

RESEARCH ARTICLE

Open Access



Very short sleep duration reveals a proteomic fingerprint that is selectively associated with incident diabetes mellitus but not with incident coronary heart disease: a cohort study

Thomas Svensson^{1,2,3*} , Akiko Kishi Svensson^{1,2,4}, Mariusz Kitlinski⁵, Gunnar Engström¹, Jan Nilsson¹, Marju Orho-Melander¹, Peter M. Nilsson¹ and Olle Melander^{1,6}

Abstract

Background The molecular pathways linking short and long sleep duration with incident diabetes mellitus (iDM) and incident coronary heart disease (iCHD) are not known. We aimed to identify circulating protein patterns associated with sleep duration and test their impact on incident cardiometabolic disease.

Methods We assessed sleep duration and measured 78 plasma proteins among 3336 participants aged 46–68 years, free from DM and CHD at baseline, and identified cases of iDM and iCHD using national registers. Incident events occurring in the first 3 years of follow-up were excluded from analyses. Tenfold cross-fit partialing-out lasso logistic regression adjusted for age and sex was used to identify proteins that significantly predicted sleep duration quintiles when compared with the referent quintile 3 (Q3). Predictive proteins were weighted and combined into proteomic scores (PS) for sleep duration Q1, Q2, Q4, and Q5. Combinations of PS were included in a linear regression model to identify the best predictors of habitual sleep duration. Cox proportional hazards regression models with sleep duration quintiles and sleep-predictive PS as the main exposures were related to iDM and iCHD after adjustment for known covariates.

Results Sixteen unique proteomic markers, predominantly reflecting inflammation and apoptosis, predicted sleep duration quintiles. The combination of PSQ1 and PSQ5 best predicted sleep duration. Mean follow-up times for iDM ($n = 522$) and iCHD ($n = 411$) were 21.8 and 22.4 years, respectively. Compared with sleep duration Q3, all sleep duration quintiles were positively and significantly associated with iDM. Only sleep duration Q1 was positively and significantly associated with iCHD. Inclusion of PSQ1 and PSQ5 abrogated the association between sleep duration Q1 and iDM. Moreover, PSQ1 was significantly associated with iDM (HR = 1.27, 95% CI: 1.06–1.53). PSQ1 and PSQ5 were not associated with iCHD and did not markedly attenuate the association between sleep duration Q1 with iCHD.

Conclusions We here identify plasma proteomic fingerprints of sleep duration and suggest that PSQ1 could explain the association between very short sleep duration and incident DM.

*Correspondence:

Thomas Svensson
thomas.svensson@med.lu.se

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Incident coronary heart disease, Incident diabetes, Inflammation, Lasso, Machine learning, Proteomic markers, Sleep duration

Background

The associations between short or long sleep durations and diabetes mellitus (DM) [1] and coronary heart disease (CHD) [2], respectively, have been established in multiple studies over several decades and in different populations. Despite overwhelming evidence and the known epidemiological associations between short and long sleep durations and cardiometabolic disease, much less is known about the potential biological intermediates that may explain these associations. One prospective study showed that short sleep duration is associated with inflammatory markers such as C-reactive protein (CRP) [3] and interleukin 6 (IL-6) [4], and a meta-analysis found that when compared to a sleep duration of 7–8 h, sleep duration > 8 h is associated with increased CRP and IL-6 [5]. These associations are not inconsequential as systemic inflammatory markers such as CRP [6] and IL-6 [7] in turn are associated with cardiometabolic outcomes, e.g., CHD. One limitation, however, of existing studies is the high probability of reverse causality where the disease process impacts sleep. Another important limitation may be the inflammatory markers themselves; they are non-specific to sleep duration or the health outcome and thus offer only limited information about possible pathways that link sleep duration with adverse health outcomes.

The advent of proteomic assays allows for a new approach to investigate potential biological pathways through which short or long sleep durations are associated with cardiometabolic outcomes. We have previously identified that plasma concentrations of caspase 8, an enzyme in the tumor necrosis factor receptor pathway, is associated with short sleep duration and incident DM [8]. An important recent study [9] further found that five proteomic markers related to cardiovascular risk were associated with sleep duration. Unfortunately, this study could not further investigate the prospective association between the identified proteomic markers and disease outcomes.

The aim of the present study is to identify proteomic markers associated with specific sleep duration categories. The secondary aim is to create proteomic sleep scores using the proteomic markers predictive of each sleep duration category and to further investigate the associations between the proteomic sleep scores with incident DM and incident CHD, respectively. We hypothesize that the proteomic sleep scores will be associated with both incident DM and incident CHD. Moreover, we hypothesize that the proteomic scores will abrogate any

association between their corresponding phenotype and said outcomes. The proteomic scores are thus hypothesized to be on the pathway between aberrant sleep duration and incident DM and incident CHD, respectively.

Methods

The Malmö Diet and Cancer (MDC) study is a population-based, prospective study in the city of Malmö, Sweden. Men and women aged 45–73 years were randomly selected between the years 1991 and 1996 and recruited for a baseline examination. Participants provided anthropometric data and blood samples and answered a detailed lifestyle questionnaire on heredity, socioeconomic variables, social network, occupation, physical activity, alcohol consumption, smoking, diseases, and medication. Details of the study have been described elsewhere [10]. Between 1991 and 1994, 6103 individuals from the MDC study were randomly selected to participate in the MDC Cardiovascular Cohort (MDC-CC). The purpose of the MDC-CC is to study the epidemiology of carotid artery disease [11]. MDC-CC participants underwent detailed examinations, including ultrasonography of the carotid artery, and provided plasma for the measure of novel proteomic markers.

For the purpose of the present study, MDC-CC participants were excluded if they had prevalent DM ($n=293$) or a fasting whole blood glucose concentration ≥ 6.1 mmol/l ($n=261$) or prevalent CHD ($n=100$) at baseline (Fig. 1). Prevalent DM was defined as: having a measured fasting whole blood glucose ≥ 6.1 mmol/l (corresponding to fasting plasma glucose concentration ≥ 7 mmol/l) at the MDC baseline examination, a self-reported history of physician diagnosed DM, the use of DM medication according to the MDC baseline questionnaire, or being diagnosed and registered in any of the local or national diabetes registries as described elsewhere [12]. Participants were further excluded if they had provided incomplete information on sleep duration ($n=294$) or if their sleep duration represented outlier values of more than 3 interquartile ranges below or above the first and fourth quartiles, respectively ($n=15$), had missing data on kidney function, as measured by cystatin C ($n=677$), or missing data for any of the covariates ($n=267$). In order to minimize reverse causation bias, participants who were diagnosed with incident DM ($n=41$) or incident CHD ($n=24$) in the first 3 years of follow-up were also excluded. Finally, those with missing data for any of the proteomic markers ($n=552$) or who

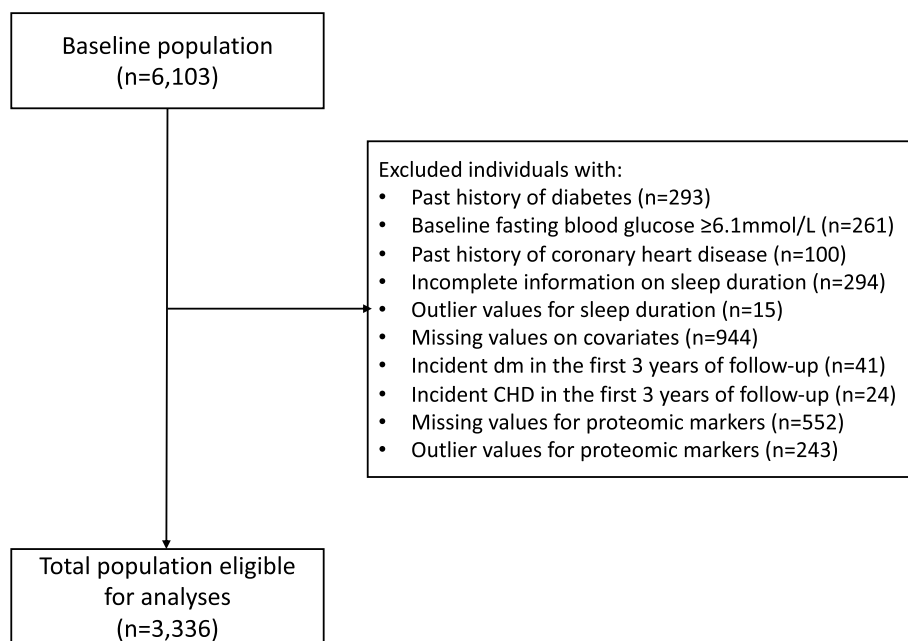


Fig. 1 Flowchart of the inclusion and exclusion of study participants

had levels of proteomic markers that represented outlier values of more than 3 interquartile ranges below or above the first and fourth quartiles, respectively ($n=243$), were also excluded. The final sample for analyses consisted of 3336 participants. All participants were followed from starting point until December 31, 2018, with person-years calculated from starting point to incident DM and incident CHD, date of death, or end of follow-up period, whichever came first. Participants who were diagnosed with both incident DM and incident CHD were included in the analyses of both outcomes unless the incident CHD event was fatal while preceding incident DM.

