

Phenome-wide associations of sleep characteristics in the Human Phenotype Project

Received: 10 March 2024

Accepted: 18 December 2024

Published online: 27 January 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Sleep tests commonly diagnose sleep disorders, but the diverse sleep-related biomarkers recorded by such tests can also provide broader health insights. In this study, we leveraged the uniquely comprehensive data from the Human Phenotype Project cohort, which includes 448 sleep characteristics collected from 16,812 nights of home sleep apnea test monitoring in 6,366 adults (3,043 male and 3,323 female participants), to study associations between sleep traits and body characteristics across 16 body systems. In this analysis, which identified thousands of significant associations, visceral adipose tissue (VAT) was the body characteristic that was most strongly correlated with the peripheral apnea–hypopnea index, as adjusted by sex, age and body mass index (BMI). Moreover, using sleep characteristics, we could predict over 15% of body characteristics, spanning 15 of the 16 body systems, in a held-out set of individuals. Notably, sleep characteristics contributed more to the prediction of certain insulin resistance, blood lipids (such as triglycerides) and cardiovascular measurements than to the characteristics of other body systems. This contribution was independent of VAT, as sleep characteristics outperformed age, BMI and VAT as predictors for these measurements in both male and female participants. Gut microbiome-related pathways and diet (especially for female participants) were notably predictive of clinical obstructive sleep apnea symptoms, particularly sleepiness, surpassing the prediction power of age, BMI and VAT on these symptoms. Together, lifestyle factors contributed to the prediction of over 50% of the sleep characteristics. This work lays the groundwork for exploring the associations of sleep traits with body characteristics and developing predictive models based on sleep monitoring.

Sophisticated sleep monitoring methods have been developed in the past decades to evaluate sleep disorders. Starting with in-laboratory polysomnography and, more recently, home sleep apnea testing (HSAT), these sleep tests employ wearable sensors to monitor the patient's respiration, heart, oxygen saturation dynamics and sometimes more signals continuously throughout the night. These methods enable assessment of sleep architecture and overall sleep health, often allowing determination of the severity of obstructive

sleep apnea (OSA), a widespread health issue with an estimated global prevalence of almost 1 billion adults affected¹ and associated with various comorbidities, such as a 2–3-fold increased risk of cardiovascular diseases and metabolic disease². Although sleep data have been and still are extensively investigated in the context of sleep disorders^{3,4}, we think that they are a valuable source of physiological signals recordings to provide insights into broader aspects of human health.

Despite the availability of numerous databases from multiple cohorts in the National Sleep Research Resource⁵, these include, at most, polysomnography, actigraphy and questionnaire-based sleep data. However understanding the intricate interplay between sleep and overall well-being requires a comprehensive examination, necessitating large-scale, deeply phenotyped cohorts. A few phenotype-wide association studies (PheWASs) have used either self-reported⁶ or limited objective measurements^{7–9} of sleep to evaluate the relationship between sleep duration and health outcomes. However, these studies lack the inclusion of a wide variety of high-resolution objective sleep measurements.

In the present work, we leveraged the sleep data from the Human Phenotype Project (HPP)¹⁰—in total, 16,812 nights of HSAT monitoring from 6,366 individuals (2.42 ± 0.70 nights per individual per research stage)—along with various modalities to explore associations of sleep with 16 other body systems. We analyzed 100 objective physiological features extracted from HSAT, which included respiratory, snoring, sleep position and stage measurements, as well as 348 nocturnal pulse rate variability (PRV) metrics derived from the resting signals (collected in the HSAT device). PRV, as a surrogate of heart rate variability, has been shown to assess the overall health and well-being of an individual¹¹, therefore providing potential competing or complementary insights into human health.

Acknowledging the well-documented night-to-night variability in sleep metrics¹², our approach emphasizes the use of aggregated values across multiple nights to enhance reliability. Although some research has focused on the association of sleep disorders with specific traits independently, such as obesity¹³, atrial fibrillation^{14,15} or type 2 diabetes mellitus¹⁶, a holistic view of all body systems associated with sleep has not yet been investigated. Thus, our aim was to map all the phenotypes associated with sleep characteristics and rank them by importance across diverse body systems. Additionally, we investigated the ability of sleep features to predict medical conditions beyond sleep or cardiovascular disorders.

Results

Cohort characteristics

To perform this work, we analyzed data from the ongoing HPP¹⁰, a large, deeply phenotyped cohort, which, at the time of this study, included only Israelis, most of whom were healthy, of European (Ashkenazi) Jewish descent and had relatively high levels of education compared to the general population (Methods).

This cohort includes various clinical, behavioral, physiological and multi-omics profiling that can be regrouped into 17 body systems: sleep and 16 others (Fig. 1a). Figure 1b depicts the 17 body systems considered in this study, showcasing their features projected onto a 2D space to illustrate how close or far apart they are from each other and how they group together according to the body system types (Methods). The cluster map reveals, for instance, the close proximity between sleep characteristics and the cardiovascular system—an association that has been extensively explored in previous research¹⁷.

We focused on the multi-night sleep metrics of 6,748 participants to investigate their associations with other body systems data measured within a period of ± 6 months (Methods and Fig. 1c). After exclusion of invalid recordings and participants who did not meet the project inclusion criteria, 6,940 data points, from 6,366 participants and 16,812 nights, were included (Extended Data Fig. 1) in this work (see Extended Data Table 1 for sample sizes of all body systems). From the total individuals included, 47.8% (3,043/6,366) were male and 52.2% (3,323/6,366) were female, with an average age of 52.4 ± 7.7 years (mean \pm s.d.) and mean BMI of $26.1 \pm 4.1 \text{ kg m}^{-2}$.

Reference values and trends of the sleep-derived features

Sleep analysis consisted of monitoring peripheral blood oxygen saturation levels, pulse rates, respiratory events, snoring levels, sleep position

and stages using the WatchPAT 300 (ZOLL Itamar) home sleep test device, as well as post-processing calculations¹⁸ (Methods), in three nights within a 2-week time period. To assess the distribution of these features among our dataset, we calculated their mean and s.d. across all data points (Methods); results for some key features are shown in Table 1.

We examined the progression of the sleep-derived features with age (see details in Methods) on male and female participants separately. Figure 2 shows the age–gender reference plots for peripheral apnea–hypopnea index (pAHI), mean nadir oxygen saturation during sleep desaturations (mean nadir SpO₂) and the percentage of light and deep sleep. In this analysis, we used the pAHI, a peripheral measurement since assessed without airflow sensor, as a surrogate of AHI and as an indicator of sleep apnea severity¹⁹. The results illustrate that, when using multi-night monitoring via HSAT, pAHI increases considerably with each additional year of age, aligning with previous work²⁰. Although the results show a more important change with age in pAHI for female participants, we can see that 10% of the female population reached a pAHI corresponding to moderate to severe OSA around the age of 55 years, whereas 10% of the male population reached that threshold already at the age of 40 years. Similarly, the mean nadir SpO₂ decreases per additional year of age in both sexes. Furthermore, changes in sleep architecture can be observed in both sexes with age progression, characterized by a reduction of $0.13 \pm 0.01\%$ (mean \pm s.d.) and $0.14 \pm 0.01\%$ per year in the percentage of deep sleep and an increase of $0.18 \pm 0.02\%$ and $0.23 \pm 0.02\%$ per year in the percentage of light sleep for female and male participants, respectively. For both pAHI and mean nadir SpO₂, an interesting difference between sexes emerges: whereas the progression of these sleep measurements in male participants is almost linear across the entire age scale, female participants exhibit a change in progression around the age of 55 years, likely reflecting the menopausal status change at that age. To evaluate this hypothesis, we subsampled our initial cohort of female participants, creating two subgroups ($n = 173$ each): one for premenopausal female participants and one for postmenopausal female participants, matched by age. As shown in Extended Data Fig. 2, we observed that the pAHI measurements in the premenopausal group increase slightly more rapidly with age compared to the postmenopausal group, reaching similar pAHI levels after the age of 50 years. However, under the age of 50 years, the premenopausal group shows lower pAHI measurements compared to the postmenopausal group of the same age. This suggests that the previously observed nonlinearity is not related to age but, rather, to menopausal status.

Within-individuals longitudinal analysis

In total, 574 individuals in our cohort underwent two sleep monitoring series (Extended Data Fig. 1). We used this subset to analyze the progression, within individuals, of key sleep-derived features presented in Table 1 over a 2-year period. For this longitudinal analysis, we calculated the percentage of change in sleep-derived measurements from baseline within individuals, as described in the Methods section. This analysis revealed very high consistency over time for features related to oxygen saturation and snoring levels, moderate consistency for sleep architecture and larger variations for features such as pAHI and PRV (Extended Data Fig. 3). Furthermore, we observed that, within individuals, there was a gain in light sleep ($+1.01 \pm 13.65\%$ median \pm s.d., $P = 0.001$) and a loss in deep sleep ($-1.14 \pm 31.64\%$ median \pm s.d., $P = 0.02$) over the 2-year period in all individuals, consistent with the age-related changes observed in the overall population.

Pairwise correlations of sleep-derived features to other body systems features

To identify specific feature–feature associations, we explored the pairwise correlations between key sleep-derived features and features from all other body systems. We selected key sleep-derived features

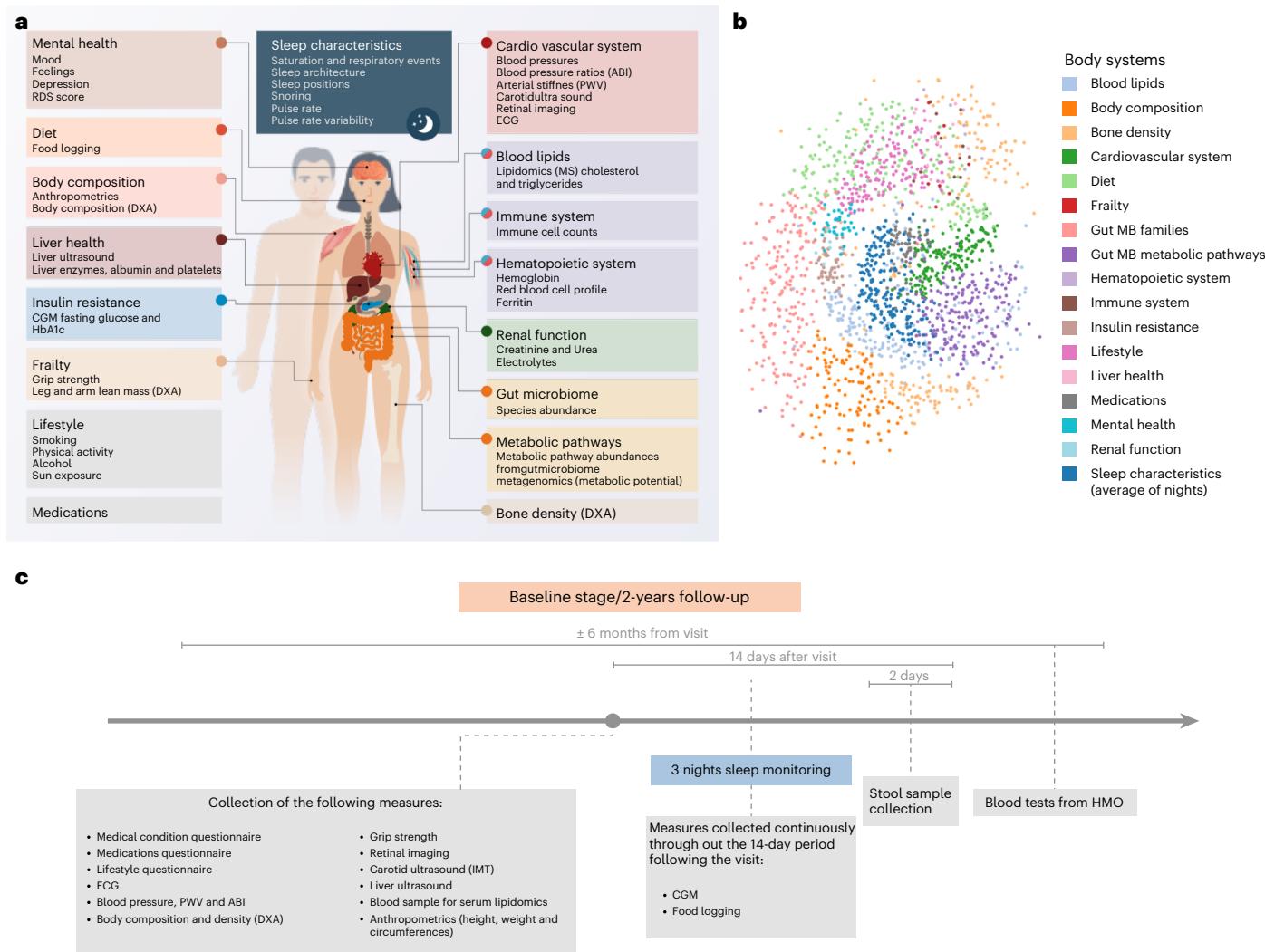


Fig. 1 | Illustration of the HPP study data used in this work. **a**, Sleep characteristics were extracted from multi-night monitoring of sleep and tested for associations with age, BMI and 16 other body system-level categories from comprehensive clinical, physiological, behavioral and multi-omic profiling data collected in the HPP study. **b**, 2D mapping of the body system features included in the HPP cohort performed using uniform manifold approximation and projection (UMAP). Body systems were color coded for better visualization of body system similarity and clustering. **c**, Timing of sleep monitoring with respect

to all other phenotypes. The body system characteristics were measured within a period of ±6 months from the visit, with the sleep monitoring performed in three nights within a 2-week time period after that visit. ABI, ankle–brachial index; CGM, continuous glucose monitoring; DXA, dual-energy X-ray absorptiometry; ECG, electrocardiogram; HMO, health maintenance organization; IMT, intima-media thickness; MB, microbiome; MS, mass spectrometry; PWV, pulse wave velocity; RDS, recent depressive symptoms.

previously associated with certain phenotypes to provide a holistic view and quantitative ranking among a wide range of body characteristics. These sleep-derived features were chosen to represent each type of variable collected using HSAT: pAHI for respiratory events; mean snoring level for continuous snoring monitoring; mean oxygen saturation nadir for continuous pulse oximetry monitoring; sleep time as an objective measurement (rather than self-reported); sleep efficiency calculated from sleep staging; and PRV during the night as a potential measurement of autonomic nervous system activity¹¹. As shown previously, sex and age have a wide effect on sleep characteristics, and on pAHI specifically, but, also, BMI is known to be significantly associated with pAHI^{13,21}. Therefore, to explore the relationship with other body systems, we used age-, sex- and BMI-adjusted Spearman correlations (Methods). In total, 12,404 out of 25,164 correlations tested were found to be significant for the six key sleep-derived features in all body systems; the largest 10 correlations for each body system are presented in Fig. 3. In particular, our findings indicate that the correlation of pAHI with body composition features, with a Spearman correlation of 0.52

($P < 3 \times 10^{-48}$) for the total scan VAT area, surpasses those observed with sex, age and BMI, emerging as the highest correlation among all the body systems. Similarly, using mediation analysis, VAT had the largest average direct effect on pAHI when considering sex, age and BMI as potential mediators (Extended Data Fig. 4). Blood lipids and cardiovascular system-derived features were also found to be highly correlated with pAHI when compared to other body systems. Notably, blood triglycerides among blood lipids and the ribs area among bone density measurements remained highly correlated with pAHI even after mediation analysis. Although each of these associations was observed in previous works separately^{22–25}, we provide here a quantitative ranking between them. This ranking revealed, for example, that the previously hypothesized correlations of diet^{26,27} or microbiome^{28,29} with sleep characteristics were very limited, when compared to other body systems, or that daytime sleepiness³⁰ (from lifestyle features) association with pAHI was ranked only after the metabolic body systems. To further explore these robust associations, we examined the linear relationship between pAHI and specific blood biomarkers, including triglycerides,

