7 mL of acetone and 0.2 mL of 5% NH_4HCO_3 were added to the tube and vigorously vortexed for 30 s. The tubes were heated at 55°C for 60 min and were strongly vortexed for 30 s every 20 min. After the membrane was fully dissolved, the liquid was gently mixed using a rotary mixer at 4°C for 2 h to precipitate the protein. Then, the mixture was centrifuged at 12,000 x g for 15 min at 18°C. After the supernatant was removed, the tube was placed upside down on filter paper for 5 min at room temperature. The precipitate was re-suspended in lysis buffer (8 M urea, 2 M thiourea, 50 mM Tris and 25 mM DTT), and the protein concentration was determined using the Bradford method.

Tryptic digestion and LC-MS/MS analysis

A mass of 200 µg of urinary proteins eluted from each Urimem sample was digested with trypsin (Trypsin Gold, mass spectrometry grade, Promega) using the filter-aided sample preparation method with 10-kDa filtration devices (Pall, NY, USA) at 37°C overnight. The peptide mixtures were then desalted using Oasis HLB cartridges (Waters, Milford, MA, USA), dried by vacuum evaporation and re-dissolved in 0.1% formic acid.

For the LC-MS/MS analysis, 1 µg of peptides from each individual sample was loaded onto a trap column (75 µm × 2 cm, 3 mm, C18, 100 Å) and was separated on a reverse-phase analytical column (50 µm × 150 mm, 2 µm, C18, 100 Å) using the EASY-nLC 1200 HPLC system (Thermo Fisher Scientific). The elution for the analytical column was performed over 60 min at a flow rate of 300 µL/min. Then, the eluted peptides were analyzed using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). MS data were acquired in the high-sensitivity mode using the following parameters: data-dependent MS/MS scans per full scan with top-speed mode (3 s), MS scans at a resolution of 120,000 and MS/MS scans at a resolution of 30,000 in Orbitrap, 30% HCD collision energy, charge-state screening (+2 to +7), dynamic exclusion (exclusion duration 30 s) and a maximum injection time of 45 ms.

Label-free quantification

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Label-free quantitation of the proteomic data was performed using Progenesis LC-MS software (version 4.1, Nonlinear, Newcastle upon Tyne, UK) according to the procedure described by Hauck et al. (Hauck et al., 2010). Twenty-seven sample features were automatically aligned according to their retention times to maximally overlay all the two-dimensional feature maps (m/z and retention time). Features with charge states of +2 to +4 were selected for further analysis. After alignment and feature exclusion, the samples were divided into three groups (Haikou, Xi'an and Xining) using a between-subject design, and the raw abundances of all features were normalized. Normalization corrects for factors resulting from experimental variation and was automatically calculated over all features in all samples.

All MS/MS spectra were exported from the progenesis software, and the data were searched against the SwissProt Human database (08/2013, containing 20,267 sequences) using Mascot software (version 2.5.1, Matrix Science, London, UK). The parent ion tolerance was set at 10 ppm, and the fragment ion mass tolerance was set to 0.05 Da. A maximum of two missed cleavage sites in the trypsin digestion was allowed. Carbamidomethylation of cysteines was set as a fixed modification, and the oxidation of methionine was considered a variable modification. For quantification, all peptides with an ion score ≥ 30 and $p < 0.01$ were included, and the total cumulative abundance of a specific protein was calculated by summing the abundances of all peptides allocated to it. Calculations of the protein p value (one-way ANOVA) were then performed on the sum of the normalized abundances across all runs. ANOVA values of $p < 0.05$ and additionally regulation of $\geq$ 2-fold were regarded as significant for all further results.

Functional and pathway analyses

Functional annotation of the differentially expressed proteins among the three areas was analyzed using DAVID (Huang da et al., 2009). Protein classification was performed using gene ontology (GO) analysis for molecular functions and biological processes. Differentially expressed proteins were also uploaded into the Kyoto Encyclopedia of Genes and Genomes (KEGG) for pathway analysis.

