

# Urinary peptidomic profiles to address age-related disabilities: a prospective population study

Dries S Martens, Lutgarde Thijs, Agnieszka Latosinska, Sander Trenson, Justyna Siwy, Zhen-Yu Zhang, Congrong Wang, Joachim Beige, Antonia Vlahou, Stefan Janssens, Harald Mischak, Tim S Nawrot, Jan A Staessen, for the FLEMENGHO investigators\*



## Summary

**Background** The Global Burden of Diseases, Injuries, and Risk Factors Study 2019 called for innovation in addressing age-related disabilities. Our study aimed to identify and validate a urinary peptidomic profile (UPP) differentiating healthy from unhealthy ageing in the general population, to test the UPP predictor in independent patient cohorts, and to search for targetable molecular pathways underlying age-related chronic diseases.

**Methods** In this prospective population study, we used data from participants in the Flemish Study on Environment, Genes and Health Outcomes (FLEMENGHO), done in northern Belgium from 1985 to 2019, and invited participants to a follow-up examination in 2005–10. Participants were eligible if their address was within 15 km of the examination centre and if they had not withdrawn consent in any of the previous examination cycles (1985–2004). All participants (2005–10) were also invited to an additional follow-up examination in 2009–13. Participants who took part in both the 2005–10 follow-up examination and in the additional 2009–13 follow-up visit constituted the derivation dataset, which included their 2005–10 data, and the time-shifted internal validation dataset, which included their 2009–13 data. The remaining participants who only had 2005–10 data constituted the synchronous internal validation dataset. Participants were excluded from analyses if they were incapacitated, had not undergone UPP, or had either missing or outlying (three SDs greater than the mean of all consenting participants) values of body-mass index, plasma glucose, or serum creatinine. The UPP was assessed by capillary electrophoresis coupled with mass spectrometry. The multidimensional UPP signature reflecting ageing was generated from the derivation dataset and validated in the time-shifted internal validation dataset and the synchronous validation dataset. It was further validated in patients with diabetes, COVID-19, or chronic kidney disease (CKD). In FLEMENGHO, the mortality endpoints were all-cause, cardiovascular, and non-cardiovascular mortality; other endpoints were fatal or non-fatal cancer and musculoskeletal disorders. Molecular pathway exploration was done using the Reactome and Kyoto Encyclopedia of Genes and Genomes databases.

**Findings** 778 individuals (395 [51%] women and 383 [49%] men; aged 16.2–82.1 years; mean age 50.9 years [SD 15.8]) from the FLEMENGHO cohort had a follow-up examination between 2005 and 2010, of whom 559 participants had a further follow-up from Oct 28, 2009, to March 19, 2013, and made up the derivation (2005–10) and time-shifted internal validation (2009–13) datasets. 219 were examined once and constituted the synchronous internal validation dataset (2005–10). With correction for multiple testing and multivariable adjustment, chronological age was associated with 210 sequenced peptides mainly showing downregulation of collagen fragments. The trained model relating chronological age to UPP, derived by elastic net regression, included 54 peptides from 17 proteins. The UPP-age prediction model explained 76.3% ( $r=0.87$ ) of chronological age in the derivation dataset, 54.4% ( $r=0.74$ ) in the time-shifted validation dataset, and 65.3% ( $r=0.81$ ) in the synchronous internal validation dataset. Compared with chronological age, the predicted UPP-age was greater in patients with diabetes (chronological age 50.8 years [SE 0.37] vs UPP-age 56.9 years [0.30]), COVID-19 (53.2 years [1.80] vs 58.5 years [1.67]), or CKD (54.6 years [0.97] vs 62.3 years [0.85]; all  $p<0.0001$ ). In the FLEMENGHO cohort, independent of chronological age, UPP-age was significantly associated with various risk markers related to cardiovascular, metabolic, and renal disease, inflammation, and medication use. Over a median of 12.4 years (IQR 10.8–13.2), total mortality, cardiovascular mortality, and osteoporosis in the population was associated with UPP-age independent of chronological age, with hazard ratios per 10 year increase in UPP-age of 1.54 (95% CI 1.22–1.95) for total mortality, 1.72 (1.20–2.47) for cardiovascular mortality, and 1.40 (1.06–1.85) for osteoporosis and fractures. The most relevant molecular pathways informed by the proteins involved deregulation of collagen biology and extracellular matrix maintenance.

**Interpretation** The UPP signature indicative of ageing reflects fibrosis and extracellular matrix remodelling and was associated with risk factors and adverse health outcomes in the population and with accelerated ageing in patients. Innovation in addressing disability should shift focus from the ontology of diseases to shared disease mechanisms, in particular ageing-related fibrotic degeneration.

**Funding** European Research Council, Ministry of the Flemish Community, OMRON Healthcare.

Lancet Healthy Longev 2021; 2: e690–703

Published Online  
October 14, 2021  
[https://doi.org/10.1016/S2666-7568\(21\)00226-9](https://doi.org/10.1016/S2666-7568(21)00226-9)

\*Investigators are listed in the appendix (p 5)

Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium (D S Martens PhD, C Wang MSc, Prof T S Nawrot PhD); Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Diseases (L Thijs MSc, Prof Z-Y Zhang MD), Research Unit Environment and Health, Department of Public Health and Primary Care (Prof T S Nawrot), and Biomedical Sciences Group, Faculty of Medicine (Prof J A Staessen MD), University of Leuven, Leuven, Belgium; Mosaïques-Diagnostics, Hannover, Germany (A Latosinska PhD, J Siwy PhD, Prof H Mischak PhD); Division of Cardiology, Sint-Jan Hospital, Bruges, Belgium (S Trenson MD); Martin Luther University of Halle-Wittenberg, Halle (Saale), Germany (Prof J Beige MD); Systems Biology Center, Biomedical Research Foundation, Academy of Athens, Athens, Greece (A Vlahou PhD); Division of Cardiology, University Hospitals Leuven, Leuven, Belgium (Prof S Janssens MD); Institute of Cardiovascular and Medical Sciences, Glasgow, UK (Prof H Mischak); Alliance for the Promotion of Preventive Medicine, Mechelen, Belgium (Prof J A Staessen)

Correspondence to: Prof Jan A Staessen, Alliance for the Promotion of Preventive Medicine, BE-2800 Mechelen, Belgium [jan.staessen@appremed.org](mailto:jan.staessen@appremed.org) @jasta49

See Online for appendix

Copyright © 2021 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 License.

## Introduction

As highlighted by the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2019,<sup>1</sup> longer life expectancy is characterised by a shift from premature mortality to functional disability caused by stagnant age-specific rates of living with a disability. Health-care expenditure mirrors this transition.<sup>1</sup> In the USA, health-care costs for musculoskeletal disorders are greater than the expenditure related to cardiovascular diseases or cancer.<sup>2</sup> Epigenetic, transcriptomic, metabolomic, and proteomic approaches, routine laboratory tests, or a combination of these methods have been used to develop ageing clocks, which have generated novel insights into the biological mechanisms of accelerated ageing that lead to disabilities. However, according to the GBD consortium,<sup>1</sup> innovation in addressing the main disabling ageing-related conditions is poor.

Urine contains more than 20 000 endogenous peptides, which are partly generated along the nephron or which pass into the tubular fluid through the glomerular barrier.<sup>3</sup> Sequencing these urinary peptides identifies the parental proteins.<sup>3</sup> Thus, the urinary peptidomic profile (UPP) provides a body-wide molecular signature of ongoing pathophysiological processes. UPP-generated molecular signatures have been shown to be associated with the preclinical phase of heart failure,<sup>4</sup> chronic kidney disease (CKD),<sup>5</sup> diabetic nephropathy,<sup>5</sup> and a variety of other disorders. However, few studies have addressed the relationship between accelerated ageing and the UPP, and none have reported on how the UPP predicts ageing-associated disabilities or adverse health outcomes. To fill this knowledge gap, we aimed to identify and validate a UPP signature that differentiates healthy from unhealthy ageing in a population cohort

### Research in context

#### Evidence before this study

We searched PubMed for studies published in English up to March 9, 2021, using the terms “proteome” AND “ageing” AND “clock”; “proteome” AND “ageing” AND “signature”; “urinary” AND “proteome” AND “ageing”; and “urine” AND “proteome” AND “ageing”. We identified 440 records of which 48 publications were duplicates and 44 reviews without original data. The remaining 348 records included 29 studies in humans on the relation between chronological age and ageing clocks, of which the full-text article and supplementary materials were read in detail. The reference lists of these 29 studies identified an additional ten studies of potential interest, which were also read. In summary, the literature review (39 articles) included 13 studies based on DNA methylation, two on routine laboratory tests, one on the gut microbiome, 11 on serum or plasma proteomics, four on urinary proteomics, six on a combination of various omics approaches, and two on epigenetics combined with telomere length. All methylation studies relied on the Illumina platform (San Diego, CA, USA) and all but two of the proteomics studies on Somalogic technology (Boulder, CO, USA). 17 studies were population-based or included a population sample. Although published as independent articles, 21 articles reused data from two to nine other studies. 11 studies included patients with various diseases, frailty, or cognitive dysfunction as a proxy of ageing, 12 focused on the lower or upper tail of the human lifespan, and 23 did not include a formal replication. No study, including the 20 studies based on epigenetic, transcriptomic, metabolomic, proteomic, other biomarkers, or a combination thereof, gave a detailed account of the association of chronological age with the urinary peptidomic profile (UPP-age) or provided evidence associating risk biomarkers or age-related disease with UPP-age.

#### Added value of this study

This study produced and validated a robust and reproducible urinary peptidomic biomarker, which reflects accelerated ageing in the general population, in patients with chronic age-related diseases, such as diabetes or chronic kidney disease, and in patients with an acute viral infection shifting life course from healthy to unhealthy ageing. Sequencing of the urinary peptides identified the parental proteins. Molecular pathway analysis indicated that the targetable molecular pathways principally involved the biology of collagen and the extracellular matrix.

