Multiplex LC-MS/MS testing for early detection of kidney injury: a next-generation alternative to conventional immunoassays?

Duijl, T.T. van; Ruhaak, L.R.; Kooten, C. van; Fijter, J.W. de; Cobbaert, C.M.

Multiplex LC-MS/MS Testing for Early Detection of Kidney Injury: A Next-Generation Alternative to Conventional Immunoassays?

Tirsa T. van Duijl [a,*] L. Renee Ruhaak [a] Cees van Kooten [b] Johan W. de Fijter [b] and Christa M. Cobbaert [a]

Background: LC-MS/MS has enabled the translation of many novel biomarkers to the clinical laboratory, but its potential for measurement of urinary proteins is still unexplored. In this study we examined the correlation and agreement between immunoassay and LC-MS/MS in the quantitation of kidney injury biomarkers and evaluated the application of technical LC-MS/MS meta-data assessment to ensure test result validity.

Methods: NGAL, IGFBP7, TIMP2, and KIM-1 were quantified in 345 urine samples with one multiplex lab-developed test that combines immunocapture with mass spectrometry read-out and 4 singleplex sandwich-type immunoassays. Assay performance and imprecision were monitored by 2 urine-based quality controls. Ion ratios, signal intensity, and retention time were monitored over all study samples.

Results: The LC-MS/MS retention time drift was $\leq$1.2%, ion ratios were within 20% of the target values at concentrations of $>$100 pmol/L, and peptides originating from the same protein were in agreement (slopes between 1.03 and 1.41). The interassay CV was between 9.3% and 19.1% for LC-MS/MS analysis and between 4.2% and 10.9% for immunoassay. Direct LC-MS/MS analysis was correlated with immunoassay in the quantitation of NGAL ($r = 0.93$; range: 0.01–37 nmol/L), IGFBP7 ($r = 0.80$; range: 0.01–2.6 nmol/L), TIMP2 ($r = 0.85$; range: 0.01–6.3 nmol/L), and KIM-1 ($r = 0.70$; range 0.01–0.4 nmol/L), but the analytical methodologies differed in measurands and calibration strategies.

Conclusions: LC-MS/MS is explored as a next-generation technology for multiplex urinary protein measurement. It has great potential to overcome nonselectivity and lack of standardization because of its capability of directly measuring well-defined molecular proteins.

[a]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, The Netherlands; [b]Department of Nephrology, Leiden University Medical Center, Leiden, The Netherlands.
*Address correspondence to this author at: Albinusdreef 2, 2333 ZA Leiden, The Netherlands. E-mail t.t.van_duijl@lumc.nl.
Received December 29, 2021; accepted March 8, 2022.
https://doi.org/10.1093/jalm/jfac024

IMPACT STATEMENT

Mass spectrometry technology has potential for direct and multiplex quantitation of urinary proteins. The technology quality assessment of test results using technical meta-data (e.g., ion ratio, retention time, internal standard signal intensity monitoring) and allows unequivocal molecular detection of the measurands of interest and dissolution of calibration bias; that is, the effects of calibration bias and sample specific bias caused by ill-defined measurands. Future standardization of mass spectrometry–based tests will allow metrological traceability of urinary marker test results to standards of higher order.

INTRODUCTION

LC-MS/MS has been proposed as a next generation analytical strategy to substitute immunoassay (IA)-based technologies for protein biomarker quantitation in complex biological matrices (1). The main benefits of LC-MS/MS are analytical selectivity, unequivocal characterization and detection of the measurand, multiplexing capability, and the use of embedded technical meta-data (e.g., ion ratio, internal standard monitoring) and interpeptide agreement evaluation in the postanalytical phase to validate test results (2, 3). Previously, we used LC-MS/MS to quantify multiple proteins in complex biological matrices to target unmet clinical needs in kidney injury diagnosis by analyzing urinary NGAL, IGFBP7, TIMP2, and KIM-1 (4). Understanding the degree of correlation and discordances of direct and indirect measurements of these urinary proteins made by both LC-MS/MS and IA enables comparison between studies and laboratories using these methods.