Statistical analysis

Statistical analyses were performed with GraphPad Prism version 7.0 (GraphPad, San Diego, CA, USA). The clinical characteristics data are expressed as the means ± SDs. Comparisons among the three groups were conducted using a one-way ANOVA followed by multiple comparisons with the least significant differences (LSD) test. Group differences resulting in p-values < 0.05 were considered statistically significant.

Compliance and ethics

The study was approved by the local Ethics Committee of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (No. 029-2013), and signed informed consent was obtained from all participants. The authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2016 YFC 1306300), the National Basic Research Program of China (973 Program) (No. 2013CB530805, No. 2013FY114100), the Beijing Natural Science Foundation (No. 7173264, 7172076), the Fundamental Research Funds for the Central Universities (No.10300-31042110), and Beijing Normal University (No. 11100704).

We thank Wei Sun (Core Facility of Instrument, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences) for the technical assistance with the proteomic data analysis.

References

Bakun, M., Senatorski, G., Rubel, T., Lukasik, A., Zielenkiewicz, P., Dadlez, M., and Paczek, L. (2014). Urine proteomes of healthy aging humans reveal extracellular matrix (ECM) alterations and immune system dysfunction. Age (Dordr) 36, 299-311

Castagna, A., Olivieri, O., Milli, A., Dal Bosco, M., Timperio, A.M., Zolla, L., Gunasekaran, M., Raffaelli, R., Pizzolo, F., and Cecconi, D. (2011). Female urinary proteomics: New insight into exogenous and physiological hormone-dependent changes. Proteomics Clin Appl 5, 343-353

Chutipongtanate, S., Changtong, C., Weeraphan, C., Hongeng, S., Srisomsap, C., and Svasti, J. (2015). Syringe-push membrane absorption as a simple rapid method of urine preparation for clinical proteomics. Clin Proteomics 12, 15

Gao, Y. (2013). Urine-an untapped goldmine for biomarker discovery? Sci China Life Sci 56, 1145-1146

Gao, Y. (2015). Using fibers for rapid extraction of proteins from urine. Proteomics Clin Appl 9, 445-446

Gu, S. (2016). The change of lymphocyte subsets, platelet and coagulation function in healthy Person at Different Altitudes. Chinese Journal of Blood Transfusion 29, 700-703 (Article in Chinese)

Guo, Z., Zhang, Y., Zou, L., Wang, D., Shao, C., Wang, Y., Sun, W., and Zhang, L. (2015). A Proteomic Analysis of Individual and Gender Variations in Normal Human Urine and Cerebrospinal Fluid Using iTRAQ Quantification. PLoS One 10, e0133270

Hauck, S.M., Dietter, J., Kramer, R.L., Hofmaier, F., Zipplies, J.K., Amann, B., Feuchtinger, A., Deeg, C.A., and Ueffing, M. (2010). Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. Mol Cell Proteomics 9, 2292-2305

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57

Ismail, S., Mayah, W., Battia, H.E., Gaballah, H., Jiman-Fatani, A., Hamouda, H., Afifi, M.A., Elmashad, N., and Saadany, S.E. (2015). Plasma nuclear factor kappa B and serum peroxiredoxin 3 in early diagnosis of hepatocellular

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

carcinoma. Asian Pac J Cancer Prev 16, 1657-1663

Jia, L., Liu, X., Liu, L., Li, M., and Gao, Y. (2014). Urimem, a membrane that can store urinary proteins simply and economically, makes the large-scale storage of clinical samples possible. Sci China Life Sci 57, 336-339

Klovaite, J., Nordestgaard, B.G., Tybjaerg-Hansen, A., and Benn, M. (2013). Elevated fibrinogen levels are associated with risk of pulmonary embolism, but not with deep venous thrombosis. Am J Respir Crit Care Med 187, 286-293