The MDC study was approved by the ethics committee at Lund University, and all participants provided written informed consent.

Proteomic markers

The 92 proteomic markers investigated in this study were from the Olink Proseek Multiplex CVD 1 panel (Additional file 1: Table S1). The proteomic markers were measured in stored fasting plasma specimens from the MDC-CC baseline examination. The specimens were immediately frozen to -80°C following collection. Plasma concentrations of the proteomic markers were quantified using a validated high-specificity immunoassay, the Proximity Extension Assay (PEA), which has been described in detail elsewhere [13, 14].

Concentrations of proteomic markers were provided on a logarithmic (\log_2) scale. Any markers that were below the limit of detection were provided a missing

value. For the purpose of this study, 14 protein markers with missing information for $\geq 10\%$ ($n=413$) of the study population were excluded from analyses (details in Additional file 1: Table S1). The concentrations of the 78 proteomic markers included in analyses were standardized, where the standard score of each marker represented the number of standard deviations above or below the mean.

Sleep duration

Habitual sleep duration was assessed through two open questions asking participants how long they slept on weekdays and weekends, respectively. The two questions were formulated as: (a) "How many hours do you usually sleep per night during a typical week (Monday-Friday)?" and (b) "How many hours do you usually sleep per night during a typical weekend (Saturday-Sunday)?" A weighted average sleep duration $[(\text{weekday} \times 5) + (\text{weekend} \times 2)]/7$ was calculated for all participants, thereby allowing the subsequent construction of a categorical variable representing sleep duration quintiles (Quintile 1 [Q1]: 4.00–6.57 h; Q2: 6.64–7.14 h; Q3: 7.21–7.57 h; Q4: 7.64–8.00 h; Q5: 8.14–11.00 h). Q3 was chosen as the reference category for two main reasons: first, the sleep duration range in this group was most similar to other studies which often use 7–8 h as the reference, thus facilitating comparisons and second, selection of the mid-quintile allowed for the investigation of J-shaped or U-shaped associations that are often reported between sleep duration and the respective outcomes e.g., [2, 15].

Incident diabetes mellitus

Incident DM was defined as new onset DM in individuals without prevalent DM at the MDC baseline examination. All incident DM events were identified through linkage of a 10-digit national personal identification number with 6 local and national registers: the Malmö HbA_{1c} register, the Regional Diabetes 2000 register of the Scania region [16], the Swedish National Diabetes Register [17], the Swedish National Inpatient Register [18], the Swedish Cause of Death Register [19], and the Swedish Prescribed Drug Register [20]. Participants with incident DM could also be identified by having a fasting plasma glucose concentration ≥ 7 mmol/l or a 120-min plasma glucose value of > 11.0 mmol/l in subpopulations of the MDC participating in a MDC re-examination [21] or the Malmö Preventive Project re-examination [22]. The detailed process of DM end-point retrieval has been described elsewhere [23].

Incident coronary heart disease

Incident CHD event was defined as a first fatal or non-fatal myocardial infarction, coronary artery bypass graft (CABG), or percutaneous coronary intervention (PCI). All incident CHD events were identified through linkage of a ten-digit national personal identification number with three registries validated for classification of outcomes as described elsewhere [24, 25]: the Swedish National Discharge Registry, the Swedish National Cause of Death Registry, and the Swedish Coronary Angiography and Angioplasty Registry. CABG and PCI were classified using the national classification of surgical procedures operation codes (KKÅ or Op6): 3065, 3066, 3068, 3080, 3092, 3105, 3127, 3158 for CABG, and FNG02 and FNG05 for PCI. Coronary event was defined according to the International Classification of Diseases, ninth (ICD-9) and tenth (ICD-10) revisions with fatal or non-fatal myocardial infarction (MI) or death due to CHD corresponding to codes 410, 412, and 414 (ICD-9), and I21-I23, and I25 (ICD-10).

Covariates

Covariates in the statistical models included age at baseline (continuous), sex, cystatin C (continuous in mg/l), education (elementary school or higher than elementary school), and physical activity defined as quartiles of leisure time physical activity based on 18 items adapted from the Minnesota Leisure Time Physical Activity instrument which has been described in detail elsewhere [26]. Smoking was defined as never, past, and current smoker (< 20 cig./day, or ≥ 20 cig./day); alcohol consumption was considered as quartiles with non-consumption as the reference value (none, 0.02–1.70, 1.70–6.87, 6.88–14.36, or > 14.36 g ethanol/day), and shift work was

a binary variable (yes/no). Symptoms of insomnia were based on four items: difficulty initiating sleep, difficulty maintaining sleep, early morning awakening, and not feeling rested after sleep; the score of each included item (0–3 points) were summed to an overall discrete insomnia score (0–12 points). Body mass index (BMI) was considered in categories (< 18.5 kg/m², 18.5–24.9 kg/m², 25.0–29.9 kg/m², or ≥ 30.0 kg/m²). Waist circumference (in cm), low-density lipoprotein cholesterol (LDL-C, in mmol/l), high-density lipoprotein cholesterol (HDL-C, in mmol/l), triglycerides (TG, in mmol/l), and hemoglobin A_{1c} (HbA_{1c}, in %) were all treated as continuous variables. HbA_{1c} was used as a stratifying variable given that its inclusion as a covariate violated the assumptions of proportional hazards in semi-parametric survival analyses with incident DM as the outcome.

Statistical analyses

Several of the proteomic markers in the Olink panels are expected to be highly correlated. Multicollinearity of predictors necessitates a model selection procedure retaining only markers that are relevant for the outcome of interest and have the largest effect. Lasso, least absolute shrinkage and selection operator, is an analysis method that produces sparse models by shrinking the coefficients of some predictors while setting the coefficients of other predictors to “0” [27]. This allows for improved prediction accuracy and improved interpretation [28]. However, lasso does not provide accurate standard errors of its estimate [27] which in turn precludes significance testing [29]. Cross-fit partialing-out lasso, also known as double machine learning [30], is a method that produces both coefficients and standard errors of predictors while controlling for relevant covariates. Cross-fit partialing out lasso splits samples into multiple folds selecting covariates and estimating post-selection coefficients [31]. The final results are obtained by averaging the results of multiple estimates acquired across the folds [30].

Ten-fold cross-fit partialing-out lasso logistic regression controlling for age and sex was used to estimate beta coefficients, standard errors, and *p*-values for each of the 78 proteomic markers' associations with specific quintiles of sleep duration. Each tenfold cross-fit partialing-out lasso logistic regression was performed against a binary outcome variable where sleep duration Q3 was considered the reference value and where, in four separate analyses, the remaining four sleep duration quintiles were considered the respective outcomes of interest, i.e., cross-fit lasso 1: Q1 vs. Q3; cross-fit lasso 2: Q2 vs. Q3; cross-fit lasso 3: Q4 vs. Q3; and cross-fit lasso 4: Q5 vs. Q3. Proteomic markers that were significantly associated with Q1, Q2, Q4, and Q5, respectively, were retained (*p* for retention: *p* < 0.05). Using their beta coefficients as

weights, the retained proteomic markers were then used to create four separate proteomic scores, one for each sleep quintile compared to Q3.

The four proteomic scores were included as independent variables in linear regression analyses with continuous sleep duration as the dependent variable. All possible combinations ($n=15$) of the 4 proteomic scores were considered and each model was compared to the remaining models using the Akaike Information Criterion (AIC). The proteomic scores from the model with the lowest AIC were retained for the semi-parametric survival analyses.

Cox proportional hazards regression was used to determine hazard ratios (HR) and 95% confidence intervals (CI) for the associations of sleep duration and proteomic risk scores with incident DM and incident CHD, respectively. Primary analyses investigated the association between sleep duration with the respective outcomes. Secondary analyses considered identical models to the primary analyses with proteomic risk scores as additional main exposures, i.e., the models were adjusted for both sleep duration and proteomic score. All statistical models were stratified by HbA_{1c} using the Stata option “strata()” to account for different baseline hazards across concentrations of HbA_{1c}. The minimally adjusted model (model 1) was adjusted for age, and sex; model 2 was additionally adjusted for cystatin C, education, physical activity, smoking, alcohol consumption, shift work, and symptoms of insomnia; and model 3 was additionally adjusted for BMI, waist circumference, LDL-C, HDL-C, and TG. No sex-specific stratification was done as there were no significant interactions between sex and sleep duration.

The HR of proteomic sleep duration scores are expressed in terms of incremental increases per SD.

Global tests for proportionality were conducted for the final multivariable models and did not reveal any significant deviation from the proportional hazards assumption.

A proteomic score was considered a probable mediator for the association between sleep duration and incident DM or incident CHD if the following conditions were met: (1) the inclusion of the proteomic score in the final multivariable model abrogated any significant associations between sleep duration and the outcome and (2) the proteomic score was significantly associated with the outcome. Mediating effects (survival functions for the follow-up period and the proportion mediated including 95% CI) were determined using the Stata “standsurv” post-estimation command on a fitted flexible parametric survival model that included only the main exposure (the relevant sleep duration quintile) and the probable mediator (protein score). The fitted models used three degrees of freedom for the baseline hazard and considered the

sleep duration quintile as a time varying effect with three degrees of freedom.

