Exploring the Effects of Metformin on the Body Through Urinary Proteomics

Authors: Yuzhen Chen, Wang Haitong, Yang Minhui, Shen Ziyun, Gao Youhe, Gao Youhe

Date: 2024-12-27T00:00:00+00:00

Abstract

Metformin is currently the first-line medication for type 2 diabetes mellitus, with over 200 million patients taking it daily, and it exhibits extensive effects. Can we explore its actions and potential mechanisms through urine proteomics? The results of this study show that after 5 consecutive days of metformin administration at 150 mg/kg to rats, a total of 166 differential proteins were identified, including complement component C6, pyruvate kinase, coagulation factor X, growth differentiation factor 15, carboxypeptidase A4, chymotrypsin-like elastase family member 1, L-lactate dehydrogenase C chain, among others, several of which have been reported to be directly affected by metformin or associated with its therapeutic efficacy. Proteins harboring differential proteins or differentially modified peptides could be enriched in multiple biological pathways reported to be related to metformin, such as glutathione metabolism, negative regulation of gluconeogenesis, and the renin-angiotensin system, etc. Meanwhile, some significantly enriched biological pathways have not been previously reported in association with metformin efficacy, potentially providing clues for investigating metformin's underlying mechanisms. In summary, urine proteomics enables comprehensive and systematic exploration of both known and unknown drug effects, opening a new window for research into metformin's mechanisms of action.

Full Text

Preamble

Exploring the Effects of Metformin on the Body Through Urine Proteomics
Yuzhen Chen, Haitong Wang, Minhui Yang, Ziyun Shen, Youhe Gao*
Gene Engineering Drug and Biotechnology Beijing Key Laboratory, College of Life Sciences, Beijing Normal University, Beijing 100875, China

Abstract

Metformin is currently the first-line medication for the treatment of type 2 diabetes mellitus (T2DM), with over 200 million patients taking it daily. Its effects are extensive and play a positive role in multiple areas. Can we explore its effects and potential mechanisms through urine proteome? In this study, a total of 166 differential proteins were identified after rats were given a dose of 150 mg/(kg · d) of metformin for 5 consecutive days, including complement component C6, pyruvate kinase, coagulation factor X, growth differentiation factor 15 (GDF15), carboxypeptidase A4, chymotrypsin-like elastase family member 1, and L-lactate dehydrogenase C chain (LDH-C).

Several of these proteins have been reported to be directly affected by metformin or associated with the effects of metformin. Several biological pathways enriched by the differential proteins or proteins where the differentially modified peptides are located have been reported to be associated with metformin, including glutathione metabolic process, negative regulation of gluconeogenesis, and renin-angiotensin system. Additionally, some significantly enriched biological pathways that have not been reported to be related to the effects of metformin may provide clues for the study of metformin's potential mechanisms. In conclusion, the application of urine proteome offers a comprehensive and systematic approach to explore both the known and unknown effects of drugs, thus opening a new window to study the mechanisms of metformin.

Keywords: Urine; Proteomics; Post-translational modifications; Metformin

Funding: Beijing Normal University (11100704)
Corresponding Author: Youhe Gao (born June 1964), male, professor, doctoral supervisor. Main research interests: urine proteomics and urine biomarkers. E-mail: gaoyouhe@bnu.edu.cn.

1 Introduction

Metformin is currently the first-line medication for the treatment of type 2 diabetes mellitus (T2DM), with over 60 years of clinical application and more than 200 million T2DM patients taking it daily worldwide [**?**]. Beyond its glucose-lowering effects, metformin also plays positive roles in improving cognitive function [**?**], anti-tumor activity [**?**], cardiovascular protection [**?**], anti-aging [**?**], and weight reduction [**?**]. In the treatment of type 2 diabetes, metformin demonstrates excellent safety and tolerability [**?**], does not increase the risk of hypoglycemia when used alone [**?**], and offers favorable cost-effectiveness compared to other hypoglycemic agents. Consequently, it is recommended by multiple guidelines as the foundational therapy for glycemic control in T2DM patients [**?**, **?**].

In recent years, researchers have continuously explored metformin's effects and underlying mechanisms. Studies have revealed that metformin promotes the biosynthesis of N-lactoyl-phenylalanine (Lac-Phe), thereby suppressing appetite

and inducing weight loss [**?**, **?**]. It also increases the abundance of beneficial gut bacteria such as *Akkermansia muciniphila*, modulates host inflammation-related pathways, reduces plasma pro-inflammatory cytokine levels, and improves cognitive function [**?**]. Additionally, metformin reprograms tryptophan metabolism to drive immune-mediated anti-tumor effects [**?**] and activates the transcription factor Nrf2, which possesses antioxidant capacity, thereby delaying neuronal and brain aging [**?**]. Despite its widespread use, metformin's mechanisms have not been fully and clearly elucidated.

With the development of high-throughput sequencing technologies, proteomics research has advanced considerably. By analyzing protein structure, expression, post-translational modifications, and protein-protein interactions, proteomics can reveal the composition and dynamic changes of proteins in cells or organisms [**?**]. Urine is not strictly regulated by homeostatic mechanisms and can accommodate and accumulate more extensive changes, thereby reflecting alterations in all organs and systems earlier and more sensitively [**?**]. Moreover, urinary proteins rarely originate directly from drugs, so changes in the urine proteome directly reflect the body's response to drug treatment. Therefore, utilizing urine proteomics enables comprehensive and systematic reflection of a drug's overall impact on the organism.

