Received: 22 May 2023 | Revised: 8 September 2023 | Accepted: 11 September 2023

DOI: 10.1002/pmic.202300227

Proteomics
Proteomics and Systems Biology

RESEARCH ARTICLE

Multiple urinary peptides display distinct sex-specific distribution

Ioanna K. Mina[1,2] | Emmanouil Mavrogeorgis[1,2] | Justyna Siwy[1] | Riste Stojanov[3] | Harald Mischak[1] | Agnieszka Latosinska[1] | Vera Jankowski[2]

[1]Mosaiques Diagnostics GmbH, Hannover, Germany

[2]Institute for Molecular Cardiovascular Research, University Hospital RWTH Aachen, Aachen, Germany

[3]Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Skopje, North Macedonia

Correspondence
Vera Jankowski, Institute for Molecular Cardiovascular Research (IMCAR), University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany.
Email: vjankowski@ukaachen.de

Funding information
European Union's Horizon Europe Marie Skłodowska-Curie Actions Doctoral Networks—Industrial Doctorates; HORIZON—MSCA – 2021 – DN-ID, Grant/Award Number: 101072828; European Union's Horizon 2020 research and innovation program; H2020-MSCA-ITN-2019 "STRATEGY-CKD", Grant/Award Number: 860329; Deutsche Forschungsgemeinschaft' (DFG, German Research Foundation); Transregional Collaborative Research Centre; TRR 219; Project-ID, Grant/Award Number: 322900939; subproject S-03, Grant/Award Number: INST 948/4S-1; CRU 5011 project, Grant/Award Number: 445703531; Cost-Action CA, Grant/Award Number: 21165; IZKF Multiorgan complexity in Friedreich Ataxia; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 322900939, 445703531; European Cooperation in Science and Technology, Grant/Award Number: 21165; HORIZON EUROPE Marie Sklodowska-Curie Actions, Grant/Award Numbers: 101072828, 860329

Abstract

Previous studies have established the association of sex with gene and protein expression. This study investigated the association of sex with the abundance of endogenous urinary peptides, using capillary electrophoresis-coupled to mass spectrometry (CE-MS) datasets from 2008 healthy individuals and patients with type II diabetes, divided in one discovery and two validation cohorts. Statistical analysis using the Mann–Whitney test, adjusted for multiple testing, revealed 143 sex-associated peptides in the discovery cohort. Of these, 90 peptides were associated with sex in at least one of the validation cohorts and showed agreement in their regulation trends across all cohorts. The 90 sex-associated peptides were fragments of 29 parental proteins. Comparison with previously published transcriptomics data demonstrated that the genes encoding 16 of these parental proteins had sex-biased expression. The 143 sex-associated peptides were combined into a support vector machine-based classifier that could discriminate males from females in two independent sets of healthy individuals and patients with type II diabetes, with an AUC of 89% and 81%, respectively. Collectively, the urinary peptidome contains multiple sex-associated differences, which may enable a better understanding of sex-biased molecular mechanisms and the development of more accurate diagnostic, prognostic, or predictive classifiers for each individual sex.

KEYWORDS
biomarker, peptidomics, proteomics, sex, urine

Abbreviations: AKAP12, A-kinase anchor protein 12; BMI, body mass index; CD99, CD99 antigen; CE-MS, capillary electrophoresis-coupled to mass spectrometry; COL1A1, collagen alpha-1(I) chain; COL1A2, collagen alpha-2(I) chain; COL2A1, collagen alpha-1(II) chain; COL3A1, collagen alpha-1(III) chain; COL4A3, collagen alpha-3(IV) chain; COL5A2, collagen alpha-2(V) chain; COL6A1, collagen alpha-1(VI) chain; COL9A2, collagen alpha-2(IX) chain; eGFR, estimated glomerular filtration rate; GTEx, Genotype-Tissue Expression; HBA, hemoglobin subunit alpha; PIGR, polymeric immunoglobulin receptor; SHAP, Shapley additive explanations; SVM, support vector machine; TMSB4X, thymosin beta-4; UMOD, uromodulin; XG, glycoprotein Xg.

1 | INTRODUCTION

Various phenotypic characteristics, including anthropomorphic traits, sexual development, hormone regulation, immune responses, disease features, and life expectancy, are associated with sex [1–3]. Although societal, behavioral, and environmental factors may contribute to some of these differences, it is essential to identify the underlying biological and molecular mechanisms behind them. Genomic studies have revealed that genes of the X and Y chromosomes are unequally represented within the human male and female genome [1]. Furthermore, Oliva et al. [4] observed that many genes, located on sex and autosomal chromosomes, show sex-associated differences in their transcriptomic expression across human tissues. Several studies identified proteomic differences in human tissues and body fluids between males and females [5]. Examining the differences in urinary peptides and proteins between the sexes was expected to further add to our understanding of sex-biased protein abundance and turnover.

Urinary peptides are, to a large degree, specific protein degradation products collected in urine as a result of either glomerular filtration of blood or being derived from the kidneys and/or bladder [6]. Thus, the urinary peptidome reflects systemic and local (patho)physiology. The urine peptidome has been investigated in multiple studies using capillary electrophoresis-coupled to mass spectrometry (CE-MS). Its association with the pathology of major chronic diseases, such as diabetes, chronic kidney disease, cardiovascular disease, and different tumor types has been well-established [7–14]. Recently, a highly significant association of urine peptides with age and mortality was also demonstrated [15, 16]. Thus, the value of urinary peptides as biomarkers is firmly established. Surprisingly, though, the association with sex has not been investigated in much detail. Some differences in urinary peptidome signature between healthy males and females were reported previously by Dakna et al. [17] however the cohort used in that study was small and consisted of only of 67 male and 67 female healthy individuals. In this study, we investigated the differences in the urinary peptidome content between males and females in three large cohorts of more than 2000 subjects in total covering healthy participants, individuals from the general population, and patients with type II diabetes. Comparison of peptide abundance after matching for age and estimated glomerular filtration rate (eGFR) revealed a sex-associated signature of 90 peptides, with a consistent abundance trend in both the general population and the diabetic cohort. As such, urinary peptidome displays sex-associated differences, which may enable a better understanding of sex-biased molecular mechanisms and the development of more accurate diagnostic, prognostic, or predictive classifiers for each individual sex.

Significance Statement

In the current study, we established that the urinary peptidome displays sex-associated differences, which may enable a better understanding of sex-biased molecular mechanisms. Specifically, the association of 90 urinary peptides with sex could be verified across the general population and in patients with type II diabetes. These peptides were derived from 29 parental proteins, many of which have been previously associated with sex, supporting the validity of the approach. Differences in the levels of peptides from CD99 antigen (CD99), thymosin beta-4 (TMSB4X), polymeric immunoglobulin receptor (PIGR), uromodulin (UMOD), and hemoglobin subunit alpha (HBA) likely reflect the sex-biased expression of the respective genes. On the other hand, the differences in the abundances of collagen peptides may indicate the sex-biased degradation of collagens. Moreover, a classifier based on urinary peptides was capable of discriminating between males and females with good accuracy, independently of the presence of diabetes, further supporting the strong association of urinary peptides with sex. Furthermore, since urinary peptides have already been established as valuable biomarkers for many diseases, sex-associated molecular features may enable the development of more accurate diagnostic, prognostic, or predictive classifiers for each individual sex.

2 | MATERIALS AND METHODS

2.1 | CE-MS data

The CE-MS analysis procedures, including sample preparation, measurement, and data analysis, have been previously described in detail [18]. In brief, a P/ACE MDQ CE (Beckman Coulter, Fullerton, CA, USA) coupled to a micro-TOF-MS (Bruker Daltonic, Bremen, Germany) was used for CE-MS analysis. Technical aspects of the analytical process have been reported in details previously [6, 18, 19]. The raw CE-MS data were evaluated using the MosaFinder software assigning the detected signal in silico to the list of 21,559 peptides, with >5000 having information about the amino acid sequence, as described elsewhere [20]. Twenty-nine collagen fragments generally unaffected in disease served as internal standards for normalizing peptide intensities [21].

2.2 | Study population

Anonymized urinary peptidome, and phenotypic data were extracted from the Human Urinary Proteome Database [20]. Datasets from healthy individuals were selected based on the following criteria: (i) availability of phenotypic data, that is, sex, age, and eGFR (CKD-EPI equation [22]) and (ii) age ≥18 years old. Only the urinary peptidome datasets representing one sample per patient (baseline sample, in the case multiple specimens were available per individual) were considered. Following the same criteria, data from patients with type II diabetes were extracted. We decided to use the PRIORITY trial [13], as it is the most extensive study of patients with type II diabetes,

without detectable end-organ damage, in the Human Urinary Proteome Database. The subjects were divided into the discovery and two validation phases. The discovery cohort comprised healthy participants from numerous studies ($N = 454$) [23–29]. The findings were validated in the healthy subjects from the general population-based cohort FLEMENGO ($N = 377$) [30] and in patients with type II diabetes from the multicenter randomized controlled trial (PRIORITY) ($N = 1886$). Ethical review and approval were waived for this study by the ethics committee of the Hannover Medical School, Germany (no. 3116-2016), due to all data being fully anonymized.