#### Implications of all the available evidence

Ageing clocks focus on moving emphasis from treating to preventing chronic age-related diseases, thereby improving the sustainability of health care. In contrast to epigenetic, proteomic, or metabolomic ageing clocks, urinary peptidomic profiling, registered as an in-vitro diagnostic, is minimally intrusive and requires only a 10 mL urine sample, which can be processed without complex manipulation and from which all previously described disease-specific UPP classifiers can be generated. Furthermore, drugs specifically designed to inhibit fibrosis at the extracellular or intracellular level would fill a crucial unmet medical need. Our findings suggest that with ageing, collagen deposition exceeds its breakdown. Focus should therefore shift from drugs designed to inhibit collagen synthesis, which also interfere with the normal response to tissue injury, to agents promoting the resolution of fibrosis. A root cause of the slow progress in designing antifibrotic drugs might be that the current disease ontology focuses on specific organs rather than on generalised pathophysiological processes, with excessive fibrosis being a major body-wide actor that transcends disease classifications.

with long-term follow-up; to replicate the trained UPP-ageing clock in patients and to validate the clock in the general population as a correlate or predictor of adverse health outcomes; and to detect the molecular pathways implicated in unhealthy ageing.

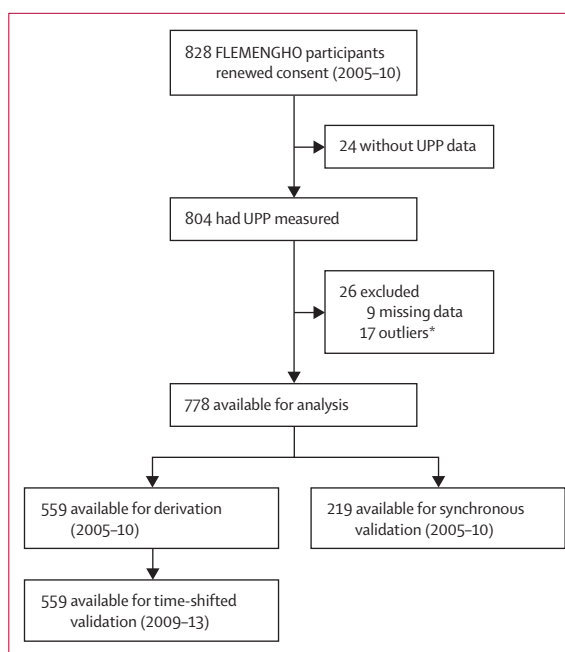
## Methods

### Study design and population

This prospective population study used data from the Flemish Study on Environment, Genes, and Health Outcomes (FLEMENGHO), which complies with the Helsinki declaration and is registered at the Belgian Data Protection Authority (III 11/1234/13; Aug 22, 2013). The ethics committee of the University Hospitals Leuven, Belgium, approved the secondary use of FLEMENGHO data (B32220083510).<sup>6</sup> From Aug 20, 1985, to Dec 14, 1990, a random sample of the households living in a geographically defined area of northern Belgium was investigated with the goal to recruit an equal number of participants in each of six subgroups stratified by sex and age (20–39 years, 40–59 years, and  $\geq 60$  years). All household members aged 20 years or older were invited to take part, provided that the quota of their sex–age group had not yet been satisfied. From April 3, 1996, to May 12, 2007, recruitment of families continued, including young people aged 10–19 years. Participants younger than 18 years provided informed assent and their parents or custodians gave informed consent (appendix p 33). Of 4286 people invited to participate in FLEMENGHO, 3343 consented (participation rate 78.0%).

From May 30, 2005, until May 31, 2010, participants were invited to a follow-up examination if their last known address was within 15 km of the local examination centre (Eksel, Belgium) and if they had not withdrawn consent in any of the previous examination cycles (1985–2004). Participants were excluded from analyses if they were incapacitated, had not undergone UPP profiling, or had either missing or outlying (three SDs greater than the mean of all consenting participants) values of body-mass index, plasma glucose, or serum creatinine. All participants who had a follow-up visit from 2005 to 2010 were invited to an additional follow-up examination from Oct 28, 2009, until March 19, 2013. Three datasets were therefore derived from the FLEMENGHO cohort. Participants who took part in both the 2005–10 follow-up examinations and in the additional 2009–13 follow-up visit constituted the derivation dataset, which included their 2005–10 data, and the time-shifted internal validation dataset, which included their 2009–13 data. The remaining participants who only had 2005–10 data constituted the synchronous internal validation dataset (figure 1). In each examination cycle, participants were examined only once. The 2005–10 data constituted the baseline for the prospective follow-up of adverse health outcomes until June 30, 2019 (appendix p 33).

For external validation, we included patients with type 1 or type 2 diabetes but without diabetic nephropathy (date



**Figure 1: Flowchart describing analysis of FLEMENGHO participants** FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. UPP=urinary peptidomic profile. \*Outlying values (three SDs greater than the mean of all consenting participants) for body-mass index, plasma glucose, or serum creatinine.

of UPP profiling May 1, 1984, to Oct 30, 2015),<sup>7</sup> patients who had recovered from PCR-confirmed COVID-19 infection WHO infection stages 1–3 (date of UPP profiling June 23, 2020, to Jan 11, 2021),<sup>8</sup> and patients with CKD but without diabetes (date of UPP profiling May 29, 2001, to Nov 17, 2017).<sup>9</sup> The patient data were retrieved from the Urine Proteome Database maintained by Mosaiques Diagnostics (Hannover, Germany). These patients were selected since diabetes and CKD are archetypes of chronic age-related diseases and a COVID-19 infection is an acute event often with long-term consequences that might alter an individual's life course from healthy to unhealthy ageing. Patients in the external datasets with all required clinical and biochemical variables available were matched to the age distribution of the FLEMENGHO derivation dataset.<sup>10</sup> Information on medication or prospective disease outcomes was not available in the Urine Proteome Database.

To assess the impact of study attrition in FLEMENGHO, we compared the 2005–10 data with those collected in previous examination cycles spanning 1996–2005 (appendix p 32).

### Measurements

For the FLEMENGHO datasets (appendix p 33), covariables, risk factors, information of medication use, and urine samples for UPP analysis were all collected on the same day: from 2005 until 2010 (for the derivation dataset and synchronous internal validation datasets) and

from 2009 until 2013 (for the time-shifted internal validation dataset). Blood pressure was the average of five consecutive auscultatory readings obtained after participants had rested for 5 min in a seated position. Study nurses administered a standardised questionnaire about each participant's medical history, smoking and drinking habits, and intake of medications. Anti-hypertensive drugs included diuretics (thiazides, loop diuretics, and aldosterone antagonists),  $\beta$  blockers, inhibitors of the renin-angiotensin system (angiotensin-converting enzyme inhibitors and angiotensin receptor blockers, including or excluding  $\beta$  blockers), vasodilators (calcium-channel blockers and  $\alpha$  blockers), and other blood pressure-lowering agents. The other major drug classes analysed were oral and parenteral antidiabetic medications, non-steroidal anti-inflammatory drugs, and lipid-lowering agents (fibrates, ezetimibe, cholestyramine, and statins). Waist circumference was measured to the nearest cm. Body fat percentage was measured using bioelectrical impedance (Bodystat QuadScan 4000, Bodystat, Isle of Man). Venous blood samples were obtained after 8 h of fasting, and participants collected a 24 h urine sample. After centrifugation and aliquoting, blood-derived specimens were stored at  $-80^{\circ}\text{C}$  and 10 mL urine samples at  $-40^{\circ}\text{C}$ . A single certified laboratory assessed the routine biochemistry, using quality-controlled automated methods. The estimated glomerular filtration rate (eGFR) was computed from serum creatinine,<sup>11</sup> measured by a modification of Jaffe's methods with isotope-dilution calibration.<sup>12</sup> According to the Chronic Kidney Disease Epidemiology Collaboration equation,<sup>12</sup>  $\text{eGFR} = 141 \times \text{minimum} (\text{Scr}/\kappa, 1)^{\alpha} \times \text{maximum} (\text{Scr}/\kappa, 1)^{-1.209} \times 0.993^{\text{age}} \times 1.018$  (if female), where Scr is the serum creatinine concentration in  $\mu\text{mol/L}$ ,  $\kappa$  is 61.9 for women and 79.6 for men, and  $\alpha$  is  $-0.329$  for women and  $-0.411$  for men; minimum indicates the minimum of  $\text{Scr}/\kappa$  or 1 and maximum indicates the maximum of  $\text{Scr}/\kappa$  or 1. Homeostatic Model Assessment for Insulin Resistance<sup>13</sup> was calculated from fasting glucose and insulin. Based on the literature, circulating biomarkers reflecting vascular function, lipid metabolism, inflammation, and renal injury were selected from panels, measured within 1 h of defrosting on a single serum aliquot, using Bio-Chip Array Technology by Randox Laboratories (County Antrim, UK). Desphospho-uncarboxylated matrix Gla protein (dpucMGP), a biomarker reflecting vitamin K status,<sup>14</sup> was measured on citrated plasma by ELISA.

#### Adverse health outcomes

At annual intervals, the vital status of FLEMENGHO participants was ascertained by record linkage with the National Population Registry in Brussels, Belgium. The ICD codes for cause of death were obtained from the Flemish Registry of Death Certificates. For all participants, data on the incidence of non-fatal endpoints were collected by a standardised questionnaire at follow-up

visits, via structured telephone interviews, and by searching participants' medical records at the four regional hospitals and University Hospitals Leuven, all serving the catchment area. The endpoints of interest were all-cause, cardiovascular, and non-cardiovascular mortality. Fatal and non-fatal cancer included cancer deaths and new-onset cancer. Musculoskeletal disorders included surgical interventions (arthroplasty and joint replacement), surgical decompression of spinal nerve roots, and osteoporosis (ICD8 code 723.0, non-accidental fractures, invasive osteosynthesis, spine stabilising surgery, or initiation of bisphosphonate treatment). For analysis of the incidence of adverse health outcomes in FLEMENGHO participants (appendix p 33), the date of urine collections in 2005–10 served as the baseline. For all endpoints, participants were censored at occurrence of the first event or on June 30, 2019, if no event had occurred.