All statistical analyses were performed using Stata/MP 17.0 (StataCorp LP, College Station, TX). The significance level was set as $p < 0.05$.

Data sharing statement

Due to ethical and legal restrictions related to the Swedish Biobanks in Medical Care Act (2002:297) and the Personal Data Act (1998:204), data are available upon request from the data access group of Malmö Diet and Cancer study by contacting Anders Dahlin (anders.dahlin@med.lu.se).

Role of the funding source

The sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Results

Table 1 shows baseline characteristics of participants according to quintiles of habitual sleep duration. Individuals in quintile 3 (Q3; referent category in survival analyses) had the lowest average age, score of insomnia symptoms, and waist circumference. They further had the lowest proportions of participants in the highest physical activity level, non-consumers of alcohol and shift workers, and the highest proportions of participants with education greater than elementary school and with high alcohol intake. Individuals with the shortest habitual sleep duration (Q1) had the highest proportion of non-consumers of alcohol, the highest insomnia score, and the highest LDL-C.

Additional file 1: Table S2 shows the beta coefficients, standard errors, z-scores, p-values, and 95% confidence intervals of all 78 proteomic markers for their association with each sleep duration quintile when compared to sleep duration Q3 using the cross-fit partialing-out lasso logistic regression. A total of 16 unique proteomic markers were significantly associated with sleep duration quintiles Q1, Q2, Q4, and Q5 when compared to Q3 (Table 2); 6 proteomic markers were significantly associated with sleep duration Q1; 4 proteomic markers were significantly associated with Q2; 5 proteomic markers were significantly associated with Q4; and 6 proteomic markers were significantly associated with Q5. Thirteen of the markers were significantly associated only with one specific sleep quintile; 1 proteomic marker was associated with both short sleep duration quintiles (Q1 and Q2); and 2 proteomic markers were associated with both short and long sleep duration quintiles.

Table 1 Baseline characteristics according to habitual sleep duration quintiles

| Variable | Sleep duration quintile (Q; hours) | | | | | p-value ^a |
|--|------------------------------------|----------------|----------------|----------------|-----------------|----------------------|
| | Q1 (4.00–6.57) | Q2 (6.64–7.14) | Q3 (7.21–7.57) | Q4 (7.64–8.00) | Q5 (8.14–11.00) | |
| Number of individuals | 715 | 668 | 804 | 764 | 435 | |
| Proportion of population (%) | 21.4 | 18.5 | 24.1 | 22.9 | 13.0 | |
| Age [mean (years ± s.d.)] | 57.1 ± 5.8 | 58.5 ± 5.9 | 55.3 ± 5.5 | 58.5 ± 5.9 | 57.3 ± 5.9 | < 0.001 |
| Men (%) | 36.5 | 38.7 | 38.6 | 35.2 | 32.2 | 0.15 |
| Cystatin C [mean (mg/L ± s.d.)] | 0.780 ± 0.133 | 0.774 ± 0.120 | 0.754 ± 0.124 | 0.773 ± 0.129 | 0.766 ± 0.125 | 0.001 |
| Education (%) | | | | | | 0.003 |
| Higher than elementary school | 56.5 | 55.2 | 61.9 | 52.8 | 53.3 | |
| Physical activity (%) | | | | | | 0.05 |
| Quartile 1 (low physical activity) | 26.0 | 23.8 | 24.3 | 24.2 | 27.4 | |
| Quartile 2 | 22.9 | 26.5 | 29.1 | 23.4 | 25.1 | |
| Quartile 3 | 24.3 | 24.3 | 26.0 | 24.4 | 22.1 | |
| Quartile 4 (high physical activity) | 26.7 | 25.4 | 20.7 | 28.0 | 25.5 | |
| Smoking status (%) | | | | | | 0.44 |
| Never | 40.1 | 43.5 | 43.7 | 45.0 | 39.5 | |
| Past | 33.7 | 32.0 | 30.6 | 31.9 | 34.3 | |
| Current < 20 cigarettes/day | 13.6 | 12.3 | 14.3 | 11.9 | 11.3 | |
| Current ≥ 20 cigarettes/day | 12.6 | 12.1 | 11.4 | 11.1 | 14.9 | |
| Alcohol consumption (%) | | | | | | 0.03 |
| None | 18.0 | 15.7 | 12.2 | 16.1 | 17.2 | |
| 0.02–1.70 g ethanol/day | 10.8 | 9.7 | 9.8 | 10.7 | 12.2 | |
| 1.70–6.87 g ethanol/day | 22.5 | 25.2 | 22.6 | 27.4 | 25.3 | |
| 6.89–14.36 g ethanol/day | 24.5 | 23.6 | 27.1 | 24.6 | 22.5 | |
| ≥ 14.36 g ethanol/day | 24.2 | 25.7 | 28.2 | 21.2 | 22.8 | |
| Shift work (%) | 28.8 | 30.1 | 22.8 | 27.1 | 32.2 | 0.002 |
| Insomnia symptoms [median (score, iqr.)] | 4 (5) | 3 (4) | 2 (3) | 3 (4) | 3 (3) | < 0.001 |
| Waist circumference [mean (cm ± s.d.)] | 82.9 ± 12.7 | 83.4 ± 12.0 | 81.6 ± 11.5 | 82.1 ± 12.1 | 81.9 ± 12.3 | 0.048 |
| LDL-C [mean (mmol/L ± s.d.)] | 4.3 ± 1.0 | 4.2 ± 0.9 | 4.1 ± 1.0 | 4.2 ± 1.0 | 4.1 ± 0.9 | 0.01 |
| HDL-C [mean (mmol/L ± s.d.)] | 1.4 ± 0.4 | 1.4 ± 0.4 | 1.4 ± 0.4 | 1.4 ± 0.4 | 1.4 ± 0.4 | 0.89 |
| Triglycerides [mean (mmol/L ± s.d.)] | 1.3 ± 0.6 | 1.3 ± 0.6 | 1.3 ± 0.6 | 1.3 ± 0.6 | 1.3 ± 0.7 | 0.81 |
| HbA1c [mean (% ± s.d.)] | 4.8 ± 0.4 | 4.8 ± 0.4 | 4.7 ± 0.4 | 4.8 ± 0.4 | 4.8 ± 0.4 | 0.14 |
| BMI | | | | | | 0.16 |
| < 18.5 kg/m ² | 1.1 | 1.0 | 0.6 | 1.1 | 0.7 | |
| 18.5–25.0 kg/m ² | 47.7 | 45.3 | 52.4 | 48.4 | 49.4 | |
| 25.0–30.0 kg/m ² | 38.6 | 42.2 | 39.3 | 39.1 | 38.6 | |
| ≥ 30.0 kg/m ² | 12.6 | 11.5 | 7.7 | 11.4 | 11.3 | |

BMI body mass index, iqr interquartile range, s.d. standard deviation

^a Chi-square test for categorical variables, ANOVA for continuous variables (age, cystatin C, waist circumference, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, and hemoglobin A1c), Kruskal–Wallis test for insomnia symptoms

Linear regression analyses of the 15 combinations of the 4 proteomic scores (4 models including 1 score each; 6 models combining variations of 2 scores; 4 models combining variations of 3 scores; and 1 model including all scores) revealed that the model that included proteomic scores for Q1 and Q5 had the lowest AIC and thus best predicted sleep duration (data not shown).

Sleep duration, proteomic scores, and incident diabetes mellitus

Mean follow-up time for incident DM was 21.8 years (72,565 person-years) during which there were 522 cases of incident DM. In the age- and sex adjusted model (model 1), all sleep duration quintiles were significantly and positively associated with incident DM when compared with Q3 (Q1: hazard ratio [HR] = 1.41,