The urine proteome is inevitably influenced by various factors including age [**?**], genetics [**?**], sex [**?**], diet [**?**], and exercise [**?**]. Minimizing interference from irrelevant factors is crucial in experiments, and animal models with controllable genetic and environmental conditions represent an excellent choice [**?**].

Multiple studies have demonstrated that drug effects can be reflected in the urine proteome. For instance, research exploring the effects of the $ $1-receptor blocker prazosin on the rat urine proteome identified 775 differential proteins, approximately half of which were related to prazosin treatment [**?**]. Similarly, after altering rat coagulation status with anticoagulants, enrichment analysis revealed activation of the coagulation system, intrinsic and extrinsic coagulation pathways, and even enabled differentiation of coagulation changes induced by drugs with different mechanisms through urine proteomics [**?**].

Furthermore, protein post-translational modifications (PTMs) play crucial roles in regulating protein function by altering chemical properties, structure, or function, thereby affecting activity, localization, folding, and interactions with other proteins. PTMs regulate numerous physiological and pathological processes and are essential for increasing proteomic diversity and maintaining cellular homeostasis [**?**]. Drug-regulated protein PTMs can serve as molecular markers for target engagement, helping to identify drug-modulated pathways and revealing drug targets and mechanisms of action [**?**]. However, the impact of drugs on protein PTMs has received limited attention.

Given metformin's extensive effects and large user population, can we leverage the advantages of urine to comprehensively, systematically, and sensitively reflect the body's status to explore metformin's effects and potential mechanisms

in detail? This study established a metformin-administered rat model to explore metformin's impact on the body through urine proteomics, investigating its known and unknown mechanisms of action and opening a new window for metformin mechanism research (Figure 1 [Figure 1: see original paper]).

2 Materials and Methods

2.1 Urine Sample Collection

Fourteen healthy male Sprague Dawley (SD) rats (200$\pm$20g), $aged$6 − 7$weeks$, $werepurchased from BeijingVitalRiverLaboratoryAnimalTechnologyCo., Ltd.Theratswerehousedir$ g. Animal experiments were reviewed and approved by the Ethics Committee of the College of Life Sciences, Beijing Normal University (Approval No.: CLS-EAW-2020-034).

Metformin was dissolved in Wahaha purified water and administered via gavage at the same time each day for 5 consecutive days. The experimental group (n=9) received a dose of 150 mg/kg, which has been reported to achieve plasma concentrations in rats similar to those in humans [**?**]. The control group (n=5) received Wahaha water via gavage. After 5 days of gavage, all rats were placed in metabolic cages with fasting and water restriction, and 12-hour urine samples were collected and stored at -80°C.

2.2 Urine Sample Processing

(1) Urinary Protein Extraction and Quantification

Collected rat urine was centrifuged at 12,000 × g for 40 min at 4°C. The supernatant was transferred to new centrifuge tubes, and three volumes of ice-cold absolute ethanol were added. After thorough mixing, proteins were precipitated overnight at -20°C. Following centrifugation at 12,000 × g for 30 min at 4°C, the supernatant was discarded, and the protein pellet was resuspended in appropriate lysis buffer (8 mol/L urea, 2 mol/L thiourea, 25 mmol/L dithiothreitol, 50 mmol/L Tris). After complete dissolution, samples were centrifuged at 12,000 × g for 30 min at 4°C, and the supernatant was transferred to new EP tubes to obtain urinary protein extracts. Protein concentration was measured using the Bradford method.

(2) Urinary Protein Digestion

One hundred micrograms of urinary protein sample were transferred to 1.5 mL centrifuge tubes, and 25 mmol/L NH_4HCO_3 solution was added to a final volume of 200 L. Twenty mmol/L dithiothreitol (DTT, Sigma) was added, vortexed, and heated in a metal bath at 97°C for 10 min, then cooled to room temperature. Fifty mmol/L iodoacetamide (IAA, Sigma) was added, vortexed, briefly centrifuged, and incubated at room temperature in the dark for 40 min. Two hundred microliters of UA solution (8 mol/L urea, 0.1 mol/L Tris-HCl, pH 8.5) were added to a 10 kD ultrafiltration tube (Pall, Port Washington, NY,

USA) and centrifuged at 14,000 × g for 5 min at 18°C for two washes. The treated protein sample was added and centrifuged at 14,000 × g for 40 min at 18°C. Two hundred microliters of UA solution were added, vortexed, and centrifuged at 14,000 × g for 40 min at 18°C; the filtrate was discarded, and this step was repeated once. Two hundred microliters of 25 mmol/L NH_4HCO_3 solution were added, vortexed, and centrifuged at 14,000 × g for 40 min at 18°C; the filtrate was discarded, and this step was repeated once. The collection tube was replaced, 100 L of NH_4HCO_3 solution was added, and trypsin (Trypsin Gold, Promega, USA) was added at a 1:50 enzyme-to-protein mass ratio for overnight digestion at 37°C. The peptide-containing solution was collected and desalted using HLB columns (Waters, Milford, MA), dried in a vacuum concentrator, and stored at -80°C.

