# **Osteoporosis in Relation to a Bone-Related Aging Biomarker Derived from the Urinary Proteomic Profile: A Population Study**

**Yu-Ling Yu, Dries S. Martens, De-Wei An, Babangida Chori, Agnieszka Latosinska, Justyna Siwy, Augustine N. Odili, Katarzyna Stolarz-Skrzypek, Gladys E. Maestre, Kei Asayama, Yan Li, Peter Verhamme, Karel Allegaert, Harald Mischak, Tim S. Nawrot, Jan A. Staessen**

#### *Supplementary Data*

#### *Osteoporosis in Relation to a Bone-Related Aging Biomarker Derived From the Urinary Proteomic Profile: A Population Study*

Yu-Ling Yu, Dries S. Martens, De-Wei An, Babangida Chori, Agnieszka Latosinska, Justyna Siwy, Augustine N. Odili, Katarzyna Stolarz-Skrzypek, Gladys E. Maestre, Kei Asayama, Yan Li, Peter Verhamme, Karel Allegaert, Harald Mischak, Tim S. Nawrot, Jan A. Staessen

#### **Table of Contents**



#### **Urinary Proteomics**

#### *Sample preparation and CE-MS analysis*

As described in detail elsewhere [1], urine aliquots were thawed and 700 μl mixed with 700 μL of 2 M urea, 10 mM NH<sub>4</sub>OH containing 0.02 % SDS. Subsequently, samples were ultrafiltered using a Centristat 20 kDa cut-off centrifugal filter device (Satorius, Göttingen, Germany) to eliminate high molecular weight proteins. The obtained filtrate was desalted using a PD 10 gel filtration column (GE Healthcare Bio Sciences, Uppsala, Sweden) to remove urea, electrolytes and salts as well as to enrich the polypeptides. The samples were lyophilized and stored at 4°C before usage. Shortly before CE-MS analysis, the samples were re-suspended in 10  $\mu$ L HPLC-grade H<sub>2</sub>O. Samples were injected into CE-MS with 2 psi for 99 sec, resulting in injection volumes of ~280 nL.

A P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA) was coupled with a Micro-TOF MS (Bruker Daltronic, Bremen, Germany). A solution of 20% acetonitrile (Sigma-Aldrich, Taufkirchen, Germany) in HPLC-grade H<sub>2</sub>O (Roth, Karlsruhe, Germany) supplemented with 0.94% formic acid (Sigma-Aldrich) was used as running buffer. For CE-MS analysis, the electrospray ionization interface from Agilent Technologies (Palo Alto, CA) was set to a potential of -4.0 to -4.5 kV. Spectra were recorded over an *m*/*z* range of 350-3000 and accumulated every 3 s.

#### *CE-MS data processing*

After the CE-MS analysis, mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaFinder software [2]. Only signals with z>1 observed in a minimum of 3 consecutive spectra with a signal-to-noise ratio of at least 4 were considered. The resulting peak list characterizes each polypeptide by its mass and migration time. Data were calibrated utilizing 3151 internal standards as reference data points for mass and migration time by applying global and local linear regression, respectively as described previously. Reference signals of 29 abundant peptides were used as internal standards for calibration of signal intensity using local linear regression [3]. This procedure was shown to be an easy and reliable method to address both analytical and dilution variances and differences in glomerular filtration in a single calibration step. This procedure was shown to be highly reproducible and to address both analytical and dilution variances in a single calibration step [3]. The obtained peak list characterizes each polypeptide by its calibrated molecular mass (Da), calibrated CE migration time (min) and normalized signal intensity. All detected peptides were deposited, matched, and annotated in a Microsoft SQL database allowing further statistical analysis.