Table 1 | Summary of key sleep-derived measurements

	Characteristic, mean (s.d.)	All visits			First visit			Second visit		
		Male (n=3,334)	Female (n=3,606)	All (n=6,940)	Male (n=3,043)	Female (n=3,323)	All (n=6,366)	Male (n=291)	Female (n=283)	All (n=574)
Sleep duration (h)	Average between nights	5.96 (0.99)	6.20 (1.00)	6.08 (1.00)	5.96 (0.99)	6.20 (1.00)	6.08 (1.00)	5.95 (1.00)	6.15 (1.03)	6.05 (1.02)
	Variation between nights	0.61 (0.43)	0.64 (0.47)	0.62 (0.45)	0.71 (0.51)	0.75 (0.55)	0.73 (0.53)	0.76 (0.59)	0.74 (0.56)	0.75 (0.57)
pAHI (events per h)	Average between nights	12.56 (10.75)	8.81 (8.50)	10.61 (9.83)	12.68 (10.82)	8.75 (8.41)	10.63 (9.83)	11.32 (9.92)	9.50 (9.48)	10.42 (9.74)
	Variation between nights	3.10 (3.06)	2.45 (2.58)	2.77 (2.84)	3.67 (3.57)	2.88 (3.03)	3.26 (3.33)	3.48 (4.16)	3.05 (3.11)	3.27 (3.69)
Mean SpO ₂ (%)	Average between nights	94.94 (1.11)	95.16 (1.19)	95.05 (1.16)	94.96 (1.11)	95.20 (1.18)	95.08 (1.15)	94.74 (1.13)	94.71 (1.26)	94.72 (1.20)
	Variation between nights	0.35 (0.31)	0.37 (0.34)	0.36 (0.33)	0.41 (0.37)	0.43 (0.40)	0.42 (0.39)	0.43 (0.35)	0.45 (0.37)	0.44 (0.36)
Mean nadir SpO ₂ (%)	Average between nights	92.05 (1.44)	91.96 (1.67)	92.00 (1.57)	92.06 (1.45)	92.00 (1.67)	92.03 (1.57)	91.91 (1.37)	91.47 (1.64)	91.69 (1.53)
	Variation between nights	0.69 (0.66)	0.89 (0.95)	0.79 (0.82)	0.83 (0.79)	1.10 (1.14)	0.97 (0.99)	0.84 (0.78)	1.05 (1.13)	0.94 (0.97)
Deep sleep (%)	Average between nights	17.49 (4.57)	17.93 (4.61)	17.72 (4.60)	17.50 (4.56)	17.96 (4.60)	17.74 (4.58)	17.42 (4.67)	17.50 (4.79)	17.46 (4.73)
	Variation between nights	2.62 (1.70)	2.57 (1.63)	2.59 (1.66)	3.10 (1.99)	3.04 (1.93)	3.07 (1.96)	3.19 (2.11)	2.97 (1.86)	3.09 (1.99)
Light sleep (%)	Average between nights	59.19 (8.79)	57.53 (9.14)	58.33 (9.01)	59.16 (8.82)	57.46 (9.16)	58.27 (9.04)	59.47 (8.56)	58.38 (8.93)	58.93 (8.75)
	Variation between nights	5.13 (3.23)	4.88 (3.18)	5.00 (3.21)	6.09 (3.78)	5.77 (3.76)	5.93 (3.77)	6.03 (4.03)	5.92 (3.82)	5.98 (3.92)
Mean snoring level (dB)	Average between nights	40.99 (1.47)	40.73 (1.25)	40.85 (1.37)	41.00 (1.48)	40.72 (1.26)	40.85 (1.38)	40.93 (1.34)	40.80 (1.05)	40.87 (1.21)
	Variation between nights	0.39 (0.51)	0.33 (0.45)	0.36 (0.48)	0.47 (0.61)	0.40 (0.54)	0.43 (0.57)	0.40 (0.55)	0.40 (0.49)	0.40 (0.52)
Sleep efficiency (%)	Average between nights	87.96 (5.14)	88.79 (4.61)	88.39 (4.89)	87.93 (5.18)	88.79 (4.63)	88.38 (4.92)	88.25 (4.77)	88.83 (4.39)	88.54 (4.59)
	Variation between nights	2.40 (1.88)	2.39 (1.91)	2.40 (1.90)	2.85 (2.23)	2.81 (2.28)	2.83 (2.26)	2.64 (2.14)	2.99 (2.05)	2.81 (2.10)
PRV during night time (RMSSD) (ms)	Average between nights	123.99 (81.82)	94.80 (67.13)	108.65 (75.87)	123.47 (80.87)	95.57 (68.18)	108.73 (75.72)	129.32 (90.95)	85.83 (52.61)	107.77 (77.51)
	Variation between nights	38.72 (57.03)	29.64 (51.89)	34.04 (54.62)	46.88 (67.43)	35.70 (62.13)	41.02 (64.94)	52.95 (77.33)	31.94 (58.82)	42.58 (69.54)
PRV during wake time (RMSSD) (ms)	Average between nights	108.28 (64.03)	93.96 (57.17)	100.84 (60.98)	107.64 (62.18)	94.28 (57.23)	100.67 (60.01)	115.03 (80.85)	90.13 (56.49)	102.78 (70.98)
	Variation between nights	34.18 (40.41)	29.75 (35.51)	31.91 (38.04)	45.97 (49.46)	39.22 (43.67)	42.48 (46.68)	45.04 (48.77)	38.71 (35.80)	41.78 (42.65)

Average and variability between nights values of key sleep-derived features for all visits, the first visit and the second visit. Data are shown as mean (s.d.) across male, female and all participants in the HPP cohort. For example, '5.96 (0.99)' in the first cell represents an average sleep duration of 5.96 h with an s.d. of 0.99 h, averaged across nights for all male participants. Similarly, '0.61 (0.43)' in the next cell indicates an average variability in sleep duration between nights of 0.61 h with an s.d. of 0.43 h for the same group. RMSSD, root mean square of successive differences between normal heartbeats.

hemoglobin A1C (HbA1C) and fasting glucose levels (Supplementary Fig. 1). For instance, our analysis showed that 49.7% of the male population with triglyceride levels exceeding the recommended threshold of 175 mg dL⁻¹ (ref. 31) and 55.5% and 53.9% of the male population with diabetes, based on fasting glucose and HbA1c thresholds³², respectively, have a pAHI indicative of moderate OSA. Interestingly, lifestyle habits, such as watching TV or smoking, showed stronger positive correlations with pAHI when compared to self-reported daytime sleepiness and negative correlations with sleep efficiency, whereas physical activity was negatively correlated with pAHI and positively correlated with PRV. Snoring levels and nadir oxygenation presented similar associations as for the pAHI, whereas sleep duration and efficiency presented poorer and fewer correlations with the other phenotypes. Finally, PRV was found to be positively correlated with sex, frailty and bone density features, as previously hypothesized^{33,34}.

Body system-level associations

Although a specific phenotype defined by a single feature may be highly correlated with a particular sleep-derived characteristic, this does not necessarily indicate a strong association between the entire body system related to that phenotype and sleep. Conversely, a lack of correlation between two phenotypes does not indicate that there is no association between a body system, when considered as a set of features, and a specific trait. Additionally, some phenotypes may exhibit nonlinear relationships with sleep characteristics that are not detectable through correlation analysis. To address this, we assessed the ability of a body system to predict a specific phenotype using both linear and nonlinear models (see Methods for the models used and calculation of predictive power). As sex, age and BMI have wide effects on human physiology and behavior, we trained and evaluated our models on male and female participants separately and adjusted them for age

and BMI. Additionally, as VAT was shown to be a strong predictor of sleep disorder breathing, the models based on sleep test measurements or predicting them were also adjusted for VAT.

Thus, we asked whether the sleep body system as a whole can predict other body system-derived features and how much it differs between male and female participants. We separated the sleep body system into two subgroups: sleep test measurements, which includes features related to respiratory events, sleep architecture, snoring and sleep positions (Methods and Supplementary Fig. 2), and PRV, which includes a set of calculations reflecting the autonomic variation in the nervous system (Methods and Supplementary Fig. 3). Figure 4 shows that sleep test measurements as a set of features adjusted for age, BMI and VAT provided the largest contribution in predicting insulin resistance and blood lipids with linear and nonlinear relationships, respectively (Extended Data Table 2), validating what was observed in previous studies^{16,22}. Sleep test measurements were significantly associated with 46% (23/50) and 50% (25/50) of the insulin resistance features and with 16% (472/3,042) and 11% (339/3,040) of the blood lipids tested for male and female participants, respectively. For example, the first quantile of glucose levels (from continuous glucose monitoring) was predicted based on sleep test measurements, age, BMI and VAT, with a Pearson correlation of 0.29 ± 0.02 (median \pm s.d.) and 0.34 ± 0.02 , compared to 0.11 ± 0.02 and 0.21 ± 0.02 when based only on age, BMI and VAT for male and female participants, respectively. This suggests that VAT does not mediate the association between sleep and insulin resistance, as opposed to what was hypothesized in previous work³⁵. Our results also demonstrated that certain body systems, such as diet, microbiome or mental health, now rank higher in their association with sleep test measurements, when considered as a comprehensive set of features and using nonlinear relationships, compared to their ranking in phenotype–phenotype pairwise relationships. No significant prediction was found for medications using the sleep test measurements set of features in both male and female participants. Interestingly, PRV as a set of features combined with age and BMI could also significantly contribute to explaining the variance of a large amount of blood lipids, above 15% (470/3,068) of them for male participants and 12% (400/3,068) for female participants. PRV could also be significantly associated with body composition, density, frailty and lifestyle, in addition to the predictable associations with the cardiovascular system (Extended Data Fig. 5).

In the opposite direction, we asked which body system can best predict sleep test measurements. We found that all body systems, when taken as a set of features, could contribute, to some extent, to the prediction of sleep test measurements above age and BMI in both sexes (Extended Data Fig. 6). Lifestyle provided the largest contribution, above age, BMI and VAT, in predicting some sleep test measurements, such as total time in bed or total arousal time for both male and female participants, which can be easily explained because questions related to sleep habits were included in the lifestyle features (Supplementary Table 1). Additionally, the cardiovascular system, insulin resistance, blood lipid levels and hematopoietic function (particularly in male participants), as well as mental health (especially in female participants), were identified as significant predictors of sleep test measurements. These findings are consistent with the associations observed throughout this study.

Although some sleep characteristics correlate with OSA, at no point did we train our models to specifically classify OSA. Because AHI, as a single biomarker, is recognized as a poor marker of OSA in healthy community cohorts³⁶, and because pAHI was only moderately correlated with symptoms of daytime sleepiness in our cohort (Extended Data Fig. 7), we defined a new variable based on both pAHI and self-reported symptoms of excessive daytime sleepiness to estimate symptomatic clinical OSA (Methods), as suggested in previous works³⁷. We evaluated the ability of each body system to classify individuals with or without clinical OSA. We found that 10 out of 16 tested

body systems had significant associations with clinical OSA in at least one sex (Extended Data Fig. 8), after adjusting for age, BMI and VAT. Notably, body systems that were poorly correlated with pAHI emerged as contributing more than other body systems to clinical OSA classification beyond age, BMI and VAT—for example, gut microbiome metabolic pathways for both female participants (area under the receiver operating characteristic curve (AUC) = 0.614, 95% confidence interval (CI): 0.612–0.617) and male participants (0.631 (0.629–0.632)) and diet for female participants (0.611 (0.608–0.613)). This suggests that daytime sleepiness may mediate their association with clinical OSA or that, when these body system features are taken as a whole, they contribute more in pAHI prediction compared to each feature separately.

Associations with medical diagnoses

The HPP cohort already includes 127 self-reported medical diagnoses at the baseline research stage, and more is expected to be gathered in the follow-up period. We first explored the associations of sleep test measurements and PRV features with these medical diagnoses self-reported at the time of the sleep test, in male and female participants separately, using a logistic regression model (Methods). The results presented in Fig. 5a show a total of 22 and 19 significant medical diagnosis associations for male and female participants, respectively, with either sleep test measurements adjusted for age, BMI and VAT or PRV features adjusted for age and BMI. One of the largest contributions of the sleep test measurements above age, BMI and VAT was for predicting sleep apnea in male participants, as expected because it included measurements such as pAHI and desaturation measurements that are commonly used for diagnosing this disease³⁸, however with relatively low prediction power (AUC = 0.517 ± 0.002 (median \pm s.d.)). Furthermore, it also contributed to predicting many other medical conditions and diagnoses, such as anxiety (0.507 ± 0.003 for male and 0.518 ± 0.007 for female participants) and hyperlipidemia (0.544 ± 0.004 for male and 0.522 ± 0.004 for female participants), aligning with the associated body systems mentioned earlier. We found sex-specific associations of sleep test measurements, such as hypertension, osteopenia and prediabetes, for female participant, whereas, for male participants, it was associated with allergy and back pain. We applied the same analysis on the PRV features, measured during night, and found associations with medical diagnoses that were already associated with sleep test measurements, such as hyperlipidemia for both sexes. In addition, PRV was associated with medical conditions that were not associated with sleep test measurements and differed between sexes. In male participants, PRV was associated with hypertension, atopic dermatitis and hearing loss, whereas, in female participants, PRV was associated with anxiety, osteoporosis and asthma, among other conditions. Second, we explored the longitudinal associations of sleep test measurements and PRV features at the baseline research stage with incident events of cardiovascular, metabolic and endocrinology diseases, self-reported in follow-up, using a Cox proportional hazard model (Methods and Extended Data Fig. 1). The results presented in Fig. 5b show that sleep architecture-related features, such as sleep duration and the percentage of deep sleep, are associated with increased risks for endocrinology diseases, whereas pAHI was more a predictor for cardiovascular and metabolic diseases.

Discussion

In this work, we analyzed sleep characteristics, including PRV calculated during night, in a cohort of 6,366 individuals with uniquely comprehensive phenotyping. Here we present the progression with age of the key features used for diagnosis of OSA, such as pAHI, showcasing the increasing prevalence of OSA with age in both male and female individuals corresponding to previous observations^{39,40} and with menopausal status in female participants. This last hypothesis is supported by previous works showing how menopause affects sleep characteristics and additional biomarkers such as blood lipids^{41,42}.

