



Full Length Articles

Sleep reverts changes in human gray and white matter caused by wake-dependent training



Giulio Bernardi^{a,b,c}, Luca Cecchetti^{b,c}, Francesca Siclari^a, Andreas Buchmann^a, Xiaoqian Yu^a, Giacomo Handjaras^{b,c}, Michele Bellesi^a, Emiliano Ricciardi^{b,c}, Steven R. Keckskemeti^d, Brady A. Riedner^a, Andrew L. Alexander^{a,d,e}, Ruth M. Benca^a, M. Felice Ghilardi^f, Pietro Pietrini^{b,c,g,*}, Chiara Cirelli^{a,**}, Giulio Tononi^{a,**}

^a Dept. of Psychiatry, University of Wisconsin, Madison, WI 53719, USA

^b Laboratory of Clinical Biochemistry and Molecular Biology, University of Pisa, Pisa 56126, Italy

^c Clinical Psychology Branch, University of Pisa, AOUP Santa Chiara, Pisa 56126, Italy

^d Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin, Madison, WI 53705, USA

^e Dept. of Medical Physics, University of Wisconsin, Madison, WI 53705, USA

^f Dept. of Physiology and Pharmacology, City University of New York Medical School, New York, NY 10031, USA

^g IMT School for Advanced Studies Lucca, Lucca 55100, Italy

ARTICLE INFO

Article history:

Received 20 October 2015

Accepted 9 January 2016

Available online 23 January 2016

Keywords:

Mean diffusivity

DWI

MRI

Sleep deprivation

Extracellular space

ABSTRACT

Learning leads to rapid microstructural changes in gray (GM) and white (WM) matter. Do these changes continue to accumulate if task training continues, and can they be reverted by sleep? We addressed these questions by combining structural and diffusion weighted MRI and high-density EEG in 16 subjects studied during the physiological sleep/wake cycle, after 12 h and 24 h of intense practice in two different tasks, and after post-training sleep. Compared to baseline wake, 12 h of training led to a decline in cortical mean diffusivity. The decrease became even more significant after 24 h of task practice combined with sleep deprivation. Prolonged practice also resulted in decreased ventricular volume and increased GM and WM subcortical volumes. All changes reverted after recovery sleep. Moreover, these structural alterations predicted cognitive performance at the individual level, suggesting that sleep's ability to counteract performance deficits is linked to its effects on the brain microstructure. The cellular mechanisms that account for the structural effects of sleep are unknown, but they may be linked to its role in promoting the production of cerebrospinal fluid and the decrease in synapse size and strength, as well as to its recently discovered ability to enhance the extracellular space and the clearance of brain metabolites.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Sleep deprivation has long been known to result in longer and/or deeper sleep. Recent studies, however, show that sleep need increases not only with the duration of wake, but also with its “intensity”, and specifically with the amount of experience-dependent plasticity and learning, a finding confirmed in insects, rodents, and humans (Tononi and Cirelli, 2014). Fruit flies, for instance, sleep longer after being awake in an enriched environment than in isolation (Bushey et al., 2011; Donlea et al., 2009; Ganguly-Fitzgerald et al., 2006). In mammals,

slow wave activity (SWA), the EEG power between 0.5 and 4.5 Hz during non-rapid eye movement (NREM) sleep, is an established marker of sleep need and intensity, since it increases with wake duration, declines in the course of sleep, and is positively correlated with arousal threshold during sleep (Vyazovskiy et al., 2011). Rats that spent time exploring new objects later show higher NREM SWA relative to rats that ignored the objects, even though wake duration was the same in all animals (Huber et al., 2007). In humans, high-density EEG (hd-EEG) experiments also show that SWA can be regulated locally, depending on the specific wake experience. For example, SWA peaks in left frontal cortex after training in a language task, and in parietal regions after learning a visuo-motor task (Huber et al., 2004; Hung et al., 2013). Thus, there is strong electrophysiological evidence that wake-related learning and sleep need are linked.

Long or enriched wake also leads to structural changes in neurons. In the fly brain, dendritic branches and synaptic puncta increase with wake and decrease with sleep (Bushey et al., 2011; Donlea et al.,

* Correspondence to: P. Pietrini, IMT School for Advanced Studies Lucca, Piazza San Ponziano, 6, I-55100 Lucca, Italy.

** Corresponding authors at: Department of Psychiatry, University of Wisconsin—Madison, 6001 Research Park Blvd, 53719 Madison, WI, USA.

E-mail addresses: pietro.pietrini@imtlucca.it (P. Pietrini), ccirelli@wisc.edu (C. Cirelli), gtononi@wisc.edu (G. Tononi).

2009; Donlea et al., 2011). In the adolescent mouse cortex, wake leads to net spine formation while sleep results in net spine elimination (Maret et al., 2011; Yang and Gan, 2012). Electron microscopy studies also show that wake/sleep dependent structural changes also occur in astrocytes. Thus, astrocytic processes move closer to the synaptic cleft after short sleep deprivation, and astrocytic coverage of cortical spines increases after chronic sleep loss (Bellesi et al., 2015). While direct evidence for similar changes in the human brain cannot be easily achieved, diffusion weighted imaging (DWI) is widely used to derive several indices reflecting the micron-scale density and organization of brain tissues (Sagi et al., 2012). In particular, mean diffusivity (MD) – a measure of tissue density based on the rate of water diffusion – has been proposed as a potential marker for the detection of relatively rapid changes in the microstructure of gray matter (GM) and white matter (WM). Indeed, careful investigations have shown that MD decreases in hippocampus, parahippocampus and fornix after just a few hours of visuo-spatial training (Hofstetter et al., 2013; Sagi et al., 2012; Tavor et al., 2013). Similar changes occur in the rat hippocampus, presumably due to an increase in glial cell volume and/or a decrease in extracellular space (Hofstetter et al., 2013; Sagi et al., 2012; Tavor et al., 2013). These findings, however, raise some fundamental questions. If the brain's ultra-structure can be altered so quickly after a few hours of training, what happens if subjects continue to practice? Do these structural changes continue to accumulate if subjects are kept awake and continue to practice a task at night, when they would normally be asleep? And if so, can sleep revert them? To address these questions we performed structural Magnetic Resonance Imaging (MRI) and DWI during the physiological sleep/wake cycle, after 12–24 h of intense task training, and after post-training sleep.

Material and methods

Participants

Sixteen healthy volunteers (age 24.0 ± 3.4 years, 8 females; 13 right-handed) were recruited from the University of Wisconsin-Madison campus. All participants had sleep duration of ~ 7 h/night, consistent bed/rise times, no daytime nap habit, no excessive daytime sleepiness (total scores in the Epworth Sleepiness Scale ≤ 10) and no history of sleep, medical, or psychiatric disorders as assessed by a clinical interview and by one 8 h night sleep recording with hd-EEG. Polysomnographic parameters (see Table 4), including total sleep time and the percentage of different sleep stages were comparable to those of healthy individuals of similar age (Ohayon et al., 2004). Sleep scoring was performed over 30 s epochs according to standard criteria by a sleep medicine board certified physician (Silber et al., 2007). Subjects were asked to maintain a regular sleep–wake schedule for at least one week before each experiment, and compliance was verified with sleep diaries and wrist-worn actimeters (Actiwatch 64, MiniMitter). Use of

alcohol and caffeine-containing beverages was prohibited starting the day of the first MRI scan and throughout each experiment. The study was approved by the local IRB. Each participant signed an IRB approved informed consent form before enrollment into the study.

Experimental design

This study was part of a project assessing the effects of extended wake with training and post-training sleep on EEG, behavioral, and structural measures (Bernardi et al., 2015). Each subject participated in two experiments (DS and EF, see below), spaced at least 2 weeks apart (Fig. 1). Each experiment included 5 consecutive MRI sessions (every ~ 12 h) with both functional and structural scans, all occurring in quiet wake: 1) W^B (wake baseline) at ~ 7 pm, after a wake day spent outside the lab without any specific training; 2) S^B (sleep baseline) the next morning at ~ 8 am, after subjects slept at home as usual; 3) W^{T12} (wake with training) ~ 8 pm, after 12 h of wake with extensive training in the lab; 4) W^{T24} (extended wake with training) ~ 8 am, after 24 h of continuous wake with extensive training in the lab; 5) S^R (sleep recovery) ~ 8 pm, after ~ 8 h of recovery sleep with hd-EEG recording in the lab (256 channels; Electrical Geodesics Inc.; recovery sleep onset ~ 10 am). During the 24 h of continuous wake all subjects completed six 2 h-training sessions (12 h total of training) of either a mouse-controlled driving simulation game (DS experiment), or a battery of tasks based on impulse control, decision-making and conflict resolution (*executive functions*, EF experiment). As discussed in previous work (Bernardi et al., 2015), the two tasks were selected to involve cognitive domains and brain cortical networks that were as distinct as possible, namely a bilateral occipito-parietal and motor network for DS and a network that includes inferior frontal gyrus, medial prefrontal cortex, cingulate cortex and pre-supplementary motor area for EF. The order of the two experiments was randomly assigned and counterbalanced across subjects. During the 24 h of wake, subjects alternated between 2 h-training sessions of task practice (DS or EF) and ~ 1 h blocks of behavioral tests and hd-EEG recordings. Each test-block included two 4 min eyes-open and eyes-closed recordings, a 5 min psychomotor vigilance test (PVT), 3 trials of a response inhibition test, 3 trials of a visuo-motor coordination test, and self-rating questionnaires that were used to assess subjective sleepiness (Bernardi et al., 2015). Two experimenters took turns attending to the participants to prevent them from falling asleep and to ensure adherence to the protocol throughout the experiment.