2.3 LC-MS/MS Tandem Mass Spectrometry Analysis

Digested samples were resuspended in 0.1% formic acid, and peptide concentration was quantified using a BCA kit and diluted to 0.5 g/ L. Fourteen microliters of each sample were pooled to create a mixed peptide sample, which was fractionated using a high pH reversed-phase peptide separation kit (Thermo Fisher Scientific) into 10 fractions. The fractions were dried in a vacuum concentrator and resuspended in 0.1% formic acid. iRT (Biognosis) was added at a 10:1 sample-to-iRT volume ratio. One microgram of each sample was analyzed using an EASY-nLC1200 chromatography system (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) for proteomic data acquisition.

2.4 Database Search and Data Processing

To generate spectral libraries, the 10 fractions were analyzed in Data Dependent Acquisition (DDA) mode. DDA results were imported into Proteome Discoverer software for database searching, and the search results were used to establish DIA acquisition methods by calculating window width and number based on m/z distribution density. Individual samples were analyzed in Data Independent Acquisition (DIA) mode, with pooled peptides serving as quality control throughout the analysis. DIA data were processed using Spectronaut Pulsar software, with protein quantification based on the summed peak areas of all fragment ions from secondary peptides. Highly confident proteins were defined as those with peptide Q value < 0.01 and containing at least 2 unique peptides per protein.

Protein modification information was obtained using pFind Studio software (version 3.2.0, Institute of Computing Technology, Chinese Academy of Sciences) for label-free quantitative analysis of mass spectrometry data. The target database was the *Rattus norvegicus* database downloaded from UniProt (updated to September 2024). Search parameters included: instrument type HCD-FTMS, trypsin digestion with maximum 2 missed cleavages, and open search. Filtering criteria were: false discovery rate (FDR) $\leq$ 1% at

the peptide level and Q value $\leq 1\%$ at the protein level. A Python script "pFind_{{{protein}}*{{contrast}}}*{script}.py" was used to extract peptide spectrum match information from pFind Studio analysis results [**?**, **?**], and modified peptides with $\geq 50\%$ intra-group reproducibility were selected.

2.5 Data Analysis

Differential proteins and modified peptides between experimental and control groups were identified using the criteria: fold change (FC) ≥ 1.5 or ≤ 0.67, and two-tailed unpaired t-test $P < 0.05$. Unsupervised hierarchical cluster analysis (HCA) and principal component analysis (PCA) were performed using the Bioinformatics platform (http://www.bioinformatics.com.cn/). Biological analysis was conducted using the UniProt website (https://www.uniprot.org/) and DAVID database (https://david.ncifcrf.gov/), with literature searches in PubMed (https://pubmed.ncbi.nlm.nih.gov/) for functional analysis of differential proteins and proteins containing differentially modified peptides.

3 Results

3.1 Urine Proteome Analysis

(1) Urinary Protein Identification

LC-MS/MS analysis was performed on 14 samples from experimental and control groups after gavage. A total of 1,542 proteins were identified using the criteria of ≥ 2 unique peptides per protein and protein-level FDR $< 1\%$. Comparison of post-gavage urine between the two groups identified 166 differential proteins using the criteria of FC ≥ 1.5 or ≤ 0.67 and $P < 0.05$, including 88 downregulated and 78 upregulated proteins. Detailed information for differential proteins is listed in Supplementary Table 1 .

Unsupervised hierarchical cluster analysis was performed on all identified proteins (1,542) and differential proteins (166) (Figure 2 [Figure 2: see original paper]), and principal component analysis was conducted on differential proteins (Figure 3 [Figure 3: see original paper]). Both analyses successfully distinguished experimental and control groups.

(2) Random Grouping Validation

To determine the likelihood that the identified differential proteins arose by chance, random grouping validation was performed on all proteins. The 14 samples from control and experimental groups were randomly shuffled and divided into two new groups, yielding 2,002 possible combinations. Using the same criteria (FC ≥ 1.5 or ≤ 0.67, $P < 0.05$), the average number of differential proteins generated was 38.46, indicating that at least 76.83% of the differential proteins were not randomly produced. This confirms the high reliability of the 166 differential proteins identified through experimental versus control group comparison.

(3) Differential Protein Analysis

Functional analysis of the 166 differential proteins using the PubMed database revealed that 4 proteins were reported to be directly affected by metformin, 2 proteins belonged to families whose other members were directly affected by metformin, and 27 proteins, while not reported to be directly influenced by metformin, had functions related to metformin's therapeutic effects. Detailed information for each protein is as follows:

Proteins Directly Affected by Metformin (Table 1)

Table 1: Differential proteins directly affected by metformin

UniProt ID	Protein name	Trend	P-value	References
A0A0H2UI07	Pyruvate kinase	↓	3.44E-05	[8,30-36]
G3V8K5	Cystathionine gamma-lyase	↑	3.70E-04	[37-42]
A0A0G2K3C5	Growth differentiation factor 15	↑	1.06E-03	[43,44]
F1LMN1	Cytochrome P450	↑	1.52E-02	[45]

1) **Pyruvate kinase**: Metformin enhances pyruvate kinase activity in hepatocytes and inhibits gluconeogenesis [**?**], which represents one of the primary pathways for metformin's glucose-lowering effects [**?**]. Researchers have proposed that metformin potentiates the allosteric activation of pyruvate kinase by fructose-1,6-bisphosphate, suggesting pyruvate kinase as a site of metformin action [**?**]. Additionally, pyruvate kinase M2 (PKM2), a downstream molecule in the PI3K/AKT/mTOR signaling pathway, is overexpressed in nearly all tumor types and plays a key role in the Warburg effect [**?**, **?**]. Metformin reduces PKM2 expression in gastric [**?**], esophageal [**?**], and breast cancer cells [**?**], and modulating PKM2 activity or expression has been reported as a potential strategy to enhance metformin's anti-cancer efficacy.

