nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00964-y

There are amendments to this paper

Macro and micro sleep architecture and cognitive performance in older adults

Ina Djonlagic[1,2], Sara Mariani[2,3], Annette L. Fitzpatrick[4], Veerle M. G. T. H. Van Der Klei[3], Dayna A. Johnson[5], Alexis C. Wood[6], Teresa Seeman[7], Ha T. Nguyen[8,14], Michael J. Prerau[2,3], José A. Luchsinger[9], Joseph M. Dzierzewski[10], Stephen R. Rapp[11], Gregory J. Tranah[12], Kristine Yaffe[13], Katherine E. Burdick[2,3], Katie L. Stone[12], Susan Redline[1,2,3] and Shaun M. Purcell[2,3] ✉

We sought to determine which facets of sleep neurophysiology were most strongly linked to cognitive performance in 3,819 older adults from two independent cohorts, using whole-night electroencephalography. From over 150 objective sleep metrics, we identified 23 that predicted cognitive performance, and processing speed in particular, with effects that were broadly independent of gross changes in sleep quality and quantity. These metrics included rapid eye movement duration, features of the electroencephalography power spectra derived from multivariate analysis, and spindle and slow oscillation morphology and coupling. These metrics were further embedded within broader associative networks linking sleep with aging and cardiometabolic disease: individuals who, compared with similarly aged peers, had better cognitive performance tended to have profiles of sleep metrics more often seen in younger, healthier individuals. Taken together, our results point to multiple facets of sleep neurophysiology that track coherently with underlying, age-dependent determinants of cognitive and physical health trajectories in older adults.

Both sleep problems[1,2] and cognitive impairment[3,4] increase with age. As well as disruptions of the sleep–wake cycle, polysomnographic studies have revealed marked age-dependent changes in sleep macro architecture (including stage duration[5,6]), micro architecture (including spindle activity[7]) and rates of sleep-disordered breathing[8]. Given the growing body of literature that points to a major role of sleep in supporting cognitive processing[9–12], it is natural to ask whether changes in sleep may be associated with accelerated cognitive decline and impairment in older adults. A large (but not always consistent) literature has indeed addressed these issues from several perspectives, from large epidemiological studies of self-reported sleep to focused laboratory studies of sleep-dependent memory consolidation and experimental sleep deprivation[13]. Observational studies have reported various aspects of sleep to be associated with cognitive performance, mild cognitive impairment and/or incident dementia in older adults, including sleep duration[14–21] and quality[22,23], excessive daytime sleepiness[15,24–26], elements of non-rapid eye movement (NREM)[27–31] and rapid eye movement (REM)[28,32–36] sleep neurophysiology, and sleep-disordered breathing[37]. However, many of the studies based on objective sleep measures suffered from small sample sizes and inadequate control of multiple testing, making it challenging to draw robust conclusions about population-level links between sleep and cognition.

In this study, we sought to identify the aspects of sleep neurophysiology most strongly associated with cognitive performance in older adults. Our data-driven approach aimed to assess a broad range of sleep measures, coupled with large sample sizes, control of multiple testing and replication across cohorts to ensure rigour. The initial cohort was the Multi-Ethnic Study of Atherosclerosis (MESA)[38,39]—an ethnically diverse population-based sample of late-middle aged and elderly adults recruited from six sites in the United States. The MESA Sleep ancillary study collected unattended level 2 polysomnography (PSG) and subjective reports of sleep quality on a subset of this sample[40]. We extended analyses to a second, independent cohort, the Osteoporotic Fractures in Men Study (MrOS)—a study of community-dwelling men 65 years or older who underwent level 2 PSG.

For details of the sleep and cognitive measures assessed, see the Methods and Supplementary Methods. In brief, we scored each individual's sleep on 173 metrics in MESA (168 in MrOS) organized in 11 domains: (1) subjective sleep problems; (2) sleep macro architecture; (3) chronotype; (4) spectral power metrics; (5) alternative time- and frequency-domain metrics; (6) a data-driven modelling of power spectra, labelled principal spectral component (PSC) analysis; (7) spindle occurrence; (8) spindle morphology; (9) slow oscillation (SO) occurrence; (10) SO morphology; and (11) spindle–SO coupling. All objective PSG metrics (the primary focus of this report) were identically calculated in MESA and MrOS, whereas some self-report measures were not available in both studies. In both cohorts, neuropsychological measures were available that collectively indexed global cognitive functioning, processing speed, working memory, attention and psychomotor functioning. Specifically, MESA employed four measures: the Digit Symbol Coding Test (DSCT)[41]; Cognitive Abilities Screening Instrument (CASI)[42]; and Digit Span Test (forward (DS_F) and backward (DS_B))[43].

[1]Beth Israel Deaconess Medical Center, Boston, MA, USA. [2]Harvard Medical School, Boston, MA, USA. [3]Brigham and Women's Hospital, Boston, MA, USA. [4]University of Washington, Seattle, WA, USA. [5]Emory University, Atlanta, GA, USA. [6]USDA/ARS Children's Nutrition Center, Baylor College of Medicine, Houston, TX, USA. [7]University of California, Los Angeles, Los Angeles, CA, USA. [8]Wake Forest School of Medicine, Winston-Salem, NC, USA. [9]Columbia University Medical Center, New York, NY, USA. [10]Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA. [11]Wake Forest Baptist Medical Center, Winston-Salem, NC, USA. [12]California Pacific Medical Center Research Institute, San Francisco, CA, USA. [13]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [14]Deceased: Ha T. Nguyen. ✉e-mail: smpurcell@bwh.harvard.edu

MrOS employed three: Trails B[44]; a modified, expanded version of the Mini-Mental State Examination (3MS)[45]; and the Digit Vigilance Test (DVT)[46].

When considering age-dependent associations between sleep and cognition, a number of potential mediating or confounding factors should be borne in mind. Poor cardiometabolic health, including hypertension and diabetes, is associated with sleep architecture[47] as well as cognitive decline[48]. Similarly, depressed mood is associated with sleep disturbances[49] and considered a risk factor for dementia[50]. Common medications associated with these conditions, including antidepressants[51,52] and beta blockers[53], can also impact sleep. In some individuals, sleep apnoea could be an underlying factor that connects poor sleep with cardiometabolic and cardiovascular health[37]. In secondary analyses, we therefore considered potential mediating and confounding factors that have been associated with sleep and/or cognition.

Results

The final analytical MESA sample comprised 1,595 adults between 54 and 93 years of age, of whom 759 (48%) were male. The MrOS sample comprised 2,224 males between 67 and 96 years of age (Supplementary Fig. 1 and Supplementary Tables 1 and 2). Sleep electroencephalography (EEG) signals were recorded centrally (C4/M1) in both studies. After first removing individuals with high levels of EEG artefacts, we set values that were statistical outliers as missing (Supplementary Table 3 and Supplementary Data 1 and 2). Below, all references to NREM sleep imply N2 + N3 combined. Unless otherwise explicitly noted, all statistics are regression coefficients (b) and corresponding significance values (P) from multiple linear regressions, always including at least a baseline set of covariates (age, sex in MESA, race/ethnicity and collection site) and typically with cognition as the dependent variable and a single sleep metric as the predictor.

As many of the objective sleep metrics have not been reported in these samples before, below we first describe the distributions and demographic associations of key metrics. We report associations between sleep metrics and cognitive performance, first in MESA and then expanding to MrOS. Finally, for the top cognition-associated objective sleep metrics, we characterize and contextualize their associations in a series of secondary analyses.

Demographic and other correlates of cognitive performance. As expected, cognitive performance was negatively associated with age in each cohort (Fig. 1a and Supplementary Tables 1 and 2). In MESA, DSCT ($b = -0.46$ s.d. units per decade of age; 95% confidence interval (CI) = −0.50 to −0.41; $P < 10^{-15}$) and CASI ($b = -0.26$; 95% CI = −0.31 to −0.21; $P < 10^{-15}$) showed stronger age-related declines in performance compared with DS_F ($b = -0.14$; 95% CI = −0.18 to −0.09; $P = 1 \times 10^{-8}$) and DS_B ($b = -0.16$; 95% CI = −0.21 to −0.11; $P = 3 \times 10^{-10}$). In MrOS, all measures showed a marked decline in performance with higher age, with $b = -0.63$ (95% CI = −0.70 to −0.55), −0.57 (95% CI = −0.65 to −0.50) and −0.37 (95% CI = −0.45 to −0.30) for Trails B*, 3MS and DVT*, respectively (all $P < 10^{-15}$; note that for timed measures, an asterisk denotes the sign-reversed metric, such that negative coefficients indicate worse performance consistently for all tests). Other demographic factors (including sex, race/ethnicity and educational attainment) and health factors (including depressive symptoms and cardiometabolic disease) showed associations with several cognitive measures (Supplementary Tables 1 and 2) and, along with collection site, these were included as covariates in subsequent secondary analyses.

Sleep macro architecture and spectral band power. On average, MESA participants had longer N1 sleep than MrOS participants (Fig. 1b; 47.7 versus 23.2 min) but shorter N2 sleep (206.4 versus 224.7 min), REM sleep (65.8 versus 71.2 min) and wake after sleep onset (WASO; 90.7 versus 112.5 min). Stage duration, expressed as a percentage of total sleep time (TST), showed a similar pattern of differences, except the relative proportion of N2 increased with age (Fig. 1b). Conditional on age and sex, MESA and MrOS showed broadly similar distributions for key stage-dependent micro-architecture metrics, including spectral power during N2 (Fig. 1c) and N3 (Fig. 1d), as well as spindle (Fig. 2b), SO (Fig. 2d) and spindle–SO coupling (Fig. 2b,e) metrics, with the exception of slow (<1 Hz) power (Fig. 1c,d) and SO duration (Fig. 2d), potentially reflecting differential high-pass filtering during acquisition or export of the original MrOS recordings. These minor differences in stage and slow power distributions notwithstanding, we note that our current focus is on within-cohort association. That is, as none of the analyses below directly combine measurements across cohorts, systematic between-cohort differences cannot, by definition, induce sleep–cognition associations spuriously.

Sleep spindles, SOs and their coupling. We targeted fast (centre frequency (F_C) = 15 Hz) and slow (F_C = 11 Hz) spindles in our primary analyses (Fig. 2a and Supplementary Figs. 2 and 3). On average, we detected 1.90 slow spindles and 2.46 fast spindles per minute of N2 sleep in MESA. In MrOS, these values were 1.75 and 2.03, respectively. Fast and slow spindle density estimates were significantly correlated with each other, although the majority (~90%) of variation in spindle density was unique to either fast or slow spindles, rather than shared across types: adjusted for baseline covariates, $r = 0.33$ (95% CI = 0.29 to 0.38; $P < 10^{-15}$) in MESA and $r = 0.32$ (95% CI = 0.28 to 0.36; $P < 10^{-15}$) in MrOS.

We detected SOs based on two complementary heuristics, using either an adaptive or an absolute amplitude threshold. Metrics of SO morphology were similar and highly correlated between threshold definitions, and similar between cohorts and stages (Supplementary Table 4 and Supplementary Fig. 4). For example, the mean durations of SOs detected during N3 only varied between 1.24 s (0.81 Hz) and 1.28 s (0.78 Hz) across definitions and cohorts. During N2 + N3 sleep combined, the SO frequency was ~0.85 Hz under the absolute definition but above 0.9 Hz under the adaptive definition (Supplementary Table 4). However, estimates of SO occurrence varied more between the two definitions (Supplementary Table 4). For MESA and MrOS, respectively, under the absolute definition, there were 7.50 and 4.25 SOs per minute during N3, compared with 1.95 and 1.02 during N2 + N3 combined. In contrast, under the adaptive definition, which as expected reduces between-cohort differences somewhat, there were 2.83 and 2.90 SOs per minute during N3 sleep, which rose to 4.26 and 3.96 during N2 + N3 combined. That the SO density counterintuitively increased when N2 sleep was included reflects the implicit lowering of the adaptive threshold (see the Supplementary Methods for a discussion of the complementary features and interpretative challenges of adaptive and absolute thresholds).

With respect to spindle–SO coupling, we observed a clear clustering of individuals' mean phase angles, whereby fast spindles tended to have their peak on the rising slope of SO positive peaks (labelled 'P' in Fig. 2c), in both MESA and MrOS. In contrast, slow spindles tended to peak on or after the positive SO peak. For coupling magnitude metrics that were dependent on SO definitions, we observed generally high correlations and similar means between adaptive (the default) and absolute SO definitions (Supplementary Table 5). We also considered spindle coupling with respect to continuous slow-wave activity (SWA), irrespective of detected SOs (called spindle–SWA coupling below). When considering spindle–SWA coupling, 80% of MESA individuals had a significant (empirical $P < 0.05$; see Methods for details) fast spindle–SO phase coupling metric, whereas 52% did for slow spindles (Supplementary Table 6). In MrOS, these proportions were 76 and 41% for fast and slow

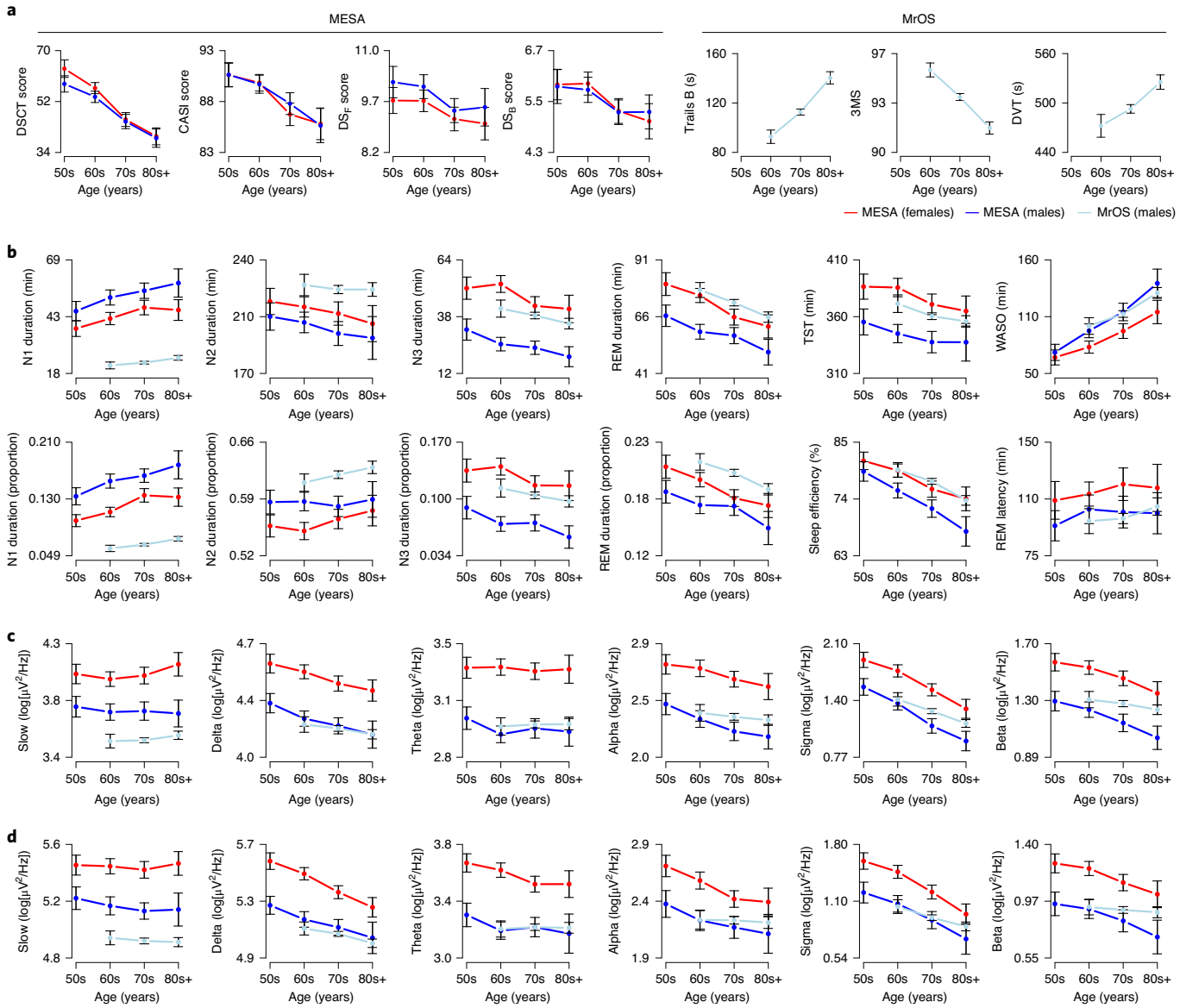


Fig. 1 | Demographic profiles of primary sleep and cognitive measures. a, Mean cognitive measure values in MESA, stratified by decade of age and sex, for DSCT, CASI, DS_F and DS_B (left); and mean cognitive measure values in MrOS, stratified by decade of age, for Trails B, 3MS and DVT (right). Note that here we present the raw values of Trails B and DVT, but association results are given for Trails B* and DVT*, which are those variables multiplied by −1. **b**, Key sleep macro architecture mean values stratified by decade of age, study and sex for N1, N2, N3 and REM duration (in minutes and as a proportion of TST), TST, WASO, sleep efficiency and REM latency. **c**, Absolute log-scaled N2 spectral band power, stratified by age decade, study and sex, for slow (<1 Hz), delta (1–4 Hz), theta (4–8 Hz), alpha (8–11 Hz), sigma (11–15 Hz) and beta (15–30 Hz) power, each divided by the total power to yield relative power metrics. **d**, As for **c**, but for N3 sleep. Error bars represent 95% CIs around the estimate of the mean.

spindles, respectively. We observed similar patterns of spindle–SO coupling, considering only the subset of spindles that overlapped a detected SO, although proportions tended to be higher for the adaptive versus absolute SO definition (Supplementary Table 6). In terms of spindle–SO gross overlap, we observed above-chance (empirical $P > 0.05$) coupling in approximately 40–60% of individuals for both fast and slow spindles, based on the adaptive SO definition (the default in our primary spindle–SO coupling analyses). Again, proportions tended to be lower when using the absolute definition (Supplementary Table 6).

Although not included as a metric in our primary analyses, to assess the performance of asymptotic significance tests of non-random spindle–SO and spindle–SWA coupling, for each individual, we calculated the proportion of shuffled replicates that were nominally significant at $P < 0.05$. Expected to be approximately 5% under the null, we observed many individuals for whom this proportion was at least doubled; in both MESA and MrOS, the median proportion of significant nulls could be as high as 10% for fast spindles, and 5% of individuals had type I error rates over 20% (Supplementary Table 6). These results underscore the potential for bias when relying on certain asymptotic statistical tests of coupling, and the need for robust approaches including non-parametric, empirical methods, as employed here.