Table 2 Proteomic markers significantly associated with specific sleep duration quintiles when compared to referent sleep duration (quintile 3) using tenfold cross-fit partialing out lasso logistic regression analyses

| Sleep duration quintile | Proteomic marker | Beta coefficient | SE | z | p-value | 95% confidence interval |
|--------------------------|------------------|------------------|-------|-------|---------|-------------------------|
| Quintile 1 vs Quintile 3 | TRANCE | -0.199 | 0.084 | -2.37 | 0.018 | -0.363 to -0.035 |
| | MMP-7 | 0.193 | 0.082 | 2.37 | 0.018 | 0.033 to 0.353 |
| | MMP-10 | -0.160 | 0.069 | -2.32 | 0.020 | -0.295 to -0.025 |
| | Follistatin | 0.206 | 0.091 | 2.25 | 0.024 | 0.027 to 0.385 |
| | E-selectin | 0.201 | 0.091 | 2.22 | 0.027 | 0.023 to 0.378 |
| | TRAIL-R2 | 0.266 | 0.133 | 2.00 | 0.046 | 0.005 to 0.528 |
| Quintile 2 vs Quintile 3 | TRAIL-R2 | 0.431 | 0.143 | 3.02 | 0.0026 | 0.151 to 0.711 |
| | FAS | -0.457 | 0.156 | -2.92 | 0.0035 | -0.763 to -0.150 |
| | Kallikrein-6 | -0.284 | 0.111 | -2.57 | 0.010 | -0.500 to -0.067 |
| | U-PAR | -0.302 | 0.140 | -2.16 | 0.031 | -0.576 to -0.0280 |
| Quintile 4 vs Quintile 3 | FAS | -0.452 | 0.135 | -3.34 | 0.00084 | -0.717 to -0.187 |
| | Renin | 0.186 | 0.070 | 2.64 | 0.0082 | 0.048 to 0.324 |
| | HB-EGF | -0.329 | 0.135 | -2.44 | 0.015 | -0.594 to -0.064 |
| | MMP-7 | 0.189 | 0.081 | 2.34 | 0.019 | 0.031 to 0.348 |
| | CXCL6 | 0.210 | 0.097 | 2.16 | 0.031 | 0.019 to 0.400 |
| Quintile 5 vs Quintile 3 | Prolactin | 0.261 | 0.081 | 3.22 | 0.0013 | 0.102 to 0.420 |
| | MMP-7 | 0.296 | 0.097 | 3.05 | 0.0023 | 0.106 to 0.486 |
| | CXCL1 | 0.245 | 0.100 | 2.44 | 0.015 | 0.0483 to 0.441 |
| | FAS | -0.402 | 0.170 | -2.37 | 0.018 | -0.735 to -0.070 |
| | tPA | 0.256 | 0.110 | 2.32 | 0.020 | 0.0395 to 0.472 |
| | HSP27 | -0.284 | 0.130 | -2.19 | 0.029 | -0.538 to -0.030 |

TRANCE tumor necrosis factor-related activation-induced cytokine, MMP-7 matrix metalloproteinase-7, MMP-10 matrix metalloproteinase-10, TRAIL-R2 tumor necrosis factor-related apoptosis-inducing ligand receptor 2, FAS tumor necrosis factor receptor superfamily member 6, U-PAR urokinase plasminogen activator surface receptor, HB-EGF proheparin-binding epidermal growth factor-like growth factor, CXCL6 C-X-C motif chemokine 6, CXCL1 C-X-C motif chemokine 1, tPA tissue-type plasminogen activator, HSP27 heat shock 27 kDa protein

95% CI: 1.08–1.84; Q2: HR=1.40, 95% CI: 1.06–1.85; Q4: HR=1.32, 95% CI: 1.01–1.72; Q5: HR=1.50, 95% CI: 1.11–2.03) (Table 3). The associations were only minimally attenuated in model 2, with all but one sleep duration quintile remaining significantly associated with incident DM in model 3 (Q1: HR=1.32, 95% CI: 1.00–1.74; Q2: HR=1.33, 95% CI: 1.00–1.76; Q5: HR=1.48,

95% CI: 1.09–2.00). The addition of proteomic scores for sleep duration Q1 and Q5 attenuated the associations in model 1; in models mutually adjusted for sleep duration and proteomic scores, sleep duration Q1 (HR=1.33, 95% CI: 1.02–1.74), Q2 (HR=1.34, 95% CI: 1.01–1.77), and Q5 (HR=1.46, 95% CI: 1.08–1.97) were significantly associated with incident DM (Table 4). The addition of

Table 3 Cox proportional hazards regression models for the association between sleep duration and incident diabetes mellitus

| | Model 1 | | Model 2 | | Model 3 | |
|----------------------------------|-----------|--------------|-----------|--------------|-----------|--------------|
| | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Sleep duration Q1 (4.00–6.57 h) | 1.41* | (1.08, 1.84) | 1.32* | (1.01, 1.74) | 1.32* | (1.00, 1.74) |
| Sleep duration Q2 (6.64–7.14 h) | 1.40* | (1.06, 1.85) | 1.40* | (1.05, 1.85) | 1.33* | (1.00, 1.76) |
| Sleep duration Q3 (7.21–7.57 h) | Reference | | Reference | | Reference | |
| Sleep duration Q4 (7.64–8.00 h) | 1.32* | (1.01, 1.72) | 1.31* | (1.00, 1.72) | 1.30 | (0.99, 1.70) |
| Sleep duration Q5 (8.14–11.00 h) | 1.50** | (1.11, 2.03) | 1.47* | (1.09, 1.99) | 1.48* | (1.09, 2.00) |

Italic values denote statistically significant results

Model 1 is stratified by HbA1c concentration and adjusted for age and sex

Model 2 is additionally adjusted for cystatin C, education, physical activity, smoking, alcohol consumption, shift work, and insomnia symptoms

Model 3 is additionally adjusted for body mass index, waist circumference, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides

* $p < 0.05$, ** $p < 0.01$

Table 4 Cox proportional hazards regression models for the association between proteomic sleep score and sleep duration with incident diabetes mellitus. Analyses are mutually adjusted for proteomic sleep score and sleep duration

| | Model 1 | | Model 2 | | Model 3 | |
|----------------------------------|----------------|---------------------|----------------|---------------------|---------------|---------------------|
| | HR | (95% CI) | HR | 95% CI | HR | 95% CI |
| Proteomic score Q1 ^a | <i>1.52***</i> | <i>(1.28, 1.81)</i> | <i>1.48***</i> | <i>(1.24, 1.77)</i> | <i>1.27**</i> | <i>(1.06, 1.53)</i> |
| Proteomic score Q5 ^a | <i>1.22*</i> | <i>(1.01, 1.48)</i> | <i>1.25*</i> | <i>(1.03, 1.52)</i> | 1.20 | (0.99, 1.45) |
| Sleep duration Q1 (4.00–6.57 h) | <i>1.33*</i> | <i>(1.02, 1.74)</i> | 1.26 | (0.96, 1.66) | 1.27 | (0.96, 1.67) |
| Sleep duration Q2 (6.64–7.14 h) | <i>1.34*</i> | <i>(1.01, 1.77)</i> | <i>1.34*</i> | <i>(1.01, 1.77)</i> | 1.29 | (0.97, 1.71) |
| Sleep duration Q3 (7.21–7.57 h) | Reference | | Reference | | Reference | |
| Sleep duration Q4 (7.64–8.00 h) | 1.27 | (0.97, 1.66) | 1.26 | (0.96, 1.65) | 1.25 | (0.95, 1.64) |
| Sleep duration Q5 (8.14–11.00 h) | <i>1.46*</i> | <i>(1.08, 1.97)</i> | <i>1.42*</i> | <i>(1.05, 1.93)</i> | <i>1.42*</i> | <i>(1.04, 1.93)</i> |

Italic values denote statistically significant results

Model 1 is stratified by HbA1c concentration and adjusted for age and sex

Model 2 is additionally adjusted for cystatin C, education, physical activity, smoking, alcohol consumption, shift work, and insomnia symptoms

Model 3 is additionally adjusted for body mass index, waist circumference, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides

^a Expressed as the HR and 95% CI of the incremental increase per SD of the proteomic sleep score

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

covariates in model 2 abrogated the association between sleep duration Q1 and incident DM and attenuated the association between sleep duration Q5 and incident DM. In model 3, only sleep duration Q5 remained significantly and positively associated with incident DM (HR=1.42, 95% CI: 1.04–1.93). The proteomic score for sleep duration Q1 was attenuated with the inclusion of covariates in models 2 and 3 but remained significantly and positively associated with incident DM in all models (model 1, proteomic score Q1: HR=1.52, 95% CI: 1.28–1.81; model 2, proteomic score Q1: HR=1.48, 95% CI: 1.24–1.77; model 3, proteomic score Q1: HR=1.27, 95% CI:

1.06–1.53). The proteomic score for sleep duration Q5 was significantly and positively associated with incident DM in model 1 (HR=1.22, 95% CI: 1.01–1.48) and model 2 (HR=1.25, 95% CI: 1.03–1.52), but was abrogated following the inclusion of covariates in model 3 (HR=1.20, 95% CI: 0.99–1.45).

Mediation analyses found that when compared to referent sleep duration Q3, the proteomic risk score for sleep duration Q1 significantly mediated between 32 and 53% of the association between sleep duration Q1 and incident DM during follow-up years 11.1 to 27.2 (Fig. 2).

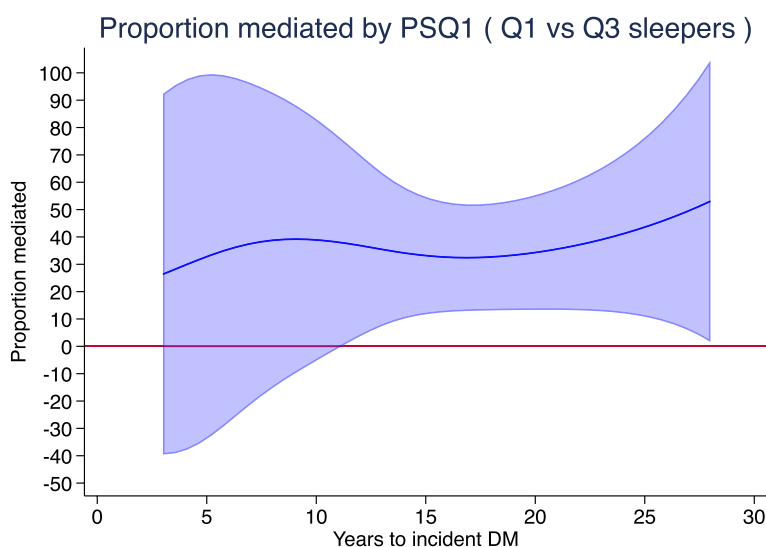


Fig. 2 Proportion of the association between sleep duration Q1 and incident diabetes mellitus (DM) mediated by the proteomic risk score for sleep duration Q1 (PSQ1). Light blue area indicates 95% confidence intervals; horizontal line is a reference indicator for 0% mediation. Confidence intervals above the reference line represent statistically significant mediation

The protein score was not a significant mediator during follow-up years 3.0 to 11.0.

