**Text-e1. Supplemental information about data acquisition, data processing and statistical analysis.**

Data acquisition.

In all participants, 11 minutes of resting-state fMRI was acquired on the same 3T magnetic resonance scanner (Intera Achieva; Philips, Best, the Netherlands). Two hundred fifty multislice T2\*-weighted fMRIs were obtained with a gradient echo-planar sequence using slice orientation (37 slices; vowel size: 2 x 2 x 3.5 mm3); repetition time = 2.600 ms; echo time = 30 ms; flip angle = 90°; field of view = 240 mm). A high-resolution anatomical image, using 3-dimensional T1-weighted sequence (in-plane resolution 1 x 1 x 1 mm, 160 contiguous slices) was also acquired in the same session. Diffusion Tensor Imaging (DTI) parameters: B-value (S/mm2 ) = 1000, TR/TE (ms) = 7000/77, Bandwidth (Hz) = 2438, Matrix = 112 x 109, FOV = 22 x 22.

Data processing.

*Structural brain markers*

Gray matter morphometry was applied on standard MRI sequences, such as T1-weighted images. We specifically focused on voxel-based morphometry, which enables the investigation of voxel-wise differences in the local grey matter volume/topography without priori information about the location of these possible differences. Image processing was performed using FSL v5 (www.fmrib.ox.ac.uk/fsl/) and an in-house developed software in Matlab (version 6.5, The MathWorks). First, all the T1 images where brain extracted then segmented into images of gray matter (GM), white matter (WM), and cerebro-spinal fluid (CSF). GM images where then non-linearly registered to a gray matter ICBM-152 template, concatenated and averaged. Second, GM images of all participants were non-linearly registered to this template. Gray matter density (GMD) maps were calculated from individual GM registered data. As for DWI, a diffusion tensor imaging (DTI) model was fit at each voxel, generating fraction anisotropy (FA), mean diffusivity (MD) and radial diffusivity (RD) maps (FMRIB's Diffusion Toolbox) (19). The FA maps were then registered to brain-extracted whole-brain volumes from T1-weighted images using a full affine (correlation ratio cost function) alignment with nearest-neighbor resampling. The calculated transformation matrix was then applied to the MD maps with identical resampling options. The white matter (i.e. cingulum) ROI was taken from an atlas based on an fMRI guided DTI tractography between parietal and frontal targeted nodes (20) (posteromedial parietal cortex – PMC - and medial prefrontal cortex – mPFC -, respectively), which provides a group probability map, more conservative than the raw connection counts. The original cingulum ROI was divided into five sub-regions equally distributed along an anteroposterior axis (Figure 1). These sub-regions were used to compare the WM differences between the groups. To assess grey matter changes, we extracted the mean values of GMD and MD from 11 mPFC sub-regions and 3 PMC sub-regions, which were defined according to fMRI segmentation (see below). To assess white matter changes, we extracted the mean values of FA, MD and RD from 5 cingulum sub-regions.

*Functional brain markers*

Resting-state functional (rs-fMRI) data was pre-processed using Statistical Parametric Mapping (version SPM 12; http://www.fil.ion.ucl.ac.uk/spm/). The fMRI images were realigned, slice-time corrected, coregistered to each subject’s T1-weighted image and normalized to standard stereotaxic anatomical Montreal Neurological Institute (MNI) space. The rs-fMRI data was further analyzed using the CONN toolbox (v.16a; http://www.nitrc.org/projects/conn). We have also performed ART outlier and non-neuronal sources of noise were estimated and removed using the CompCor method integrated in the CONN toolbox. Principal components of the signals from the WM and the CSF voxels (using normalized T1 segmented masks), alongside the motion parameters (estimated during realignment) and between-scan motion outliers (ART toolbox), were removed with regression. Finally, a temporal band-pass filter was applied to the residual blood oxygen level–dependent (BOLD) time course in order to obtain a low-frequency range of 0.01 to 0.1 Hz. We targeted our analysis on three sub-regions of the PMC, and 11 sub-regions of the mPFC, as described in functional atlas (Figure 1). The analysis was done using each of these sub-regions separately, and using the total mean values, averaged separately for the PMC and the mPFC regions.

*Statistical analysis*

The mean values of GMD, MD, FA, FC were compared using a repeated-measures multivariate analysis of variance (MANOVA) with two factors (group, sub-region). We also performed a “whole-region” analyses averaging the values in the subregions for mPFC, PMC and the cingulum (text e-1). In case of significant results (group effect or group by sub-region interaction), a one-way ANOVA with group as factor was applied to each parameter and each structure. To complement these univariate analyses, we have performed discriminative and predictive analyses using machine learning techniques.