PSC analysis. We also applied a multivariate statistical approach (PSC analysis) to capture some of the complex and inter-related features that biological signals often show (for example, between spindle and other rhythmic activity)[54]. Whereas standard measures

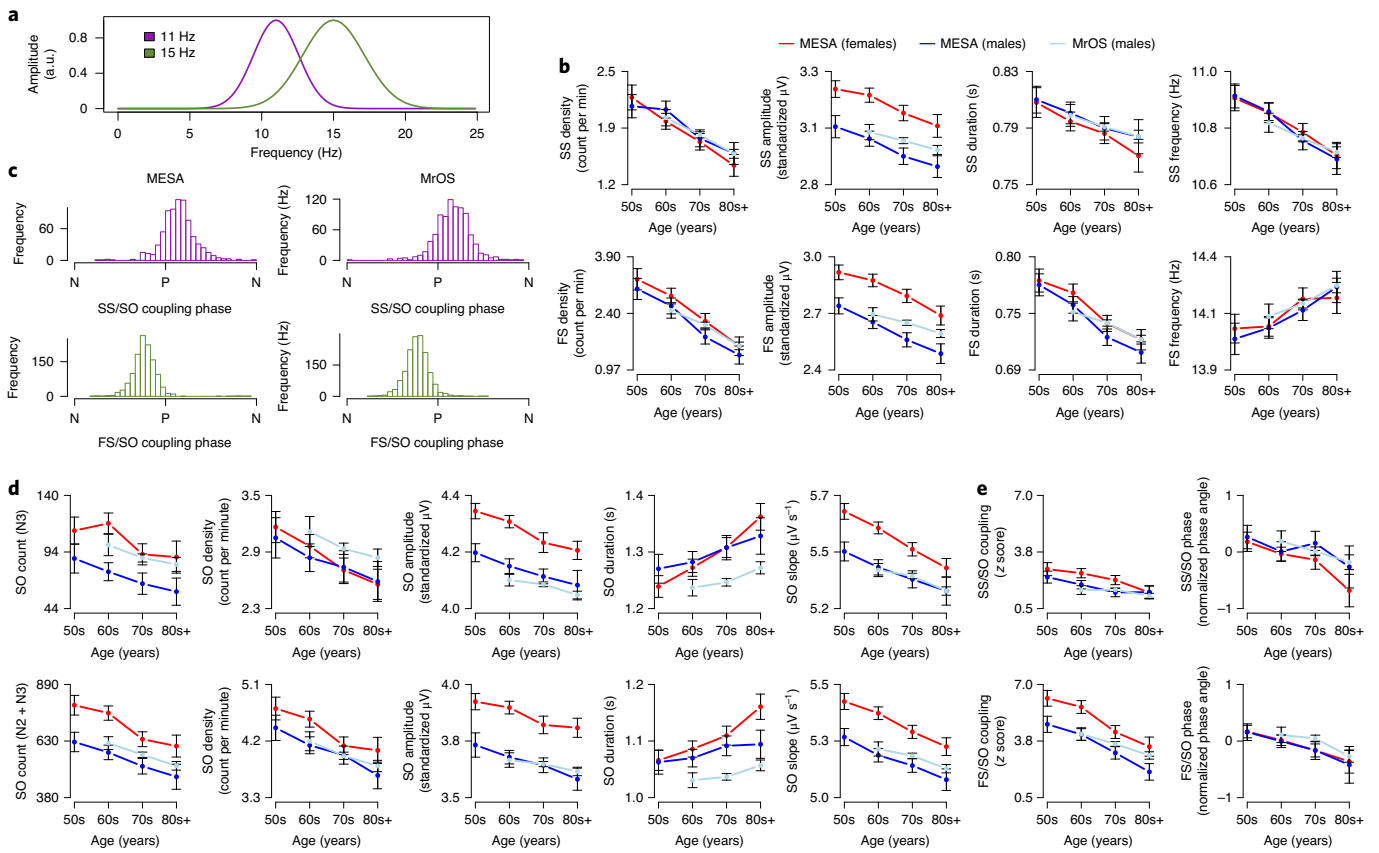


Fig. 2 | Spindles, SOs and their coupling. a, Spindle wavelet amplitude when using the default bandwidth setting (seven cycles), for $F_C = 11$ Hz and $F_C = 15$ Hz wavelets, where each wavelet detects spindles primarily within approximately ±2 Hz of the target frequency. **b**, Key N2 spindle metrics, stratified by age (decade), cohort and sex, for spindle density, amplitude, duration and frequency, separately for slow spindles (SS; top row) and fast spindles (FS; bottom row). **c**, Distributions of individuals' mean SO phase at the spindle peak for slow spindles (top) and fast spindles (bottom) in MESA (left) and MrOS (right). The SO phase angle is oriented on the *x* axis from one negative peak (N) to the subsequent one. Slow (fast) spindles tend to cluster just after (just before) the positive SO peak (P). **d**, Key SO occurrence and morphology mean values, stratified by age (decade), cohort and sex, for SO count, density, amplitude, duration and slope, in both N3 (top row) and N2 + N3 combined (bottom row). **e**, Spindle–SO coupling metrics (magnitude *Z* score and normalized SO angle) for slow (top) and fast spindles (bottom) in N2 + N3 sleep. Error bars in **b**, **d** and **e** represent 95% CIs around the estimate of the mean.

of power primarily focus on predefined frequency intervals one at a time (for example, 11–15 Hz 'sigma' power), PSCs can capture coordinated changes across the entire power spectra in a more flexible, data-driven manner, as well as providing an alternative to normalization based on relative power (see Methods). Applied to N2 and N3 sleep separately, PSC analysis reduced the joint variation across all 1,595 MESA power spectra to ten orthogonal components that explained over 99% of the total variance (Supplementary Fig. 5); similar curves were observed for MrOS (data not shown). Taking the ten largest components from analyses performed separately in MESA and MrOS, we observed a clear correspondence between the *k*th MESA component and the *k*th MrOS component (Fig. 3a,b). Broadly speaking, the first components (PSC1) captured individual differences in total power, whereas the second components reflected differences in the general steepness of the 1/*f* slope (Fig. 3a). The third and fourth components reflected individual differences yoked across alpha, sigma and beta bands (Supplementary Figs 6 and 7 show the equivalent plots for all N2 and N3 PSCs). Nonetheless, to ensure comparability between MESA and MrOS, our primary analysis used components based on a projection of the MESA PSC analysis to compute MrOS components. These MESA-derived MrOS components were highly similar to the components obtained from an MrOS-only analysis (Fig. 3b).

Within the same cohort, there was clear correspondence between the *k*th N2 component and the *k*th N3 component (Fig. 3c,d, top two plots). This pairwise correspondence of components did not arise obligatorily: for example, when applied to REM sleep spectra, fewer REM components strongly correlated with a single N2 component (Fig. 3c,d, bottom two plots), reflecting the greater degree of structured, stereotypical waveforms that occur during NREM but not REM sleep that are reflected by different PSCs. In contrast, that the top three components showed greater concordance across REM and NREM sleep might indicate a lower likelihood of reflecting neural oscillatory activity that is specific to sleep, as opposed to general EEG phenomena including sources of gross anatomical or technical variability.

Supplementary Fig. 8 shows all nominally ($P < 0.05$) significant correlations between N2 PSCs and other sleep metrics (after adjustment for age, sex, race and site). PSCs did not straightforwardly correspond to traditional band power metrics (either absolute or relative) in a one-to-one manner; different PSCs also showed qualitatively distinct patterns of correlation with spindle and SO metrics. For example, PSC4 correlated positively with both fast and slow spindle density, and positively with slow spindle frequency, but negatively with fast spindle frequency, underscoring the complex patterns of individual differences in the EEG not necessarily captured by standard metrics. In general, Supplementary Fig. 8 points

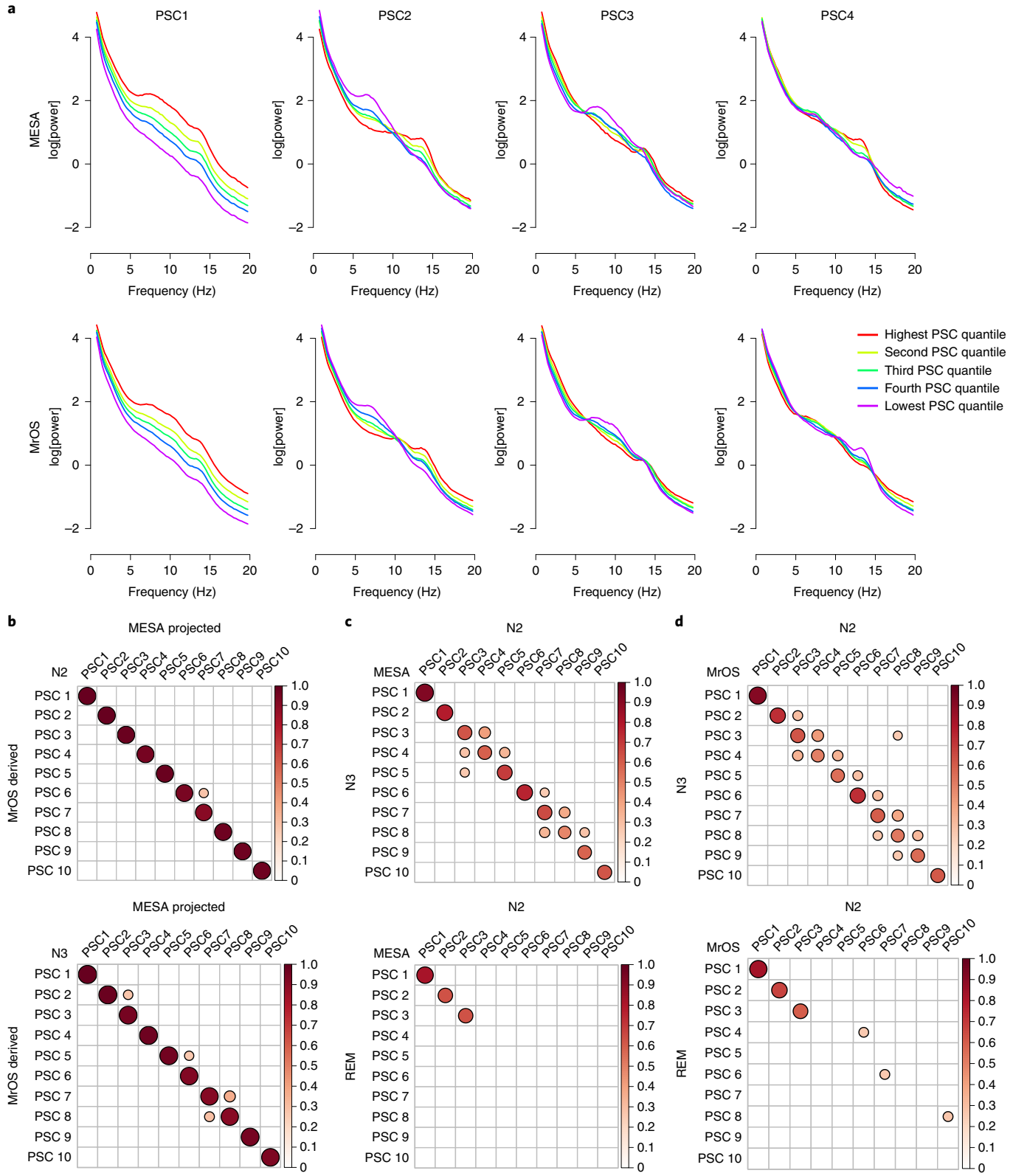


Fig. 3 | PSC analysis. a, Median absolute log-scaled power for five groups defined by the quintiles of the corresponding PSC during N2 sleep. Results are shown for the first four components, for MESA (top) and MrOS (bottom). See Supplementary Figs 6 and 7 for the first ten, in both N2 and N3 sleep. **b**, Absolute Pearson correlation coefficients from an intra-MrOS comparison of the top ten N2 (top) and N3 PSCs (bottom), comparing those derived from analysis within MrOS versus those projected from the MESA-derived components. The results show a very strong correspondence between the PSC structures in MESA and MrOS. **c**, Within-cohort, between-sleep-stage comparisons for N2 versus N3 (top row) and N2 versus REM (bottom row) PSCs for MESA, showing absolute Pearson correlation coefficients. Whereas N2 and N3 show a clear correspondence of PSCs, as ordered by the rank of singular values, REM only shows strong correspondence for the first three PSCs. **d**, Same as **c**, but showing the results from MrOS. For clarity of presentation, in all correlation plots (**b**–**d**), only correlations where $P < 10^{-10}$ and $|r| > 0.25$ are shown. The colour and size of each displayed point are proportional to the magnitude of the correlation coefficient; color bar indicates the absolute correlation coefficient, $|r|$.

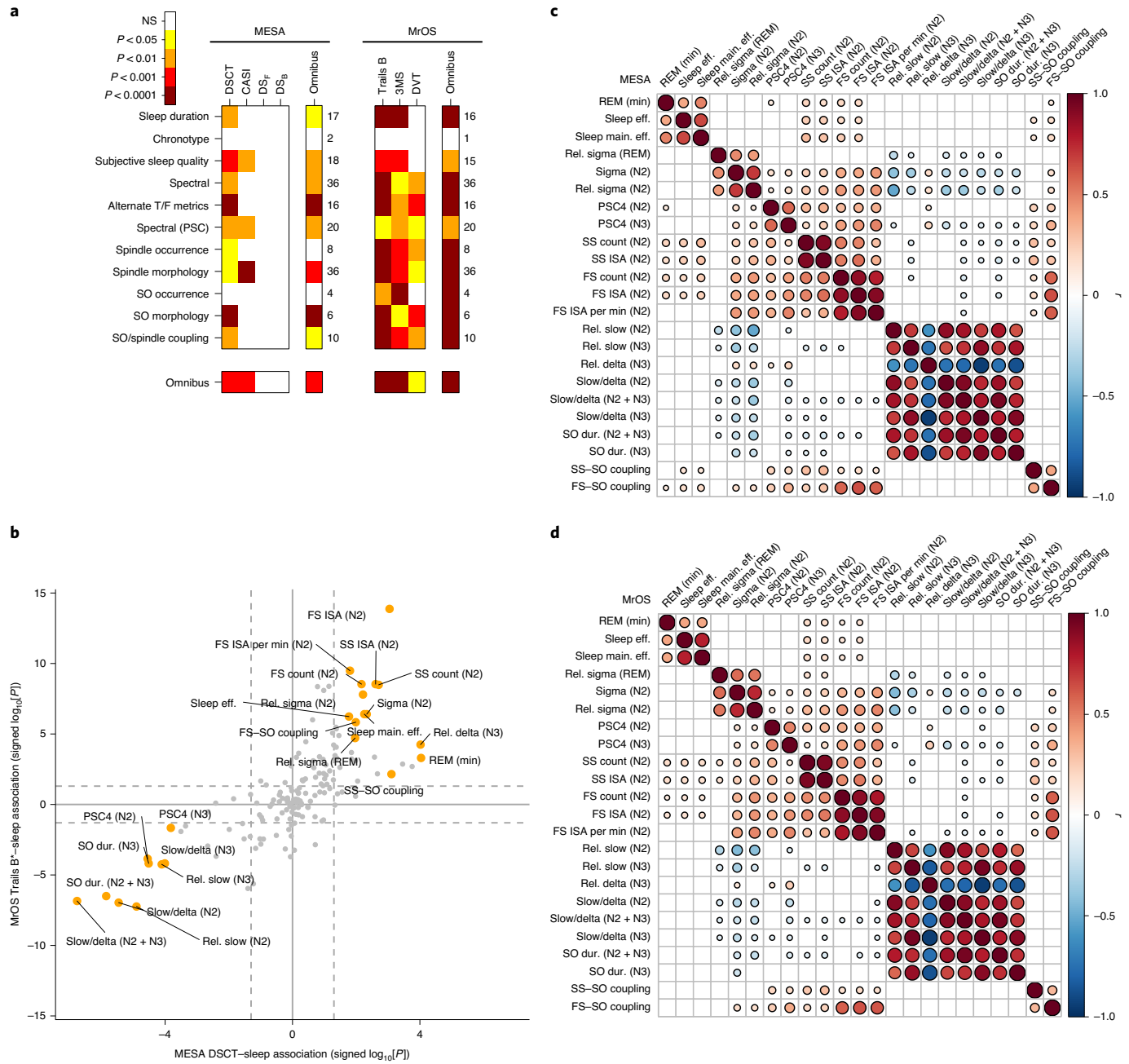


Fig. 4 | Primary sleep–cognition association results for omnibus tests and the 23 selected objective metrics, in MESA and MrOS. a, Domain-based omnibus results in MESA and MrOS. The colour-coded empirical significance values, based on 20,000 permutations, were calculated using the baseline covariate model and the max test statistic to control for multiple testing (see Methods for further details). NS, not significant. T/F metrics, time- and frequency-domain metrics. **b**, Correspondence of cognition–sleep associations in MESA and MrOS, showing $\log_{10}$-scaled *P* values, signed by the direction of effect, for DSCT in MESA (*x* axis) versus Trails B* in MrOS (*y* axis). The labelled orange points indicate the 23 metrics that replicated across cohorts. **c**, Correlation coefficients (Pearson's *r*) in MESA between key sleep metrics, adjusted for baseline covariates, and only showing cells where $P < 0.01$ and $|r| > 0.1$. Positive correlations are shown in red, negative correlations in blue; the size of each cell reflects the absolute magnitude of the correlation. **d**, As in **c**, but for MrOS. dur., duration; eff. efficiency; main., maintenance; rel., relative.

to PSCs as reflecting a complex but data-driven re-parameterization of dozens of inter-related sleep metrics, and one that was stable between MESA and MrOS.

Correlational structure of sleep metrics. As described below, we originally organized sleep metrics into 11 prespecified domains, simply to provide some higher-order structure for evaluating association results (Fig. 4a). An alternative approach, and one that may offer insight into how different aspects of sleep are inter-related, is to adopt a data-driven, empirical clustering of similar metrics. Based on distances derived from the correlation matrices for all metrics (Supplementary Fig. 9), we applied dimension reduction for visualization and clustering (uniform manifold approximation and projection (UMAP) and hierarchical density-based spatial clustering of applications with noise (HDBSCAN); see Methods) to MESA and MrOS separately. This analysis yielded ten clusters in MESA plus a

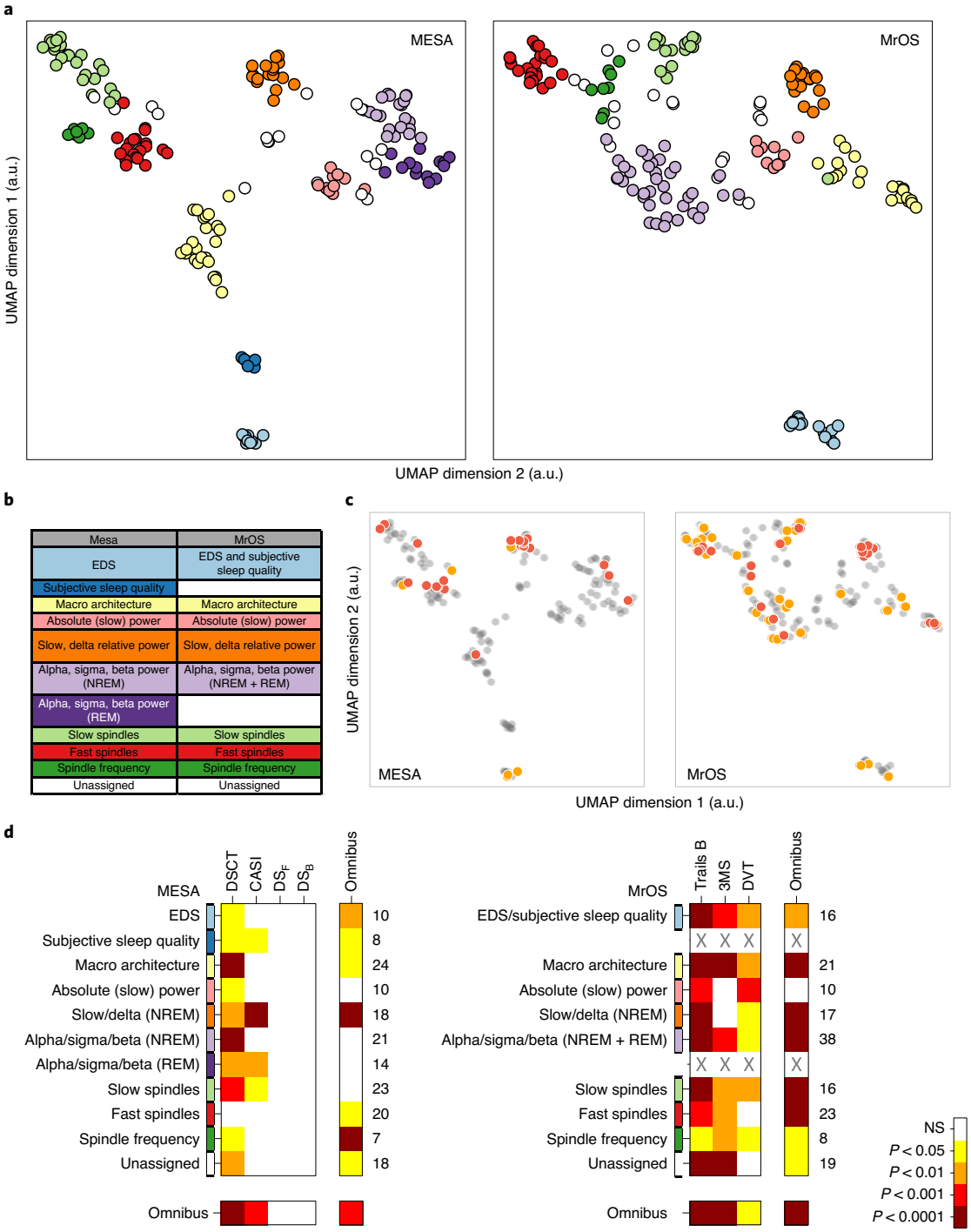


Fig. 5 | Empirical clustering of sleep metrics. a, Results from UMAP dimension reduction (based on $1-|r_{ij}|$, where r_{ij} is the Pearson correlation coefficient between the *i*th and *j*th sleep metrics), extracting two components for visualization, in MESA and MrOS separately. The colours correspond to the cluster assignment from HDBSCAN cluster analysis based on ten UMAP components (a.u., arbitrary units). **b**, Approximate labels of clusters from HDBSCAN analysis (see Supplementary Fig. 10b for a full tabulation). MESA yielded a ten-cluster solution, whereas MrOS yielded an eight-cluster solution. Clusters that are similar between MESA and MrOS (based on assigned metrics) are listed next to each other. EDS, excessive daytime sleepiness. **c**, Indication of cognition-associated sleep metrics in the UMAP two-dimensional projection of MESA and MrOS sleep metrics. Red indicates one of the 23 replicated hits. Orange indicates metrics with $P<10^{-3}$ for DSCT (MESA) or Trails B (MrOS). **d**, Omnibus association empirical significance values for groupings of sleep metrics based on data-driven clustering, for the max statistic.