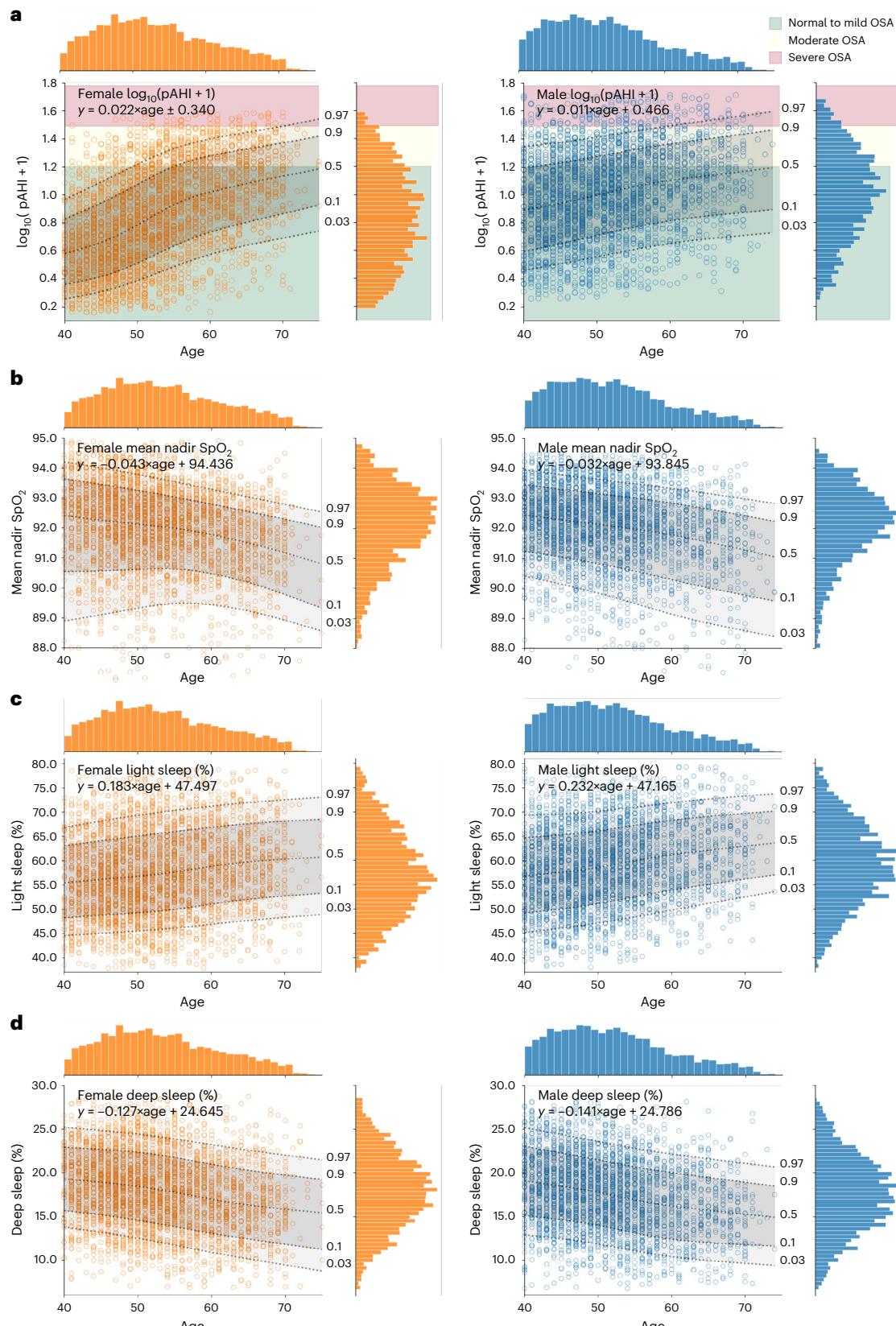
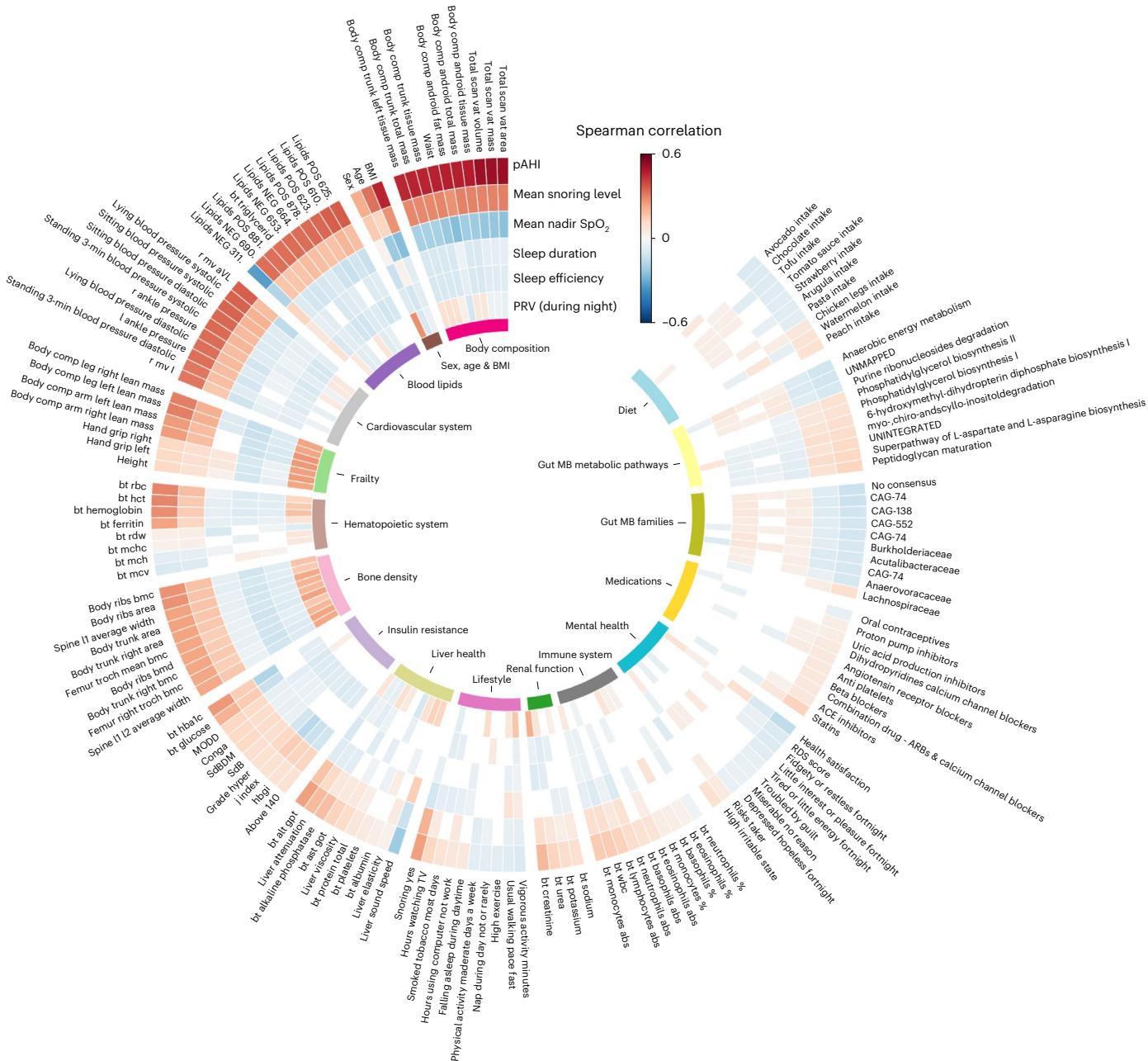


Fig. 2 | Association of sleep features with age in male and female participants. **a**, pAHI in logarithmic scale versus age for female (orange) and male (blue) participants separately. OSA severity levels are marked in green, yellow and red for normal-mild ($p\text{AHI} < 15$), moderate ($15 < p\text{AHI} < 30$) and severe ($p\text{AHI} > 30$) OSA, respectively. **b**, Mean nadir SpO_2 during sleep versus age for female (orange) and male (blue) participants separately. **c**, As in **b** for the percentage of time spent in light sleep. **d**, As in **b** for the percentage of time spent in deep sleep. For

all panels (**a-d**), the robust linear regression equation is shown in the top left of each graph, in which 'y' is the pAHI in logarithmic scale (**a**), mean nadir SpO_2 (**b**), percentage of light sleep (**c**) and percentage of deep sleep (**d**). The 3rd, 10th, 50th, 90th and 97th percentiles obtained using LOWESS regression are shown in dotted black lines on each graph. The histograms of the x and y axis values are shown on the top and right of each graph, respectively. LOWESS, locally weighted scatterplot smoothing.



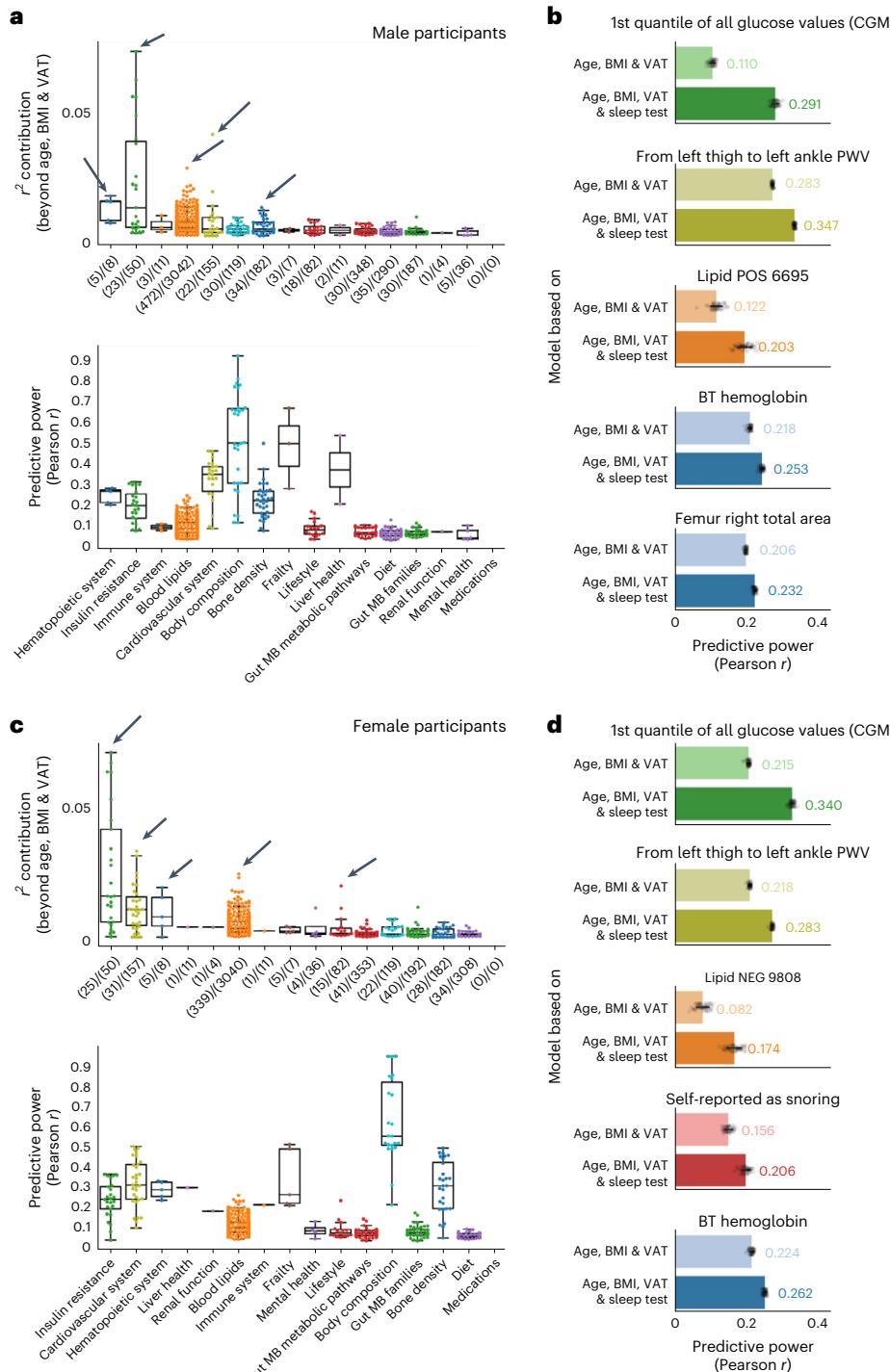


Fig. 4 | Body system feature prediction based on sleep test measurements averaged over three nights. **a,c**, Box and swarm plots (center, median; box, interquartile range (IQR); whiskers, $1.5 \times$ IQR) comparing the performance of models based on sleep test measurements averaged over three nights in predicting body system features, as compared to the performance of a baseline model, based on age, BMI and VAT, for male (**a**) and female (**c**) participants. Predictive power of each model was evaluated using fivefold cross-validation, repeated over 50 iterations (each using a different random seed), reporting the Pearson correlation r between predicted and actual values for the held-out set in each iteration. In **a** and **c**, the bottom graphs show the median predictive power obtained from these iterations for the models based on sleep test data, whereas the top graphs show the added variance that was explained (difference in r^2) using the models based on sleep test data, when compared to the baseline model. The number of features with significantly improved predictions using

the sleep test measurements (significant difference from the performance of a baseline model, with $P < 0.001$ in a two-sided t -test), which corresponds to the number of dots (n) included in these distributions, are shown for each body system in brackets on the x axis as (n /total number of features in the body system). Features are grouped by body system, and the body systems are ordered from left to right along the x axis in descending order of the median explained variance (box center) shown in the upper graphs of **a** and **c**. **b,d**, Bar and dot plots ($n = 50$; center, mean; error bar, s.d.) comparing the performances (each of the cross-validation r is plotted as a dot) of models based on age, BMI and VAT versus corresponding models based on age, BMI, VAT and sleep test measurements averaged over three nights for predicting the body system measurements marked with an arrow in plots **a** and **c**, respectively, for male (**b**) and female (**d**) participants. The mean predictive power value of each model is annotated on the right of each bar. BT, blood test.

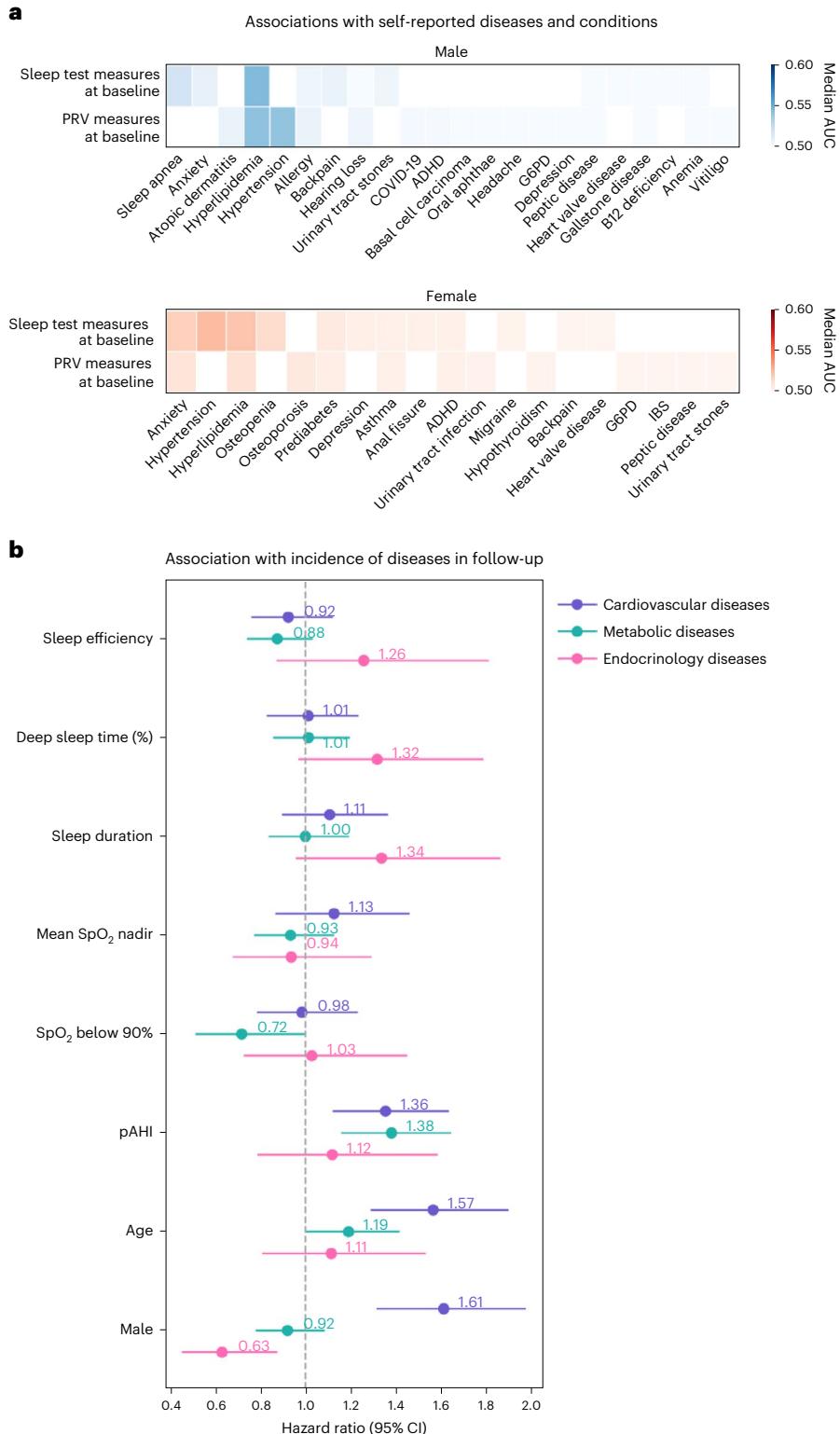


Fig. 5 | Sleep data-derived predictions of medical conditions and diseases. **a**, Heatmaps showing the predictive power of a logistic regression classifier based on either sleep test measurements at baseline or PRV features at baseline, combined with age and BMI, in predicting medical conditions and diseases as reported by male (top) or female (bottom) participants. Predictive power of each model was evaluated using fivefold cross-validation, repeated over 50 iterations, reporting the AUC for the held-out set in each iteration. The median prediction power (median AUC) obtained from these iterations is shown in these heatmaps for each of the conditions and diseases. Conditions or diseases for which the model based on sleep data combined with age and BMI did not provide a significant improvement compared to a model based only on age and

BMI (baseline model) are either not presented in this graph or masked (white cells). Conditions and diseases are sorted from left to right along the x axis from larger to smaller improvement in median AUC when using models based on either sleep test or PRV data as compared to the baseline model. **b**, The forest plot displays hazard ratios (HRs) in dots with 95% CI as error bars for various sleep characteristics at baseline in relation to the risk of developing cardiovascular (purple, $n = 4,555$), metabolic (green, $n = 3,651$) and endocrine (pink, $n = 4,624$) diseases in the following 2 years. HRs greater than 1 indicate increased risk, whereas HRs less than 1 suggest a protective effect. Substantial associations are highlighted. ADHD, attention-deficit/hyperactivity disorder; G6PD, glucose-6-phosphate dehydrogenase; IBS, irritable bowel syndrome.

that share a common physiological basis instead of being based on a single biomarker. We focused on two distinct subgroups: sleep test measurements from multi-night monitoring and nocturnal PRV. These subgroups demonstrated significant predictive power for most body system features with notable gender-specific variations. We chose to distinguish between these two subgroups to reflect the physiological aspect of sleep on the one hand, with the well-established parameters measured in a standard sleep study alone and with the exception of direct cardiac measurements such as pulse rate statistics, and, on the other hand, the autonomic nervous system activity reflected by the PRV measured during night.