The application of IA or LC-MS/MS for protein quantitation each have particular benefits and drawbacks. IA is known for its high precision, robustness, and suitability for automation for time-efficient analysis but is more prone to specificity errors, especially when multiplexing low-abundance proteins (5, 6). Moreover, the selectivity of the measurand in IA principally relies on the antibodies used, since optical or chemiluminescence read-out systems indirectly detect the protein measurand. In contrast, the analytical selectivity of LC-MS/MS is achieved by direct detection of the mass:charge ratio (m/z) of the proteotypic peptide measurand(s). Moreover, an additional level of selectivity in quantitative protein mass spectometry (MS) may be achieved through evaluation of interpeptide agreement during the postanalytical assessment. Therefore, LC-MS/MS–based protein analysis strategies may reveal IA selectivity flaws that would otherwise remain unnoticed (6). However, protein quantitation by LC-MS/MS is technically complex and has multiple error-sensitive sample preparation steps and a time-consuming preanalytical phase.

TIMP2, KIM-1, NGAL, and IGFBP7 are low-abundance proteins, present in picomolar to nanomolar concentrations, and demand a sensitive and selective analytical platform for quantitation. To date, concentrations of these biomarkers can be determined by either lab-developed tests (LDTs), research use–only sandwich ELISAs, or CE-marked IA-based tests (7, 8). We developed a LDT for multiplex NGAL, IGFBP7, TIMP2, and KIM-1 quantitation to facilitate parallel biomarker evaluation and translation towards the clinical laboratory (9). This LDT combines immunocapture with an LC-MS/MS read-out to achieve the desired level of analytical sensitivity and selectivity. In this study, we aim to determine the correlation and agreement between

the multiplex LDT and 4 singleplex ELISAs in the quantitation of NGAL, IGFBP7, TIMP2, and KIM-1 from urine samples of renal allograft donors and recipients. By exploring method transferability and LC-MS/MS data validity, we aim to unveil the opportunities and limitations of multiplex kidney injury biomarker analysis by LC-MS/MS.

MATERIALS AND METHODS

To cover the expected wide urinary concentration range of TIMP2, KIM-1, NGAL, and IGFBP7, 343 deidentified urine samples were selected from the REnal Protection Against Ischemia-Reperfusion in transplantation (REPAIR) clinical trial (10). Ethical approval for the study in the United Kingdom was given by the Joint University College London/University College London Hospital Committees on the Ethics of Human Research in June 2009 (reference number 09/H0715/48). In the Netherlands, the trial was approved by the central Medical Ethical Committee. Between January 2010 and April 2013, spot urine samples were collected and centrifuged at 400 g for 10 min and the supernatant was stored at −80°C until analysis. Samples underwent 2 and 3 freeze–thaw cycles before ELISA and LC-MS/MS analysis, respectively.

Urinary proteins were quantified in singleplex by 4 research-use only ELISA kits: NGAL (R&D Systems), IGFBP7 (BOSTER Biological Technology), TIMP2 (Quantikine, BioTechne), and KIM-1 (R&D Systems) according to the manufacturer's instructions. The kits included calibrators based on recombinant protein. Detailed information about the reagents and analytical procedures can be found in the online supplementary information (Supplemental methods).

Proteins were quantified in multiplex in a total of five LC-MS/MS analysis batches by the LDT (9). In each batch, 2 urine-based internal quality controls and 5 urine-based external calibrators were prepared and analyzed together with the samples. For NGAL, TIMP-2, and KIM-1, calibrators were value-assigned in pmol/L by recombinant protein spiking, whereas IGFBP7 was value-assigned by ELISA.

ELISA results were converted to molar units using the molecular weight based on the canonical amino acid sequence in the UniProt database (11). Mass Hunter Workstation software, version 10.0 (Agilent Technologies), was used for LC-MS/MS peak integration. Statistical analysis and graphics were generated in R (version 4.0.2). Passing-Bablok regression with Spearman rank correlation coefficient are provided.

RESULTS

NGAL, IGFBP7, TIMP2, and KIM-1 were quantified by IA and LC-MS/MS, both using antibody-based enrichment but different read-out technology, in 343 urine samples from kidney transplant donors and recipients and were found to be present in the picomolar to nanomolar range (Table 1). IGFBP7, TIMP2, and NGAL were detected by LC-MS/MS in >93% of urine specimens, whereas KIM-1 could be detected in 61% of the samples (limit of detection ± 1 pmol/L). IA analysis for NGAL, IGFBP7, TIMP2, and KIM-1 provided results for all urine samples (all >5 pmol/L). The interassay analytical imprecision of 10 LC-MS/MS measurements over 5 days was between 9.3% to 25.0%, depending on the concentration and peptide, while the interassay imprecision by ELISA was 3.6% at 172 pmol/L and 4.2% at 244 pmol/L ($n = 15$) for TIMP2 and 9.4% at 757 pmol/L and 10.9% at 2053 pmol/L for IGFBP7 ($n = 20$) (Table 1). The mean intraassay of KIM-1 and NGAL ELISAs was obtained by calculating the percentage difference between duplicates in the calibration curves and were 5.9% and 7.3%, respectively.