set of metrics not confidently assigned to any one cluster, and eight clusters in MrOS plus a further unassigned set (Supplementary Fig. 10). We observed broadly comparable clustering between MESA and MrOS (Fig. 5a; note that the precise, global locations of where each cluster appears are largely arbitrary and should not be directly compared between plots; rather the local grouping of metrics is the more relevant aspect). We gave descriptive labels to clusters (Fig. 5b) based on the assigned metrics (Supplementary Fig. 11),

but note that these are necessarily approximations. Although pre-specified domains did not relate in a one-to-one manner with empirical clusters, there were clear points of correspondence (Supplementary Fig. 12). On the broadest level, subjective macro- and micro-architecture metrics formed largely distinct clusters (in particular, the independence of subjective and objective metrics was clear from the raw correlation matrices; Supplementary Fig. 9). In terms of micro-architecture metrics, perhaps the biggest difference between classifications was that fast and slow spindles were shown to be separable categories in both MESA and MrOS (Fig. 5a,b), in contrast with our primary partitioning by occurrence and morphology. The PSC metrics spanned a broad range of the space of micro-architecture metrics, as one might expect given the orthogonality inherent in their definition (Supplementary Fig. 11).

Although at the highest level these cluster solutions arguably have face validity, there are limitations inherent in these analyses (outlined below in the Discussion) and it would be a mistake to place too much weight on the specifics of any one solution. We therefore use these empirically derived clusters, as well as the original domains, to structure the omnibus association testing below.

Demographic correlates of sleep micro architecture. As others have reported[29,55–57], many sleep metrics showed marked age-related trends and sex differences. In general, objective sleep metrics showed larger demographic correlations compared with subjective measures. Age-related trends for objective sleep metrics were generally consistent between MESA and MrOS (Supplementary Fig. 13a). Of all of the metrics, fast spindle density (including the related metrics spindle count, integrated spindle activity (ISA) per minute (ISA_M) and total ISA (ISA_T); see Methods) showed the largest age-related changes, in both MESA ($b=-0.37$ s.d. units per decade; 95% CI = −0.42 to −0.32; $P<10^{-15}$) and MrOS ($b=-0.42$; 95% CI = −0.49 to −0.34; $P<10^{-15}$), although multiple slow spindle occurrence, sigma power and spindle morphology metrics also showed substantial age-related changes in both cohorts (Supplementary Data 1 and 2). For example, both fast and slow spindles during N2 sleep showed reduced chirp (intra-spindle change in frequency) with increasing age, in both MESA ($b=0.22$ (95% CI = 0.17 to 0.27; $P<10^{-15}$) and $b=0.12$ (95% CI = 0.06 to 0.17; $P=1\times10^{-5}$) for slow and fast spindles, respectively) and MrOS ($b=0.33$ (95% CI = 0.25 to 0.41; $P<10^{-15}$) and $b=0.17$ (95% CI = 0.09 to 0.25; $P=3\times10^{-5}$)). In all cases, the mean spindle chirp was negative, meaning that over the course of a single spindle, oscillations tended to slow down. The positive coefficients above indicate that older individuals tended to have spindles that slowed down less. A second age-related change in spindle morphology involved spindle frequency (Fig. 2b). Although fast and slow spindles were defined in terms of fixed-target F_C (11 or 15 Hz), there was still considerable person-to-person variability in the observed frequencies of spindles detected for a given F_C (for example, slow spindles may have had a mean frequency of 10.5 Hz in one individual versus 11.5 Hz in another). In both MESA and MrOS, with increasing age, slow spindles grew slower ($b=-0.28$ (95% CI = −0.34 to −0.23; $P<10^{-15}$) and $b=-0.26$ (95% CI = −0.34 to −0.18; $P=2\times10^{-10}$) for MESA and MrOS, respectively), whereas fast spindles grew faster ($b=+0.19$ (95% CI = 0.14 to 0.24; $P=3\times10^{-12}$) and $b=+0.23$ (95% CI = 0.16 to 0.31; $P=5\times10^{-9}$)). Similar directional effects were observed during N3 sleep and several PSCs exhibited large correlations of opposite signs between fast and slow spindle frequency, including PSC4, the component with the strongest age association in both MESA and MrOS (Supplementary Fig. 8).

Aside from spindle activity, a number of other sleep metrics showed marked age-related changes. In terms of sleep macro architecture, WASO increased with age in both MESA ($b=19.3$ min per decade; 95% CI = 16.3 to 22.2; $P<10^{-15}$) and MrOS ($b=21.0$; 95% CI = 16.3 to 25.7; $P<10^{-15}$), with corresponding decreases in sleep efficiency. Both MESA and MrOS showed age-related increases in N1 sleep (absolute duration in minutes), as well as decreases in N3 and REM sleep duration. We observed marked age-related reductions in the magnitude of spindle–SO coupling for both fast and slow spindles, as well as age-related shifts in the mean coupling phase angle, such that spindles occurred earlier in the SO in older individuals, in both MESA and MrOS (Supplementary Data 1 and 2). In contrast, although spindle–SO gross overlap decreased with age in MESA, there was no evidence for an association with age in MrOS (perhaps reflecting that coupling had already broken down for a greater proportion of individuals in the older MrOS sample; Supplementary Table 6).

Age-related changes in spectral power, including decreasing absolute sigma N2 power ($b=-0.33$ s.d. per decade (95% CI = −0.38 to −0.29; $P<10^{-15}$) and $b=-0.30$ (95% CI = −0.38 to −0.22; $P=4\times10^{-14}$) in MESA and MrOS, respectively) were generally statistically stronger in MESA despite the smaller sample size (potentially reflecting its greater age range). We observed age-related reductions in delta power (for example, $b=-0.21$ (95% CI = −0.27 to −0.15; $P=1\times10^{-11}$) for relative delta power during N3 in MESA and $b=-0.20$ (95% CI = −0.29 to −0.11; $P=6\times10^{-6}$) in MrOS, with broadly similar results for N2 and absolute power metrics too). In contrast, we observed qualitatively different results for slow (<1 Hz) power, which did not exhibit credible evidence for age-related decline during NREM sleep. In fact, in only one instance was absolute slow power significantly associated with age, in which case it increased with increasing age (absolute slow power during N2 in MrOS; $b=0.12$; 95% CI = 0.05 to 0.20; $P=0.002$). Relative slow power also tended to increase with increasing age (for example, $b=+0.19$ (95% CI = 0.13 to 0.24; $P=3\times10^{-12}$) for relative slow power during N2 in MESA and $b=+0.24$ (95% CI = 0.16 to 0.32; $P=1\times10^{-9}$) in MrOS).

In terms of sex differences within MESA, the most marked reflected males having generally lower absolute spectral power during both N2 and N3, across all power bands. This association was captured by a large difference in PSC1 ($b=-0.78$ s.d. units; 95% CI = −0.87 to −0.70; $P<10^{-15}$). Males also had shorter scored N3 ($b=-23.2$ min; 95% CI = −26.1 to −20.3; $P<10^{-15}$) and REM sleep ($b=-12.6$ min; 95% CI = −15.3 to −9.8; $P<10^{-15}$), as well as TST ($b=-37.1$ min; 95% CI = −44.5 to −29.8; $P<10^{-15}$). Consistent with our previous report in independent cohorts[7], we observed sex differences in spindle density for fast spindles ($b=-0.18$ s.d. units; 95% CI = −0.27 to −0.08; $P=2\times10^{-4}$) but failed to reject the null hypothesis of no sex differences for slow spindles ($b=0.07$; 95% CI = −0.02 to 0.17; $P=0.12$). A similar pattern generally held for most fast and slow spindle morphology metrics too. Fast spindles further showed a significant sex difference in spindle–SO gross overlap, with males showing reduced fast spindle overlap ($b=-0.31$ s.d. units; 95% CI = −0.40 to −0.21; $P<6\times10^{-10}$), whereas we did not observe credible evidence of such an effect for slow spindles ($b=-0.05$; 95% CI = −0.14 to 0.05; $P=0.35$). Finally, despite excellent power to detect relatively small effects (see Supplementary Methods) and the general tendency for males to have objectively worse sleep (in the sense of male/female differences more often corresponding to old/young rather than young/old differences for a given metric), a number of metrics showed qualitatively distinct patterns of age and sex differences. For example, whereas the most statistically significant sex difference in MESA was for higher N2 absolute theta power in females (0.81 s.d. units; 95% CI = 0.72 to 0.90; $P=10^{-64}$), there was no credible evidence for an association with age ($b=-0.002$; 95% CI = −0.006 to 0.003; $P=0.53$). In contrast, metrics including slow spindle density and frequency showed marked age-related change but did not show credible evidence for any sex differences. Finally, as alluded to above, other metrics showed marked changes with respect to both age and sex (for example, SO slope during N2 + N3 sleep). Future work will be needed to resolve the extent of overlap

in the mechanisms of age-related changes versus sex differences in sleep neurophysiology.

SO occurrence, morphology and coupling metrics showed generally similar profiles of demographic association under both adaptive and absolute threshold definitions (Supplementary Data 3 and 4), albeit with weaker and more variable age-related changes in MrOS compared with MESA. SO count, density, amplitude and slope all declined with increasing age, whereas SO duration increased, meaning that SOs became slower with age. However, threshold definition had a considerable impact on a handful of comparisons. Most prominently, in MESA, whereas males had markedly lower SO density during N3 compared with females ($b = -0.71$ s.d. units; 95% CI = −0.81 to −0.60; $P < 10^{-15}$) under the absolute threshold, there was no credible evidence for a similar effect under the adaptive definition ($b = -0.045$; 95% CI = −0.16 to 0.07; $P = 0.43$; Supplementary Data 3), precisely due to the within-individual normalization. In terms of precise spindle–SO phase coupling, all analyses (that is, whether for spindle–SWA coupling that included all spindles or spindle–SO coupling based on adaptive or absolute definitions that included only a subset of spindles, and whether for fast or slow spindles) tended to yield similar patterns of results—namely, that the magnitude of coupling decreased with increasing age, and spindles tended to peak earlier during the SO in older individuals.

Primary sleep–cognition analyses. *Statistical approach to association analysis.* As the sleep–cognition analyses involved hundreds of significance tests, we ensured robustness using two approaches: (1) control for multiple testing by use of a permutation-based omnibus association framework; and (2) testing for replication in an independent sample. Specifically, for each cohort, we generated 20,000 permuted datasets in which individuals' sleep metrics were randomly reassigned. In each null dataset, we refit all sleep–cognition regression models (for example, 692 in MESA), recording both the most significant result and average result (in actuality, the maximum and sum of the 692 log-scaled *P* values from the regressions). Considering all 20,000 randomized datasets, in this way, we estimated the maximum and typical results that would be expected to occur by chance alone, while controlling for the correlations between tests. For each of the two aggregate metrics, the proportion of the 20,000 randomized datasets with equal or larger values than originally observed gives an empirical omnibus *P* value, labelled below as max and sum (that is, average) tests. As well as controlling for multiple testing, we used this framework to contextualize broad patterns of association between sleep and cognitive performance, by further considering subsets of tests (for example, for one particular cognitive measure) rather than across all combinations of sleep metric and cognitive measure. With respect to replication, we initially applied this approach to MESA, and then considered replication in MrOS, at both the individual test and omnibus levels. Finally, we identified individual metrics showing consistently replicated patterns of association across both cohorts (see Methods and Supplementary Methods for more details).

Omnibus and domain-level analyses in MESA. In MESA, the global null hypothesis of no association with cognition for any sleep metrics was rejected under both max (omnibus empirical $P = 0.001$) and sum (omnibus empirical $P = 0.0002$) permutation-based omnibus statistics, adjusted for age, sex, race/ethnicity and field centre (site) (Fig. 4a and Supplementary Fig. 12a). Of the 11 sleep domains tested, seven were globally significant at an omnibus empirical *P* value of <0.05 under both statistics (Fig. 4a and Supplementary Fig. 12a): subjective sleep problems; sleep macro architecture; spectral and alternate time–frequency metrics; PSC metrics; SO morphology; and spindle–SO coupling. Additionally, spindle morphology had an omnibus empirical *P* value of 0.0002 for the max statistic, but an omnibus empirical *P* value of 0.08 for the sum statistic. In terms of cognitive measures, DSCT and CASI showed significant omnibus association across all sleep metrics (all omnibus empirical *P* values were <0.05 for the max and sum statistics), whereas neither digit span measure showed credible evidence for an association (DS_F and DS_B; all omnibus empirical *P* values were >0.15; Fig. 4a and Supplementary Fig. 12a). For the chronotype domain (which only comprised two metrics) and SO occurrence domain, we failed to reject the null hypothesis of omnibus association in MESA. However, the overwhelming picture, supported by a series of significant omnibus tests (Fig. 4a and Supplementary Fig. 12a)—from global to subsets of sleep domains and cognitive measures—points to statistically robust evidence for correlation between sleep metrics and cognition, after controlling for age, sex, race/ethnicity and site.

Metric-level replication in MrOS. Given omnibus-level associations across multiple domains, we next sought to identify and characterize associations at the level of individual sleep metrics. Guided by the omnibus tests, we identified 32 sleep metrics that were significant for either DSCT or CASI at $P < 10^{-2}$ (based on nominal, asymptotic baseline model *P* values) in MESA and present in the independent MrOS cohort for a replication attempt. Of the 32 significant metrics that spanned subjective, macro-architecture, spectral power, PSC, spindle and SO domains, 30 (94%) were nominally significant in MrOS for Trails B*, 3MS or DVT*, all of which had the same direction of effect (Supplementary Data 1 and 2). Overall, these results suggested a clear correspondence between associations in MESA and MrOS.

Omnibus and domain-level analyses in MrOS. Given the high rate of replication in MrOS, we next sought to characterize the patterns of association in MrOS more generally, adopting the same omnibus testing framework as was applied to MESA (Fig. 4a and Supplementary Fig. 12a). The global omnibus test was significant for both max (omnibus empirical $P = 5 \times 10^{-5}$, implying that none of 20,000 null replicates equalled or exceeded the observed statistic) and sum statistics (omnibus empirical $P = 5 \times 10^{-5}$). Omnibus tests for all three cognitive measures were significant ($P < 0.05$) for both statistics, and ten of the 11 sleep domains were significant ($P < 0.05$) for both statistics, the exception being the (single-item) chronotype domain (Fig. 4a and Supplementary Fig. 12a). That is, we observed unambiguous statistical support at the omnibus level to suggest that sleep metrics across multiple domains of sleep were associated with the three cognitive measures in MrOS. Overall, the results were statistically stronger in MrOS than MESA, potentially reflecting the larger sample size, different demographics (all male; on average, 7.9 years older) and/or different cognitive measures employed.

Omnibus cluster-level results in MESA and MrOS. Re-running the omnibus tests but with sleep metrics grouped by empirical cluster instead of prespecified domain, we observed broadly similar patterns of results (Fig. 5d and Supplementary Fig. 12b). For our primary test, based on max statistics, of the ten clusters and one unassigned set, all but one (fast spindles) were associated with DSCT in MESA. In MrOS, all nine groups were associated with Trails B (Fig. 5d). That is, consistent with the domain-based grouping, testing based on empirical clusters pointed to multiple, independent aspects of sleep macro and micro architecture as being associated with cognitive performance, rather than only one or two tightly interconnected subsets of sleep metrics.

Identifying consistently associated sleep metrics across MESA and MrOS. As suggested by the above replication analyses, it seemed probable that many true MrOS associations would also be present in MESA and vice versa. To represent our final set of cognition-associated sleep metrics, we therefore identified all metrics that were consistently and significantly associated across both

Table 1 | Metric-level cognition/sleep association results based on discovery in MESA, with replication attempted in MrOS

	MESA; discovery $P<0.05/(155\times2)$				MROS; replication $P<0.05/3$		
Sleep metric	**Outcome**	***b* (95% CI)**	***P***	**Replication?**	**Outcome**	***b* (95% CI)**	***P***
Slow/delta (N2 + N3)	DSCT	−0.11 (−0.16 to −0.072)	2×10^{-7}	Yes	Trails B*	−0.11 (−0.15 to −0.068)	1×10^{-7}
SO wavelength (N2 + N3)	DSCT	−0.11 (−0.15 to −0.063)	1×10^{-6}	Yes	Trails B*	−0.1 (−0.14 to −0.064)	3×10^{-7}
Slow/delta (N2)	DSCT	−0.1 (−0.15 to −0.06)	3×10^{-6}	Yes	Trails B*	−0.11 (−0.15 to −0.07)	1×10^{-7}
Slow (N2 relative)	DSCT	−0.098 (−0.14 to −0.054)	1×10^{-5}	Yes	Trails B*	−0.11 (−0.15 to −0.071)	6×10^{-8}
PSC4 (N2)	DSCT	0.092 (0.049 to 0.14)	3×10^{-5}	Yes	Trails B*	0.079 (0.039 to 0.12)	0.00014
SO wavelength (N3)	DSCT	−0.11 (−0.16 to −0.057)	3×10^{-5}	Yes	Trails B*	−0.089 (−0.13 to −0.045)	7×10^{-5}
Slow (N3 relative)	DSCT	−0.1 (−0.15 to −0.051)	8×10^{-5}	Yes	Trails B*	−0.089 (−0.13 to −0.046)	6×10^{-5}
REM duration (min)	DSCT	0.086 (0.043 to 0.13)	9×10^{-5}	Yes	Trails B*	0.071 (0.031 to 0.11)	0.00050
Delta (N3 relative)	DSCT	0.1 (0.05 to 0.15)	1×10^{-4}	Yes	Trails B*	0.091 (0.047 to 0.13)	6×10^{-5}
Slow/delta (N3)	DSCT	−0.1 (−0.15 to −0.05)	1×10^{-4}	Yes	Trails B*	−0.089 (−0.13 to −0.045)	7×10^{-5}
PSC4 (N3)	DSCT	0.096 (0.047 to 0.15)	0.00015	Yes	DVT*	0.067 (0.021 to 0.11)	0.0040
Fast spindle frequency (N2)	CASI	−0.1 (−0.15 to −0.057)	9×10^{-6}	No	Trails B*	−0.045 (−0.085 to −0.0045)	0.030
Fall asleep while sitting and talking	DSCT	−0.087 (−0.13 to −0.046)	4×10^{-5}	N/A	.	.	.