#### Urinary peptidomics

The same laboratory measured the urinary peptidome in FLEMENGHO participants and in participants in the external validation datasets, using a standardised protocol. The methods for urine sample preparation; capillary electrophoresis coupled with mass spectrometry (CE-MS); peptide sequencing; and for the analysis, calibration, and quality control of the mass spectrometric data have been published<sup>3</sup> and are described in the appendix (pp 2–4). Sequenced urinary peptides with a detectable signal in at least 30% of study participants are commonly analysed when defining classifiers differentiating cases from controls using support vector modelling.<sup>15</sup> However, in our study, a continuously distributed variable (age) was analysed in relation to a linear combination of sequenced urinary peptides. To avoid excessive variability due to a large number of missing values, the more stringent 70% threshold was applied. Thus, after rank normalisation with the value set to the distribution minimum if undetectable,<sup>16</sup> sequenced urinary peptides with a detectable signal in at least 70% of FLEMENGHO participants were analysed.

#### Statistical analysis

Database management and statistical analysis were done using SAS, version 9.4 (maintenance level 5). The distributions of circulating biomarkers were rank normalised, by sorting measurements from the smallest to the highest value and then applying the inverse cumulative normal function. Between-group means were compared using the large-sample Z test, proportions were compared by Fisher's exact test, and longitudinal changes in proportions were compared by McNemar's test. Statistical methods also included single-adjusted and multivariable-adjusted linear and proportional hazard regression. The intraclass correlation coefficients modelling aggregation of traits between related individuals were estimated, using analysis of variance with unrelated participants excluded from analyses.

Significance was based on p values or q values of 0·05 or less, with adjustment for multiple testing if indicated.

The analysis was done according to predefined steps (appendix p 34). First, the association between chronological age and urinary peptides was investigated in the 2005–10 derivation dataset (n=559), with cumulative adjustments applied for sex (categorical), mean arterial pressure (diastolic blood pressure plus a third of the difference between systolic and diastolic blood pressure introduced as a continuously distributed variable), body-mass index (continuous), plasma glucose (continuous),  $\gamma$ -glutamyltransferase as an index of alcohol intake (continuous), current smoking status (categorical), the total-to-HDL serum cholesterol ratio (continuous), and eGFR (continuous). The second analysis step involved peptides remaining associated with chronological age after correction for the false discovery rate. These peptides were analysed by elastic net regression to construct a model predicting age (UPP-age). The L1 and L2 regression penalties were determined by 10-fold random cross-validation. This procedure was bootstrapped 1000 times to obtain the final estimates of L1 and L2 and the 95% CI of the regression coefficients, linking chronological age with each protein fragment in the reduced set of urinary peptides.

Validation in the next analysis step was as follows: (1) applying the UPP-age prediction model to the time-shifted internal validation dataset; (2) applying the UPP-age prediction model to the synchronous internal validation dataset; and (3) external validation in patients with diabetes, COVID-19 infection, or CKD without diabetes. To study the associations of biomarkers and adverse health outcomes with molecular ageing, as captured by UPP-age, over and beyond chronological age, the residual of UPP-age regressed on chronological age (UPP-age-R) was computed (appendix p 35).<sup>17</sup> UPP-age-R reflects rapid (accelerated) ageing. Molecular pathway analysis, using the databases of Reactome and the Kyoto Encyclopedia of Genes and Genomes (KEGG) was the final step in the analysis. Enriched pathways included at least three proteins identified from the sequenced urinary peptides, had a significant q value after false discovery rate correction, and were visualised using R software to create dot plots.<sup>18</sup> Enrichment analysis, also known as over-representation analysis, is a statistical method to establish whether gene products (proteins) belonging to a specific Reactome or KEGG pathway are present more than would be expected by chance in a dataset.

### Role of the funding source

The funders of this study had no role in study design, data collection, data analysis, data interpretation, or the writing of the report.

### Results

From May 30, 2005, to May 31, 2010, 1208 FLEMENGHO participants were invited for a follow-up examination. Of

	Derivation and time-shifted validation dataset (n=559)			Synchronous validation dataset (n=219)	
	2005–10 (baseline)	2009–13 (follow-up)	p value*	2005–10	p value†
Women	282 (50%)	282 (50%)	..	113 (52%)	0·81
Menopausal women	145 (26%)	178 (32%)	<0·0001	65 (30%)	0·048
Surgical menopause	13 (2%)	17 (3%)	<0·0001	8 (4%)	0·11
Men	277 (50%)	277 (50%)	..	106 (48%)	0·81
Smokers	110 (20%)	88 (16%)	0·099	49 (22%)	0·43
Hypertension‡	230 (41%)	287 (51%)	0·0018	95 (43%)	0·57
Treated hypertension	138 (25%)	177 (32%)	0·012	60 (27%)	0·46
Diabetes§	17 (3%)	31 (6%)	0·054	11 (5%)	0·20
History of cardiovascular disease	22 (4%)	27 (5%)	0·56	17 (8%)	0·043
History of cancer	1 (<1%)	3 (1%)	0·62	2 (1%)	0·19
Age, years	50·3 (14·8)	55·0 (14·8)	<0·0001	52·7 (17·9)	0·11
Body-mass index, kg/m <sup>2</sup>	26·2 (3·8)	27·0 (4·0)	0·00075	26·6 (4·5)	0·29
Waist circumference, cm	89·5 (11·8)	95·2 (12·2)	<0·0001	90·2 (13·4)	0·51
Systolic blood pressure, mm Hg	128·3 (17·0)	132·1 (17·0)	0·00022	131·5 (19·5)	0·035
Diastolic blood pressure, mm Hg	79·7 (9·4)	82·3 (9·6)	<0·0001	79·5 (10·3)	0·92
Mean blood pressure, mm Hg	95·9 (10·4)	98·8 (10·2)	<0·0001	96·9 (11·5)	0·25
Serum creatinine, $\mu$ mol/L	85·6 (13·1)	88·8 (18·1)	0·00065	87·2 (15·2)	0·16
eGFR, mL/min per 1·73 m <sup>2</sup>	81·8 (14·7)	76·6 (15·2)	<0·0001	79·6 (18·1)	0·12
Plasma glucose, mmol/L	4·87 (0·47)	4·92 (0·70)	0·19	4·98 (0·86)	0·085
Total serum cholesterol, mmol/L	5·27 (0·96)	5·03 (0·94)	0·00010	5·20 (1·00)	0·34
HDL serum cholesterol, mmol/L	1·44 (0·35)	1·47 (0·38)	0·25	1·38 (0·35)	0·021
Total-to-HDL serum cholesterol ratio	3·82 (1·00)	3·69 (2·54)	0·28	3·95 (1·07)	0·12
FSH in women, U/L	20·0 (4·10–55·0)	40·7 (6·2–64·9)	<0·0001	29·4 (6·0–57·2)	<0·0001
Daily tobacco use in smokers, g	12 (7–17)	10 (6–15)	0·040	10 (10–20)	0·44
10 year Framingham risk score, %	7·28 (2·50–15·2)	..	..	8·71 (2·90–20·9)	0·14
dpucMGP, pmol/L	4·22 (2·92–5·88)	..	..	4·57 (2·89–6·39)	0·23
$\gamma$ -glutamyltransferase, U/L	22·0 (15·0–32·0)	22·0 (16·0–33·0)	0·29	22·0 (16·0–33·0)	0·74