MRI data acquisition

During each MRI session (3 T scanner, Discovery MR750, GE Healthcare) subjects underwent a 5 min eyes-closed EPI resting-state scan (as reported in previous work Bernardi et al., 2015) and a high-resolution 3D inversion-prepared fast spoiled gradient echo (IR-

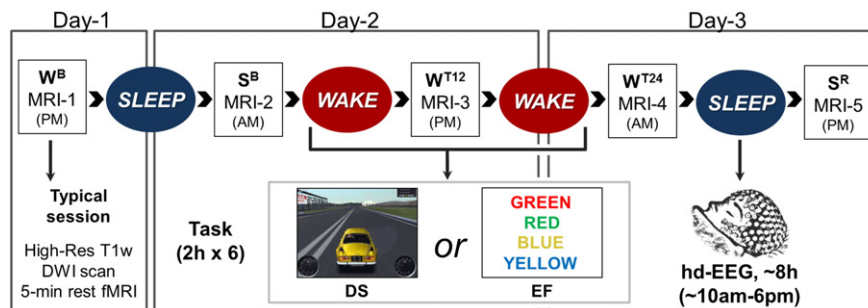


Fig. 1. Experimental design. Each subject completed two experiments (DS and EF), spaced at least 2 weeks apart. During both experiments MRI sessions were performed every ~ 12 h: W^B (wake baseline), after a wake day spent outside the lab without any specific training; S^B (sleep baseline) the next morning, after subjects slept at home as usual; W^{T12} (wake with training), after 12 h of wake with extensive training in the lab; W^{T24} (extended wake with training), after 24 h of continuous wake with extensive training in the lab; S^R (sleep recovery), after ~ 8 h of recovery sleep with hd-EEG recording in the lab. Each MRI session included a high-resolution anatomical scan, a 5-min eyes-closed resting state functional scan, and a diffusion weighted scan (DWI).

fSPGR) T1-weighted (T1w) anatomical scan (inversion time: 450 ms, repetition time: 8.2 ms, echo time: 3.2 ms, flip angle: 12°, voxel size: $1 \times 1 \times 1$ mm, in-plane matrix: 256×256 , number of slices: 156). DWI data were acquired using repetition time = 7000 ms, echo time = 66.3 ms, flip angle = 90°, acquisition matrix = 96×96 , field of view = 230 mm, in-plane resolution = 2.396×2.396 mm (resolution after on-scanner interpolation = 0.898×0.898 mm), and slice thickness = 2.3 mm with no gap. Diffusion-sensitizing gradient encoding was applied in 52 directions with three different diffusion-weighted factors, corresponding to $b = 400, 800$ and 1200 s/mm². Six images (b_0 image) were acquired without use of a diffusion gradient. For each encoding direction, 61 axial images were acquired to cover the entire brain. Due to technical problems, MRI data were not obtained in subjects S05 and S14 (males, 1 left handed) during experiment EF. In addition, structural data were not obtained in S^B of subjects S07 and S^R of subject S15 during experiment EF (females, right handed), and in session W^B of subject S08 (female, right handed) during experiment DS.

Structural data preprocessing

High-resolution T1w images were automatically processed using the *Freesurfer* longitudinal pipeline (Reuter et al., 2012). Differently from common procedures for structural analysis, the longitudinal approach allows to obtain more reliable cortical and subcortical morphological measurements by incorporating temporal information. Specifically, an unbiased within-subject template was created using robust, inverse consistent registration (Reuter et al., 2010). Then, subsequent preprocessing steps, including skull stripping, standard-space transformation, atlas registration and spherical surface maps generation and parcellations were performed using common information from the within-subject template (Reuter et al., 2012). Thus, for each subject and time-point, the software automatically assigned neuroanatomical labels to each brain location using probabilistic information estimated from both geometric data derived from the cortical model and neuroanatomical convention obtained from a pre-labeled training set (Desikan et al., 2006; Fischl et al., 2004). Importantly, this parcellation strategy allowed to minimize potential biases related to inter-subject anatomical differences and alignment issues caused by data transformation in a common reference system, as well as to take into account relative within-subject structural variations.

Analysis of diffusion images was performed using the *FSL* software package (Smith et al., 2004). For each session and subject, all diffusion weighted and b_0 images were affinely coregistered to the b_0 image of the first repetition using *FLIRT* (*FMRIB's Linear Image Registration Tool*; Jenkinson and Smith, 2001), to correct both for eddy current induced distortion (*eddy_correct tool*) and subject's motion effects. Moreover, computed motion parameters were used to adjust the direction of gradient vectors. A brain mask was created from the first b_0 image using *BET* (*Brain extraction Tool*; Smith, 2002) and used to constrain the tensor fitting within voxels of interest. A *linear least squares* (LLS) approach was used to fit the tensor models at each voxel (*FDT*, *FMRIB's Diffusion Toolbox*; Behrens et al., 2003) and compute the MD maps (of note, partially different absolute MD values, but analogous statistical results were obtained if the *weighted linear least squares* approach was used instead). Finally, DWI data were aligned to the *Freesurfer* within-subject template using a two-step procedure. First, using an affine linear registration (*FLIRT*, 12 degrees of freedom), the average b_0 image of each DWI scan was coregistered to the b_0 image obtained in the same MRI session of the T1w image representing the reference of the anatomical template. Then, *3dQwarp* (Cox, 2012) was used to perform a constrained non-linear registration to the obtained b_0 images and match the within-subject anatomical template. Importantly, a cost function that is insensitive to contrast differences (*mutual information*) was adopted to optimize correction of geometrical distortions in DWI data (Gholipour et al., 2006). Resulting transformation matrices and

deformation fields were then concatenated and applied to individual MD maps (Fig. S1).

Brain structural changes associated with intensive practice and sleep deprivation

Global and regional changes in structural measures

MD was measured in four large regions of interest (ROIs; Fig. 2): cortical gray matter (GM), subcortical GM, white matter (WM) and ventricles. In the same ROIs we also measured volume (cortical and subcortical GM, WM and ventricles). Cortical GM thickness was measured instead of cortical volume because the latter is affected both by cortical thickness and by area, and thus its variations may be more difficult to interpret (Winkler et al., 2010). These measures were extracted for each subject, experimental condition (DS, EF) and available time point (1–5). To improve the accuracy of MD calculation for each ROI, structure-specific masks were created for each time-point and subject, using the segmentation information previously obtained in *Freesurfer*. In addition, to minimize biases related to potential changes in the number of voxels included in each time point, we created a unique conjunction mask (logical AND) for each subject and structure of interest (Table S1). Thus, only voxels for which the 'structural labeling' did not change across scans were included in the MD analysis. Mean MD values were subsequently extracted from the obtained masks. For measures of MD, volume, and cortical thickness, we performed a three-way repeated measures (rm)-ANOVA in SPSS Statistics 21 (IBM Corporation), including training condition (W^B, S^B vs. W^{T12}, W^{T24}), time-of-day (8 am vs. 8 pm), and task (DS vs. EF) as within-subjects factors. To ensure a balanced design with no missing data, only the 12 subjects having all time points were included in these analyses. However, to better characterize reliability of potential findings, analyses were also repeated using analytical models allowing for the inclusion of all subjects ($N = 16$) and time-points (linear mixed effect analysis, LME). Results of these auxiliary analyses are reported in Table S2. Planned post-hoc comparisons tested the effects of normal sleep (W^B vs. S^B), 12 h of wake with training (W^B vs. W^{T12}), and sleep deprivation with training (W^{T12} vs. W^{T24}), as well as differences between morning and evening time points (morning: S^B vs. W^{T24} ; evening: W^B vs. W^{T12}). Statistical significance accounting for multiple comparisons was assessed by applying the Bonferroni–Holm adjustment. Finally, the effect of recovery sleep was investigated through paired t -tests comparing W^{T24} and S^R .

To investigate smaller scale structural modifications and evaluate local effects related to the practiced task we also performed additional ROI-based analyses using the 40 cortical and subcortical regions of the Desikan–Killiany Atlas (Fischl et al., 2004). Given that no lateralization of the effects of interest was expected, extracted values of MD, cortical thickness, or volume were averaged across homolog areas of the two hemispheres. Then, independent rmANOVAs were performed at each ROI for either thickness/volume or MD, and the obtained p -value of

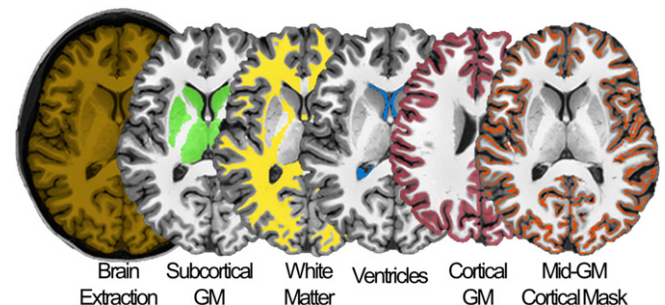


Fig. 2. Large-scale ROIs used for the analyses of global brain changes. The four ROIs (here shown in a representative subject) are: [green] subcortical gray matter (GM), [yellow] white matter (WM), [blue] ventricles and [red] cortical GM. Moreover, a mid-GM cortical mask was generated in order to eliminate voxels characterized by a high probability of containing mixed tissues (see text for details).

each tested effect was subsequently adjusted to account for multiple comparisons. Specifically, a Bonferroni–Holm correction was applied across the 40 ROIs and the significance threshold was set to *corrected* $p < 0.05$.