Using all methods and for both sexes, sleep test measurements were found to be associated, beyond age and BMI, with body composition, specifically with VAT. VAT is a hormonally active component of the total body fat reflected by the intra-abdominal fat, known to be higher in male individuals, potentially explaining the higher prevalence of OSA in male individuals. Additionally, sleep test measurements were associated with insulin resistance measurements, validating in the Israeli population what was previously observed in China, the USA, Japan and Korea^{45,46}. Blood lipids association with sleep test measurements, and OSA in particular, was recently explored in different cohorts^{47,48}; in our results, the blood lipids also emerged as being significantly more predictable using sleep test measurements rather than only age, BMI and even VAT in both male and female participants. The associations of sleep with both blood lipids and insulin resistance, corresponding to the components of the metabolic syndrome, align with the findings of various additional works^{16,46,49}, without being mediated by visceral fat as previously hypothesized³⁵. A potential future work could include clinical trials to evaluate interventions, such as diet, exercise or medications, aimed at reducing VAT, which emerged as a key risk factor, and investigate how improving insulin resistance affects sleep apnea severity. Given the link between insulin resistance and sleep disturbances independent of VAT, future studies should explore whether early intervention in insulin resistance, through medication or lifestyle changes, improves sleep outcomes, especially in individuals at risk of OSA, and whether a more aggressive approach to treating OSA can improve glucose control in prediabetes and diabetes.

However, our analysis also highlighted the lack of strong associations between sleep and many phenotypes for which associations have been postulated^{50–52}. For example, associations between sleep and PRV domains with measurements of renal function or liver health were found to be lower compared to other body systems. These findings suggest that these relationships may be weaker or more complex than previously thought. Moreover, despite the substantial focus on nocturnal respiration, a limitation of this study is the absence of respiratory function assessments (for example, spirometry), which could provide a more complete picture ranking the impact of respiratory health on sleep.

When analyzing the associations of self-reported medical conditions and diagnoses with sleep characteristics with medical conditions, we confirmed very sex-specific associations, such as osteopenia and osteoporosis for female individuals^{53,54}. These observations may here again indicate the effect of the menopausal status, where these medical conditions typically appear after menopause due to reduction in estrogen levels⁵⁵. Similarly, the increase in the apnea events can be partly attributed to the postmenopause accumulation of intra-abdominal fat^{56,57}. Although the influence of sleep test measurements on human health is well established, nocturnal PRV remains largely unexplored in connection with most medical diagnoses, except for mental health⁵⁸. Overall, the relatively low prediction power observed in this analysis might be attributed to noise in the self-reported data, which may not accurately reflect the true medical state of the participants. For example, sleep apnea is frequently underdiagnosed⁵⁹, and, in our cohort, 85 participants (1.85%) reported being diagnosed with sleep apnea, and 1,134 participants (24.67%) had an averaged pAHI above 15 in this

same dataset. Another limitation for the associations with medical conditions presented in this work is the inclusion of medical conditions reported in very few samples, reducing the significance of some associations. This was done to include self-reported sleep apnea in the analysis, serving as a sanity check for our models because we anticipated finding an association with this condition.

Further research should focus on determining the causality behind the identified associations, such as whether metabolic, cardiovascular or dietary factors directly influence sleep outcomes or if sleep disturbances contribute to dysfunctions in these systems. Clarifying these causal relationships could enhance both preventive and therapeutic strategies. Additionally, artificial intelligence (AI)-based predictive models using deep learning on raw sleep monitoring data could uncover novel sleep features associated with health beyond the current sleep measurements included in our study.

Finally, the study population may be affected by selection bias due to the inclusion of healthy volunteers and from a single country, limiting the generalizability of the results for other populations. However, these findings can potentially serve as hypotheses for future research and validation, particularly when long-term follow-up data, as well as data from non-healthy participants and participants from diverse populations, become available within the HPP cohort.

The main novelty of our work lies in the comprehensive and simultaneous evaluation of high-resolution sleep characteristics with a wide range of phenotypes spanning across 16 body systems and the quantitative ranking of their associations within a single large adult cohort. We think that this holistic approach may assist in future research, discovery of biomarkers, development of predictive models or understanding of the underlying metabolic mechanisms related to sleep.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-024-03481-x>.

References

1. Benjafield, A. V. et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir. Med.* **7**, 687–698 (2019).
2. Gottlieb, D. J. & Punjabi, N. M. Diagnosis and management of obstructive sleep apnea: a review. *JAMA* **323**, 1389–1400 (2020).
3. Redline, S. & Purcell, S. M. Sleep and big data: harnessing data, technology, and analytics for monitoring sleep and improving diagnostics, prediction, and interventions—an era for Sleep-Omics? *Sleep* **44**, zsab107 (2021).
4. Levy, J., Álvarez, D., Del Campo, F. & Behar, J. A. Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry. *Nat. Commun.* **14**, 4881 (2023).
5. Zhang, G.-Q. et al. The National Sleep Research Resource: towards a sleep data commons. *J. Am. Med. Inform. Assoc.* **25**, 1351–1358 (2018).
6. Dashti, H. S. et al. Sleep health, diseases, and pain syndromes: findings from an electronic health record biobank. *Sleep* **44**, zsaa189 (2021).
7. Brooks, T. G. et al. Diurnal rhythms of wrist temperature are associated with future disease risk in the UK Biobank. *Nat. Commun.* **14**, 5172 (2023).
8. Zheng, N. S. et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program. *Nat. Med.* **30**, 2648–2656 (2024).
9. Viswanath, V. K. et al. Five million nights: temporal dynamics in human sleep phenotypes. *NPJ Digit. Med.* **7**, 150 (2024).

10. Shilo, S. et al. 10 K: a large-scale prospective longitudinal study in Israel. *Eur. J. Epidemiol.* **36**, 1187–1194 (2021).
11. Thayer, J. F. & Sternberg, E. Beyond heart rate variability: vagal regulation of allostatic systems. *Ann. N. Y. Acad. Sci.* **1088**, 361–372 (2006).
12. Chouraki, A., Tournant, J., Arnal, P., Pépin, J.-L. & Bailly, S. Objective multi-night sleep monitoring at home: variability of sleep parameters between nights and implications for the reliability of sleep assessment in clinical trials. *Sleep* **46**, zsac319 (2023).
13. Patil, S. P. & Kirkness, J. P. Pathogenesis of obstructive sleep apnea in obesity. In *Obesity and Lung Disease: A Guide to Pathophysiology, Evaluation, and Management* (eds Dixon, A. E. & Forno, E.) 125–150 (Springer International Publishing, 2024).
14. Zhang, D., Ma, Y., Xu, J. & Yi, F. Association between obstructive sleep apnea (OSA) and atrial fibrillation (AF): a dose-response meta-analysis. *Medicine (Baltimore)* **101**, e29443 (2022).
15. Moula, A. I. et al. Obstructive sleep apnea and atrial fibrillation. *J. Clin. Med.* **11**, 1242 (2022).
16. Reutrakul, S. & Mokhlesi, B. Obstructive sleep apnea and diabetes: a state of the art review. *Chest* **152**, 1070–1086 (2017).
17. Korostovtseva, L., Bochkarev, M. & Sviryayev, Y. Sleep and cardiovascular risk. *Sleep Med. Clin.* **16**, 485–497 (2021).
18. Diament, A. et al. A multimodal dataset of 21,412 recorded nights for sleep and respiratory research. Preprint at <https://doi.org/10.48550/arxiv.2311.08979> (2023).
19. Goyal, M. & Johnson, J. Obstructive sleep apnea diagnosis and management. *Mo Med.* **114**, 120–124 (2017).
20. Fietze, I. et al. Prevalence and association analysis of obstructive sleep apnea with gender and age differences—results of SHIP-Trend. *J. Sleep Res.* **28**, e12770 (2019).
21. Dong, Z. et al. Association of overweight and obesity with obstructive sleep apnoea: a systematic review and meta-analysis. *Obes. Med.* **17**, 100185 (2020).
22. Nadeem, R. et al. Effect of obstructive sleep apnea hypopnea syndrome on lipid profile: a meta-regression analysis. *J. Clin. Sleep Med.* **10**, 475–489 (2014).
23. Harada, Y. et al. Differences in associations between visceral fat accumulation and obstructive sleep apnea by sex. *Ann. Am. Thorac. Soc.* **11**, 383–391 (2014).
24. Zheng, C. et al. Visceral adipose tissue indices independently correlated with obstructive sleep apnea in patients with type 2 diabetes. *J. Diabetes Res.* **2022**, 4950528 (2022).
25. Strassberger, C. et al. Beyond the AHI—pulse wave analysis during sleep for recognition of cardiovascular risk in sleep apnea patients. *J. Sleep Res.* **30**, e13364 (2021).
26. Peuhkuri, K., Sihvola, N. & Korpela, R. Diet promotes sleep duration and quality. *Nutr. Res.* **32**, 309–319 (2012).
27. Godos, J. et al. Association between diet and sleep quality: a systematic review. *Sleep Med. Rev.* **57**, 101430 (2021).
28. Cai, Y., Juszczak, H. M., Cope, E. K. & Goldberg, A. N. The microbiome in obstructive sleep apnea. *Sleep* **44**, zsab061 (2021).
29. Li, Q. et al. Obstructive sleep apnea is related to alterations in fecal microbiome and impaired intestinal barrier function. *Sci. Rep.* **13**, 778 (2023).
30. Ulander, M., Hedner, J., Stillberg, G., Sunnergren, O. & Grote, L. Correlates of excessive daytime sleepiness in obstructive sleep apnea: results from the nationwide SESAR cohort including 34,684 patients. *J. Sleep Res.* **31**, e13690 (2022).
31. Grundy, S. M. et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J. Am. Coll. Cardiol.* **73**, e285–e350 (2019).
32. American Diabetes Association Professional Practice Committee. 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2024. *Diabetes Care* **47**, S20–S42 (2024).
33. Zhang, Y. et al. Research for correlation between heart rate variability parameters and bone mineral density in patients of type 2 diabetes mellitus. *J. Endocrinol. Invest.* **46**, 79–88 (2023).
34. Arantes, F. S. et al. Heart rate variability: a biomarker of frailty in older adults? *Front. Med. (Lausanne)* **9**, 1008970 (2022).
35. Bozkurt, N. C. et al. Visceral obesity mediates the association between metabolic syndrome and obstructive sleep apnea syndrome. *Metab. Syndr. Relat. Disord.* **14**, 217–221 (2016).
36. Pevernagie, D. A. et al. On the rise and fall of the apnea-hypopnea index: a historical review and critical appraisal. *J. Sleep Res.* **29**, e13066 (2020).
37. Malhotra, A. et al. Metrics of sleep apnea severity: beyond the apnea-hypopnea index. *Sleep* **44**, zsab030 (2021).
38. Körkuyu, E. et al. The efficacy of Watch PAT in obstructive sleep apnea syndrome diagnosis. *Eur. Arch. Otorhinolaryngol.* **272**, 111–116 (2015).
39. Vicini, C. et al. The aging effect on upper airways collapse of patients with obstructive sleep apnea syndrome. *Eur. Arch. Otorhinolaryngol.* **275**, 2983–2990 (2018).
40. Cunningham, J. et al. The prevalence and comorbidities of obstructive sleep apnea in middle-aged men and women: the Busselton Healthy Ageing Study. *J. Clin. Sleep Med.* **17**, 2029–2039 (2021).
41. Ameratunga, D., Goldin, J. & Hickey, M. Sleep disturbance in menopause. *Intern. Med. J.* **42**, 742–747 (2012).
42. Beyene, H. B. et al. High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: evidence from two large population cohort studies. *PLoS Biol.* **18**, e3000870 (2020).
43. Nozawa, S., Urushihata, K., Machida, R. & Hanaoka, M. Sleep architecture of short sleep time in patients with obstructive sleep apnea: a retrospective single-facility study. *Sleep Breath.* **26**, 1633–1640 (2022).
44. Boulos, M. I. et al. Normal polysomnography parameters in healthy adults: a systematic review and meta-analysis. *Lancet Respir. Med.* **7**, 533–543 (2019).
45. Lee, S. W. H., Ng, K. Y. & Chin, W. K. The impact of sleep amount and sleep quality on glycemic control in type 2 diabetes: a systematic review and meta-analysis. *Sleep Med. Rev.* **31**, 91–101 (2017).
46. Alshehri, M. M. et al. Associations between ankle-brachial index, diabetes, and sleep apnea in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) database. *BMC Cardiovasc. Disord.* **20**, 118 (2020).
47. Nakatsuka, Y. et al. Hyperfructosemia in sleep disordered breathing: metabolome analysis of Nagahama study. *Sci. Rep.* **13**, 12735 (2023).
48. Zhang, Y. et al. Development and validation of a metabolite index for obstructive sleep apnea across race/ethnicities. *Sci. Rep.* **12**, 21805 (2022).
49. Wang, F. et al. The association between obstructive sleep apnea syndrome and metabolic syndrome: a confirmatory factor analysis. *Sleep Breath.* **23**, 1011–1019 (2019).
50. Marjot, T., Ray, D. W., Williams, F. R., Tomlinson, J. W. & Armstrong, M. J. Sleep and liver disease: a bidirectional relationship. *Lancet Gastroenterol. Hepatol.* **6**, 850–863 (2021).
51. Parikh, M. P., Gupta, N. M. & McCullough, A. J. Obstructive sleep apnea and the liver. *Clin. Liver Dis.* **23**, 363–382 (2019).
52. Mazidi, M., Shekoohi, N., Katsiki, N. & Banach, M. Longer sleep duration may negatively affect renal function. *Int. Urol. Nephrol.* **53**, 325–332 (2021).
53. Lucassen, E. A. et al. Poor sleep quality and later sleep timing are risk factors for osteopenia and sarcopenia in middle-aged men and women: the NEO study. *PLoS ONE* **12**, e0176685 (2017).

54. Ochs-Balcom, H. M. et al. Short sleep is associated with low bone mineral density and osteoporosis in the Women's Health Initiative. *J. Bone Miner. Res.* **35**, 261–268 (2020).
55. Sipilä, S. et al. Muscle and bone mass in middle-aged women: role of menopausal status and physical activity. *J. Cachexia Sarcopenia Muscle* **11**, 698–709 (2020).
56. Greendale, G. A. et al. Changes in regional fat distribution and anthropometric measures across the menopause transition. *J. Clin. Endocrinol. Metab.* **106**, 2520–2534 (2021).
57. Ambikairajah, A., Walsh, E., Tabatabaei-Jafari, H. & Cherbuin, N. Fat mass changes during menopause: a metaanalysis. *Am. J. Obstet. Gynecol.* **221**, 393–409 (2019).
58. Schiweck, C., Piette, D., Berckmans, D., Claes, S. & Vrieze, E. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. *Psychol. Med.* **49**, 200–211 (2019).
59. Young, T., Evans, L., Finn, L. & Palta, M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* **20**, 705–706 (1997).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Sarah Kohn  ¹, **Alon Diament**  ², **Anastasia Godneva** ¹, **Raja Dhir** ³, **Adina Weinberger**  ^{1,2}, **Yotam Reisner**  ^{1,2},
Hagai Rossman ² & **Eran Segal**  ^{1,2,4} 

¹Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. ²Pheno.AI, Ltd., Tel Aviv, Israel. ³Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland. ⁴Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE.  e-mail: eran.segal@weizmann.ac.il

Methods

Description of cohort

The data presented in this paper were collected between January 2019 and December 2022, from a total of 6,748 participants aged between 40 years and 75 years, who were enrolled as part of the HPP study and who underwent at least one home sleep test as part of the study. This study was approved by the institutional review board of the Weizmann Institute of Science (reference no. 1719-1), and all participants were self-assigned volunteers with informed consent. The cohort in this study is one of the largest longitudinal studies established in Israel, with a population originating from several different ancestries who reside in a relatively small geographic region and, therefore, share a relatively similar environment and habits. The population is largely composed of educated European (Ashkenazi) Jews, healthy at the time of recruitment (that is, severe medical conditions and diseases were defined as exclusion criteria), who have follow-up visits every 2 years for 25 years. For the full study design, see ref. 10.