Kohler, M., Walpurgis, K., Thomas, A., de Maree, M., Mester, J., Schanzer, W., and Thevis, M. (2010). Effects of endurance exercise on the urinary proteome analyzed by 2-D PAGE and Orbitrap MS. Proteomics Clin Appl 4, 568-576

Kronborg, C.S., Allen, J., Vittinghus, E., and Knudsen, U.B. (2007). Pre-symptomatic increase in urine-orosomucoid excretion in pre-eclamptic women. Acta Obstet Gynecol Scand 86, 930-937

Kurutas, E.B., Gumusalan, Y., Cetinkaya, A., and Dogan, E. (2015). Evaluation of method performance for oxidative stress biomarkers in urine and biological variations in urine of patients with type 2 diabetes mellitus and diabetic nephropathy. Biol Proced Online 17, 3

Maffei, F., Angeloni, C., Malaguti, M., Moraga, J.M., Pasqui, F., Poli, C., Colecchia, A., Festi, D., Hrelia, P., and Hrelia, S. (2011). Plasma antioxidant enzymes and clastogenic factors as possible biomarkers of colorectal cancer risk. Mutat Res 714, 88-92

Narita, T., Sasaki, H., Hosoba, M., Miura, T., Yoshioka, N., Morii, T., Shimotomai, T., Koshimura, J., Fujita, H., Kakei, M., and Ito, S. (2004). Parallel increase in urinary excretion rates of immunoglobulin G, ceruloplasmin, transferrin, and orosomucoid in normoalbuminuric type 2 diabetic patients. Diabetes Care 27, 1176-1181

Qin, W., and Gao, Y. (2015). [Elution of urinary proteins preserved on nitrocellulose membrane with heating]. Sheng Wu Gong Cheng Xue Bao 31, 1387-1392

Rucksaken, R., Charoensuk, L., Pinlaor, P., Pairojkul, C., Khuntikeo, N., and Pinlaor, S. (2017). Plasma orosomucoid 2 as a potential risk marker of cholangiocarcinoma. Cancer Biomark 18, 27-34

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, M111 010975

Shi, L., Wu, L.L., Yang, J.R., Chen, X.F., Zhang, Y., Chen, Z.Q., Liu, C.L., Chi, S.Y., Zheng, J.Y., Huang, H.X., Yu, F.J., and Lin, X.Y. (2014). Serum peroxiredoxin3 is a useful biomarker for early diagnosis and assessemnt of prognosis of hepatocellular carcinoma in Chinese patients. Asian Pac J Cancer Prev 15, 2979-2986

Smalley, D.M., Sheman, N.E., Nelson, K., and Theodorescu, D. (2008). Isolation and identification of potential urinary microparticle biomarkers of bladder cancer. J Proteome Res 7, 2088-2096

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tse, B.W.C., Volpert, M., Ratther, E., Stylianou, N., Nouri, M., McGowan, K., Lehman, M.L., McPherson, S.J., Roshan-Moniri, M., Butler, M.S., Caradec, J., Gregory-Evans, C.Y., McGovern, J., Das, R., Takhar, M., Erho, N., Alshalafa, M., Davicioni, E., Schaeffer, E.M., Jenkins, R.B., Ross, A.E., Karnes, R.J., Den, R.B., Fazli, L., Gregory, P.A., Gleave, M.E., Williams, E.D., Rennie, P.S., Buttyan, R., Gunter, J.H., Selth, L.A., Russell, P.J., Nelson, C.C., and Hollier, B.G. (2017). Neuropilin-1 is upregulated in the adaptive response of prostate tumors to androgen-targeted therapies and is prognostic of metastatic progression and patient mortality. Oncogene 36, 3417-3427

Vasiljevic, N., Ahmad, A.S., Thorat, M.A., Fisher, G., Berney, D.M., Moller, H., Foster, C.S., Cuzick, J., and Lorincz, A.T. (2014). DNA methylation gene-based models indicating independent poor outcome in prostate cancer. BMC Cancer 14, 655