Sleep duration, proteomic scores, and incident coronary heart disease

Mean follow-up time for incident CHD was 22.4 years (74,681 person-years) during which there were 411 cases of incident CHD. In the age- and sex adjusted model (model 1), sleep duration Q1 and Q2 were significantly and positively associated with incident CHD when compared to Q3 (Q1: HR=1.53, 95% CI: 1.14–2.06; Q2: HR=1.37, 95% CI: 1.01–1.88) (Table 5). The association between sleep duration Q2 and incident CHD was abrogated in model 2 and model 3, whereas the association between sleep duration Q1 and incident CHD was only minimally attenuated in model 2 (HR=1.39, 95% CI: 1.02–1.88) and model 3 (HR=1.37, 95% CI: 1.01–1.86). The addition of proteomic scores of sleep durations Q1

and Q5 (i.e., models mutually adjusted for sleep duration and proteomic scores) abrogated the association between sleep duration Q2 and incident CHD in model 1, whereas sleep duration Q1 remained significantly associated with incident CHD in all models (model 1: HR=1.51, 95% CI: 1.12–2.04; model 2: HR=1.38, 95% CI: 1.02–1.88; model 3: HR=1.37, 95% CI: 1.01–1.86) (Table 6). Proteomic scores for sleep durations Q1 and Q5 were not significantly associated with incident CHD in any model.

Discussion

The key results of this study are first, sleep duration as the single main exposure is associated with both incident DM and incident CHD. Second, 16 unique proteomic markers significantly predict specific sleep duration quintiles when compared to referent sleep duration. Third, proteomic scores, created using the weights of individual proteomic markers, were differentially associated

Table 5 Cox proportional hazards regression models for the association between sleep duration and incident coronary heart disease

| | Model 1 | | Model 2 | | Model 3 | |
|----------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|
| | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Sleep duration Q1 (4.00–6.57 h) | <i>1.53**</i> | (1.14, 2.06) | <i>1.39*</i> | (1.02, 1.88) | <i>1.37*</i> | (1.01, 1.86) |
| Sleep duration Q2 (6.64–7.14 h) | <i>1.37*</i> | (1.01, 1.88) | 1.34 | (0.98, 1.84) | 1.35 | (0.99, 1.85) |
| Sleep duration Q3 (7.21–7.57 h) | Reference | | Reference | | Reference | |
| Sleep duration Q4 (7.64–8.00 h) | 1.21 | (0.89, 1.64) | 1.20 | (0.88, 1.64) | 1.22 | (0.89, 1.66) |
| Sleep duration Q5 (8.14–11.00 h) | 1.24 | (0.86, 1.78) | 1.19 | (0.83, 1.72) | 1.23 | (0.85, 1.77) |

Italic values denote statistically significant results

Model 1 is stratified by HbA1c concentration and adjusted for age and sex

Model 2 is additionally adjusted for cystatin C, education, physical activity, smoking, alcohol consumption, shift work, and insomnia symptoms

Model 3 is additionally adjusted for body mass index, waist circumference, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides

* $p < 0.05$, ** $p < 0.01$

Table 6 Cox proportional hazards regression models for the association between proteomic sleep score and sleep duration with incident coronary heart disease. Analyses are mutually adjusted for proteomic sleep score and sleep duration

| | Model 1 | | Model 2 | | Model 3 | |
|----------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|
| | Coefficient | 95% CI | Coefficient | 95% CI | Coefficient | 95% CI |
| Proteomic score Q1 ^a | 1.13 | (0.93, 1.38) | 1.03 | (0.84, 1.25) | 1.03 | (0.84, 1.26) |
| Proteomic score Q5 ^a | 1.00 | (0.80, 1.24) | 1.02 | (0.82, 1.26) | 1.02 | (0.82, 1.26) |
| Sleep duration Q1 (4.00–6.57 h) | <i>1.51**</i> | (1.12, 2.04) | <i>1.38*</i> | (1.02, 1.88) | <i>1.37*</i> | (1.01, 1.86) |
| Sleep duration Q2 (6.64–7.14 h) | 1.36 | (0.99, 1.86) | 1.34 | (0.98, 1.83) | 1.35 | (0.98, 1.85) |
| Sleep duration Q3 (7.21–7.57 h) | Reference | | Reference | | Reference | |
| Sleep duration Q4 (7.64–8.00 h) | 1.20 | (0.88, 1.63) | 1.20 | (0.88, 1.63) | 1.21 | (0.89, 1.66) |
| Sleep duration Q5 (8.14–11.00 h) | 1.24 | (0.86, 1.78) | 1.19 | (0.82, 1.71) | 1.22 | (0.85, 1.77) |

Italic values denote statistically significant results

Model 1 is stratified by HbA1c concentration and adjusted for age and sex

Model 2 is additionally adjusted for cystatin C, education, physical activity, smoking, alcohol consumption, shift work, and insomnia symptoms

Model 3 is additionally adjusted for body mass index, waist circumference, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides

^a Expressed as the HR and 95% CI of the incremental increase per SD of the proteomic sleep score

* $p < 0.05$, ** $p < 0.01$

with incident DM and incident CHD thereby suggesting independent pathways for the association between sleep duration with specific cardiometabolic diseases. Fourth, the proteomic score for sleep duration Q1 was a significant mediator of the association between sleep duration Q1 and incident DM. The results are discussed in detail below.

Association between sleep duration and cardiometabolic disease

The association between sleep duration and incident DM indicate an approximately 30% increased risk of DM with sleep durations shorter than the referent 7.21–7.57 h and an increased risks of approximately 50% for those with sleep durations 8.14–11.00 h. These results and the J-shaped association are in line with our hypothesis and consistent with several studies published to date, including a large meta-analysis [1]. One major difference, however, when comparing with other studies are the similar risks seen in this study for Q1 (HR: 1.32) and Q2 (HR: 1.33) when compared with Q3. The most probable explanation for this is the decision to, in the present study, use statistically defined cutoffs (i.e., quintiles) rather than discrete hourly cutoffs, in order to get similarly sized groups. This approach may dilute the relative risks in the shortest sleep duration quintile given that the reference category is approximately 26 min short of 8 h which in many studies is considered referent sleep duration. Despite this, the findings confirm the independent and positive association between both short (Q1 and Q2) and long (Q5) sleep durations with incident DM.

The association between sleep duration and incident CHD indicates a 37% increased risk of CHD only for those with the shortest sleep duration (Q1) when compared to Q3. These results are very similar to our findings from a larger sample of the MDC cohort which showed 41% increased risk of CHD for men ($n=6966$) and 46% increased risk for women ($n=9378$) when comparing <6 h of sleep with a referent 7–8 h [32]. Our obtained results for the association between short sleep duration and CHD are also confirmed by meta-analyses [2, 33]. The absence of any association between the longest sleep quintile (Q5) with CHD in the present study could be explained by the decision to not stratify by sex. Indeed, in a larger sample from the same cohort, sleep duration ≥ 9 h was associated with a 33% increased risk of incident CHD among men but not women. Another possible explanation for this discrepancy could be the categorization of sleep duration into quintiles where Q5 represents a relatively broad sleep duration range (8.14–11.0 h). The National Sleep Foundation guidelines recommend a sleep duration of 7–9 h [34].

Proteomic markers predictive of sleep duration

Sixteen unique proteomic markers were found to be significantly associated with sleep duration when comparing four sleep duration quintiles to referent sleep duration (Q3). The combination of proteomic scores for Q1 and Q5 best predicted sleep duration.