The data include various clinical, physiological, behavioral and multi-omic profiling data, which we categorized into 17 groups⁶⁰: sleep characteristics and other 16 body systems representing major physiological systems and environmental exposures (Fig. 1).

Body system-derived features

The following groups represent the baseline characteristics and system-level categories used for this work. The exact number of individuals, data points and features available in each group can be found in Extended Data Table 1.

Sex, age and BMI (baseline characteristics): this category was defined to set a baseline category, including basic physiological features known as covariates, for most health conditions and status: sex, age and BMI.

Blood lipids: includes 3,098 lipid clusters as computed by lipidomics, using a Waters ACQUITY UPLC system coupled to a Vion IMS QToF mass spectrometer (Waters Corporation) and in-house processing, as well as the following blood tests: high-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol and triglycerides.

Body composition: includes 108 measurements of fat and lean mass assessments for the legs, arms, trunk, gynoid and android regions. These measurements are derived from dual-energy X-ray absorptiometry imaging. Additionally, anthropometric parameters, such as weight, height and hip and waist circumference, were also included. BMI was not part of this category because it was already accounted for in the baseline group.

Bone density: includes 182 measurements of the mineral content in different parts of a variety of skeletal components, based on dual-energy X-ray absorptiometry imaging.

Cardiovascular system: includes measurements from various tests: blood pressure measurements, blood pressure ratios computed using the ankle–brachial index test, arterial stiffness estimated by pulse wave velocity using a Falcon device (Viasonix) and carotid intima-media thickness computed from the carotid ultrasound using a SuperSonic Aixplorer MACH 30 (Hologic), vascular parameters averaged over both eyes computed from retinal imaging using an iCare DRSpplus confocal fundus imaging system (iCare) and the Python AutoMorph package⁶¹, as well as electric activity of the heart as captured in the 12-lead resting ECG using a PC-ECG 1200 machine (NORAV).

Diet: includes daily mean consumption of 322 foods over a maximal period of 16 days of self-logging. Logging days of fewer than 500 kcal were excluded from the analysis. Foods that were not logged by at least 5% of all participants were not included. Logging data were clipped at 10 s.d., as computed using the middle 99th percentile of the data.

Frailty: includes measurements of hand grip strength as well as the lean mass of the arms and legs (overlapping with body composition).

Gut microbiome: includes relative abundances of 627 families.

We used the method for metagenomic reads extraction and bacterial abundances estimation previously described⁶² at the family level, in combination with the previously published improved human gut microbiome reference set for mapping⁶³. Bacterial families that were not identified in at least 5% of the samples were not included. Missing data were assumed to represent missing abundance or abundances that are below the detection limit. Thus, missing data were imputed by a minimum value of 0.0001. In this work, the log₁₀-transformed values of the resulting abundances per sample were used.

Gut microbiome metabolic pathways: microbial pathway abundances from gut microbiome metagenomic data, using HUMAnN3 (ref. 64) functional profiling.

Hematopoietic system: includes the following blood laboratory test measurements, which are characteristics and components of red blood cells: ferritin, hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, red blood cells, hematocrit, mean corpuscular hemoglobin and red blood cell distribution width.

Immune system: includes the following immune cell measurements from complete blood count: white blood cells, absolute eosinophils, percentage eosinophils, absolute monocytes, percentage monocytes, absolute lymphocytes, percentage lymphocytes, absolute basophils, percentage basophils, absolute neutrophils and percentage neutrophils.

Insulin resistance: includes 49 parameters describing blood glucose variability computed as previously described⁶⁵. In addition, this category includes fasting glucose and HbA1C from laboratory tests.

Lifestyle: assessed by 46 questions from questionnaires, following the structure and content of the UK Biobank⁶⁶, addressing the following topics: smoking, alcohol consumption, physical activity, employment, sleep, sun exposure and electronic device use (full list in Supplementary Table 1).

Liver health: assessed by parameters from a liver ultrasound and 2D shear wave elastography (2D-SWE), including measurements of viscosity, elasticity, attenuation and sound speed. In addition, the following liver enzymes and related blood tests are included: alkaline phosphatase, alanine transaminase, aspartate transaminase, total protein, total bilirubin, platelets and albumin.

Medications: includes a total of 60 medication types, with self-reports relating to if the medication was taken between the study enrollment and the data point or not (1 or 0, respectively).

Mental health: assessed by 35 questions from a questionnaire following the structure and content of the UK Biobank⁶⁶ equivalent questionnaire relating to the individual's mood, satisfaction and depressive symptoms. In addition, recent depressive symptoms (RDS) score was computed as the sum of the self-reported scores of the following questions; each self-reported score is a number between 1 and 4 (1 indicates not at all, 4 indicates nearly every day) and, therefore, ranges from 4 to 16:

Over the past two weeks, how often have you felt down, depressed or hopeless?

Over the past two weeks, how often have you had little interest or pleasure in doing things?

Over the past two weeks, how often have you felt tense, fidgety or restless?

Over the past two weeks, how often have you felt tired or had little energy?

The list of the 36 total features for this body system can be found in Supplementary Table 2.

Renal function: includes creatinine, urea and the electrolytes sodium and potassium, all derived from blood laboratory tests.

Sleep characteristics: includes 448 features regrouped into two subgroups, sleep test measurements and PRV features, extracted as described in the following section.

Sleep-derived features extraction

Each participant included in this cohort underwent one or two series of home sleep monitoring tests, each series consisting of three nights, within a timeframe of 2 weeks, of continuous sleep monitoring using the WatchPAT 300 device (ZOLL Itamar). In case a same participant underwent two series of testing, a gap of 2 years differentiates between them.

One data point was considered as one series of sleep monitoring of a single participant. For each participant, all the data collected within a window of ± 6 months from its sleep data were considered as the same data point. Detailed information for the timing of the sleep monitoring series with respect to all other phenotypes collected at a certain datapoint (that is, baseline research stage or second visit after 2 years) can be found in Fig. 1c.

In total, 20,288 nights of monitoring were collected; among these, a total of 16,812 nights from 6,366 individuals, resulting in 6,490 sleep data points, were included in this work (see detailed selection process in Extended Data Fig. 1).

For each night of monitoring, peripheral blood oxygen saturation levels, pulse rates, respiratory events, snoring levels, discrete sleep positions and stages were derived from the signals collected, and corresponding overnight features were extracted from Itamar medical software, validated against gold standard polysomnography^{67,68}.

Oxygen saturation levels and pulse rates were determined by finger pulse oximetry, when oxygen desaturation events are determined with a minimal oxygen desaturation of 4%. Peripheral apnea–hypopnea events and respiratory effort-related arousals were determined by the WatchPAT 300's automatic algorithm analyzing the peripheral arterial tonometry signal amplitude alongside the pulse rate and oxygen saturation, with minimal oxygen desaturation of 3% as default. A RERA is defined as an arousal from sleep that follows 10 s or more of increased respiratory effort but does not meet the criteria for apnea or hypopnea. Subsequently, pAHI, respiratory disturbance index (including both apnea–hypopnea events and respiratory effort-related arousals) and oxygen desaturation index were estimated as the number of corresponding events per hour. pAHI based on 4% desaturation events was also estimated and referred to as ‘pAHI 4%’, and ‘pAHI’ refers to the default calculation based on 3% desaturation events. Additionally, the snoring sensor included in the WatchPAT 300 was used to provide snoring levels and to determine if respiratory events were obstructive or central. The body position sensor was used to determine the discrete body positions and to identify if sleep apnea had a positional component. Finally, the discrete sleep stages were determined by the WatchPAT 300's algorithm analyzing the peripheral arterial tonometry signal amplitude and actigraphy.

Large measurements of events per hour (apnea–hypopnea, oxygen desaturation and respiratory disturbance indexes) were clipped at 50.

In addition, sleep efficiency and variability between sleep stages, as sleep architecture-related measurements, were computed as follows:

$$\text{Sleep efficiency} = 100 \times \frac{\text{Time asleep}}{\text{Total time in bed}}$$

Sleep stages variability

$$= \sqrt{\frac{1}{3} \left((\text{SD}_{\% \text{deep sleep}})^2 + (\text{SD}_{\% \text{light sleep}})^2 + (\text{SD}_{\% \text{REM sleep}})^2 \right)}$$

We regrouped all these measurements—excluding the pulse rate measurements because they are very redundant with the ECG measurements—into a subgroup denoted as the sleep test measurements, listing in total 100 features (Supplementary Table 3).

Furthermore, another subgroup, related to PRV, was defined with 348 features (Supplementary Table 4). The features were extracted from the device raw peripheral arterial tonometry signal⁶⁹, using the NeuroKit2 Python package⁷⁰, computing a signal quality score and 86 PRV features spanning five feature families: time domain, frequency

domain, nonlinear, complexity/entropy and fractal dimension. Each of these features was computed on four segments of the recordings per night: entire night, longest non-REM sleep segment, longest REM segment and longest wake segment. Each segment was at least 5 min long.

For all analyses performed in this work, we compared different aggregation methods to combine the feature measurements of all nights into one value per data point—averaging the measurements across nights or taking the measurements of the longest night of sleep. The first method outperformed the second; therefore, all results presented in this work were using sleep features averaged over all nights of monitoring of the same data point, unless stated otherwise.

Body systems cluster map

To better visualize the high-dimensional data included in this work, and in the HPP cohort in general, we regrouped all the data points that had simultaneous measurements for all the 17 body systems depicted in Fig. 1a—in total, 1,309 data points with 5,956 features each (sum of the features numbers in Extended Data Table 1)—and mapped the features into a 2D embedding space. For each body system having more than 200 features, we first performed dimensionality reduction using principal component analysis (PCA)⁷¹ to 200 components implemented in the scikit-learn⁷² Python package. Then, we mapped each of the body system features or PCAs, each represented by 1,309 data points, in a 2D space using the UMAP Python package, which implements an algorithm based on Riemannian geometry and algebraic topology to perform dimension reduction and data visualization⁷³.

Association of sleep-derived features with age

To assess the relationship between sleep-derived features and age, a robust linear regression was performed using Hubber regression implemented in the statsmodels Python package⁷⁴ on male and female participants separately.

Association of sleep characteristics with menopausal status

We classified female participants as premenopausal or postmenopausal based on self-report questionnaires. A female who reported that she was no longer menstruating and had been experiencing amenorrhea for a year was identified with menopause.

As the menopausal transition usually takes place within a narrow range of age, we subsampled our initial cohort for female participants around the typical age of menopause, performing a 1:1 matching between premenopausal and postmenopausal female participants over age. We then applied a robust linear regression to assess the relationship between the sleep-derived feature and age (as mentioned above) in the two female subgroups separately.

Longitudinal analysis

To assess the trends in some sleep characteristics when measured 2 years apart in the same individual (Extended Data Fig. 2), we calculated the percentage of change from first measurement in the 574 individuals who presented such repeated sleep test as follows:

$$\text{Percentage of change} = 100$$

$$\times \frac{(\text{averaged measure in 2nd series} - \text{average measure in 1st series})}{\text{averaged measure in 1st series}}$$

To determine if these longitudinal changes distribute significantly around a mean value different from zero, we computed the *P* values of a one-sample two-sided *t*-test using the statsmodels Python package⁷⁴.

Correlation of sleep-derived features with body systems

To identify the body system-derived features associated with pAHI, snoring level, mean SpO₂, sleep time, sleep efficiency and PRV measured during night, we performed a Spearman correlation between each of the above-mentioned sleep-derived features with each of the

body system-derived features separately, with sex, age, gender and BMI as covariates.

For this, each body system dataset was paired with the sleep characteristic dataset, including both male and female participants. For each dataset, the data points that did not have a sleep measurement from the same individual at the same research stage were excluded from the analysis. The resulting paired subset sample sizes are shown in the 'All' column of Extended Data Table 1. Features that had fewer than 500 valid values in the resulting paired datasets were also removed from the analysis.

Spearman partial correlation coefficients and *P* values were computed using the Pingouin Python package^{75,76}. We applied false discovery rate (FDR) correction with 10% error rate, for each paired dataset separately, using the MNE Python package^{77,78}.

Predictive models

To identify associations between sleep characteristics and other body systems, we applied regression models (see detailed method in 'Comparison of linear model with gradient boosting decision trees') on the sleep test measurements dataset and the PRV dataset to predict the features of each body system, for each sex separately. We also applied these models of each of the body system datasets to predict sleep test measurements, for each sex separately.

Definition of covariates. As a default, we used age and BMI as covariates for the analyses stratified by sex. However, for the models predicting traits based on sleep test measurements, or predicting sleep test measurements based on other traits, we added VAT as a covariate in addition to age and BMI. This is based on the high correlation found between pAHI and VAT, beyond age, sex and BMI.

Pre-processing. We paired each predicting dataset with the defined covariates for male and female participants separately (the resulting subset sample sizes can be found in the 'Males' and 'Females' columns of Extended Data Table 1). To ensure sufficient variability in the data, we mandated a minimum of 200 samples with values that differed from the most prevalent value for each feature in the resulting sex-specific paired dataset; if this requirement was not satisfied, the feature was removed from the dataset. We imputed all missing values from the predicting features using a multivariate imputer that estimates each feature from the 10 nearest other predicting features. We then standardized all the predicting features (both operations were using the scikit-learn⁷² Python package). The following inference and validation methods were performed per feature from the predicted dataset—that is, the targeted trait—and per sex.

Inference. We ran a hyperparameter grid search using five-fold cross-validation; the detailed list of hyperparameter combination tested can be found in 'Regression hyperparameters search' below. Then, using the best hyperparameters found, we ran 50 repeated five-fold cross-validations to compute 50 scores. For each iteration, the data were split in a random five-fold, each sample in the data belonging to exactly one test set, and its prediction was computed with an estimator fitted on the corresponding training set. Then, we computed the model score of each such iteration as the Pearson correlation between all the predictions and their actual value. The predictive power of a model was estimated as the median of its 50 scores.

Validation. To evaluate the results, the traits were additionally inferred using the defined covariates only, from the same subset, resulting in an additional 50 scores. We performed a *t*-test to compare the two groups of 50 scores each—one group for the model based on the predicting dataset with age, BMI and VAT and the second group for the model based on only age, BMI and VAT as baseline. A dataset was considered to significantly improve the predictions of a specific trait if the scores

distribution of this dataset was found to be significantly higher than the scores distribution of the baseline model (*P*<0.001 in a two-sided *t*-test).

Comparison of linear versus nonlinear models

For each domain–domain association (that is, predicting features from a specific body system based on sleep test or PRV measurements or predicting sleep test measurements based on a specific body system feature), we evaluated whether the relationship was linear or nonlinear. We used a least absolute shrinkage and selection operator (LASSO) regression model (using the scikit-learn⁷² Python package) for linear associations and a gradient boosting decision tree model (using the LightGBM⁷⁹ Python package) for nonlinear associations. We then compared the performance of the two model types at the domain–domain level, retaining only the best-performing model for the results presented (Extended Data Table 2). This involved determining whether more significant associations were identified with linear or nonlinear models or, if an equivalent number of significant associations was found, which model provided higher predictive power.

Association with current diseases and medical conditions

To identify associations between body systems and clinical OSA, or between sleep characteristics and diseases or medical conditions, we applied a logistic regression model with an elastic net regularization (using the scikit-learn⁷² Python package), for male and female participants separately. Imputation of missing feature values and scaling were performed as mentioned above (see 'Predictive models' subsection). Inference and validation were performed as mentioned above, with the exception of the model score calculation, which was computed as the AUC for each iteration, to adapt to binary classification.

Associations with clinical OSA. To predict clinical OSA from body systems, we defined individuals with clinical OSA using their pAHI and self-reported daytime sleepiness from the lifestyle questionnaire, based on the following criteria:

- pAHI averaged over three nights is greater than 15 and
- Daytime sleepiness is not reported as rare or not at all

Because this definition was based on sleep and lifestyle data, we excluded the daytime sleepiness-related questions from the lifestyle body system when evaluating its role in predicting clinical OSA. In addition to sleep, we assessed the association of the remaining 16 body systems (along with age, BMI and VAT) with clinical OSA, analyzing male and female participants separately.