LC-MS/MS analysis enables postanalytical data validity assessment for clinical samples utilizing embedded technical metadata, such as

Table 1. General imprecision for an in-house developed LC-MS/MS vs research-use only ELISA.

	LC-MS/MS					ELISA				
		Concentration		Interday imprecision		Concentration		Interday imprecision		
Biomarker	Measuring range (pmol/L)	IQC1[a] (pmol/L)	IQC2 (pmol/L)	IQC1 %CV	IQC2 %CV	IQC1 (pmol/L)	IQC2 (pmol/L)	IQC1 %CV	IQC2 %CV	Intraday imprecision CAL curve[b] %CV
NGAL	10–36 875	3890	2864	14.3	17.4	—	—	[c]	[c]	7.3[c]
IGFBP7	10–2615	294	537	13.0	25.0	757	2053	9.4	10.9	[d]
TIMP2	10–6285	569	1424	10.9	9.3	172	244	3.6	4.2	[d]
KIM-1	10–355	97	1798	19.1	16.4	—	—	[c]	[c]	5.9

[a]Internal quality control.
[b]Calibration curve.
[c]Interassay imprecision data of IQC samples not available.
[d]Not applicable, quality control data available.

measurand retention time, ion ratio, and internal standard signal intensity. The robustness of liquid chromatography performance, specified by retention time drift over all urine samples, ranged from CV = 0.66% to 1.21%. The analytical specificity was assessed by ion ratio monitoring, which is the ratio of qualifying product ion peak area over the quantifying product ion peak area. Ion ratios deviate ≤20% from the target at concentrations >100 pmol/L (Supplemental Fig. 1). An equal amount of internal standard was added to all samples to correct for variances introduced in the (pre)analytical phase; internal standard peak areas were monitored in clinical samples to identify outliers and/or matrix effects. Two samples were excluded based on internal standard signal (peak area cutoff ≤2000 counts) (Supplemental Fig. 2). Finally, the validity of the results was assessed through interpeptide comparisons for peptides originating from the same protein. For NGAL, IGFBP7, and TIMP2, results were in agreement, but the qualifying peptide of KIM-1 performed less well, yielding only 35 results eligible for comparison (Supplemental Fig. 3).

Concentrations obtained by LC-MS/MS and IA were compared and correlation coefficients and slopes for plots comparing the paired measurements were r = 0.926, slope = 2.41 (95% CI: 2.31, 2.70) for NGAL; r = 0.800, slope = 1.08 (95% CI: 0.95, 1.20) for IGFBP7; r = 0.846, slope = 0.43 (95% CI: 0.38, 0.47) for TIMP2; and r = 0.697, slope = 0.42 (95% CI: 0.38, 0.49) for KIM-1 (Fig. 1). Deviations from slope = 1 are largely due to the calibrators used for converting results into concentrations in molar or mass units, while lower r values may indicate variations in the measurand or lower measurement precision. The LC-MS/MS calibrator of IGFBP7 was value-assigned by the ELISA kit, resulting in a slope closer to 1 compared to the other 3 proteins, for which different calibrator strategies were used. The poor correlation and data scatter at concentrations <100 pmol/L for all biomarkers are likely due to increased measurement uncertainty. This is supported by unstable ion ratios observed at concentrations below approximately 100 pmol/L in LC-MS/MS analysis.

DISCUSSION

To explore the value and analytical performance of LC-MS/MS in biomarker translation, we made use of LC-MS/MS metadata for quality assessment and present a head-to-head comparison of IA-based ELISAs and MS-based test results for translating promising urinary biomarkers from