Investigation of partial volume effects

MD measurements are known to be potentially affected by the so called “partial volume effect”, that is, the presence of multiple tissue types within the same voxels (Alexander et al., 2001). Specifically, CSF contamination in GM or WM may be responsible for erroneous estimations of diffusivity measures. We addressed this issue using a combination of different approaches. First, we performed a serial visual inspection of MD distributions in each ROI to exclude the possible influence of partial volume effects. Specifically, for each subject and time point we divided MD values in $0.1 \times 10^{-3} \text{ mm}^2/\text{s}$ bins (from 0 to $3 \times 10^{-3} \text{ mm}^2/\text{s}$) and determined the percentage of voxels included in each bin. The distributions of the group-averaged values in the baseline condition without training (S^B) and in the experimental condition with training (W^{T12} , W^{T24}) were plotted and qualitatively compared. Second, in order to further evaluate the possible influence of partial volume artifacts on our results, we repeated the cortical GM analysis using an approach based on a “skeletonized” cortical mid-GM mask excluding voxels with a high probability of containing multiple tissues (Ball et al., 2013). Specifically, segmentation maps obtained in *Freesurfer* were initially used to identify the WM-GM and the GM-CSF boundaries. Then, a mid-GM mask was defined as the line passing at equal distance from the two boundaries (i.e., 50% of cortical thickness) in each point. Finally, voxels lying along this line, and included in the previously defined conjunction cortical mask, were used to create a final “mid-GM mask” (Figs. S2 and S3). This procedure allowed to retain only voxels relatively distant from both the WM-GM and the GM-CSF boundaries, and thus characterized by a minimal probability of containing mixed tissues. The obtained mid-GM mask included $60.4 \pm 1.4\%$ less voxels than the conjunction cortical mask. Cortical MD values were extracted from this new mask and analyzed as previously described. Moreover, given that a few voxels affected by CSF contamination may have remained included also in the mid-GM mask, the above calculations were repeated after further discarding voxels containing MD values greater than $1.0 \times 10^{-3} \text{ mm}^2/\text{s}$.

Investigation of other potential confounding factors

Recent findings suggest that the human brain may undergo morning-to-evening size variations and it has been suggested that these changes may depend on a redistribution of body fluids (Nakamura et al., 2015). However, global or structure-specific volumetric variations may also reflect other underlying phenomena potentially associated with changes in water diffusivity (e.g., compression). In these conditions MD and volumetric measures are expected to covary. These observations highlighted the need to evaluate global brain volumetric changes in our samples and to explore their possible relationship with MD measures. Importantly, however, recent studies also suggested that head movement could lead to erroneous estimations of diffusivity-based measures and recommended to include head motion as a nuisance variable in statistical analyses to reduce the risk of biases related to this potential confounding factor (Yendiki et al., 2014). Therefore, statistical LME models were specifically computed with the inclusion of covariates represented by the total in-scanner head movement (estimated from the coregistration of DWI volumes using the RMS deviation measure; Jenkinson, 1999) and either the total brain volume (calculated here as the sum of the volumes of cortical GM, subcortical GM and WM) or the total volume of each examined structure (e.g., cortical GM). The models included the same within subject factors introduced in the rmANOVAs (task, time-of-day, training condition). Finally, relative global volumetric changes were also investigated using a rmANOVA, as previously described.

Relationship between sleep parameters and structural recovery

We tested whether global changes in MD, volume or cortical thickness that occurred after recovery sleep (S^R vs. W^{T24}) correlated with different sleep parameters, including total SWA and changes in SWA and slow wave amplitude. EEG recordings were first-order high-pass filtered (0.1 Hz) and band-pass filtered between 0.5 and 58 Hz. For scoring purposes, four of the 256 electrodes placed at the outer canthi of the eyes were used to monitor eye movements (electro-oculography), while electrodes located in the chin-cheek region were used to evaluate muscular activity (electromyography). Due to technical problems during the recordings, EEG data were not available in subjects S03 and S07 (experiments EF and DS, respectively). Bad channels were visually identified, rejected, and replaced with data interpolated from nearby channels using spherical splines (NetStation, Electrical Geodesics Inc.). SWA activity was calculated for each NREM epoch as the spectral power in the range between 0.5 and 4.5 Hz. Specifically, after excluding electrodes located on the neck/face region, the signal of each channel was re-referenced to the average of the remaining 185 electrodes, and the power spectral density estimates were computed using the Welch's method (*pwelch* function, MATLAB signal processing toolbox) in 2 s data segments (Hamming windows, 8 sections, 50% overlap). The resulting power spectral densities in the SWA range were then averaged across the 185 electrodes and within each epoch.

For the slow wave detection procedure, preprocessed EEG signals from each NREM epoch and channel were initially referenced to the average of the two mastoid electrodes. Then, an automatic detection algorithm adapted from a previous study (Siclari et al., 2014) was applied. Specifically, we first created a single timing reference by calculating the negative-going signal envelope, defined as the 0.025 quantile of the signal values detected across all channels for each point in time. The resulting signal was broadband filtered (0.5–40 Hz, stop-band at 0.1 and 60 Hz) prior to the application of the wave detection. Only slow waves with a duration of 0.25–1.25 s between consecutive zero crossings were further evaluated. Additional criteria were applied to exclude negative signal deflections of potential artifactual origin. Specifically, for each slow wave we first calculated the scalp involvement as the mean signal achieved in the 20 ms around the wave peak in each channel. Then, the top 5% electrodes showing the lowest EEG signal values (maximal involvement) were identified. A slow wave was discarded if at least half of these electrodes were located in the neck/face regions (possible muscular artifact) or around both eyes but not in the medial and lateral frontal areas (possible ocular artifact).

Finally, for each subject and experimental condition, we calculated the total number of detected slow waves, the total SWA (defined as the mean SWA computed across all NREM epochs), and the variation in slow wave amplitude and SWA from the first to the last NREM sleep cycle (difference last-first). The Pearson's coefficient was used to investigate the potential correlation between structural changes following recovery sleep and examined sleep parameters ($p < 0.05$, Bonferroni–Holm correction). The correlations with total sleep time and N2, N3 and REM time were also examined.

Analysis of the relationship between structural changes and behavioral correlates

Relationship between structural changes and vigilance levels

Vigilance levels were measured by calculating the mean reaction time during PVT (Bernardi et al., 2015) for test blocks completed in temporal proximity with each MRI scan except the first one (PVT was not performed before the first scan). A machine learning procedure was developed using MATLAB (The MathWorks, Inc.) and LibSVM (Wang et al., 2011) to assess the potential relationship between structural changes and variations in vigilance levels. Specifically, using support vector regression (SVR) machines (Drucker et al., 1997), PVT reaction times were predicted from the 8 measures derived from the global structural analyses (MD, volumes and cortical thickness for the 4 ROIs) in both DS

and EF experiments. Reaction times and MRI measures were normalized (subtracting the mean and dividing it by the standard deviation) across sessions and within each subject. A leave-one-subject-out cross-validation procedure (based on the removal of all time-points of the test subject) was then used to train and test linear SVR machines, which resulted in predicted reaction times for the left-out subject across MRI sessions. The mean squared error (MSE) was estimated by comparing real and predicted data. Moreover, to assess the goodness of the prediction, a procedure based on permutation tests was developed (1000 repetitions). Specifically, a MSE null distribution was obtained by training and testing SVR machines with data shuffled across subjects and within each time point. Collected MSE values from the permutation procedure were compared to the MSE estimated from real data with a one-tailed rank test ($p < 0.05$). Importantly, the leave-one-subject-out procedure allowed excluding a possible bias of subject-specific values on the prediction accuracy. Moreover, the null distribution obtained from the permutation test allowed ruling out a possible generalized mean effect related to the examined time points (i.e., an identical effect of the time-of-day in all subjects). The R^2 coefficient between predicted and original data was calculated for each subject.

Relationship between structural changes and cognitive performance

A machine learning procedure similar to the one applied to explore the relationship between structural changes and vigilance level was used on behavioral parameters reflecting the individual performance in the response inhibition test and in the visuo-motor test (Bernardi et al., 2015). As described in previous work, the response inhibition test consisted of a classical Go/NoGo test during which a stream of visual stimuli, that are capital letters X or Y, was presented in alternating order (1 Hz). Subjects were requested to press a button for every stimulus that followed a different stimulus (Go), and to withhold their response each time two identical stimuli followed each other (NoGo). The proportion of *commission errors* (i.e., cases in which the subject responded despite a NoGo stimulus was presented) and the *intraindividual coefficient of variation* (ICV, defined as the standard deviation of reaction time divided by the individual mean) were used as indices of inhibitory efficiency. Differently, during the visuo-motor test, participants were required to perform straight, out and back movements of a tracker, held with the dominant hand, from a central starting area to one of 8 radial targets (time interval = 1.5 s). The *movement time* (time from movement onset to reversal) and the *linear error* (distance of the reversal point from the center of the target) of each movement were collected and used as measures of visuo-motor control efficiency. As described above, impulse control and visuo-motor performance levels were calculated in test blocks completed in temporal proximity with each MRI scan except the first one.