Associations with self-reported diseases and conditions. For predictions of diseases and medical conditions from sleep characteristics (that is, sleep test and PRV measurements), we used a subset of the sleep test and PRV measurements for this analysis, using only samples from the baseline research stage, resulting in 2,180 male and 2,416 female participants. This is to be able to further validate these models with the longitudinal data that will be available in the future for this HPP study. For each pair—that is, sleep features of a specific sex at baseline and a singular diagnosis among the 127 self-reported medical diagnoses—we applied the logistic regression only if the singular diagnosis was reported by at least 18 individuals. This threshold was set to include sleep apnea diagnosis in the analysis. In total, 41 and 49 medical diagnoses were analyzed for male and female participants, respectively.

Association with incidence of diseases in follow-up

For this longitudinal analysis, we estimated hazard ratios for incidence of cardiovascular, metabolic and endocrinology diseases based on baseline measurements (*n*=5,154 observations; Extended Data Fig. 1), fitting a Cox proportional hazard model⁸⁰, using the lifelines Python package. Baseline measurements included sex, age, pAHI,

desaturations below 90%, mean SpO₂ nadir, sleep duration, the percentage of deep sleep and sleep efficiency. Imputation of missing baseline features values and scaling were performed as mentioned above (see ‘Predictive models’ subsection). The Cox model was trained to fit the time of incidence of events related to cardiovascular, metabolic and endocrinology diseases, self-reported during the 2-year follow-up.

Regression hyperparameters search

For the LightGBM models, a grid search was applied on all combinations of the following parameters:

- ‘objective’: [‘regression’]
- ‘min_child_weight’: [0.02 * 3000, 0.03 * 3000, 0.04 * 3000, 0.05 * 3000]
- ‘max_depth’: [3, 4]
- ‘n_estimators’: [1000]
- ‘num_leaves’: [500, 1000, 1500]
- ‘feature_fraction’: [0.1, 0.15, 0.2]
- ‘bagging_fraction’: [0.7]
- ‘alpha’: [0.1, 0.2, 0.5, 0.8]
- ‘lambda_l2’: [0.1, 0.2, 0.5, 0.8]

For LASSO regression models:

- ‘alpha’: [0.1, 1.0, 5.0, 10.0, 50.0]

For logistic regression models:

- ‘penalty’: [‘elasticnet’]
- ‘solver’: [‘saga’]
- ‘l1_ratio’: [0, 0.5, 1]
- ‘max_iter’: [1000]

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data in this paper are part of the Human Phenotype Project (HPP) and are accessible to researchers from universities and other research institutions at <https://humanphenotyperproject.org/data-access>.

The HPP data include personal information and, in compliance with institutional review board regulations, cannot be made publicly available. Interested bona fide researchers should contact info@pheno.ai to obtain instructions for accessing the data, which is typically granted within a few days.

Code availability

Code used in this study is available at the following GitHub link: <https://github.com/SarahKohn/SleepAssociationsHPP>.

References

60. Reicher, L. et al. Phenome-wide associations of human aging uncover sex-specific dynamics. *Nat. Aging* **4**, 1643–1655 (2024).
61. Zhou, Y. et al. Automorph: automated retinal vascular morphology quantification via a deep learning pipeline. *Transl. Vis. Sci. Technol.* **11**, 12 (2022).
62. Rothschild, D. et al. An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. *PLoS ONE* **17**, e0265756 (2022).
63. Leviatan, S., Shoer, S., Rothschild, D., Gorodetski, M. & Segal, E. An expanded reference map of the human gut microbiome reveals hundreds of previously unknown species. *Nat. Commun.* **13**, 3863 (2022).
64. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).
65. Keshet, A. et al. CGMap: characterizing continuous glucose monitor data in thousands of non-diabetic individuals. *Cell Metab.* **35**, 758–769 (2023).
66. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
67. Hedner, J. et al. Sleep staging based on autonomic signals: a multi-center validation study. *J. Clin. Sleep Med.* **7**, 301–306 (2011).
68. Yalamanchali, S. et al. Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* **139**, 1343–1350 (2013).
69. Schnall, R. P., Sheffy, J. K. & Penzel, T. Peripheral arterial tonometry-PAT technology. *Sleep Med. Rev.* **61**, 101566 (2022).
70. Makowski, D. et al. NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behav. Res. Methods* **53**, 1689–1696 (2021).
71. Tipping, M. & Bishop, C. Mixtures of probabilistic principal component analysers. *Neural Comput.* **11**, 443–482 (1999).
72. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
73. Healy, J. & McInnes, L. Uniform manifold approximation and projection. *Nat. Rev. Methods Primers* **4**, 82 (2024).
74. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In *Proc. of the 9th Python in Science Conference* 92–96 (SciPy, 2010).
75. Kim, S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665–674 (2015).
76. Vallat, R. Pingouin: statistics in Python. *J. Open Source Softw.* **3**, 1026 (2018).
77. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
78. Gramfort, A. et al. MNE software for processing MEG and EEG data. *Neuroimage* **86**, 446–460 (2014).
79. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In *Proc. of the 31st International Conference on Neural Information Processing Systems* <https://dl.acm.org/doi/pdf/10.5555/3294996.3295074> (Curran Associates, 2017).
80. Royston, P. & Parmar, M. K. B. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* **21**, 2175–2197 (2002).

Acknowledgements

We thank E. Litman from ZOLL Itamar for help on this project and all Segal laboratory group members for fruitful discussions, specifically N. Bar and L. Reicher for body systems definitions and inspiration for some figures and D. Krongauz and A. Keshet for metabolic pathway processing. E.S. is supported by the Crown Human Genome Center, by D. L. Schwarz, J. N. Halpern and L. Steinberg and by grants funded by the European Research Council and the Israel Science Foundation.

Author contributions

S.K. conceived the project, designed and conducted all analyses, interpreted the results and wrote the manuscript. A.D. performed the pulse rate variability calculations and reviewed the manuscript. N.G. performed data processing. A.W. developed protocols and oversaw sample collection and processing. Y.R. and R.D. interpreted the results and reviewed the manuscript. H.R. reviewed the results and the manuscript. E.S. conceived and directed the project and analyses, designed the analyses and reviewed the results and the manuscript.

Competing interests

A.D., Y.R. and H.R. are employees of Pheno.AI, Ltd., a biomedical data science company in Tel Aviv, Israel. A.W. and E.S. are paid consultants of Pheno.AI, Ltd. Other authors declare no competing interests.

Additional information

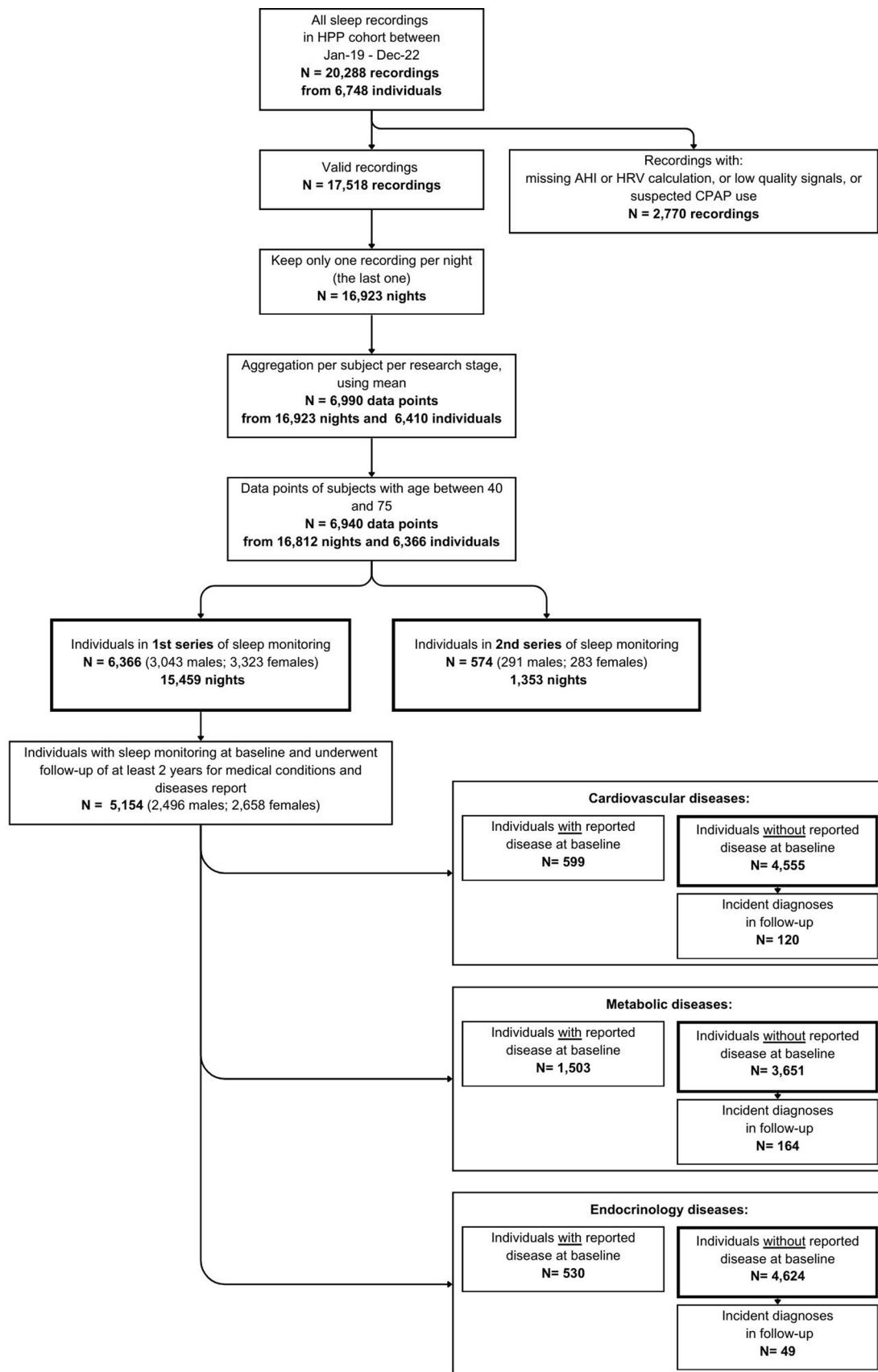
Extended data is available for this paper at
<https://doi.org/10.1038/s41591-024-03481-x>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-024-03481-x>.

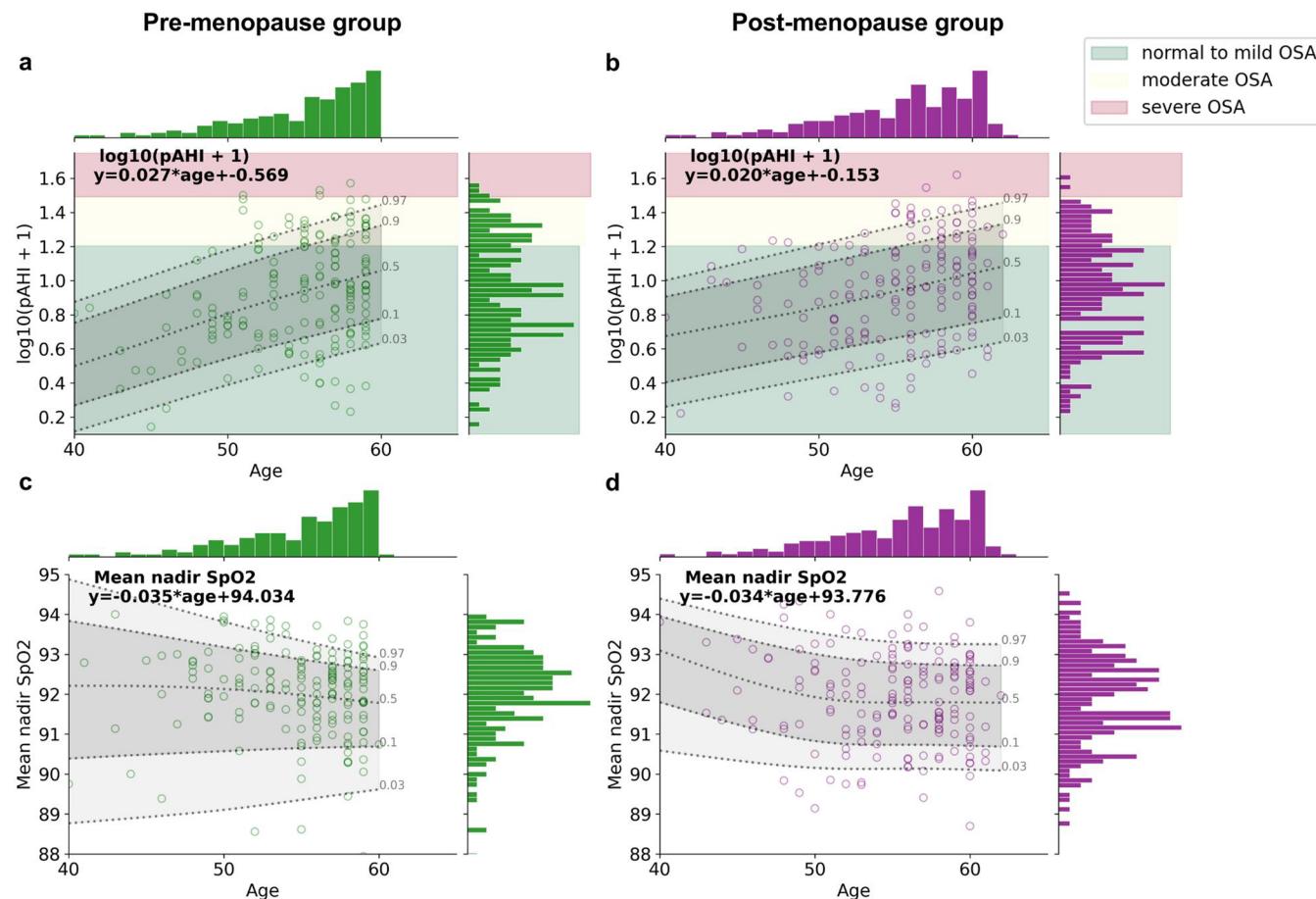
Correspondence and requests for materials should be addressed to Eran Segal.

Peer review information *Nature Medicine* thanks Neomi Shah, Diego Mazzotti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Michael Basson, in collaboration with the *Nature Medicine* team.

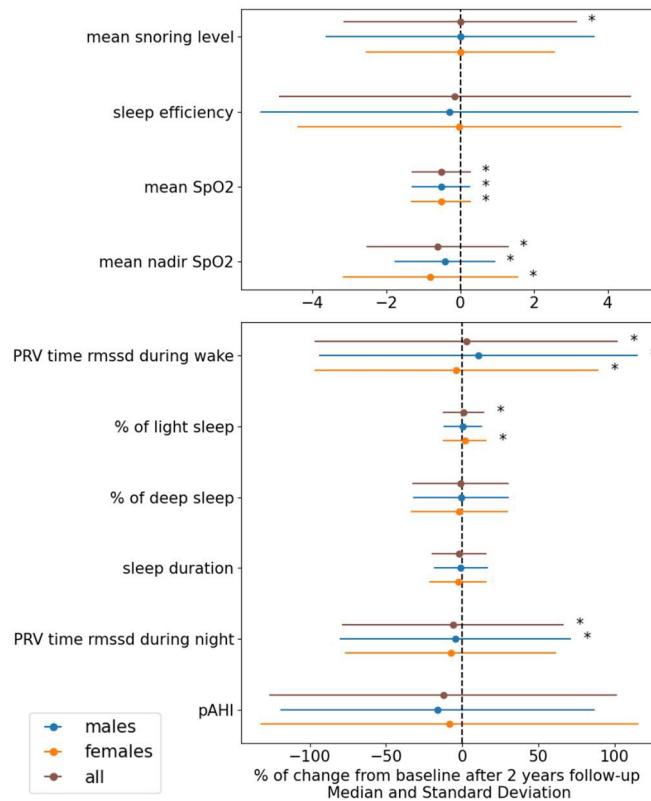
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Data Selection and sample sizes. Flow chart showing the different sample sizes of the data available at the time of this work, after selection process, per research stage, and stratified by self-reported cardiovascular, metabolic and endocrine diseases.