2.3 | Case–control matching

Male and female individuals were matched based on age and eGFR in R ("MatchIt" package, default parameters) using the nearest neighbor method [31]. The matching was assisted toward more similar variable distributions between the two groups, through a series of participant removal experiments. The matching resulted in the selection of 116 male and 113 female individuals in the discovery cohort, 169 male and 168 female individuals in the FLEMENGHO cohort, and 721 male and 721 female individuals in the PRIORITY cohort. Peptidome data from these matched cohorts were investigated to define the sex-associated urinary peptides. The cohort determination process, including the matching step, is depicted in Figure S1. For each cohort, body mass index (BMI), systolic, and diastolic blood pressure were available for >50% of the matched patients.

2.4 | Statistical analysis

Analysis was limited to the 5071 peptides of known amino acid sequence. All missing abundance values were replaced with 0, as applied in previous peptidome studies [32]. For the discovery cohort, peptides of a frequency <30% in both groups (females and males) were excluded. The frequency threshold of 30% has been utilized in previous peptidome studies to deal with variables below the measurement threshold, while maintaining high coverage [32]. In addition, while biomarkers with lower occurrences might possess limited individual clinical utility, they may contribute positively to classifiers by enhancing their stability. For the validation cohorts, no frequency threshold was applied. Statistical analysis was performed between males and females for each cohort using the non-parametric Mann–Whitney test [17]. Subsequently, the calculated p-values were adjusted based on the Benjamini–Hochberg method [33]. Adjusted p-values < 0.05 were considered statistically significant. All the aforementioned calculations were completed in R (R version 4.2.1, R Foundation for Statistical Computing, Vienna, Austria) using the wilcox_test() function of the coin package and the p.adjust() function of the stats package, as applicable. The effect sizes were calculated as the absolute value of the test statistic divided by the square root of the sample size.

The Spearman's rank correlation analysis of the peptide fold changes (calculated by dividing the average value of peptide intensity between males and females) across the discovery as well as the two validation cohorts was conducted using MedCalc for Windows version 12.7.5.0 (MedCalc Software, Ostend, Belgium).

2.5 | Cross-omics analysis

For the cross-omics analysis, results of two previous studies that had examined sex-biased gene expression based on transcriptomic data from the Genotype-Tissue Expression (GTEx) project were investigated [34]. The top 500 significantly sex-biased genes per tissue were retrieved across 44 tissues from Oliva et al., 2020 [4]. Because extracellular matrix proteins generate most urinary peptides [6], 315 significantly sex-biased matrisome genes across 37 tissues were also retrieved from Nieuwenhuis et al., 2021 [35].

2.6 | Classifier development

Peptides significantly associated with sex in the discovery cohort were combined in a support vector machine (SVM)-based classifier with radial basis function kernel. Kernel functions project data from a lower-dimensional space into a higher-dimensional one, enhancing their potential to reveal distinct patterns [36]. The radial basis function operates by measuring the similarity between data points based on their relative distances. Points that are closer to each other exhibit higher similarity, enabling the model to effectively identify non-linear decision boundaries [37]. The classifier was trained to distinguish between males and females in the discovery cohort and its classification performance was evaluated in the FLEMENGHO and PRIORITY cohorts by calculating the AUC for each of them. To gain insights into how much the individual peptides contributed to the output of the classifier, Shapley additive explanations (SHAP) were used [38]. SHAP is a novel technique for interpreting machine learning models, which employs the concept of Shapley values from game theory to provide local explanations and allocate credit optimally. SHAP values are calculated by considering all possible combinations of peptides, observing how the model's predictions change for each combination, and then using weighted averages to determine the individual contributions of each peptide to the model's final prediction. All the above analysis was conducted in Python version 3.9.16 with the libraries sklearn v1.2.2, pandas v1.5.3, numpy v1.22.4, scipy v1.10.1, and shap v0.41.0. The code for these analyses is available at the following link: https://colab.research.google.com/drive/1k-vmRkfcCry1RtdCtouLfddL1HhheCHB?usp = sharing.

3 | RESULTS

3.1 | Cohort characteristics

The cohort determination process is graphically depicted in Figure S1. The clinical characteristics of male and female

participants in the discovery and validation cohorts are shown in Table S1. The average age of the participants was 56.94 (±14.34) years in the discovery cohort, 41.25 (±12.39) and 61.92 (±8.44) years in the FLEMENGHO and PRIORITY validation cohorts, respectively. Individuals in all three cohorts had an average eGFR > 80 mL/min/1.73 m^2, with the participants from the FLEMENGHO cohort showing the highest average eGFR, 92.49 (±13.62) mL/min/1.73 m^2, and the PRIORITY showing the lowest average eGFR, 84.38 (±17.48) mL/min/1.73 m^2. Male and female participants were matched within each cohort, ensuring no significant differences in average age and eGFR between the groups. As a result, any observed differences in peptide abundance between males and females are not affected by between group differences in these variables, providing a solid foundation for the subsequent analysis.

Individuals from the general population had the lowest average BMI, 25.29 (±4.0) kg/m^2, and as expected, diabetic individuals had the highest average BMI 30.64 (±5.28) kg/m^2. Participants of the FLEMENGHO study were also characterized with the lowest average systolic and diastolic blood pressures, 118.92 (±9.86) and 76.32 (±7.47) mmHg respectively, while the participants of the discovery cohort showed the highest average blood pressures, with 141.5 (±20.41) mmHg systolic and 80.97 (±10.77) mmHg diastolic blood pressure. BMI and blood pressure differed between males and females, mainly in the validation cohorts, as shown in Table S1.

3.2 | Sex-associated urinary peptides

The differences in urinary peptides between the discovery cohort's healthy male and female individuals were investigated. In a total dataspace of 5071 peptides (restricted to urine peptides with frequency ≥30% in at least one sex), we identified 143 peptides that showed statistically significant differences between the two sexes. Subsequently validation analysis was carried out to verify the association of these peptide with sex. Specifically, these peptides were investigated in two other cohorts consisting of healthy individuals (FLEMENGHO) and patients with type II diabetes (PRIORITY), respectively. This analysis resulted in the verification of 90 peptides that were found to be significantly associated with sex in at least one of the validation cohorts and showed agreement in the regulation trend across all three cohorts (Table S2). The fold changes of the 90 sex-associated peptides were significantly correlated ($p < 0.0001$) across the discovery and validation cohorts, with correlation coefficients ranging between 0.87 and 0.92 (Figure S2). The list of 20 peptides providing the most significant discrimination in the discovery cohort between males and females are listed in Table 1. The study design is depicted in Figure 1.

Additionally, the process of discovery followed by validation was repeated, using different frequency thresholds: 0%, 15%, 50%, and 70% in at least one sex. The analysis revealed that the majority of peptides present in the 90 sex-associated peptide signature, obtained using the 30% frequency threshold, were reproduced in the sex-associated signatures resulting from the analysis using various frequency thresholds. Notably, all the peptides reproduced across the thresholds maintained consistent regulation. As expected, the utilization of more conservative frequency thresholds led to the identification of a smaller number of sex-associated peptides, limiting the coverage and interpretation of the findings in the context of biology (Table S3).

Furthermore, the potential impact of menopause on the 90 sex-associated peptides was investigated in two distinct groups of healthy individuals, aged below 45 years and over 55 years, respectively. The expectation was that post-menopausal females in the >55 age group would exhibit greater similarity to males, resulting in fewer sex-specific differences in urinary peptides among individuals over 55 years old. Surprisingly, out of the 90 sex-associated peptides, 51 were found to be associated with sex in individuals below 45 years old, and 66 peptides were associated with sex in individuals over 55 years old, indicating that sex-specific differences become more pronounced after menopause (Table S4). Additionally, 32 peptides remained associated with sex independently of menopausal status. The average abundances of the top 10 sex-associated peptides in the two age groups are depicted in Figure 2. While these results provide insights into the menopausal impact on the sex-association of urinary peptides, further analysis is warranted within a well-defined cohort of females, both with and without menopause, based on comprehensive clinical records. The verified 90 sex-associated urinary peptides were fragments of 29 different proteins (Figure 3). Specifically, 15 non-collagen parental proteins were responsible for generating 34 sex-associated peptides, while the remaining 56 sex-associated peptides were derived from 14 different collagen proteins. In addition, a consistent regulation trend among all peptides belonging to the same parental protein, for proteins that give rise to at least two peptides, was found for CD99 antigen (CD99), uromodulin (UMOD), thymosin beta-4 (TMSB4X), collagen alpha-2(I) chain (COL1A2), collagen alpha-2(V) chain (COL5A2), and collagen alpha-3(IV) chain (COL4A3). Urinary levels of all peptides derived from CD99 ($n = 13$), UMOD ($n = 2$), and COL4A3 ($n = 2$) were found consistently higher in males in comparison to females. The abundance of peptides from TMSB4X ($n = 2$), COL1A2 ($n = 4$), and COL5A2 ($n = 2$) was consistently higher in females than males. Moreover, CD99 and TMSB4X fragments were among the 20 most significant findings (Table 1). Collagen alpha-1(I) chain (COL1A1), collagen alpha-1(II) chain (COL2A1), collagen alpha-1(III) chain (COL3A1), collagen alpha-1(VI) chain (COL6A1), collagen alpha-2(IX) chain (COL9A2), and polymeric immunoglobulin receptor (PIGR) were also responsible for the generation of at least two sex-associated peptides each. However, peptides derived from these proteins were not fully consistent in their regulation trends. A high number of sex-associated peptides were derived from COL1A1 ($n = 24$), with almost half of them (54%) being more abundant in females than males. From the COL2A1 fragments ($n = 3$), 67% were more abundant in females than males. From the COL3A1 fragments ($n = 11$), 73% were more abundant in males in comparison to females. Half of the peptides from COL6A1 ($n = 2$) and COL9A2 ($n = 2$) were more abundant in each sex. In addition, from the peptides derived from PIGR ($n = 6$), 67% were more abundant in females than males.