Vermassen, T., Van Praet, C., Poelaert, F., Lumen, N., Decaestecker, K., Hoebeke, P., Van Belle, S., Rottey, S., and Delanghe, J. (2015). Diagnostic accuracy of urinary prostate protein glycosylation profiling in prostatitis diagnosis. Biochem Med (Zagreb) 25, 439-449

von Zur Muhlen, C., Koeck, T., Schiffer, E., Sackmann, C., Zurbig, P., Hilgendorf, I., Reinohl, J., Rivera, J., Zirlik, A., Hehrlein, C., Mischak, H., Bode, C., and Peter, K. (2016). Urine proteome analysis as a discovery tool in patients with deep vein thrombosis and pulmonary embolism. Proteomics Clin Appl 10, 574-584

Wang, M., Liu, X., Jia, L., Sun, Y., Gao, Y., and Li, M. (2014a). [Comparison of cost-effectiveness between Urimem and direct freezing for urinary protein preservation]. Sheng Wu Gong Cheng Xue Bao 30, 1128-1133

Wang, X., Li, X., Jia, L., and Gao, Y. (2014b). [Comparison of two urinary protein preparation methods: nitrocellulose membrane preservation and acetone precipitation]. Sheng Wu Gong Cheng Xue Bao 30, 982-989

Wu, J., and Gao, Y. (2015). Physiological conditions can be reflected in human urine proteome and metabolome. Expert Rev Proteomics 12, 623-636

Wu, J., Guo, Z., and Gao, Y. (2017a). Dynamic changes of urine proteome in a Walker 256 tumor-bearing rat model. Cancer Medicine, doi: 10.1002/cam4.1225.

Wu, J., Li, X., and Gao, Y. (2017b). Early detection in urinary proteome for the effective early treatment of bleomycin-induced pulmonary fibrosis in a rat model. http://www.paper.edu.cn/releasepaper/content/201701-209

Yin, B.W., Xiangling; Zhao, Lihua; Yue, Tianhai; Li, Lihua; Li, Burong; Fan, Xiaoyan; Zhang, Tong (2008). The investigation of reference values of serum proteins within healthy crowd in xi'an area. Hainan Medical Sciences 19, 11-13 (Article in Chinese)

Yin, W., Qin, W., and Gao, Y. (2017). Urine glucose levels are disordered before blood glucose level increase was observed in Zucker diabetic fatty rats. Sci China Life Sci, doi: 10.1007/s11427-017-9134-6.

Yuan, Y., Zhang, F., Wu, J., Shao, C., and Gao, Y. (2015). Urinary candidate biomarker discovery in a rat unilateral ureteral obstruction model. Sci Rep 5,

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9314

Zhang, F., Cheng, X., Yuan, Y., Wu, J., and Gao, Y. (2015). Urinary microRNA can be concentrated, dried on membranes and stored at room temperature in vacuum bags. PeerJ 3, e1082

Zhang, F., Ni, Y., Yuan, Y., Yin, W., and Gao, Y. (2017). Early urinary candidate biomarker discovery in a rat thioacetamide-induced liver fibrosis model. bioRxiv 125120

Zhao, M., Wu, J., and Gao, Y. (2016). The Specific alpha1-Adrenergic Receptor Antagonist Prazosin Influences the Urine Proteome. PLoS One 11, e0164796

Zi, J., Zi, Y., Li, Q., Zhang, Y., Tang, H. (2014). Analysis on biochemical indexes of healthy population in different altitude areas in Yunnan province of China. Journal of Plateau Medicine 24, 35-38 (Article in Chinese)

Figure legends

Figure 1. Geographical distribution of the participants enrolled in this study.
Xining (longitude: 101.74, latitude: 36.56, elevation: 2250 m); Xi'an (longitude: 108.95, latitude: 34.27, elevation: 415 m); Haikou (longitude: 110.35, latitude: 20.02, elevation: 8 m).