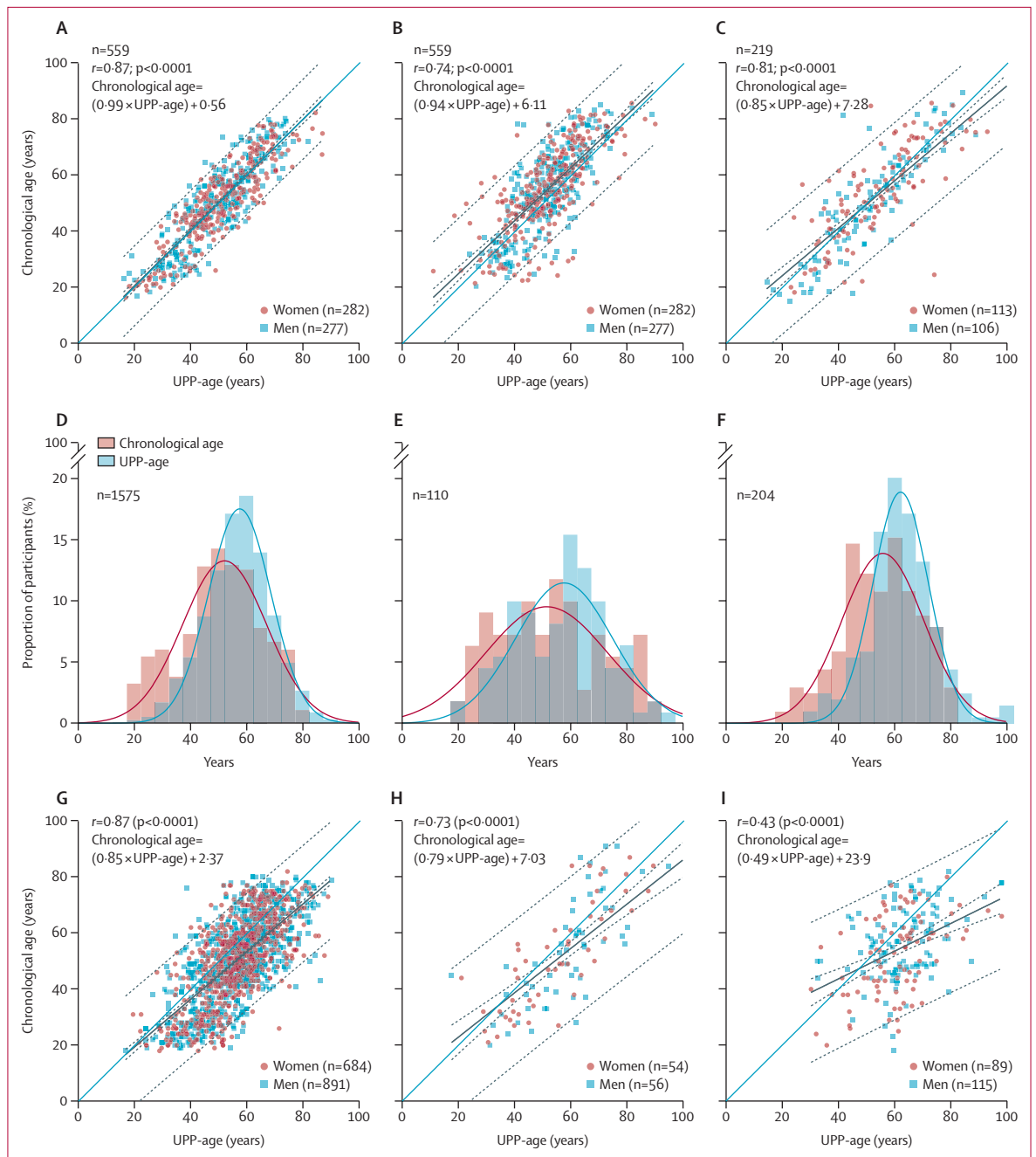
Data are n (%), mean (SD), or median (IQR). To convert eGFR from mL/min to mL/s per 1·73 m<sup>2</sup>, multiply by 0·0167. To convert dpucMGP from pmol/L to  $\mu$ g/L, multiply by 94·299. In FLEMENGHO, all participants were White and European. dpucMGP=desphospho-uncarboxylated matrix Gla protein. eGFR=glomerular filtration rate derived from serum creatinine, using the Chronic Kidney Disease Epidemiology Collaboration equation. FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. FSH=follicle-stimulating hormone. \*p values for the difference between baseline and follow-up. †p values for the difference between the derivation and synchronous validation cohorts (2005–10 data). ‡Blood pressure of  $\geq$ 140 mm Hg systolic or  $\geq$ 90 mm Hg diastolic or use of antihypertensive drugs. §Fasting plasma glucose of  $\geq$ 7·0 mmol/L, a self-reported diagnosis, diabetes documented in practice or hospital records, or use of antidiabetic drugs.

Table 1: Characteristics of FLEMENGHO participants

these participants, 26 had died, 27 were incapacitated, and 100 had moved out of the catchment area. Of the remaining 1055 participants, 828 (77·9%) renewed written informed consent (figure 1). 50 further participants were excluded because they did not meet inclusion criteria, resulting in 778 participants with follow-up examinations from 2005 to 2010. Of these, 559 (71·9%) had a further follow-up examination from Oct 28, 2009, to March 19, 2013

For the Reactome website see <https://reactome.org>

For the Kyoto Encyclopedia of Genes and Genomes see <https://www.genome.jp/kegg>



**Figure 2: Correlations between observed chronological age and age predicted by the urinary proteome in FLEMENGHO participants and patients**  
 Correlations for the derivation dataset (2005–10; A), time-shifted internal validation dataset (2009–13; B), and synchronous internal validation dataset (2005–10; C) in FLEMENGHO participants. Distributions of chronological age and UPP-age (superimposed on the UPP-age distribution in the FLEMENGHO derivation dataset [grey bars]) for patients with diabetes (D), COVID-19 (E), and chronic kidney disease (F). Correlations between chronological age and UPP-age for patients with diabetes (G), COVID-19 (H), and chronic kidney disease (I). Regression lines (solid black) are given with 95% CIs (dotted lines) for predicting mean chronological age (narrow band) and chronological age in individual participants (broad band). The blue line in the correlation plots is the identity line. FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. UPP=urinary peptidomic profile. UPP-age=age as predicted by the UPP.

(figure 1) and were included in derivation and time-shifted internal datasets; the remaining 219 participants were included in the synchronous internal validation dataset.

The derivation and time-shifted internal validation cohort (n=559) included 282 (50%) women and 277 (50%)

men (table 1). In these 559 participants, mean age at baseline (2005–10) was 50.3 years (SD 14.8; range 16.2–82.1) and mean age was 55.0 years (14.8) at follow-up (2009–13), which took place after a median interval of 4.9 years (IQR 4.4–5.2). Ageing from baseline to

follow-up was accompanied by increases in mean body-mass index (+0.8 kg/m<sup>2</sup>), waist circumference (+5.7 cm), and systolic and diastolic blood pressure (+3.8 mm Hg and +2.6 mm Hg), a reduction in mean eGFR (−5.2 mL/min per 1.73 m<sup>2</sup>), reduced median use of tobacco (−2 g), and a higher number of menopausal women (+33 women). The characteristics of the derivation (n=559) and synchronous internal validation (n=219) cohorts at baseline were broadly similar (table 1).

In the derivation database, the UPP included 21560 peptides. Of 4080 sequenced peptides, 635 identifying 67 parental proteins had a detectable signal in at least 70% of participants and were used for further analyses. With adjustments applied for sex, mean arterial pressure, body-mass index, plasma glucose,  $\gamma$ -glutamyltransferase, smoking status, the total-to-HDL serum cholesterol ratio, and eGFR, and with false discovery rate corrected significance, chronological age was associated with 210 sequenced peptides, identifying 39 parental proteins (appendix pp 7–18). Associations of age with multiple urinary peptide fragments originating from the same protein were all directionally similar for the top-ranked peptides, with the exception of one of 104 COL1A1 fragments, one of 41 COL3A1 fragments, and one of three COL5A2 peptides (appendix pp 7–18). For the peptide with the highest significance level representing one of the 39 parental proteins in the derivation dataset, associations of chronological age with the urinary peptides were confirmed for all in the time-shifted internal validation dataset and the synchronous internal validation dataset (appendix pp 19–20) on the basis of a one-sided p value of 0.05 (given the prior probability established by the significance in the derivation dataset). For 25 (64%) of the 39 proteins, confirmation was established in the time-shifted and the synchronous internal validation datasets, for two (5%) in the time-shifted validation dataset, and for six (15%) in the synchronous validation set. Thus, the association with chronological age remained unconfirmed for six (15%) of 39 proteins from the derivation dataset (appendix pp 19–20). The fragments that were not replicated originated from CD99, PIGR, KRT10, PCDH7, COL7A1, and UBAC2. Findings were largely similar when we assessed the consistency of the relationship with chronological age for all peptide fragments derived from the 39 proteins that had passed the false discovery rate-corrected threshold of significance in the derivation dataset (appendix pp 21–22). Volcano plots (appendix p 37) show the high consistency of the finding across the FLEMENGHO datasets (derivation vs time-shifted and synchronous datasets).

In FLEMENGHO, of 210 sequenced peptides (appendix pp 7–18) submitted to elastic net regression, 54 (26%) originating from 17 proteins were retained in the UPP-age prediction model (appendix pp 23–26). UPP-age explained 76.3% of chronological age in the derivation dataset ( $r=0.87$ ), with similar estimates in women

(78.5% [ $r=0.89$ ]) and men (74.5 [ $r=0.86$ ]), 54.4% ( $r=0.74$ ) of chronological age in the time-shifted internal validation dataset, and 65.3% ( $r=0.81$ ) in the synchronous internal validation dataset (figure 2). Chronological age and UPP-age were similar in the derivation dataset (chronological age 50.3 years [SE 0.63] vs UPP-age 50.3 years [0.55],  $p=0.99$ ) and in the synchronous internal validation dataset (52.5 years [1.21] vs 53.4 years [1.15],  $p=0.18$ ). However, in the time-shifted internal validation dataset, chronological age was higher than UPP-age (55.0 years [0.62] vs 51.8 years [0.49],  $p<0.0001$ ). Furthermore, compared with the derivation dataset (figure 2A), the regression line slope of chronological age on UPP-age was similar in the time-shifted internal validation dataset (figure 2B;  $p=0.29$ ), but slightly smaller in the synchronous internal validation data (figure 2C;  $p=0.0074$ ). The estimates of

	Chronological age		UPP-age		UPP-age-R	
	r	p value	r	p value	r	p value
<b>Vascular biomarkers</b>						
Systolic blood pressure	0.52	<0.0001	0.49	<0.0001	0.10	0.0042
Diastolic blood pressure	0.17	<0.0001	0.19	<0.0001	0.07	0.037
VEGF	0.14	<0.0001	0.18	<0.0001	0.10	0.0048
PAI1	0.10	0.0041	0.14	<0.0001	0.10	0.0042
<b>Renal biomarkers</b>						
eGFR	−0.71	<0.0001	−0.65	<0.0001	−0.09	0.011
NGAL	0.19	<0.0001	0.22	<0.0001	0.11	0.0015
<b>Insulin resistance</b>						
Plasma glucose	0.30	<0.0001	0.28	<0.0001	0.05	0.17
Serum insulin	−0.01	0.77	0.03	0.33	0.08	0.022
HOMA-IR	0.05	0.16	0.08	0.018	0.08	0.028
<b>Lipid biomarkers</b>						
Body-mass index	0.27	<0.0001	0.29	<0.0001	0.11	0.0013
Waist circumference	0.33	<0.0001	0.34	<0.0001	0.11	0.0029
Body fat	0.36	<0.0001	0.34	<0.0001	0.08	0.22
Leptin	0.09	0.0081	0.14	<0.0001	0.12	0.00067
Resistin	0.08	0.031	0.11	0.0028	0.08	0.029
<b>Inflammation biomarkers</b>						
C-reactive protein	0.12	0.0011	0.14	<0.0001	0.08	0.029
TNF $\alpha$	0.28	<0.0001	0.31	<0.0001	0.12	0.00049
TNFR1	0.36	<0.0001	0.36	<0.0001	0.11	0.0016
<b>Other markers</b>						
Daily tobacco use in smokers	−0.07	0.052	−0.054	0.13	0.01	0.78
Follicle-stimulating hormone in women (n=395)	0.64	<0.0001	0.55	<0.0001	0.06	0.26
Framingham risk score	0.88	<0.0001	0.76	<0.0001	0.03	0.39
dpucMGP	0.42	<0.0001	0.42	<0.0001	0.11	0.0027

The distributions of VEGF, PAI1, NGAL, HOMA-IR, body fat, leptin, resistin, C-reactive protein, TNF $\alpha$ , TNFR1, daily use of tobacco, follicle-stimulating hormone, Framingham risk score, and dpucMGP were rank normalised. r is the Pearson correlation coefficient. dpucMGP=desphospho-uncarboxylated matrix Gla protein. eGFR=glomerular filtration rate derived from serum creatinine. FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. HOMA-IR=Homeostatic Model Assessment for Insulin Resistance. UPP=urinary peptidomic profile. UPP-age=age as predicted by the UPP. UPP-age-R=residual of the regression of UPP-age on chronological age and reflects accelerated ageing as predicted by the UPP-age, independent of chronological age.