TABLE 1 The top 20 peptides showing the most significant association with sex.

Peptide sequence	Protein symbol	Fold change	*p*-value
DGVSGGEGKGGSDGGGSHRKEGEEADAPGVIPGIVGAVV	CD99	1.925	3.57E-06
sDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	TMSB4X	0.138	1.52E-05
GADGQpGAKGEpGDAGAKGDAGPpGPAGPAGPpGPIG	COL1A1	0.575	2.45E-05
VKQADSGSSEEKQLYNKYPDAVAT	SPP1	0.037	2.45E-05
DpGEAGPIGPKGYRGDEGPPGSEGARGAPGP	COL6A1	0.287	2.89E-05
DAGAPGApGGKGDAGApGERGpPG	COL3A1	1.533	2.03E-04
TGSpGSpGPDGKTGPpGP	COL1A1	0.619	2.49E-04
GADGQpGAKGEpGDAGAKGDAGPpGPAGP	COL1A1	0.409	2.49E-04
EPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPPGppGL	COL1A2	0.528	2.49E-04
sDKPDmAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	TMSB4X	0.132	2.49E-04
DRGEpGPpGP	COL7A1	0.536	2.67E-04
DGQPGAKGEPGDAGAKGDAGPpGP	COL1A1	1.666	3.82E-04
ApGDRGEPGPpGP	COL1A1	1.484	8.80E-04
ASVDSGSSEEQGGSSRALVSTLVPLG	PIGR	0.390	1.39E-03
ERTEEQKTEVEETAGSVPAEEL	AKAP12	5.327	1.44E-03
ADGVSGGEGKGGSDGGGSHRKEGEEADAPGVIPG	CD99	3.250	1.44E-03
DADLADGVSGGEGKGGSDGGGSHRKEGEEADAPGVIPGIVGAVV	CD99	1.632	1.44E-03
TGSpGSpGPDGKTGPpGPAG	COL1A1	0.805	1.57E-03
NTGApGSPGVSGPKGDAGQpGEKGSpGAQGpPGAPGPLG	COL3A1	1.343	1.63E-03
DGVSGGEGKGGSDGGGSHRKEGEEADAPGVIPG	CD99	1.540	2.10E-03

Peptides were selected from the list of 90 sex-associated peptides based on the lowest adjusted *p*-values in the discovery cohort. Fold change: fold changes in the discovery cohort, calculated by dividing average value of peptide intensity in males and average value of peptide intensity in females. *p*-value: adjusted *p*-values in the discovery cohort. In the peptide sequences N-terminal acetylated serine and hydroxyproline, and oxidized methionine, are indicated with small p, s, and m, respectively. AKAP12, A-kinase anchor protein 12; CD99, CD99 antigen; COL1A1, collagen alpha-1(I) chain; COL1A2, collagen alpha-2(I) chain; COL3A1, collagen alpha-1(III) chain; COL6A1, collagen alpha-1(VI) chain; COL7A1, collagen alpha-1(VII) chain; PIGR, polymeric immunoglobulin receptor; SPP1, osteopontin; TMSB4X, thymosin beta-4.

3.3 | Sex-biased parental proteins' expression across different tissues

We investigated whether the parental proteins that generate the sex-associated peptides show differential expression between males and females. For that purpose, the top 500 significantly sex-associated genes across 44 tissues, retrieved from Oliva et al., 2020 [4] and the 315 significantly sex-associated matrisome genes across 37 tissues, retrieved from Nieuwenhuis et al., 2021 [35], were examined. Seventeen genes encoding 16 parental proteins of the sex-associated peptides were differentially expressed between males and females at the transcription level in at least one tissue (Table 2).

Among the non-collagen parental proteins, the most prominent transcriptomic findings concerned CD99, glycoprotein Xg (XG) and hemoglobin subunit alpha (HBA). *CD99* showed higher expression in males than in females in all 44 tissues [4], in agreement with the consistently higher abundance of CD99 urinary peptides in males. *XG* was expressed higher in females in agreement with the higher abundance of one XG urinary fragment in females. The two genes encoding HBA, *HBA1* and *HBA2*, were expressed higher in males in agreement with the higher abundance of one HBA urinary fragment in males. Regarding collagens, only *COL1A2* and *COL6A1* showed consistent upregulation in males in at least two tissues. However, COL1A2 urinary fragments were found consistently more abundant in females than in males and COL6A1 derived fragments did not show consistent regulation between the two sexes.

3.4 | Sex classification based on the urinary peptidome signature

To investigate the association of the overall urinary peptidome signature with sex, the 143 peptides that were associated with sex in the discovery cohort were combined in an SVM classifier, trained in the discovery cohort and validated in the FLEMENGHO and PRIORITY cohorts. The AUCs of the classifier were 0.89 (95% CI 0.7975–0.87) and 0.81 (95% CI 0.7745–0.822), for the FLEMENGHO and PRIORITY cohort, respectively (Figure 4). To further validate the impact of the individual peptides for sex prediction, SHAP analysis with permutation explainer on the FLEMENEGHO and PRIORITY cohorts using the SVM