2) **Growth differentiation factor 15 (GDF15)**: Multiple studies have shown that metformin administration increases GDF15 levels [**?**]. GDF15 is a cytokine with anti-inflammatory properties that increases insulin sensitivity, suppresses appetite, reduces body weight in diabetic and non-diabetic patients, and improves prognosis in diabetic patients. GDF15 levels are also associated with the progression of diabetic complications, including thrombosis, diabetic nephropathy, diabetic neuropathy, and diabetic retinopathy [**?**]. GDF15 has been reported as a novel biomarker for metformin in patients with dysglycemia, with its concentration reflecting metformin dosage [**?**]. Due to its anti-inflammatory and appetite-suppressing effects, GDF15 holds great therapeutic potential for various metabolic diseases including obesity, type 2 diabetes, non-alcoholic fatty liver disease, cardiovascular disease, and cancer cachexia [**?**].

3) **Cystathionine gamma-lyase (CSE)**: Metformin alleviates atherosclerosis by regulating CSE expression and promoting hydrogen sulfide (H_2S)

production [**?**]. In rats exposed to bisphenol A (BPA), metformin upregulates CSE and cystathionine β-synthase (CBS) expression, reduces serum homocysteine levels, and protects against BPA-induced liver injury [**?**].

4) **Cytochrome P450**: In rats, metformin is primarily metabolized by hepatic microsomal cytochrome P450 (CYP) isoenzymes CYP2C11, 2D1, and 3A1/2 [**?**].

Proteins Whose Family Members Are Directly Affected by Metformin (Table 2)

Table 2: Differential proteins whose family members are directly affected by metformin

UniProt ID	Protein name	Trend	P-value	References
A0A0G2JVF2	Solute carrier family 22, member 21	↓	4.66E-02	[46-49]
D3ZHS5	Carboxypeptidase A4	↑	8.39E-03	[50]

1) **Solute carrier family 22 member 21 (SLC22A21)**: The solute carrier family 22 is an organic cation transporter involved in transporting various endogenous and exogenous substances. Metformin's transporters are primarily solute carrier family 22 member 1 (OCT1) and member 4 (OCTN1) [**?**], with genetic polymorphisms in OCT1 affecting metformin pharmacokinetics and gastrointestinal intolerance, thereby influencing individual responses to metformin [**?**].

2) **Carboxypeptidase A4 (CPA4)**: Carboxypeptidase A4 (FC = 49.79, p = 8.39$×10^{-3}$) showed the highest fold change among the 166 differential proteins identified in this study. Genetic variation in carboxypeptidase A6 (CPA6) has been reported to be associated with metformin response in type 2 diabetes patients [**?**].

Proteins Not Directly Reported to Be Affected by Metformin but Functionally Related to Its Therapeutic Effects (Table 3)

Table 3: Differential proteins not directly reported to be affected by metformin but functionally related to its therapeutic effects

UniProt ID	Protein name	Trend	FC	P-value	Effect/Associated with a disease
D3ZUR5	Secreted Ly6/Plaur domain containing 2	↑	1.54	5.67E-04	psoriasis [51-55]

UniProt ID	Protein name	Trend	FC	P-value	Effect/Associated with a disease
A0A0G2JX26	Chymotrypsin-like elastase family member 1	↑	25.11	3.91E-02	emphysema [56,57]
Q63471	BPI fold-containing family A member 2	↑	24.76	4.66E-03	acute kidney injury [58,59]
Q5XIM9	T-complex protein 1 subunit beta (TCP-1-beta)	↑	19.23	3.68E-03	diabetic nephropathy [60-63]
G3V709	Nicotinate phosphoribosyltransferase	↑	19.52	1.61E-02	aging [64]
Q07009	Coagulation factor XII	↑	2.29	2.91E-03	coagulation factor
P16446	Coagulation factor X	↑	2.29	2.08E-04	coagulation factor
Q5BJU0	Calpain-2 catalytic subunit	↓	0.21	9.08E-04	atrial fibrillation [68]
A0A0G2JVZ6	Integrin subunit alpha V	↓	0.27	4.92E-02	high blood glucose/diabetes [69]
Q4KLZ6	5-oxoprolinase	↓	0.61	2.41E-02	high blood glucose/diabetes [70]
A0A0G2K2B6	Riokinase/FMN cyclase	↓	0.61	2.45E-02	high blood glucose/diabetes [71]
D4A0W2	Regenerating family member 3 beta	↑	1.84	3.68E-03	high blood glucose/diabetes [72]
P07824	Lysozyme f1	↑	1.90	1.76E-02	high blood glucose/diabetes [73]
P97608	Arginase-1	↑	2.04	2.98E-02	high blood glucose/diabetes [74]