**Table 2: Risk biomarkers in relation to age in FLEMENGHO participants examined from 2005 to 2010**

	n/N (%)	Chronological age		UPP-age		UPP-age-R	
		HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
<b>Mortality</b>							
Total	71/778 (9%)	3.62 (2.87–4.57)	<0.0001	2.23 (1.97–2.52)	<0.0001	1.54 (1.22–1.95)	0.00032
Cardiovascular	24/778 (3%)	7.27 (4.32–12.23)	<0.0001	2.58 (2.10–3.16)	<0.0001	1.72 (1.20–2.47)	0.0033
Non-cardiovascular	43/778 (6%)	2.71 (2.08–3.53)	<0.0001	2.00 (1.69–2.37)	<0.0001	1.35 (0.96–1.91)	0.083
Fatal and non-fatal cancer	71/778 (9%)	2.00 (1.67–2.39)	<0.0001	1.70 (1.47–1.96)	<0.0001	1.10 (0.82–1.49)	0.53
<b>Musculoskeletal disorders</b>							
Osteoarticular complications	162/778 (21%)	1.03 (0.93–1.14)	0.53	1.07 (0.96–1.19)	0.24	1.13 (0.92–1.37)	0.24
Osteoporosis and fractures							
All	68/778 (9%)	1.24 (1.06–1.46)	0.0083	1.34 (1.14–1.57)	0.00046	1.40 (1.06–1.85)	0.018
Women	48/395 (12%)	1.55 (1.26–1.91)	<0.0001	1.53 (1.28–1.83)	<0.0001	1.43 (1.28–1.97)	0.029
Men	20/383 (5%)	0.82 (0.62–1.08)	0.19	0.86 (0.62–1.19)	0.36	1.20 (0.68–2.11)	0.54
HRs, given with 95% CI, express the relative risk per 10 year increment in chronological age, UPP-age, and UPP-age-R. Cause of death was not documented in four participants. FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. HR=hazard ratio. n=incident endpoints N=participants at risk. UPP=urinary peptidomic profile. UPP-age=age as predicted by the UPP. UPP-age-R=residual of the regression of UPP-age on chronological age and reflects accelerated ageing as predicted by the UPP-age, independent of chronological age.							
<b>Table 3: Risk of adverse health outcomes in relation to age in FLEMENGHO participants</b>							

the increases in chronological age for a 10 year increment in UPP-age were 9.9 years (95% CI 9.4–10.3) in the derivation dataset, 9.4 years (8.7–10.2) in the time-shifted internal validation dataset, and 8.5 years (7.6–9.3) in the synchronous internal validation dataset.

The external validation cohorts, age-matched to the derivation dataset, included 1575 patients with diabetes without nephropathy (mean eGFR 97.7 mL/min per 1.73 m<sup>2</sup>; (figure 2G), 110 patients with PCR-confirmed COVID-19 (eGFR 75.4 mL/min per 1.73 m<sup>2</sup>; figure 2H), and 204 patients with CKD without diabetes (eGFR 43.3 mL/min per 1.73 m<sup>2</sup>; figure 2I). Compared with chronological age (figure 2D–F), predicted UPP-age was higher in patients with diabetes (chronological age 50.8 years [SE 0.37] vs UPP-age 56.9 years [0.30]), COVID-19 (53.2 years [1.80] vs 58.5 years [1.67]), or CKD (54.6 years [0.97] vs 62.3 years [0.85]; all p<0.0001). Using the trained model, the slopes of chronological age regressed on UPP-age were smaller in all patient groups than in the derivation sample (p<0.0001; figure 2G–I). The increases in chronological age predicted by 10 year UPP-age increments were 8.5 years (95% CI 8.1–8.9) in patients with diabetes, 7.9 years (6.5–9.3) in those with COVID-19, and 4.9 years (3.5–6.4) in those with CKD (figure 2G–I). Thus, in patients, the slope of chronological age on UPP-age was smaller than the age predicted by the trained model in the FLEMENGHO derivation dataset, thereby resulting in greater UPP-age than chronological age, reflecting accelerated ageing in the patients compared with the general population.

The coefficient of variation of UPP-age was established in three healthy individuals not included in the population or patient cohorts, who each had UPP profiling on mid-morning urine samples collected on 6 different days within a 1 month time span. The within-individual coefficients of variation ranged from 2.7% to 11.3%, with

a mean of 7.3% (SE 2.5), indicating satisfactory short-term reproducibility (appendix p 36).

Overall, the FLEMENGHO study population consisted of 114 unrelated participants and 664 related participants, belonging to 17 single-generation families and 61 multigeneration pedigrees. The intraclass correlation coefficients among related FLEMENGHO participants were 0.15 (95% CI 0.07 to 0.23; p=0.00029) for chronological age, 0.23 (0.15 to 0.30; p<0.0001) for UPP-age, and 0.19 (0.11 to 0.26; p<0.0001) for UPP-age-R. Biomarkers reflecting vascular risk factors including blood pressure, renal dysfunction, insulin resistance, obesity, disproportional distribution of fat tissue, deregulation of the lipid metabolism, inflammation, and poor vitamin K status, as reflected by dpucMGP, were significantly associated with UPP-age and UPP-age-R (table 2). Accounting for the intrafamilial clustering of risk factors weakened the associations of risk factors with chronological age, UPP-age, and UPP-age-R (appendix p 27), but the correlations of UPP-age-R remained significant for NGAL (r=0.09; p=0.036), body-mass index (r=0.10; p=0.021), waist circumference (r=0.09; 0.022), leptin (r=0.09; 0.027), and dpucMGP (r=0.10; p=0.019). 298 (38%) of 778 FLEMENGHO participants reported taking antihypertensive, antidiabetic, non-steroidal anti-inflammatory, and lipid-lowering drugs. These participants had higher chronological age and UPP-age than participants not using any of these drug classes (p<0.0001; appendix p 6). However, UPP-age-R was only higher in patients using diuretics (n=78; 2.07 years; 95% CI 0.35 to 3.79; p=0.018), patients taking inhibitors of the renin-angiotensin system including  $\beta$  blockers (n=21; 3.34 years; 0.22 to 6.46; p=0.036), and patients on a single antihypertensive drug (n=41; 2.65 years; 0.32 to 4.98; p=0.026). There were no significant differences in patients taking antidiabetic drugs (n=22;



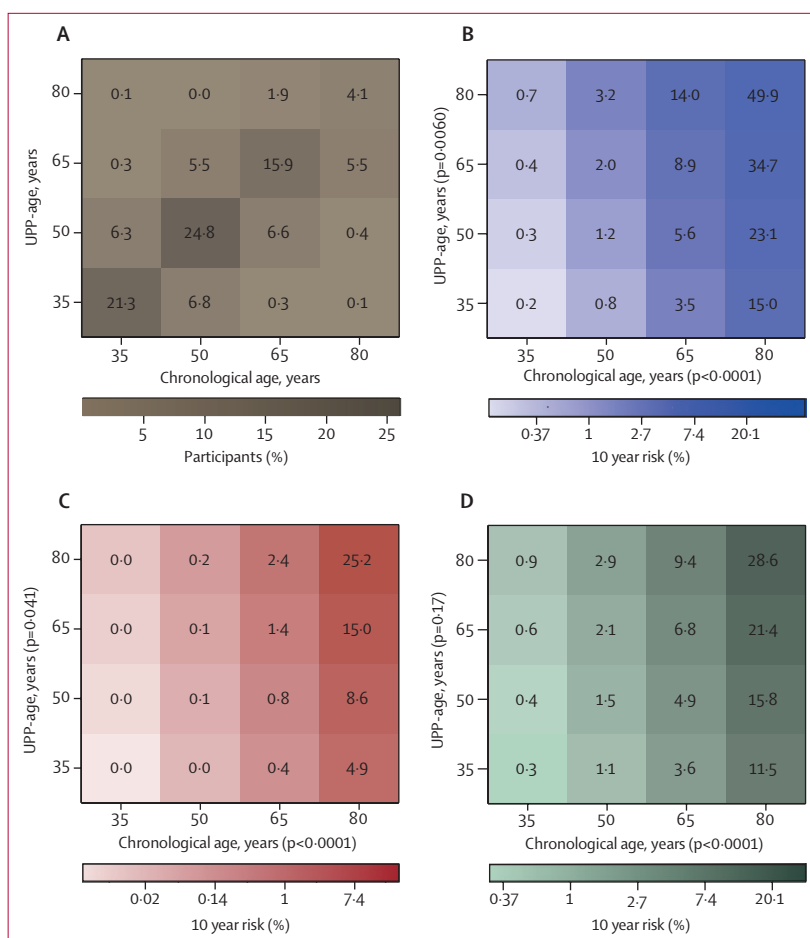
2.86 years;  $-0.17$  to  $5.88$  years;  $p=0.064$ ) or in patients taking any combination of the aforementioned drug classes ( $n=298$ ; 1.00 year;  $-0.10$  to  $2.10$ ;  $p=0.075$ ) compared with participants not taking any drug (appendix p 6).