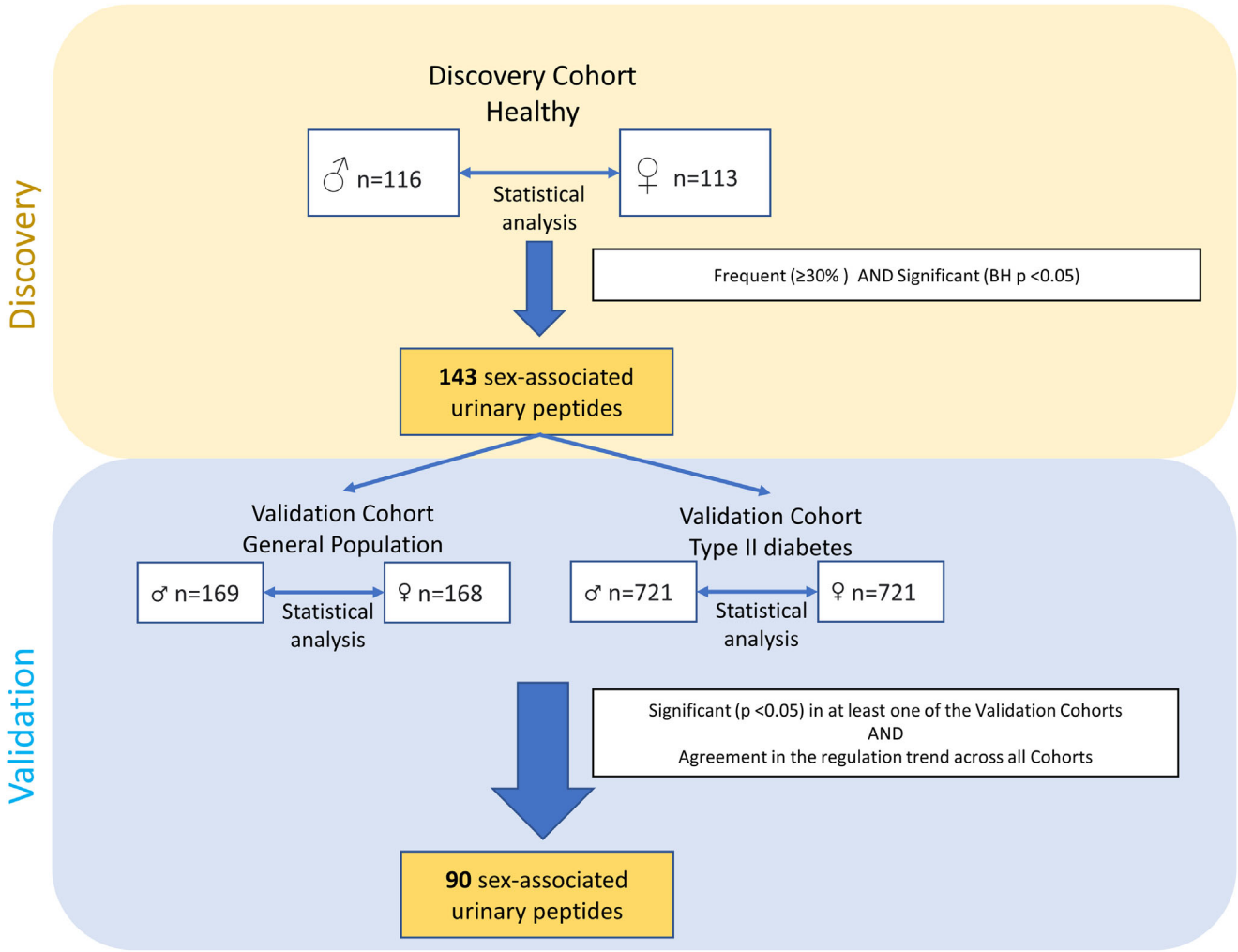

FIGURE 1 Study design. After selection of peptides with frequency ≥30% in at least one sex, statistical analysis in the discovery cohort resulted in the identification of 143 peptides with significant differences between males (♂) and females (♀) ($p < 0.05$). Subsequently, separate statistical analyses in each validation cohort showed that of these 143 peptides, 90 were significantly associated with sex ($p < 0.05$) in at least one of the validation cohorts and had the same regulation trend (i.e., consistent increase or decrease of peptide abundance in males in comparison to females) across the discovery and validation cohorts. BH, Benjamini-Hochberg.

classifier was conducted (Table S5). There was a very high overlap between the top 10 peptides, with the highest impact observed in FLEMENGHO and PRIORITY cohorts. Fragments from CD99, COL1A1, COL1A2, COL3A1, collagen alpha-1(VII) chain (COL7A1), and A-kinase anchor protein 12 (AKAP12) were among the top 10 peptides that impacted mostly the output of the classifier in both cohorts. Fragments from XG and COL6A1 were among the 10 most impactful peptides on the output in the FLEMENGHO and PRIORITY cohort, respectively (Figure 4).

3.5 | Comparison of sex-associated urinary peptides in health and diabetes

We further investigated differences between sex-associated urinary peptides defined in healthy and diabetic individuals. Out of the 100 most significant sex-associated peptides in healthy subjects, 75 were also significantly associated with sex in patients with type II diabetes. Of these, 73 peptides showed the same regulation trend between males and females across healthy individuals and patients with type II diabetes. Similarly, when investigating the top 100 most significant sex-associated peptides in patients with type II diabetes, 62 were also significant in the pooled analysis of healthy subjects, with 58 of them displaying a consistent regulation trend (Table S6).

These findings indicate that a considerable number of significant peptides associated with sex overlap between healthy and diabetic subjects, supporting further their association with sex rather than underlying disease. Of note: the lack of significance in certain cases could be attributed to the varying statistical power of the analysis when investigating healthy and diabetic subjects. Therefore, while some peptides may appear to be significant only in either the healthy or diabetic group, the interpretation of these findings should be considered with caution, considering the differences in cohort size.

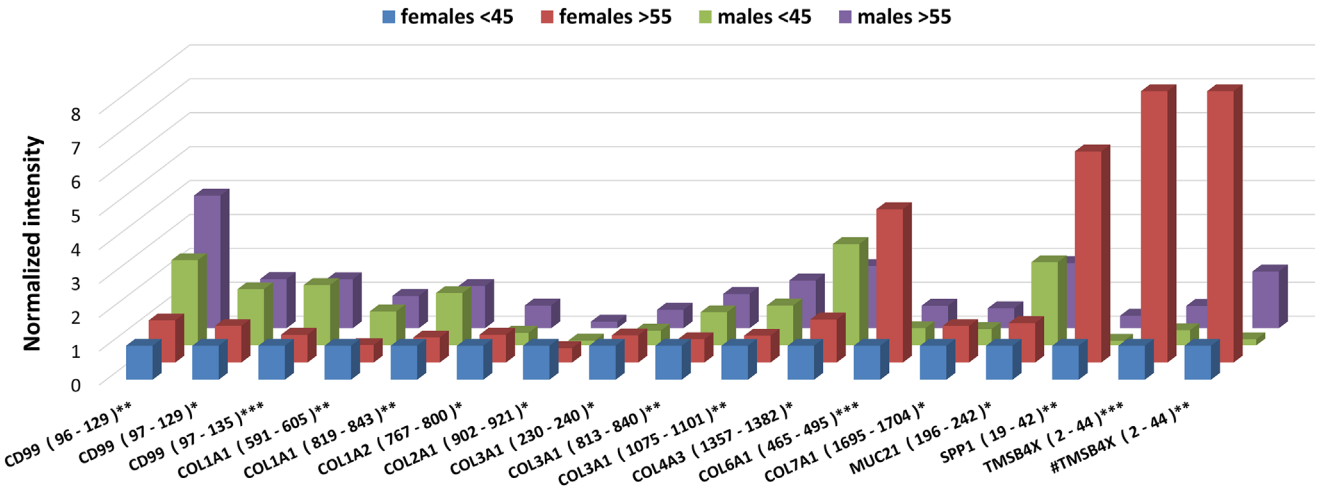

FIGURE 2 Menopausal impact on intensity values of the 90 sex-associated peptides. Bar graph of the intensity values of the top 10 statistically most significant sex-associated peptides within healthy individuals with age <45 (*) and >55 (**). Notably, the three peptides that rank as highly significant in both age groups are denoted with ***. The intensity values have been normalized based on the intensity values from females under the age of 45. The peptide TMSB4X with an oxidized methionine is denoted by the symbol "#". CD99, CD99 antigen; COL1A1, collagen alpha-1(I) chain; COL1A2, collagen alpha-2(I) chain; COL2A1, collagen alpha-1(II) chain; COL3A1, collagen alpha-1(III) chain; COL4A3, collagen alpha-3(IV) chain; COL6A1, collagen alpha-1(VI) chain; COL7A1, collagen alpha-1(VII) chain; MUC21, mucin-21; SPP1, osteopontin; TMSB4X, thymosin beta-4.

4 | DISCUSSION

The results of the current study demonstrated that sex is associated with a specific urinary peptidome signature. Specifically, 90 sex-associated peptides with preserved regulation across independent cohorts of healthy individuals and patients with type II diabetes were identified. Of these 90 peptides, 51 were more abundant in males and 39 were increased in females. These peptides were fragments of 29 parental proteins, from which 15 originated from non-collagen proteins and 14 from collagens.

One of the most striking findings was a consistent and highly significant increase in the abundance of all 13 CD99 peptides detected in urine of males compared to females. CD99 is a broadly expressed cell surface protein participating in various biological processes, including cell adhesion, migration, death, differentiation, and diapedesis, and impacts inflammation and immune responses [39]. CD99 is encoded by a pseudoautosomal *CD99* gene located near the boundary of the PAR1 region, found in both X and Y chromosomes. It is known that pseudoautosomal genes typically escape X chromosome inactivation in females [40]. However, given *CD99* location, spreading the X chromosome inactivation signal across the boundary may likely result in the partial silencing of the *CD99* in females [40], thus, higher expression in males. Along these lines, Johnston and colleagues [40] observed a higher *CD99* expression in lymphoblastoid cell lines derived from males (n = 105) than those derived from females (n = 105). The extensive recent transcriptomics study [4] by Oliva et al. also supports higher levels of *CD99* expression in males across a wide range of investigated tissue. Furthermore, Lefèvre and colleagues [41] found an increased percentage of monocytes expressing CD99 in healthy males (n = 12) compared to females (n = 12). These observations align with our findings and may suggest that the male-biased CD99 urinary peptidome signature results from the higher expression of CD99 in males.