This table includes metric-level association statistics in MESA (as the discovery sample), under the baseline covariate model, for the 13 metrics significant at $P<1.6\times10^{-4}$ (that is, $P<0.05$ after Bonferroni correction for 155 metrics present in both MESA and MrOS × two cognitive tests), considering only results for DSCT and CASI (that is, based on prior omnibus results). Of these 13, 11 replicated in MrOS at $P<0.0166$ (that is, $P<0.05$ after Bonferroni correction for three cognitive tests). One subjective measure was not present in MrOS. The other objective measure (fast spindle frequency) was nominally significant in MrOS ($P=0.03$) but did not meet the Bonferroni-adjusted threshold. All replicated associations showed the same direction of effect. 'Outcome' indicates the most significant cognitive test (DSCT or CASI for MESA; Trails B*, 3MS or DVT* for MrOS).

MESA and MrOS, as follows. For two of the four cognitive tests in MESA (DS_F and DS_B), the omnibus analyses failed to reject the global null hypothesis; we therefore did not consider those two tests further, to reduce the testing burden by a factor of two in this cohort. In contrast, associations appeared more diffusely spread with respect to the type of sleep metric in both MESA and MrOS, whether considering either domains or empirical clusters, for both max and sum statistics. As such, the omnibus results did not provide a clear basis for restricting the sets of sleep metrics to be advanced to individual testing. We therefore adopted a simple and stringent metric-level procedure to obtain a prioritized set of cognition-associated sleep metrics, independent of domain or cluster.

Specifically, we focused on 155 metrics that were common to MESA and MrOS (essentially all objective metrics plus the Epworth Sleepiness Scale (ESS)), for the two omnibus-significant cognitive measures in MESA (DSCT and CASI) and all three cognitive measures in MrOS (Trails B, 3MS and DVT). For this set, we identified metrics meeting Bonferroni-corrected significance at both discovery and replication. That is, for discovery, we required $P<0.05/(155\times t)$ for at least one cognitive measure in either MESA or MrOS, thereby allowing for 155 sleep metrics × t cognitive measures, where $t=2$ in MESA (implying $P<1.6\times10^{-4}$) and $t=3$ in MrOS (implying $P<1.1\times10^{-4}$). For metrics meeting the discovery threshold, we further required a replication $P<0.05/t$ and a directionally consistent association, with one or more cognitive measure in the second cohort. This yielded 23 sleep metrics, henceforth called the associated metrics (Tables 1 and 2 and Fig. 4b). Nonetheless, there are still likely to be true positives not included in this set of 23, which were selected only to represent a manageable subset with the strongest evidence for association.

Considering these 23 associated metrics, cognitive performance was related to sleep across macro architecture and multiple spectral, spindle, SO and spindle–SO coupling domains. Based on a visual inspection of their correlational structure (Fig. 4c,d), associated metrics fell across at least three broad classes: (1) sleep duration and continuity; (2) spindle activity and spindle–SO coupling; and (3) SWA. Consistent with the omnibus results, these 23 metrics were also widely dispersed across UMAP component space (Fig. 5c). Briefly, in terms of macro architecture, increased REM duration, sleep efficiency and sleep maintenance efficiency were associated with better cognitive performance (Fig. 4b and Tables 1 and 2). In terms of spindle activity, higher count and ISA were associated with better cognitive performance for both fast and slow N2 spindles, as well as higher fast spindle ISA per minute and higher absolute and relative sigma power during N2, but also higher sigma power during REM. Higher values of PSC4 (which was correlated with higher sigma but lower beta power, as well as multiple of spindle metrics, and showed a marked age-related decrease) were also associated with better cognition during both N2 and N3 sleep (Figs. 4b and 6a and Supplementary Fig. 8). With respect to SWA, we observed eight highly inter-related cognition-associated metrics (Fig. 4c,d). Higher relative slow (<1 Hz) power (during both N2 and N3) but

Table 2 | Metric-level cognition–sleep association results based on discovery in MrOS, with replication attempted in MESA

	MrOS; discovery $P < 0.05/(155 \times 3)$				MESA; replication $P < 0.05/2$		
Sleep metric	**Outcome**	***b* (95% CI)**	***P***	**Replication?**	**Outcome**	***b* (95% CI)**	***P***
Fast spindle total ISA (N2)	Trails B*	0.16 (0.12 to 0.2)	1×10^{-14}	Yes	DSCT	0.079 (0.033 to 0.13)	0.00090
Fast spindle ISA per minute (N2)	Trails B*	0.13 (0.091 to 0.17)	3×10^{-10}	Yes	DSCT	0.058 (0.011 to 0.1)	0.016
Fast spindle count (N2)	Trails B*	0.12 (0.084 to 0.17)	3×10^{-9}	Yes	DSCT	0.062 (0.017 to 0.11)	0.0068
Slow spindle total ISA (N2)	Trails B*	0.12 (0.083 to 0.16)	3×10^{-9}	Yes	DSCT	0.069 (0.024 to 0.11)	0.0025
Slow spindle count (N2)	Trails B*	0.12 (0.083 to 0.16)	3×10^{-9}	Yes	DSCT	0.069 (0.026 to 0.11)	0.0020
Sigma (N2 relative)	Trails B*	0.12 (0.077 to 0.16)	2×10^{-8}	Yes	DSCT	0.063 (0.018 to 0.11)	0.0061
Slow (N2 relative)	Trails B*	−0.11 (−0.15 to −0.071)	6×10^{-8}	Yes	DSCT	−0.098 (−0.14 to −0.054)	1×10^{-5}
Sleep efficiency	3MS	0.11 (0.069 to 0.15)	9×10^{-8}	Yes	CASI	0.059 (0.013 to 0.11)	0.012
Slow/delta (N2)	Trails B*	−0.11 (−0.15 to −0.07)	1×10^{-7}	Yes	DSCT	−0.1 (−0.15 to −0.06)	3×10^{-6}
Slow/delta (N2 + N3)	Trails B*	−0.11 (−0.15 to −0.068)	1×10^{-7}	Yes	DSCT	−0.11 (−0.16 to −0.072)	2×10^{-7}
SO wavelength (N2 + N3)	Trails B*	−0.1 (−0.14 to −0.064)	3×10^{-7}	Yes	DSCT	−0.11 (−0.15 to −0.063)	1×10^{-6}
Sleep maintenance efficiency	Trails B*	0.11 (0.065 to 0.15)	4×10^{-7}	Yes	DSCT	0.064 (0.019 to 0.11)	0.0056
Sigma (N2 absolute)	Trails B*	0.1 (0.064 to 0.14)	4×10^{-7}	Yes	DSCT	0.07 (0.021 to 0.12)	0.0047
Fast spindle–SO magnitude (N2 + N3)	Trails B*	0.1 (0.059 to 0.14)	1×10^{-6}	Yes	DSCT	0.061 (0.015 to 0.11)	0.010
Slow spindle–SO magnitude (N2 + N3)	3MS	0.091 (0.051 to 0.13)	8×10^{-6}	Yes	DSCT	0.074 (0.031 to 0.12)	0.00078
Sigma (REM relative)	Trails B*	0.087 (0.047 to 0.13)	2×10^{-5}	Yes	DSCT	0.057 (0.013 to 0.1)	0.011
Slow (N3 relative)	Trails B*	−0.089 (−0.13 to −0.046)	6×10^{-5}	Yes	DSCT	−0.1 (−0.15 to −0.051)	8×10^{-5}
Delta (N3 relative)	Trails B*	0.091 (0.047 to 0.13)	6×10^{-5}	Yes	DSCT	0.1 (0.05 to 0.15)	1×10^{-4}
SO wavelength (N3)	Trails B*	−0.089 (−0.13 to −0.045)	7×10^{-5}	Yes	DSCT	−0.11 (−0.16 to −0.057)	3×10^{-5}
Slow/delta (N3)	Trails B*	−0.089 (−0.13 to −0.045)	7×10^{-5}	Yes	DSCT	−0.1 (−0.15 to −0.05)	1×10^{-4}
Fast spindle–SWA magnitude (N2 + N3)	Trails B*	0.12 (0.081 to 0.16)	4×10^{-9}	No	DSCT	0.042 (−0.0037 to 0.087)	0.072
Fast spindle density (N2)	Trails B*	0.12 (0.083 to 0.16)	4×10^{-9}	No	DSCT	0.032 (−0.013 to 0.078)	0.16
Fast spindle ISA per spindle (N2)	Trails B*	0.12 (0.079 to 0.16)	8×10^{-9}	No	DSCT	0.038 (−0.0078 to 0.084)	0.10
WASO	3MS	−0.11 (−0.15 to −0.067)	2×10^{-7}	No	DSCT	−0.047 (−0.092 to −0.0022)	0.040
SO count (N2 + N3)	3MS	0.1 (0.065 to 0.14)	3×10^{-7}	No	DSCT	0.047 (0.002 to 0.091)	0.041
Fast spindle count (N3)	Trails B*	0.11 (0.067 to 0.16)	1×10^{-6}	No	DSCT	0.033 (−0.021 to 0.086)	0.23
N1 duration (%)	Trails B*	−0.096 (−0.14 to −0.056)	2×10^{-6}	No	DSCT	−0.043 (−0.087 to 0.0016)	0.059

Continued

Table 2 | Metric-level cognition/sleep association results based on discovery in MrOS, with replication attempted in MESA (Continued)

	MrOS; discovery $P < 0.05/(155 \times 3)$				MESA; replication $P < 0.05/2$		
Sleep metric	**Outcome**	***b* (95% CI)**	***P***	**Replication?**	**Outcome**	***b* (95% CI)**	***P***
Slow spindle density (N2)	Trails B*	0.097 (0.056 to 0.14)	3×10^{-6}	No	DSCT	0.046 (0.0014 to 0.091)	0.043
Slow spindle ISA per minute (N2)	Trails B*	0.096 (0.055 to 0.14)	4×10^{-6}	No	DSCT	0.047 (0.0017 to 0.093)	0.042
Fast spindle ISA per minute (N3)	Trails B*	0.1 (0.057 to 0.15)	8×10^{-6}	No	DSCT	0.051 (−0.0011 to 0.1)	0.055
Fast spindle total ISA (N3)	Trails B*	0.099 (0.055 to 0.14)	1×10^{-5}	No	DSCT	0.058 (0.0042 to 0.11)	0.035
Fast spindle density (N3)	Trails B*	0.093 (0.049 to 0.14)	4×10^{-5}	No	DSCT	0.023 (−0.028 to 0.074)	0.38
SO density (N2 + N3)	3MS	0.083 (0.043 to 0.12)	5×10^{-5}	No	CASI	0.023 (−0.023 to 0.069)	0.33
Sigma (N3 relative)	Trails B*	0.089 (0.046 to 0.13)	6×10^{-5}	No	DSCT	0.05 (−0.00016 to 0.1)	0.051
Beta (REM relative)	DVT*	0.084 (0.043 to 0.13)	8×10^{-5}	No	DSCT	0.044 (0.00078 to 0.088)	0.046
Sigma (N3 absolute)	Trails B*	0.087 (0.044 to 0.13)	9×10^{-5}	No	DSCT	0.056 (0.002 to 0.11)	0.042
Fast spindle–SWA angle (N2 + N3)	3MS	0.089 (0.044 to 0.13)	1×10^{-4}	No	DSCT	−0.036 (−0.084 to 0.013)	0.15
PSQI (efficiency)	Trails B*	−0.089 (−0.13 to −0.049)	1×10^{-5}	N/A			
FOSQ (productivity)	Trails B*	0.084 (0.045 to 0.12)	3×10^{-5}	N/A			

This table shows results similar to those presented in Table 1, except with MrOS as the discovery cohort and MESA as the replication, with suitably adjusted significance thresholds representing the different number of cognitive tests considered in each study. Here, 39 metrics were significant after correction for multiple testing in the MrOS discovery cohort. Of these, 20 were significant (corrected for the two tests in MESA) and had the same direction of effect in MESA. Of the remainder, two subjective metrics were not present in MESA and so could not be tested for replication. Of the other 17, seven were nominally significant at $P < 0.05$ but did not meet the stricter threshold of $P < 0.05/3$. Overall, Tables 1 and 2 list 23 unique sleep metrics that were discovered in one cohort and replicated in the second cohort. PSQI, Pittsburgh Sleep Quality Index; FOSQ, Functional Outcomes of Sleep Questionnaire.

lower relative delta (1–4 Hz) power (N3) were associated with worse cognitive performance). Correspondingly, higher slow/delta ratios (for N2, N3 and N2 + N3) also predicted worse cognitive performance. Based on individual SOs (detected under default, adaptive thresholds), decreased SO duration (for N3 as well as N2 + N3) was associated with better cognitive performance. Associations for SO duration were similar when based on absolute thresholds (Supplementary Data 3 and 4). SO duration was very highly correlated with slow/delta ratio in both MESA ($r = 0.947$ (95% CI = 0.942 to 0.952; $P < 10^{-15}$) during N2 + N3 sleep) and MrOS ($r = 0.952$; 95% CI = 0.948 to 0.956; $P < 10^{-15}$). As per-individual mean SO durations varied between approximately 0.8 and 1.5 s in both cohorts, with a mean around 1.1 s, slower SOs contributed relatively more to slow power (<1 Hz) than to delta power (>1 Hz). Finally, a stronger magnitude of spindle–SO coupling was associated with better cognitive performance, for both fast and slow spindles (Fig. 5b).

Considering these hits, in almost all cases, DSCT and Trails B were the most highly associated cognitive measures in MESA and MrOS, respectively (Tables 1 and 2 and Supplementary Data 1 and 2). To simplify subsequent follow-up analyses, we therefore concentrate primarily on DSCT and Trails B.

Joint analyses of associated metrics. That a particular metric was associated independent of the others is challenging to state definitively. We fit penalized Lasso regression models to all 23 metrics, with DSCT and Trails B as the dependent variables for MESA and MrOS respectively. All models were forced to include baseline covariates (see Supplementary Methods). For MESA ($n = 945$ due to list-wise deletion for missing data and the presence of N3 metrics), ten metrics with non-zero Lasso coefficients were: REM duration; sleep maintenance efficiency; relative slow power (N2); relative sigma power (REM); PSC4 (both N2 and N3); slow spindle count (N2); fast spindle ISA (N2); SO duration (N2 + N3); and slow spindle coupling magnitude (N2 + N3). For MrOS ($n = 1{,}571$ due to list-wise deletion for missing data), the 15 selected metrics were: REM duration; sleep efficiency; sleep maintenance efficiency; relative slow power (N2); relative delta power (N3); relative sigma power (REM); PSC4 (both N2 and N3); slow spindle ISA (N2); fast spindle ISA and ISA per minute (N2); slow/delta ratio (N2); SO duration (N2 + N3); and fast and slow spindle–SO coupling magnitude (N2 + N3).

While unlikely to truly prioritize optimal metrics, the above analyses indicate independent contributions of sleep macro architecture as well as spectral features of the sleep EEG, spindle activity, SO activity and spindle–SO coupling in predicting cognitive performance. To address this more directly, we asked whether EEG-based micro-architecture metrics (spectral metrics and spindle–SO activity and coupling) were informative over and above simpler indices that could be assayed via actigraphy or self-report. We fit a series of baseline regression models additionally controlling for TST, WASO, sleep efficiency and subjectively reported typical sleep duration and sleep quality (Fig. 6e,f). When predicting cognitive performance controlling for these indices of sleep continuity, satisfaction, disturbance and sleepiness, we found that, as one might expect, metrics that by definition depend on TST (that is, REM duration, spindle count and total ISA) showed *P* values that were orders of magnitude

smaller, in both MESA and MrOS (Fig. 6e,f). In MESA, both sleep efficiency metrics, slow spindle count/total ISA and fast spindle count, as well as fast spindle–SO coupling magnitude, were no longer nominally significant ($P > 0.05$) in the adjusted model. In MrOS, REM duration, both sleep efficiency metrics and PSC4 (N3) were no longer nominally significant. However, for all other metrics, association statistics remained highly significant, suggesting that most micro-architecture metrics captured more than simply reflecting sleep duration or poor sleep on a gross level (Fig. 6e,f).

Profiles of age- and cognition-related sleep metric associations. Many cognition-associated metrics also showed marked age-related changes (Fig. 7a). Furthermore, even though all models controlled for chronological age, associations between objective sleep metrics and cognitive performance appeared to be yoked inversely to their associations with age in both MESA (Fig. 7a and Supplementary Fig. 13a) and MrOS (Fig. 7a and Supplementary Fig. 13b). That is, if higher values were associated with increased age (for example, SO duration), higher values were also associated with decreased cognitive performance, even after controlling for age. Conversely, if lower values were associated with increased age (that is, REM duration), lower values were associated with decreased cognitive performance. More broadly, across all objective sleep metrics, and irrespective of statistical significance, we observed considerable negative correlations between coefficients for age and cognition from a joint model with the sleep metric as the dependent variable, along with baseline covariates and cardiometabolic disease status (Fig. 7b).