Median follow-up of the all FLEMENGHO participants was 12.4 years (IQR 10.8–13.2). Total mortality, cardiovascular mortality, non-cardiovascular mortality, and fatal plus non-fatal cancer were associated with increasing chronological age, with hazard ratios (HRs) per 10 year increment ranging from 2.00 (95% CI 1.67–2.39) to 7.27 (4.32–12.23; all  $p<0.0001$ ), and were associated with increasing UPP-age, with HRs ranging from 1.70 (1.47–1.96) to 2.58 (2.10–3.16; all  $p<0.0001$ ; table 3, figure 3). Total and cardiovascular mortality were also significantly associated with UPP-age-R, with HRs of 1.54 (1.22–1.95;  $p=0.00032$ ) for total mortality and 1.72 (1.20–2.47;  $p=0.0033$ ) for cardiovascular mortality (table 3). The combined endpoint of osteoporosis and fractures (table 3) was significantly associated with chronological age, UPP-age, and UPP-age-R across all participants and in women but not in men (table 3). In all participants, the 10 year HRs were 1.24 (95% CI 1.06–1.46) for chronological age, 1.34 (1.14–1.57) for UPP-age, and 1.40 (1.06–1.85) for UPP-age-R; the corresponding HRs in women only were 1.55 (1.26–1.91), 1.53 (1.28–1.83), and 1.43 (1.28–1.97). Adjustment for family clusters produced similar results (appendix p 28) to the analyses unadjusted for family clusters.

Using the Reactome and KEGG databases, the UPP reflected 27 (Reactome) and ten (KEGG) over-represented pathways ( $q$  value  $<0.01$ ; figure 4; appendix pp 29–31). The top pathways identified by Reactome included the biology of collagen, the extracellular matrix (integrin and non-integrin interactions), and cellular structural regulation. The highest-ranking KEGG pathways were protein digestion and absorption (*hsa04974*), and in line with the Reactome analysis, pathways involved in extracellular matrix structure and functioning (*hsa04512* and *hsa04510*). Information on the effects of study attrition was available for all 778 FLEMENGHO participants included in the derivation and synchronous internal validation cohorts (2005–10) and for 360 of the 413 participants not included in analyses, because they had died, were institutionalised, had moved away from the study area, declined further participation, or were excluded from analysis. In 1996–2005, participants included in analyses were 7.4 years older than those not included, which explained most of the other between-group differences.

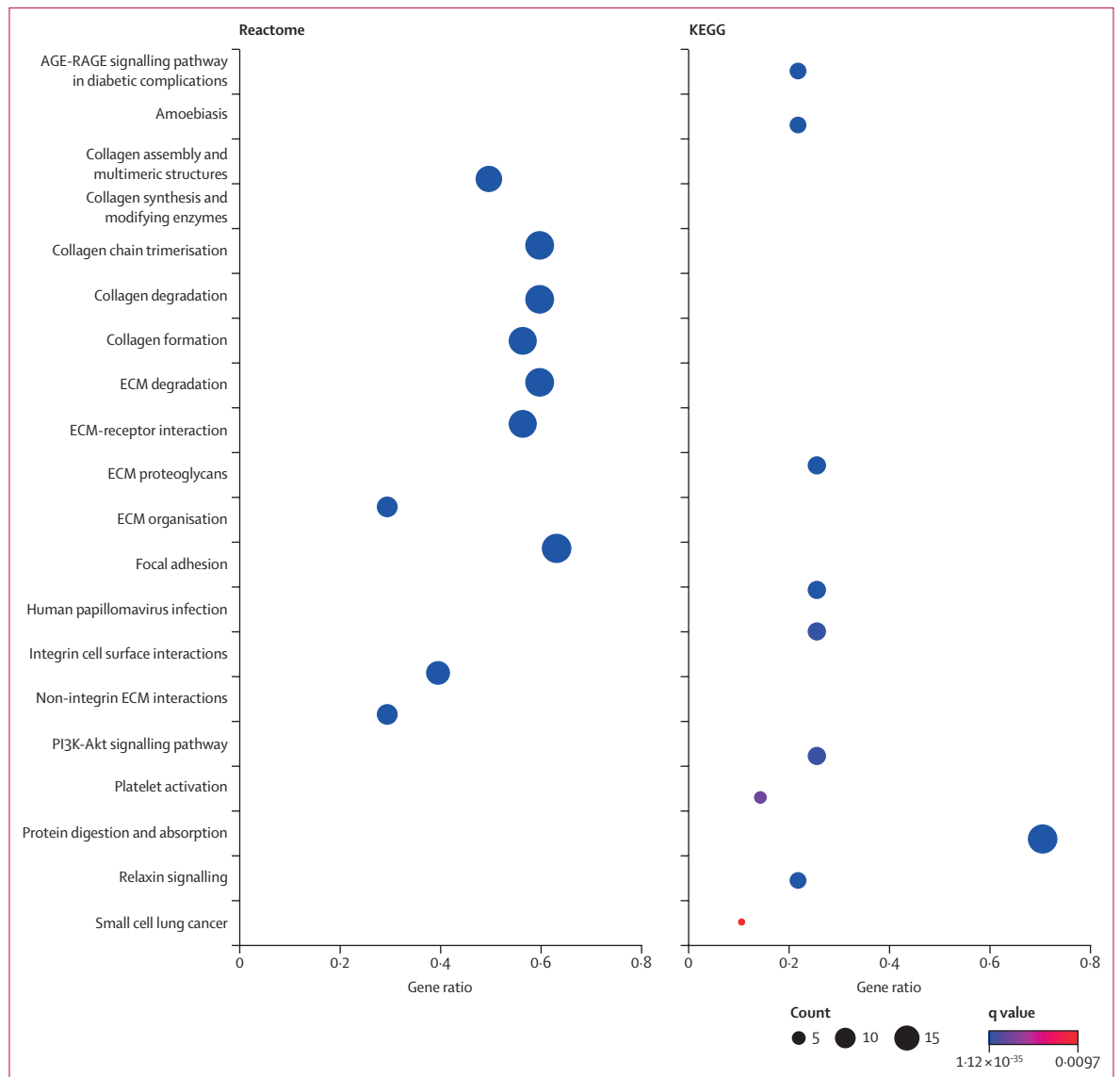
## Discussion

The main finding of this study was that chronological age was associated with a specific change in the UPP signature. This relationship between chronological age and UPP-age was replicated within the FLEMENGHO population using the derivation and internal validation



**Figure 3:** Heat maps relating mortality to age in 778 FLEMENGHO participants. Proportion of participants by UPP-age and chronological age cross-classification (A). Cox proportional hazards regression showing 10 year risk of death for total (B), cardiovascular (C), and non-cardiovascular (D) mortality in relation to chronological age and UPP-age, derived from the 2005–10 baseline examinations (appendix p 33). FLEMENGHO=Flemish Study on Environment, Genes, and Health Outcomes. UPP=urinary peptidomic profile. UPP-age=age as predicted by the UPP.

datasets. In patients with diabetes, PCR-confirmed COVID-19 infection, or CKD, there was a change in UPP towards a profile characteristic of a chronological age higher than predicted by the trained model indicating that the UPP-age was higher than chronological age, and suggesting accelerated ageing in the patient cohorts. In the FLEMENGHO population, risk factors were associated with UPP-age, independent of chronological age, but adjustment for family clusters weakened these associations, probably because of the high intrafamilial correlation of UPP-age-R, a marker of accelerated ageing, and perhaps because of the simultaneous assessment of risk factors and the UPP. However, the positive associations of NGAL, body-mass index, waist circumference, leptin, and dpucMGP withstood adjustment for familial aggregation. The differences in UPP-age-R between patients taking medication to treat chronic age-related disorders (appendix p 6) and patients not taking medication must be carefully interpreted. We previously



**Figure 4: Pathway analysis**

AGE-RAGE=advanced glycation end product-receptor for advanced glycation end products. ECM=extracellular matrix. KEGG=Kyoto Encyclopedia of Genes and Genomes. q values are the Benjamini-Hochberg adjusted p values.

showed that most drug classes, including antirejection drugs,<sup>19</sup> do not materially alter the UPP. Therefore, the results of the medication analyses are likely not to reflect the use of medications per se but rather the conditions for which these medications are prescribed. Indeed, the prescription of medication might signal greater disease severity, hence higher UPP-age. Furthermore, irrespective of adjustment for familial aggregation, the incidence of total and cardiovascular mortality and osteoporosis combined with non-accidental fractures was related to UPP-age independent of chronological age. The most relevant molecular pathways informed by the proteins involved deregulation of collagen biology and extracellular matrix maintenance.

In keeping with the pathway analysis (appendix pp 29–31), the decrease with age in the abundance of the urinary collagen fragments, mainly originating from fibrillar collagens (types I, II, III, V, and XI) and molecular bridge collagens (types IX, XIV, XIX, and XXII), was a striking finding. This change in UPP reflects the body-wide fibrosis of vital organs and extracellular matrix reorganisation, commonly associated with ageing as well as with age-related chronic diseases, whereby the rate of collagen deposition exceeds the rate of its breakdown.<sup>20</sup> Stiffening of tissues caused by fibrosis and extracellular matrix remodelling is the root cause of organ dysfunction, particularly in the heart, lung, and kidney and is a major player in rheumatic disease and osteoarthritis. The finding

that urinary fragments derived from collagens and MGP were dysregulated and were associated with cardiovascular mortality suggest potential involvement of arterial and left ventricular stiffening and microvascular dysfunction in accelerated ageing.<sup>14</sup> A urinary fragment of COL25A1, of which the tissue-specific expression and functional role are currently not well defined, was consistently upregulated in the derivation and validation datasets. The COL25A1 gene plays a role in collagenous Alzheimer amyloid plaque formation, an age-related condition.<sup>21</sup> Furthermore, insulin resistance was significantly associated with UPP-age. Via a non-enzymatic process, dysregulation of glucose metabolism and diabetes promote the generation of advanced glycation end products (AGEs), which contribute to cross-linking of collagens in the basement membranes, leading to fibrosis, particularly in the vasculature.<sup>22</sup> The PI3K pathway (appendix pp 29–31) regulates insulin and glucose metabolism, immune responses, and cell proliferation.<sup>23</sup> Both the advanced glycation end product-receptor for advanced glycation end products (AGE-RAGE) signalling<sup>22</sup> and PI3K<sup>23</sup> pathways are affected by nutritional status and medications, which vary in an age-dependent way.