#### *Sequencing of peptides*

Candidate biomarkers were sequenced using CE-MS/MS or LC-MS/MS analysis, as described in detail [4]. MS/MS experiments were using an Ultimate 3000 nano-flow system (Dionex/LC Packings, Thermo Fisher Scientific Inc., Waltham, MA) or a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA), both connected to an LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific GmbH, Bremen, Germany) equipped with a nano-electrospray ion source. The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. Survey full-scan MS spectra (from *m*/*z* 300–2,000) were acquired in the Orbitrap. Ions were sequentially isolated for fragmentation. Data files were searched against the

UniProt human nonredundant database, using Proteome Discoverer 2.4 and the SEQUEST search engine. Relevant settings were: no fixed modifications, oxidation of methionine and proline as variable modifications. The minimum precursor mass was set to 790 Da, maximum precursor mass to 6000 Da with a minimum peak count of 10. The high-confidence peptides were defined by cross-correlation  $(Xcorr) > 1.9$  and rank = 1. Precursor mass tolerance and fragment mass tolerance were 5 ppm and 0.05 Da, respectively. For further validation of obtained peptide sequences, the correlation between peptide charge at the working pH of 2 and CE-migration time was utilized to minimize incorrect sequence assignment [5]. Calculated CE-migration time of the sequence candidate based on its peptide sequence (number of basic amino acids) was compared to the experimental migration time.

#### **References**

- 1. Mischak H, Vlahou A, Ioannidis JPA (2013). Technical aspects and inter-laboratory variability in native peptide profiling: the CE-MS experience. Clin Biochem*,* 46: 432-43.
- 2. 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: 2294-308.
- 3. Jantos-Siwy J, Schiffer E, Brand K, Schumann G, Rossing K, Delles C, et al. (2009). Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. J Proteome Res*,* 8: 268-81.
- 4. Klein J, Papadopoulos T, Mischak H, Mullen W (2014). Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. Electrophoresis*,* 35: 1060-4.
- 5. Zürbig P, Renfrow MB, Schiffer E, Novak J, Walden M, Wittke S, et al. (2006). Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platformindependent separation. Electrophoresis*,* 27: 2111-25.



**Supplementary Table 1.** Bone-related function of 19 parental protein identified by 395 sequenced urinary peptides.

For analysis, 395 sequenced urinary peptides derived from 19 parental proteins with a detectable signal in ≥70% FLEMENGHO participants were selected (see Figure 1). Accession numbers refer to the Uniprot database [\(http://www.uniprot.org/uniprot\)](http://www.uniprot.org/uniprot), which substantiated the bone-related function of the parental proteins.



**Supplementary Table 2.** Association of 99 bone-related sequenced urinary peptides with chronological age in the derivation dataset (starts)



**Supplementary Table 2** Association of 99 bone-related sequenced urinary peptides with chronological age in the derivation dataset (continued)



**Supplementary Table 2.** Association of 99 bone-related sequenced urinary peptides with chronological age in the derivation dataset (continued)



**Supplementary Table 2.** Association of 99 bone-related sequenced urinary peptides with chronological age in the derivation dataset (continued)

**Supplementary Table 2.** Association of 99 bone-related sequenced urinary peptides with chronological age in the derivation dataset (ends)



FLEMENGHO: Flemish Study on Environment, Genes, and Health Outcomes. The lower-case letters in the amino-acid sequence identify post-translational modifications: c, disulphide bridges; m, oxidized methionine; n, deaminated asparagine; p, hydroxylated proline; k, hydroxylated lysine; q, deaminated glutamine. Accession numbers refer to the Uniprot database [\(http://www.uniprot.org/uniprot\)](http://www.uniprot.org/uniprot). Association sizes (β) between urinary peptides and chronological age were expressed per 10-year increment and were adjusted for the clustering within families, sex, body mass index, the glomerular filtration rate calculated from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration equation, current smoking, γ-glutamyltransferase, fasting plasma glucose, physical activity, use of diuretics (yes *vs* no), and socioeconomic status (High or middle *vs* low). *p*-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate for 395 tested associations.





Only urinary peptides with the highest significance per identified protein were selected for this analysis to demonstrate consistency of the associations in 519 participants included in the derivation and the time-shifted validation datasets and in the 187 participants included in the synchronous validation dataset. Association sizes (β) of chronological age with the urinary peptides were expressed per 10-year increment and were adjusted for the clustering within families, sex, body mass index, the glomerular filtration rate calculated from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration equation, current smoking, γ-glutamyltransferase, fasting plasma glucose, physical activity, use of diuretics (yes *vs* no), and socioeconomic status (high or middle *vs* low). The *p*-values indicate the validation status ( $p < 0.10$ , given the prior probability established by the significance in the derivation dataset). NS indicates not significant.