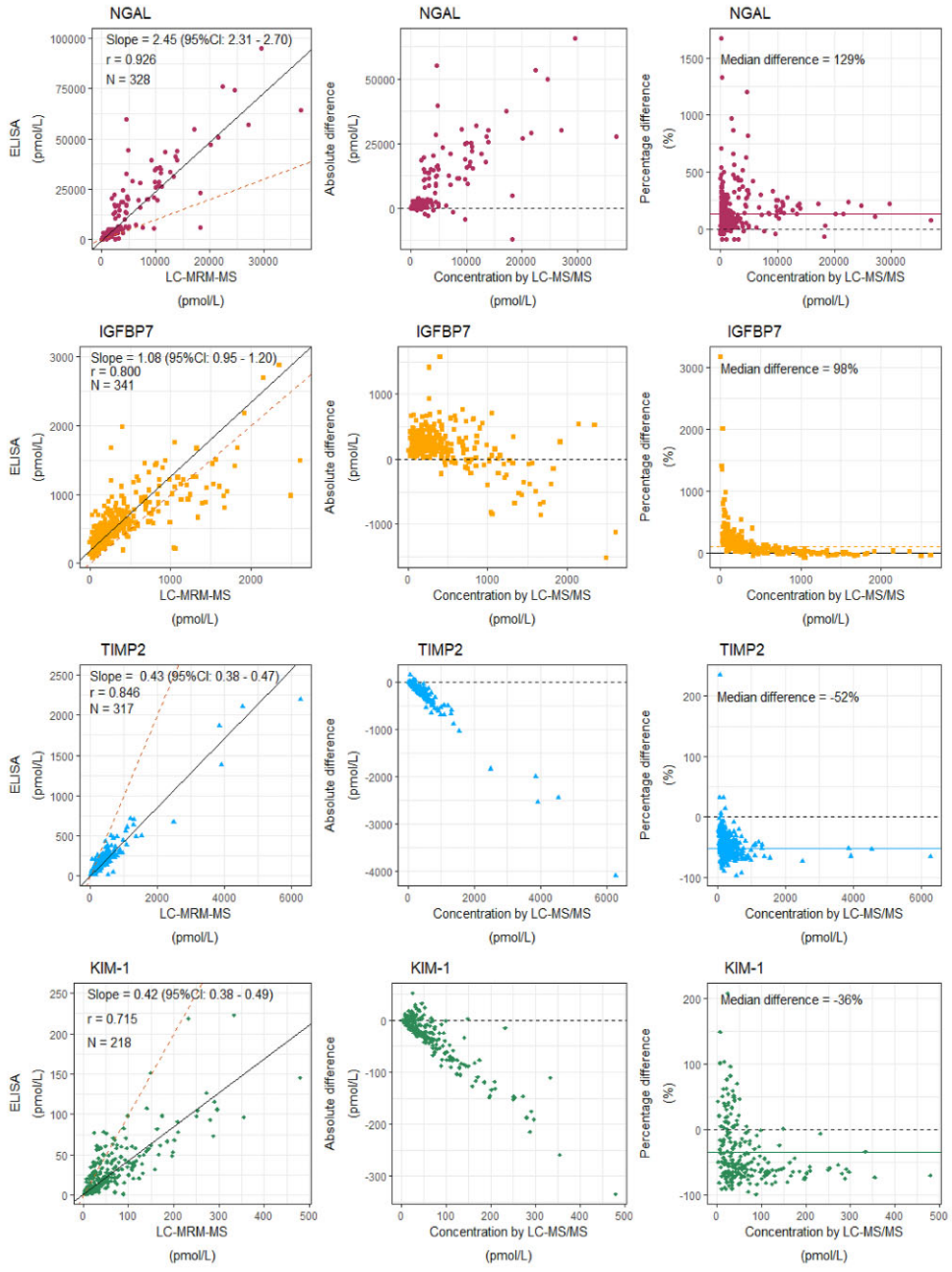

Fig. 1. Measurement procedure comparison of multiplex LC-MS/MS and singleplex research-use only ELISA kits for the quantitation of NGAL, IGFBP7, TIMP2, and KIM-1 from urine. Left column: Passing-Bablok regression (solid line) and line of identity (dashed line). Middle column: Difference plots, showing absolute differences between the paired measurements plotted against the results obtained by LC-MS/MS. Right column: Difference plots, showing percentage differences between the paired measurements plotted against the results obtained by LC-MS/MS, with the median percentage difference represented by solid lines.

bench to bedside. While LC-MS/MS allows confirmation of peptide (and hence protein) identities through ion ratio monitoring, the spectrophotometric detection of ELISAs does not; instead, in IAs whole protein complexes are measured and therefore, by design, analytically less selective than LC-MS/MS. However, it is important to note that the MS-based application described here is still in its exploratory phase. While it is sufficiently developed for use as a second-tier test (12), its imprecision and turnaround time cannot match the performance of commercially available ELISAs.