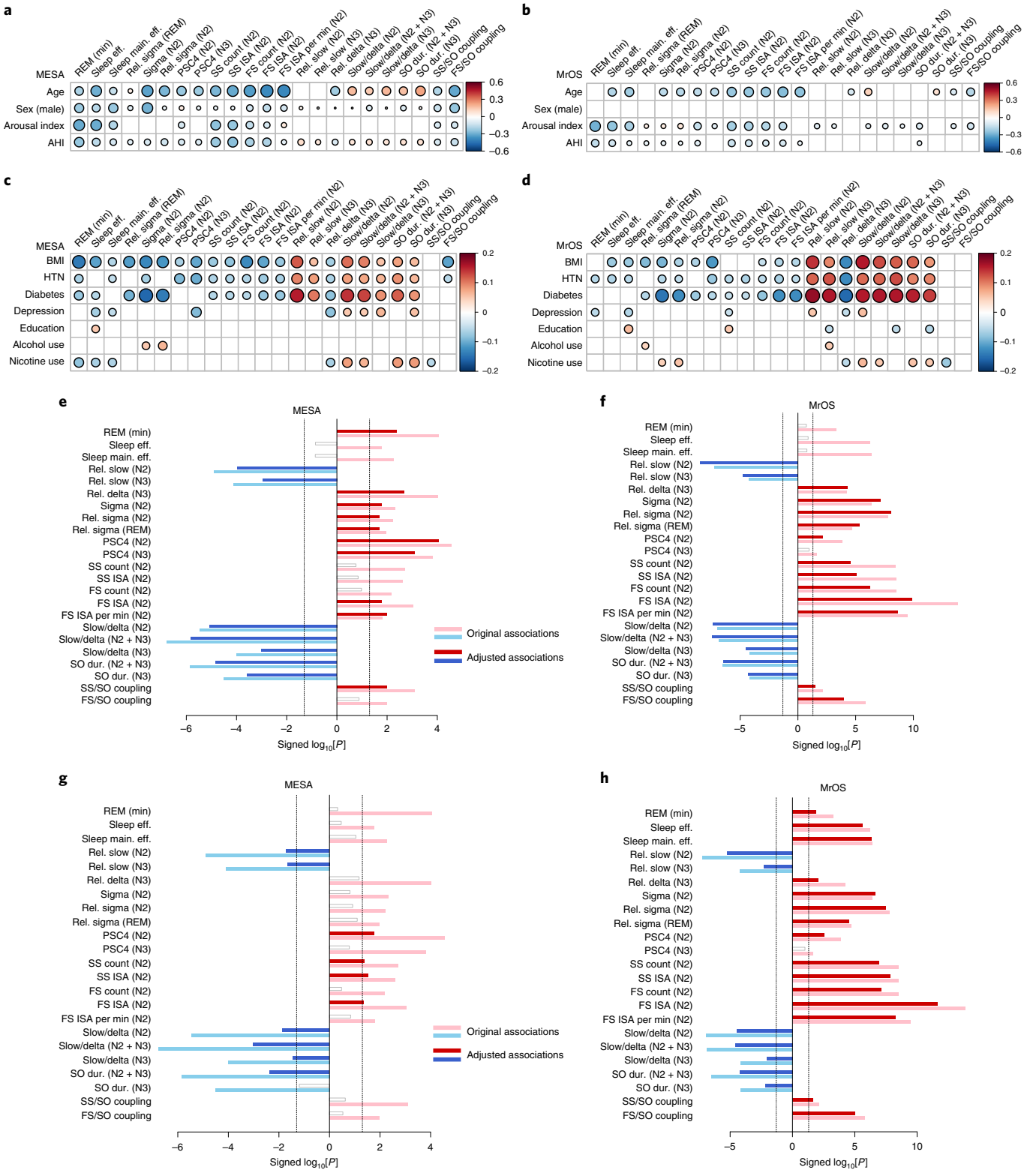
Correlates of cognition-associated sleep metrics. Other factors linked to both sleep and cognition may account for one or more of the associations presented above. Across MESA and MrOS, we observed that depressed mood, hypertension, diabetes, alcohol consumption and smoking—but not arousal index, apnoea–hypopnoea index (AHI) or body mass index (BMI)—showed significant associations with cognitive performance (Supplementary Tables 1 and 2). Many of the 23 associated sleep metrics were also correlated with one or more of these factors (Fig. 6a–d). In particular, arousal index, AHI and BMI, although not highly correlated with cognition in this sample, were quite highly correlated with several macro- and micro-architecture metrics. Between cohorts, we also saw consistent patterns of associations for other putative confounders and/or mediators and a number of sleep metrics. For example, in MESA, longer SO duration was associated with type 2 diabetes ($b = 0.31$ s.d. units; 95% CI = 0.18 to 0.43; $P = 1 \times 10^{-6}$) and hypertension ($b = 0.20$; 95% CI = 0.10 to 0.30; $P = 1 \times 10^{-4}$), with these associations being of comparable statistical significance to the association between SO duration and DSCT, with all analyses controlling for baseline covariates. SO duration continued to predict DSCT when additionally controlling for diabetes and hypertension status, albeit with a wider confidence interval and higher estimated effect size ($b = -0.68$ s.d. units; 95% CI = −1.0 to −0.35; $P = 8 \times 10^{-5}$), and in this joint model both diabetes ($b = -0.17$; 95% CI = −0.29 to −0.06; $P = 2 \times 10^{-3}$) and hypertension ($b = -0.20$; 95% CI = −0.29 to −0.11; $P = 2 \times 10^{-5}$) remained significant. We also observed a similar pattern of associations between SO duration, diabetes/hypertension and Trails B in MrOS, suggesting that although inter-related, these aspects of cardiometabolic health did not substantively confound or mediate this particular relationship between sleep neurophysiology and cognitive performance. More generally, the patterns of associations between objective metrics and diabetes/hypertension reflected age-related changes in those metrics (Fig. 7a and Supplementary Table 7). That is, based on models that regressed each sleep metric on disease state, controlling for age and other baseline covariates, individuals with diabetes/hypertension tended to have sleep metrics that looked more similar to those seen in older but healthier individuals, particularly for the NREM slow-activity metrics. Even considering all 155 sleep metrics common to MESA and MrOS, and estimating regression coefficients from a series of joint models with the sleep metric as the dependent variable and cognition (DSCT or Trails B*) and cardiometabolic disease state (hypertension or diabetes) as predictors along with baseline covariates, we observed strong negative correlations between the sleep–cognition associations on the one hand versus sleep–age and sleep–cardiometabolic disease associations on the other hand (Fig. 7b). There were also numerous correlations between commonly used medications (including sleeping pills, antidepressants and diabetes/hypertension medications) and the 23 sleep metrics (Supplementary Fig. 14a). For example, beta blocker use was associated with significantly longer SO duration in both MESA ($b = 0.23$ s.d. units; 95% CI = 0.09 to 0.36; $P = 0.001$) and MrOS ($b = 0.14$; 95% CI = 0.05 to 0.24; $P = 0.004$), controlling for diabetes, hypertension and other baseline covariates. In general, however, it will be challenging to disentangle the effects of medications from the underlying conditions that they treat (Supplementary Fig. 14b).

To establish whether the 23 associated sleep metrics showed associations with cognitive performance that were statistically independent of these other factors, we re-ran the regressions with an augmented covariate model, including the factors considered in Fig. 6a–d and Supplementary Fig. 14 (that is, educational attainment, arousal index, AHI, BMI, diabetes, hypertension, depression symptoms and alcohol/nicotine use, as well as common medication use; see Supplementary Methods). In MESA, only ten of the 23 metrics remained a nominally significant predictor of DSCT independent of these factors, including SO duration, slow power metrics, PSC4 and fast and slow spindle ISA, although even these associations were attenuated (Fig. 6g). In MrOS, however, 22 of the 23 metrics remained significant predictors of Trails B after adjustment, albeit with slight reductions in statistical support (Fig. 6h). That the MrOS sample was more homogeneous, larger and potentially better powered to detect associations might account for this apparent difference. In MESA, there was no single additional covariate that clearly accounted for the attenuated sleep–cognition associations

Fig. 6 | Correlations between cognition-associated metrics and other putative confounding and mediating variables. a–d, Pearson correlations of sleep metrics versus age, sex and other factors. Correlations with variables other than age and sex are adjusted for age, sex, race/ethnicity and collection site (being based on the baseline model residuals rather than raw sleep metrics). The size of each circle is proportional to the absolute magnitude of the correlation; the colour reflects the direction (positive correlations are red; negative correlations are blue). Only significant ($P < 0.05$) correlations are shown. Color bars indicate the correlation coefficient. **a**, Correlations with age, sex, arousal index and AHI in MESA. **b**, As in **a**, but for MrOS. **c**, Correlations with other health and behavioural measures in MESA. Note that, for clarity of presentation, as age, sex, arousal index and AHI show large correlations with some sleep metrics, the scale is reduced here (±0.2 in **c** and **d** versus ±0.6 in **a** and **b**). HTN, hypertension. **d**, As in **c**, but for MrOS. **e**, Association results (direction-of-effect signed $-\log_{10}[P\text{values}]$) in MESA for the 23 selected metrics, adjusted for gross sleep metrics (namely, ESS[96], the Women's Health Initiative Insomnia Rating Scale[97] and two MESA Sleep Questionnaire items in MESA; see Supplementary Methods). **f**, As in **e**, but for MrOS, adjusting for ESS, the Pittsburgh Sleep Quality Index[99] and the Functional Outcomes of Sleep Questionnaire[100]. **g**, Fully adjusted association results for the selected 23 metrics under the augmented covariate model and including medication-use covariates in MESA. **h**, As in **g**, but for MrOS. In **e**–**h**, significant ($P < 0.05$) positive associations are shown in red, negative associations are shown in blue, original associations are shown in light shades, and adjusted associations are shown in dark shades; dashed vertical lines indicate the nominal $P < 0.05$ level of association, with nonsignificant associations shown in white.

(Supplementary Fig. 15). Educational attainment, in particular, is often used to adjust neurocognitive test scores; however, we found that adjusting for educational attainment had little or no impact on our primary baseline results in either MESA (Supplementary Fig. 15b) or MrOS. Unlike cardiometabolic disease, educational attainment did not show credible evidence of an association with most of the 23 sleep metrics (Supplementary Table 7). Overall, further work will be needed to unpack this set of inter-related associations between cognition, cardiometabolic health and medication, and specific elements of sleep neurophysiology, including SO morphology.

Objective and subjective sleep duration metrics and inverted-U association models. Objectively measured TST did not meet our criteria for association in either study, although there was a nominally significant result for DSCT in MESA ($b = 0.06$; 95% CI = 0.01 to



0.10; $P=0.01$). Nonetheless, it is important to note that objectively measured TST tends to capture qualitative as well as quantitative differences in sleep (Supplementary Fig. 16). For example, presumably reflecting the typical structure of successive sleep cycles, individuals with longer TST had a lower percentage of N1 sleep ($r=-0.13$ (95% CI = −0.18 to −0.08; $P=2\times10^{-7}$) in MESA; $r=-0.20$ (95% CI = −0.24 to −0.16; $P<10^{-15}$) in MrOS) but a higher percentage of REM sleep ($r=0.22$ (95% CI = 0.18 to 0.27; $P<10^{-15}$) in MESA; $r=0.17$ (95% CI = 0.13 to 0.21; $P<10^{-15}$) in MrOS). Such distinctions may be significant when interpreting studies of self-reported or actigraphy-based sleep duration, and may also have implications when considering putative interventions or behavioural modifications centred around sleep duration (that is, the quality or micro architecture of sleep, and not simply the quantity, may be the causally relevant factor).

In our cohorts, the available self-report measures of typical sleep duration were only moderately correlated with objective TST from the PSG night, in both MESA ($r=0.21$; 95% CI = 0.16 to 0.25; $P<10^{-15}$) and MrOS ($r=0.17$; 95% CI = 0.13 to 0.21; $P<10^{-15}$). As others have noted[58], objective TST was also substantially shorter (~6 h in both MESA and MrOS) compared with self-reported sleep duration (7.8 and 8.1 h for weekdays and weekends, respectively, in MESA; and 6.7 h in MrOS based on a four-point scale of <5, 5–6, 6–7 and >7 h, coded as 4.5, 5.5, 6.5 and 7.5 h, respectively)—differences that may arise for multiple reasons, including reporting bias and misclassification, as well as the atypical context of having a PSG administered[58].

Comparing objective and subjective sleep duration, we also observed differences in the MESA associations with DSCT. As noted above, longer objective TST showed a modest association with improved DSCT performance, whereas longer subjective sleep duration was modestly associated with worse DSCT ($b=-0.04$; 95% CI = −0.09 to −0.0004; $P=0.048$ for weekday sleep) and CASI performance ($b=-0.05$ (95% CI = −0.10 to −0.01; $P=0.02$) and $b=-0.06$ (95% CI = −0.10 to −0.01; $P=0.013$) for weekday and weekend sleep, respectively). One explanation for this apparent discrepancy is the possibility of nonlinear effects: whereas our primary analyses considered only linear associations, there have been reports of nonlinear, inverted-U associations between sleep duration and poorer cognitive outcomes[20,21]. As linear models may obscure estimates of such effects, for both subjective and objective measures of sleep duration, we fit additional models that included an orthogonal second-order polynomial term for sleep duration as a predictor. In MrOS, for both subjective and objective sleep duration measures, we observed significant quadratic terms in models predicting Trails B, 3MS and DVT (Supplementary Fig. 17b), although the results were less clear for MESA (Supplementary Fig. 17a). In both MESA and MrOS, objective and subjective sleep times showed significant nonlinear associations with other measures of sleep quality, including ESS- and PSG-derived sleep efficiency and WASO (Supplementary Fig. 18). In general, compared with individuals with a typical self-reported sleep duration, those who reported either shorter or longer sleep tended to have worse-quality sleep. As others have suggested[59], this makes interpretation of associations with sleep duration (nonlinear or otherwise) difficult to interpret without more detailed knowledge of sleep architecture.

Testing for effect modification by sex and *APOE* genotype. Finally, in MESA, we tested the 23 associated metrics for evidence of effect modification by sex or *APOE* genotype. There was no credible evidence for interaction with sex (all P values were > 0.1). Previous analyses in MESA identified effect modification by *APOE* (carriers versus non-carriers of an E4 allele) of the association between DSCT and ESS[60]. We confirmed this result in our analytical sample and testing framework, controlling for ancestry by inclusion of the first ten principal components[61] as well as self-reported race/ethnicity ($P=2\times10^{-3}$ for the ESS × *APOE* interaction on DSCT), whereby a significant DSCT–ESS association was observed in carriers ($n=406$; $b=-0.14$; 95% CI = −0.22 to −0.06; $P=5\times10^{-4}$) but not non-carriers ($n=1{,}118$; $b=0.005$; 95% CI = −0.05 to 0.06; $P=0.86$). We also observed a significant ESS × *APOE* interaction on CASI ($P=4\times10^{-4}$), although, unlike DSCT, this was primarily driven by a positive association in non-carriers ($b=0.11$; 95% CI = 0.05 to 0.16; $P=1\times10^{-4}$) combined with the absence of a nominally significant association in carriers. Beyond ESS, however, there was no strong and consistent evidence to suggest either an association with or effect moderation by *APOE* genotype for any of the 23 cognition-associated sleep metrics on DSCT or CASI. However, reduced power to detect interactions compared with main effects makes it difficult to definitively rule out possible involvement of *APOE* in the links between sleep and cognition.

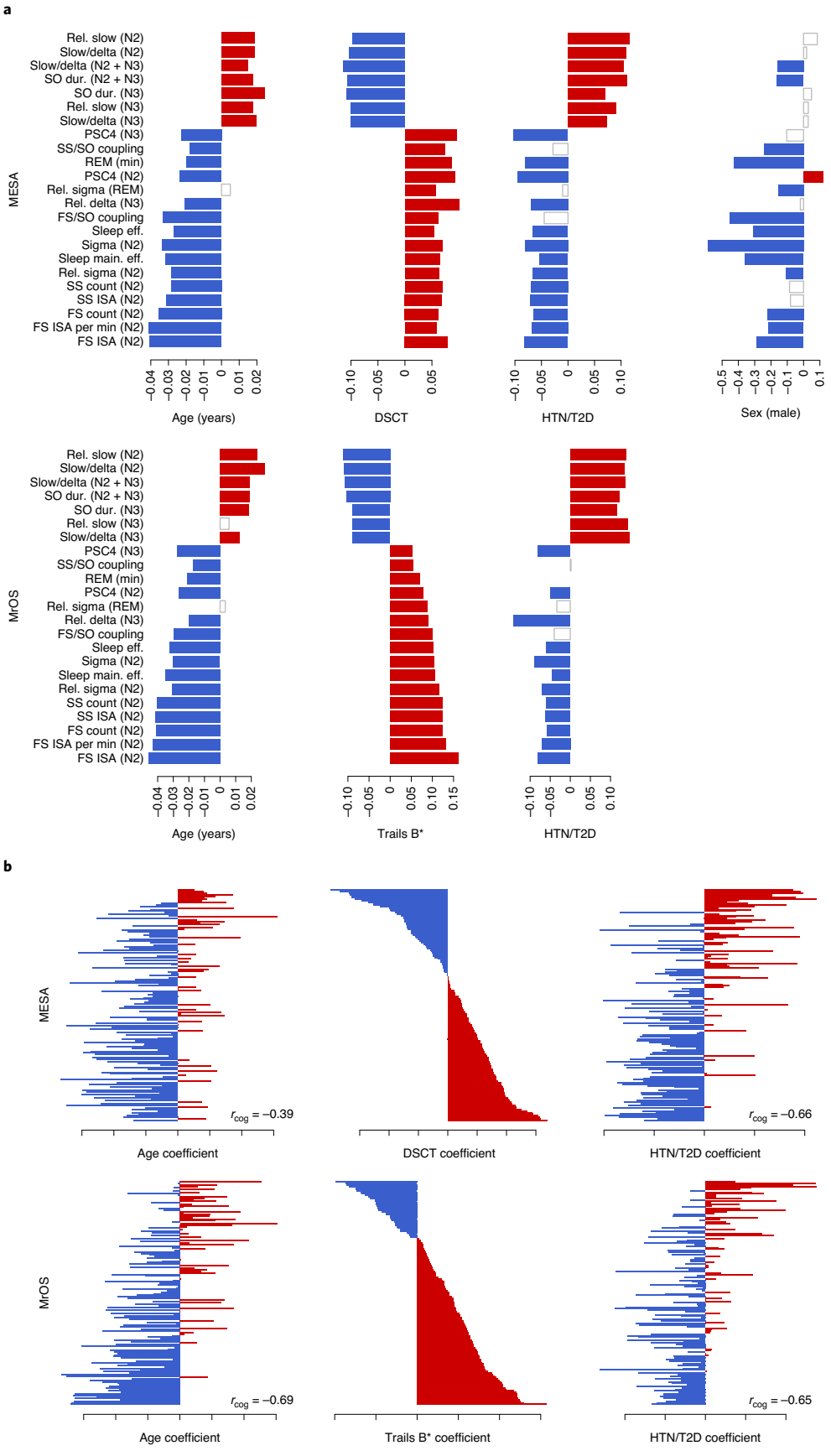
Discussion

Sleep quality and quantity change markedly over the life course. Emerging evidence suggests that these changes are predictive of (and potentially causally related to) age-related cognitive decline and impairment. Here, we adopted an explicitly data-driven and exploratory approach with the aim of broadly cataloguing individual differences in sleep neurophysiology and their relationships with cognition. In two independent community samples of older adults, we identified multiple inter-related properties of sleep, in terms of macro-architectural as well as spectral, spindle and SO micro-architectural elements, that were associated with cognitive performance, and with processing speed and executive functioning in particular. Whereas multiple other studies have reported associations between aspects of sleep micro architecture and cognitive performance, our study is distinguished by the scope of metrics comprehensively tested. Despite evaluating a large number of metrics, we attempted to ensure the rigour of our data-driven

Fig. 7 | Patterns of associations between sleep metrics and age, cognition, cardiometabolic disease and sex. a, Standardized regression coefficients for age (per year), cognitive performance (DSCT and Trails B* in MESA and MrOS, respectively, per s.d. unit), a binary indicator variable denoting the presence of hypertension or type 2 diabetes (T2D), and sex (1 = male; 0 = female). The rows correspond to the 23 replicated cognition-associated sleep metrics, ordered by increasing effect size of the cognition–sleep association in MrOS. Positive effects are shown in red, whereas negative effects are shown in blue. Effects not nominally significant ($P>0.05$) are shown in white with a grey outline. For DSCT and Trails B*, cognition was the dependent variable and the sleep metric was the predictor (that is, as per the primary analyses). For the other variables, the sleep metric was the dependent variable. All models included baseline covariates. **b**, Results for all 155 sleep metrics in common between MESA and MrOS (all objective metrics plus ESS). Unlike in **a**, here effect estimates for age, cognition and cardiometabolic disease were all jointly estimated in a single multiple linear regression, with the sleep metric as the dependent variable and with age, cognitive performance (DSCT or Trails B*) and cardiometabolic disease state as predictors (along with other baseline model covariates). All effect estimates are plotted in red or blue to denote the direction of effect ($b>0$ and $b<0$, respectively), regardless of statistical significance. Rows (sleep metrics) are sorted by the cognitive effect estimate, separately for MESA and MrOS. Across all sleep metrics, effect estimates were highly negatively correlated for cognition and cardiometabolic disease in both MESA (Pearson's $r=-0.66$; 95% CI = −0.74 to −0.57; $P<10^{-15}$) and MrOS ($r=-0.65$; 95% CI = −0.73 to −0.55; $P<10^{-15}$), as well as age in MESA ($r=-0.39$; 95% CI = −0.52 to −0.25; $P<5\times10^{-7}$) and MrOS ($r=-0.69$; 95% CI = −0.76 to −0.59; $P<10^{-15}$).

testing model by the use of large samples, formal control of multiple testing and insistence on replication in an independent sample.

Several aspects of our results are worth highlighting. First, the dozens of inter-related sleep metrics measured exhibited a higher-order underlying structure that was broadly consistent



between MESA and MrOS, with different aspects being apparent through dimension reduction/cluster analysis and PSC analysis. To a first approximation, subjective metrics were independent of objective PSG metrics. Of the objective metrics, macro-architecture metrics, including sleep duration and efficiency, clustered separately from micro-architecture metrics. Of the micro-architecture metrics, analyses in both MESA and MrOS pointed to at least six further subdivisions, including: absolute total (slow) power; slow, delta relative power; alpha, sigma, beta power; slow spindles; and fast spindles. These results suggest that many of the different metrics deployed in our study are capturing unique aspects of the sleep process.