UniProt ID	Protein name	Trend	FC	P-value	Effect/Associated with a disease
P19629	L-lactate dehydrogenase C chain (LDH-C)	↑	2.18	4.38E-02	cancer [81,82]
Q6TA48	Mucosal pentraxin	↑	2.18	4.09E-02	cancer [83-87]
Q07009	Dynein cytoplasmic 1 heavy chain 1	↑	2.19	2.34E-02	cancer [88-90]
P16446	Phosphatidylinositol transfer protein alpha isoform	↑	2.29	1.59E-03	cancer [91-93]
D3Z9E5	Sodium-coupled monocarboxylate transporter 1	↑	2.31	6.20E-03	cancer [94-97]
F1M3L7	Epidermal growth factor receptor kinase substrate 8	↑	2.41	7.70E-03	cancer [98-103]
Q66H12	Alpha-N-acetylgalactosaminidase	↑	2.57	2.21E-03	cancer [104]
Q4FZU6	Annexin A8	↑	2.78	1.60E-03	cancer [105,106]
P02793	Ferritin light chain 1	↑	3.50	2.44E-02	cancer [107]
F1LQQ8	Beta-glucuronidase	↑	3.76	3.37E-03	cancer [108]
Q3KR94	Vitronectin	↑	4.07	2.72E-02	cognitive dysfunction [116,117]
F1M7F7	Complement component C6	↑	4.26	1.97E-02	cognitive dysfunction [118,119]

1) **Secreted Ly6/Plaur domain containing 2 (SLURP-2)**: SLURP-2 (FC = 1.54, p = 5.67$×10^{-4}$) ranked 5th smallest in p-value among all differential proteins identified. SLURP-2 is a novel member of the Ly-6 superfamily that is upregulated in psoriasis patients and may participate in psoriasis pathophysiology through keratinocyte proliferation and T cell differentiation/activation [**?**]. Multiple studies have shown that metformin effectively improves treatment outcomes and metabolic syndrome in psoriasis patients [**?**], and long-term metformin use is associated with reduced psoriasis risk [**?**].

2) **Chymotrypsin-like elastase family member 1**: This protein (FC = 25.11, p = 3.91$×10^{-2}$) showed the second-highest fold change after carboxypeptidase A4. The gene encoding this protein, *Cela1*, is expressed during lung development and is closely related to stretch-dependent remodeling processes in lung physiology and pathology, with increased expression in both mouse and human alpha-1 antitrypsin deficiency emphysema, making it a potential specific therapeutic target [**?**]. Metformin also plays a potentially important role in emphysema treatment in mice and humans, particularly in slowing emphysema progression [**?**].

3) **BPI fold-containing family A member 2 (BPIFA2)**: BPIFA2 levels are higher in blood and urine of acute kidney injury patients compared to healthy individuals, serving as an early biomarker for acute kidney injury [**?**]. Even low-dose metformin can exacerbate renal ischemia-reperfusion-induced acute kidney injury and increase mortality in mice [**?**].

4) **T-complex protein 1 subunit beta (TCP-1β)**: TCP-1β can serve as a biomarker for the glomerular hyperfiltration phase of diabetic nephropathy in type 2 diabetes [**?**]. Multiple studies have shown that metformin plays an important role in alleviating diabetic nephropathy [**?**].

5) **Nicotinate phosphoribosyltransferase**: This enzyme plays an important role in trigonelline metabolism, which can be used for nicotinamide adenine dinucleotide (NAD^+) synthesis, enhances mitochondrial activity, and holds great potential for improving age-related muscle decline [**?**].

6) **Coagulation factor X and Coagulation factor XII**: Diabetic patients exhibit a hypercoagulable state compared to healthy controls, associated with high levels of coagulation factors (II, V, VII, VIII, and X) and low levels of anticoagulants (protein C) [**?**].

7) **Calpain-2 catalytic subunit**: Multiple studies have shown that metformin plays a positive role in cardiovascular protection, such as reducing cardiovascular mortality, all-cause mortality, and cardiovascular event risk in coronary heart disease patients [**?**], and is associated with reduced cardiovascular disease mortality and morbidity in type 2 diabetes patients [**?**]. Metformin also protects the heart against hypertrophic and apoptotic remodeling after myocardial infarction [**?**]. Calpain 2 is reportedly significantly elevated in atrial samples from atrial fibrillation patients compared

to sinus rhythm patients and may be associated with atrial fibrillation development in valvular heart disease and diabetic patients [**?**].

Proteins related to hyperglycemia or diabetes include:

1) **Integrin subunit alpha V**: Hyperglycemia reduces integrin αv and $ $5 expression in dermal fibroblasts, affecting fibroblast migration and wound healing, which may represent one mechanism for defective wound healing in diabetes [**?**].

2) **5-oxoprolinase (OPLAH)**: Downregulation of OPLAH in human skeletal muscle cells may cause insulin resistance and impaired glucose uptake through oxidative stress, representing a novel therapeutic target for type 2 diabetes [**?**].

3) **Triokinase/FMN cyclase**: The gene encoding this enzyme, *Tkfc*, is a novel NRF2 target gene that is upregulated in non-canonical NRF2 activation and plays an important role in promoting hepatic fructose metabolism and gluconeogenesis, affecting glucose homeostasis [**?**].

4) **Regenerating family member 3 beta (Reg3β)**: Recombinant Reg3β protein plays an important role in preventing streptozotocin-induced diabetes and pancreatic β-cell damage in mice [**?**].

5) **Lysozyme f1**: In type 2 diabetes patients, lysozyme structure and function are altered. L-lysine, as a chemical chaperone, can significantly improve lysozyme structure and function, reverse glycation-induced changes, and increase lysozyme activity, helping prevent diabetic complications [**?**].

6) **Arginase-1 (ARG1)**: Arginase 1 and 2 play important roles in regulating β-cell function, insulin resistance, and vascular complications by modulating L-arginine metabolism, nitric oxide (NO) production, inflammatory responses, and oxidative stress. Abnormal arginase expression and activity are closely related to diabetes development and complications, making arginase a potential therapeutic target [**?**].