Another important finding was the identification of two TMS4BX fragments showing a highly significant increase in abundance in females. TMS4BX is an actin-sequestering protein thought to be involved in many biological functions, including tissue regeneration, wound healing, and inflammation [42]. The TMS4BX protein is encoded by a *TMS4BX* gene in the X chromosome that probably escapes X inactivation [43]. In that case, the expression of *TMS4BX* by both X chromosomes in females and by only one X chromosome in males would lead to the higher expression of *TMS4BX* in females. Previous research [44] has shown higher abundance of TMSB4X in female amniocyte samples (n = 15) compared to male samples (n = 14). In contrast, in a recently published transcriptomics study [4], *TMS4BX* expression was reported increased in males, however, this study did not include tissues with the highest *TMSB4X* expression (bone marrow and lymphoid tissue) [45]. Collectively, it seems that increased levels of TMS4BX peptides in female urine are associated with a probable escape of *TMS4BX* from X inactivation. Interestingly, *TMSB4X* has a homolog on the Y chromosome (*TMSB4Y*), encoding a protein (TMSB4X, Y-chromosomal) with slightly different sequence, expressed at very low levels across male tissues [34, 46]. Most likely due to the low expression, TMSB4Y peptides are not detected in the urinary peptidome [6]. Given the potential protective role of TMS4BX in heart failure and chronic kidney disease [47, 48], diseases exhibiting specific sex-differences in epidemiology and pathophysiology [49, 50], as well as previous findings indicating elevated plasma TMS4BX levels specifically in women with heart failure [47], further research is warranted to explore the molecular functions of TMS4BX in these conditions, particularly in the context of sex differences. With its anti-inflammatory and anti-fibrotic properties [51], increased levels of TMS4BX in females

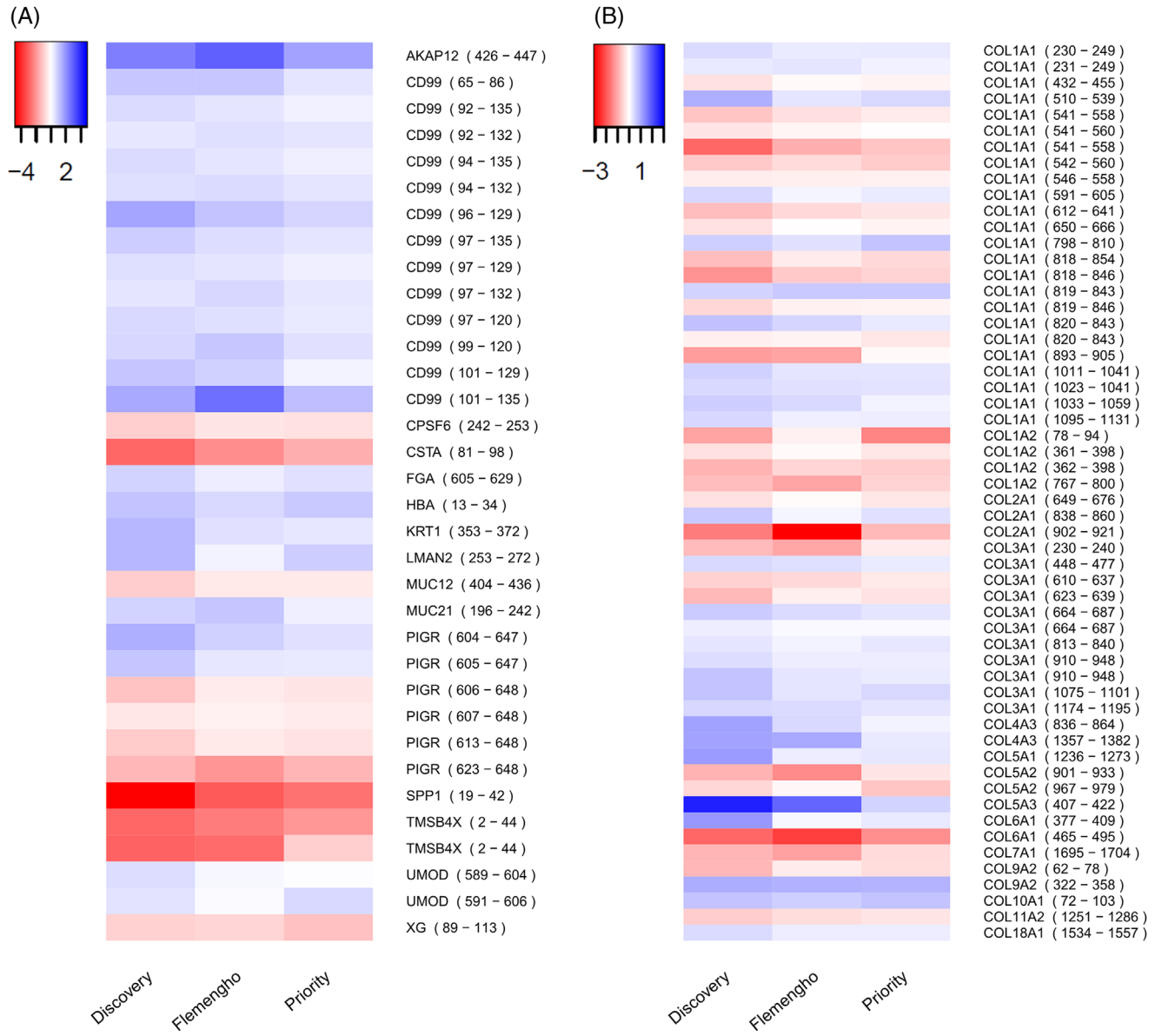

FIGURE 3 Sex specific regulation of urinary peptides. Heatmaps depicting the $\log_2$ fold change ($\log_2$FC) of the 90 sex-associated urinary peptides across the validation and discovery cohorts for non-collagen peptides (A), and collagen peptides (B). Fold change is the ratio of the average value of peptide intensity in males to the respective of females. The lower-positioned TMSB4X peptide contains an oxidized methionine. AKAP12, A-kinase anchor protein 12; CD99, CD99 antigen; COL1A1, collagen alpha-1(I) chain; COL1A2, collagen alpha-2(I) chain; COL2A1, collagen alpha-1(II) chain; COL3A1, collagen alpha-1(III) chain; COL4A3, collagen alpha-3(IV) chain; COL5A1, collagen alpha-1(V) chain; COL5A2, collagen alpha-2(V) chain; COL5A3, collagen alpha-3(V) chain; COL6A1, collagen alpha-1(VI) chain; COL7A1, collagen alpha-1(VII) chain; COL9A2, collagen alpha-2(IX) chain; COL10A1, collagen alpha-1(X) chain; COL11A2, collagen alpha-2(XI) chain; COL18A1, collagen alpha-1(XVIII) chain; CPSF6, cleavage and polyadenylation specificity factor subunit 6; CSTA, cystatin-A; FGA, fibrinogen alpha chain; HBA, hemoglobin subunit alpha; KRT1, keratin, type II cytoskeletal 1; LMAN2, vesicular integral-membrane protein VIP36; MUC12, mucin-12; MUC21, mucin-21; PIGR, polymeric immunoglobulin receptor; SPP1, osteopontin; TMSB4X, thymosin beta-4; UMOD, uromodulin; XG, glycoprotein Xg.

may offer protection against these diseases. Investigating this aspect using animal models could help elucidate the mechanisms by which TMS4BX might exert sex-specific protective effects in heart failure and chronic kidney disease.

Moreover, a consistently higher level of both sex-associated UMOD fragments was observed in males. UMOD is the most abundant protein in urine. It is produced exclusively by renal tubular epithelial cells and plays a role in many biological functions, such as renal ion transport and immunomodulation [52]. Previous research has shown sex differences in excretion levels of urinary UMOD [53], possibly leading to the increased abundance of UMOD peptides in male urine.

An additional interesting finding was the identification of 6 PIGR sex-associated fragments, most of which had increased abundance in females. PIGR is an epithelial glycoprotein that mediates the transportation of immunoglobulin A across mucosal epithelial cells [54]. PIGR is possibly regulated by sex hormones [54]. The expression

TABLE 2 Cross-omics analysis.

	Number of urinary sex-associated peptides		Number of tissues with sex-biased expression			
			Oliva et al., 2020 [4]		Nieuwenhuis et al., 2021 [35]	
Gene symbol	♂	♀	♂	♀	♂	♀
CD99	13	0	44	0		
CSTA	0	1			1	1
HBA1	1	0	14	0		
HBA2	1	0	10	0		
KRT1	1	0	0	1		
SPP1	0	1			0	1
TMSB4X	0	2	1	0		
XG	0	1	0	19		
COL1A2	0	4	4	0	2	0
COL4A3	2	0	3	0	1	2
COL5A1	1	0			1	0
COL5A2	0	2	1	0		
COL6A1	1	1	1	0	1	0
COL7A1	0	1			2	2
COL9A2	1	1	10	2	1	2
COL10A1	1	0			1	0
COL11A2	0	1			0	1

Using the results from two transcriptomic studies, genes encoding 16 parental proteins of sex-associated urinary peptides were found to have sex-biased expression. The number of sex-associated peptides with abundance increased in male (♂) or female (♀) and the number of tissues with higher expression in male (♂) or female (♀) are represented here. CD99, CD99 antigen; COL10A1, collagen alpha-1(X) chain; COL11A2, collagen alpha-2(XI) chain; COL1A2, collagen alpha-2(I) chain; COL4A3, collagen alpha-3(IV) chain; COL5A1, collagen alpha-1(V) chain; COL5A2, collagen alpha-2(V) chain; COL6A1, collagen alpha-1(VI) chain; COL7A1, collagen alpha-1(VII) chain; COL9A2, collagen alpha-2(IX) chain; CSTA, cystatin-A; HBA, hemoglobin subunit alpha; KRT1, keratin, type II cytoskeletal 1; SPP1, osteopontin; TMSB4X, thymosin beta-4; XG, glycoprotein Xg.

pattern of PIGR in the human endometrium during the estrous cycle indicates that it is positively regulated by estrogen and negatively regulated by progesterone [54], which may explain the increase of PIGR peptides in female urine.