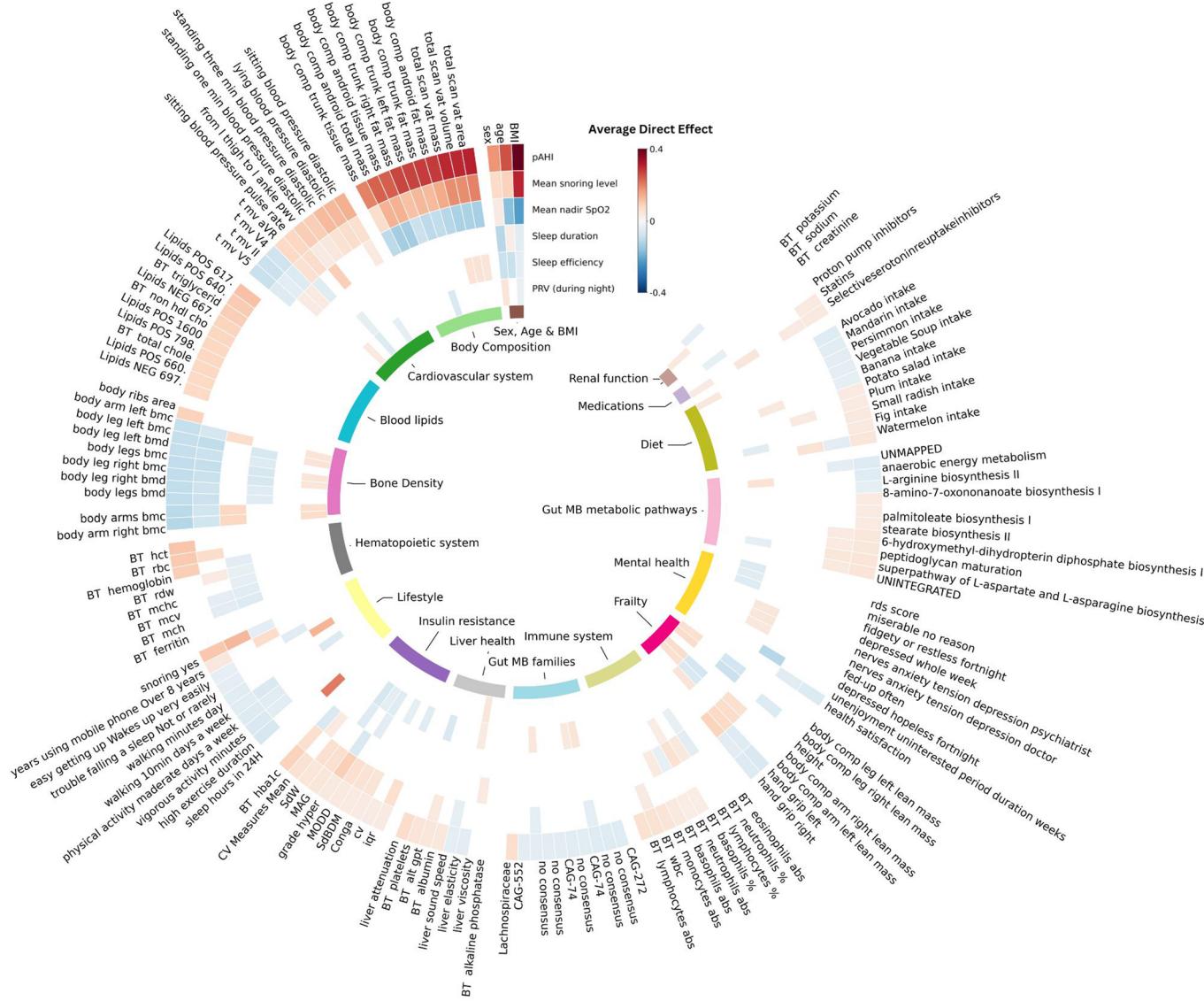


(green) and postmenopausal females (purple). For all panels (a, b, c, d), the robust linear regression equation is shown in the top left panel for each graph. The 3rd, 10th, 50th, 90th, and 97th percentiles obtained using Lowess regression are shown in dotted black lines on each graph. The histograms of the x and y axis values are shown on the top and right of each graph respectively. OSA - Obstructive Sleep Apnea. pAHI - Peripheral Apnea-Hypopnea Index.



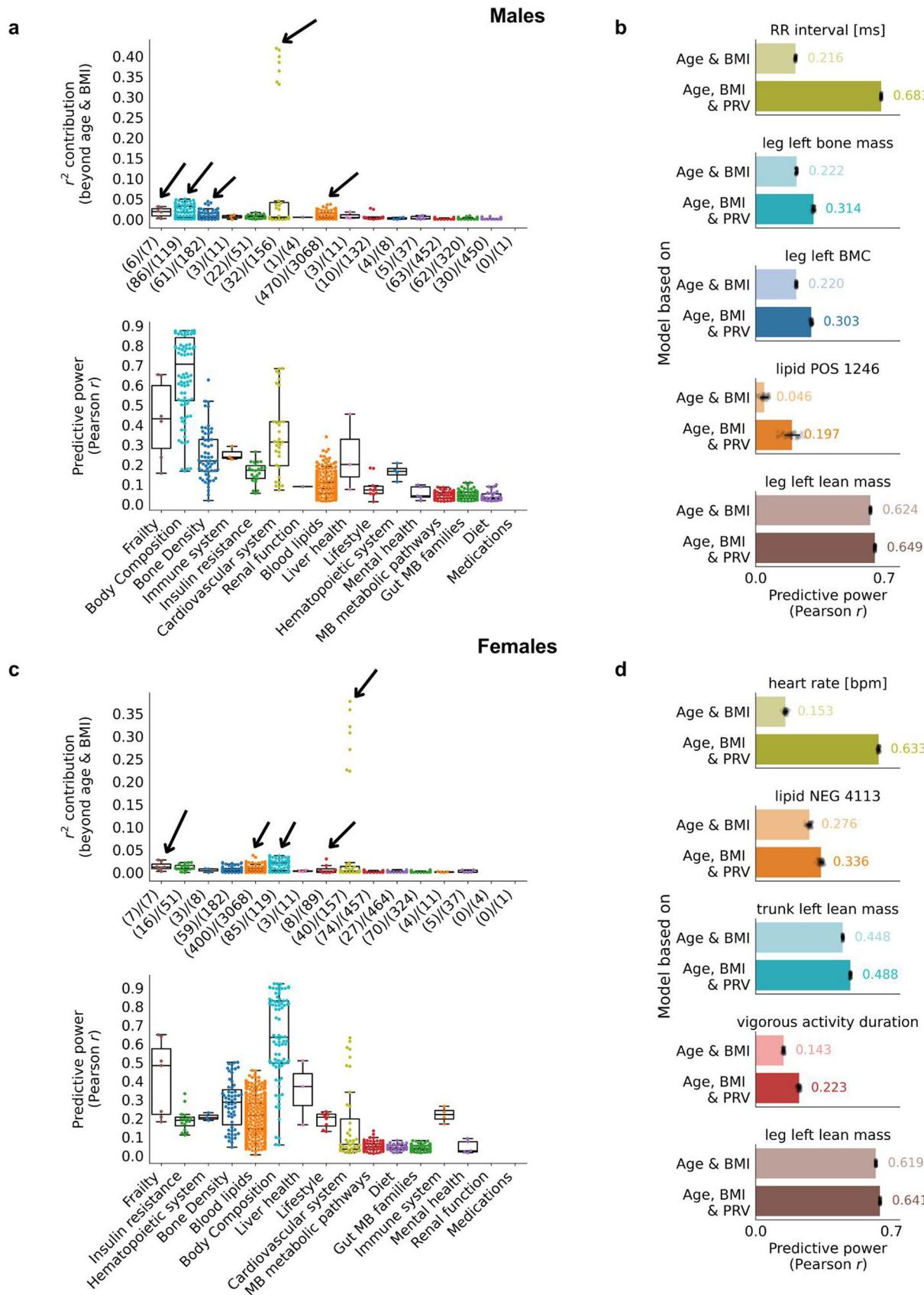
Extended Data Fig. 3 | Trends in sleep characteristics when measured 2 years apart in the same individual, as percent change in each variable between baseline and follow up. Forest plot showing the distributions (dot, median; error bars, SD) of the percentage of change in the sleep measurements between the two series of sleep monitoring undergone by each participant (n = 574, Methods),

across males (blue), females (orange) and all participants (brown), with asterisks (*) marking significant difference from zero with (P value < 0.01 in one sample two-sided t-test). pAHI - peripheral Apnea-Hypopnea Index. PRV - Pulse Rate Variability. RMSSD - Root Mean Square of Successive Differences between normal heartbeats. SD - Standard Deviation.



Extended Data Fig. 4 | Direct effect of features derived from body systems on key sleep-derived features. The circular heatmap shows the top 10 features significantly associated with pAHI for each body system (body systems names are indicated in the center of the figure), using mediation analysis with age, sex and BMI as mediators. Each slice of the heatmap represents a single body system feature, the name of which is indicated at the outer layer of the heatmap. The color code, indicated at the top of the heatmap, indicates the average direct effect (ADE) on pAHI, mean snoring level, mean SpO₂, sleep time, sleep efficiency

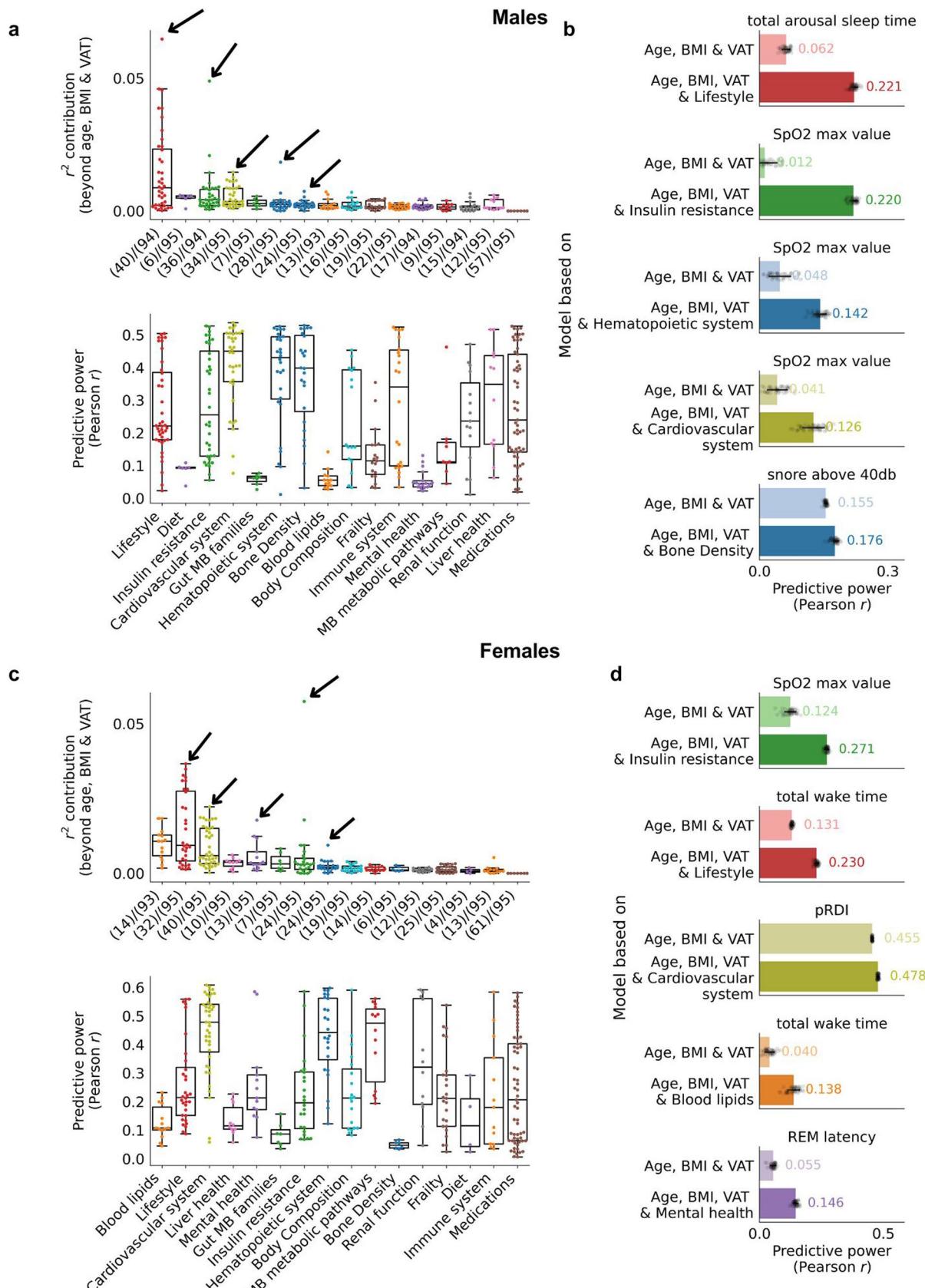
and PRV during night, respectively from the outermost to innermost layer. Direct effects that were not significant after FDR correction are shown as blank white cells. The features are grouped by their corresponding body systems they are derived from, as indicated in the innermost layer, and are ranked in an inverse clockwise order from the greater to the lowest direct effect on pAHI (based on the absolute value of the ADE). pAHI - Peripheral Apnea Hypopnea Index. SpO₂ - Oxygen Saturation. BMI - Body Mass Index. BT - Blood test. MB - Microbiome. ARB - Angiotensin receptor blockers. ACE - Angiotensin-Converting Enzyme.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Body system's features predictions based on PRV measurements. (a, c). Box and swarm plots (center, median; box, interquartile range (IQR); whiskers, $1.5 \times$ IQR) comparing the performance of models based on PRV measurements averaged over 3 nights in predicting body system features, as compared to the performance of a baseline model, based on age, and BMI, for males (a) and females (c). Predictive power of each model was evaluated using 5-fold cross-validation, repeated over 50 iterations (each using a different random seed), reporting the Pearson correlation r between predicted and actual values for the held-out set in each iteration. In (a, c), the lower graphs show the median predictive power obtained from these iterations for the models based on PRV data while the upper graphs show the added variance that was explained (difference in r^2) using the models based on PRV data, when compared to the baseline model. The number of features with significantly improved predictions using the PRV measurements (significant difference from the performance of