Along similar lines, several risk factors, in particular those remaining significantly associated with UPP-age-R after adjustment for familial aggregation—ie, NGAL,<sup>24</sup> body-mass index, waist circumference, and leptin<sup>25</sup>—are commonly associated with chronic inflammation and sympathetic activation, which are both main drivers of fibrosis. A systematic review of the age-related proteome established from various human biological media summarised 36 studies<sup>26</sup> and identified 117 proteins associated with ageing, which were replicated in at least four of the articles reviewed. Of these proteins, our study picked up eight: COL1A1, COL3A1, COL1A2, FGA, CSTB, S100A9, PIGR, COL15A1. This previous review<sup>26</sup> included four studies based on the urinary proteome.<sup>27–30</sup> Two small studies of 37 healthy volunteers (age range 19–90 years)<sup>27</sup> and 52 healthy men (19–57 years)<sup>28</sup> identified 19 and 23 age-related proteins, respectively, which were inconsistent with our current findings. One larger study reported that in 324 healthy individuals (2–73 years) most of 49 age-related urinary peptides were markers of CKD.<sup>29</sup> The most recent report of 1227 healthy individuals and 10 333 patients with a wide variety of disorders (aged 20–86 years) identified 116 age-related urinary peptides, which predominantly originated from collagen, uromodulin, and fibrinogen.<sup>30</sup> However, none of these previous publications<sup>27–30</sup> attempted to construct a robust UPP-ageing marker independent of chronological age.

Despite the diversity of underlying causes that lead to fibrosis in any given tissue, common biochemical and cellular mechanisms occur in all organs. Injury activates resident fibroblasts or recruits bone-marrow-derived circulating fibrocytes and epithelial or endothelial cells, and drives the epithelial or endothelial-to-mesenchymal

transition. Subsequently, these cells transdifferentiate into  $\alpha$ -smooth muscle actin expressing myofibroblasts that secrete the extracellular matrix components required for wound repair in acute injury, but produce excessive extracellular matrix deposition in response to persistent injury.<sup>31</sup> Antifibrotic drugs remain a crucially important unmet medical need, as nearly 45% of all natural deaths in high-income countries are attributed to the complications of chronic fibroproliferative disorders.<sup>32</sup> A detailed review of the molecular pathways of fibrosis and of the refurbished or experimental drugs inhibiting fibrosis extracellularly or intracellularly is beyond the scope of this Article and is summarised elsewhere.<sup>33</sup> However, drugs specifically designed to inhibit fibrosis at the intracellular or extracellular level, in particular by targeting collagen degradation,<sup>34</sup> would meet a pressing need in ageing populations<sup>32</sup> but, to our knowledge, few are being developed. The underlying reason for the unmet needs for antifibrotic drugs might be that the current ontology of disease focuses on specific organs rather than on generalised pathophysiological processes, with fibrosis being a major body-wide actor that transcends disease classifications, even playing a role in the treatability of solid tumours.<sup>35</sup> Multiple efforts aimed at attenuating collagen synthesis have been made but until now have not delivered benefits. Moreover, inhibiting collagen synthesis interferes with the normal responses to injury. Because our findings suggest that the rate of collagen deposition exceeds that of its breakdown, focus might shift to interventions promoting resolution of fibrosis rather than inhibition of the deposition of collagen.

The main function of type I collagen assembly in the skeletal extracellular matrix is to maintain the structural integrity and mechanical properties of the bone.<sup>36</sup> Oestrogen deficiency in women, occurring after oophorectomy or naturally with menopause, activates proinflammatory cytokines, which stimulate osteoclast activity and shift bone metabolism to bone loss,<sup>37</sup> thereby contributing to the decrease with age in urinary collagen type I fragments. The AGE-RAGE signalling pathway is potentially implicated in reducing bone density and bone quality and in suppressing bone turnover biomarkers.<sup>38</sup> Independent of chronological age, UPP-age was positively associated with dpucMGP, a marker of vitamin K deficiency, which is required for the  $\gamma$  carboxylation and activation of MGP, osteocalcin, GAS6, periostin, and protein S.<sup>39</sup> These vitamin K-dependent proteins are major players in promoting differentiation and expression of specific genes in osteoblasts and activating proteins involved in bone extracellular matrix mineralisation<sup>40</sup> and play a role in activating TAM receptors affecting inflammation and haemostasis.<sup>41</sup> Moreover, vascular smooth muscle cells and the endothelium synthesise MGP.<sup>14</sup> Activated MGP plays a key role in maintaining macrovascular and microvascular integrity and the maintenance of left ventricular and renal function.<sup>14</sup> In FLEMENGHO, plasma dpucMGP levels ranging from

1.4 µg/L to 4.6 µg/L (132–434 pmol/L) were optimal in mitigating the risk of mortality and macrovascular cardiovascular disease; the 4.6 µg/L threshold corresponded with the 65th percentile of the dpucMGP distribution, indicating that nearly a third of the general population in a high-income country has a poor vitamin K status.<sup>14</sup>

Strengths of the current study are the replication and validation of the UPP-age prediction model within the FLEMENGHO cohort, the demonstration of premature ageing in patients with diabetes, COVID-19, or CKD, and the demonstration of the clinical relevance of the UPP data in relation to risk biomarkers and, more importantly, in relation to prospectively collected health outcomes, including both fatal and non-fatal endpoints and both cardiovascular and non-cardiovascular adverse health outcomes. The UPP model is unbiased in the sense that it does not depend on predefined markers but is only determined by peptides passing the glomerular barrier or generated along the nephron, and because it is platform independent.<sup>3</sup> A 10 mL morning urine sample, as used in patient studies,<sup>7,9,10</sup> or an aliquot from 24 h urine collection, as used in FLEMENGHO,<sup>6</sup> are sufficient to run the CE-MS analysis. This provides great flexibility in collecting urine and using biobank resources. The calibration of the CE-MS procedure (appendix pp 2–4) corrects for methodological differences in urine sampling. The current study was population-based, and did not include severely ill patients as a proxy of ageing and covered a representative age range. The data that we analysed are original and were not included in any previous analyses of ageing clocks. Notwithstanding these strong points, the present study must also be interpreted within the context of its limitations. First, over nearly four decades of follow-up, as is the case in all long-term population studies, attrition of the FLEMENGHO cohort occurred. However, results of sensitivity analyses (appendix p 32) showed that participants included in analyses were older than those not included, suggesting that attrition mainly reflected the difficulty in keeping young adults in follow-up, who were pursuing secondary or higher education elsewhere or who were involved in occupational or familial responsibilities. Second, the FLEMENGHO cohort was recruited in a high-income country, while low-income and middle-income countries are transitioning rapidly from a disease burden dominated by communicable, maternal, neonatal, and nutritional causes to non-communicable diseases and injuries.<sup>1</sup> The ongoing Urinary Proteomics Combined with Home Blood Pressure Telemonitoring for Health Care Reform Trial (NCT04299529) is addressing UPP profiling in a multiethnic and multicultural context.<sup>42</sup> Third, the findings on osteoporosis and associated fractures are in keeping with the concept that women live longer than men do, but with a lower quality of life.<sup>43</sup> However, the present study did not elaborate on the health-care costs related to this issue. Fourth, there was a preponderance of short amino acid chains among the sequenced peptides, which

might be due to the reduced success of sequencing longer-chain peptides. Unknown post-translational modifications are the most likely cause for not being able to identify a peptide sequence with confidence. Peptide abundance also affects the quality of the tandem mass spectrometry spectra. However, although only 17% of the detected urinary peptides were sequenced, those currently sequenced represented 63% of the total peptide mass.<sup>44</sup> Fourth, proteins with high molecular weight, such as CILP,<sup>31</sup> do not readily pass the cell membrane, remain trapped within the extracellular matrix, and consequently might not be detected by UPP profiling, although they do play a role in pressure-induced left ventricular fibrosis and might be biomarkers of premature ageing.

In conclusion, ageing is associated with a specific and reproducible shift in the UPP signature, predominantly reflecting fibrosis. UPP-based classifiers are registered in-vitro diagnostics in Germany and the EU. Until now, UPP classifiers had specific clinical indications, such as screening for left ventricular dysfunction,<sup>4</sup> early risk stratification for progressing CKD<sup>5</sup> or diabetic nephropathy,<sup>5</sup> or predicting admission to intensive care facilities by patients with COVID-19.<sup>8</sup> In contrast to epigenetic or proteomic ageing clocks, UPP profiling is minimally intrusive and requires only a 10 mL urine sample, which can be processed without complex manipulation and from which all previously described disease-specific UPP classifiers can be generated. Thus, the UPP-ageing clock widens the clinical applicability of UPP profiling and might help in moving emphasis from treating to preventing chronic age-related diseases.<sup>42</sup> Furthermore, innovation in addressing the disability associated with ageing and age-related disorders should transcend the artificial categorisation of disease by organ system and should instead focus on treatment modalities that reduce the maladaptive fibrotic response to chronic injury across the body. Antifibrotic drugs, preferentially designed to stimulate collagen breakdown rather than to inhibit collagen biosynthesis, would allow prevention instead of treatment of established, progressive, and often irreversible disease, thereby increasing the sustainability of health care. Here, UPP profiling opens an additional window of opportunity, because in line with the suggestion of the European Medicines Agency to shorten the duration and reduce the cost of clinical trials, for instance in CKD,<sup>45</sup> a change in UPP-age could be an intermediary trial endpoint, accelerating innovation in drug discovery.