The transcriptomic analysis [4] indicated that differential expression of *HBA1/HBA2* and *XG* between the two sexes might also be the reason for the observed sex association of their peptides. The *HBA1* and *HBA2* genes encode HBA. Two HBA subunits and two HBB subunits form hemoglobin, the most abundant protein in erythrocytes, which is involved in transporting oxygen from the lungs to the tissues [55]. *HBA1* and *HBA2* were expressed consistently at higher levels in male tissues, in agreement with the higher abundance of one HBA1-derived fragment in males. XG, together with CD99, constitute the Xg blood system [56]. *XG* was consistently expressed at higher levels in female tissues, consistent with the higher abundance of one XG-derived peptide in females. XG, as CD99, is encoded by a gene in the PAR1 region and escapes X inactivation in females [57]. However, the copy of this gene in the Y chromosome does not produce a specific product [58]. The expression of the *XG* from both X chromosomes in females but only from one X chromosome in males might result in the higher expression of *XG* in females and a higher abundance of the XG-derived peptide in female urine.

Collagens, the most abundant proteins in mammals, are deposited in the extracellular matrix and contribute to the organization, shape and mechanical properties of the tissues [59]. Excessive collagen accumulation causes fibrosis, a common pathological component across many chronic diseases with significant epidemiological differences between the sexes [2]. COL1A2, COL5A2, and COL4A3 fragments showed a consistent sex-associated regulation, with the abundance of all peptides for the first two collagens being increased in females while the last one being increased in males. Other collagens (COL1A1, COL2A1, COL3A1, COL6A1, and COL9A2) were also responsible for the generation of at least two sex-associated peptides each, however, displaying an apparent inconsistent distribution: up- and downregulation were observed. The increase or decrease in peptide abundance was consistent across different cohorts, indicating validity of the finding. The exact mechanisms explaining these observations are unknown. Based on the previously published transcriptomics analysis, the observed results cannot be explained by increased or decreased gene expression [4, 35]. The urinary collagen peptides are protein degradation products, a process affected by multiple factors apart from protein expression. Collagens are characterized by extensive post-translational modifications and crosslinking and are degraded by various proteases. It is reasonable to assume that collagen, the substrate for proteolytic degradation, is in a state of saturation. Thus, the generated peptides do not reflect the abundance of collagens but rather the specific protease activity. Previous research suggests that sex hormones regulate collagen remodeling. Kehlet et al. [60] observed that women in menopause have an increased collagen turnover as compared to men of the same age. Furthermore, estrogens seem to display a protective effect against fibrosis, with in vivo and in vitro models indicating that estrogens suppress collagen synthesis and upregulate collagen degradation [2].

Based on the 143 peptides that were significantly associated with sex in the discovery cohort and using the same cohort as a training set, a sex-discriminating SVM classifier was generated. The performance of the classifier across in two independent validation sets (AUC 0.89 and 0.81) with individuals from the general population and patients with type II diabetes, respectively, highlights a strong association of the urinary peptidome with sex. Investigation of the impact of each individual peptide on the classification outcomes revealed, as expected, specific collagen and CD99 fragments as the most impactful in both validation sets.

Collectively, this study identified 90 urinary peptides displaying sex-specific differences. These peptides were derived from 29 parental proteins, many of which have been previously associated with sex, supporting the validity of the approach. Differences in the levels of peptides from CD99, TMSB4X, PIGR, UMOD, and HBA likely reflect the sex-biased expression of the respective genes. On the other hand, the differences in the abundances of collagen peptides may indicate the

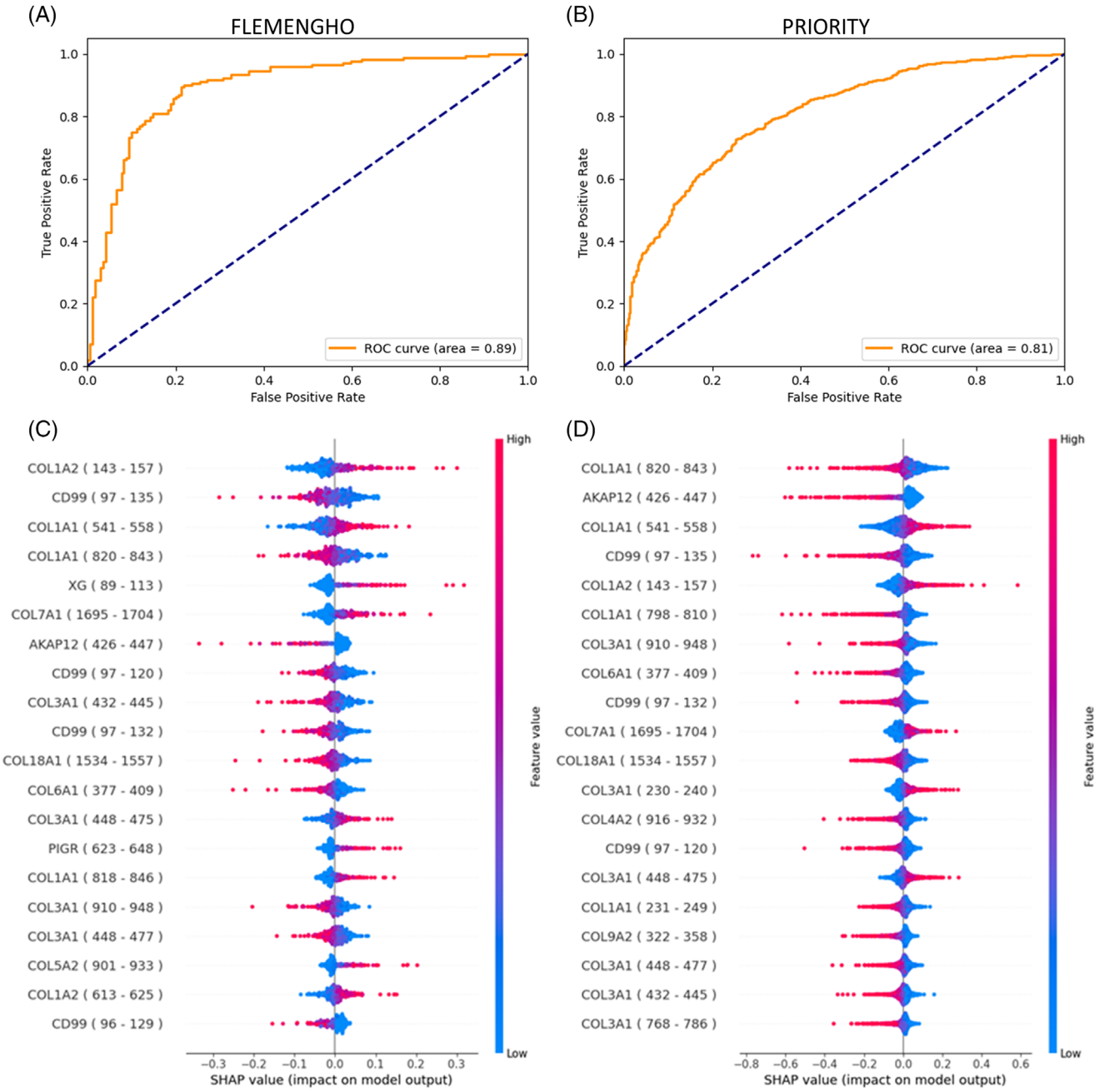

FIGURE 4 Classifier development. The performance of the sex-discriminating classifier, based on the 143 peptides that were associated with sex in the discovery cohort, was evaluated in the validation cohorts. The ROC curves of the classifier in the FLEMENGHO (A) and PRIORITY (B) cohort are shown here. Subsequently, SHAP analysis was conducted to evaluate the impact of the individual peptides for sex prediction. The plots represent the SHAP values (x-axis) of every sample, for the 20 peptides (y-axis) with the highest absolute mean SHAP values in the FLEMENGHO (C) and PRIORITY (D) cohort. The peptide values are represented with color. For example, the higher values for peptide COL1A2 (143–157) improve, while lower values have a negative impact, on the classifier performance in both cohorts. AKAP12, A-kinase anchor protein 12; CD99, CD99 antigen; COL1A1, collagen alpha-1(I) chain; COL1A2, collagen alpha-2(I) chain; COL3A1, collagen alpha-1(III) chain; COL4A2, collagen alpha-2(IV) chain; COL5A2, collagen alpha-2(V) chain; COL6A1, collagen alpha-1(VI) chain; COL7A1, collagen alpha-1(VII) chain; COL9A2, collagen alpha-2(IX) chain; COL18A1, collagen alpha-1(XVIII) chain; PIGR, polymeric immunoglobulin receptor; XG, glycoprotein Xg.

sex-biased degradation of collagens. As a result of these data, we plan to investigate the hypothesis that collagen degradation is differentiated between males and females. Furthermore, since urinary peptides have already been established as valuable biomarkers for many diseases such as kidney disease and heart failure [10, 20], which are known for displaying sex differences in epidemiology and pathophysiology, the sex-associated peptidome signature may contribute to the development of more accurate classifiers, to further improve on early diagnosis, prognosis, and prediction of treatment response for each individual sex.