Figure 2. Blood biochemical indicators that differ among three areas of China.
Three areas in China: HN (Haikou City), XA (Xi'an City), and XN (Xining City). Abbreviations: TC (serum total cholesterol), LDL-C (low-density lipoprotein cholesterol), Urea (serum urea nitrogen), IgG (Immunoglobulin G), TP (serum total protein), ALB (serum albumin), Ca (calcium) and P (phosphorus). Comparisons between independent regions were conducted using a one-way ANOVA followed by post hoc analysis with the least significant differences test. *indicates $p < 0.05$, **indicates $p < 0.01$, and ***indicates $p < 0.001$.

Figure 3. The effects of inter-regional and inter-sex differences on the urine proteome. (A) Veen diagram of the differentially expressed urine proteins (fold change > 2 and $p < 0.05$) among the different regions. (B) Clustering analysis of all the identified urine proteins from the 27 Urimem samples. HN (Haikou), XA (Xi'an), XN (Xining), M (Male), and F (Female).

Additional files

Table S1. Characteristics of 27 participants according to location.

Table S2. Total urine proteins identified from the Urimem samples using label-free identification

Table S3. Differentially expressed urine proteins among the three areas.

Table S4. Gender differences in the urine proteome.

Figure S1. Hierarchical clustering analysis of the differentially expressed proteins among the three areas. Heatmap of differentially expressed urine proteins analyzed via ANOVA ($p < 0.05$). HN (Haikou City), XA (Xi'an City), and XN (Xining City).

The supporting information (Table S2-S4 and Figure S1) are uploaded at https://figshare.com/s/130588f37a1a79846976.

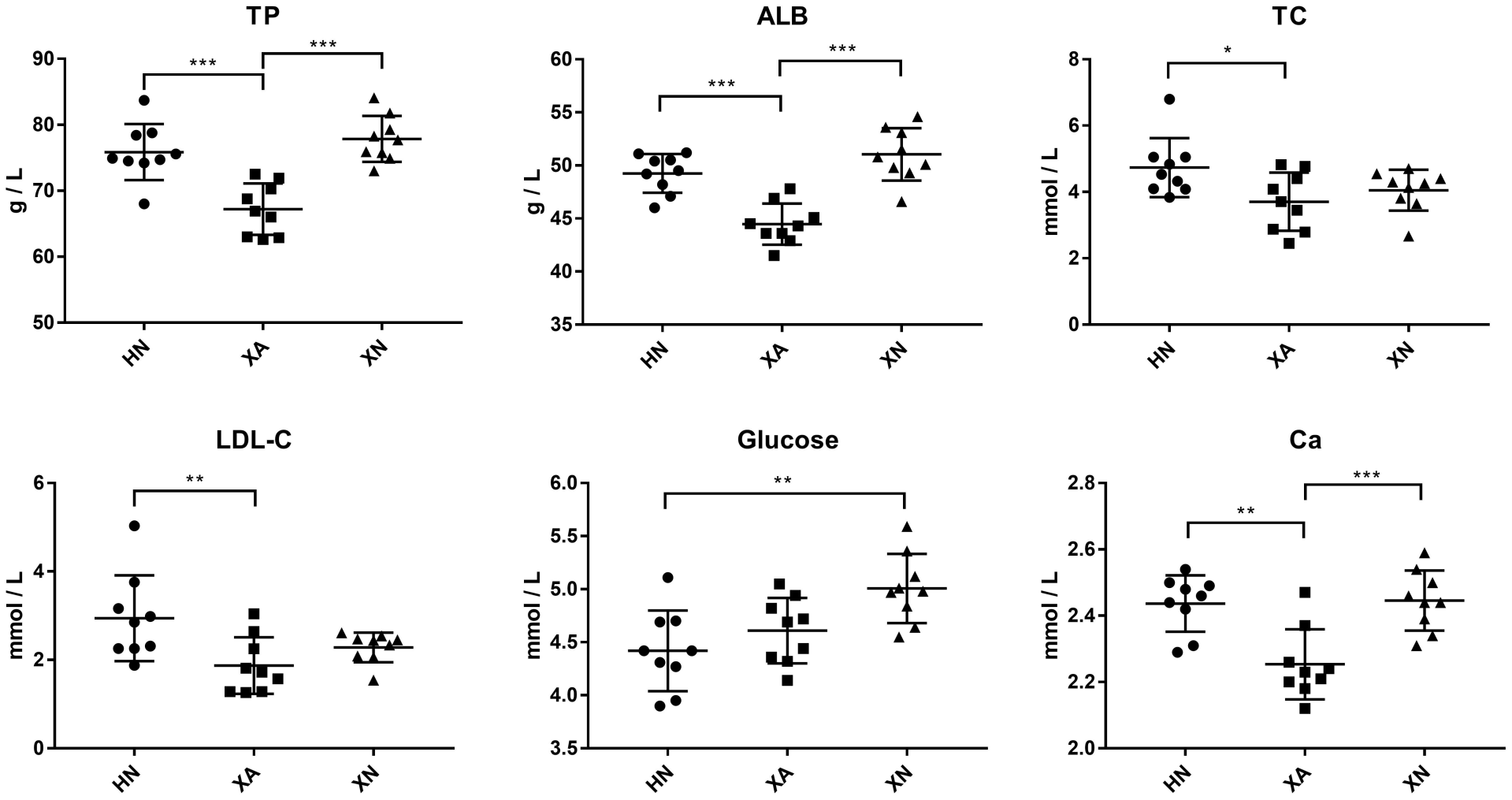
Table 1. The effects of regional differences on disease biomarkers