Second, multiple metrics were robustly associated with cognition, spanning a number of domains/clusters. The 23 flagged metrics included REM duration, sleep efficiency, PSC4 (sigma/beta ratio), both fast and slow spindle activity, SO duration (and slow/delta ratio) and the strength of spindle–SO coupling. We observed independent associations with cognition across a number of these metrics, suggesting that studies of sleep and cognition should not focus on only one aspect of sleep. Third, spectral–spindle–SO micro-architecture metrics tended to predict cognitive performance independent of more routinely collected measures of sleep duration and quality. Indeed, overall, the subjective measures of sleep were only loosely correlated with objective metrics and could not be seen as meaningful proxies for the cognition-relevant information contained in the sleep EEG. Fourth, all cognition-associated objective sleep metrics showed marked age-dependent trends, such that objective sleep metrics more commonly observed in younger individuals tended to be associated with better cognitive health, even after adjusting for chronological age. Indeed, one could relatively well predict whether a sleep metric would associate with cognition (independent of age), and in which direction, solely on the basis of its association with age. Fifth, sleep metrics associated with advanced age and worse cognitive performance tended to also be independently associated with poor cardiometabolic health; namely, diabetes and hypertension.

One possibility is that multiple facets of sleep neurophysiology may reflect more general biological aging processes that mediate age-dependent cognitive decline, rather than being specific to sleep per se. One approach that captures the spirit of these observations is to estimate a so-called brain age index from sleep signal data, whereby accelerated aging (as estimated by sleep microstructure relative to chronological age) may be a general marker of pathophysiology[62]. Our results support this general concept, although identifying the points at which a unidimensional construct such as this loses explanatory value will be an empirical question. For example, although males indeed look like older females for many sleep metrics in our data, there are many other sleep metrics for which this pattern clearly does not hold, suggesting at least one other pertinent (sex-specific) dimension, and potentially many more, that may not be optimally captured by a single number.

Impairment on standard cognitive tests can precede the onset of dementia by several years[63]. Indeed, there is an increasing interest in differentiating between normal aging versus emerging dementia[64,65], and consideration of sleep may be fruitful in this regard. Aging and Alzheimer's disease are characterized by sleep disturbances and brain regions important for sleep, and wake mechanisms are affected in early Alzheimer's disease[66,67]. Inadequate sleep duration, increased fragmentation, decreased depth and increased daytime sleepiness have been associated with a heightened risk of developing Alzheimer's disease independent of co-existing sleep disorders[16,18,68,69]. Excessive daytime sleepiness has also been found to be a marker of cognitive decline and dementia[15,25], and in brain imaging studies, changes in default mode network connectivity were associated with daytime sleepiness, distinct from the effects of aging[70,71]. With respect to total sleep duration, we found partial support for a nonlinear, inverted-U association with cognition, primarily in MrOS. Although inverted-U models have been reported for sleep duration and a range of health outcomes, subjectively reported sleep duration is a poor proxy for a complex set of underlying factors, which makes nonlinear association difficult to interpret if duration is not a unidimensional construct[59,72–74]. Our results are consistent with a model in which (self-reported) sleep duration captures variation in sleep quality as well as sleep duration, and may include increased WASO.

Primary transient elements of the sleep EEG, including spindles and SOs, are known to vary with age and correlate with cognitive functions. For example, patients with Alzheimer's disease performed better on memory testing if a greater number of fast spindles were present[75]. Although the exact neural underpinnings of fast versus slow spindles remain unclear, their differential occurrence during the SO cycle has pointed to separate generation mechanisms[76]. The occurrence of fast and slow spindles also differs over the life course: although spindle density declines with age, relatively more fast spindles are observed in older individuals compared with younger individuals[7]. Here, we further observed that, whereas fast spindles tend to become faster with age, slow spindles become slower.

Much attention has been given to the link between cognitive decline/dementia and N3 or slow-wave sleep and/or SWA, based on human and animal studies demonstrating that newly learned memory traces are strengthened through an active neuronal replay of these memory representations during slow-wave sleep[10,77,78]. Furthermore, disruption of SWA is associated with increased cerebrospinal fluid amyloid-β and, vice versa, amyloid-β pathology has been linked to a reduced generation of slow-wave oscillation and sleep-related memory consolidation[79–81]. We did not observe associations with N3 duration and our cognitive measures, whereas SO wavelength, the balance of slow (<1 Hz) to delta (1–4 Hz) spectral power and spindle–SO(SWA) coupling predicted cognitive performance. Whereas some studies[79] define SWA to encompass 0.5–4 Hz, we observed qualitative differences in the associations of <1 Hz versus >1 Hz SWA. This distinction is reflected in the metric, used here and by others[30,80], that explicitly indexes the proportion of SWA that is <1 Hz (<1 Hz relative SWA), labelled here as the slow/delta ratio. However, seemingly at odds with previous reports of reduced <1 Hz relative SWA predicting increased β-amyloid burden[30,80], we instead observed that increased <1 Hz relative SWA was associated with poorer cognitive performance. Consistent with this, increased SO duration (which is highly correlated with increased <1 Hz relative SWA) similarly predicted poorer cognitive performance and both measures increased with increasing age in both cohorts. As others have noted, increased SO duration and slope may reflect changes in synaptic density and white matter integrity that accumulate with age[56].

Our findings with respect to REM duration strengthen early polysomnographic studies[82,83] and more recent epidemiological reports on cognitive performance[16,18,28,84,85] and incident dementia[33], as well as animal studies[86,87]. Cholinergic neurons are important modulators of REM sleep, and Alzheimer's disease pathology is associated with impairment in cholinergic networks and imbalance between cholinergic and orexinergic systems; it has been suggested that, in Alzheimer's disease, these impairments may be responsible for sleep fragmentation and changes in NREM/REM sleep architecture, resulting in decreased REM sleep[88]. REM sleep, as well as other aspects of sleep neurophysiology, including spindles, is under partial circadian control[89], although in this study the REM effects on cognition were independent of available chronotype metrics.

From a methodological perspective, we applied PSC analysis to complement traditional band power metrics, uncovering dimensions of individual variation in the sleep EEG that were highly preserved across cohorts. Of ten retained components, the top three were common to both REM and NREM sleep, whereas subsequent

components were more specific to NREM sleep. In both MESA and MrOS, and in both N2 and N3 sleep, the fourth component (PSC4, higher values of which primarily indexed relatively greater sigma and lower beta power) was consistently and significantly associated with poorer cognitive performance. These effects persisted when additionally controlling for sigma power by itself (data not shown). Of all of the PSCs, PSC4 also showed the strongest age-related decline, in both MESA and MrOS and in both N2 and N3. PSCs provide an efficient partitioning of individual differences based on empirical patterns of variation, although they do not necessarily map directly and uniquely onto distinct underlying physiological processes any more than a gross measure such as sigma band power does. In the future, application of this approach to sleep studies with multiple EEG channels (potentially including cross-channel coherence spectral terms) may be particularly attractive. Following on from our use of MESA as a reference (that is, a set of singular values and vectors used to impute PSCs into MrOS participants), a further possible extension is to develop a standardized, universal reference, so that other investigators could impute the equivalent set of PSCs into new studies. Ideally, such a reference panel (or panels) would be based on data from a demographically and medically diverse range of individuals and across different stages of sleep. Likewise, separate single- and multi-channel references could be created for particular montages. Inasmuch as PSCs prove to reflect an efficient summary of power spectra, this would provide a means by which large-scale studies could empower smaller ones.

Our study has a number of further limitations that limit its scope, generalizability and basis for inference. First, we only considered cognition cross-sectionally, rather than assessing cognitive decline per se. Second, the characterization of sleep was based on a limited EEG montage, reducing our ability to detect spatially specific associations or to consider functional connectivity during sleep.

Third, we do not know the extent to which our findings on sleep in older adults will generalize in different populations, including younger participants. Although our results may have implications for Alzheimer's disease and clinically recognized mild cognitive impairment, the present study primarily reflects variation within the normal range of cognitive performance in older adults. As we were not able to assay EEG or other brain biomarkers during the waking state, the extent to which our findings reflect sleep-dependent processes specifically, versus more general neuropsychological or neurodegenerative features, cannot be directly inferred. Nonetheless, it may still be the case that the sleep EEG provides a particularly good index of these more general factors.

Fourth, although the cognitive tests used were valid and sensitive measures within their domains, the cognitive battery was not comprehensive and it did not assess all relevant cognitive domains. The lack of a learning task that measures longer-term memory retention is a weakness, as this domain is known to change with age and serve as a differential marker of normal age-related decline versus dementia. Our primary measures of executive functioning and processing speed (DSCT and Trails B) may also reflect psychomotor and visual problems. Nonetheless, our findings suggest that some measures may be less sensitive, sleep-dependent processes (namely, Digit Span in MESA).

Fifth, unravelling the precise causal relationships between sleep and cognitive functioning, which probably include bidirectional effects and pleiotropic genetic factors, is beyond the current scope. As well as a baseline covariate model, we applied an augmented covariate model, and specifically considered the role of educational attainment on the reported associations. Controlling for educational attainment itself, although strongly correlated with cognitive performance, had little impact on the results, in both MESA and MrOS. In MESA, a number of metrics that were significant under the baseline model were no longer significant under the broader augmented model. In contrast, almost all metrics remained significant under both baseline and augmented covariate models in MrOS. Interpreting such models is challenging when key aspects of the underlying causal models are unknown. In particular, the addition of an extra covariate may over-control if it is on the causal path from predictor to outcome, or if it is itself a parallel outcome.

Sixth, although chronotype metrics did not show associations with cognition, chronotype was far from comprehensively measured in this study, with only a single objective metric: sleep midpoint from a single night of in-home PSG. Nonetheless, sleep midpoint was still highly correlated with Horne–Ostberg Morningness–Eveningness Questionnaire (MEQ) morningness in MESA ($r=-0.42$; 95% CI $=-0.46$ to -0.38; $P<10^{-15}$) and so can be taken as a rough index of chronotype, as participants were instructed to go to bed at their usual time.

Seventh, our characterization of sleep was by no means fully comprehensive. For example, as well as topographical differences and cross-channel coherence, which were excluded due to the limited montage, we did not consider nocturnal (ultradian) dynamics, cross-frequency coupling or specific transient features of REM sleep.

Finally, although our top-down, domain-based approach to testing provided a useful framework for organizing our work (as well as a granular but rigorous statistical control of multiple testing, an issue that is sometimes afforded scant regard in the literature), the precise assignment of metrics to domains was in many ways arbitrary. There are other reasonable ways to structure domains. For example, instead of spindle occurrence and morphology, we could instead have tested fast and slow spindles as two separate domains. In recognition of this, we also performed a parallel, empirical clustering to group sleep metrics (which did in fact show fast and slow spindles to be separable features). Nonetheless, this approach is not without drawbacks either: there are myriad approaches to clustering and/or dimension reduction, and different approaches will give different sets of clusters. There is no gold standard to adjudicate between competing solutions and exhaustive comparison is beyond our present scope. Furthermore, dimension reduction/cluster solutions will be different in different samples, potentially making it difficult to compare between studies. For these reasons, although the broad picture is clear, we advise that not too much weight be placed on the specifics of any single grouping presented here.

In conclusion, this study contributes to a body of evidence, including strong animal and emerging human work, that indicates that sleep disturbances are associated with cognitive impairment. Identifying the specific elements of sleep neurophysiology that are associated with cognitive impairment in older adults has a broad range of potential benefits: to elucidate pathogenic mechanisms of cognitive decline; to provide intermediate phenotypes (for example, outcomes in intervention or genetic studies); and to predict high-risk individuals. At the same time, our study also highlights some of the challenges inherent in applying sleep neurophysiology metrics as clinically relevant biomarkers. While statistically highly significant, most effect sizes for individual metrics were modest, with standardized beta coefficients typically under 0.1, meaning that their predictive power is individually limited. Many metrics are inter-related with each other and can be calculated in a number of different ways, meaning it is often difficult to determine the optimal and independent set of metrics. Furthermore, we demonstrated a number of demographic differences in some key sleep metrics considered here. Although it was possible for us to control for these factors in the context of an epidemiological study, in the context of using sleep metrics from individual patients as biomarkers, it will be crucial to adjust raw metrics to account for such effects, or to identify metrics that are less susceptible to technical measurement factors, for example. Still, identifying which facets of sleep neurophysiology are relevant for cognitive performance retains the potential

to provide therapeutic targets, including the modulation of sleep oscillations. As potential therapeutic targets, either through lifestyle intervention and cognitive behavioural therapies, pharmacological approaches[90] or closed-loop auditory stimulation and transcranial direct current stimulation/transcranial magnetic stimulation[91–95], multiple aspects of sleep behaviour and neurophysiology may be modifiable, pointing to the possibility of future sleep-based initiatives to maintain and improve cognitive health. However, future longitudinal and interventional studies will be needed, to address the extent to which these putative biomarkers have direct causal influences on cognition and whether they are modifiable in ways that impact cognition, or whether they instead secondarily reflect a more general accelerated biological aging that manifests as worse cognitive functioning as well as poorer cardiometabolic health.

Methods

Cohorts and cognitive measures. MESA is a prospective observational cohort study consisting of a diverse community-based sample (38% white, 28% African–American, 22% Hispanic and 12% Chinese–American) of 6,814 men and women. Details on the study design for MESA have been published previously[38,39]. The current analysis used data from the MESA Sleep ancillary study, conducted in proximity to the fifth MESA follow-up examination, in which 2,048 individuals underwent PSG conducted between 2010 and 2013 (the clinic exams occurred from 2010–2012, whereas the PSGs were scheduled at that time but occurred between 2010 and 2013). Three neuropsychological tests were selected to assess cognitive function during the fifth visit examination (2010–2012): the CASI[42], DSCT[41] and Digit Span Test (DS_F and DS_B)[43]. Key covariates included self-reported educational attainment and income, arousal index and AHI from the PSG, BMI, self-reported diagnoses of hypertension or diabetes, depression symptoms from the Center for Epidemiologic Studies Depression Scale, current alcohol and nicotine use, and *APOE4* genotype.

The MrOS Sleep Study enrolled 3,135 participants (2003–2005) for in-home PSG. Cognitive measures and key covariates were available that were similar (but not identical) to those in MESA. The cognitive data included in this report were collected during sleep visit 1 (2003–2005): Trails B[44]; a modified, expanded version of the 3MS test[45]; and the DVT[46]. Along with age, collection site and ethnicity (white/non-white), key covariates were educational attainment, arousal index, AHI, self-reported sleep apnoea or any sleep disorder, BMI, self-reported hypertension and diabetes, symptoms of depression and anxiety rated by the Geriatric Depression Scale and Goldberg depression and anxiety scales, and current alcohol and nicotine use. Comparing cognitive measures in MESA and MrOS, for which we might expect a broad correspondence of results, we note that DSCT and Trails B are both measures of processing speed and executive functioning, whereas CASI and 3MS are both designed to be global indices sensitive to dementia. Note that Trails B and DVT were based on timed performance; for consistency of reporting, Trails B* and DVT* refer to the corresponding measure with the sign reversed, such that a negative coefficient implies worse performance across all seven cognitive measures. See the Supplementary Methods for details on cohorts and measures.

Subjective sleep metrics. In both MESA and MrOS, daytime sleepiness was measured by the ESS[96], which consists of eight items each scored from 0–3. The summed total score (range = 0–24) can be dichotomized such that scores of ten or more indicate excessive daytime sleepiness. In MESA only, insomnia symptoms were assessed with the Women's Health Initiative Insomnia Rating Scale[97] (a five-item scale that rates sleep latency, sleep maintenance insomnia, early morning awakening and sleep quality) and circadian preference towards morningness and eveningness was assessed by a five-item modified MEQ[98], where higher scores indicate a preference for morningness. Additionally, in MESA, we included two other self-report items from the MESA sleep questionnaire, asking whether the participant had sleep difficulties causing irritability, and whether they felt overly sleepy during the day (each question was asked with respect to the past 4 weeks). MrOS participants also completed the Pittsburgh Sleep Quality Index[99] and the Functional Outcomes of Sleep Questionnaire[100].

Objective sleep metrics. All PSG studies were scored manually by trained staff using standard criteria[101], and low-quality studies were identified by scorers and removed from subsequent data analysis. Epochs containing manually annotated arousals or movements were removed. All EEG analysis used the Luna C/C++ pipeline developed by S.M.P. (http://zzz.bwh.harvard.edu/luna/). In both cohorts, EEG signals from C4/M1 were resampled at 100 Hz and segmented into 30-s epochs, then bandpass filtered with transition frequencies at 0.3 and 35 Hz using a Kaiser window zero-phase filter. We extracted epochs corresponding to four sets of stages (N2, N3, N2 + N3 and REM) and, within each set, applied two consecutive approaches to artefact detection, as well as correction for potential cardiac contamination (see Supplementary Methods). Separately for each stage (or for N2 + N3 combined), we required at least ten epochs (5 min) of artefact-free sleep for that individual to be included in further spectral, spindle and SO analyses (Supplementary Table 3).

We derived a panel of metrics, grouped into a number of domains (fully enumerated in the Supplementary Methods): sleep macro architecture; chronotype; spectral power; alternative time–frequency metrics; PSCs; spindle occurrence; spindle morphology; SO occurrence; SO morphology; and spindle–SO coupling (which encompass the so-called spindle–SWA coupling metrics also). Based on a priori considerations, not all metrics were calculated for all stage sets as described below; for example, we did not consider spindles or SOs during REM sleep (see the Supplementary Methods for more details). We also derived, separately for MESA and MrOS, empirical groupings of sleep metrics based on cluster analysis, as described below. Note that these domains and clusters were only used during omnibus testing, to provide a high-level view of the patterns of results: neither domains nor empirically derived clusters were used in the selection of the final set of associated metrics.

Sleep macro architecture. The sleep architecture domain comprised 15 metrics based on manual staging: TST; stage N1, N2, N3 and REM duration (minutes and percentage); number and mean duration of NREM cycles (defined per ref. [102]); sleep efficiency (TST as a proportion of total recording time); sleep maintenance efficiency (which excludes the sleep latency period from the denominator); WASO (minutes) and REM latency (minutes; time from sleep onset to first REM epoch). In MESA, this domain included self-reported typical weekday and weekend sleep durations, whereas in MrOS this domain included a single measure of self-reported sleep duration.

Chronotype. The chronotype domain comprised a single objective metric: sleep midpoint (the middle point between sleep onset and offset based on the PSG, coded as hours past 17:00). In MESA, this domain also included self-reported chronotype, based on a modified MEQ[98].

Spectral power. We calculated spectral power using the Welch algorithm and fast Fourier transformation applied to 4-s windows, shifted by 2-s increments and tapered with a Tukey window function (taper length = 50%), yielding a frequency resolution of 0.25 Hz. The relative band power was obtained by dividing the absolute band power by the total power, where the total power was based on the band 0.5–30 Hz. This domain comprised 36 metrics: absolute log-scale and relative spectral power for slow, delta, theta, alpha, sigma and beta bands, defined as 0.5–1, 1–4, 4–8, 8–12, 12–15 and 15–30 Hz, respectively, for N2, N3 and REM sleep (two definitions × six bands × three stages = 36).

Alternative time–frequency metrics. A second domain of alternative EEG summaries comprised 16 metrics: three Hjorth parameters (activity, mobility and complexity)[103]; an index of EEG slowing for N2, N3 and REM sleep; and, following ref. [80], the absolute slow (*S*; 0.5–1 Hz) power normalized by the sum of slow and delta (*D*; 1–4 Hz) power, labelled here as the slow/delta ratio, although note that this quantity is $S/(S+D)$ not S/D. For comparability with other SO metrics (below), the slow/delta ratio was also calculated for N2 + N3 sleep combined (that is, four metrics × three stages + one metric × four stages = 16).