ACKNOWLEDGMENTS

Ioanna K. Mina was supported by a grant from European Union's Horizon Europe Marie Skłodowska-Curie Actions Doctoral Networks—Industrial Doctorates Programme (HORIZON—MSCA – 2021 – DN-ID, grant number 101072828). Emmanouil Mavrogeorgis was supported by a grant from the European Union's Horizon 2020 research and innovation program (H2020-MSCA-ITN-2019 "STRATEGY-CKD", grant number 860329). Vera Jankowski was supported by a grant from the 'Deutsche Forschungsgemeinschaft' (DFG, German Research Foundation) through the Transregional Collaborative Research Centre (TRR 219; Project-ID 322900939, subproject S-03, INST 948/4S-1); CRU 5011 project number 445703531, Cost-Action CA 21165, IZKF Multiorgan complexity in Friedreich Ataxia.

CONFLICT OF INTEREST STATEMENT

Harald Mischak is the founder and co-owner of Mosaiques Diagnostics (Hannover, Germany). Ioanna K. Mina, Emmanouil Mavrogeorgis, Justyna Siwy, and Agnieszka Latosinska are employed by Mosaiques Diagnostics. The remaining authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Proposals will be reviewed and approved by the authors with scientific merit and feasibility as the criteria. After approval of a proposal, data can be shared via a secure online platform after signing a data access and confidentiality agreement. Data will be made available for a maximum of 5 years after a data sharing agreement has been signed.

ORCID

Ioanna K. Mina https://orcid.org/0009-0006-4099-665X
Justyna Siwy https://orcid.org/0000-0003-1407-2534
Agnieszka Latosinska https://orcid.org/0000-0001-8917-2412

REFERENCES

1. Arnold, A. P. (2019). Rethinking sex determination of non-gonadal tissues. *Current Topics in Developmental Biology*, *134*, 289–315.
2. Garate-Carrillo, A., Gonzalez, J., Ceballos, G., Ramirez-Sanchez, I., & Villarreal, F. (2020). Sex related differences in the pathogenesis of organ fibrosis. *Translational Research*, *222*, 41–55.
3. Zarullia, V., Kashnitskya, I., & Vaupela, J. W. (2021). Death rates at specific life stages mold the sex gap in life expectancy. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2010588118.
4. Oliva, M., Muñoz-aguirre, M., Kim-hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S. E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E. A., Skol, A. D., Garrido-Martín, D., Reverter, F., Brown, A., … Stranger, B. E. (2020). The impact of sex on gene expression across human tissues. *Science*, 369, eaba3066.
5. Gianazza, E., Miller, I., Guerrini, U., Palazzolo, L., Parravicini, C., & Eberini, I. (2018). Gender proteomics I. Which proteins in non-sexual organs. *Journal of Proteomics*, *178*, 7–17.
6. Mavrogeorgis, E., Mischak, H., Latosinska, A., Siwy, J., Jankowski, V., & Jankowski, J. (2021). Reproducibility evaluation of urinary peptide detection using CE-MS. *Molecules (Basel, Switzerland)*, *26*, 7260.
7. Argilés, À., Siwy, J., Duranton, F., Gayrard, N., Dakna, M., Lundin, U., Osaba, L., Delles, C., Mourad, G., Weinberger, K. M., & Mischak, H. (2013). CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS ONE*, *8*, e62837.
8. Frantzi, M., Van Kessel, K. E., Zwarthoff, E. C., Marquez, M., Rava, M., Malats, N., Merseburger, A. S., Katafigiotis, I., Stravodimos, K., Mullen, W., Zoidakis, J., Makridakis, M., Pejchinovski, M., Critselis, E., Lichtinghagen, R., Brand, K., Dakna, M., Roubelakis, M. G., Theodorescu, D., … Anagnou, N. P. (2016). Development and validation of urine-based peptide biomarker panels for detecting bladder cancer in a multi-center study. *Clinical Cancer Research*, *22*, 4077–4086.
9. Frantzi, M., Metzger, J., Banks, R. E., Husi, H., Klein, J., Dakna, M., Mullen, W., Cartledge, J. J., Schanstra, J. P., Brand, K., Kuczyk, M. A., Mischak, H., Vlahou, A., Theodorescu, D., & Merseburger, A. S. (2014). Discovery and validation of urinary biomarkers for detection of renal cell carcinoma. *Journal of Proteomics*, *98*, 44–58.
10. Latosinska, A., Siwy, J., Faguer, S., Beige, J., Mischak, H., & Schanstra, J. P. (2021). Value of urine peptides in assessing kidney and cardiovascular disease. *Proteomics – Clinical Applications*, *15*, e2000027.
11. Metzger, J., Negm, A. A., Plentz, R. R., Weismüller, T. J., Wedemeyer, J., Karlsen, T. H., Dakna, M., Mullen, W., Mischak, H., Manns, M. P., & Lankisch, T. O. (2013). Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut*, *62*, 122–130.
12. Rossing, K., Mischak, H., Dakna, M., Zürbig, P., Novak, J., Julian, B. A., Good, D. M., Coon, J. J., Tarnow, L., & Rossing, P. (2008). Urinary proteomics in diabetes and CKD. *Journal of the American Society of Nephrology*, *19*, 1283–1290.
13. Tofte, N., Lindhardt, M., Adamova, K., Bakker, S. J. L., Beige, J., Beulens, J. W. J., Birkenfeld, A. L., Currie, G., Delles, C., Dimos, I., Francová, L., Frimodt-Møller, M., Girman, P., Göke, R., Havrdova, T., Heerspink, H. J. L., Kooy, A., Laverman, G. D., Mischak, H., … Kilic, C. (2020). Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): A prospective observational study and embedded randomised placebo-controlled trial. *The Lancet Diabetes & Endocrinology*, *8*, 301–312.
14. Zhang, Z. Y., Thijs, L., Petit, T., Gu, Y. M., Jacobs, L., Yang, W. Y., Liu, Y. P., Koeck, T., Zürbig, P., Jin, Y., Verhamme, P., Voigt, J. U., Kuznetsova, T., Mischak, H., & Staessen, J. A. (2015). Urinary proteome and systolic blood pressure as predictors of 5-year cardiovascular and cardiac outcomes in a general population. *Hypertension*, *66*, 52–60.
15. Martens, D. S., Thijs, L., Latosinska, A., Trenson, S., Siwy, J., Zhang, Z. Y., Wang, C., Beige, J., Vlahou, A., Janssens, S., Mischak, H., Nawrot, T. S., Staessen, J. A., Asayama, K., Bochud, M., Boggia, J., Brguljan-Hitij, J., Feng, Y. M., Gu, Y. M., … Maestre, G. E. (2021). Urinary peptidomic profiles to address age-related disabilities: A prospective population study. *The Lancet Healthy Longevity*, *2*, e690–e703.
16. Nkuipou-Kenfack, E., Latosinska, A., Yang, W. Y., Fournier, M. C., Blet, A., Mujaj, B., Thijs, L., Feliot, E., Gayat, E., Mischak, H., Staessen, J. A., Mebazaa, A., & Zhang, Z. Y. (2020). A novel urinary biomarker predicts 10-year mortality after discharge from intensive care. *Critical Care*, *24*, 1–11.
17. Dakna, M., Harris, K., Kalousis, A., Carpentier, S., Kolch, W., Schanstra, J. P., Haubitz, M., Vlahou, A., Mischak, H., & Girolami, M. (2010). Addressing the challenge of defining valid proteomic biomarkers and classifiers. *BMC Bioinformatics*, *11*, 1–16.
18. Mischak, H., Vlahou, A., & Ioannidis, J. P. A. (2013). Technical aspects and inter-laboratory variability in native peptide profiling: The CE-MS experience. *Clinical Biochemistry*, *46*, 432–443.
19. Mischak, H., Kolch, W., Aivaliotis, M., Bouyssié, D., Court, M., Dihazi, H., Dihazi, G. H., Franke, J., Garin, J., De Peredo, A. G., Iphöfer, A., Jänsch, L., Lacroix, C., Makridakis, M., Masselon, C., Metzger, J., Monsarrat, B., Mrug, M., Norling, M., … Vlahou, A. (2010). Comprehensive human