ID	Protein name	MW	Ratio	P-value	Regional difference	Biomarker use
P30048	Peroxiredoxin-3	27 kDa	7.70	0.038	Haikou / Xining	Liver cancer
P04040	Catalase	60 kDa	4.14	0.015	Haikou / Xining	Colorectal cancer
P02675	Fibrinogen beta chain	56 kDa	3.69	< 0.001	Xi'an / Xining	Bladder cancer and deep vein thrombosis (urine)
P19652	Alpha-1-acid glycoprotein 2	24 kDa	0.37	0.019	Xi'an / Haikou	Cholangiocarcinoma;
			3.02	0.009	Haikou / Xining	pre-eclampsia and diabetes (urine)
Q14117	Dihydropyrimidinase	57 kDa	2.76	0.009	Xi'an / Haikou	Prostate cancer and colon cancer
			2.07	0.033	Xi'an / Xining	
O14786	Neuropilin-1	103kDa	2.09	0.040	Xi'an / Haikou	Colorectal Cancer
Q14108	Lysosome membrane protein 2	54 kDa	2.08	0.009	Xi'an / Haikou	Oral Cancer
Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2	81 kDa	2.12	0.021	Xi'an / Haikou	Bladder cancer (urine)
O00560	Syntenin-1	32 kDa	0.47	0.002	Xi'an / Xining	Renal cell carcinoma (urine)
P20933	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	37 kDa	2.11	0.017	Xi'an / Haikou	Prostatitis (urine)