#### Contributors

DSM, HM, TSN, and JAS conceptualised the study. AL, JS, and HM did the proteomic urine analyses. AL, ST, AV, SJ, and HM provided expertise in the interpretation of the molecular pathways and metabolism of collagen and the extracellular matrix. DSM, LT, and JAS did the statistical analysis. HM and JB provided access to the patient data. JAS, LT, DSM, and HM had access to and verified the data reported in this study. DSM, LT, and JAS wrote the first draft of the manuscript. Z-YZ and CW participated in the conduct of the study. All authors interpreted the results, commented on successive drafts of the manuscript, and approved

the final version. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

#### Declaration of interests

HM is the cofounder and co-owner of Mosaïques Diagnostics (Hannover, Germany) and AL and JS are employees of Mosaïques Diagnostics. All other authors declare no competing interests.

#### Data sharing

Anonymised participants data will be made available upon request directed to the corresponding author. 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.

#### Acknowledgments

FLEMENGHO was supported by the EU (grants IC15-CT98-0329-EPOGH, LSHMCT037093-InGenious HyperCare, HEALTH-201550-HyperGenes, HEALTH-278249-EU-MASCARA, and HEALTH-305507HOMAGE), the European Research Council (advanced researcher grant 2011-294713-EPLORE and proof-of-concept grant 713601uPROPHET), the European Research Area Net for Cardiovascular Diseases (JTC2017-046-PROACT), and the Research Foundation Flanders, Ministry of the Flemish Community (Brussels, Belgium). APPREMED received a non-binding grant from OMRON Healthcare (Kyoto, Japan). DSM holds a postdoctoral grant from the Research Foundation Flanders (grant 12X9620N). Vera De Leebeek and Renilde Wolfs, of the Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, provided expert clerical assistance.

#### References

- GBD 2019 Viewpoint Collaborators. Five insights of the Global Burden of Disease Study 2019. *Lancet* 2020; **396**: 1135–59.
- Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996–2016. *JAMA* 2020; **323**: 863–84.
- Klein J, Papadopoulos T, Mischak H, Mullen W. Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. *Electrophoresis* 2014; **35**: 1060–64.
- Zhang ZY, Nkuiou-Kenfack E, Staessen JA. Urinary peptidomic biomarker for personalized prevention and treatment of diastolic left ventricular dysfunction. *Proteomics Clin Appl* 2019; **13**: e1800174.
- Toft N, Lindhardt M, Adamova K, et al. 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. *Lancet Diabetes Endocrinol* 2020; **8**: 301–12.
- Zhang ZY, Marrachelli VG, Yang WY, et al. Diastolic left ventricular function in relation to circulating metabolic biomarkers in a population study. *Eur J Prev Cardiol* 2019; **26**: 22–32.
- Maahs DM, Sivy J, Argilés A, et al. Urinary collagen fragments are significantly altered in diabetes: a link to pathophysiology. *PLoS One* 2010; **5**: e13051.
- Wendt R, Thijs L, Kalbitz S, et al. A urinary peptidomic profile predicts outcome in SARS-CoV-2-infected patients. *EclinicalMedicine* 2021; **36**: 100883.
- Pontillo C, Jacobs L, Staessen JA, et al. A urinary proteome-based classifier for the early detection of decline in glomerular filtration. *Nephrol Dial Transplant* 2017; **32**: 1510–16.
- Feng Y, Yu L. Generate non-overlapping equivalent samples to match an existing distribution in SAS. September, 2006. <https://lexjansen.com/nesug/nesug06/ap/ap17.pdf> (accessed Oct 2, 2021).
- Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med* 1999; **130**: 461–70.
- Peake M, Whiting M. Measurement of serum creatinine—current status and future goals. *Clin Biochem Rev* 2006; **27**: 173–84.
- Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985; **28**: 412–19.
- Wei FF, Trenson S, Verhamme P, Vermeer C, Staessen JA. Vitamin K-dependent matrix Gla protein as multifaceted protector of vascular and tissue integrity. *Hypertension* 2019; **73**: 1160–69.
- He T, Mischak M, Clark AL, et al. Urinary peptides in heart failure: a link to molecular pathophysiology. *Eur J Heart Fail* 2021; published online April 21. <https://doi.org/10.1002/ejhf.2195>.
- Lazar C, Gatto L, Ferro M, Bruley C, Burger T. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J Proteome Res* 2016; **15**: 1116–25.
- Robinson O, Chadeau Hyam M, Karaman I, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell* 2020; **19**: e13149.
- Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 2016; **12**: 477–79.
- Huang QF, Zhang ZY, Van Keer J, et al. Urinary peptidomic biomarkers of renal function in heart transplant recipients. *Nephrol Dial Transplant* 2019; **34**: 1336–43.
- Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* 2014; **15**: 786–801.
- Forsell C, Björk BF, Lilius L, et al. Genetic association to the amyloid plaque associated protein gene COL25A1 in Alzheimer's disease. *Neurobiol Aging* 2010; **31**: 409–15.
- Egaña-Gorroño L, López-Díez R, Yepuri G, et al. Receptor for advanced glycation end products (RAGE) and mechanisms and therapeutic opportunities in diabetes and cardiovascular disease: insights from human subjects and animal models. *Front Cardiovasc Med* 2020; **7**: 37.
- Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. The PI3K pathway in human disease. *Cell* 2017; **170**: 605–35.
- Haase M, Mertens PR, Haase-Fielitz A. Renal stress in vivo in real-time—visualised by the NGAL reporter mouse. *Nephrol Dial Transplant* 2011; **26**: 2109–11.
- Kaze AD, Musani SK, Bidulescu A, et al. Plasma leptin and blood pressure progression in Blacks: the Jackson Heart Study. *Hypertension* 2021; **77**: 1069–75.
- Johnson AA, Shokhirev MN, Wyss-Coray T, Lehallier B. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev* 2020; **60**: 101070.
- Bakun M, Senatorski G, Rubel T, et al. Urine proteomes of healthy aging humans reveal extracellular matrix (ECM) alterations and immune system dysfunction. *Age* 2014; **36**: 299–311.
- Pastushkova LK, Kononikhin AS, Tiys ES, et al. Characteristics of age-dependent changes in the urinary proteome in healthy men. *Adv Gerontol* 2016; **6**: 123–28.
- Zürbig P, Decramer S, Dakna M, et al. The human urinary proteome reveals high similarity between kidney aging and chronic kidney disease. *Proteomics* 2009; **9**: 2108–17.
- Nkuiou-Kenfack E, Bhat A, Klein J, et al. Identification of ageing-associated naturally occurring peptides in human urine. *Oncotarget* 2015; **6**: 34106–17.
- Trenson S, Hermans H, Craps S, et al. Cardiac microvascular endothelial cells in pressure overload-induced heart disease. *Circ Heart Fail* 2021; **14**: e006979.
- Bollong MJ, Yang B, Vergani N, et al. Small molecule-mediated inhibition of myofibroblast transdifferentiation for the treatment of fibrosis. *Proc Natl Acad Sci USA* 2017; **114**: 4679–84.
- Li X, Zhu L, Wang B, Yuan M, Zhu R. Drugs and targets in fibrosis. *Front Pharmacol* 2017; **8**: 855.
- Podolsky MJ, Yang CD, Valenzuela CL, et al. Age-dependent regulation of cell-mediated collagen turnover. *JCI Insight* 2020; **5**: e137519.
- Stepiowski A, Fertala A. Inhibition of collagen fibril formation. *Fibrogenesis Tissue Repair* 2012; **5** (suppl 1): S29.
- Licini C, Vitale-Brovarene C, Mattioli-Belmonte M. Collagen and non-collagenous proteins molecular crosstalk in the pathophysiology of osteoporosis. *Cytokine Growth Factor Rev* 2019; **49**: 59–69.
- Nagy EE, Nagy-Finnc C, Popoviciu H, Kovács B. Soluble biomarkers of osteoporosis and osteoarthritis, from pathway mapping to clinical trials: an update. *Clin Interv Aging* 2020; **15**: 501–18.

For more on APPREMED see <https://www.appremed.org>

- 38 Asadipooya K, Uy EM. Advanced glycation end products (AGEs), receptor for AGEs, diabetes, and bone: review of the literature. *J Endocr Soc* 2019; **3**: 1799–818.
- 39 Elshaikh AO, Shah L, Joy Mathew C, Lee R, Jose MT, Cancarevic I. Influence of vitamin K on bone mineral density and osteoporosis. *Cureus* 2020; **12**: e10816.
- 40 Akbari S, Rasouli-Ghahroudi AA. Vitamin K and bone metabolism: a review of the latest evidence in preclinical studies. *BioMed Res Int* 2018; **2018**: 4629383.
- 41 Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* 2014; **123**: 1214–17.
- 42 Thijs L, Asayama K, Maestre GE, et al. Urinary proteomics combined with home blood pressure telemonitoring for health care reform trial: rational and protocol. *Blood Press* 2021; published online Aug 30. <https://doi.org/10.1080/08037051.2021.1952061>.
- 43 Eurostat. The life of women and men in Europe—a statistical portrait. 2010. <https://webgate.ec.europa.eu/eurostat/web/products-statistical-books/-/ks-80-07-135> (accessed Oct 2, 2021).
- 44 Latosinska A, Siwy J, Mischak H, Frantzi M. Peptidomics and proteomics based on CE-MS as a robust tool in clinical application: the past, the present, and the future. *Electrophoresis* 2019; **40**: 2294–308.
- 45 European Medicines Agency. Guideline on the clinical investigation of medicinal products to prevent development/slow progression of chronic renal insufficiency. Oct 20, 2016. <https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-prevent-developmentslow-progression-chronic-renal> (accessed Oct 2, 2021).