The proteomic score for sleep duration Q1 consisted of six unique proteomic markers, however, the scientific literature reveals very little about how these proteomic markers and their corresponding proteins are related to sleep duration. Of the six markers, only two have previous studies indicating a possible association with sleep-related outcomes; follistatin (FS), upregulated in the shortest sleep quintile in our study, has an L-shaped association with sleep duration [9], and concentrations of circulating FS suggest a circadian variation [35]. The function of circulating FS is not known but is suggested to be involved in regulating energy metabolism [36]. E-selectin, a cellular adhesion molecule [37], was upregulated in the shortest sleep quintile which is supported by at least one study in which one night of sleep deprivation was associated with increased concentrations of soluble E-selectin [38]. Moreover, E-selectin concentrations are increased in obstructive sleep apnea [39] and are reduced following treatment with continuous positive airway pressure [39, 40]. Those findings suggest that concentrations of E-selectin may rapidly adjust with changes to sleep dynamics. It should be noted that two studies found no significant association between sleep duration and E-selectin [41, 42]. For the remaining four markers that were associated with the shortest sleep duration quintile in this study, the literature, to the best of our knowledge, reveals no findings linked to sleep duration or sleep-related outcomes. In the present study, tumor necrosis factor-related activation-induced cytokine (TRANCE), also known as receptor activator of nuclear factor- κ B ligand (RANKL), as well as matrix metalloproteinase (MMP)-10 were downregulated, whereas MMP-7 and tumor necrosis factor-related apoptosis-inducing ligand receptor 2 (TRAIL-R2) were upregulated with the shortest sleep duration quintile. RANKL has been implicated in vascular inflammation, vascular calcification, and angiogenesis [43]; MMP-10 is involved in the resolution of acute inflammation [44]; MMP-7 has been linked with both proapoptotic and cell proliferative pathways through cleavage of the ligand FasL and its involvement in the Fas/FasL pathway [45]; TRAIL-R2 triggers caspase-dependent apoptosis [46] and soluble TRAIL-R2 is released with apoptosis activated through the Fas/FasL pathway [47]. In summary, very short sleep duration is linked with upregulation of pro-inflammatory and pro-apoptotic markers and of markers related to

cell adhesion and with downregulation of markers related to angiogenesis and resolution of inflammation.

The longest sleep duration quintile was positively associated with prolactin, MMP-7, C-X-C motif chemokine 1 (CXCL1), tissue-type plasminogen activator (t-PA), and inversely associated with tumor necrosis factor receptor superfamily member 6 (Fas), and heat shock 27 kDa protein (HSP 27). There is only scarce information in the existing literature linking these specific markers with sleep-related outcomes. Prolactin is involved in lactation, reproduction, immune response, and angiogenesis [48], and prolactin levels have been shown to increase during sleep [49–51], which is in line with our findings. tPA is involved in endogenous fibrinolysis [52]. Acute release of tPA in men is reduced in habitually short sleepers (<7.0 h) compared to those who sleep between 7.0 h and 8.1 h [53]. Moreover, tPa has been found to be positively associated with moderate to severe obstructive sleep apnea (OSA) [39, 54] and with desaturation events during sleep [54]. These findings are also in line with our finding of tPA upregulation with the longest sleep quintile and may suggest a link between long self-reported sleep duration and OSA. Fas is a cell-surface receptor involved in caspase-dependent apoptosis [45, 55]. Although no studies link plasma concentrations of Fas with sleep outcomes, it is noteworthy that plasma concentrations of caspase-8 are themselves positively associated with short sleep duration [8] which would be in line with the downregulation of Fas with long sleep as found in the present study. CXCL1 is a chemoattractant cytokine that increases neutrophil recruitment in inflammation and promotes angiogenesis [56]. HSP27 functions as an antioxidant and an anti-apoptotic agent during oxidative and chemical stress, respectively [57]. Taken together, the longest sleep duration is linked with upregulation of markers related to angiogenesis, cell proliferation and fibrinolysis, and downregulation of markers related to apoptosis. The upregulation of MMP-7 with both short and long sleep duration could be due to its suggested involvement in both cell proliferative and proapoptotic pathways [45]. Based on the suggested function of the remaining proteins, it could be speculated that the role of MMP-7 in long sleep is related to cell proliferation.

Association between proteomic scores and cardiometabolic disease

The two proteomic scores for sleep durations Q1 and Q5 together best predicted habitual sleep duration in a multilinear regression model. This result is in line with expectations given that the two scores account for the extreme sleep duration quintiles. Consequently, the two quintiles account for the largest proportion of variance explained while keeping the number of predictors in the

model to a minimum. The inclusion of the two proteomic scores in the models for sleep duration and incident DM abrogated the association for sleep durations Q1 and Q2 in the final multivariable model. The proteomic score for Q1 remained significantly associated with incident DM in all models albeit with some attenuation following the inclusion of additional covariates in models 2 and 3. This would be expected given that several covariates are known to be associated with DM; there is therefore little reason to believe that they would not be associated also with proteomic markers of inflammation. Indeed, of the six markers included in the proteomic score for sleep duration Q1, FS [58], E-selectin [59], MMP-7 [60], and TRAIL-R2 [47] have been shown to be associated with the risk of DM, while RANKL is inversely associated with prevalent DM [61].

Mediation analyses revealed that the proteomic score for sleep duration Q1 was a significant mediator of approximately 40% of the association between sleep duration Q1, when compared to Q3, and incident DM. It is noteworthy that mediation was significant after 11 years of follow-up; however, this could be explained by the comparatively small number of cases of incident DM ($n=121$) during follow-up years 3.0–11.0 compared to the 401 cases during follow-up years 11.1–27.2. These findings strongly suggest that the proteomic score for sleep duration Q1 provides an explanation of the mechanism underlying the association between short sleep duration and incident DM and as such could be considered a proteomic fingerprint of very short sleep duration. Conversely, the proteomic score for sleep duration Q5 was only marginally attenuated by the inclusion of covariates but was itself not significantly associated with incident DM in the final model. Moreover, the corresponding sleep duration Q5 phenotype remained, with minimal attenuation, significantly associated with incident DM throughout all models. The findings thus highlight that although specific proteomic markers are predictive of the longest sleep duration, their combination does not explain the mechanism through which the longest sleep is associated with incident DM. Instead, the stable association of the longest sleep phenotype with incident DM therefore points towards residual confounding.

For the association between sleep duration and incident CHD, the inclusion of proteomic scores for sleep durations Q1 and Q5 had no effect on the association between sleep duration and incident CHD. Moreover, the proteomic markers did not attenuate the strength of the association between sleep duration Q1 and incident CHD. More surprisingly, neither of the proteomic scores were associated with incident CHD in any of the models. This is an unexpected finding that contradicts our initial hypothesis. All proteomic markers were derived

from a CVD panel with the expectation that their combination would be predominantly associated with a CVD outcome. Contrary to being a limitation of the study, this result could instead be used to argue that the specific combination of markers and their respective weights is explanatory only of incident DM. Consequently, the observed association between sleep duration Q1 and incident CHD suggests that incident DM itself may be the intermediate explanatory stage. Indeed, when considering incident DM as a time-varying covariate in the present study, the association between sleep duration Q1 and incident CHD is abrogated (results not shown). This suggests an overall pathway where very short sleep duration, through inflammatory and apoptotic processes, is associated with incident DM which in turn is an intermediate step explaining the downstream association with incident CHD. The latter part confirms the results of a recent study using the same study cohort [32].

Limitations

This study has some limitations. First, we have used only one Olink panel whereas another study [9] used multiple panels. Moreover, today's proteomic assays allow for the quantification of several thousand proteomic markers; it is important that studies with available large-scale proteomic data aim to replicate the findings. Nevertheless, the CVD I panel used in the present study is highly relevant to cardiometabolic disease and inflammation. Second, the found association between long sleep duration and incident DM is most likely driven by other factors than inflammation. Alternatively, the association could be explained by inflammatory and cardiometabolic markers other than the ones investigated in the study. Third, there is possible circadian variation of at least two proteomic markers (FS and E-selectin) [38] indicating that timing of the blood sampling may be a relevant factor. However, all samples were taken during morning (7–11 AM) hours and despite differences in timing during that period, we would expect that this is a non-differential misclassification which drives results towards the null. Fourth, the current study does not allow for any conclusion regarding the specific tissues that may be involved in the associations between sleep duration and circulating proteomic markers. Fifth, we were not able to include OSA as a covariate. OSA is associated with several of the found markers [39, 40, 54] and could potentially be associated with additional markers. However, the role of OSA would appear to be more relevant for long sleep duration. Sixth, this study did not consider hypnotics or psychotropic drugs that may well impact the exposures and outcomes of this study. However, we adjusted for symptoms of insomnia which is prevalent in many with affective disorder and may serve as a proxy for the use

of hypnotics. Seventh, due to sample size, sleep duration was categorized into quintiles to allow for relevant statistical analyses. However, larger studies with information on both sleep duration and proteomic markers would do well to also consider discrete hourly sleep duration categories. Finally, several markers (e.g., CXCL1 and MMP-7) are involved in multiple biological pathways; plasma concentrations do not elucidate the specific pathways that are involved. However, the combination of markers allows for educated speculation as to which pathways that may be relevant and is in itself a strength of the present study's approach.

Strengths

This study also has several strengths. First, lasso as a selection procedure allows for the inclusion of all relevant proteomic markers while avoiding multicollinearity. Indeed, it is expected that many of the included proteins represent similar or identical biological pathways and would therefore be concomitantly up- and down-regulated with any specific exposure of interest. Moreover, lasso iterates over several subsets of data before estimating coefficients based on average values of the selected markers. Second, the combination of individual proteomic markers into proteomic scores has the benefit of offering a more comprehensive view of any possible underlying mechanisms of the association between sleep duration and cardiometabolic outcomes. Third, the analyses excluded the first 3 years of follow-up thereby minimizing the risk of reverse causation; indeed, any illness or disease process related to inflammation would be expected to have a direct impact on sleep duration. Fourth, participants in the study were followed for an average of approximately 22 years. Fifth, both incident DM and incident CHD were collected from highly accurate nationwide registers.