Reliability of the prediction of behavioral parameters from structural data

Given that the present study design does not include independent datasets for training and testing the SVM classifier, a relative risk of “overfitting” is implied in the adopted *leave-one-out* procedure. Thus, in order to evaluate stability and reliability of obtained results, all analyses were repeated while replacing the *leave-one-out* method with a *recursive half-split* of the examined sample. Thus, for each of 1000 iterations, available subjects were equally divided in a training sample, used to train the classifier, and in a test sample, used to evaluate accuracy of the prediction of behavioral performance in the remaining participants. As described for the *leave-one-out* procedure, the goodness of the prediction was assessed using permutation tests through a shuffling of available data across subjects and within time-points. Results of this auxiliary analysis are reported in Table S9.

Results

All subjects except one participated in two experiments (S14 completed only DS; see below). Each experiment included 5 MRI scans acquired following baseline wake without training (W^B), baseline sleep (S^B), 12 h and 24 h of wake with training (W^{T12} , W^{T24}), and post-training recovery sleep (S^R) (Fig. 1). The only difference between the two experiments was in the practiced task, either a driving simulation task (DS) involving visuo-motor areas, or a battery of *executive function* tasks (EF) mainly relying on prefrontal cortex.

Global changes in MD

MD was measured in four large regions of interest (ROIs; Fig. 2, Table S1): cortical gray matter (GM), subcortical GM, white matter (WM) and ventricles. As a first step, a three-way repeated measures (rm)-ANOVA – including training condition (with, without), time-of-day (am, pm), and task (DS, EF) as within-subjects factors – was run using the 12 subjects for which time points from W^B to W^{T24} (4 in DS, 4 in EF) were available (Table 1; also see Table S2). Neither the main effect of task, nor its interactions with other factors, reached statistical significance in any of the examined ROIs. By contrast, a significant main effect of training was identified in cortical GM (*corrected* $p < 0.05$). Importantly, this result was confirmed by statistical models that included total in-scanner head movement and either total brain volume or cortical volume as covariates (Table S3). No significant effects were seen in WM, subcortical GM and ventricles (Fig. S4B–D; Table 1). Next, planned post-hoc comparisons tested the effects of sleep and wake in the absence of training (S^B vs. W^B), training during the first 12 h of wake, i.e. without sleep deprivation (W^{T12} vs. W^B), and sleep deprivation with training (W^{T24} vs. W^{T12}). Finally, the effect of recovery sleep was investigated through paired *t*-tests by comparing S^R and W^{T24} . Relative to baseline sleep (S^B), a decrease in MD was already evident in cortical GM after the first 12 h of wake with training (mean variation \pm SE, computed across subjects after averaging across ROI voxels, S^B – $W^{T12} = -0.52 \pm 0.17\%$), but not after 12 h of baseline wake (Fig. 3A, Tables S4–S5). After 24 h of training, cortical MD continued to decrease (S^B – $W^{T24} = -0.94 \pm 0.15\%$), and this trend was reverted by subsequent recovery sleep.

Of note, typical MD values are substantially different in GM ($\sim 0.8 \times 10^{-3} \text{ mm}^2/\text{s}$), WM ($\sim 0.7 \times 10^{-3} \text{ mm}^2/\text{s}$), and most notably, cerebrospinal fluid (CSF, $\sim 3.0 \times 10^{-3} \text{ mm}^2/\text{s}$). Thus, DWI analyses can be affected by confounds related to the *partial volume effect*, that is, the inclusion of voxels representing a mixture of multiple tissue types (Alexander et al., 2001), with erroneous estimations of diffusivity arising especially because of CSF contamination. However, a serial visual inspection of the distributions of MD values in the cortical ROI allowed to identify a clear shift of the histograms' peak after prolonged training: in fact, W^{T24} had a relatively higher percentage of voxels in the 0.6 – $0.8 \times 10^{-3} \text{ mm}^2/\text{s}$ range, while S^B had relatively more voxels in the 0.8 – $1.0 \times 10^{-3} \text{ mm}^2/\text{s}$ range. No clear differences were observed in the tails of the distributions, which contain voxels with higher probability of being affected by partial volume artifacts. Moreover, we repeated the rmANOVA using MD values extracted from a cortical mid-GM mask that excluded voxels along the WM–GM and the GM–CSF interfaces (Ball et al., 2013). This approach confirmed the existence of a strong and significant effect of training condition in cortical GM, which survived after the further removal of voxels with MD value greater than $1.0 \times 10^{-3} \text{ mm}^2/\text{s}$ (Table S6, Fig. 3B). Of note, the independence of detected MD variations from potential partial volume artifacts was also supported by the additional ROI analysis described below, in which we found that MD changes are relatively widespread (i.e., do not depend on few outlier regions). Finally, we found that MD changes are relatively independent from variations in brain or cortical volumes, and from relative changes in the extent of head movements (Table S3).

Table 1
Results of the rmANOVAs exploring global MD changes. Significant effects surviving after correction for multiple comparisons ($p < 0.05$, Bonferroni–Holm) are marked in bold text.

rmANOVA (MD)	Cortical GM		Subcortical GM		WM		Ventricles	
	<i>P</i>	<i>F</i> _(1,11)	<i>P</i>	<i>F</i> _(1,11)	<i>P</i>	<i>F</i> _(1,11)	<i>P</i>	<i>F</i> _(1,11)
Time (am, pm)	.1557	2.3233	.1069	3.0835	.6917	.1658	.0266	6.5406
Condition (with/out training)	.0002	29.2218	.0716	3.9756	.1619	2.2479	.0318	6.0426
Task (DS, EF)	.3076	1.1448	.2682	1.3601	.3081	1.1419	.1693	2.1642
Time × condition	.0424	5.2651	.3370	1.0080	.0358	5.7178	.1910	1.9419
Time × task	.6952	.1619	.3186	1.0911	.7971	.0694	.6905	.1672
Condition × task	.6925	.1648	.6683	.1937	.3727	.8634	.8815	.0233
Time × condition × task	.6524	.2144	.3548	.9333	.4679	.5653	.2439	1.5160

Regional changes in MD

To test for regional effects, a three-way rmANOVA (training condition, time-of-day, task) was run using the 40 cortical and subcortical regions of the Desikan–Killiany Atlas (Fischl et al., 2004). Significant effects of training condition were found in several brain areas, including superior temporal sulcus, inferior temporal cortex, middle temporal cortex, lateral and medial orbitofrontal cortex (*corrected* $p < 0.05$; Table 2; Fig. S5). Moreover, the rmANOVA revealed a significant time by condition interaction in the superior temporal cortex and in the pars triangularis of the inferior frontal gyrus.

Analyses based on LME models taking into account changes in total head movement and in brain volume were also used to evaluate the potential influence of these possible confounding factors. Importantly, results of this additional analysis confirmed the effects detected in the inferior frontal gyrus and in the middle, inferior and superior temporal cortices. Differently, the main effects of experimental condition identified by the rmANOVA in superior temporal sulcus and medial/lateral orbitofrontal cortex did not reach the set threshold for statistical significance, although a clear trend was observed (*uncorrected* $p < .008$; Table S7). On the other hand, the LME analysis identified an additional condition effect in the fusiform gyrus, and time by condition interactions in the rostral anterior cingulate cortex and in the pars orbitalis of the inferior frontal gyrus. Overall, obtained results point to a distributed effect of the experimental condition on cortical MD, with the strongest variations in temporal and prefrontal brain areas. Finally, the LME models also detected significant main effects of the task (DS, EF) in mid/posterior cingulate cortex and in superior temporal cortex, although no interactions with other examined factors emerged.

Post-hoc tests confirmed the existence of a significant MD decrease after prolonged task practice in ROIs characterized by a significant condition effect or time by condition interaction. Recovery sleep was

associated with a significant MD increase relative to the end of the sleep deprivation period.

Global changes in volume and cortical thickness

We then focused on other structural parameters, namely cortical thickness and volume of subcortical GM, WM, and ventricles. The rmANOVA identified significant main effects of training on all 3 volumetric measures, but not on cortical thickness (*corrected* $p < 0.05$; Table 3; also see Table S2). A time of day effect was present in the ventricles, while neither the main effect of task, nor its interaction with other factors, was significant. Significant time by training interactions were present for cortical thickness and ventricular volume. Planned post-hoc comparisons identified no significant change in cortical thickness between baseline sleep and the first 12 h of training (Fig. 4A, Tables S4–S5), although a relative increase was observed in most subjects ($S^B-W^{T12} = +0.26 \pm 0.19$). Similarly, there was no significant difference between 12 h and 24 h of training, although a trend towards a decrease was present ($W^{T12}-W^{T24} = -0.88 \pm 0.31\%$; $p < 0.05$, *uncorrected*). By contrast, a significant increase in cortical thickness was found after recovery sleep relative to a night of sleep deprivation (Fig. 4A)

Planned post-hoc comparisons for other structural parameters also found that relative to baseline sleep, volumetric changes were present after 12 and/or 24 h of training in all three ROIs, but in different directions, with increases in subcortical GM and WM and decreases in the ventricles (Fig. 4B–D). All these changes were reverted by recovery sleep. No regional changes were found using the ROI-based analysis of thickness and volume in cortical and subcortical structures, with the exception of the thalamus, where a significant increase in volume was found after 24 h of training ($S^B-W^{T24} = +0.64 \pm 0.13\%$).

All described structural changes were not accompanied by significant variations in total brain volume (calculated as the sum of cortical GM, subcortical GM and WM). In fact, the rmANOVA (task, time-of-day, training condition) failed to identify any significant effects (Table S8). Moreover, evening-to-morning changes – specifically explored through paired *t*-test in light of recent findings suggesting the existence of diurnal brain volumetric variations (Nakamura et al., 2015) – were not found in either the “baseline” condition (W^B-S^B : $p > 0.18$) or the “experimental” condition ($W^{T12}-W^{T24}$: $p > 0.31$).