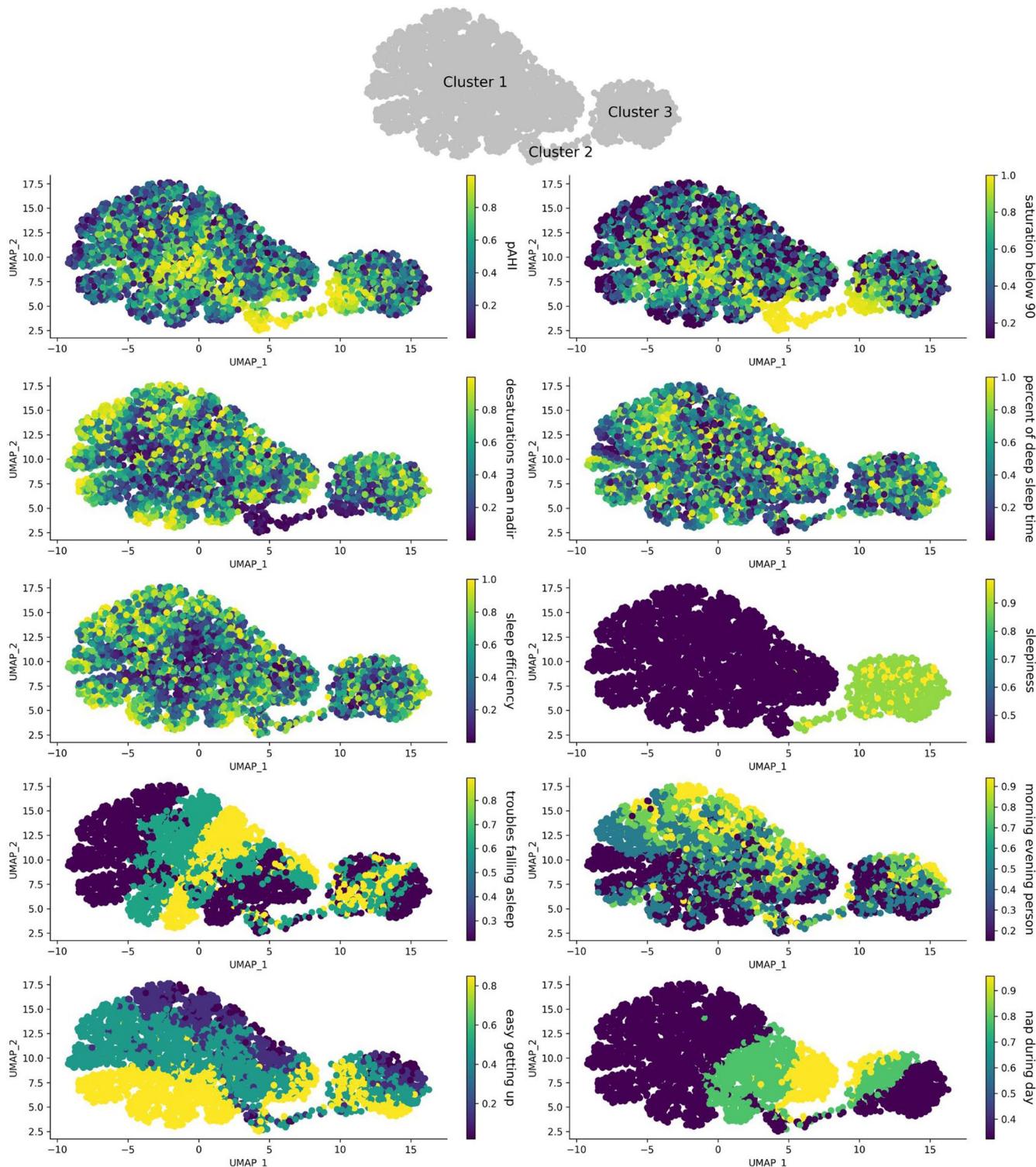
a baseline model, with P value < 0.001 in a two-sided t-test), which corresponds to the number of dots (n) included in these distributions, are shown for each body system in brackets on the x-axis as (n / total number of features in the body system). Features are grouped by body system and the body systems are ordered from left to right along the x-axis in descending order of the median explained variance (box center) shown in the upper graphs of (a) and (c). (b, d). Bar and dot plots ($n = 50$, center, mean; error bar, SD) comparing the performances (each of the cross-validation r is plotted as a dot) of models based on age and BMI versus corresponding models based on age, BMI and PRV measurements averaged over 3 nights for predicting the body system measurements marked with an arrow in plots (a) and (c) respectively, for males (b) and females (d). The mean predictive power value of each model is annotated on the right of each bar. BMC - Bone Mineral Content. BMI - Body mass index. PRV - Pulse Rate variability. PWV - Pulse Wave Velocity. SD - Standard Deviation.



Extended Data Fig. 6 | See next page for caption.

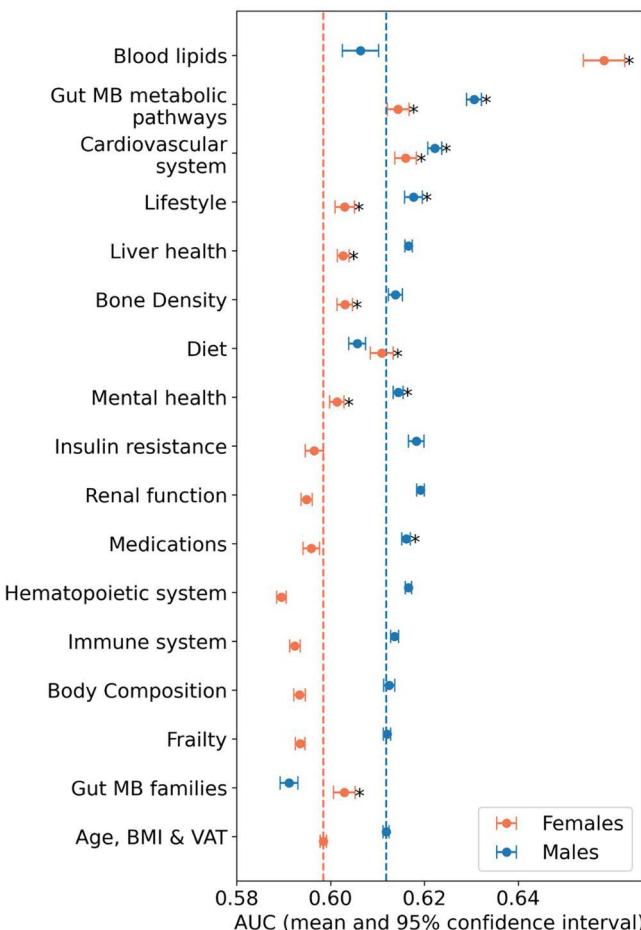
Extended Data Fig. 6 | Sleep test measurements predictions based on body systems features. (a, c). Box and swarm plots (center, median; box, interquartile range (IQR); whiskers, $1.5 \times$ IQR) comparing the performance of models based on each body system in predicting sleep test measurements, as compared to the performance of a baseline model, based on age, BMI and VAT, for males (a) and females (c). Predictive power of each model was evaluated using 5-fold cross-validation, repeated over 50 iterations (each using a different random seed), reporting the Pearson correlation r between predicted and actual values for the held-out set in each iteration. In (a, c), the lower graphs show the median predictive power obtained from these iterations for the models based on body system data while the upper graphs show the added variance that was explained (difference in r^2) using the models based on each body system data, when compared to the baseline model. The number of features with significantly improved predictions using a body system dataset (significant difference from

the performance of a baseline model, with P value < 0.001 in a two-sided t-test), which corresponds to the number of dots (n) included in these distributions, are shown in brackets on the x-axis as (n / total number of sleep test measurements). Features are grouped by the body system they were predicted from and the body systems are ordered from left to right along the x-axis in descending order of the median explained variance (box center) shown in the upper graphs of (a) and (c). (b, d). Bar and dot plots ($n = 50$, center, mean; error bar, SD) comparing the performances (each of the cross-validation r is plotted as a dot) of models based on age, BMI and VAT versus corresponding models based on age, BMI, VAT and a body system for predicting the sleep test measurements marked with an arrow in plots (a) and (c) respectively, for males (b) and females (d). The mean predictive power value of each model is annotated on the right of each bar. BMI - Body mass index. REM - Rapid Eye Movement. SD - Standard Deviation. SpO₂ - Oxygen Saturation. VAT - Visceral Adipose Tissue.



Extended Data Fig. 7 | 2D projection of 5,098 individuals' sleep traits revealing relationships between sleep traits, self-reported symptoms and habits. The population of the HPP cohort described by their objective sleep monitoring measurements (pAHI, desaturations severity, sleep efficiency and percentage of deep sleep), self-reported sleep related symptoms (daytime sleepiness, difficulty falling asleep or waking up) and sleep habits (being a morning/evening person or napping during the day), were mapped down to 2D using UMAP, and colored by sleep traits, for better visualization of the sleep traits similarity and clustering.

Three main clusters emerged: Cluster 1 represents individuals without symptoms of daytime sleepiness, while Cluster 3 includes those reporting daytime sleepiness. Cluster 2 encompasses individuals with elevated pAHI and deeper desaturations, predominantly combined with daytime sleepiness, though not exclusively. Notably, characteristics such as sleep efficiency, deep sleep percentage, and sleep habits were evenly distributed across all clusters. pAHI - peripheral Apnea-Hypopnea Index. UMAP - Uniform Manifold Approximation and Projection.



Extended Data Fig. 8 | Body systems associations with clinical OSA. Forest plot ($n = 50$) comparing the performance of sixteen models, each based on a different body system together with age, BMI and VAT, versus corresponding models based on age, BMI and VAT for predicting the estimated clinical OSA, for males (blue) and females (orange). Clinical OSA was estimated based on the pAHI measured during home sleep tests averaged over three nights (average pAHI > 15), combined with self-reported symptoms of excessive daytime sleepiness. Predictive power of each model was evaluated using 5-fold cross-validation,

repeated over 50 iterations (each using a different random seed), reporting the AUC for the held-out set in each iteration. The dots represent the mean AUC values, lines indicate the 95% confidence intervals, and asterisks (*) denote a significant difference from the performance of a classifier based solely on age, BMI, and VAT (P value < 0.001 in a two-sided t-test). The dotted lines indicate the AUC of a classifier based on only age, BMI and VAT, for males (blue) and females (orange). AUC - Area under the Receiver Operating Characteristic Curve. BMI - Body Mass index. OSA - Obstructive Sleep Apnea. VAT - Visceral Adipose Tissue.

Extended Data Table 1 | Datasets and sample sizes

Body system-level categories	Number of individuals	Number of samples	Number of features	Number of actual data points with available sleep characteristics, with age between 40 and 75		
				All	Males	Females
Blood lipids	6,164	6,164	3,103	2,805	1,347	1,458
Body composition	9,606	11,326	120	5,928	2,856	3,072
Bone density	9,606	11,327	183	5,937	2,862	3,075
Cardiovascular system	10,802	13,803	162	6,912	3,322	3,590
Diet	10832	13545	317	6,725	3,220	3,505
Frailty	9606	11326	8	5,937	2,862	3,075
Gut microbiome families	10470	13398	627	6,574	3,354	3,220
Gut microbiome metabolic pathways	10,073	12,548	535	6,391	3,109	3,282
Hematopoietic system	10579	13539	9	5,916	2,845	3,071
Immune system	10,575	13,556	12	5,943	2,857	3,086
Insulin resistance	7,577	7,577	53	4,002	1,894	2,108
Lifestyle	10602	15398	162	5,066	2,451	2,615
Liver health	9292	11171	12	5,839	2,802	3,037
Medications	11386	15347	61	6,940	3,335	3,605
Mental health	10511	15267	36	5,030	2,434	2,596
Renal function	9,019	10,349	5	4,924	2,384	2,540
Sex, age and BMI	11,545	16,720	3	6,940	3,334	3,606
Sleep characteristics	6,410	16,923	448	6,940	3,334	3,606
Baseline medical diagnosis	10,396	10,396	127	4,596	2,180	2,416

For each body system, the table shows the number of individuals, measurements and features available in the HPP cohort at the time of this work (left part of the table). The final number of data points (measurements per research stage) included in the analyses per body system—that is, for which at least one sleep test was available at the same research stage—is shown in the right part of the table for all male and female participants.

Extended Data Table 2 | Predictive model types

Body systems	Best model for predicting body characteristics based on sleep test measurements		Best model for predicting body characteristics based on PRV measurements		Best model for predicting sleep test measurements based on body systems	
	males	females	males	females	males	females
Blood lipids	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Linear (Lasso)	Non-linear (LightGBM)	Non-linear (LightGBM)
Body composition	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Bone density	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Non-linear (LightGBM)
Cardiovascular system	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Non-linear (LightGBM)	Linear (Lasso)	Linear (Lasso)
Diet	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Linear (Lasso)	Linear (Lasso)
Frailty	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Gut microbiome families	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)
Gut microbiome metabolic pathways	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Linear (Lasso)	Linear (Lasso)
Hematopoietic system	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Immune system	Non-linear (LightGBM)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Insulin resistance	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Lifestyle	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Liver health	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)	Linear (Lasso)
Medications				Linear (Lasso)	Non-linear (LightGBM)	
Mental health	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)	Linear (Lasso)
Renal function	Non-linear (LightGBM)	Non-linear (LightGBM)	Non-linear (LightGBM)			

For each task associating specific body system measurements with sleep characteristics (either sleep test measurements or PRV measurements), both linear and nonlinear models were tested. The table shows which model type performed better in each case, and this model was used for the figures and results presented in the manuscript. If, for a specific task, no significant association was found, the corresponding cell was colored in black in the table.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No data collection software was used

Data analysis Python 3.9, with packages: lifelines 0.29.0, lightgbm 4.5.0, mne 1.8.0., pingouin 0.5.5, scikit-learn 1.5.1, scipy 1.11.4, statsmodels 0.14.2, umap 0.1.1 and umap-learn 0.5.5, <https://github.com/SarahKohn/SleepAssociationsHPP.AutoMorph> software, <https://github.com/HullLab/AutoMorph>, and neurokit2 0.2.3 were used with Python 3.10.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data in this paper is part of the Human Phenotype Project (HPP) and is accessible to researchers from universities and other research institutions at: <https://humanphenotypeproject.org/data-access>

The HPP data includes personal information and, in compliance with IRB regulations, cannot be made publicly available. Interested bona fide researchers should contact info@pheno.ai to obtain instructions for accessing the data, which is typically granted within a few days.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Study population included 3,323 women and 3,043 men. Participants' sex was determined based on self-reporting. Participants were not asked about their gender, and the terms 'men' and 'women', if used in the manuscript, are used to describe human males and females, accordingly. In most of the analyses performed in the manuscript, men and women were analyzed separately. In case it was not, for example for the pairwise correlations, sex was used as a covariate in the analysis.

Reporting on race, ethnicity, or other socially relevant groupings

No socially constructed or socially relevant categorization variable was used in the work described in this manuscript. Confounding variables, such as sex, age, and body mass index (BMI), were controlled in our analyses by including them as covariates in the predictive models, correlation analyses, or mediation analyses. For predictive models, the performance of these models was compared to that of corresponding models based solely on the confounding variables, to determine whether an association was present.

Population characteristics

Samples analyzed in this study were collected as part of the '10K project', also called the Human Phenotype Project (HPP), described in details in: Shilo, Smadar, et al. "10K: a large scale prospective longitudinal study in Israel" European journal of epidemiology 36.11 (2021): 1187-1194. In brief, participants are either healthy individuals (as defined in the above mentioned manuscript) aged 40-70 or participants from the previous study described in: Zeevi, David et al. "Personalized nutrition by prediction of glycemic response." Cell 163.5 (2015): 1079-1094, who chose to join the new cohort, and thus may be younger than 40 or older than 70. In total, the study population related in the manuscript included 6,366 participants aged 40 to 75.

Recruitment

The recruitment process relied on self-assignment of volunteers who register to the 10K trial website (<https://www.project10k.org.il/en>)

Ethics oversight

The 10K cohort is conducted according to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of the Weizmann Institute of Science.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Since this is the first study to associate sleep derived features with thousands features spanning on 16 other body systems, no prior knowledge on the expected effect sizes exists, and thus a sample size calculation was not feasible. Samples included in this study are of all the participants who met the inclusion criteria detailed in Shilo et al. (2021), with available data collected to date and with the exclusions mentioned below (the resulted available sample sizes for this work were detailed in Extended data table 1). The results reported in the manuscript are those that passed the statistical significance threshold following multiple hypotheses correction. It is likely that with greater sample size additional findings of smaller effect sizes could be detected.

Data exclusions

Sample exclusions were detailed in Extended data figure 1. Additionally heart rate measurements measured using the WatchPAT-300 device were excluded from the sleep derived features since very redundant with the ECG measurements (also mentioned in the manuscript Methods section).

Replication

Since the HPP cohort is the first dataset including such a deep phenotype data to allow the kind of analysis presented in the manuscript, replication on an independent cohort was not feasible. However, a lot of the associations found in this study were also observed in previous works as referenced and detailed in the Discussion section of the manuscript.

Randomization

There was no group allocation in this study.

Blinding

There was no group allocation in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study <input checked="" type="checkbox"/> Antibodies <input checked="" type="checkbox"/> Eukaryotic cell lines <input checked="" type="checkbox"/> Palaeontology and archaeology <input checked="" type="checkbox"/> Animals and other organisms <input checked="" type="checkbox"/> Clinical data <input checked="" type="checkbox"/> Dual use research of concern <input checked="" type="checkbox"/> Plants
-----	---

Methods

n/a	Involved in the study <input checked="" type="checkbox"/> ChIP-seq <input checked="" type="checkbox"/> Flow cytometry <input checked="" type="checkbox"/> MRI-based neuroimaging
-----	---

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration NCT05817734

Study protocol The full description of the study can be found in: Shilo, Smadar, et al. "10K: a large scale prospective longitudinal study in Israel" European journal of epidemiology 36.11 (2021): 1187-1194.

Data collection The trial is being held in Rehovot, Israel. The study started on 2018-10-01 and is expected to be completed on 2045-12-31.

Outcomes Development of medical conditions based on participant's self-reporting coded to ICD11.

Plants

Seed stocks No seed stocks or other plant material was used in this study.

Novel plant genotypes No seed stocks or other plant material was used in this study.

Authentication No seed stocks or other plant material was used in this study.