Conclusions

This study found differential proteomic marker expressions for the very shortest and longest sleep duration quintiles in a general population cohort. The proteomic score of the shortest sleep duration was positively associated with incident DM but not with incident CHD. It is therefore suggested that the proinflammatory and proapoptotic pathways identify a biological mechanism that mediates and links very short sleep duration specifically to incident DM. No associations were found for the proteomic score of the longest sleep duration with incident DM or incident CHD thereby suggesting residual confounding as a possible explanation.

Abbreviations

| | |
|-----|------------------------|
| DM | Diabetes mellitus |
| CHD | Coronary heart disease |

| | |
|----------|--|
| CRP | C-reactive protein |
| IL-6 | Interleukin 6 |
| MDC | Malmö Diet and Cancer |
| MDC-CC | Malmö Diet and Cancer Cardiovascular Cohort |
| CVD | Cardiovascular disease |
| PEA | Proximity extension assay |
| CABG | Coronary artery bypass graft |
| PCI | Percutaneous coronary intervention |
| MI | Myocardial infarction |
| LDL-C | Low-density lipoprotein cholesterol |
| HDL-C | High-density lipoprotein cholesterol |
| TG | Triglycerides |
| HbA1c | Hemoglobin A1c |
| AIC | Akaike Information Criterion |
| HR | Hazard ratios |
| CI | Confidence intervals |
| FS | Follistatin |
| TRANCE | Tumor necrosis factor-related activation-induced cytokine |
| RANKL | Receptor activator of nuclear factor- κ B ligand |
| MMP-10 | Matrix metalloproteinase-10 |
| MMP-7 | Matrix metalloproteinase-7 |
| TRAIL-R2 | Tumor necrosis factor-related apoptosis-inducing ligand receptor 2 |
| CXCL1 | C-X-C motif chemokine 1 |
| t-PA | Tissue-type plasminogen activator |
| HSP 27 | Heat shock 27 kDa protein |
| OSA | Obstructive sleep apnea |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-024-03392-1>.

Additional file 1: Table S1 Proteomic markers included in the Olink Proseek Multiplex CVD 1 panel. **Table S2** Proteomic markers and their association with specific sleep duration quintiles.

Authors' contributions

TS and OM were responsible for the conception and design of the study. TS, AKS, MK, GE, JN, MOM, PN, and OM analyzed and interpreted the data. TS drafted the manuscript. AKS, MK, GE, JN, MOM, PN, and OM critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript. TS is the guarantor of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. TS had final responsibility for the decision to submit for publication. All authors read and approved the final manuscript.

Authors' Twitter handles

Thomas Svensson @precision_sleep.

Funding

Open access funding provided by Lund University. The Malmö Diet and Cancer Study was supported by Lund University Infrastructure grant "Malmö population-based cohorts" (STYR 2019/2046). O.M. was supported by the European Research Council (ADG-282255), the Swedish Heart and Lung Foundation, Swedish Research Council, Novo Nordisk Foundation, Knut and Alice Wallenberg Foundation, Göran Gustafsson Foundation, Skåne University Hospital donation funds, Medical Faculty at Lund University, governmental funding of clinical research within the national health services, Albert Pålsson Research Foundation, and Region Skåne. T.S. was supported by Ernhold Lundströms Stiftelse, and Esther Olssons stiftelse II and Anna Jönssons Minnesfond. The funders had no role in study design, in the collection, analysis and interpretation of data, in writing of the article, or in the decision to submit the article.

Availability of data and materials

Due to ethical and legal restrictions related to the Swedish Biobanks in Medical Care Act (2002:297) and the Personal Data Act (1998:204), data are available

upon request from the data access group of Malmö Diet and Cancer study by contacting Anders Dahlin (anders.dahlin@med.lu.se).

Declarations

Ethics approval and consent to participate

The MDC study was approved by the ethics committee at Lund University (LU 51–90 and Dnr 2009/633), and all participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; and no other relationships.

Author details

¹Department of Clinical Sciences, Lund University, Skåne University Hospital, CRC, Jan Waldenströms Gata 35, 20502 Malmö, Sweden. ²Precision Health, Department of Bioengineering, Graduate School of Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-8655, Japan. ³Graduate School of Health Innovation, Kanagawa University of Human Services, Kawasaki-Ku, Kawasaki-Shi, Kanagawa, Japan. ⁴Department of Diabetes and Metabolic Diseases, the University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan. ⁵Department of Cardiology, Skåne University Hospital, Malmö, Sweden. ⁶Department of Internal Medicine, Skåne University Hospital, Malmö, Sweden.

Received: 28 November 2023 Accepted: 15 April 2024

Published online: 23 April 2024

References

- Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2010;33(2):414–20.
- Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*. 2011;32(12):1484–92.
- Ferrie JE, Kivimaki M, Akbaraly TN, Singh-Manoux A, Miller MA, Gimeno D, Kumari M, Davey Smith G, Shipley MJ. Associations between change in sleep duration and inflammation: findings on C-reactive protein and interleukin 6 in the Whitehall II Study. *Am J Epidemiol*. 2013;178(6):956–61.
- Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25–33.
- Irwin MR, Olmstead R, Carroll JE. Sleep disturbance, sleep duration, and inflammation: a systematic review and meta-analysis of cohort studies and experimental sleep deprivation. *Biol Psychiatry*. 2016;80(1):40–52.
- Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med*. 2004;350(14):1387–97.
- Danesh J, Kaptoge S, Mann AG, Sarwar N, Wood A, Angleman SB, Wensley F, Higgins JP, Lennon L, Eiriksdottir G, et al. Long-term interleukin-6 levels and subsequent risk of coronary heart disease: two new prospective studies and a systematic review. *PLoS Med*. 2008;5(4): e78.
- Svensson T, Svensson AK, Kitlinski M, Almgren P, Engstrom G, Nilsson J, Orho-Melander M, Nilsson PM, Melander O. Plasma concentration of Caspase-8 is associated with short sleep duration and the risk of incident diabetes mellitus. *J Clin Endocrinol Metab*. 2018;103(4):1592–600.