Sleep parameters and structural measures of recovery

All subjects had the possibility to sleep for ~8 h during the day after the end of 24 h of training (Table 4). Although recovery sleep occurred at the wrong circadian time (sleep onset ~10 am), most of its features, including the proportion of NREM (N2 + N3) and REM stages, were not different from those of baseline sleep during the night. However, in all subjects of both DS and EF the latency of the first REM sleep episode was significantly reduced ($p < 0.05$, Bonferroni–Holm correction), a sign of high REM sleep pressure likely due to the combined effect of sleep deprivation and sleeping during the day (Dijk and Czeisler, 1995). Moreover, in both experiments the proportion of time spent in deep sleep (N3) increased in recovery sleep as compared to baseline sleep, mainly at the expenses of the proportion of light sleep (N2).

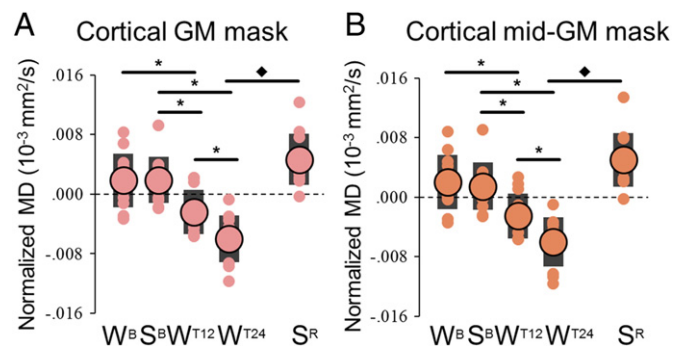


Fig. 3. Comparison of MD variations detected using the cortical GM mask (A) and the mid-GM mask (B) excluding voxels along the WM–GM and GM–CSF interfaces. On average the mid-GM mask included $60.4 \pm 1.4\%$ less voxels than the original cortical mask. Large circles represent the group-level average, while each small circle represents a different subject (black bars indicate one standard deviation from the mean). Values were zero-mean normalized by subtracting the across-time-points mean. *, significant differences for planned comparisons ($p < 0.05$, after Bonferroni–Holm adjustment; $N = 12$); ◆, significant effect of recovery sleep (paired *t*-test).

Table 2

Results of the rmANOVAs exploring regional MD changes. Only statistically significant effects surviving Bonferroni–Holm adjustment are shown ($p < 0.05$). The last column shows the mean relative MD change following 24 h spent awake ($S^B - W^{T24}$ variation). Given that no significant effects of the task were observed, the reported variation was calculated after averaging DS and EF measures. PT, Pars Triangularis.

rmANOVA	Significant effects	Significance level	% MD variation (\pm SE)
Bank superior temporal sulcus	Condition	$p = .0013$; $F_{(1,11)} = 18.498$	–1.14 (.33)
Inferior temporal cortex	Condition	$p = .0004$; $F_{(1,11)} = 25.250$	–1.17 (.25)
Middle temporal cortex	Condition	$p = .0001$; $F_{(1,11)} = 36.456$	–1.16 (.15)
Superior temporal cortex	Time \times condition	$p = .0004$; $F_{(1,11)} = 25.799$	–1.36 (.26)
Lateral orbitofrontal cortex	Condition	$p = .0002$; $F_{(1,11)} = 28.434$	–1.98 (.38)
Medial orbitofrontal cortex	Condition	$p = .0002$; $F_{(1,11)} = 29.604$	–2.70 (.57)
Inferior frontal gyrus (PT)	Time \times condition	$p = .0005$; $F_{(1,11)} = 23.521$	–1.64 (.35)

Actual sleep time, including all non-wake epochs, corresponded to 6.8 ± 1.3 h in DS, and to 7.1 ± 0.9 h in EF. Overall, no significant differences were observed between recovery sleep of DS and that of EF (all $p > 0.05$, *uncorrected*). In both experiments no correlation was found between total sleep time, or time spent in REM, N2 and N3, and any of the structural parameters that showed a significant change between 24 h of training and recovery sleep (GM thickness and MD, subcortical GM volume and MD, WM volume, ventricular volume and MD). Total SWA, total number of detected slow waves, as well as changes in SWA and slow wave amplitude between the first and the last NREM sleep cycle also did not correlate with any structural measure of recovery ($p < 0.05$, *corrected*).

Relationship between structural changes and vigilance level

Next, we examined whether global structural variations were reflected in vigilance changes, measured using the mean reaction time during the PVT (Bernardi et al., 2015), a sustained vigilance task known to be highly sensitive to sleep loss (Basner and Dinges, 2011). In the previous study that used the same subjects we confirmed that PVT performance declines in the course of the 24 h of practice and renormalizes after recovery sleep (Bernardi et al., 2015). Here mean PVT reaction times for test blocks completed in temporal proximity with each MRI scan were predicted from the 8 measures derived from the global structural analyses (MD, volumes and cortical thickness for the 4 large ROIs). The regression procedure identified a significant relationship between structural changes and variations in vigilance level in both DS (average R^2 across subjects \pm SD = 0.46 ± 0.30 ; $p < 0.001$) and EF ($R^2 = 0.47 \pm 0.31$; $p < 0.001$). Thus, variations in global structural measures were able to predict to some extent changes in sustained attention (Fig. 5; also see Table S9). Based on a qualitative evaluation across the two experimental conditions, we found that the most relevant predictors were represented by ventricular volume, MD of cortical GM and WM volume.

Relationship between structural changes and cognitive performance

Finally, by applying the same procedure described above, we tested whether global structural changes were reflected in performance changes in two tests used to track variations in impulse control and visuo-motor coordination. In the previous study (Bernardi et al., 2015), we demonstrated that performance in the response inhibition test, as measured using the number of *commission errors* or the *intraindividual coefficient of variation* in reaction time (ICV), is characterized by a time-course similar to the one described for the PVT test, suggesting a relevant influence of the vigilance level. Indeed, ICV values were successfully predicted by global structural changes in both DS ($R^2 = 0.51 \pm 0.35$; $p < 0.001$) and EF ($R^2 = 0.32 \pm 0.33$; $p = 0.017$) experiments (Fig. S6). Differently, prediction accuracy for *commission errors* reached significance in DS ($R^2 = 0.64 \pm 0.30$; $p < 0.001$) but not in EF ($R^2 = 0.26 \pm 0.27$; $p = 0.053$), perhaps due to lower statistical power in this latter condition ($N = 13$ in EF; $N = 15$ in DS). Of note, however, the quality of the prediction in EF may have been also negatively affected by local, experience-dependent changes in the impulse control network, which seem to be independent from changes in the general vigilance level. Indeed, in line with a possible dissociation between EF performance and vigilance levels, we previously showed that prolonged practice with tasks based on executive functions was associated with a relative performance impairment during the test-block corresponding to W^{T12} , although no differences in the PVT performance were present between DS and EF in the same time-point (Bernardi et al., 2015). With regard to the visuo-motor test (Fig. S7), both *linear error* (LE, a measure of movement accuracy) and *movement time* (MT, an index of eye-hand coordination efficiency) were not predicted by global structural changes in either DS (LE: $R^2 = 0.40 \pm 0.32$; $p = 0.288$ – MT: $R^2 = 0.16 \pm 0.16$; $p = 0.233$) or EF (LE: $R^2 = 0.33 \pm 0.39$; $p = 0.218$ – MT: $R^2 = 0.28 \pm 0.28$; $p = 0.207$). These findings are consistent with previous observations suggesting a lower vulnerability of the visuo-motor function to the effects of sleep deprivation (Bernardi et al.,

Table 3

Results of the rmANOVAs exploring other global structural changes. Significant effects surviving after correction for multiple comparisons ($p < 0.05$, Bonferroni–Holm) are marked in bold text.

rmANOVA (thickness, volume)	Cortical thickness		Subcort. GM volume		WM volume		Ventricular volume	
	<i>P</i>	$F_{(1,11)}$	<i>P</i>	$F_{(1,11)}$	<i>P</i>	$F_{(1,11)}$	<i>P</i>	$F_{(1,11)}$
Time (am, pm)	.1457	2.4520	.6453	.2239	.0596	4.4082	.0024	15.3479
Condition (with/out training)	.7126	.1429	.0071	10.8941	.0116	9.1358	< .0001	42.4533
Task (DS, EF)	.1564	2.3144	.1947	1.9066	.3214	1.0781	.7666	.0926
Time \times condition	.0139	8.5438	.7993	.0679	.2138	1.7408	.0110	9.3109
Time \times task	.2774	1.3058	.3346	1.0184	.4970	.4933	.3915	.7955
Condition \times task	.1768	2.0833	.4821	.5292	.0726	3.9415	.5520	.3765
Time \times condition \times task	.8476	.0387	.3051	1.1572	.3999	.7669	.4830	.5271