Furthermore, cancer has become the leading cause of death in diabetic patients in high-income countries [**?**]. A meta-analysis involving 10,695,875 type 2 diabetes patients showed that metformin use significantly reduced cancer risk compared to other hypoglycemic agents, with subgroup analysis revealing significant risk reduction for bladder, colorectal, gastric, hepatic, lung, pancreatic, and prostate cancers [**?**]. Multiple studies have also demonstrated metformin's therapeutic potential in glioma [**?**], cervical cancer [**?**], acute myeloid leukemia [**?**], breast cancer [**?**], and ovarian cancer [**?**].

Cancer-related differential proteins include:

1) **L-lactate dehydrogenase C chain (LDH-C)**: L-lactate is generated from pyruvate reduction catalyzed by lactate dehydrogenase. Many human tumor tissues have higher lactate dehydrogenase 5 (LDH5) levels than normal tissues, and the *LDHC* gene is also expressed in various

cancers including lung, melanoma, prostate, and breast cancers [**?**, **?**]. Enzymes involved in L-lactate metabolism are closely associated with pathophysiological processes in diabetes, cancer, and other diseases, potentially offering new therapeutic strategies [**?**].

2) **Mucosal pentraxin**: Heme, which is associated with red meat and colorectal cancer risk, can downregulate the rat *Mptx* gene by more than 10-fold [**?**]. Mptx may be involved in processing damaged cells in colonic mucosa, and its expression is related to colonic cell turnover, serving as a marker for diet-induced colonic mucosal stress [**?**, **?**]. Meta-analysis studies have shown that metformin use is associated with reduced colorectal cancer risk [**?**], lower overall colorectal cancer mortality, and better prognosis in colorectal cancer patients [**?**, **?**].

3) **Dynein cytoplasmic 1 heavy chain 1 (DYNC1H1)**: DYNC1H1 encodes the cytoplasmic dynein heavy chain family, which links engulfment to apoptosis and prevents various diseases including cancer, neurodegenerative diseases, and autoimmune disorders [**?**, **?**]. Studies have shown that DYNC1H1 is a novel prognostic biomarker for hepatocellular carcinoma associated with epithelial-mesenchymal transition and immune infiltration, holding great potential for early diagnosis and effective intervention [**?**]. Metformin may affect early progression of NAFLD/NASH-related hepatocellular carcinoma by modulating macrophage polarization and T cell infiltration [**?**].

4) **Phosphatidylinositol transfer protein alpha isoform (PITP-alpha)**: PITP is an abundant and ubiquitous soluble protein, with PITPα/β showing increased expression in gastric cancer tissues and correlating with poor prognosis, representing potential therapeutic targets [**?**]. Additionally, reduced PITPα expression is associated with improved pathology in Duchenne muscular dystrophy, making it a potential therapeutic target [**?**], while metformin can improve muscle function and reduce neuromuscular defects in muscular dystrophy mice, showing potential as a treatment for Duchenne muscular dystrophy patients [**?**].

5) **Sodium-coupled monocarboxylate transporter 1**: The gene encoding this transporter, *SLC5A8*, can function as a tumor suppressor in various cancers. For example, in cervical cancer, *SLC5A8* is silenced through DNA hypermethylation and histone deacetylation, serving as a diagnostic and prognostic biomarker and therapeutic target [**?**]. *SLC5A8* is identified as a tumor suppressor gene silenced by methylation in colon cancer [**?**]. In acute myeloid leukemia with MLL partial tandem duplication, tumor suppressor gene *SLC5A8* silencing promotes leukemogenesis [**?**]. *SLC5A8* is also silenced in mouse breast cancer, and reactivating its expression may represent a novel breast cancer treatment strategy [**?**].

6) **Epidermal growth factor receptor kinase substrate 8 (Eps8)**: Eps8 is highly expressed in various human tumor types including colorectal

[?], pituitary [?], oral squamous cell [?], esophageal [?], pancreatic [?], and cervical cancers [?], participating in many signal transduction pathways related to cancer development, metastasis, and proliferation, and serving as a biomarker for poor cancer prognosis.

7) **Alpha-N-acetylgalactosaminidase**: shRNA-mediated downregulation of α-N-acetylgalactosaminidase inhibits migration and invasion of breast and ovarian cancer cell lines, making it a potential anti-cancer therapeutic target [?].

8) **Annexin A8**: Annexin A8 is overexpressed in pancreatic cancer [?], and its increased expression is associated with poor prognosis in early pancreatic cancer, serving as a prognostic marker and potential therapeutic target [?].

9) **Ferritin light chain 1 (FTL)**: Hypoxia-induced FTL is a regulator of epithelial-mesenchymal transition and serves as a prognostic marker for glioma and a novel biomarker for response to the anti-tumor drug temozolomide [?]. Metformin can inhibit glioma cell stemness and epithelial-mesenchymal transition by regulating Hippo pathway effector YAP activity [?].

10) **Beta-glucuronidase**: High levels of urinary β-glucuronidase are observed in bladder cancer patients [?].

Cognitive dysfunction has been reported as one of many diabetic complications [?], with multiple studies showing that diabetic patients have increased risk of dementia such as Alzheimer's disease [?]. Metformin plays an active role in improving cognitive impairment and alleviating memory loss [?, ?, ?, ?].