9. Theorell-Haglow J, Hammar U, Lind L, Elmstahl S, Lindberg E, Fall T. Sleep duration is associated with protein biomarkers for cardiometabolic health: a large-scale population study. *J Sleep Res.* 2021;30(5):e13284.
10. Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo diet and cancer study design and feasibility. *J Intern Med.* 1993;233(1):45–51.
11. Rosvall M, Janzon L, Berglund G, Engstrom G, Hedblad B. Incidence of stroke is related to carotid IMT even in the absence of plaque. *Atherosclerosis.* 2005;179(2):325–31.
12. Enhorning S, Sjogren M, Hedblad B, Nilsson PM, Struck J, Melander O. Genetic vasopressin 1b receptor variance in overweight and diabetes mellitus. *Eur J Endocrinol.* 2016;174(1):69–75.
13. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* 2011;39(15):e102.
14. Assarsson E, Lundberg M, Holmquist G, Bjorkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE.* 2014;9(4):e95192.
15. Shan Z, Ma H, Xie M, Yan P, Guo Y, Bao W, Rong Y, Jackson CL, Hu FB, Liu L. Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies. *Diabetes Care.* 2015;38(3):529–37.
16. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh CD. Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol.* 2001;17(11):983–9.
17. Cederholm J, Eeg-Olofsson K, Eliasson B, Zethelius B, Nilsson PM, Gudbjornsdottir S. Swedish National Diabetes R: Risk prediction of cardiovascular disease in type 2 diabetes: a risk equation from the Swedish National Diabetes Register. *Diabetes Care.* 2008;31(10):2038–43.
18. Ludvigsson JF, Andersson E, Ekborn A, Feychting M, Kim JL, Reuterwall C, Heurgren M, Olausson PO. External review and validation of the Swedish national inpatient register. *BMC Public Health.* 2011;11:450.
19. Centre for Epidemiology. National Board of Health and Welfare: A finger on the pulse monitoring public health and social conditions in Sweden 1992–2002. National Board of Health and Welfare: In. Stockholm; 2003.
20. Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, Persson I, Sundstrom A, Westerholm B, Rosen M. The new Swedish Prescribed Drug Register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf.* 2007;16(7):726–35.
21. Jujic A, Nilsson PM, Persson M, Holst JJ, Torekov SS, Lyssenko V, Groop L, Melander O, Magnusson M. Atrial natriuretic peptide in the high normal range is associated with lower prevalence of insulin resistance. *J Clin Endocrinol Metab.* 2016;101(4):1372–80.
22. Magnusson M, Wang TJ, Clish C, Engstrom G, Nilsson P, Gerszten RE, Melander O. Dimethylglycine deficiency and the development of diabetes. *Diabetes.* 2015;64(8):3010–6.
23. Enhorning S, Wang TJ, Nilsson PM, Almgren P, Hedblad B, Berglund G, Struck J, Morgenthaler NG, Bergmann A, Lindholm E, et al. Plasma copeptin and the risk of diabetes mellitus. *Circulation.* 2010;121(19):2102–8.
24. Lagerqvist B, James SK, Stenestrand U, Lindback J, Nilsson T, Wallentin L, Group SS. Long-term outcomes with drug-eluting stents versus bare-metal stents in Sweden. *N Engl J Med.* 2007;356(10):1009–19.
25. National Board of Health and Welfare. valuation of diagnostic quality for acute myocardial infarction in the Swedish Inpatient Registry 1987 and 1995 [in Swedish]. Stockholm, Sweden: Epidemiologiskt Centrum, Socialstyrelsen; 2000.
26. Enhorning S, Struck J, Wirfalt E, Hedblad B, Morgenthaler NG, Melander O. Plasma copeptin, a unifying factor behind the metabolic syndrome. *J Clin Endocrinol Metab.* 2011;96(7):E1065–1072.
27. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol).* 1996;58(1):267–88.
28. Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: the lasso and generalizations. 1st ed. New York: Chapman and Hall/CRC; 2015.
29. Lasso inference intro - introduction to inferential lasso models <https://www.stata.com/manuals/lassolasso inference intro.pdf>
30. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. *Economet J.* 2018;21(1):C1–68.
31. Drukker D, Liu D. Using the lasso for inference in high-dimensional models. In, vol. 2022. Stata.com: StataCorp LLC; 2019. <https://blog.stata.com/2019/09/09/using-the-lasso-for-inference-in-high-dimensional-models/>. Accessed 6 June 2022.
32. Svensson AK, Svensson T, Kitlinski M, Almgren P, Engstrom G, Nilsson PM, Melander O. Incident diabetes mellitus may explain the association between sleep duration and incident coronary heart disease. *Diabetologia.* 2018;61(2):331–41.
33. Wang D, Li W, Cui X, Meng Y, Zhou M, Xiao L, Ma J, Yi G, Chen W. Sleep duration and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *Int J Cardiol.* 2016;219:231–9.
34. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, Hazen N, Herman J, Katz ES, Kheirandish-Gozal L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health.* 2015;1(1):40–3.
35. Anastasilakis AD, Polyzos SA, Skouvaklidou EC, Kynigopoulos G, Saridakis ZG, Apostolou A, Triantafyllou GA, Karagiozoglou-Lampoudi T, Mantzoros CS. Circulating follistatin displays a day-night rhythm and is associated with muscle mass and circulating leptin levels in healthy, young humans. *Metabolism.* 2016;65(10):1459–65.
36. Hansen JS, Plomgaard P. Circulating follistatin in relation to energy metabolism. *Mol Cell Endocrinol.* 2016;433:87–93.
37. Bevilacqua MP, Nelson RM. Selectins. *J Clin Invest.* 1993;91(2):379–87.
38. Frey DJ, Fleshner M, Wright KP Jr. The effects of 40 hours of total sleep deprivation on inflammatory markers in healthy young adults. *Brain Behav Immun.* 2007;21(8):1050–7.
39. Cederberg KLJ, Hanif U, PerisSempere V, Hedou J, Leary EB, Schneider LD, Lin L, Zhang J, Morse AM, Blackman A, et al. Proteomic biomarkers of the apnea hypopnea index and obstructive sleep apnea: insights into the pathophysiology of presence, severity, and treatment response. *Int J Mol Sci.* 2022;23(14):7983.
40. Chin K, Nakamura T, Shimizu K, Mishima M, Nakamura T, Miyasaka M, Ohi M. Effects of nasal continuous positive airway pressure on soluble cell adhesion molecules in patients with obstructive sleep apnea syndrome. *Am J Med.* 2000;109(7):562–7.
41. Dowd JB, Goldman N, Weinstein M. Sleep duration, sleep quality, and biomarkers of inflammation in a Taiwanese population. *Ann Epidemiol.* 2011;21(11):799–806.
42. Martinez-Gomez D, Eisenmann JC, Gomez-Martinez S, Hill EE, Zapatera B, Veiga OL, Marcos A, Group AS. Sleep duration and emerging cardiometabolic risk markers in adolescents The AFINOS study. *Sleep Med.* 2011;12(10):997–1002.
43. Min JK, Kim YM, Kim SW, Kwon MC, Kong YY, Hwang IK, Won MH, Rho J, Kwon YG. TNF-related activation-induced cytokine enhances leukocyte adhesiveness: induction of ICAM-1 and VCAM-1 via TNF receptor-associated factor and protein kinase C-dependent NF-kappaB activation in endothelial cells. *J Immunol.* 2005;175(1):531–40.
44. Fingleton B. Matrix metalloproteinases as regulators of inflammatory processes. *Biochim Biophys Acta Mol Cell Res.* 2017;1864(11 Pt A):2036–42.
45. Yamada A, Arakaki R, Saito M, Kudo Y, Ishimaru N. Dual role of Fas/ FasL-mediated signal in peripheral immune tolerance. *Front Immunol.* 2017;8:403.
46. Walczak H, Degli-Esposti MA, Johnson RS, Smolak PJ, Waugh JY, Boiani N, Timour MS, Gerhart MJ, Schooley KA, Smith CA, et al. TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL. *EMBO J.* 1997;16(17):5386–97.
47. Mattisson IY, Bjorkbacka H, Wigren M, Edseldt A, Melander O, Fredriksson GN, Bengtsson E, Goncalves I, Orho-Melander M, Engstrom G, et al. Elevated markers of death receptor-activated apoptosis are associated with increased risk for development of diabetes and cardiovascular disease. *EBioMedicine.* 2017;26:187–97.
48. Freeman ME, Kanyicska B, Lerant A, Nagy G. Prolactin: structure, function, and regulation of secretion. *Physiol Rev.* 2000;80(4):1523–631.
49. Sassin JF, Frantz AG, Kapen S, Weitzman ED. The nocturnal rise of human prolactin is dependent on sleep. *J Clin Endocrinol Metab.* 1973;37(3):436–40.
50. Sassin JF, Frantz AG, Weitzman ED, Kapen S. Human prolactin: 24-hour pattern with increased release during sleep. *Science.* 1972;177(4055):1205–7.
51. Parker DC, Rossman LG, Vander Laan EF. Sleep-related, nychthermeral and briefly episodic variation in human plasma prolactin concentrations. *J Clin Endocrinol Metab.* 1973;36(6):1119–24.

52. Oliver JJ, Webb DJ, Newby DE. Stimulated tissue plasminogen activator release as a marker of endothelial function in humans. *Arterioscler Thromb Vasc Biol.* 2005;25(12):2470–9.
53. Weil BR, Greiner JJ, Stauffer BL, Desouza CA. Self-reported habitual short sleep duration is associated with endothelial fibrinolytic dysfunction in men: a preliminary report. *Sleep.* 2013;36(2):183–8.
54. Ambati A, Ju YE, Lin L, Olesen AN, Koch H, Hedou JJ, Leary EB, Sempere VP, Mignot E, Taheri S. Proteomic biomarkers of sleep apnea. *Sleep.* 2020;43(11):zsaa086.
55. Waring P, Mullbacher A. Cell death induced by the Fas/Fas ligand pathway and its role in pathology. *Immunol Cell Biol.* 1999;77(4):312–7.
56. Korbecki J, Szatkowska I, Kupnicka P, Zwierello W, Barczak K, Poziomkowska-Gesicka I, Wojcik J, Chlubek D, Baranowska-Bosiacka I. The importance of CXCL1 in the physiological state and in noncancer diseases of the oral cavity and abdominal organs. *Int J Mol Sci.* 2022;23(13):7151.
57. Vidyasagar A, Wilson NA, Djamali A. Heat shock protein 27 (HSP27): biomarker of disease and therapeutic target. *Fibrogenesis Tissue Repair.* 2012;5(1):7.
58. Wu C, Borne Y, Gao R, Lopez Rodriguez M, Roell WC, Wilson JM, Regmi A, Luan C, Aly DM, Peter A, et al. Elevated circulating follistatin associates with an increased risk of type 2 diabetes. *Nat Commun.* 2021;12(1):6486.
59. Qiu S, Cai X, Liu J, Yang B, Zugel M, Steinacker JM, Sun Z, Schumann U. Association between circulating cell adhesion molecules and risk of type 2 diabetes: a meta-analysis. *Atherosclerosis.* 2019;287:147–54.
60. Bao X, Xu B, Yin S, Pan J, Nilsson PM, Nilsson J, Melander O, Orho-Melander M, Engstrom G. Proteomic profiles of body mass index and waist-to-hip ratio and their role in incidence of diabetes. *J Clin Endocrinol Metab.* 2022;107(7):e2982–90.
61. Lieb W, Gona P, Larson MG, Massaro JM, Lipinska I, Keaney JF Jr, Rong J, Corey D, Hoffmann U, Fox CS, et al. Biomarkers of the osteoprotegerin pathway: clinical correlates, subclinical disease, incident cardiovascular disease, and mortality. *Arterioscler Thromb Vasc Biol.* 2010;30(9):1849–54.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.