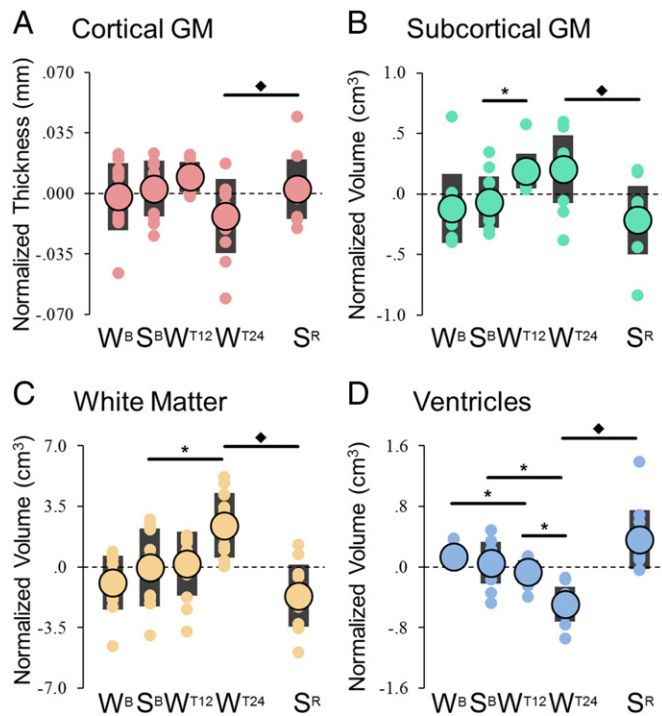


Fig. 4. Global changes in cortical thickness (A) and volume changes in subcortical GM, WM and ventricles (B–D). Large circles represent the group-level average, while each small circle represents a different subject (black bars indicate one standard deviation). Values were zero-mean normalized by subtracting the across-time-points mean. *, significant differences for planned comparisons ($p < 0.05$, after Bonferroni–Holm adjustment; $N = 12$); ♦, significant effect of recovery sleep (paired t -test).

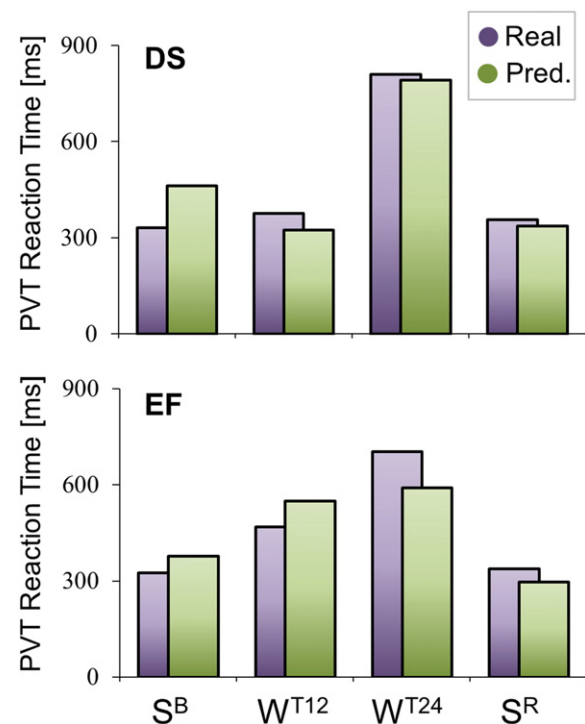


Fig. 5. Prediction of vigilance level based on structural parameters. For both DS (top) and EF (bottom), the graphs show the “real” (purple) vigilance level observed during sessions SB, WT12, WT24 and SR (measured as the reaction time during the PVT test), in comparison with the vigilance level predicted based on global structural parameters (green), in a representative subject (S06).

2015), and potentially reflect a lower dependence of these parameters on the global vigilance state.

Discussion

Extended task practice is associated with a decrease in cortical diffusivity

We found that MD declined in cortical GM after 12 h of wake with intensive task practice, both compared to wake after a night of sleep and relative to 12 h of wake without task practice. Moreover, we

Table 4

Polysomnographic parameters. Sleep parameters (group average \pm SD) during the first baseline night (BL) and the recovery sleep periods of the two experiments (DS, EF). Percentages are expressed relative to total sleep time. Latency of the first N2, N3 and REM period is expressed from the first non-wake (N1) epoch. No significant differences were observed between recovery sleep of DS and EF ($p < 0.05$, uncorrected).

	BL	DS	EF
Total sleep time [min]	415.1 \pm 83.2	406.7 \pm 81.0	423.8 \pm 53.1
N1 time [min]	23.2 \pm 15.7	13.9 \pm 9.1	16.9 \pm 14.3
N2 time [min]	214.9 \pm 59.4	176.0 \pm 46.9	175.8 \pm 45.5
N3 time [min]	83.5 \pm 31.7	116.9 \pm 32.4 ^b	120.1 \pm 18.1^a
REM time [min]	93.5 \pm 32.1	99.9 \pm 46.4	110.9 \pm 42.6
N1%	5.8 \pm 4.4	3.7 \pm 3.0	4.2 \pm 3.8
N2%	51.7 \pm 9.1	43.3 \pm 8.3 ^b	41.0 \pm 8.3^a
N3%	20.5 \pm 6.9	29.2 \pm 7.3^a	28.8 \pm 5.8^a
REM %	22.0 \pm 6.3	23.8 \pm 7.6	26.0 \pm 8.0
N2 latency [min]	4.4 \pm 2.9	13.6 \pm 17.5	12.1 \pm 27.6
N3 latency [min]	22.5 \pm 17.2	26.1 \pm 18.6	32.6 \pm 26.3
REM latency [min]	85.7 \pm 30.2	26.3 \pm 28.0^a	21.3 \pm 25.6^a

^a Significant differences between DS/EF and BL ($p < 0.05$, unpaired t -tests, Bonferroni–Holm correction for multiple comparisons).

^b $p < 0.05$ uncorrected.

observed that cortical MD declined even further after 24 h relative to 12 h of practice, and that all changes were reverted by ~ 7 h of sleep.

Recent studies in humans and rats had shown a link between short-term learning, such as practicing a visuo-spatial task for less than 2 h, and decreases of MD (Hofstetter et al., 2013; Sagi et al., 2012). These studies also demonstrated that MD changes are due to learning rather than movement and general activity (Hofstetter et al., 2013; Sagi et al., 2012). Other studies had examined MD changes in the context of chronic sleep disorders and found heterogeneous local changes, including an increase in the upper brainstem with REM sleep behavior disorder (Scherfler et al., 2011), an increase in hypothalamus and frontal cortex with narcolepsy–cataplexy (Scherfler et al., 2012), and a decrease in some cortical and subcortical areas with obstructive sleep apnea, an effect that reverts after treatment (refs in Castronovo et al., 2014).

MD reflects tissue density, and thus its decrease in GM may result from several not mutually exclusive factors including increase in synapse size, cell swelling, changes in extracellular space, and/or increase in glial cell volume. Although synaptic strength, which is correlated with size (Meyer et al., 2014), is known to increase with wake and decrease with sleep (Tononi and Cirelli, 2014), synapses contribute little to the overall GM volume, making it unlikely that rapid changes in MD after training and extended wake can be accounted for primarily by synaptic changes per se. More likely candidates are microstructural changes that are triggered by and/or are associated with synaptic activity and plasticity, such as variations in the ratio between intra- and extracellular volumes and astrocytic changes (Assaf and Pasternak, 2008). For instance, neuronal activity leads to a decrease in extracellular space (Ransom et al., 1985), and learning, sustained neuronal activity, or induction of long-term potentiation result in astrocytic hypertrophy and increased astrocytic coverage of synaptic processes (Anderson et al., 1994; Bernardinelli et al., 2014; Genoud et al., 2006; Jones and Greenough, 1996; Wenzel et al., 1991). Using serial-block-face electron microscopy, we recently found that after extended wake peripheral astrocytic processes in mouse frontal cortex move closer to the synaptic

cleft, expand, and increase their surface to volume ratio (Bellesi et al., 2015). These changes likely enhance the housekeeping functions of astrocytes and promote glutamate clearance from the cleft. At the same time, however, since the neuropil is filled with astrocytic processes, their wake-related “expansion” may impair the diffusion of water and other small molecules, potentially accounting for changes in MD.

In our paradigm, we did not detect a significant global change in MD after sleep following baseline wake. A previous study using voxel-based analyses found large local increases in brain diffusivity in the morning relative to the evening in the absence of training (4.4–5.6% increase in apparent diffusion coefficient) (Jiang et al., 2014). Voxel-based analysis may permit a more powerful detection of localized MD changes, but it can also lead to spurious results if partial volume effects or within-subject and across-subjects alignment issues are not adequately controlled for. Our analytical approaches were selected to minimize these problems.

Acute sleep deprivation and cortical thickness

While MD changes occur early during wake with practice and are sensitive to training per se, no significant global or regional changes in cortical thickness were found after 12 h of training relative to baseline (after wake and/or after sleep), although there was a tendency towards an increase. The subsequent change from 12 h to 24 h of practice was significantly different from the variation in the normal sleep night (W^B-S^B), although we detected no significant differences between individual time points. Moreover, recovery sleep was associated with a significant cortical thickness increase. Changes in cortical thickness have been observed mainly during development, aging, or in response to chronic manipulations. A case in point is adolescence, during which cortical GM shows a progressive thinning, a phenomenon that has been linked to improvement in cognitive functions and may result from selective pruning of inefficient synaptic connections and increases in myelination (Schnack et al., 2014). Changes in cortical thickness have also been reported in chronic primary insomnia, but results are inconsistent (Dang-Vu, 2013). Finally, increases in GM thickness have been described after weeks or months of training (May, 2011), but not after short training.

In our case, cortical thickness trended in opposite directions after 12 h (increase) and 24 h (decrease) of practice. This result suggests that sleep loss per se may have a more prominent influence on this parameter than synaptic plasticity, but more experiments are needed to test this hypothesis.