PSC analysis. In addition to summarizing power spectra in traditional bands, we applied an alternative, data-driven decomposition of power spectra, labelled PSC analysis. This approach uses singular value decomposition to define empirical bases of variation in power spectra between individuals; others have used similar approaches in different contexts[104–106]. Here, we used PSC analysis to reduce 77 correlated absolute power measures (0.75–20 Hz in 0.25-Hz intervals) to a smaller number of orthogonal components that captured the majority of the variation observed within each cohort. PSC analysis provides a natural way to quantify coordinated changes across power spectra, and may be expected to better capture subtle individual differences, such as a shift in peak spindle frequency within the sigma band, which will be reflected in the correlational structure of power measures. Inasmuch as broad individual differences in total power (or overall ~1/*f* slope) map onto specific components, PSC analysis also effectively normalizes power spectra, such that the remaining orthogonal components will more directly index specific oscillatory activity with aperiodic background activity subtracted out.

We applied PSC analysis to N2 and N3 power spectra separately. Based on visual review of the cumulative total variation explained, we selected the number of components to retain. As the sign/polarity of singular value decomposition components is arbitrary, we constrained comparable (highly correlated) N2 and N3 components to be positively correlated with each other, to aid the interpretation of results. If PSC analyses were applied to MESA and MrOS separately, one challenge would be to establish the comparability of the resulting components (that is, whether the PSC4 in MESA, for example, measured a similar thing as PSC4 in MrOS). For the primary analyses, rather than perform PSC analysis separately in MESA and MrOS, we therefore projected the observed power spectra in MrOS into the reduced dimensional space defined by the MESA PSC analysis, to ensure comparability of PSCs between cohorts (see the Supplementary Methods for details). We compared the resulting MESA-derived MrOS components with those

from an MrOS-specific analysis. To describe each component (labelled PSC1, PSC2 and so on) in terms of the original power spectra, we plotted the median power (at each 0.25-Hz frequency bin) for groups stratified according to the quintiles of that component.

Spindles. We detected spindles using a previously published wavelet-based approach[7], modified to determine a suitable detection threshold empirically, based on Otsu's method[107] of maximizing between-class variance in the wavelet coefficient, comparing above-threshold (putative spindle) and below-threshold (non-spindle) intervals (Supplementary Fig. 2), which selected a multiplicative threshold of six times the individual's median wavelet power. Based on previous work[108–110], we assumed that a frequency-dependent dichotomization between fast and slow spindles captured important distinctions in spindle properties. Indeed, this assumption was supported by an analysis in which we fit, separately for each individual, a Gaussian mixture model to the multi-modal frequency distribution of spindles detected under an untargeted, frequency-agnostic approach (see Supplementary Methods and Supplementary Fig. 3a,b).

Our primary analyses targeted two classes of spindle: fast ($F_C = 15$ Hz) and slow ($F_C = 11$ Hz), in each case detecting spindles within approximately ±2 Hz of the target frequency (Fig. 2a). Given these two sets of detected spindles, we estimated each individual's mean for several spindle morphology metrics. The spindle occurrence domain comprised eight metrics: spindle count and density for fast and slow spindles during either N2 or N3 sleep (two measures × two spindle types × two stages = 8). The spindle morphology domain comprised 36 metrics: mean spindle duration, frequency, number of oscillations, an index of ISA per spindle (ISA_S), per minute (ISA_M) or in total (ISA_T), intra-spindle frequency change (chirp) and symmetry and asymmetry indices for fast and slow spindles during N2 and N3 sleep (nine measures × two spindle types × two stages = 36). See Supplementary Methods for details.

SOs. We used a heuristic to detect individual SOs in the sleep EEG. Multiple quantifications of SO have been used in the literature[28,111,112] and there is no consensus on optimal definitions, especially in elderly individuals who tend to have decreased slow-wave sleep, potentially leading to methodological difficulties[113]. We therefore adopted two approaches (using either adaptive (relative) or absolute (fixed) amplitude thresholds) and assessed the robustness of the results to this choice. In primary analyses, we used adaptive thresholds, defined relative to each individual's baseline level of slow activity (see Supplementary Methods). We initially focused on SO during N3 sleep only, but extended this to consider N2 + N3 sleep combined, as a nontrivial number of older individuals did not have a sufficient duration of N3 sleep. The SO occurrence domain comprised four metrics: SO count and density (counts per minute), during either N3 sleep or N2 + N3 sleep combined. The SO morphology domain comprised six metrics: mean peak-to-peak amplitude (μV; log-scaled); duration or wavelength (s); and upward slope of the negative peak (μV s^{-1}; log-scaled) during either N3 sleep or N2 + N3 sleep combined (three measures × two stage groups = 6).

Spindle–SO coupling. The primary quantification of spindle–SO coupling was based on the phase of SWA at spindle peaks (that is, the points of maximum spindle oscillation (peak-to-peak amplitude), typically near the spindle's centre). We estimated three measures of coupling: gross overlap; coupling magnitude; and coupling phase angle. As defined below, overlap measured whether spindles and SO tended to occur concurrently, whereas phase coupling measured whether spindles tended to peak at particular times during the SO, thereby indexing a more precise temporal relationship.

Overlap was defined as the number of spindle peaks that fell within a detected SO. We determined the null distribution of this metric empirically, by randomly shuffling spindle peaks and recalculating overlap 100,000 times. For each individual, we normalized the overlap metric as a Z score, given the mean and standard deviation of the null distribution. We also calculated an empirical P value for above-chance overlap. One potential concern when using this approach, especially when considering N2 + N3 sleep combined, is that apparent above-chance spindle–SO overlap may simply reflect that relatively more spindles occur during N2 but relatively more SOs occur during N3, whereas stage-agnostic shuffling does not preserve this aspect of the data. We therefore used a modified scheme, whereby each spindle peak was randomly shuffled only within the 30-s epoch that spanned it.

Coupling magnitude and phase angle metrics were based on the instantaneous phase from a Hilbert transform of EEG after bandpass filtering in the 0.3–4 Hz range. Although we were primarily interested in SOs in the ~1 Hz frequency range, using a broader bandpass (up to 4 Hz) helped to not impose an artificially sinusoidal shape on SO waveforms. That is, the criteria to define SOs identified waveforms that were, on average, of <1 Hz frequency (that is, Supplementary Fig. 4 and Supplementary Table 4). We calculated intra-trial phase consistency (ITPC)[114] as a measure of the strength of spindle–SO coupling. Briefly, if the Hilbert-derived phase angle at each spindle peak is assumed to be a unit vector on a circle, with an angle matching the phase angle, the ITPC is the magnitude of the average of these complex vectors. Under the null hypothesis of no coupling, an asymptotic P value can be computed as $\exp(-n \times \mathrm{ITPC}^2)$. An individual's coupling phase angle is given by the angle of the complex mean of these vectors. Separately for fast and slow spindles, we calculated two ITPC statistics: (1) for all spindles; and (2) only for spindles with a peak that overlapped an SO. We refer to these two classes of coupling metrics as spindle–SWA and spindle–SO, respectively. As ITPC statistics and asymptotic P values can show bias or noise when based on a small number of spindle–SO events, or on non-sinusoidal waveforms, our primary analyses generated empirical null distributions to normalize them. Under the former spindle–SWA scenario, we used the within-epoch shuffling scheme described above. Under the latter spindle–SO scenario, which only considered spindles overlapping detected SOs, in order to ensure that the phase coupling metric was not confounded by differences in gross spindle–SO overlap, we required null replicates to contain the same number of SO-overlapping spindle peaks (that is, to have an identical extent of overlap). We therefore shuffled each spindle peak by a random offset, between zero seconds and the duration of the spanning SO, wrapping as necessary. This within-SO shuffling scheme preserved the total number of spindles, SOs and their gross overlap in each null replicate, but randomized only the precise relationships between spindle peaks and SO phase.

Spindle–SWA and spindle–SO analyses were restricted to N2 + N3 sleep combined, because spindles are predominant during N2 whereas SOs are predominant during N3. Many individuals had too few N3 spindles to assess coupling in N3 alone (Supplementary Table 3). For primary association analyses with cognitive measures, the spindle–SO coupling domain comprised ten metrics: overlap, phase coupling magnitude and mean phase at spindle peak, separately for fast and slow spindles. Furthermore, phase coupling magnitude and angle metrics were calculated separately for: (1) all spindles (that is, spindle–SWA coupling); versus (2) only spindles whose peak overlapped a detected SO (that is, spindle–SO coupling) (one measure × two spindle types + two measures × two spindle types × two spindle filters = 10).

Empirical clustering of sleep metrics. To complement the prespecified domains, we also sought to determine groupings of sleep metrics empirically in MESA and MrOS. We defined the distance between metrics i and j as $1 - |r_{ij}|$ and applied UMAP[115], as implemented in the umap R package, allowing for nine nearest neighbours, a minimum distance of 0.1 and extracting either ten components (for clustering) or two components (for visualization). We then applied HDBSCAN[116] as implemented in the dbscan R package, setting the minimum size of clusters to six. We manually labelled clusters by visual inspection of the assigned metrics. Note that UMAP dimension reduction does not fully preserve the global structure of the data (that is, versus local clustering) and the orientation of components/axes is arbitrary, meaning that direct comparisons between MESA and MrOS projections are not necessarily meaningful.

Statistical approach to association analysis. As enumerated above, in MESA alone, we tested 173 sleep metrics × four cognitive measures, yielding 692 tests. Performing multiple tests necessitates consideration of statistical false positives. We therefore adopted a permutation-based top-down approach to control false positive rates while accounting for the correlational structure between sleep predictors and between cognitive outcomes. We generated 20,000 permuted datasets in which individuals' sleep metrics were randomly reassigned. In each null dataset, we refit the 692 linear regression models of covariate-residualized cognitive measures on sleep metrics and recorded the $-\log_{10}[P\text{value}]$ from each. For each replicate (randomized dataset), we calculated two additional omnibus statistics based on these asymptotic P values: the maximum and the sum of the $-\log_{10}[P\text{value}]$, labelled max and sum, respectively. We rejected the global null hypothesis that all sleep metrics were unassociated with all cognitive measures given an empirical P value of <0.05 for either the max or sum statistic, based on their null distributions. Specifically, if R is the number of times a statistic from a null replicate was equal to or exceeded the observed statistic, the empirical P value is $(R+1)/(N+1)$, where $N = 20{,}000$ replicates. The max statistic is expected to be more powerful under the scenario in which a small number of metrics are highly associated with cognitive performance, whereas the sum statistic will be more powerful when a larger proportion of sleep metrics show more modest associations with cognitive performance.

Following a top-down approach, and contingent on a significant global omnibus test, we next asked whether specific subsets of tests had significant max or sum statistics, considering each of the 11 sleep metric domains (for example, all spindle morphology metrics for all cognitive measures) and each of the four MESA cognitive measures (for example, all sleep metrics for DSCT). Finally, we tested each sleep domain for each cognitive measure separately (for example, all spindle morphology metrics for DSCT). In this way, as well as controlling for multiple testing, this framework helped to contextualize broader patterns of association between sleep and cognitive performance.

We initially designed this study to focus only on the MESA cohort, but subsequently sought to replicate our findings in the MrOS cohort. Respecting this context, we first applied the omnibus testing framework to MESA. Second, we attempted replication in the independent MrOS sample for the specific metrics that showed the strongest association with cognitive performance in MESA. Third, we repeated omnibus testing for MrOS and considered the broader correspondence of results between cohorts. Fourth, with a focus on objective

rather than subjective sleep metrics, we identified those showing significant and consistent patterns of association across both cohorts. Finally, for the set of replicated cognition-associated sleep metrics, we considered potential confounder or mediator variables in a series of secondary analyses.

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

Data availability

All PSG data are freely available via the National Sleep Research Resource (http://sleepdata.org). The MESA dataset is available at https://doi.org/10.25822/n7hq-c406. The MrOS dataset is at https://doi.org/10.25822/kc27-0425. Full PSG and clinical/covariate data are available for all interested parties pending completion of a Data Access and Use Agreement and Institutional Review Board approval, as outlined on the National Sleep Research Resource website.

Code availability

Sleep EEG data were processed using the Luna package developed by S.M.P. (http://zzz.bwh.harvard.edu/luna/). The C/C++ code is available at the following GitHub repository: http://github.com/remnrem/luna-base/. Specifically, the analysis presented in this manuscript used Luna to derive measures of sleep architecture (HYPNO command) and to perform epoch-level artefact detection (SIGSTATS), signal filtering (FILTER), spectral estimation (PSD) and spindle–SO detection (SPINDLES).

Received: 25 January 2019; Accepted: 15 September 2020;
Published online: 16 November 2020

References

1. Foley, D. J. et al. Sleep complaints among elderly persons: an epidemiologic study of three communities. *Sleep* **18**, 425–432 (1995).
2. Foley, D., Ancoli-Israel, S., Britz, P. & Walsh, J. Sleep disturbances and chronic disease in older adults: results of the 2003 National Sleep Foundation Sleep in America Survey. *J. Psychosom. Res.* **56**, 497–502 (2004).
3. Petersen, R. C. et al. Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* **56**, 303–308 (1999).
4. Li, S.-C. et al. Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychol. Sci.* **15**, 155–163 (2004).
5. Dijk, D. J., Duffy, J. F. & Czeisler, C. A. Contribution of circadian physiology and sleep homeostasis to age-related changes in human sleep. *Chronobiol. Int.* **17**, 285–311 (2000).
6. Ohayon, M. M., Carskadon, M. A., Guilleminault, C. & Vitiello, M. V. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep* **27**, 1255–1273 (2004).
7. Purcell, S. M. et al. Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource. *Nat. Commun.* **8**, 15930 (2017).
8. Redline, S. et al. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch. Intern. Med.* **164**, 406–418 (2004).
9. Diering, G. H. et al. Homer1a drives homeostatic scaling-down of excitatory synapses during sleep. *Science* **355**, 511–515 (2017).
10. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
11. Stickgold, R. & Walker, M. P. Sleep-dependent memory consolidation and reconsolidation. *Sleep Med.* **8**, 331–343 (2007).
12. De Vivo, L. et al. Ultrastructural evidence for synaptic scaling across the wake/sleep cycle. *Science* **355**, 507–510 (2017).
13. Scullin, M. K. & Bliwise, D. L. Sleep, cognition, and normal aging: integrating a half century of multidisciplinary research. *Perspect. Psychol. Sci.* **10**, 97–137 (2015).
14. Ohayon, M. M. & Vecchierini, M.-F. Normative sleep data, cognitive function and daily living activities in older adults in the community. *Sleep* **28**, 981–989 (2005).
15. Keage, H. A. D. et al. What sleep characteristics predict cognitive decline in the elderly? *Sleep Med.* **13**, 886–892 (2012).
16. Song, Y. et al. Relationships between sleep stages and changes in cognitive function in older men: the MrOS Sleep Study. *Sleep* **38**, 411–421 (2015).
17. Cavuoto, M. G. et al. Objective but not subjective sleep predicts memory in community-dwelling older adults. *J. Sleep Res.* **25**, 475–485 (2016).
18. Blackwell, T. et al. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men sleep study. *J. Am. Geriatr. Soc.* **59**, 2217–2225 (2011).
19. Spira, A. P. et al. Actigraphic sleep duration and fragmentation in older women: associations with performance across cognitive domains. *Sleep* **40**, zsx073 (2017).
20. Devore, E. E. et al. Sleep duration in midlife and later life in relation to cognition. *J. Am. Geriatr. Soc.* **62**, 1073–1081 (2014).
21. Ramos, A. R. et al. Sleep duration and neurocognitive function in the Hispanic Community Health Study/Study of Latinos. *Sleep* **39**, 1843–1851 (2016).
22. Blackwell, T. et al. Association of sleep characteristics and cognition in older community-dwelling men: the MrOS Sleep Study. *Sleep* **34**, 1347–1356 (2011).
23. Potvin, O. et al. Sleep quality and 1-year incident cognitive impairment in community-dwelling older adults. *Sleep* **35**, 491–499 (2012).
24. Jaussent, I. et al. Excessive sleepiness is predictive of cognitive decline in the elderly. *Sleep* **35**, 1201–1207 (2012).
25. Foley, D. et al. Daytime sleepiness is associated with 3-year incident dementia and cognitive decline in older Japanese–American men. *J. Am. Geriatr. Soc.* **49**, 1628–1632 (2001).
26. Carvalho, D. Z. et al. Association of excessive daytime sleepiness with longitudinal β-amyloid accumulation in elderly persons without dementia. *JAMA Neurol.* **75**, 672–680 (2018).
27. Mander, B. A., Winer, J. R., Jagust, W. J. & Walker, M. P. Sleep: a novel mechanistic pathway, biomarker, and treatment target in the pathology of Alzheimer's disease? *Trends Neurosci.* **39**, 552–566 (2016).
28. Lafortune, M. et al. Sleep spindles and rapid eye movement sleep as predictors of next morning cognitive performance in healthy middle-aged and older participants. *J. Sleep Res.* **23**, 159–167 (2014).
29. Helfrich, R. F., Mander, B. A., Jagust, W. J., Knight, R. T. & Walker, M. P. Old brains come uncoupled in sleep: slow wave–spindle synchrony, brain atrophy, and forgetting. *Neuron* **97**, 221–230.e4 (2018).
30. Winer, J. R. et al. Sleep as a potential biomarker of tau and β-amyloid burden in the human brain. *J. Neurosci.* **39**, 6315–6324 (2019).
31. Muehlroth, B. E. et al. Precise slow oscillation–spindle coupling promotes memory consolidation in younger and older adults. *Sci. Rep.* **9**, 1940 (2019).
32. Prinz, P. N. et al. Sleep, EEG and mental function changes in senile dementia of the Alzheimer's type. *Neurobiol. Aging* **3**, 361–370 (1982).
33. Pase, M. P. et al. Sleep architecture and the risk of incident dementia in the community. *Neurology* **89**, 1244–1250 (2017).
34. Varga, A. W. et al. Apnea-induced rapid eye movement sleep disruption impairs human spatial navigational memory. *J. Neurosci.* **34**, 14571–14577 (2014).
35. Bjorness, T. E., Riley, B. T., Tysor, M. K. & Poe, G. R. REM restriction persistently alters strategy used to solve a spatial task. *Learn. Mem.* **12**, 352–359 (2005).
36. Smith, C. & Rose, G. M. Evidence for a paradoxical sleep window for place learning in the Morris water maze. *Physiol. Behav.* **59**, 93–97 (1996).
37. Yaffe, K. et al. Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *J. Am. Med. Assoc.* **306**, 613–619 (2011).
38. Burke, G., Lima, J., Wong, N. D. & Narula, J. The multiethnic study of atherosclerosis. *Glob. Heart* **11**, 267–268 (2016).
39. Bild, D. E. et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am. J. Epidemiol.* **156**, 871–881 (2002).
40. Chen, X. et al. Racial/ethnic differences in sleep disturbances: the Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep* **38**, 877–888 (2015).
41. Wechsler, D. *Adult Intelligence Scale-III (WAIS-III)* (Psychological Corporation/Harcourt, 1996).
42. Teng, E. L. et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int. Psychogeriatr.* **6**, 45–58 (1994).
43. Fitzpatrick, A. L. et al. Sociodemographic correlates of cognition in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am. J. Geriatr. Psychiatry* **23**, 684–697 (2015).
44. Reitan, R. M. Validity of the trail making test as an indicator of organic brain damage. *Percept. Mot. Skills* **8**, 271–276 (1958).
45. Teng, E. L. & Chui, H. C. The Modified Mini-Mental State (3MS) examination. *J. Clin. Psychiatry* **48**, 314–318 (1987).
46. Kelland, D. Z. & Lewis, R. F. The Digit Vigilance Test: reliability, validity, and sensitivity to diazepam. *Arch. Clin. Neuropsychol.* **11**, 339–344 (1996).
47. Cappuccio, F. P. & Miller, M. A. Sleep and cardio-metabolic disease. *Curr. Cardiol. Rep.* **19**, 110 (2017).
48. Köhler, S. et al. Temporal evolution of cognitive changes in incident hypertension: prospective cohort study across the adult age span. *Hypertension* **63**, 245–251 (2014).
49. Roberts, R. E. & Duong, H. T. The prospective association between sleep deprivation and depression among adolescents. *Sleep* **37**, 239–244 (2014).
50. Byers, A. L. & Yaffe, K. Depression and risk of developing dementia. *Nat. Rev. Neurol.* **7**, 323–331 (2011).
51. Kluge, M., Schüssler, P. & Steiger, A. Duloxetine increases stage 3 sleep and suppresses rapid eye movement (REM) sleep in patients with major depression. *Eur. Neuropsychopharmacol.* **17**, 527–531 (2007).