Cognitive dysfunction-related differential proteins include:

1) **Secernin-2**: Six genes—*SCRN2*, *LCMT1*, *LRRC46*, *MRPL10*, *SP6*, and *OSBPL7*—are significantly associated with amyloid-β standardized uptake value ratio in the brain. SNPs in these genes are also associated with reduced hippocampal volume and decreased cognitive scores, making these six genes potential novel therapeutic targets for Alzheimer's disease [?].

2) **Vitronectin (VTN)**: VTN is a multifunctional glycoprotein whose receptors are associated with various diseases including tumors, coagulation disorders, inflammatory diseases, and multiple neurodegenerative diseases. VTN plays important roles in neuronal function and neurodegenerative diseases, participating in neuronal differentiation, neurotrophic support, and neurogenesis, regulating axon size, supporting and guiding neurite extension, and protecting the brain by reducing blood-brain barrier permeability through interaction with integrin receptors in vascular endothelial cells [?].

3) **Complement component C6**: Complement component C6 (FC = 1.96,

p = 2.66$×10^{-5}$) showed the smallest p-value among all differential proteins identified. Complement is an important factor in neurodegenerative disease progression, including Alzheimer's disease, amyotrophic lateral sclerosis, and schizophrenia [**?**]. Activation of innate immune responses, particularly the terminal pathway of the complement system, can lead to membrane attack complex (MAC) formation and delay peripheral nervous system regeneration, representing a key cause of neuronal injury. Complement component C6 plays an important role in complement system activation and MAC formation, and C6 deficiency is beneficial for neuronal recovery after trauma [**?**].

(3) Biological Pathway Analysis

Biological process and molecular function enrichment analysis of differential proteins was performed using the DAVID database (Figure 4 [Figure 4: see original paper]). These differential proteins were mainly involved in cellular response to cytochalasin B, apical protein localization, regulation of norepinephrine uptake, regulation of transmembrane transporter activity, proteolysis, glycoside catabolism, regulation of cyclin-dependent protein serine/threonine kinase activity, coagulation, glutathione metabolism, establishment of endothelial barrier, and negative regulation of gluconeogenesis. Among these, coagulation, glutathione metabolism, establishment of endothelial barrier, and negative regulation of gluconeogenesis have been reported to be associated with metformin's therapeutic effects. Diabetic patients exhibit a hypercoagulable state compared to healthy controls [**?**]. Metformin can regulate glutathione metabolism and affect thyroid cancer progression [**?**]. The vascular endothelium is closely related to cardiovascular function regulation, and multiple studies have shown that metformin plays an important role in improving endothelial function [**?**]. Inhibition of gluconeogenesis is one of the primary pathways for metformin's glucose-lowering effects [**?**].

In terms of molecular function, these differential proteins mostly exhibited protein binding, hydrolase activity, structural constituent of postsynaptic actin cytoskeleton, serine-type endopeptidase activity, Tat protein binding, integrin binding, receptor binding, small molecule binding, nitric-oxide synthase binding, and lysozyme activity. Nitric-oxide synthase plays an important role in oxidative stress and vascular disease [**?**], while endothelial nitric-oxide synthase (eNOS) uncoupling can partially explain the pathogenesis of diabetic vascular disease through reduced endothelial progenitor cell levels and impaired function [**?**]. Metformin can promote angiogenesis and neurological function recovery after spinal cord injury in aged mice by activating the AMPK/eNOS signaling pathway [**?**]. In type 2 diabetes patients, lysozyme structure and function are altered, and reversing glycation-induced changes to increase lysozyme activity helps prevent diabetic complications [**?**].

KEGG pathway enrichment analysis (Figure 5 [Figure 5: see original paper]) revealed significant enrichment in complement and coagulation cascades, proteoglycans in cancer, focal adhesion, regulation of actin cytoskeleton,

dilated cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy, ECM-receptor interaction, hypertrophic cardiomyopathy, glycerolipid metabolism, platelet activation, and PI3K-Akt signaling pathway. Among these, glycerolipid metabolism, platelet activation, and PI3K-Akt signaling pathway have been reported to be associated with metformin's therapeutic effects. Type 2 diabetes can induce inactivation of the glycerolipid metabolism pathway, which is significantly reversed by vanillin with anti-diabetic activity [**?**]. Compared to healthy individuals, platelet surface receptors and platelet activation markers show significant differences in type 2 diabetes patients [**?**], and platelet activation is also associated with the pathogenesis of chronic diseases such as atherosclerosis, coronary artery disease, and cerebrovascular disease [**?**]. The PI3K-Akt signaling pathway is the most frequently activated pathway in cancer. Under physiological conditions, this pathway is activated in response to insulin, growth factors, and cytokines, regulating key metabolic processes such as glucose metabolism and macromolecule biosynthesis to maintain systemic metabolic homeostasis [**?**].

3.2 Urine Proteome Modification Analysis

(1) Identification of Differentially Modified Peptides

Using label-free quantitative proteomics and LC-MS/MS analysis of 14 samples, peptide spectrum match information was obtained through open-pFind searching, including protein localization and modification types. Modified peptides with $\geq$ 50% intra-group reproducibility were selected from experimental and control groups, yielding a total of 3,206 modified peptides. Using the criteria of FC $\geq$ 1.5 or $\leq$ 0.67 and P $<$ 0.05, 285 differentially modified peptides were identified. Detailed information is provided in Supplementary File 1, including peptide sequences, modification types, and host proteins.