Volumetric changes potentially reflect an altered circulation of interstitial fluids

Training and extended wake also led to increased volume in subcortical GM and WM and decreased volume in the ventricles, and all changes were reverted by recovery sleep. These findings may be related to the recently discovered role of sleep in modulating circulation of interstitial fluids (ISF) (Xie et al., 2013). In fact, CSF is known to interchange with the brain ISF, and their combined movement allows the clearance of solutes from the brain (Brinker et al., 2014; Iliff et al., 2012). CSF production as measured by phase-contrast MRI also shows a strong circadian pattern in humans, being lower during the day and peaking at night (Nilsson et al., 1992). Importantly, the CSF-ISF movement is favored by sleep and impaired by wake (Xie et al., 2013) and sleep deprivation (Plog et al., 2015). Given these premises, sleep deprivation at night may have significantly reduced CSF-ISF movement and CSF production, impairing the clearance of brain metabolites and causing a reduction in ventricular volume and a compensatory expansion, or a swelling, of nearby structures such as subcortical GM and WM. Recovery sleep would revert described changes by promoting CSF-ISF movement and CSF production.

Importantly, all described structural modifications do not simply reflect alterations determining global, uniform variations in brain size, as indicated by the absence of significant changes in total brain volume throughout the experiments. The apparent contrast with recent work suggesting the existence of brain diurnal volumetric fluctuations (Nakamura et al., 2015) may depend on several differences in experimental design, since we only included healthy young adult volunteers who were not assuming any medications, and who were kept in homogeneous and controlled (although “extreme”) experimental conditions. On the other hand, we cannot exclude that modest global volumetric variations could have remained undetected in our dataset due to our smaller sample size.

Relationship between structural changes and sleep or behavioral parameters

Structural changes associated with recovery sleep did not correlate with any single sleep parameter that we tested. This may suggest that sleep as a whole, or at least several sleep features together, contributed to the effect. Other factors however, cannot be ruled out, including a “ceiling effect” caused by the high efficiency of sleep in all subjects, most likely due to the combination of 24 h of continuous wake with the intense task practice.

Previous studies suggested that specific structural measures taken at rest may reflect the degree of individual cognitive vulnerability to sleep deprivation (Cui et al., 2015; Rocklage et al., 2009). Here, we found that structural changes occurring during extended wake can be used to predict individual vigilance levels, and may partially contribute to predict variation in specific behavioral measures related to performance in an impulse control test. Thus, our results confirm and support the existence of a link between microstructural alterations and cognitive performance, and suggest that some individuals may be more resilient to sleep loss because of an optimal initial “structural reserve” that makes them less vulnerable to the microstructural alterations occurring during extended wake. Overall, they suggest that sleep’s ability to counteract performance deficits is linked to its effects on the brain microstructure.

Methodological considerations

Our study has several limitations. First, MRI-based approaches can only provide indirect measures of microstructural changes occurring within the human brain. Consequently, hypotheses regarding the link between variations in MRI-based parameters and the underlying biological mechanisms will require verification in different experimental models. On the other hand, several previous studies support the reliability of MRI-derived indices for studying a variety of physiological processes, including experience-dependent learning (Johansen-Berg et al., 2012; Zatorre et al., 2012), thus providing a solid foundation for the interpretation of present findings.

MRI-based parameters, and in particular measures related to water diffusivity, can be influenced by many confounds, including the partial volume effect (Alexander et al., 2001). While there are currently no commonly accepted gold-standard methods to address this problem, here we applied several strategies to minimize the potential impact of artifacts related to the partial volume effect, including the exclusion of voxels with a high probability of CSF contamination, and the adjustment of statistical models based on volumetric changes (associated to relative variations of the GM-CSF boundary). The obtained results suggest that changes in CSF contamination cannot account, alone, for the observed variations in cortical MD. This conclusion is further supported by the observation, during prolonged wakefulness, of parallel changes in cortical MD and thickness, and by the detected relationship between changes in cortical MD and in vigilance level. We acknowledge, however, that future studies using additional methods to limit the partial volume effect are required to provide an independent validation of the present results.

Another potential issue is the occurrence of wake–sleep transitions during the acquisition of DWI-scans, especially because deep sleep is known to be associated with relative changes in participants' movements and in brain temperature (Franken et al., 1992; Ogilvie, 2001), which, in turn, may affect diffusivity measures. The most accurate approach to control for this confound is MRI-EEG co-registration. However, this solution was not appropriate in the context of the present study for at least two reasons: i) the complexity of the experimental setup, together with technical and temporal constraints, limited the possibility to prepare a MRI-compatible EEG registration in our samples; ii) the EEG-net is known to affect the MRI signal, potentially leading to alterations of the estimation of structural parameters (Klein et al., 2015; Luo and Glover, 2012). On the other hand, several observations suggest that changes in behavioral state (wake/sleep) cannot account for our findings. First, the inclusion of head motion as a nuisance variable in statistical analyses had no relevant impact on our results. Moreover, the influence of a general decrease in brain temperature appears implausible, because relative changes in MD occurred with different timings in different structures. Of note, the initial MD decrease in cortical GM was also observed at a time-of-day typically associated with the circadian minimum in sleep pressure (Lavie, 1986). In summary, several empirical and theoretical considerations indicate that the possible occurrence of a transition to sleep in some of the subjects would not be sufficient to account for our results, leaving the change in tissue microstructure as the most plausible explanation.

Acknowledgments

Supported by NIH (grant R01MH099231 to CC and GT), McDonnell Foundation (to MFG and GT), the Waisman Brain Imaging Core (supported by NIH P30 HD003352), the Swiss National Foundation and the Swiss Foundation for Medical–Biological Grants (Grants 139778 and 145763 to FS), and Fondazione Cassa di Risparmio di Lucca (Lucca, Italy). The authors thank Juan Benzo, Ching-Sui Hung, Jeffrey Guokas and Corinna Zennig for help with data collection and technical assistance.

Conflict of interest

G. Tononi is involved in a research study in humans supported by Philips Respironics. This study is not related to the work presented in the current manuscript. R. Benca has served as a consultant to Merck and Jazz, and receives research funding from Merck. A. Alexander is part owner of insert MRI, Inc. This company is not related to the work and did not sponsor this research. The other authors have indicated no financial conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.01.020>.

References

Alexander, A.L., Hasan, K.M., Lazar, M., Tsuruda, J.S., Parker, D.L., 2001. Analysis of partial volume effects in diffusion-tensor MRI. *Magn. Reson. Med.* 45, 770–780.

Anderson, B.J., Li, X., Alcantara, A.A., Isaacs, K.R., Black, J.E., Greenough, W.T., 1994. Glial hypertrophy is associated with synaptogenesis following motor-skill learning, but not with angiogenesis following exercise. *Glia* 11, 73–80.

Assaf, Y., Pasternak, O., 2008. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *J. Mol. Neurosci.* 34, 51–61.

Ball, G., Srinivasan, L., Aljabar, P., Counsell, S.J., Durighel, G., Hajnal, J.V., Rutherford, M.A., Edwards, A.D., 2013. Development of cortical microstructure in the preterm human brain. *Proc. Natl. Acad. Sci.* 110, 9541–9546.

Basner, M., Dinges, D.F., 2011. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep* 34, 581–591.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.

Bellesi, M., de Vivo, L., Tononi, G., Cirelli, C., 2015. Effects of sleep and wake on astrocytes: clues from molecular and ultrastructural studies. *BMC Biol.* 13, 66.

Bernardi, G., Siclari, F., Yu, X., Zennig, C., Bellesi, M., Ricciardi, E., Cirelli, C., Ghilardi, M.F., Pietrini, P., Tononi, G., 2015. Neural and behavioral correlates of extended training during sleep deprivation in humans: evidence for local, task-specific effects. *J. Neurosci.* 35, 4487–4500.

Bernardinelli, Y., Randall, J., Janett, E., Nikonenko, I., Konig, S., Jones, E.V., Flores, C.E., Murai, K.K., Bochet, C.G., Holtmaat, A., Muller, D., 2014. Activity-dependent structural plasticity of perisynaptic astrocytic domains promotes excitatory synapse stability. *Curr. Biol.* 24, 1679–1688.

Brinker, T., Stopa, E., Morrison, J., Klinge, P., 2014. A new look at cerebrospinal fluid circulation. *Fluids Barriers CNS* 11, 10.

Bushey, D., Tononi, G., Cirelli, C., 2011. Sleep and synaptic homeostasis: structural evidence in *Drosophila*. *Science* 332, 1576–1581.

Castro-Alamancos, V., Scifo, P., Castellano, A., Aloia, M.S., Iadanza, A., Marelli, S., Cappa, S.F., Strambi, L.F., Falini, A., 2014. White matter integrity in obstructive sleep apnea before and after treatment. *Sleep* 37, 1465–1475.

Cox, R.W., 2012. AFNI: what a long strange trip it's been. *NeuroImage* 62, 743–747.

Cui, J., Tkachenko, O., Gogel, H., Kipman, M., Preer, L.A., Weber, M., Divatia, S.C., Demers, L.A., Olson, E.A., Buchholz, J.L., Bark, J.S., Rosso, I.M., Rauch, S.L., Killgore, W.D., 2015. Microstructure of frontoparietal connections predicts individual resistance to sleep deprivation. *NeuroImage* 106, 123–133.

Dang-Vu, T.T., 2013. Structural brain modifications in primary insomnia: myth or reality? *Sleep* 36, 965–966.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980.