Calibration biases were observed for these urinary biomarkers in this method comparison, which can be explained by different (recombinant protein-based) calibration strategies and the lack of international recognized reference materials. The nonequivalence of test results between the IA- and MS-based analytical strategies emphasizes the need for test standardization from the initial development to increase transferability of results. Correspondingly, Ji et al. reported moderate NGAL test agreement with a slope of 0.71 (95% CI: −0.67, 0.77) between a particle-enhanced turbidimetric IA and an LC-MS/MS-based LDT (13).

Beyond calibration bias, the correlation between direct and indirect IA-based proteins tests was suboptimal. This sample-specific bias may be explained by differences in measurand and measurement uncertainty, which was relatively larger in our LC-MS/MS analyses. In IA, a mixture of different proteoforms from a single protein may be measured and the recovery is a function of theantibody types that have been used. The LC-MS/MS test, on the other hand, combines immunocapture, using polyclonal antibodies with variable epitope specificities, with a peptide-specific detection method. Importantly, antibody affinity and avidity are influenced by tertiary and quaternary protein structures, and while the kidney injury biomarker proteins are assumed to be soluble and freely present in urine, they are often part of protein complexes. Specifically, only the soluble cleaved ectodomain of transmembrane protein KIM-1 is quantified by IA (14). Furthermore, NGAL may be present in urine as monomer, homodimer, and heterodimer with MMP9 (15), while TIMP2 is detected in complex with its active substrate MMP9, and IGFBP7 is typically bound to insulin and its growth factors (16). The molecular presentation of the proteins of interest affect their affinities for the antibodies used in IAs, potentially leading to variable results (15).

Strengths of this study are the unveiling of both calibration and sample specific bias in LC-MS/MS test results compared to IA. In addition, MS-based tests will, by design, allow adequate characterization of calibrator(s) and selective measurement of potentially clinically relevant proteoforms in biological specimens.

There are limitations to this study that need to be acknowledged. First, the LC-MS/MS methodology was not compared with commercially available regulatory approved tests and therefore could only be considered for research use only. Because the current LC-MS/MS test can be considered as a second-tier test, it is too preliminary to compare the methodology to commercially available CE-marked medical tests. If the current LC-MS/MS test would be developed into a first-tier diagnostic test, comparison to commercially available tests is needed. Second, internal quality controls were used to investigate the between-day variances of NGAL, IGFBP7, TIMP2, and KIM-1, but imprecision data were incomplete for ELISA-based KIM-1 and NGAL measurements. Third, the urine specimens were stored long-term, up to 10 years at −80°C, prior to analysis, and an additional freeze–thaw cycle was required for LC-MS/MS analyses. Urine biomarker instability and/or degradation and unfolding is an important determinant of the preanalytical phase, which affects biomarker recovery. Besides the measurand, the analytical platform used in the analytical phase determines the susceptibility to biomarker recovery. Moreover, the epitope accessibility and

integrity determine the antibody-target affinity and resident time and protein recovery. The interplay between measurand intactness and technology can affect biomarker recovery. Previous studies evaluating biomarker stability were generally performed using IA. For instance, Pennemans et al. studied urinary KIM-1 stability and stated that freeze–thaw cycles should to be avoided (17), whereas Schuh et al. reported a decrease of <3% in KIM-1 and NGAL concentrations after 3 consecutive freeze–thaw cycles (18). In our hands, protein recovery by LC-MS/MS was considered stable for up to 4 freeze–thaw cycles (Supplemental Table 1).

In conclusion, we demonstrate that LC-MS/MS has clear potential as a next-generation measurement platform for absolute quantitation of urinary proteins. The advantages it offers over IAs are its analytical selectivity, a molecular definition of the measurands, and its potential for ab initio medical test standardization. Moreover, the versatility and multiplexing capability of MS facilitates efficient translational biomarker research and the development of in-house tests based on clinical needs. Finally, LC-MS/MS has the potential to capture biological complexity for future precision diagnostics.

SUPPLEMENTAL MATERIAL

Supplemental material is available at *The Journal of Applied Laboratory Medicine* online.

Nonstandard Abbreviations: IA, immunoassay; MS, mass spectometry; LDT, lab-developed test.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

All authors discussed the results and contributed to the final manuscript.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest*: **Employment or Leadership:** None declared. **Consultant or Advisory Role:** None declared. **Stock Ownership:** None declared. **Honoraria:** None declared. **Research Funding:** This work was partially funded by Roche Diagnostics and the Leiden University Medical Center. **Expert Testimony:** None declared. **Patents:** None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors thank Esther N.M. de Rooij for database management, Marte E. Koelemaij for LC-MS/MS analysis, and Fred P.H.T.M. Romijn and Sandra van der Kooij for specimen storage and ELISA analyses. We are grateful to Zain Odho for careful reading of our manuscript.