Unsupervised hierarchical cluster analysis was performed on all modified peptides (3,206) and differentially modified peptides (285) (Figure 6 [Figure 6: see original paper]), and principal component analysis was conducted on differentially modified peptides (Figure 7 [Figure 7: see original paper]). Both analyses successfully distinguished experimental and control samples. Principal component analysis results for both differential proteins and differentially modified peptides (Figures 3 and 7) showed that experimental group samples were more dispersed than control group samples, indicating individual variation. Genetic polymorphisms in genes encoding metformin transporters such as OCT1, OCT2, MATE1, and MATE2 are significantly associated with metformin efficacy and toxicity [**?**]. Additionally, individual responses to metformin are influenced by DNA methylation levels [**?**], individual health status [**?**, **?**], sex [**?**], and other factors. This suggests that individual differences should be considered in drug research.

(2) Random Grouping Validation

To determine the likelihood that the identified differentially modified peptides

arose by chance, random grouping validation was performed on all modified peptides. The 14 samples from control and experimental groups were randomly shuffled and divided into two new groups, yielding 2,002 possible combinations. Using the same criteria (FC $\geq$ 1.5 or $\leq$ 0.67, P $<$ 0.05), the average number of differentially modified peptides generated was 132.8, indicating that at least 53.4% of the differentially modified peptides were not randomly produced.

(3) Biological Pathway Analysis of Proteins with Differentially Modified Peptides

Proteins containing differentially modified peptides were identified, yielding 127 proteins. Detailed information is provided in Supplementary File 2. Biological process and molecular function enrichment analysis of these proteins was performed using the DAVID database (Figure 8 [Figure 8: see original paper]). Proteins with differentially modified peptides were mainly involved in acute-phase response, proteolysis, carbohydrate metabolism, zymogen activation, positive regulation of endothelial cell proliferation, response to bacteria, response to nutrients, positive regulation of receptor-mediated endocytosis, regulation of systemic arterial blood pressure, cellular oxidant detoxification, positive regulation of oligodendrocyte progenitor cell proliferation, coagulation, intracellular phosphate ion homeostasis, negative regulation of amyloid fibril formation, and negative regulation of blood coagulation.

In terms of molecular function, these proteins mostly exhibited pheromone binding, small molecule binding, endopeptidase inhibitor activity, protease binding, serine-type endopeptidase inhibitor activity, serine-type endopeptidase activity, antioxidant activity, protein folding chaperone binding, and hemoglobin binding.

KEGG enrichment analysis (Figure 9 [Figure 9: see original paper]) revealed significant enrichment in complement and coagulation cascades, lysosome, renin-angiotensin system, *Staphylococcus aureus* infection, protein digestion and absorption, ECM-receptor interaction, endocrine and other factor-regulated calcium reabsorption, glycosaminoglycan degradation, galactose metabolism, starch and sucrose metabolism, focal adhesion, cholesterol metabolism, sphingolipid metabolism, and sphingolipid signaling pathway. Angiotensin II, a key component of the renin-angiotensin system (RAS), is an important target for effectively lowering blood pressure and preventing cardiovascular disease progression and renal injury in diabetic patients [**?**].

(4) Differentially Modified Peptides with Appearance/Disappearance

Differentially modified peptides with significant changes were selected based on the following criteria: identified in more than half of control group samples but absent in all experimental group samples, or identified in more than half of experimental group samples but absent in all control group samples. A total of 100 such peptides were identified, including 9 present in control but absent in experimental groups, and 91 present in experimental but absent in control groups. Detailed information is provided in Supplementary File 3, including

peptide sequences, modification types, host proteins, and spectrum counts for each sample.

Proteins containing these appearance/disappearance peptides were identified, yielding 57 proteins. Detailed information is provided in Supplementary File 4. Biological process and molecular function enrichment analysis of these proteins was performed using the DAVID database (Figure 10 [Figure 10: see original paper]). These proteins were mainly involved in proteolysis, zymogen activation, response to lipopolysaccharide, cellular oxidant detoxification, systemic arterial blood pressure regulation, acute-phase response, response to nutrients, zinc ion response, negative regulation of membrane attack complex activation, positive regulation of lipoprotein transport, acylglycerol homeostasis, very-low-density lipoprotein particle remodeling, response to triglyceride, lipoprotein catabolism, positive regulation of CoA-transferase activity, reverse cholesterol transport, peripheral nervous system axon regeneration, negative regulation of lipid biosynthesis, and high-density lipoprotein particle remodeling. Long-term metformin use can reduce cholesterol and low-density lipoprotein levels in mice [**?**]. In terms of molecular function, these proteins mostly exhibited pheromone binding, small molecule binding, endopeptidase inhibitor activity, antioxidant activity, serine-type endopeptidase activity, serine-type endopeptidase inhibitor activity, phosphatidylcholine-sterol O-acyltransferase activator activity, hemoglobin binding, and carbohydrate binding.

KEGG enrichment analysis (Figure 11 [Figure 11: see original paper]) revealed significant enrichment in renin-angiotensin system, complement and coagulation cascades, lysosome, and cholesterol metabolism.

4 Conclusion

Utilizing urine proteomics facilitates comprehensive and systematic exploration of both known and unknown drug effects, opening a new window for metformin mechanism research.

Note: Due to automatic loss of some spectrum count information during modified peptide screening with Excel in the first version, the proteome modification analysis section contained errors. This version has corrected this error.

Note: Figure translations are in progress. See original paper for figures.