urine standards for comparability and standardization in clinical proteome analysis. *Proteomics – Clinical Applications*, *4*, 464–478.
20. Latosinska, A., Siwy, J., Mischak, H., & Frantzi, M. (2019). Peptidomics and proteomics based on CE-MS as a robust tool in clinical application: The past, the present, and the future. *Electrophoresis*, *40*, 2294–2308.
21. Jantos-Siwy, J., Schiffer, E., Brand, K., Schumann, G., Rossing, K., Delles, C., Mischak, H., & Metzger, J. (2009). Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. *Journal of Proteome Research*, *8*, 268–281.
22. Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro 3rd, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., & Coresh, J., CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). (2009). A new equation to estimate glomerular filtration rate. *Annals of Internal Medicine*, *150*, 604–612.
23. Campbell, R. T., Jasilek, A., Mischak, H., Nkuipou-Kenfack, E., Latosinska, A., Welsh, P. I., Jackson, C. E., Cannon, J., Mcconnachie, A., Delles, C., & Mcmurray, J. J. V. (2020). The novel urinary proteomic classifier HF1 has similar diagnostic and prognostic utility to BNP in heart failure. *ESC Heart Failure*, *7*, 1595–1604.
24. Cannon, J. A., Moffitt, P., Perez-Moreno, A. C., Walters, M. R., Broomfield, N. M., McMurray, J. J. V., & Quinn, T. J. (2017). Cognitive Impairment in Heart Failure: Systemativ Review and Meta-Analysis. *J Card Fail.*, *23*(6), 464–475.
25. Catanese, L., Siwy, J., Mavrogeorgis, E., Amann, K., Mischak, H., Beige, J., & Rupprecht, H. (2021). A novel urinary proteomics classifier for non-invasive evaluation of interstitial fibrosis and tubular atrophy in chronic kidney disease. *Proteomes*, *9*, 32.
26. Neisius, U., Koeck, T., Mischak, H., Rossi, S. H., Olson, E., Carty, D. M., Dymott, J. A., Dominiczak, A. F., Berry, C., Oldroyd, K. G., & Delles, C. (2016). Urine proteomics in the diagnosis of stable angina. *BMC Cardiovascular Disorders*, *16*, 1–8.
27. Schanstra, J. P., Zürbig, P., Alkhalaf, A., Argiles, A., Bakker, S. J. L., Beige, J., Bilo, H. J. G., Chatzikyrkou, C., Dakna, M., Dawson, J., Delles, C., Haller, H., Haubitz, M., Husi, H., Jankowski, J., Jerums, G., Kleefstra, N., Kuznetsova, T., Maahs, D. M., … Vanholder, R. (2015). Diagnosis and prediction of CKD progression by assessment of urinary peptides. *Journal of the American Society of Nephrology*, *26*, 1999–2010.
28. Siwy, J., Zürbig, P., Argiles, A., Beige, J., Haubitz, M., Jankowski, J., Julian, B. A., Linde, P. G., Marx, D., Mischak, H., Mullen, W., Novak, J., Ortiz, A., Persson, F., Pontillo, C., Rossi, P., Rupprecht, H., Schanstra, J. P., Vlahou, A., … Vanholder, R. (2017). Noninvasive diagnosis of chronic kidney diseases using urinary proteome analysis. *Nephrology Dialysis Transplantation*, *32*, 2079–2089.
29. Snell-Bergeon, J. K., Maahs, D. M., Ogden, L. G., Kinney, G. L., Hokanson, J. E., Schiffer, E., Rewers, M., & Mischak, H. (2009). Evaluation of urinary biomarkers for coronary artery disease, diabetes, and diabetic kidney disease. *Diabetes Technology and Therapeutics*, *11*, 1–9.
30. Zhang, Z., Staessen, J. A., Thijs, L., Gu, Y., Liu, Y., Jacobs, L., Koeck, T., Zürbig, P., Mischak, H., & Kuznetsova, T. (2014). Left ventricular diastolic function in relation to the urinary proteome: A proof-of-concept study in a general population. *International Journal of Cardiology*, *176*, 158–165.
31. Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*, 1–28.
32. An, D. W., Yu, Y. L., Martens, D. S., Latosinska, A., Zhang, Z. Y., Mischak, H., Nawrot, T. S., & Staessen, J. A. (2023). Statistical approaches applicable in managing OMICS data: Urinary proteomics as exemplary case. Mass Spectrometry Reviews.
33. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300.
34. Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*, 580–585.
35. Nieuwenhuis, T. O., Rosenberg, A. Z., Mccall, M. N., & Halushka, M. K. (2021). Tissue, age, sex, and disease patterns of matrisome expression in GTEx transcriptome data. *Scientific Reports*, *11*, 21549.
36. Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*, 1565–1567.
37. Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and Knowledge Discovery*, *2*, 121–167.
38. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*, 4766–4775.
39. Pasello, M., Manara, M. C., & Scotlandi, K. (2018). CD99 at the crossroads of physiology and pathology. *Journal of Cell Communication and Signaling*, *12*, 55–68.
40. Johnston, C. M., Lovell, F. L., Leongamornlert, D. A., Stranger, B. E., Dermitzakis, E. T., & Ross, M. T. (2008). Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genetics*, *4*, e9.
41. Lefèvre, N., Corazza, F., Duchateau, J., Desir, J., & Casimir, G. (2012). Sex differences in inflammatory cytokines and CD99 expression following in vitro lipopolysaccharide stimulation. *Shock (Augusta, Ga.)*, *38*, 37–42.
42. Wang, H., Qian, Y., Wang, J. W., Fang, Y., Fan, Y. C., Liu, H. H., & Wang, K. (2023). Hypomethylation of thymosin β4 promoter is associated with glucocorticoid therapy in patients with acute-on-chronic hepatitis B-induced liver failure. *International Health*, *15*, 19–29.
43. Wainer Katsir, K., & Linial, M. (2019). Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics*, *20*, 1–17.
44. Liu, Y., Zhang, X., Zhang, L., Zhu, H., Chen, J., Lin, Z., Zhou, B., Liu, S., Wang, H., & Sun, H. (2022). Sex differences in protein expression and their perturbations in amniotic fluid cells of down syndrome fetuses. *ACS Omega*, *7*, 35981–35992.
45. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., & Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, *347*, 1260419.
46. UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023 – Google Scholar. *Nucleic Acids Research*, *51*, 523–531.
47. Drum, C. L., Tan, W. K. Y., Chan, S. P., Pakkiri, L. S., Chong, J. P. C., Liew, O. W., Ng, T. P., Ling, L. H., Sim, D., Leong, K. T. G., Yeo, D. P. S., Ong, H. Y., Jaufeerally, F., Wong, R. C. C., Chai, P., Low, A. F., Davidsson, P., Liljeblad, M., Söderling, A. S., … Richards, A. M. (2017). Thymosin beta-4 is elevated in women with heart failure with preserved ejection fraction. *Journal of the American Heart Association*, *6*, e005586.
48. Vasilopoulou, E., Riley, P. R., & Long, D. A. (2018). Thymosin-beta4: A key modifier of renal disease. *Expert Opinion on Biological Therapy*, *18*, 185–192.
49. Carrero, J. J., Hecking, M., Chesnaye, N. C., & Jager, K. J. (2018). Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. *Nature Reviews Nephrology*, *14*, 151–164.
50. Taylor, A. L. (2015). Heart failure in women. *Current Heart Failure Reports*, *12*, 187–195.
51. Xing, Y., Ye, Y., Zuo, H., & Li, Y. (2021). Progress on the function and application of thymosin beta4. *Frontiers in Endocrinology*, *12*, 767785.
52. Schaeffer, C., Devuyst, O., & Rampoldi, L. (2021). Uromodulin: Roles in health and disease. *Annual Review of Physiology*, *83*, 477–501.
53. Steubl, D., Buzkova, P., Ix, J. H., Devarajan, P., Bennett, M. R., Chaves, P. H. M., Shlipak, M. G., Bansal, N., Sarnak, M. J., & Garimella, P. S. (2020). Association of serum and urinary uromodulin and their correlates in

older adults—The Cardiovascular Health Study. *Nephrology*, *25*, 522–526.
54. Kaetzel, C. S. (2005). The polymeric immunoglobulin receptor: Bridging innate and adaptive immune responses at mucosal surfaces. *Immunological Reviews*, *206*, 83–99.
55. Barrera-Reyes, P. K., & Tejero, M. E. (2019). Genetic variation influencing hemoglobin levels and risk for anemia across populations. *Annals of the New York Academy of Sciences*, *1450*, 32–46.
56. Thornton, N. M., & Grimsley, S. P. (2019). Clinical significance of antibodies to antigens in the ABO and Xg blood group systems. *Immunohematology*, *35*, 95–101.
57. Helena Mangs, A., & Morris, B. (2007). The human pseudoautosomal region (PAR): Origin, function and future. *Current Genomics*, *8*, 129–136.
58. Weller, P. A., Critcher, R., Goodfellow, P. N., German, J., & Ellis, N. A. (1995). The human Y chromosome homologue of XG: Transcription of a naturally truncated gene. *Human Molecular Genetics*, *4*, 859–868.
59. Ricard-blum, S. (2011). The collagen family. *Cold Spring Harbor Perspectives in Biology*, *3*, a004978.
60. Kehlet, S. N., Willumsen, N., Armbrecht, G., Dietzel, R., Brix, S., Henriksen, K., & Karsdal, M. A. (2018). Age-related collagen turnover of the interstitial matrix and basement membrane: Implications of age- and sex-dependent remodeling of the extracellular matrix. *PLoS ONE*, *13*, e0194458.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202300227 in the Supporting Information section at the end of the article.

How to cite this article: Mina, I. K., Mavrogeorgis, E., Siwy, J., Stojanov, R., Mischak, H., Latosinska, A., & Jankowski, V. (2024). Multiple urinary peptides display distinct sex-specific distribution. *Proteomics*, *24*, e2300227. https://doi.org/10.1002/pmic.202300227