Dijk, D.J., Czeisler, C.A., 1995. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J. Neurosci.* 15, 3526–3538.

Donlea, J.M., Ramanan, N., Shaw, P.J., 2009. Use-dependent plasticity in clock neurons regulates sleep need in *Drosophila*. *Science* 324, 105–108.

Donlea, J.M., Thimgan, M.S., Suzuki, Y., Gottschalk, L., Shaw, P.J., 2011. Inducing sleep by remote control facilitates memory consolidation in *Drosophila*. *Science* 332, 1571–1576.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1997. Support vector regression machines. *Advances in neural information processing systems* 9, pp. 155–161.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.

Franken, P., Tobler, I., Borbély, A.A., 1992. Cortical temperature and EEG slow-wave activity in the rat: analysis of vigilance state related changes. *Pflügers Archiv. Eur. J. Phys.* 420, 500–507.

Ganguly-Fitzgerald, I., Donlea, J., Shaw, P.J., 2006. Waking experience affects sleep need in *Drosophila*. *Science* 313, 1775–1781.

Genoud, C., Quairiaux, C., Steiner, P., Hirling, H., Welker, E., Knott, G.W., 2006. Plasticity of astrocytic coverage and glutamate transporter expression in adult mouse cortex. *PLoS Biol.* 4, e343.

Gholipour, A., Kehtarnavaz, N., Gopinath, K., Briggs, R., Devous, M., Haley, R., 2006. Distortion correction via non-rigid registration of functional to anatomical magnetic resonance brain images. *Image Processing, 2006 IEEE International Conference on. IEEE*, pp. 1181–1184.

Hofstetter, S., Tavor, I., Tzur Moryosef, S., Assaf, Y., 2013. Short-term learning induces white matter plasticity in the fornix. *J. Neurosci.* 33, 12844–12850.

Huber, R., Ghilardi, M.F., Massimini, M., Tononi, G., 2004. Local sleep and learning. *Nature* 430, 78–81.

Huber, R., Tononi, G., Cirelli, C., 2007. Exploratory behavior, cortical BDNF expression, and sleep homeostasis. *Sleep* 30, 129–139.

Hung, C.S., Sarasso, S., Ferrarelli, F., Riedner, B., Ghilardi, M.F., Cirelli, C., Tononi, G., 2013. Local, experience-dependent changes in the wake EEG after prolonged wakefulness. *Sleep* 36, 59–72.

Iliff, J.J., Wang, M., Liao, Y., Plogg, B.A., Peng, W., Gundersen, G.A., Benveniste, H., Vates, G.E., Deane, R., Goldman, S.A., Nagelhus, E.A., Nedergaard, M., 2012. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid beta. *Sci. Transl. Med.* 4, 147ra111.

Jenkinson, M., 1999. Measuring transformation error by RMS deviation. *Rep. TR99MJ1, Oxford Center for Functional Magnetic Resonance Imaging of the Brain (FMRIB)*.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

Jiang, C., Zhang, L., Zou, C., Long, X., Liu, X., Zheng, H., Liao, W., Diao, Y., 2014. Diurnal microstructural variations in healthy adult brain revealed by diffusion tensor imaging. *PLoS One* 9, e84822.

Johansen-Berg, H., Baptista, C.S., Thomas, A.G., 2012. Human structural plasticity at record speed. *Neuron* 73, 1058–1060.

Jones, T.A., Greenough, W.T., 1996. Ultrastructural evidence for increased contact between astrocytes and synapses in rats reared in a complex environment. *Neurobiol. Learn. Mem.* 65, 48–56.

Klein, C., Hänggi, J., Luechinger, R., Jäncke, L., 2015. MRI with and without a high-density EEG cap—what makes the difference? *NeuroImage* 106, 189–197.

Lavie, P., 1986. Ultraslow sleep-waking schedule. III. 'Gates' and 'forbidden zones' for sleep. *Electroencephalogr. Clin. Neurophysiol.* 63, 414–425.

Luo, Q., Glover, G.H., 2012. Influence of dense-array EEG cap on fMRI signal. *Magn. Reson. Med.* 68, 807–815.

- Maret, S., Faraguna, U., Nelson, A., Cirelli, C., Tononi, G., 2011. Sleep and wake modulate spine turnover in the adolescent mouse cortex. *Nat. Neurosci.* 14, 1418–1420.
- May, A., 2011. Experience-dependent structural plasticity in the adult human brain. *Trends Cogn. Sci.* 15, 475–482.
- Meyer, D., Bonhoeffer, T., Scheuss, V., 2014. Balance and stability of synaptic structures during synaptic plasticity. *Neuron* 82, 430–443.
- Nakamura, K., Brown, R.A., Narayanan, S., Collins, D.L., Arnold, D.L., Alzheimer's Disease Neuroimaging, I., 2015. Diurnal fluctuations in brain volume: statistical analyses of MRI from large populations. *NeuroImage* 118, 126–132.
- Nilsson, C., Stahlberg, F., Thomsen, C., Henriksen, O., Herning, M., Owman, C., 1992. Circadian variation in human cerebrospinal fluid production measured by magnetic resonance imaging. *Am. J. Phys.* 262, R20–R24.
- Ogilvie, R.D., 2001. The process of falling asleep. *Sleep Med. Rev.* 5, 247–270.
- Ohayon, M.M., Carskadon, M.A., Guilleminault, C., Vitiello, M.V., 2004. 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.
- Plog, B.A., Dashnaw, M.L., Hitomi, E., Peng, W., Liao, Y., Lou, N., Deane, R., Nedergaard, M., 2015. Biomarkers of traumatic injury are transported from brain to blood via the glymphatic system. *J. Neurosci.* 35, 518–526.
- Ransom, B.R., Yamate, C.L., Connors, B.W., 1985. Activity-dependent shrinkage of extracellular space in rat optic nerve: a developmental study. *J. Neurosci.* 5, 532–535.
- Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. *NeuroImage* 53, 1181–1196.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61, 1402–1418.
- Rocklage, M., Williams, V., Pacheco, J., Schnyer, D.M., 2009. White matter differences predict cognitive vulnerability to sleep deprivation. *Sleep* 32, 1100–1103.
- Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., Assaf, Y., 2012. Learning in the fast lane: new insights into neuroplasticity. *Neuron* 73, 1195–1203.
- Scherfner, C., Frauscher, B., Schocke, M., Iranzo, A., Gschliesser, V., Seppi, K., Santamaria, J., Tolosa, E., Hogg, B., Poewe, W., Group, S., 2011. White and gray matter abnormalities in idiopathic rapid eye movement sleep behavior disorder: a diffusion-tensor imaging and voxel-based morphometry study. *Ann. Neurol.* 69, 400–407.
- Scherfner, C., Frauscher, B., Schocke, M., Nocker, M., Gschliesser, V., Ehrmann, L., Niederreiter, M., Esterhammer, R., Seppi, K., Brandauer, E., Poewe, W., Hogg, B., 2012. White and gray matter abnormalities in narcolepsy with cataplexy. *Sleep* 35, 345–351.
- Schnack, H.G., van Haren, N.E., Brouwer, R.M., Evans, A., Durston, S., Boomsma, D.I., Kahn, R.S., Hulshoff Pol, H.E., 2014. Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cereb. Cortex*.
- Siclari, F., Bernardi, G., Riedner, B.A., LaRocque, J.J., Benca, R.M., Tononi, G., 2014. Two distinct synchronization processes in the transition to sleep: a high-density electroencephalographic study. *Sleep* 37, 1621–1637.
- Silber, M.H., Ancoli-Israel, S., Bonnet, M.H., Chokroverty, S., Grigg-Damberger, M.M., Hirshkowitz, M., Kapen, S., Keenan, S.A., Kryger, M.H., Penzel, T., Pressman, M.R., Iber, C., 2007. The visual scoring of sleep in adults. *J. Clin. Sleep Med.* 3, 121–131.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 (Suppl. 1), S208–S219.
- Tavor, I., Hofstetter, S., Assaf, Y., 2013. Micro-structural assessment of short term plasticity dynamics. *NeuroImage* 81, 1–7.
- Tononi, G., Cirelli, C., 2014. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron* 81, 12–34.
- Vyazovskiy, V.V., Cirelli, C., Tononi, G., 2011. Electrophysiological correlates of sleep homeostasis in freely behaving rats. *Prog. Brain Res.* 193, 17–38.
- Wang, H.W., Lin, Y.C., Pai, T.W., Chang, H.T., 2011. Prediction of B-cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification. *J. Biomed. Biotechnol.* 2011, 432830.
- Wenzel, J., Lammert, G., Meyer, U., Krug, M., 1991. The influence of long-term potentiation on the spatial relationship between astrocyte processes and potentiated synapses in the dentate gyrus neuropil of rat brain. *Brain Res.* 560, 122–131.
- Winkler, A.M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P.T., Duggirala, R., Glahn, D.C., 2010. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage* 53, 1135–1146.
- Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., Takano, T., Deane, R., Nedergaard, M., 2013. Sleep drives metabolite clearance from the adult brain. *Science* 342, 373–377.
- Yang, G., Gan, W.B., 2012. Sleep contributes to dendritic spine formation and elimination in the developing mouse somatosensory cortex. *Dev. Neurobiol.* 72, 1391–1398.
- Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., Fischl, B., 2014. Spurious group differences due to head motion in a diffusion MRI study. *NeuroImage* 88, 79–90.
- Zatorre, R.J., Fields, R.D., Johansen-Berg, H., 2012. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15, 528–536.