REFERENCES

1. Hoofnagle AN, Cobbaert CM, Delatour V, Kelleher NL, Lowenthal MS, Shuford CM. Should LC-MS/MS be the reference measurement procedure to determine protein concentrations in human samples? Clin Chem 2021;67:466–71.
2. Čuklina J, Lee CH, Williams EG, Sajic T, Collins BC, Rodríguez Martínez M, et al. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. Mol Syst Biol 2021;17:e10240.
3. Smit NPM, Ruhaak LR, Romijn F, Pieterse MM, van der Burgt YEM, Cobbaert CM. The time has come for quantitative protein mass spectrometry tests that target unmet clinical needs. J Am Soc Mass Spectrom 2021;32:636–47.
4. van Duijl TT, Soonawala D, de Fijter JW, Ruhaak LR, Cobbaert CM. Rational selection of a biomarker panel targeting unmet clinical needs in kidney injury. Clin Proteomics 2021;18:10.
5. Jones A, Dhanapala L, Kankanamage RNT, Kumar CV, Rusling JF. Multiplexed immunosensors and immunoarrays. Anal Chem 2020;92:345–62.
6. Hoofnagle AN, Wener MH. The fundamental flaws of immunoassays and potential solutions using tandem

mass spectrometry. J Immunol Methods 2009;347: 3–11.
7. Uettwiller-Geiger DL, Vijayendran R, Kellum JA, Fitzgerald RL. Analytical characteristics of a biomarker-based risk assessment test for acute kidney injury (AKI). Clin Chim Acta 2016;455:93–8.
8. Lippi G, Aloe R, Storelli A, Cervellin G, Trenti T. Evaluation of NGAL Test™, a fully-automated neutrophil gelatinase-associated lipocalin (NGAL) immunoassay on Beckman Coulter AU 5822. Clin Chem Lab Med 2011;50:1581–4.
9. van Duijl TT, Ruhaak LR, Smit NPM, Pieterse MM, Romijn F, Dolezal N, et al. Development and provisional validation of a multiplex LC-MRM-MS test for timely kidney injury detection in urine. J Proteome Res 2021;20:5304–14.
10. Veighey KV, Nicholas JM, Clayton T, Knight R, Robertson S, Dalton N, et al. Early remote ischaemic preconditioning leads to sustained improvement in allograft function after live donor kidney transplantation: long-term outcomes in the REnal Protection Against Ischaemia-Reperfusion in transplantation (REPAIR) randomised trial. Br J Anaesth 2019;123:584–91.
11. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, et al. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2004;32: D115–9.
12. Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol Cell Proteomics 2014;13: 907–17.
13. Ji H, Xu L, Su J, Shen L, Wang H, Wang J, et al. Absolute quantification of urinary neutrophil gelatinase-associated lipocalin by ultra-high-performance liquid chromatography/tandem mass spectrometry and the diagnostic efficacy of acute kidney injury. Rapid Commun Mass Spectrom 2020;34:e8637.
14. Zhang Z, Humphreys BD, Bonventre JV. Shedding of the urinary biomarker kidney injury molecule-1 (KIM-1) is regulated by MAP kinases and juxtamembrane region. J Am Soc Nephrol 2007;18:2704–14.
15. Makris K, Rizos D, Kafkas N, Haliassos A. Neurophil gelatinase-associated lipocalin as a new biomarker in laboratory medicine. Clin Chem Lab Med 2012;50: 1519–32.
16. Jin L, Shen F, Weinfeld M, Sergi C. Insulin growth factor binding protein 7 (IGFBP7)-related cancer and IGFBP3 and IGFBP7 crosstalk. Front Oncol 2020;10:727.
17. Pennemans V, Rigo JM, Penders J, Swennen Q. Collection and storage requirements for urinary kidney injury molecule-1 (KIM-1) measurements in humans. Clin Chem Lab Med 2011;50:539–43.
18. Schuh MP, Nehus E, Ma Q, Haffner C, Bennett M, Krawczeski CD, Devarajan P. Long-term stability of urinary biomarkers of acute kidney injury in children. Am J Kidney Dis 2016;67:56–61.