**Supplementary Table 4.** Number of replicated bone-related urinary peptides identifying parental proteins

Replication indicates similar directionality in the association with chronological age with a *p-value* of <0.10 (given the prior probability established by the significance in the derivation dataset).



**Supplementary Table 5.** Sex difference of bone-related urinary peptides in the derivation dataset (starts)



**Supplementary Table 5.** Sex difference of bone-related urinary peptides in the derivation dataset (ends)

NS: not significant. Estimate was the association size of difference in urinary peptide level in women compared to men (men as reference) derived from the ranknormalized data in the unadjusted and adjusted general linear models. Adjusted models accounted for the clustering within families, age, body mass index, the glomerular filtration rate calculated from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration equation, current smoking, γglutamyltransferase as index of alcohol intake, blood glucose, physical activity, use of diuretics (yes *vs.* no), and the middle or high socioeconomic status (yes vs. no). The 99 urinary peptides associated with chronological age were included in the analysis in the derivation dataset. *P-values* were corrected for the Benjamini-Hochberg false discovery rate for 99 tested association. 37 (9 parental proteins) was identified in the unadjusted models and 32 urinary peptides (9 parental proteins) in the adjusted models, respectively.



**Supplementary Table 6.** Weights assigned by elastic net regression to the multidimensional UPP biomarker reflecting aging (starts)



**Supplementary Table 6.** Weights assigned by elastic net regression to the multidimensional UPP biomarker reflecting aging (ends).

The derivation dataset included the baseline data of 519 participants in FLEMENGHO. Accession numbers refer to the Uniprot database [\(http://www.uniprot.org/uniprot\)](http://www.uniprot.org/uniprot). Association sizes (β) between chronological age and the levels of the urinary peptides were derived by elastic net regression, which was bootstrapped 1000 times. The regression slopes (βs) reflect the weight of each of the 29 peptide fragments in the construction of the multidimensional UPPost-age biomarker. Of 99 peptides entering the data reduction analysis, 29 entered in at least 60% of models and identified 10 proteins. The 95% confidence intervals of the βs were derived from the bootstrap distribution.



#### **Supplementary Table 7.** Risk of having HASD ≤ -3 cm in FLEMENGHO participants at baseline (2005-2010).

Abbreviations: HASD: the difference of body height minus arm span; n/N: number of participants with  $HASD \leq -3$  cm/total number of participants; UPPost-age: age as predicted by the urinary proteomic profile; OR: odds ratio; 95% CI: 95% confidence interval. ORs, given with 95% CI, express the relative risk per 10-year increment in chronological age or UPPost-age with adjustment for clustering within families, sex (in all participants only), current smoking, γ-glutamyltransferase, fasting plasma glucose, physical activity, use of diuretics (yes *vs* no) and socioeconomic status (high and middle *vs* low).



**Supplementary Figure 1. Correlation between height loss and body-height-to-arm-span ratio in 656 FLEMENGHO participants.** HASR: body-height-to-arm-span ratio. *r* is the Pearson correlation coefficient. Height change was the differences between two body height measurement examined in the interval from 1985 until 2008. During the median time interval of 9.56 years (interquartile range: 6.54, 15.5 years), the mean height change was 0.46 cm (SD: 1.51 cm) in 656 participants (50.6% women; median age at the first measurement: 40.5 year). Height change was significantly correlated with HASR  $(r = -0.17, p < 0.0001)$ . The sex-specific correlation coefficients were  $-0.24$  in women (red circles) and  $-$ 0.095 in men (green squares), respectively ( $p = 0.057$  for the sex difference).

### SUPPLEMENTARY DATA



**Supplementary Figure 2. Regression plot of dp-ucMGP with chronological age (A) and UPPost-age (B) in 597 FLEMENGHO participants.** *r* is the Pearson correlation coefficient. dp-ucMGP denotes desphospho-uncarboxylated matrix Gla protein. UPPost-age is age as predicted by the urinary proteomic profile.

## SUPPLEMENTARY DATA