* Single markers and two-markers combination

First, we wanted to test the hypothesis that single MRI indexes in isolation could discriminate between healthy controls and patients in a coma, between TBI and anoxic patients as well as predict the clinical outcome measured 3-months after the MRI acquisition. Moreover, we were interested in testing the discriminative and predictive power of canonical anatomical regions of interest (ROIs) and more spatially specific ROIs. We extracted the functional connectivity between medial PFC and the PMC, the grey matter volume from the medial PFC and the PMC, and the MD, RD and FA from the PFC, the PMC and the cingulum. We also used more spatially specific ROIs that subdivide the structure of interest in functional/anatomical units to test the hypothesis that increased spatial specificity could improve the discriminative/predictive performance. As for outcome variables we used the following variables: coma patients vs healthy controls (binary), TBI vs anoxic patient (binary), recovered (minimally conscious state - MCS - encompassing both MCS (-) and MCS (+) which were defined according to the identification of command-following, intelligible verbalization or intentional communication abilities)(20-22) vs non-recovered (vegetative state / unresponsive wakefulness syndrome - VS/UWS) patients (binary) and CRS-R (continuous variable). We used logistic regression without penalty for binary variables and linear regression for continuous ones. We used the scikit-learn module in Python 3.3 for implementing discriminative/predictive pipeline: we used repeated 10-folds cross validation:. the sample was divided in 10 equally sized subsamples stratified by label/outcome, each subsample was used once as testing sample while the other 9 subsamples were used for fitting, this procedure was repeated 1000 times with different partitions of the sample. On the basis of these repetitions we calculated mean and 95% confidence interval of the area under the curve for binary outcomes and mean and 95% confidence interval of the correlation between predictions and actual values for the continuous outcome. Note that since there were no hyperparameters to optimize, the 10 folds cross-validation scheme is enough to ensure unbiased predictions.

* Predictive models with all indexes and regions

To confirm the discriminative and predictive power of the different markers/regions we also ran analyses using all markers and regions at the same time, while recording the performance of the models and the number of times a certain marker/region were selected. We combined supporting vector machine (SVM (linear support vector classification for binary variables and linear support vector regression for continuous variables, in the scikit-learn implementation, with default hyperparameters –*L1* penalty, C = 1 for classification; ε = 0.1, C = 1 for regression) with a univariate filter for features selection . SVM was chosen due to its relative independence from sample size. As for the univariate filter, we used an F-score filter to only retain the k features with the highest association with the outcome variables before fitting the model. The parameter k (i.e. the proportion of features to retain, set to be [.9, .8, .7, .6, .5, .4, .3, .2, .1]) was selected using a 10-folds cross-validation scheme nested into a bootstrap procedure. Note that the performance of the models was evaluated using out-of-bag methods (i.e. using the subjects that were not resampled in the bootstrap procedures as testing set at each iteration) to obtain unbiased prediction. We repeated the bootstrap procedure 1000 times and calculated the mean and 95% confidence interval of the performance of the model (balanced accuracy for the binary variables, correlation between the predictions and the actual value as well as absolute mean error for the continuous variables). The entire procedure was repeated shuffling the labels/outcomes of the subjects in order to obtain an approximation of the null distribution of performances (i.e. the distribution of performances when there is no association between features and label/outcome). To test the significance of the performance obtained with the original labels/outcome we counted the number of times a model fitted with shuffled labels reached an accuracy equal or higher than the median of the performance obtained with the true labels and divided this number by the total number of shuffling (i.e. 1000).

* Nested cross-validation model

To assess the stability of our results to different methods, we have repeated performed an additional analysis using all indexes and regions. In particular, we adopted a nested cross-validation method in place of the boot-strapping procedure. This means that we embedded an inner 10-fold cross validation for choosing the best parameter k for the F-score filter (as in the above method) within an outer 10-fold cross-validation loop used to obtain unbiased prediction. The nested-cross-validation was repeated 1000 times for each analysis. Below we report the performance of the different models.

When discriminating between controls and comatose patients, the nested-cross validation reached a mean accuracy of .85 (bootstrap analysis performance .83).

When discriminating aetiologies, the nested-cross validation reached a mean accuracy of .61 (bootstrap analysis performance .75). The performance of the nested cross-validation is lower than the that of the bootstrap model, but still above chance level.

When discriminating between dichotomous outcome (favourable/unfavourable), the nested-cross validation model reached a mean accuracy of XX (bootstrap analysis performance .81)

When predicting outcome as continuous CRS-R the nested cross validation model reached a mean correlation of .59 (bootstrap analysis performance .67).

Overall the results of the nested-cross validation procedures are congruent with those of the bootstrap procedure.