52. DeMartinis, N. A. & Winokur, A. Effects of psychiatric medications on sleep and sleep disorders. *CNS Neurol. Disord. Drug Targets* **6**, 17–29 (2007).
53. Scheer, F. A. J. L. et al. Repeated melatonin supplementation improves sleep in hypertensive patients treated with beta-blockers: a randomized controlled trial. *Sleep* **35**, 1395–1402 (2012).
54. Laventure, S. et al. Beyond spindles: interactions between sleep spindles and boundary frequencies during cued reactivation of motor memory representations. *Sleep* **41**, zsy142 (2018).
55. Dubé, J. et al. Cortical thinning explains changes in sleep slow waves during adulthood. *J. Neurosci.* **35**, 7795–7807 (2015).
56. Carrier, J. et al. Sleep slow wave changes during the middle years of life. *Eur. J. Neurosci.* **33**, 758–766 (2011).
57. Martin, N. et al. Topography of age-related changes in sleep spindles. *Neurobiol. Aging* **34**, 468–476 (2013).
58. Jackson, C. L., Patel, S. R., Jackson, W. B., Lutsey, P. L. & Redline, S.Agreement between self-reported and objectively measured sleep duration among white, Black, Hispanic, and Chinese adults in the United States: Multi-Ethnic Study of Atherosclerosis. *Sleep* **41**, zsy057 (2018).
59. Bianchi, M. T., Thomas, R. J. & Westover, M. B. An open request to epidemiologists: please stop querying self-reported sleep duration. *Sleep Med.* **35**, 92–93 (2017).
60. Johnson, D. A. et al. Greater cognitive deficits with sleep-disordered breathing among individuals with genetic susceptibility to Alzheimer disease. The Multi-Ethnic Study of Atherosclerosis. *Ann. Am. Thorac. Soc.* **14**, 1697–1705 (2017).
61. Liu, Y. et al. Methylomics of gene expression in human monocytes. *Hum. Mol. Genet.* **22**, 5065–5074 (2013).
62. Sun, H. et al. Brain age from the electroencephalogram of sleep. *Neurobiol. Aging* **74**, 112–120 (2019).
63. Jorm, A. F., Masaki, K. H., Petrovitch, H., Ross, G. W. & White, L. R. Cognitive deficits 3 to 6 years before dementia onset in a population sample: the Honolulu–Asia Aging Study. *J. Am. Geriatr. Soc.* **53**, 452–455 (2005).
64. Brody, J. A. & Schneider, E. L. Diseases and disorders of aging: an hypothesis. *J. Chronic Dis.* **39**, 871–876 (1986).
65. Toepper, M. Dissociating normal aging from Alzheimer's disease: a view from cognitive neuroscience. *J. Alzheimers Dis.* **57**, 331–352 (2017).
66. Wang, J. L. et al. Suprachiasmatic neuron numbers and rest–activity circadian rhythms in older humans. *Ann. Neurol.* **78**, 317–322 (2015).
67. Swaab, D. F., Fliers, E. & Partiman, T. S. The suprachiasmatic nucleus of the human brain in relation to sex, age and senile dementia. *Brain Res.* **342**, 37–44 (1985).
68. Blackwell, T. et al. Poor sleep is associated with impaired cognitive function in older women: the Study of Osteoporotic Fractures. *J. Gerontol. A Biol. Sci. Med. Sci.* **61**, 405–410 (2006).
69. Lim, A. S. P. et al. Sleep is related to neuron numbers in the ventrolateral preoptic/intermediate nucleus in older adults with and without Alzheimer's disease. *Brain* **137**, 2847–2861 (2014).
70. Ward, A. M. et al. Daytime sleepiness is associated with decreased default mode network connectivity in both young and cognitively intact elderly subjects. *Sleep* **36**, 1609–1615 (2013).
71. Carvalho, D. Z. et al. Excessive daytime sleepiness and fatigue may indicate accelerated brain aging in cognitively normal late middle-aged and older adults. *Sleep Med.* **32**, 236–243 (2017).
72. Knutson, K. L. & Turek, F. W. The U-shaped association between sleep and health: the 2 peaks do not mean the same thing. *Sleep* **29**, 878–879 (2006).
73. Patel, S. R., Malhotra, A., Gottlieb, D. J., White, D. P. & Hu, F. B. Correlates of long sleep duration. *Sleep* **29**, 881–889 (2006).
74. Grandner, M. A. & Drummond, S. P. A. Who are the long sleepers? Towards an understanding of the mortality relationship. *Sleep Med. Rev.* **11**, 341–360 (2007).
75. Rauchs, G. et al. Is there a link between sleep changes and memory in Alzheimer's disease? *NeuroReport* **19**, 1159–1162 (2008).
76. Mölle, M., Bergmann, T. O., Marshall, L. & Born, J. Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing. *Sleep* **34**, 1411–1421 (2011).
77. Feld, G. & Diekelmann, S.Sleep smart—optimizing sleep for declarative learning and memory. *Front. Psychol.* **6**, 622 (2015).
78. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. Local sleep and learning. *Nature* **430**, 78–81 (2004).
79. Ju, Y.-E. S. et al. Slow wave sleep disruption increases cerebrospinal fluid amyloid-β levels. *Brain* **140**, 2104–2111 (2017).
80. Mander, B. A. et al. β-amyloid disrupts human NREM slow waves and related hippocampus-dependent memory consolidation. *Nat. Neurosci.* **18**, 1051–1057 (2015).
81. Wang, C. & Holtzman, D. M. Bidirectional relationship between sleep and Alzheimer's disease: role of amyloid, tau, and other factors. *Neuropsychopharmacology* **45**, 104–120 (2020).
82. Feinberg, I., Koresko, R. L. & Heller, N. EEG sleep patterns as a function of normal and pathological aging in man. *J. Psychiatr. Res.* **5**, 107–144 (1967).
83. Spiegel, R., Herzog, A. & Köberle, S. Polygraphic sleep criteria as predictors of successful aging: an exploratory longitudinal study. *Biol. Psychiatry* **45**, 435–442 (1999).
84. Kim, S. J., Lee, J. H., Lee, D. Y., Jhoo, J. H. & Woo, J. I. Neurocognitive dysfunction associated with sleep quality and sleep apnea in patients with mild cognitive impairment. *Am. J. Geriatr. Psychiatry* **19**, 374–381 (2011).
85. Della Monica, C., Johnsen, S., Atzori, G., Groeger, J. A. & Dijk, D.-J. Rapid eye movement sleep, sleep continuity and slow wave sleep as predictors of cognition, mood, and subjective sleep quality in healthy men and women, aged 20–84 years. *Front. Psychiatry* **9**, 255 (2018).
86. Markowska, A. L. et al. Individual differences in aging: behavioral and neurobiological correlates. *Neurobiol. Aging* **10**, 31–43 (1989).
87. Stone, W. S., Altman, H. J., Berman, R. F., Caldwell, D. F. & Kilbey, M. M. Association of sleep parameters and memory in intact old rats and young rats with lesions in the nucleus basalis magnocellularis. *Behav. Neurosci.* **103**, 755–764 (1989).
88. Liguori, C. et al. Orexinergic system dysregulation, sleep impairment, and cognitive decline in Alzheimer disease. *JAMA Neurol.* **71**, 1498–1505 (2014).
89. Czeisler, C. A., Zimmerman, J. C., Ronda, J. M., Moore-Ede, M. C. & Weitzman, E. D. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep* **2**, 329–346 (1980).
90. Roh, J. H. et al. Potential role of orexin and sleep modulation in the pathogenesis of Alzheimer's disease. *J. Exp. Med.* **211**, 2487–2496 (2014).
91. Astori, S., Wimmer, R. D. & Lüthi, A. Manipulating sleep spindles—expanding views on sleep, memory, and disease. *Trends Neurosci.* **36**, 738–748 (2013).
92. Marshall, L., Helgadóttir, H., Mölle, M. & Born, J. Boosting slow oscillations during sleep potentiates memory. *Nature* **444**, 610–613 (2006).
93. Wilckens, K. A., Ferrarelli, F., Walker, M. P. & Buysse, D. J. Slow-wave activity enhancement to improve cognition. *Trends Neurosci.* **41**, 470–482 (2018).
94. Bellesi, M., Riedner, B. A., Garcia-Molina, G. N., Cirelli, C. & Tononi, G. Enhancement of sleep slow waves: underlying mechanisms and practical consequences. *Front. Syst. Neurosci.* **8**, 208 (2014).
95. Papalambros, N. A. et al. Acoustic enhancement of sleep slow oscillations and concomitant memory improvement in older adults. *Front. Hum. Neurosci.* **11**, 109 (2017).
96. Johns, M. W. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* **14**, 540–545 (1991).
97. Levine, D. W. et al. Validation of the Women's Health Initiative Insomnia Rating Scale in a multicenter controlled clinical trial. *Psychosom. Med.* **67**, 98–104 (2005).
98. Horne, J. A. & Ostberg, O. A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *Int. J. Chronobiol.* **4**, 97–110 (1976).
99. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
100. Weaver, T. E. et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep* **20**, 835–843 (1997).
101. Rechtschaffen, A. & Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (United States Government Printing Office, 1968).
102. Feinberg, I. & Floyd, T. C. Systematic trends across the night in human sleep cycles. *Psychophysiology* **16**, 283–291 (1979).
103. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* **29**, 306–310 (1970).
104. Tenke, C. E. & Kayser, J. Reference-free quantification of EEG spectra: combining current source density (CSD) and frequency principal components analysis (fPCA). *Clin. Neurophysiol.* **116**, 2826–2846 (2005).
105. Miller, K. J. et al. Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Comput. Biol.* **8**, e1002655 (2012).
106. Duffy, F. H. & Als, H. A stable pattern of EEG spectral coherence distinguishes children with autism from neuro-typical controls—a large case control study. *BMC Med.* **10**, 64 (2012).
107. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
108. Ayoub, A. et al. Differential effects on fast and slow spindle activity, and the sleep slow oscillation in humans with carbamazepine and flunarizine to antagonize voltage-dependent Na^+ and Ca^{2+} channel activity. *Sleep* **36**, 905–911 (2013).

109. Zeitlhofer, J. et al. Topographic distribution of sleep spindles in young healthy subjects. *J. Sleep Res.* **6**, 149–155 (1997).
110. De Gennaro, L. & Ferrara, M. Sleep spindles: an overview. *Sleep Med. Rev.* **7**, 423–440 (2003).
111. Massimini, M., Huber, R., Ferrarelli, F., Hill, S. & Tononi, G. The sleep slow oscillation as a traveling wave. *J. Neurosci.* **24**, 6862–6870 (2004).
112. Dang-Vu, T. T. et al. Spontaneous neural activity during human slow wave sleep. *Proc. Natl Acad. Sci. USA* **105**, 15160–15165 (2008).
113. Muehlroth, B. E. & Werkle-Bergner, M. Understanding the interplay of sleep and aging: methodological challenges. *Psychophysiology* **57**, e13523 (2020).
114. Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice* (The MIT Press, 2014).
115. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018).
116. Campello, R. J. G. B., Moulavi, D. & Sander, J. in *Advances in Knowledge Discovery and Data Mining* Vol. 7819 (eds Pei, J. et al.) 160–172 (Springer, 2013).

Acknowledgements

MESA and the MESA SNP Health Association Resource project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the NHLBI, and by grants UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, DK063491, K24 AG045334, P30AG059303 and R01AG058969. Funding support for the sleep PSG dataset was provided by grant HL56984. Funding for SNP Health Association Resource genotyping was provided by NHLBI contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, United States) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, United States) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank the other investigators, the staff and the participants of the MESA study for valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. The MrOS is supported by National Institutes of Health (NIH) funding. The National Institute on Aging (NIA), National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), National Center for Advancing Translational Sciences (NCATS) and NIH Roadmap for Medical Research provide support under the following grant numbers: U01 AG027810; U01 AG042124; U01 AG042139; U01 AG042140; U01 AG042143; U01 AG042145; U01 AG042168; U01 AR066160; and UL1 TR000128. The NHLBI provides funding for the MrOS Sleep ancillary study 'Outcomes of Sleep Disorders in Older Men' under the following grant numbers: R01 HL071194; R01 HL070848; R01 HL070847; R01 HL070842; R01 HL070841; R01 HL070837; R01 HL070838; and R01 HL070839. In addition, this work was also supported by NIH/National Institute of Mental Health grant R03 MH108908 (to S.M.P.), NIH/NHLBI grant R01 HL146339 (to S.M.P.), NIH/NHLBI grant R21 HL145492 (to S.M.P.), NIH/National Institute on Minority Health and Health Disparities grant R21 MD012738 (to S.M.P.), NIH grant K23AG049955 (to J.M.D.), NIH/NHLBI grant K01HL138211 (to D.J.), NIH/NHLBI grant R35 HL135818 (to S.R.), NIH/National Institute of Neurological Disorders and Stroke grant R01 NS096177 (to M.J.P.), NIH/National Institute on Aging grant R01 AG054081 (to M.J.P.), K24 AG045334 (to J.A.L.), a Beth Israel Deaconess Medical Center Neurology Department Grant (to I.D.) and NIH/NHLBI grant R24 HL114473 (to S.R., S.M. and S.M.P.). This work is, in part, also a publication of the US Department of Agriculture/Agricultural Research Service (USDA/ADS) Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine (Houston, Texas), funded in part by the USDA/ADS (cooperative agreement 58-3092-5-001). The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products or organizations imply endorsement from the US Government. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

I.D., S.R. and S.M.P. conceived of and planned the study. S.R.R., A.L.F., A.C.W., T.S., H.T.N., J.A.L. and S.R. collected the primary sleep and cognitive data in MESA. K.L.S., G.J.T., K.Y. and S.R. collected the primary sleep and cognitive data in MrOS. S.M.P. developed the analytical software and approach. I.D., S.M., M.J.P., V.M.G.T.H.V.D.K., D.J., J.M.D. and K.E.B. discussed the analytical approach/results. S.M.P., I.D. and S.R. drafted the manuscript. All authors reviewed and commented on the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00964-y.

Correspondence and requests for materials should be addressed to S.M.P.

Peer review information Primary Handling Editor: Marike Schiffer.

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

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

nature research

Corresponding author(s): Shaun M. Purcell

Last updated by author(s): Aug 17, 2020

Reporting Summary

Nature Research 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 Research 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	Profusion (version 3) software was used to record, annotate and manually score the polysomnograms
Data analysis	Luna (v0.23), an open-source C/C++ software package for the analysis of sleep signal data (written by S.M.P.) and R (version 3.6.3) were used to analyze the data. Luna is available at http://zzz.bwh.harvard.edu/luna/ and http://github.com/remnrem/luna-base/

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

MESA and MrOS datasets (polysomnography and clinical/covariate data) are available through the National Sleep Research Resource (NSRR). These data are available to all interested parties, pending completion of a Data Access and Use Agreement and IRB, as described on the NSRR web site (http://sleepdata.org). Specifically, MESA data (DOI: 10.25822/n7hq-c406) are available from this URL: https://sleepdata.org/datasets/mesa. MrOS data (DOI: 10.25822/kc27-0425) are available from this URL: https://sleepdata.org/datasets/mros.

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	Sample size was not determined a priori based on statistical considerations; rather, we obtained two large cohorts with available sleep and cognitive data available, providing a total sample size considerably greater than observed in the literature on sleep EEG, cognition and aging. We have 80% to detect correlation coefficients (Pearson's r) as low as 0.07 and 0.06 in MESA and MrOS respectively, for nominal p<0.05 significance levels.
Data exclusions	Individuals with poor quality sleep signal data were excluded from analysis (465 of 2060 from MESA and 687 of 2911 from MrOS). See Supplementary Table S3 for full details.
Replication	Results from the MESA sample were tested for replication in the independent MrOS sample. As we observed a high degree of correspondence between MESA and MrOS results, the final analyses were based on a joint analysis of both samples, selecting only metrics with consistent and significant results in both samples.
Randomization	Randomization was not relevant to this study, as there was no experimental or therapeutic group/condition.
Blinding	For all manual scoring and annotation of polysomnograms, technicians were blind to the cognitive and medical status of participants. In the primary statistical analyses, as this study was not based on allocation to experimental group, blinding was not relevant. An identical analytic pipeline was applied to all polysomnograms to extract derived sleep metrics, and this automated pipeline was agnostic to outcome (i.e. cognitive status).

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The Multi-Ethnic Study of Atherosclerosis (MESA) is a prospective observational cohort study consisting of a diverse population sample of 6,814 men and women between the ages of 45 and 84 years, who were free of clinically apparent cardiovascular disease at the time of enrollment, between 2000 and 2002. The study intended to prospectively investigate risk factors for progression of subclinical atherosclerosis to clinical disease. Of the recruited participants, 38% were white, 28% African-American, 22% Hispanic, and 12% Asian, predominantly of Chinese descent. They were recruited from six communities in the United States: i) Baltimore City and Baltimore County, Maryland; ii) Chicago, Illinois; iii) Forsyth County, North Carolina; iv) Los Angeles County, California; v) Northern Manhattan and the Bronx, New York; vi) St. Paul, Minnesota. The MrOS Sleep Study is an ancillary study of the Osteoporotic Fractures in Men Study, which was initially designed to examine the risk factors related to osteoporosis and osteoporotic fractures in older men and to determine the effect of fractures on their quality of life. In the MrOS study, baseline examinations were conducted from 2000 to 2002. A total of 5,994 community-dwelling men aged 65 y or older were enrolled at six clinical centers in the United States: Birmingham, AL; Minneapolis, MN; the Monongahela Valley near Pittsburgh, PA; Palo Alto, CA; Portland, OR; and San Diego, CA. Older men were eligible for the parent study if they could walk without assistance, had no history of bilateral hip replacement, resided near a clinical site for the duration of the study, and had no medical condition that would result in imminent death. From

	December 2003 through March 2005, MrOS participants were invited to participate in the ancillary study, the MrOS Sleep Study. Its purpose was to investigate the relationships between sleep, incident cardiovascular disease, and other age-related outcomes in older men. Men were screened for nightly use of mechanical devices during sleep including pressure mask for sleep apnea (e.g., continuous positive airway pressure or bilevel positive airway pressure), mouthpiece for snoring or sleep apnea, or oxygen therapy and were excluded if they could not forgo use of these devises during PSG recording. A total of 3,135 participants were recruited for sleep measures. All participants provided written informed consent, and the study was approved by the Institutional Review Board at each site.
Recruitment	Both MESA and MrOS were designed to be community-based, epidemiology samples that are broadly representative of the population of interest (older adults). MESA: Each of six sites aimed to recruit 1,100 participants, equally divided between men and women. Wake Forest, Johns Hopkins, Minnesota, and Northwestern all plan to start by creating community awareness of the study and enlisting the support and endorsement of community-based organizations and leadership. All sites implemented techniques that have been used successfully in other studies to recruit minority populations. Columbia worked closely with the 1199 National Benefit Fund during recruitment, including using study staff hired through the union for recruitment, retention, and study publicity. UCLA will recruit using random-digit dialing. All sites, which are recruiting Hispanics, will employ staff fluent in Spanish, and sites recruiting Chinese-Americans will employ staff fluent in Cantonese and Mandarin. Prior to recruitment, the purpose, rationale, and design of the study was publicized to residents of target areas at each site. Successive efforts were directed at targeted individuals, and included mailings of letters and brochures, followed by personal contacts via telephone or in person. Sites modified these materials slightly to meet unique aspects of the source population and recruitment strategy. Standard press releases were written, and templates developed for participant letters, brochures, and scripts across sites. MrOS: Key methods included mailings using community and provider contact lists; regional and senior newspaper advertisements; and presentations targeted to seniors. Sites used a centrally developed recruitment brochure.
Ethics oversight	IRBs at each MESA and MrOS site approved this research.

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