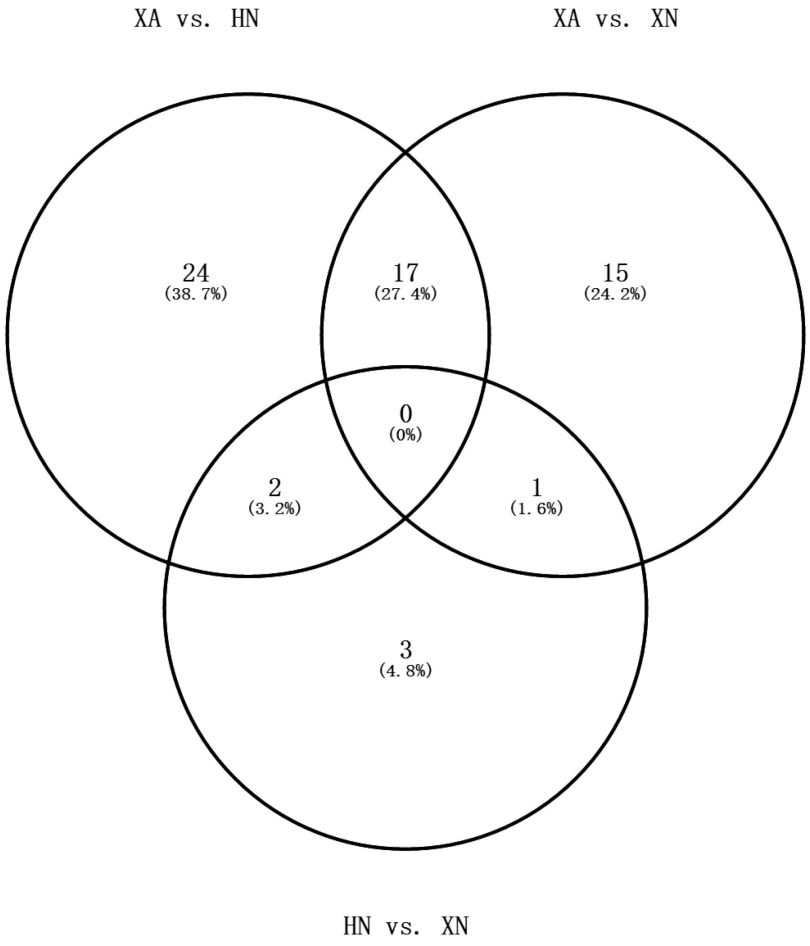
Footnote: MW-molecular weight

bioRxiv preprint doi: https://doi.org/10.1101/190710; this version posted October 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

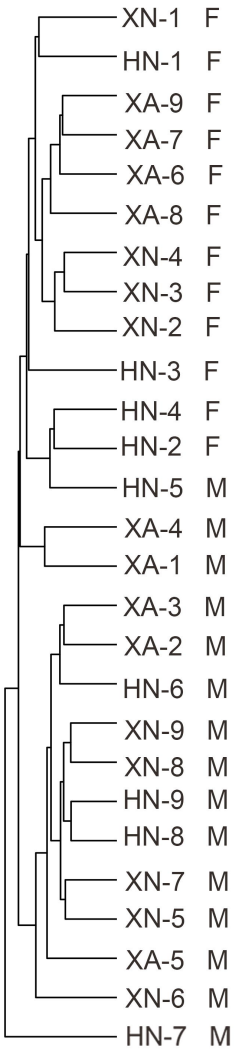
Xining
Xi'an
Haikou


TP
g / L
90
80
70
60
50
***
***
HN
XA
XN
ALB
g / L
60
55
50
45
40
35
***
***
HN
XA
XN
TC
mmol / L
8
6
4
2
0
*
HN
XA
XN
LDL-C
mmol / L
6
4
2
0
**
HN
XA
XN
Glucose
mmol / L
6.0
5.5
5.0
4.5
4.0
3.5
**
HN
XA
XN
Ca
mmol / L
2.8
2.6
2.4
2.2
2.0
**
***
HN
XA
XN

A


XA vs. HN
XA vs. XN
HN vs. XN
24
(38.7%)
17
(27.4%)
15
(24.2%)
0
(0%)
2
(3.2%)
1
(1.6%)
3
(4.8%)

B


XN-1 F
HN-1 F
XA-9 F
XA-7 F
XA-6 F
XA-8 F
XN-4 F
XN-3 F
XN-2 F
HN-3 F
HN-4 F
HN-2 F
HN-5 M
XA-4 M
XA-1 M
XA-3 M
XA-2 M
HN-6 M
XN-9 M
XN-8 M
HN-9 M
HN-8 M
XN-7 M
XN-5 M
XA-5 M
XN-6 M
HN-7 M